

IBM Content Collector
Version 3.0.0.2

Guide d'indexation



IBM Content Collector
Version 3.0.0.2

Guide d'indexation



Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 231.

Deuxième édition - Décembre 2012

Réf. US : SH12-6979-01

LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE.

Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. Les informations qui y sont fournies sont susceptibles d'être modifiées avant que les produits décrits ne deviennent eux-mêmes disponibles. En outre, il peut contenir des informations ou des références concernant certains produits, logiciels ou services non annoncés dans ce pays. Cela ne signifie cependant pas qu'ils y seront annoncés.

Pour plus de détails, pour toute demande d'ordre technique, ou pour obtenir des exemplaires de documents IBM, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial.

Vous pouvez également consulter les serveurs Internet suivants :

- <http://www.fr.ibm.com> (serveur IBM en France)
- <http://www.can.ibm.com> (serveur IBM au Canada)
- <http://www.ibm.com> (serveur IBM aux Etats-Unis)

*Compagnie IBM France
Direction Qualité
17, avenue de l'Europe
92275 Bois-Colombes Cedex*

© Copyright IBM France 2012. Tous droits réservés.

© **Copyright IBM Corporation 2008, 2012.**

Table des matières

Avis aux lecteurs canadiens v

ibm.com et ressources associées vii

Comment envoyer vos commentaires viii

Contacteur IBM viii

Partie 1. Indexation dans IBM

Content Manager 1

Indexation avec IBM Content Collector

Text Search Support 3

Nouvelles fonctionnalités d'indexation dans Content

Collector Text Search Support. 3

Indexeur de recherche de texte 6

Installation de Content Collector Text Search Support

pour IBM Content Manager 8

Installation de Content Collector Text Search

Support en mode interface graphique 8

Installation de Content Collector Text Search

Support en mode console 9

Formulaire de configuration de Net Search Extender 9

Principales étapes de l'indexation 10

Configuration d'un référentiel IBM Content

Manager 11

Activation d'un référentiel IBM Content Manager

à des fins de traitement par l'indexeur pour la

recherche de texte 15

Activation de types d'éléments qui n'ont pas été

créés dans Content Collector. 16

Création d'un fichier de configuration de

l'indexeur pour le type d'élément 20

Indexation des éléments d'un type d'élément 25

Vérification du code retour du processus de

l'indexeur 36

Vérification du statut des éléments indexés. 37

Indexation des types d'élément non créés sous

Content Collector 43

Indexation des éléments qui ont été archivés à

l'aide de IBM CommonStore. 43

Indexation des éléments dans un type d'élément

créé sous IBM Content Manager 45

Surveillance de l'indexation via IBM Content

Collector Text Search Support 46

Sélection des informations à surveiller 47

Fichiers journaux créés pendant le processus de

l'indexeur 48

Fichiers de trace créés lors de l'exécution des

outils de l'indexeur 50

Statistiques créées pendant la procédure de

l'indexeur 50

Informations de date/heure créées pendant

l'exécution du processus de l'indexeur 53

Opérations et paramètres avancés 54

Préparation à l'indexation 54

Fichier modèle 54

Utilisation de TIEFLAG lors de l'indexation dans

IBM Content Collector. 61

Variables d'environnement de l'indexeur de texte 63

Modification de la configuration par défaut de

l'indexeur 66

Contraintes lors de l'indexation en mode de

traitement par lots 88

Utilisation de plusieurs processus à unités

d'exécution multiples durant le traitement et

l'indexation de documents 89

Exécution de afuIndexer en mode continu 92

Traitement des courriers électroniques sans pièces

jointes 94

Traitement simultané de plusieurs types

d'élément 95

Traitement des pièces jointes imbriquées dans des

courriers électroniques et des documents 95

Exemple de script pour l'évaluation des codes

retour du processus de l'indexeur 100

Réindexation de documents 103

Opérations à effectuer en cas d'arrêt brutal du

processus de l'indexeur 105

Éléments ayant généré une erreur grave 106

Débogage du traitement d'un élément 108

Arrêt prématuré, ou à un moment spécifique,

d'un processus de l'indexeur IBM Content

Collector 111

Réorganisation d'index 112

Recherche de chaînes d'erreur 114

Validation d'index et opérations de réparation 116

Exécution de l'outil de contrôle de la

configuration 130

Recréation d'un index 134

Correction de la structure du fichier CSN dans

des courriers électroniques Lotus Domino MIME 140

Sauvegarde des fichiers d'index 145

Présentation de la table de base de données

pour les tables de tâches de l'indexeur 145

Suppression d'un type d'élément 146

Messages d'erreur de l'indexeur de recherche de

texte 146

Identification et résolution des problèmes de

l'indexeur de recherche de texte 150

Configuration Net Search Extender incorrecte 150

L'outil de l'indexeur ne parvient pas à traiter le

type d'élément 151

Impossible d'activer les types d'élément de

courrier électronique IBM CommonStore via

afuEnableItemType 151

Impossible de charger une bibliothèque DB2

native 152

Impossible de localiser le répertoire

d'installation DB2 153

Escalade de verrous DB2 153

Index introuvable après sa recréation 154

Net Search Extender s'arrête de manière inattendue à l'émission de COMMIT	154
Net Search Extender échoue lorsque le processus de l'indexeur est en cours d'exécution.	155
Recherche d'un document introuvable	155
Le processus de l'indexeur s'arrête car la fonction définie par l'utilisateur AFUFetch n'est pas accessible	157
Obtention du numéro de port DB2	158
Utilisation d'instructions SQL longues ou complexes	158
Utilisation d'une configuration JDBC incorrecte pour les processus de l'indexeur	159
Chargement impossible du pilote JDBC de type 2	159
Les éléments dont le contenu textuel est supérieur à 1 Mo ne sont pas indexés lors de l'utilisation de DB2 version 9.5.	160
Le processus de l'indexeur s'arrête car l'index est introuvable	160
Modifications du répertoire par défaut et des paramètres de nom de fichier de la configuration de la fonction définie par l'utilisateur AFUFetchFile	161
Erreur lors de l'accès aux fichiers XML dans la fonction définie par l'utilisateur AFUFetchFile	162
Erreur lors de l'accès aux fichiers XML dans la fonction définie par l'utilisateur AFUFetchFile si l'ID d'élément est constitué de points d'interrogation	163
Le processus de travail de l'indexeur s'arrête et indique une situation de mémoire saturée.	164

Partie 2. Indexation dans IBM FileNet P8	165
Indexation dans IBM FileNet P8	167

Nouvelles fonctions d'indexation de IBM FileNet P8	167
Indexation avec IBM Legacy Content Search Engine	168
Concepts d'indexation dans IBM Legacy Content Search Engine	168
Activation de l'indexation en texte intégral avec IBM Legacy Content Search Engine	169
Identification des incidents d'indexation lors de l'indexation avec IBM Legacy Content Search Engine	179
Réindexation de documents	181
Remplacement d'un index	182
Contrôle de l'indexation avec IBM Legacy Content Search Engine	183
Indexation avec IBM Content Search Services.	185
Concepts d'indexation dans IBM Content Search Services	185
Activation de l'indexation en texte intégral avec IBM Content Search Services	186
Etapes de traitement Content Search Services Support	213
Identification et résolution des incidents d'indexation lors de l'indexation avec IBM Content Search Services	214
Contrôle de l'indexation avec IBM Content Collector P8 Content Search Services Support	222
Identification et résolution des problèmes liés à Content Search Services Support	226

Partie 3. Annexes	229
Remarques	231
Index	235

Avis aux lecteurs canadiens

Le présent document a été traduit en France. Voici les principales différences et particularités dont vous devez tenir compte.

Illustrations

Les illustrations sont fournies à titre d'exemple. Certaines peuvent contenir des données propres à la France.

Terminologie

La terminologie des titres IBM peut différer d'un pays à l'autre. Reportez-vous au tableau ci-dessous, au besoin.

IBM France	IBM Canada
ingénieur commercial	représentant
agence commerciale	succursale
ingénieur technico-commercial	informaticien
inspecteur	technicien du matériel

Claviers

Les lettres sont disposées différemment : le clavier français est de type AZERTY, et le clavier français-canadien de type QWERTY.








OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

- les pages de codes 850 (multilingue) et 863 (français-canadien),
- le code pays 002,
- le code clavier CF.

Nomenclature

Les touches présentées dans le tableau d'équivalence suivant sont libellées différemment selon qu'il s'agit du clavier de la France, du clavier du Canada ou du clavier des États-Unis. Reportez-vous à ce tableau pour faire correspondre les touches françaises figurant dans le présent document aux touches de votre clavier.

France	Canada	Etats-Unis
 (Pos1)		Home
Fin	Fin	End
 (PgAr)		PgUp
 (PgAv)		PgDn
Inser	Inser	Ins
Suppr	Suppr	Del
Echap	Echap	Esc
Attn	Intrp	Break
Impr écran	ImpEc	PrtSc
Verr num	Num	Num Lock
Arrêt défil	Défil	Scroll Lock
 (Verr maj)	FixMaj	Caps Lock
AltGr	AltCar	Alt (à droite)

Brevets

Il est possible qu'IBM détienne des brevets ou qu'elle ait déposé des demandes de brevets portant sur certains sujets abordés dans ce document. Le fait qu'IBM vous fournisse le présent document ne signifie pas qu'elle vous accorde un permis d'utilisation de ces brevets. Vous pouvez envoyer, par écrit, vos demandes de renseignements relatives aux permis d'utilisation au directeur général des relations commerciales d'IBM, 3600 Steeles Avenue East, Markham, Ontario, L3R 9Z7.

Assistance téléphonique

Si vous avez besoin d'assistance ou si vous voulez commander du matériel, des logiciels et des publications IBM, contactez IBM direct au 1 800 465-1234.

ibm.com et ressources associées

Le support et la documentation produit sont disponibles sur [ibm.com](http://www.ibm.com).

Support et assistance

Le support produit est disponible sur le Web. Il vous suffit de cliquer sur Support à partir du site Web correspondant au produit.

IBM® Content Collector

<http://www-01.ibm.com/software/data/content-management/content-collector/>

Solution IBM Email Archive and eDiscovery

<http://pic.dhe.ibm.com/infocenter/email/v3r0m0/index.jsp>

IBM CommonStore for Exchange Server

<http://www.ibm.com/software/data/commonstore/exchange/>

IBM CommonStore for Lotus Domino

<http://www.ibm.com/software/data/commonstore/lotus/>

IBM Content Manager

<http://www.ibm.com/software/data/cm/cmgr/mp/>

IBM FileNet P8

<http://www.ibm.com/software/data/content-management/filenet-p8-platform/>

IBM Enterprise Records

<http://www.ibm.com/software/data/content-management/filenet-records-manager/>

IBM Records Manager

<http://www.ibm.com/software/data/cm/cmgr/rm/>

IBM WebSphere Application Server

<http://www.ibm.com/software/webservers/appserv/was/>

Lotus Notes et Domino

<http://www.ibm.com/software/lotus/notesanddomino/>

Centre de documentation

Vous pouvez afficher la documentation du produit IBM Content Collector dans un centre de documentation Eclipse. Consultez le centre de documentation à l'adresse <http://pic.dhe.ibm.com/infocenter/email/v3r0m0/index.jsp>.

Publications au format PDF

Vous pouvez afficher une version PDF du guide d'installation et de configuration d'IBM Content Collector à l'aide d'Adobe Acrobat Reader pour votre système d'exploitation. Ce guide est disponible dans le Centre de publications IBM. Si Acrobat Reader n'est pas installé, vous pouvez le télécharger à partir du site Web d'Adobe, à l'adresse <http://www.adobe.com>.

Comment envoyer vos commentaires

Vos commentaires nous aident à vous fournir un niveau d'informations précis et de qualité.

Envoyez vos commentaires en remplissant le formulaire en ligne disponible à l'adresse suivante : https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf.

Contacteur IBM

Pour prendre contact avec le centre de support IBM aux Etats-Unis ou au Canada, appelez le 1-800-IBM-SERV (1-800-426-7378).

Pour en savoir plus sur les options de service disponibles, appelez l'un des numéros suivants :

- Aux Etats-Unis : 1-888-426-4343
- Au Canada : 1-800-465-9600

Pour plus d'informations sur la manière de contacter IBM, reportez-vous au site Web de contact IBM à l'adresse <http://www.ibm.com/contact/us/>.

Partie 1. Indexation dans IBM Content Manager

Indexation avec IBM Content Collector Text Search Support

IBM Content Collector Text Search Support est un composant logiciel d'indexation indépendant qui permet de traiter les documents archivés dans les répertoires IBM Content Manager afin d'activer la fonction de recherche dans ces documents.

Le traitement de documents et les opérations d'indexation dans IBM Content Collector Text Search Support sont désignés sous le terme générique d'*indexeur de recherche de texte*.

L'indexeur de recherche de texte remplace et combine les modules de l'exit utilisateur de recherche de texte et de l'indexeur rapide utilisés dans les versions d'IBM Content Collector antérieures à la version 2.1.1 et dans IBM CommonStore. Il exécute en parallèle les processus d'extraction de documents et de conversion de contenu, ce qui augmente sensiblement la vitesse d'indexation de texte intégral.

IBM Content Collector Text Search Support peut être installé, configuré et utilisé sans avoir à installer IBM Content Collector.

Nouvelles fonctionnalités d'indexation dans Content Collector Text Search Support

L'indexation dans IBM Content Manager à l'aide d'IBM Content Collector Text Search Support propose les nouvelles fonctions suivantes.

Fonctions d'indexation ajoutées dans IBM Content Collector V2.2.0.2

Traitement supplémentaire dans afuEnableItemType pour prendre en charge la reconnaissance de la valeur TIEFLAG dans les types d'éléments de ressource IBM CommonStore

Traitement supplémentaire dans afuEnableItemType pour prendre en charge la reconnaissance de la valeur TIEFLAG dans les types d'éléments de ressource IBM CommonStore. Lorsque vous exécutez l'outil de type d'élément appelé afuEnableItemType sur les types d'éléments de ressource IBM CommonStore, la table des tâches terminées est automatiquement renseignée avec tous les éléments qui ont déjà été indexés. La valeur IDXRC qui est attribuée à ces éléments correspond à la valeur TIEFLAG qui définit les composants d'élément textuels.

Nouvelle option de configuration ajoutée au processus de l'indexeur qui reconnaît la valeur TIEFLAG utilisée dans IBM CommonStore

Une nouvelle option de configuration de l'indexeur a été ajoutée au processus de l'indexeur qui remplit la colonne TIEFLAG dans la table des composants du type d'élément dans IBM Content Manager.

Modifications apportées à l'argument de ligne de commande -reindexwarnings utilisé avec afuIndexer

L'argument -reindexwarnings de l'outil afuIndexer ignore les éléments qui ont été indexés par l'indexeur rapide ou l'indexeur standard IBM Content Manager avec l'exit utilisateur IBM Text Search et ont une valeur IDXRC importée entre 10 et 19.

Nouvelle option de ligne de commande pour afuEnableItemType pour modifier la taille de la mémoire tampon définie par l'utilisateur

Une nouvelle option de ligne de commande appelée `-udftransferbuffersize` a été ajoutée à `afuEnableItemType` afin de pouvoir l'utiliser si vous devez spécifier une taille différente de la mémoire tampon qu'`AFUFetchFile` utilise pour charger les fichiers XML temporaires qui seront accessibles par `Net Search Extender`.

Nouvelle option de configuration de l'indexeur pour gérer les éléments présentant des erreurs serveur

En définissant l'option de configuration `IdxProcessSevereErrors` sur 1, les éléments susceptibles d'avoir provoqué l'arrêt inattendu du processus de l'indexeur ne sont pas déplacés vers la table des tâches terminées avec le code `IDXRC 200`, mais seront à nouveau traités sans les pièces jointes intégrées à la prochaine exécution d'`afuIndexer`.

Exécution de la validation de l'index et des opérations de réparation

L'indexeur de l'outil d'index de recherche de texte appelé `afuIndexTool` fournit des opérations d'index utiles qui peuvent être appliquées à un index afin de rechercher les incohérences ou mettre à jour les tables de la base de données d'index afin de prendre en charge la fonction `TIEFLAG IBM CommonStore`.

Réindexation de documents de courrier électronique Lotus Domino archivés qui n'ont pas été indexés correctement

Pour identifier les documents qui peuvent être affectés et réindexés, utilisez l'outil appelé `afuRepairCSN`. Cet outil doit être exécuté sur tous les types d'élément contenant des documents de courrier électronique Lotus Domino qui ont été archivés à l'aide d'une version `IBM Content Collector Server` comprise entre 2.1.1.1 LA006 et 2.1.1.3 LA006.

Recherche de courriers électroniques chiffrés dans l'index

Une chaîne de message de notification de recherche d'avertissement appelée `IcmFceWarning: IcmDocIsEncrypted` est indexée lorsque des courriers électroniques chiffrés (`Exchange`, `Domino` et `SMTP/MIME`) sont traités par l'indexeur. Le contenu des courriers électroniques chiffrés ne peut pas être indexé. À l'aide de la chaîne du message de recherche, vous pouvez rechercher tous les courriers électroniques chiffrés, les déchiffrer et réindexer le courrier électronique si vous souhaitez indexer leur contenu.

Indexation des fichiers MSG intégrés

Le contenu textuel des fichiers `MSG` intégrés, y compris des messages `MSG` intégrés de manière récursive, peut désormais être indexé. Cela signifie que la chaîne de notification `IcmFceWarning: IcmUnhandledEmbeddedMsg` n'est plus utilisée et ne peut pas être recherchée. Une fois que vous avez appliqué le groupe de correctifs, recherchez dans l'index tous les éléments ayant la chaîne de notification `IcmFceWarning: IcmUnhandledEmbeddedMsg` et réindexez-les.

Indexation des types d'élément de document IBM CommonStore Content Manager

Vous pouvez indexer les éléments contenus dans les types d'élément `IBM CommonStore` pour le modèle de document `IBM Content Manager GENERIC_MULTIDOC` et `GENERIC_MULTIPART` et le type d'archivage entier et pièce jointe. Avant de pouvoir utiliser un type d'élément de document `CommonStore` dans `IBM Content Collector`, ce type d'élément doit être activé pour l'utilisation dans `Content Collector`.

Fonctions d'indexation ajoutées dans IBM Content Collector V3.0

Nouveaux arguments de ligne de commande de l'indexeur pour la réindexation des éléments indexés à l'aide de chaînes de recherche

Réindexe uniquement les documents qui ont été indexés pendant l'exécution d'une procédure d'indexation précédente et pour lesquels la chaîne indiquée a été indexée au document.

Nouveau mode d'indexation qui traite les éléments présentant des erreurs graves uniquement

Vous pouvez exécuter `afuIndexer` dans un mode spécial dans lequel seuls les éléments qui ont été traités dans une indexation précédente et ont généré une erreur grave sont traités à l'aide des paramètres de configuration optimisés pour la gestion des situations d'erreur, et non pas pour la performance et un débit élevé.

Prise en charge supplémentaire pour les types d'élément contenant les documents IBM Connections

L'indexeur de recherche de texte prend en charge le traitement et l'indexation des documents IBM Connections.

Prise en charge des types d'élément Microsoft SharePoint créés à l'aide du modèle de données avec des pièces jointes intégrées

L'indexeur de recherche de texte prend en charge le traitement et l'indexation des documents Microsoft SharePoint dans les types d'élément créés à l'aide d'un nouveau modèle de données prenant en charge la gestion des pièces jointes intégrées.

La validation d'index s'exécute dans un mode parallèle

Pour optimiser les performances, l'outil de validation d'index `afuIndexTool` exécute des opérations de validation d'index en parallèle.

Fonctions d'indexation ajoutées à IBM Content Collector V3.0.0.1

Utilisation de plusieurs processus à unité d'exécution simple durant le traitement et l'indexation de documents

Cette nouvelle fonction d'indexation permute le traitement de l'indexeur entre plusieurs unités d'exécution regroupées sous un processus de traitement unique et plusieurs processus de traitement comportant chacun une unité d'exécution unique.

Sur le système d'exploitation AIX, la création de plusieurs processus de traitement comportant chacun une unité d'exécution unique est l'option par défaut. Sur tous les autres systèmes d'exploitation, l'option par défaut consiste à créer plusieurs unités d'exécution au sein d'un seul processus de traitement.

Lorsque vous spécifiez l'option `afuIndexer -usemultipleprocessing`, les ressources de gestion des documents ne sont pas partagées par plusieurs documents sous le même processus. Au lieu de cela, chaque processus de traitement concentre toutes les ressources disponibles, telles que la mémoire, sur le traitement d'un document à un moment donné. De même, les problèmes qui se produisent dans le contexte d'un traitement à unités d'exécution multiples et dans `Filters Oracle Outside In Technology` peuvent être évités.

Nouvel attribut de courrier électronique Lotus Notes

L'attribut Lotus Notes suivant a été ajouté aux modèles de configuration de messagerie :

```
$INetPrincipal = /icc_document/icc_email/icc_from
```

Dans les versions antérieures d'IBM Content Collector, seul le principal de l'e-mail était indexé sous forme canonique (Principal = /icc_document/icc_email/icc_from). L'adresse électronique ou Internet du propriétaire principal, par exemple jdoe@company.com, n'était pas été indexée du fait que l'attribut était manquant dans les fichiers de configuration de la messagerie électronique.

Désormais, le propriétaire de l'e-mail, l'adresse e-mail du propriétaire et l'adresse électronique de la personne déléguée qui a effectivement envoyé le courrier électronique sont indexées sous icc_from.

Pour pouvoir rechercher l'adresse SMTP du propriétaire de boîte aux lettres dans le courrier électronique de délégation, l'index recherché doit avoir été créé en utilisant l'attribut \$INetPrincipal. Seuls les index créés via IBM Content Collector V3.0.0.1 et versions ultérieures contiennent cet attribut. Les index créés avant IBM Content Collector V3.0.0.1 ne contiennent pas cette information dans l'index. Si vous souhaitez rechercher l'adresse SMTP de la boîte aux lettres du propriétaire dans un courrier électronique de délégation contenu dans des index anciens, vous devez ajouter cette information à la configuration du type d'élément Content Manager existante, puis réindexer tous les documents de messagerie.

Fonctions d'indexation ajoutées à IBM Content Collector V3.0.0.2

Recherche de courrier électronique avec un corps vide dans l'index

Une chaîne de message de notification de recherche d'avertissement appelée IcmFceWarning: IcmEmptyBody est indexée lorsque des courriers électroniques contiennent un corps vide. Un corps de message vide implique que le corps ne contient aucun caractère, ou seulement des espaces. Vous pouvez, avec la chaîne du message de recherche, trouver tous les messages électroniques dont les sections Body sont vides.

Indexeur de recherche de texte

L'indexeur de recherche de texte IBM Content Collector prépare tous les documents à l'indexation et configure le moteur d'indexation, Net Search Extender, pour exécuter l'indexation dans IBM Content Manager.

Tous les documents archivés dans IBM Content Manager via IBM Content Collector sont stockés au sein de collections de documents appelées "types d'élément" dont le texte doit pouvoir faire l'objet de recherches. Tous les documents de ces types d'élément sont marqués pour indexation dès qu'ils sont stockés dans le type d'élément.

Avant de commencer le traitement d'un type d'élément, les tâches suivantes sont effectuées pour configurer le système et prendre en charge l'indexeur de recherche de texte. Ces tâches sont effectuées pendant la configuration initiale, la première fois que vous installez IBM Content Collector, ou après l'installation avec les outils de configuration d'IBM Content Collector. Ces tâches comprennent :

- Création de types d'élément ou reconfiguration de types d'élément existants pour activer le traitement à l'aide de l'indexeur de recherche de texte IBM Content Collector
- Ajustement de l'indexeur pour la configuration de la recherche de texte utilisée par le type d'élément. Cette opération inclut la configuration du moteur d'indexation appelé IBM Net Search Extender.

- Préparation du type d'élément pour l'indexeur de recherche de texte (auparavant processus FetchContentSetup exécuté séparément). Cette opération est maintenant effectuée automatiquement et inclut les procédures suivantes pour chaque type d'élément :
 - Création d'une table de base de données pour les tâches d'indexation ouvertes
 - | La table des tâches ouvertes contient les éléments qui n'ont pas été traités et
 - | indexés. La table est automatiquement renseignée lorsque des articles sont
 - | archivés ou réindexés.
 - Création d'une table de base de données pour les tâches d'indexation terminées
 - | La table des tâches terminées contient à la fois les éléments qui ont été traités
 - | et indexés avec succès et les éléments qui n'ont pas pu être traités, quelle
 - | qu'en soit la raison. Chaque élément de cette table a un code de retour
 - | d'indexation (IDXRC).
 - | La table est remplie par l'opération d'indexation et constitue le point de
 - | départ de l'opération de réindexation.
 - Création de déclencheurs qui créent des lignes dans la table des tâches ouvertes pour chaque élément archivé

Une fois les tâches de configuration exécutées et si le type d'élément contient des documents, l'indexation peut commencer.

Ce processus d'indexation appelé processus afuIndexer (anciennement processus IndexDocuments) extrait les documents archivés du gestionnaire de ressources pour un type d'élément spécifique, puis extrait toutes les informations textuelles stockées dans chaque document. afuIndexer est un outil Java autonome. Cette fonction n'est pas un exit utilisateur et elle n'emploie pas de fonctions définies par l'utilisateur. Elle ne comporte aucune dépendance distincte.

Les informations textuelles extraites sont stockées dans un fichier XML temporaire, qui est lu par le moteur d'indexation Net Search Extender avec l'aide d'une fonction définie par l'utilisateur (UDF) appelée AFUFetchFile. Net Search Extender analyse le contenu de chaque fichier XML et crée l'index documentaire.

Les étapes afuIndexer sont les suivantes :

- Lecture et vérification de la configuration de l'indexeur, notamment des paramètres propres aux types
- Vérification de l'espace disque disponible et attribution de droits d'accès exclusifs à l'index
- Définition automatique de la taille du lot des éléments (documents) à traiter pendant l'exécution d'une indexation
- Création d'un processus de travail avec un pool d'unités d'exécution pour traiter le lot d'éléments
- Lecture des éléments de la table des tâches ouvertes
- Extraction de documents du gestionnaire de ressources
- Extraction de toutes les informations textuelles stockées dans les éléments et copie de ces informations dans des fichiers XML. Toutes les pièces jointes imbriquées dans des courriers électroniques, les pièces jointes dans IBM Connections et tous les documents Microsoft SharePoint et Système de fichiers archivés sont analysés à l'aide des Filtres Oracle Outside In Technology.

- Déclenchement de Net Search Extender pour traiter les fichiers XML à l'aide de la fonction définie par l'utilisateur AFUFetchfile et ajout du contenu du fichier à l'index documentaire
- Déplacement de tous les éléments indexés de la table des tâches ouvertes vers la table des tâches terminées

Installation de Content Collector Text Search Support pour IBM Content Manager

Vous pouvez rechercher des mots, des phrases et des chaînes de caractères dans le contenu des documents archivés. Si vous utilisez IBM Content Manager comme référentiel, installez Content Collector Text Search Support, qui est fourni avec IBM Content Collector.

Avant de commencer :

- Vérifiez les conditions requises pour l'installation.
- Si vous passez d'IBM CommonStore à IBM Content Collector ou si vous effectuez la mise à niveau à partir d'une version de Content Collector V2.1.0.x, il est recommandé de ne pas désinstaller le composant d'exit utilisateur de recherche de texte sur le serveur Content Manager avant d'avoir installé IBM Content Collector Text Search Support et activé tous les types d'éléments Content Collector ou CommonStore existants pour le traitement par l'indexeur de recherche de texte.
- Si vous mettez à niveau une version de Content Collector V2.1.1.x, vous n'avez pas à désinstaller l'installation précédente de la prise en charge de recherche de texte. Pendant l'exécution du programme d'installation, aucune information ne vous est demandée. Le composant est automatiquement installé à l'emplacement de l'installation précédente. Si des fichiers doivent être écrasés, autorisez cette opération.
- Le programme d'installation tente d'installer IBM Content Collector Text Search Support sans arrêter DB2. Si ce n'est pas possible, l'assistant vous demande l'autorisation d'arrêter DB2. Si vous refusez, le processus d'installation est annulé. Si l'assistant ne parvient pas à arrêter DB2, le processus d'installation est également annulé. Dans ce dernier cas, arrêtez manuellement DB2 et réexécutez l'assistant. Redémarrez DB2 après l'installation.

Vous pouvez installer IBM Content Collector Text Search Support en mode interface graphique ou en mode console.

IBM Content Collector Text Search Support doit être installé sur le serveur IBM Content Manager. Aucun autre composant IBM Content Collector n'a besoin d'être installé pour prendre en charge la recherche de texte.

Installation de Content Collector Text Search Support en mode interface graphique

Pour installer IBM Content Collector Text Search Support en mode interface graphique, suivez la procédure ci-après.

Pour installer Content Collector Text Search Support en mode interface graphique, procédez comme suit.

1. Insérez le DVD du produit dans l'ordinateur sur lequel vous voulez installer le composant de recherche de texte. Vous pouvez également télécharger et extraire le module d'installation approprié.

2. Démarrez l'assistant d'installation :
 - Sous Linux et UNIX, exécutez **install.bin** qui se trouve dans le répertoire `/CM8Text/système_exploitation`. *système_exploitation* correspond au nom du système d'exploitation, par exemple, Linux.
 - Sous Windows, exécutez **install.exe** qui se trouve dans le répertoire `\CM8Text\WIN`.
3. Le cas échéant, répondez aux autres invites.
 - Voici les dossiers d'installation par défaut :

Système d'exploitation	Chemin
Windows (32 bits)	C:\Program Files\IBM\ICCIndexer
Windows (64 bits)	c:\Program Files(x86)\IBM\ICCIndexer
Linux et UNIX	/opt/IBM/ICCIndexer

- Si vous installez le composant sous Linux ou UNIX, indiquez l'ID administrateur DB2.

Installation de Content Collector Text Search Support en mode console

Pour installer IBM Content Collector Text Search Support en mode console, suivez la procédure ci-après.

Pour installer IBM Content Collector Text Search Support en mode console :

1. Insérez le DVD du produit dans l'ordinateur sur lequel vous voulez installer le composant de recherche de texte. Vous pouvez également télécharger et extraire le module d'installation approprié.
2. Ouvrez une fenêtre d'invite de commande et accédez au répertoire `CM8Text` du DVD ou à l'image d'installation.
3. Démarrez le programme d'installation :
 - Sous Linux et UNIX, entrez `install.bin -i console`, qui se trouve dans le répertoire `/CM8Text/système_exploitation`. *système_exploitation* correspond au nom du système d'exploitation, par exemple, Linux.
 - Sous Windows, entrez `install.exe -i console`, qui se trouve dans le répertoire `\CM8Text\WIN`.
4. Suivez les instructions fournies dans la fenêtre d'invite de commande.
 - Voici les dossiers d'installation par défaut :

Système d'exploitation	Chemin
Windows (32 bits)	C:\Program Files\IBM\ICCIndexer
Windows (64 bits)	c:\Program Files(x86)\IBM\ICCIndexer
Linux et UNIX	/opt/IBM/ICCIndexer

- Si vous installez le composant sous AIX, Linux et Linux on System z, indiquez l'ID racine.

Formulaire de configuration de Net Search Extender

Utilisez les feuilles de planification suivantes pour collecter les informations dont vous avez besoin pour travailler avec la prise en charge de recherche de texte IBM Content Collector et Net Search Extender.

Configuration d'IBM Content Manager Net Search Extender

Tableau 1. Informations relatives à la configuration d'IBM Content Manager Text Search Support dans IBM Content Collector

Informations	Remarques	Enregistrez votre valeur ici
ID utilisateur	ID utilisateur exécutant tous les outils de l'indexeur. Il correspond à l'ID utilisateur du propriétaire de l'instance DB2.	
Chemin XML à indiquer dans l'option de configuration GenXMLDirectory	Chemin absolu du répertoire dans lequel sont stockés les fichiers XML de l'indexeur. Cette valeur DOIT être identique pour toutes les configurations de type d'élément. Exemple de valeur : C:\ICCIndexer\XML	
Chemin de fichier journal à indiquer dans l'option de configuration GenLogDirectory	Chemin absolu du répertoire dans lequel sont stockés les fichiers journaux de l'indexeur. Cette valeur DOIT être identique pour toutes les configurations de type d'élément. Exemple de valeur : C:\ICCIndexer\log	
Chemin du fichier de vidage à indiquer dans l'option de configuration GenDumpDirectory	Chemin absolu du répertoire dans lequel sont stockés les documents d'origine extraits du gestionnaire de ressources IBM Content Manager. Cette valeur doit être identique pour toutes les configurations de type d'élément. Exemple de valeur : C:\ICCIndexer\dump	
Chemin de fichier de travail à indiquer dans l'option de configuration GenWorkingDirectoryForItemTypes	Chemin absolu du répertoire dans lequel sont écrits les fichiers de travail requis par le processus de l'indexeur. Cette valeur doit être identique pour toutes les configurations de type d'élément. Exemple de valeur : C:\ICCIndexer\work	

Principales étapes de l'indexation

Vous pouvez démarrer des éléments d'indexation dans le type d'élément immédiatement après la création de la configuration du type d'élément et dès que le lot d'éléments à indexer dépasse la taille minimale définie requise pour démarrer l'indexation.

Les principales étapes d'indexation comprennent :

1. Création d'un nouveau type d'élément ou activation de types d'élément existants pour le traitement par l'indexeur de recherche de texte. Ignorez cette étape si vous avez déjà créé des types d'élément ou activé des types d'élément existants pour le traitement par l'indexeur de recherche de texte.
2. Création d'une configuration pour le type d'élément.
3. Indexation d'éléments.
4. Vérification du code retour du processus de l'indexeur
5. Vérification du statut des éléments indexés

Configuration d'un référentiel IBM Content Manager

Tous les documents archivés à l'aide d'IBM Content Collector sont stockés dans un type d'élément dans IBM Content Manager. Vous devez disposer d'au moins un type d'élément IBM Content Manager pour chaque système source que vous configurez dans IBM Content Collector.

La première fois que vous installez Content Collector et exécutez sa configuration initiale, l'assistant de configuration vous guide dans la procédure de création d'un type d'élément pour chaque système source qui a été sélectionné pendant l'installation. Par exemple, si vous avez sélectionné Microsoft Exchange, Système de fichiers et Microsoft SharePoint, vous serez guidé tout au long du processus de création de types d'élément pour ces trois systèmes source.

Au cours de la configuration initiale de Content Collector, des types d'élément sont créés comme suit.

- Types d'élément de ressource pour les systèmes source de courrier électronique et les sources Système de fichiers
- Types d'élément de modèle de document pour les systèmes source IBM Connections et Microsoft SharePoint

Pour créer des types d'élément supplémentaires après la première installation de Content Collector, utilisez les outils de configuration de Content Collector.

Pour les sources Système de fichiers, Content Collector prend également en charge les types d'élément de modèle de document IBM Content Manager. Toutefois, vous devez créer ces types d'élément manuellement dans IBM Content Manager.

Pour les systèmes source de courrier électronique, la création de type d'élément varie en fonction du modèle de données de courrier électronique sélectionné :

- Si vous mettez à niveau vers IBM Content Collector V3 et que vous utilisez le modèle de données intégré pour les types d'élément de courrier électronique, vous pouvez sélectionner ces modèles de données de courrier électronique pour l'archivage à partir de boîtes aux lettres Lotus Domino ou Microsoft Exchange :

Modèle de données composé

Ce modèle archive les documents en parties. Il existe depuis IBM Content Collector V2.1.1.

S'il est sélectionné, deux types d'élément sont créés lorsque le type d'élément est configuré ; un *type d'élément de courrier électronique* pour le corps du courrier et les métadonnées, et un *type d'élément de pièce jointe* pour les pièces jointes imbriquées dans les courriers. Un type d'élément

de pièce jointe peut être partagé par plusieurs types d'élément de courrier électronique. Cette division améliore considérablement le dédoublement des pièces jointes.

Modèle de données intégré

Ce modèle archive toutes les données, y compris les pièces jointes imbriquées, sous la forme d'une seule unité dans un type d'élément unique. Il a été utilisé dans toutes les versions de Content Collector antérieures à la version V2.1.1. Si vous effectuez une mise à niveau à partir d'une version Content Collector antérieure à la version ayant créé des types d'élément de type intégré, vous pouvez choisir de créer des types d'élément correspondant aux deux modèles de données.

Cependant, l'amélioration du dédoublement des pièces jointes étant pris en charge uniquement à l'aide du nouveau modèle de données composé, il est fortement conseillé d'utiliser le modèle de données composé lorsque vous créez de nouveaux types d'élément pour l'archivage et l'indexation des courriers électroniques. C'est particulièrement le cas lorsque vous divisez des types d'élément ; activez le traitement de types d'éléments existants par l'indexeur IBM Content Collector pour la recherche de texte, et, lorsque le type d'élément suivant doit être créé, passez au nouveau modèle.

- S'il s'agit d'une première installation d'IBM Content Collector ou si vous procédez à une mise à niveau d'une version V2.1.1.x d'IBM Content Collector, vous ne pouvez pas sélectionner le modèle de données Courrier électronique.

Les autres systèmes source ne nécessitent pas un modèle de données personnalisé.

Tous les nouveaux types d'élément sont activés par défaut pour le traitement par l'indexeur de recherche de texte IBM Content Collector, sauf si vous en décidez autrement. Ces types d'élément activés peuvent uniquement être traités par l'indexeur de recherche de texte IBM Content Collector. Ils ne peuvent pas être traités par l'indexeur rapide ou l'indexeur IBM Content Manager standard avec l'exit utilisateur IBM Text Search.

Pour créer un type d'élément IBM Content Manager :

1. Planifiez la création de votre type d'élément.
2. Sélectionnez **Outils de configuration > Configuration d'un référentiel CM**
3. Sélectionnez le système source pour lequel vous voulez créer un type d'élément.
4. Sélectionnez le serveur IBM Content Manager.
5. Entrez l'ID et le mot de passe administrateur.
6. Si vous souhaitez rechercher le contenu du document, choisissez d'activer le type d'élément pour la recherche de texte et entrez le chemin dans lequel le module IBM Content Collector Text Search Support a été installé.
7. Fournissez les informations répertoriées dans le tableau ci-après.

Système source	Tâche
Systèmes de messagerie source : <ul style="list-style-type: none"> • Lotus Domino • Microsoft Exchange 	<ol style="list-style-type: none"> 1. Sélectionnez le modèle de données pour le type d'élément. Si vous êtes un nouvel utilisateur d'IBM Content Collector, ou si vous mettez à niveau à partir d'une version V2.1.1.x d'IBM Content Collector, vous ne pouvez pas sélectionner un modèle de données. Seuls les types d'élément du modèle de données composé sont créés. 2. Entrez le nom du type d'élément de courrier électronique. 3. Sélectionnez la création d'un nouveau type d'élément de pièce jointe et entrez un nom, ou sélectionnez l'utilisation d'un type d'élément de pièce jointe et choisissez-le. 4. Si le type d'élément est activé pour la recherche de texte, entrez le chemin d'accès relatif au répertoire d'index et au répertoire de travail d'index.
Courrier électronique via SMTP ou Système de fichiers	<ol style="list-style-type: none"> 1. Entrez le nom du type d'élément. Tous les types d'élément SMTP et Système de fichiers sont créés automatiquement comme des types d'élément de ressource. 2. Si le type d'élément est activé pour la recherche de texte, entrez le chemin d'accès relatif au répertoire d'index et au répertoire de travail d'index.
IBM Connections ou Microsoft SharePoint	<ol style="list-style-type: none"> 1. Entrez le nom du type d'élément. Tous les types d'élément IBM Connections et Microsoft SharePoint sont automatiquement créés comme des types d'élément de modèle de document avec un composant de document de base ICMBASETEXT. Vous pouvez ajouter d'autres composants dans IBM Content Manager. Notez toutefois que vous ne pouvez utiliser que les composants ICMBASETEXT et ICMBASE lorsque vous configurez des routes de tâches dans le Gestionnaire de configuration IBM Content Collector. 2. Si le type d'élément est activé pour la recherche de texte, entrez le chemin d'accès relatif au répertoire d'index et au répertoire de travail d'index.

Planification de votre création de type d'élément

Avant de créer un type d'élément, vous devez choisir de conserver ou non la description d'attribut par défaut d'un type d'élément défini par IBM Content Collector ou d'ajouter d'autres attributs personnalisés et des attributs IBM Content Manager supplémentaires au type d'élément. La description d'attribut ne peut pas être modifiée après la création du type d'élément sans recréer l'index de texte intégral. Envisagez également de créer plusieurs types d'élément pour l'archivage afin d'éviter que les fichiers d'index ne deviennent trop volumineux. Les recherches sur des index plus petits donnent de meilleurs résultats.

Les attributs représentent le contenu du document que vous pouvez ensuite analyser lorsque vous recherchez le document.

Vous pouvez étendre l'index de texte complet Net Search Extender en ajoutant des informations supplémentaires à partir de documents ou d'attributs IBM Content

Manager. Lorsque vous ajoutez des définitions de zone contenant ces informations à des fichiers de configuration, les données de ces zones sont ajoutées à l'index lorsque l'index est créé. Les informations supplémentaires de l'index étendent la portée de la recherche, ce qui signifie que des informations d'autres zones de document peuvent être utilisées pour rechercher des documents archivés.

Les attributs de document sont définis dans un fichier modèle. Les fichiers modèle suivants existent et peuvent être utilisés par le moteur d'indexation Net Search Extender. Les fichiers modèle contiennent l'ensemble complet de définitions de zone requis pour l'indexation de recherche de texte par IBM Content Collector.

Tableau 2. Fichiers modèle utilisés par IBM Content Collector

Type d'élément	Fichier modèle	Version du produit
Type d'élément de courrier électronique Content Collector groupé	<i>répertoire_TextSearchSupport/model/afu_mail_entire.xml</i>	Créé à l'aide de l'IBM Content Collector V2.1.0
Type d'élément de courrier électronique Content Collector composé	<i>répertoire_TextSearchSupport/model/afu_mail_entire_cmp.xml</i>	Créé à l'aide d'IBM Content Collector V2.1.1 et V2.2
Type d'élément de courrier électronique Content Collector composé	<i>répertoire_TextSearchSupport/model/afu_mail_entire_cmp2.xml</i>	Créé à l'aide d'IBM Content Collector V3
Types d'élément de courrier électronique IBM CommonStore pour le modèle de données BUNDLED et le type d'archivage ENTIRE (BRI)	<i>répertoire_TextSearchSupport/model/cs_mail_entire.xml</i>	Créé à l'aide d'IBM CommonStore V8.4
Types d'élément de système de fichiers	<i>répertoire_TextSearchSupport/model/afu_generic.xml</i>	Toutes les versions d'IBM Content Collector
Types d'élément Content Collector créés pour des données IBM Connections	<i>répertoire_TextSearchSupport/model/afu_conn.xml</i>	Créé à l'aide d'IBM Content Collector V3
Types d'élément Microsoft SharePoint	<i>répertoire_TextSearchSupport/model/afu_generic.xml</i>	Créé avant IBM Content Collector V3
Types d'élément Microsoft SharePoint avec des pièces jointes intégrées	<i>répertoire_TextSearchSupport/model/afu_sp.xml</i>	Créé à l'aide d'IBM Content Collector V3

Pour les documents Lotus Domino, vous pouvez ajouter des définitions de zone pour les zones du type suivant :

- Zone de texte
- Zone de liste de texte
- Zone de texte RFC822

Pour les documents Exchange, toutes les zones définies par l'utilisateur peuvent être indexées.

De plus, tous les attributs définis dans IBM Content Manager peuvent être ajoutés à l'index pour tous les types de document.

Important :

1. Si vous définissez des attributs supplémentaires à ajouter à l'index, vous devez ajouter les mêmes définitions au fichier modèle et au fichier de configuration de l'indexeur pour le type d'élément.
2. Le fichier de configuration de l'indexeur est créé lorsque vous exécutez `afuConfig`. Pour indexer les éléments d'un type d'élément, vous devez mettre à jour manuellement le fichier de configuration de l'indexeur pour inclure les mêmes définitions d'attribut que celles ajoutées au fichier modèle. Lorsque le processus de l'indexeur est démarré sur le type d'élément, il vérifie que le fichier modèle et le fichier de configuration de l'indexeur sont cohérents.

Avant de créer un type d'élément, vous devez :

1. Déterminez quelles sont les données utilisateur supplémentaires à ajouter comme attributs personnalisés au fichier modèle. Par exemple, vous pouvez, si vous le souhaitez, stocker le numéro d'enregistrement de l'entreprise pour l'utiliser dans des requêtes de recherche ultérieures.
2. Ajoutez des attributs au fichier modèle correspondant au modèle de données du type d'élément que vous souhaitez créer.

Activation d'un référentiel IBM Content Manager à des fins de traitement par l'indexeur pour la recherche de texte




Tous les types d'élément existants créés dans des versions antérieures à Content Collector 2.1.1 doivent être activés pour le traitement par l'indexeur pour la recherche de texte pour que l'indexation puisse être exécutée sur ces types d'élément.

Important : Le type d'élément doit être activé pour le traitement par l'indexeur de recherche de texte pour pouvoir utiliser un élément créé avant Content Collector 2.1.1 dans un version Content Collector ultérieure à 2.1.1. Avant d'activer le type d'élément existant, vous devez arrêter l'archivage pour ce type d'élément et vérifier que tous ses éléments archivés ont été indexés à l'aide de l'indexeur rapide ou de l'indexeur IBM Content Manager standard. La table de journal Net Search Extender doit être vide.

Si vous choisissez de ne pas activer un type d'élément existant pour le traitement par l'indexeur de recherche de texte, vous ne pouvez le traiter qu'à l'aide de l'indexeur rapide ou de l'indexeur IBM Content Manager standard associé à l'exit utilisateur IBM Text Search. Vous ne pouvez pas traiter le type d'élément à l'aide de l'indexeur IBM Content Collector pour la recherche de texte.

Seuls des types d'élément Content Collector existants peuvent être activés à l'aide des outils de configuration Content Collector. Pour activer des types d'élément IBM CommonStore pour le traitement par l'indexeur de recherche de texte, voir Indexation d'éléments qui ont été archivés à l'aide d'IBM(r) CommonStore.

Pour activer des types d'élément IBM Content Manager existants pour le traitement par l'indexeur IBM Content Collector pour la recherche de texte :

1. Sélectionnez **Outils de configuration > Activation d'un référentiel CM**
2. Sélectionnez le serveur IBM Content Manager sur lequel se trouvent les types d'élément et entrez l'ID d'administrateur et le mot de passe.
3. Sélectionnez les types d'élément à activer pour la recherche de texte dans la liste des types d'élément chargés.
 - Les types d'élément marqués par  peuvent être sélectionnés.
 - Les types d'élément marqués par  sont déjà activés et ne peuvent pas être sélectionnés à nouveau.
 - Les types d'élément marqués par  ne peuvent pas être activés car la table de journalisation Net Search Extender pour le type d'élément n'est pas vide. Cela signifie que le type d'élément n'a pas été totalement indexé par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Avant d'avoir activé le type d'élément pour le traiter avec l'indexeur de recherche de texte d'IBM Content Collector, vous devez terminer l'indexation du type d'élément avec l'indexeur rapide ou l'indexeur IBM Content Manager standard.

La sélection unique et la sélection multiple sont prises en charge.

4. Pour activer les types d'élément sélectionnés, cliquez sur **Activer les types d'éléments sélectionnés**.
5. Pour régénérer la liste des types d'éléments, cliquez sur **Recharger les types d'éléments**.

Important :

- Les types d'élément activés ne peuvent être traités que par l'indexeur de recherche de texte IBM Content Collector. Ils ne peuvent pas être traités par l'indexeur rapide ou l'indexeur IBM Content Manager standard avec l'exit utilisateur IBM Text Search.
- Après avoir activé un type d'élément existant pour le traiter avec l'indexeur de recherche de texte, vous pouvez continuer à utiliser le même index de texte complet que celui utilisé pour le type d'élément avec l'indexeur rapide ou l'indexeur IBM Content Manager standard associé à l'exit utilisateur IBM Text Search. Vous n'avez pas besoin de recréer les index de texte intégral.
- L'indexeur de recherche de texte IBM Content Collector peut coexister avec l'indexeur rapide ou l'indexeur IBM Content Manager standard associé à l'exit utilisateur IBM Text Search sur la même machine IBM Content Manager et peut s'exécuter simultanément mais pas sur le même type d'élément.

Activation de types d'éléments qui n'ont pas été créés dans Content Collector

Les types d'élément qui ont été directement créés dans IBM Content Manager pour l'archivage dans IBM CommonStore, ou à d'autres fins, doivent être activés pour l'indexation à l'aide de l'indexeur IBM Content Collector de recherche de texte pour pouvoir indexer des éléments dans ces types d'élément.

Vous pouvez continuer à utiliser le même index de recherche que celui utilisé pour le type d'élément avec l'indexeur rapide ou l'indexeur IBM Content Manager standard associé à l'exécutif utilisateur IBM Text Search. Vous n'avez pas besoin de recréer les index de texte intégral.

Les types d'élément qui ont été créés directement dans IBM Content Manager for archiving in IBM CommonStore ne peuvent pas être activés à l'aide des outils de configuration IBM Content Collector. Vous devez exécuter l'outil `afuEnableItemType` manuellement pour activer l'indexation à l'aide de l'indexeur Content Collector.

L'activation d'un type d'élément équivaut à préparer le type d'élément pour l'indexation à l'aide de l'indexeur Content Collector. Ceci inclut la création de la table des tâches ouvertes et de la table des tâches terminées pour le type d'élément. Lorsque vous activez un type d'élément et si celui-ci a été créé pour l'archivage dans IBM CommonStore, la table des tâches terminées est renseignée avec tous les éléments et le code d'état d'indexation, la valeur `IDXRC` de chaque élément est définie selon que l'élément a été indexé à l'aide de l'indexeur rapide ou de l'indexeur IBM Content Manager standard. Le renseignement de la table des tâches terminées avec une valeur `IDXRC` permet de suivre les éléments qui n'ont pas correctement indexés à l'aide d'IBM CommonStore. Ces éléments peuvent être retraités à l'aide de l'indexeur Content Collector. Pour plus de détails sur les valeurs `IDXRC` définies pour les éléments CommonStore, voir «Utilisation de `TIEFLAG` lors de l'indexation dans IBM Content Collector», à la page 61.

Si vous souhaitez vérifier les éléments qui n'ont pas été indexés correctement, vous pouvez exécuter `afuIndexTool` avec l'argument `-opValidate -deepvalidation`. Par la suite, vous pouvez réindexer les éléments à l'aide de l'outil `afuIndexer` avec l'un des arguments de réindexation.

Pour activer des types d'élément CommonStore existants pour l'indexation à l'aide de l'indexeur IBM Content Collector de recherche de texte, exécutez le script suivant :

- Sous UNIX et Linux, exécutez :
`afuEnableItemType.sh`
- Sous Windows :
`afuEnableItemType.cmd`

Remarque : Lorsque vous exécutez le processus `afuEnableItemType` :

- Tous les arguments de ligne de commande font une distinction entre les majuscules et les minuscules.
- Les arguments de ligne de commande peuvent apparaître dans n'importe quel ordre.
- Le nom du type d'élément ne fait pas de distinction entre les majuscules et les minuscules.

Si vous exécutez `afuEnableItemType` sans indiquer d'argument, ou si vous exécutez `afuEnableItemType` avec l'argument `-?`, les éléments suivants sont affichés :

```
afuEnableItemType: Syntaxe :
[-?|-h|-copyright]
-it <itemtype>
-database <database>
[-user <userID|@environmentVariable|:filename>]
[-password <password|@environmentVariable|:filename>]
[-schema <schema>]
```

```
[-tablespace <tablespace>]
[-udftransferbuffersize <buffer size in MB>]
[-verbose] [-showsql] [-trace] [-simulate]
[-quiet]
```

où :

-? Répertoire tous les arguments que vous pouvez transmettre à l'outil de l'indexeur `afuEnableItemType`.

-h Répertoire tous les arguments que vous pouvez transmettre à l'outil de l'indexeur `afuEnableItemType`.

-copyright

Affiche la notice de copyright et les informations de version.

-it *<typed'élément>*

Obligatoire.

Type d'élément Content Manager (collection de documents). Les noms de type d'élément ne sont pas sensibles à la casse.

-database *<basededonnées>*

Obligatoire.

Nom de la base de données du serveur de bibliothèques où le type d'élément est stocké. Le nom de la base de données peut être indiqué dans deux formats qui déterminent comment l'outil de l'indexeur se connecte à la base de données. La connexion à la base de données est mise en oeuvre à l'aide de JDBC (Java Data Base Connectivity) qui définit comment un client Java™ peut accéder à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Pour utiliser une connexion JDBC de type 4, le nom de la base de données de l'argument `-database` doit inclure un nom de serveur, un numéro de port et un nom de base de données au format `<serveurbasededonnées>:<numérodeport>/<nombasededonnées>`, par exemple `-database localhost:50000/icm1sdb`. Il est recommandé d'utiliser JDBC type 4 comme méthode d'accès à la base de données mais ce mécanisme requiert la spécification complète de l'ID utilisateur et le mot de passe.

Pour utiliser une connexion JDBC type 2, la spécification du nom de la base de données de l'argument `-database` doit correspondre à un nom de connexion DB2 existant. En d'autres termes, le nom de la base de données doit être catalogué dans DB2, par exemple `-database icm1sdb`. Comme JDBC type 2 utilise une connexion à une base de données existante, il est inutile d'indiquer un ID utilisateur et un mot de passe. Toutefois, une version 32 bits native du pilote JDBC type 2 DB2 (bibliothèque partagée) doit être disponible et définie dans le chemin d'accès aux classes pour l'utilisation de la connexion JDBC type 2.

-user *<userID|@environmentVariable|:filename>*

Facultatif.

Indiquez un ID utilisateur. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un ID utilisateur de trois manières :

1. Indiquez l'ID utilisateur directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle l'ID utilisateur est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_USERID=user8
```

```
afuEnableItemType -user @INDEXER_USERID
```

3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuEnableItemType -user :myUserID.txt
```

Le fichier texte doit contenir une seule ligne avec l'ID utilisateur (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer l'ID utilisateur dans toutes les informations d'état des processus.

L'utilisateur doit disposer de droits d'accès à IBM Content Manager.

-password <userID|@environmentVariable|:filename>

Facultatif.

Mot de passe de l'ID utilisateur doté des droits administrateur de la base de données. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un mot de passe de trois manières :

1. Indiquez le mot de passe directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle le mot de passe est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_PASSWORD=12345678
```

```
afuEnableItemType -password @INDEXER_PASSWORD
```

3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuEnableItemType -password :myPassword.txt
```

Le fichier texte doit contenir une seule ligne avec le mot de passe (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer le mot de passe dans toutes les informations d'état des processus.

Les mots de passe ne sont pas consignés dans un format de texte normal lisible.

-schema <schéma>

Facultatif.

Nom du schéma de la base de données s'il est différent du schéma de la base de données en cours.

-tablespace <tablespace>

Facultatif.

Permet d'indiquer quel espace de table utiliser pour les tables utilisées par le processus de l'indexeur. Si cet argument n'est pas transmis, l'espace de table DB2 par défaut est utilisé.

-udftransferbuffersize <taille de mémoire tampon en Mo>

Facultatif.

Permet d'indiquer la taille de la mémoire tampon utilisée par la fonction définie par l'utilisateur AFUFetchFile pour charger les fichiers XML temporaires pour être accessibles à Net Search Extender. La taille de mémoire tampon de transfert par défaut est de 32 Mo. Si vous devez modifier la taille de la

mémoire tampon de transfert définie par l'utilisateur après l'activation du type d'élément, voir «Modification de la taille de la mémoire tampon de transfert par défaut de la fonction définie par l'utilisateur AFUFetchFile», à la page 87.

-verbose

Facultatif.

Affiche des informations détaillées sur le traitement. L'utilisation de l'option **-verbose** est utile au début pour obtenir des informations détaillées sur l'ensemble des traitements. Si **-verbose** est indiqué, les fichiers journaux contiennent également des informations supplémentaires.

-showsq1

Facultatif.

Affiche toutes les commandes SQL DB2 exécutées par le système. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de tous les traitements système DB2. Si cet argument est indiqué, toutes les instructions SQL sont également ajoutées au fichier journal.

-trace

Facultatif.

Fournit des informations plus détaillées que l'option **-verbose** en affichant des données supplémentaires dans les fichiers journaux pour effectuer l'identification des erreurs. Cette opération est obligatoire pour le service de support logiciel IBM uniquement et requiert l'utilisation du fichier des propriétés de la fonction de trace.

-simulate

Facultatif.

Exécute le processus de création de l'index sans mettre à jour les bases de données. Utilisez ce paramètre pour valider la configuration d'index.

-quiet

Aucune étape de traitement n'est affichée sur la console.

Création d'un fichier de configuration de l'indexeur pour le type d'élément

Pour indexer les éléments d'un type d'élément à l'aide de l'indexeur de recherche de texte, vous devez d'abord créer une configuration de l'indexeur associée. Une configuration de l'indexeur de type d'élément peut être distribuée à plusieurs fichiers ou être partagée entre plusieurs types d'élément.

Si le type d'élément est un type d'élément composé de courrier électronique, créez uniquement une configuration d'indexeur pour le type d'élément de courrier électronique. Ne créez pas de configuration ni ne tentez d'indexer le type d'élément de pièce jointe correspondant. Les types d'élément de pièce jointe ne peuvent pas être traités seuls ; ils sont traités automatiquement comme partie du type d'élément de courrier électronique.

Vous pouvez créer un fichier de configuration de l'indexeur en appelant l'outil `afuConfig` ou en créant manuellement un fichier de configuration à l'aide de votre éditeur de texte favori. `afuConfig` reconnaît automatiquement s'il s'agit d'un type d'élément Content Collector ou CommonStore activé et crée une configuration par défaut en fonction du modèle de données du type d'élément. Le nouveau fichier de configuration s'appelle `<nom_type_élément>.ini`.

La configuration par défaut contient toutes les options de configuration nécessaires à l'indexation et n'a pas besoin d'être modifiée. Les modifications apportées aux options de configuration sont nécessaires uniquement pour optimiser le processus de l'indexeur ou pour déboguer ou réindexer des données. L'indexation peut être exécutée à l'aide des valeurs par défaut.

Les seules options n'ayant pas de valeurs par défaut car elles dépendent de l'utilisateur sont les valeurs des chemins de fichiers absolus qui vous sont demandées lorsque vous exécutez `afuConfig` pour la première fois :

- **GenXmlDirectory** : répertoire dans lequel les fichiers XML intermédiaires, qui sont les entrées de l'indexation, sont stockés.
Cette valeur DOIT être identique pour toutes les configurations de type d'élément.
- **GenLogDirectory** : chemin de fichier dans lequel les fichiers journaux de l'indexeur sont stockés.
Cette valeur DOIT être identique pour toutes les configurations de type d'élément.
- **GenDumpDirectory** : répertoire dans lequel les documents d'origine extraits du gestionnaire de ressources IBM Content Manager sont éventuellement stockés.
Il est recommandé d'avoir une valeur identique pour toutes les configurations de type d'élément.
- **GenWorkingDirectoryForItemTypes** : emplacement où les fichiers de travail requis par le processus de l'indexeur sont écrits.
Il est recommandé que cette valeur soit identique pour toutes les configurations de type d'élément.

Les informations entrées relatives au chemin du fichier sont valables pour toutes les configurations de type d'élément et vous n'êtes invité à les fournir que la première fois que vous créez une configuration de type d'élément. L'outil `afuConfig` vérifie que les répertoires que vous avez indiqués existent réellement et que les noms de répertoires contiennent uniquement les caractères alphanumériques pris en charge. Si les répertoires n'existent pas, vous êtes invité à indiquer si vous souhaitez les créer.

Important : Si vous modifiez les chemins de fichier **GenLogDirectory** ou **GenXmlDirectory** ultérieurement, vous devez mettre à jour toutes les configurations de type d'élément car ces chemins de fichier doivent être cohérents avec tous les types d'élément. Le processus de l'indexeur enregistre automatiquement ces modifications et crée de nouveaux répertoires si nécessaire.

Si les valeurs de **GenLogDirectory** ou de **GenXmlDirectory** sont modifiées, la configuration de la fonction définie par l'utilisateur `AFUFetchFile` est automatiquement mise à jour. Ces informations sont indiquées dans un message pouvant nécessiter le redémarrage de DB2 pour que les modifications soient prises en compte.

Ces répertoires étant partagés entre tous les processus de l'indexeur simultanément en cours d'exécution et leur taille pouvant devenir très importante, vous devez vous assurer que l'espace disque qui leur est alloué est suffisant, tout particulièrement pour les fichiers XML temporaires dans le répertoire indiqué par l'option de configuration **GenXMLDirectory**.

Pour éviter d'avoir à conserver de nombreuses configurations de type d'élément similaires, il est recommandé de distribuer la configuration de l'indexeur d'un type

d'élément à plusieurs fichiers de configuration. Toutefois, vous ne pouvez pas partager les configurations de l'indexeur entre les types d'élément qui utilisent des modèles de données différents. Pour plus d'informations, voir «Modification de la configuration par défaut de l'indexeur», à la page 66.

Pour créer un fichier de configuration de l'indexeur pour le type d'élément :

- Sous UNIX et Linux, exécutez
afuConfig.sh
- Sous Windows, exécutez
afuConfig.cmd

Remarque : Lorsque vous exécutez le processus afuConfig :

- Tous les arguments de ligne de commande font une distinction entre les majuscules et les minuscules.
- Les arguments de ligne de commande peuvent apparaître dans n'importe quel ordre.
- Le nom du type d'élément ne fait pas de distinction entre les majuscules et les minuscules.

Si vous exécutez afuConfig et que vous n'indiquez pas d'argument ou que vous entrez la commande suivie de `-?`, les données suivantes s'affichent :

```
afuConfig: Syntaxe :
[-?|-h|-copyright]
[-it <typed'élément>]
-database <database>
[-user <userID|@environmentVariable|:filename>]
[-password <password|@environmentVariable|:filename>]
[-schema <schema>]
[-verbose] [-showsql] [-trace]
[-quiet]
```

où :

- ? Répertorie tous les arguments que vous pouvez transmettre à l'outil afuConfig.
- h Répertorie tous les arguments que vous pouvez transmettre à l'outil afuConfig.

-copyright

Affiche la notice de copyright et les informations de version.

-it <typed'élément>

Facultatif.

Nom du type d'élément IBM Content Manager (collection de documents). Si afuConfig est appelé sans le nom de type d'élément, l'utilisateur est invité à l'indiquer.

Important : Si votre type d'élément est un type d'élément composé de courrier électronique, n'entrez pas le type d'élément de pièce jointe. Les types d'élément de pièce jointe ne peuvent pas être traités seuls mais sont traités automatiquement comme partie du type d'élément de courrier électronique.

-database <basededonnées>

Obligatoire.

Nom de la base de données du serveur de bibliothèques où le type d'élément est stocké. Le nom de la base de données peut être indiqué dans deux formats qui déterminent comment l'outil de l'indexeur se connecte à la base de données. La connexion à la base de données est mise en oeuvre à l'aide de

JDBC (Java Data Base Connectivity) qui définit comment un client Java peut accéder à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Pour utiliser une connexion JDBC de type 4, le nom de la base de données de l'argument `-database` doit inclure un nom de serveur, un numéro de port et un nom de base de données au format `<serveurbasededonnées>:<numérodeport>/<nombasededonnées>`, par exemple `-database localhost:50000/icm1sdb`. Il est recommandé d'utiliser JDBC type 4 comme méthode d'accès à la base de données mais ce mécanisme requiert la spécification complète de l'ID utilisateur et le mot de passe.

Pour utiliser une connexion JDBC type 2, la spécification du nom de la base de données de l'argument `-database` doit correspondre à un nom de connexion DB2 existant. En d'autres termes, le nom de la base de données doit être catalogué dans DB2, par exemple `-database icm1sdb`. Comme JDBC type 2 utilise une connexion à une base de données existante, il est inutile d'indiquer un ID utilisateur et un mot de passe. Toutefois, une version 32 bits native du pilote JDBC type 2 DB2 (bibliothèque partagée) doit être disponible et définie dans le chemin d'accès aux classes pour l'utilisation de la connexion JDBC type 2.

-user `<userID|@environmentVariable|:filename>`

Facultatif.

Indiquez un ID utilisateur. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un ID utilisateur de trois manières :

1. Indiquez l'ID utilisateur directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle l'ID utilisateur est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_USERID=user8
```



```
afuConfig -user @INDEXER_USERID
```
3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuConfig -user :myUserID.txt
```

Le fichier texte doit contenir une seule ligne avec l'ID utilisateur (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer l'ID utilisateur dans toutes les informations d'état des processus.

L'utilisateur doit disposer de droits d'accès à IBM Content Manager.

-password `<password|@environmentVariable|:filename>`

Facultatif.

Mot de passe de l'ID utilisateur doté des droits administrateur de la base de données. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un mot de passe de trois manières :

1. Indiquez le mot de passe directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle le mot de passe est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_PASSWORD=12345678
```

```
afuConfig -password @INDEXER_PASSWORD
```

3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,
afuConfig -password :myPassword.txt

Le fichier texte doit contenir une seule ligne avec le mot de passe (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer le mot de passe dans toutes les informations d'état des processus.

Les mots de passe ne sont pas consignés dans un format de texte normal lisible.

-schema <*schéma*>

Facultatif.

Nom du schéma de la base de données s'il est différent du schéma de la base de données par défaut.

-verbose

Facultatif.

Affiche des informations détaillées sur le traitement. L'utilisation de l'option **-verbose** est utile au début pour obtenir des informations détaillées sur l'ensemble des traitements. Si **-verbose** est indiqué, les fichiers journaux contiennent également des informations supplémentaires.

-showsql

Facultatif.

Affiche toutes les commandes SQL DB2 exécutées par le système. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de tous les traitements système DB2. Si cet argument est indiqué, toutes les instructions SQL sont également ajoutées au fichier journal.

-trace

Facultatif.

Fournit des informations plus détaillées que l'option **-verbose** en affichant des données supplémentaires dans les fichiers journaux pour effectuer l'identification des erreurs. Cette opération est obligatoire pour le service de support logiciel IBM uniquement et requiert l'utilisation du fichier des propriétés de la fonction de trace.

-quiet

Aucune étape de traitement n'est affichée sur la console.

Si vous avez apporté des modifications aux fichiers modèles de type d'élément lors de la création des types d'élément, vous devez également les ajouter au fichier de configuration de l'indexeur pour le type d'élément avant de procéder à l'indexation.

Si vous venez de créer une nouvelle configuration pour un type d'élément CommonStore activé et que vous avez défini des attributs propres dans le fichier de configuration d'exit utilisateur CommonStore icmfce_config.ini, vous devez ajouter ces entrées d'attributs définis par l'utilisateur au nouveau fichier de configuration de l'indexeur en utilisant la syntaxe de configuration de l'indexeur appropriée.

Pour reconnaître les attributs définis par l'utilisateur :

Ajoutez les définitions d'attribut au fichier de configuration d'index.

Référence associée :

«Options de configuration de l'indexeur», à la page 67

«Options de configuration de l'indexeur», à la page 67

Indexation des éléments d'un type d'élément

Si le type d'élément a été activé pour être traité par l'indexeur de recherche de texte et que le processus de l'indexeur et le type d'élément ont été configurés, le processus de l'indexeur peut être démarré.

Si vous avez apporté des modifications au fichier modèle du type d'élément lors de la création de ce dernier, vous devez ajouter ces modifications au fichier de configuration de l'indexeur du type d'élément pour pouvoir commencer l'indexation.

Si vous venez de créer une configuration pour un type d'élément CommonStore actif, et que vous avez déjà défini vos propres attributs dans le fichier de configuration icmfce_config.ini de l'exécutif utilisateur CommonStore, vous devez ajouter ces entrées d'attribut définies par l'utilisateur au nouveau fichier de configuration de l'indexeur afin de vous assurer que ces attributs sont également ajoutés à l'index.

L'indexeur pour le processus de recherche de texte (anciennement processus IndexDocuments) extrait les documents archivés du gestionnaire de ressources IBM Content Manager pour un type d'élément spécifique et extrait tout le contenu textuel de chaque document. Ce contenu textuel est stocké dans un fichier XML temporaire, qui est lu par le moteur d'indexation Net Search Extender avec l'aide de la fonction définie par l'utilisateur (UDF) appelée AFUFetchFile. Net Search Extender analyse le contenu de chaque fichier XML et crée l'index documentaire.

Vous pouvez exécuter l'indexeur de recherche de texte de l'une des manières suivantes :

- Mode de traitement par lots : utilisez un planificateur, tel que cron ou le Planificateur de tâches Windows, pour exécuter le processus d'indexation sur un type d'élément dans une planification.

Le mode de traitement par lots est le mode par défaut. Il convient d'utiliser le mode de traitement par lots si vous souhaitez traiter un lot d'articles seulement, par exemple l'ensemble des éléments qui ont été archivés à la fin d'une journée de travail. Il convient également d'utiliser le mode de traitement par lots si le mode continu n'est pas possible parce que le système se trouve immobilisé pour quelques heures, par exemple pour les besoins de la maintenance.

Lorsque vous utilisez le mode de traitement par lots, vous devez définir soigneusement la période d'inactivité et ne pas omettre de calculer le temps nécessaire à l'achèvement du traitement d'un lot actif. Vous devez surveiller attentivement les processus de traitement précédents afin de parvenir à une estimation fiable du temps nécessaire au traitement d'un lot complet. Pour plus de détails sur les éléments à prendre en compte lors de la configuration du mode de traitement par lots, voir «Contraintes lors de l'indexation en mode de traitement par lots», à la page 88.

- Mode continu : exécutez le processus d'indexation dans une boucle sans fin jusqu'à interruption par une commande d'arrêt explicite.

Il convient d'utiliser le mode continu si le processus d'indexation s'exécute sur une longue période, voire sur plusieurs jours. Dans ce cas, le processus

d'indexation continue de traiter tous les nouveaux articles au fur et à mesure de leur archivage et de mettre à jour l'index à intervalles définis. Pour plus d'informations, voir «Exécution de afuIndexer en mode continu», à la page 92.

Pour traiter aussi rapidement que possible des commandes en attente, vous devez toujours essayer d'utiliser le mode continu, car il réduit le temps d'inactivité, même si l'exécution en mode continu est restreinte par la nécessité d'exécuter des sauvegardes régulières et qu'elle augmente la charge du système durant les heures de bureau.

- Mode d'erreurs graves : exécutez le processus d'indexation dans un mode spécial à l'aide de paramètres optimisés afin de retraiter uniquement les éléments qui ont été précédemment traités et qui se sont soldés par une erreur grave. Ce mode n'est pas un mode de traitement classique et ne doit être utilisé que pour retraiter les éléments qui ont généré une erreur grave pendant un traitement précédent et qui n'ont pas été indexés.

Le processus de l'indexeur de recherche de texte ne peut être exécuté que sur un seul type d'élément à la fois mais plusieurs types d'élément peuvent être traités simultanément. Pour plus d'informations, voir «Traitement simultané de plusieurs types d'élément», à la page 95.

Si le type d'élément est un type d'élément composé de courrier électronique, exécutez uniquement l'indexeur de recherche de texte sur le type d'élément de courrier électronique. N'indexez pas le type d'élément de pièce jointe correspondant. Les types d'élément de pièce jointe ne peuvent pas être traités seuls ; ils sont traités automatiquement comme partie du type d'élément de courrier électronique.

Si vous prévoyez d'exécuter une seule instance de afuIndexer sur un type d'élément à la fois, vous devez configurer le processus d'indexation de la manière suivante pour améliorer les performances :

- Définissez **XmlDeleteFilesAfterNseProcessing** sur la valeur False.
- Supprimez le répertoire de stockage des fichiers XML générés (défini par **GenXmlDirectory**) dans le cadre du fichier batch ou du script afuIndexer à l'aide des commandes DOS DEL ou UNIX rm.

La suppression des fichiers XML dans le répertoire spécifié par **GenXmlDirectory** à l'aide de la commande du système d'exploitation est beaucoup plus rapide et se traduit par une amélioration significative des performances. Si plusieurs processus d'indexation sont en cours d'exécution sur des types d'éléments différents, vous ne devez pas supprimer les fichiers XML, car le répertoire spécifié par **GenXmlDirectory** est partagé par tous les processus d'indexation.

Avant de commencer l'indexation, assurez-vous que suffisamment d'espace disque est alloué aux fichiers XML temporaires dans le répertoire spécifié par l'option de configuration **GenXMLDirectory**, car ce répertoire est partagé par tous les processus de l'indexeur exécutés simultanément pour plusieurs types d'élément et ils peuvent donc devenir trop volumineux. Cette règle s'applique également si vous exécutez une seule instance d'afuIndexer sur un type d'élément en mode continu et que le processus s'exécute pendant longtemps.

L'indexation requiert un fichier de configuration de l'indexeur pour le type d'élément. Par défaut, l'outil de l'indexeur recherche un fichier de configuration par défaut appelé <nom_type_élément>.ini dans le répertoire en cours et le charge. Si aucun fichier de configuration par défaut n'est trouvé et qu'aucun fichier de

configuration n'est transmis en tant qu'argument lors du démarrage du processus d'indexation, le processus de l'indexeur affiche un message d'erreur et arrête le traitement.

Important :

- Le processus de l'indexeur démarre uniquement le traitement si le nombre d'éléments à indexer dans le lot d'éléments suivant correspond au moins au nombre indiqué par l'option de configuration **IdxItemMinBatchSize** ou au nombre indiqué par l'argument de ligne de commande de l'indexeur `-minbatchsize`.

La valeur par défaut de **IdxItemMinBatchSize** est 0 ; elle impose au processus de calculer automatiquement la taille minimale du lot d'éléments en attente de traitement.

- Seul le nombre d'éléments indiqué par l'option de configuration **IdxItemMaxBatchSize** ou par l'argument de ligne de commande de l'indexeur `-maxbatchsize` est traité chaque fois que l'exécution de l'indexeur est planifiée. Pour plus d'informations sur l'interaction entre **IdxItemMinBatchSize** et **IdxItemMaxBatchSize**, voir «Contraintes lors de l'indexation en mode de traitement par lots», à la page 88.

Pour extraire, convertir et indexer des documents archivés, procédez comme suit.

- Sous UNIX et Linux, exécutez :
`afuIndexer.sh`
- Sous Windows :
`afuIndexer.cmd`

Remarque : Lorsque vous exécutez le processus `afuIndexer` :

- Tous les arguments de ligne de commande font une distinction entre les majuscules et les minuscules.
- Les arguments de ligne de commande peuvent apparaître dans n'importe quel ordre.
- Utilisez les délimiteurs suivants pour les arguments qui peuvent transmettre plusieurs valeurs :
 - Sous Windows, un point-virgule (;)
 - Sous UNIX et Linux, deux-points (:)
- Le nom du type d'élément ne fait pas de distinction entre les majuscules et les minuscules.
- Les options de configuration de l'indexeur transmises via un argument ne sont pas sensibles à la casse.

Le type d'élément et le nom de la base de données du serveur de bibliothèques dans laquelle le type d'élément est stocké doit être transmis sous la forme d'arguments obligatoires au processus d'indexation.

Si vous exécutez `afuIndexer` sans indiquer d'argument, ou si vous exécutez `afuIndexer` avec l'argument `-?`, les éléments suivants sont affichés :

```
afuIndexer: Syntaxe :
[-?|-h|-copyright]
-it <type_élément>
-database <base_de_données>
[-user <ID_utilisateur|@variable_environnement|:nom_fichier>]
[-password <mot_de_passe|@variable_environnement|:nom_fichier>]
[-schema <schéma>]
```

```

[-port <numéro_port>]
[-maxbatchsize <nombre_éléments>]
[-minbatchsize <nombre_éléments>]
[-continuousmode]
[-severeerrorsmode]
[-reindexdate <aaaammjj>]
[-reindexwarnings] [-reindexerrors] [-reindexsevereerrors]
[-reindexrc <IDXRC>]
[-reindexitems <ID_élément>;<...>]
[-reindexsearch <termerecherche>]
[-reindexsearchwarnings] [-reindexsearcherrors]
[-debug <ID_élément>;<...>]
[-usemultiprocessing]
[-nseindexreorg]
[-shutdown complete|force]
[-useconfig <fichier.ini>;<...>]
[-usesettings "<option=valeur>;<...>"]
[-exportconfig <fichier.ini>]
[-listconfig]
[-verbose] [-showsql] [-trace]
[-simulate]
[-quiet]

```

Où :

- ? Répertoire tous les arguments que vous pouvez transmettre à l'outil de l'indexeur afuIndexer.
- h Répertoire tous les arguments que vous pouvez transmettre à l'outil de l'indexeur afuIndexer.

-copyright

Affiche la notice de copyright et les informations de version.

-it <type_élément>

Obligatoire.

Type d'élément Content Manager (collection de documents). Les noms de type d'élément des modèles de données intégré et composé ne font pas de distinction entre les majuscules et les minuscules.

Important : Si votre type d'élément est un type d'élément composé de courrier électronique, n'entrez pas le type d'élément de pièce jointe. Les types d'élément de pièce jointe ne peuvent pas être traités seuls mais sont traités automatiquement comme partie du type d'élément de courrier électronique.

-database <base_de_données>

Obligatoire.

Nom de la base de données du serveur de bibliothèques où le type d'élément est stocké. Le nom de la base de données peut être indiqué dans deux formats qui déterminent comment l'outil de l'indexeur se connecte à la base de données. La connexion à la base de données est mise en oeuvre à l'aide de JDBC (Java Data Base Connectivity) qui définit comment un client Java peut accéder à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Pour utiliser une connexion JDBC de type 4, le nom de la base de données de l'argument -database doit inclure un nom de serveur, un numéro de port et un nom de base de données au format <serveurbasededonnées>;<numérodeport>/<nombasededonnées>, par exemple -database localhost:50000/icm1sdb. Il est recommandé d'utiliser JDBC type 4 comme méthode d'accès à la base de données mais ce mécanisme requiert la spécification complète de l'ID utilisateur et le mot de passe.

Pour utiliser une connexion JDBC type 2, la spécification du nom de la base de données de l'argument `-database` doit correspondre à un nom de connexion DB2 existant. En d'autres termes, le nom de la base de données doit être catalogué dans DB2, par exemple `-database icm1sdb`. Comme JDBC type 2 utilise une connexion à une base de données existante, il est inutile d'indiquer un ID utilisateur et un mot de passe. Toutefois, une version 32 bits native du pilote JDBC type 2 DB2 (bibliothèque partagée) doit être disponible et définie dans le chemin d'accès aux classes pour l'utilisation de la connexion JDBC type 2.

-user *<ID_utilisateur|@variable_environnement|:nom_de_fichier>*
Facultatif.

Indiquez un ID utilisateur. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un ID utilisateur de trois manières :

1. Indiquez l'ID utilisateur directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle l'ID utilisateur est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_USERID=user8
```



```
afuIndexer -user @INDEXER_USERID
```
3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuIndexer -user :myUserID.txt
```

Le fichier texte doit contenir une seule ligne avec l'ID utilisateur (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer l'ID utilisateur dans toutes les informations d'état des processus.

L'utilisateur doit disposer des droits nécessaires pour mettre à jour les index de recherche de texte. Pour connaître les droits utilisateur requis pour mettre à jour les index de recherche de texte, voir <http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html> pour DB2 Net Search Extender version 9.5 et <http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html> pour DB2 Net Search Extender version 9.7.

-password *<mot_de_passe|@variable_environnement|:nom_fichier>*
Facultatif.

Mot de passe de l'ID utilisateur doté des droits administrateur de la base de données. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un mot de passe de trois manières :

1. Indiquez le mot de passe directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle le mot de passe est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_PASSWORD=12345678
```



```
afuIndexer -password @INDEXER_PASSWORD
```
3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuIndexer -password :myPassword.txt
```

Le fichier texte doit contenir une seule ligne avec le mot de passe (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer le mot de passe dans toutes les informations d'état des processus.

Les mots de passe ne sont pas consignés dans un format de texte normal lisible.

-schema <*schéma*>

Facultatif.

Nom du schéma de la base de données s'il est différent du schéma de la base de données en cours.

-port <*numéro_port*>

Facultatif.

Le processus de l'indexeur communique avec le processus de traitement via un port.

Le numéro de port par défaut est 4242 si aucune valeur n'est indiquée à l'aide de -port ou si la variable d'environnement **AFU_INDEXER_RMI_PORT** n'est pas définie pour l'instance.

Cependant, pour traiter plusieurs types d'éléments à la fois, chaque instance du processus de l'indexeur doit utiliser un port unique et libre ; soit en transmettant un numéro de port unique à l'aide de l'argument de ligne de commande -port, soit en définissant la variable d'environnement **AFU_INDEXER_RMI_PORT** pour chaque instance avant le démarrage de afuIndexer.

-maxbatchsize <*nombre_éléments*>

Facultatif.

Nombre maximal de documents d'un type d'élément traités pendant une exécution d'index. Une fois le nombre de documents indiqués traités, Net Search Extender est appelé. La valeur remplace la taille du lot indiquée dans l'option de configuration **IdxItemMaxBatchSize** et désactive le calcul automatique de la valeur.

-minbatchsize <*nombre_éléments*>

Facultatif.

Nombre minimal d'éléments du type d'élément requis avant l'exécution du processus de l'indexeur. La valeur remplace la taille minimale du lot indiquée dans l'option de configuration **IdxItemMinBatchSize** et désactive le calcul automatique de la valeur.

-continuousmode

Facultatif.

Le processus de l'indexeur s'exécute en boucle sans s'arrêter jusqu'à ce qu'une demande d'arrêt explicite soit émise pour le type d'élément. Pour plus d'informations, voir «Exécution de afuIndexer en mode continu», à la page 92.

-severeerrorsmode

Facultatif.

Bascule le traitement de l'indexeur vers un mode spécial dans lequel seuls les éléments ayant déjà été traités et ayant généré une valeur IDXRC 200 ou 9 sont traités de nouveau à l'aide d'une table des tâches ouvertes différente et de

paramètres de configuration optimisés pour la gestion des situations d'erreur, et ne sont pas optimisés pour les performances et un débit élevé.

Pour plus d'informations, voir «Retraitement des éléments ayant généré une erreur grave en mode spécial», à la page 106.

Les arguments suivants sont tous liés à la réindexation des documents. Ils ne peuvent pas être utilisés si l'indexeur est exécuté en mode débogage. L'ensemble des traitements de réindexation peuvent prendre un certain temps. Pour plus d'informations, reportez-vous à la rubrique sur la réindexation des documents dans la section Rubriques connexes.

-reindexdate <aaaammjj>

Facultatif.

Réindexe les documents qui ont déjà été indexés mais qui doivent l'être une nouvelle fois. Tous les documents dont la date d'archivage est <aaaammjj> ou une date ultérieure sont réindexés.

-reindexwarnings

Facultatif.

Réindexe tous les documents qui ont été indexés pendant l'exécution d'une procédure d'indexation précédente et qui ont généré un avertissement IDXRC compris entre 1 et 99.

Notez que si vous exécutez `afuIndexer` sur un type d'élément IBM CommonStore actif, l'argument `-reindexwarnings` ignore l'ensemble des éléments qui ont été indexés par l'indexeur rapide ou l'indexeur IBM Content Manager standard via l'exit utilisateur IBM Text Search et qui possèdent une valeur IDXRC importée comprise entre 10 et 19.

Pour réindexer les éléments avec une valeur IDXRC comprise entre 10 et 19, exécutez `afuIndexer` avec l'argument de ligne de commande `-reindexrc <IDXRC>`. Compte tenu que tous les éléments importés à partir d'un type d'élément IBM CommonStore actif possèdent généralement la même valeur TIEFLAG, notamment 17 ou 18, une seule exécution de l'indexeur `afuIndexer` avec l'argument de ligne de commande `-reindexrc 17` ou `-reindexrc 18` doit être suffisante.

-reindexerrors

Facultatif.

Réindexe tous les documents qui ont été indexés pendant une procédure d'indexation précédente et qui ont généré une erreur IDXRC comprise entre 100 et 199.

-reindexsevereerrors

Facultatif.

Réindexe tous les documents qui ont été indexés pendant l'exécution d'une procédure d'indexation précédente et qui ont généré une erreur grave (IDXRC = 200). Pour plus d'informations, reportez-vous à la rubrique indiquant les étapes à suivre si l'indexation s'arrête brutalement.

Cet argument de ligne de commande remplace l'argument de ligne de commande `-reindexfatalerrors` utilisé dans IBM Content Collector V2.1.1.

-reindexrc <IDXRC>

Facultatif.

Réindexe tous les documents qui ont été indexés pendant l'exécution d'une procédure d'indexation précédente et qui ont généré une valeur <IDXRC> spécifique.

-reindexitems <ID_élément>;<ID_élément>

Facultatif.

Réindexe uniquement les documents répertoriés. Répertoriez les documents (<ID_élément>) séparés par le délimiteur de votre système d'exploitation. Seuls les éléments du type qui est traité peuvent être réindexés.

Les ID d'élément peuvent être répertoriés avec et sans numéro de version. Si le numéro de version est omis, -V1 (version 1) est la valeur par défaut. La gestion de versions n'est pas appropriée pour les courriers électroniques car ceux-ci n'ont toujours qu'une seule version. Cependant, pour des documents d'application, par exemple, ou des documents Microsoft SharePoint et Système de fichiers, le numéro de version est approprié et doit être indiqué dans l'ID de l'élément pour s'assurer que le bon élément est traité.

-reindexsearch <termerecherche>

Facultatif.

Réindexe uniquement les documents qui ont été indexés pendant l'exécution d'une procédure d'indexation précédente et pour lesquels la chaîne indiquée a été indexée au document.

-reindexsearchwarnings

Facultatif.

Réindexe tous les documents qui ont été indexés pendant l'exécution d'une procédure d'indexation précédente avec une chaîne de notification d'avertissement commençant par IcmFceWarning:.

Pour plus d'informations sur les chaînes de notification d'avertissement, voir «Recherche de chaînes d'erreur», à la page 114.

-reindexsearcherrors

Facultatif.

Réindexe tous les documents qui ont été indexés pendant l'exécution d'une procédure d'indexation précédente avec une chaîne de notification d'erreur commençant par IcmFceError:.

Pour plus d'informations sur les chaînes de notification d'avertissement, voir «Recherche de chaînes d'erreur», à la page 114.

Les arguments facultatifs suivants peuvent également être transmis à l'outil de l'indexeur :

-debug <ID_élément>;<ID_élément>

Facultatif.

Bascule le traitement en mode débogage et permet de transmettre des documents individuels. Répertoriez les documents (<ID_élément>) séparés par le délimiteur de votre système d'exploitation. L'option de débogage ne peut pas être exécutée en mode de traitement par lots car l'intervention de l'utilisateur est nécessaire. Pour plus d'informations, voir «Débogage du traitement d'un élément», à la page 108.

Les ID d'élément peuvent être répertoriés avec et sans numéro de version. Si le numéro de version est omis, -V1 (version 1) est la valeur par défaut. La gestion de versions n'est pas appropriée pour les courriers électroniques car ceux-ci n'ont toujours qu'une seule version. Cependant, pour des documents

d'application, par exemple, ou des documents Microsoft SharePoint et Système de fichiers, le numéro de version est approprié et doit être indiqué dans l'ID de l'élément pour s'assurer que la bonne version de l'élément est traitée.

-usemultipleprocessing

Facultatif. Cet argument de ligne de commande peut être utilisée dans IBM Content Collector V3.0.0.1 et versions supérieures.

Permute le traitement de l'indexeur entre plusieurs unités d'exécution regroupées sous un processus de traitement unique et plusieurs processus de traitement comportant chacun une unité d'exécution unique.

Sur le système d'exploitation AIX, la création de plusieurs processus de traitement comportant chacun une unité d'exécution unique est l'option par défaut. Sur tous les autres systèmes d'exploitation, l'option par défaut consiste à créer plusieurs unités d'exécution au sein d'un seul processus de traitement.

Lorsque vous spécifiez l'option `-usemultipleprocessing`, les ressources de gestion des documents ne sont pas partagées par plusieurs documents sous le même processus. Au lieu de cela, chaque processus de traitement concentre toutes les ressources disponibles, telles que la mémoire, sur le traitement d'un document à un moment donné. De même, les problèmes qui se produisent dans le contexte d'un traitement à unités d'exécution multiples et dans Filtrage Oracle Outside In Technology peuvent être évités.

Lorsque vous basculez entre l'utilisation d'unités d'exécution multiples sur un processus de traitement unique, si l'un des processus s'arrête de manière inattendue en raison d'une pièce jointe incorrecte ou endommagée dont le traitement par Filtrage Oracle Outside In Technology n'est pas possible, un autre processus de traitement démarre automatiquement et poursuit le traitement du lot.

L'inconvénient de l'utilisation de processus de traitement multiples est que la mémoire virtuelle requise sur la machine augmente en même temps que le nombre de processus de traitement. De plus, comme le planificateur du système d'exploitation est responsable de l'attribution du temps processeur à chaque processus de traitement, la distribution des documents qui seront traités par ces processus de traitement ne peut pas être contrôlé précisément par les principaux processus de l'indexeur. La répartition des documents entre les processus de travail peut donc être inégale.

Pour plus d'informations, voir «Utilisation de plusieurs processus à unités d'exécution multiples durant le traitement et l'indexation de documents», à la page 89.

-nseindexreorg

Facultatif.

Effectue une réorganisation immédiate de l'index Net Search Extender dès que Net Search Extender a terminé l'indexation. Cet argument de ligne de commande remplace tous les calculs de réorganisation d'index définis dans d'autres emplacements.

Les arguments de ligne de commande `-nseindexreorg` et `-debug` ne peuvent pas être utilisés simultanément.

Les arguments de ligne de commande `-nseindexreorg` et `-continuousmode` ne peuvent pas être utilisés simultanément.

-shutdown *complete|force*

Facultatif.

Demande l'arrêt d'une instance de l'indexeur en cours d'exécution pour le type d'élément indiqué.

Sélectionnez *force* pour arrêter le traitement dès que possible. Si Net Search Extender n'a pas encore été démarré, il n'est pas appelé et l'ensemble des traitements effectués sont perdus car aucun élément n'est indexé. Si Net Search Extender est déjà lancé, les éléments en attente d'indexation sont tout d'abord indexés, puis le traitement est arrêté.

Sélectionnez *complete* pour arrêter le traitement dès que les éléments restants en attente de traitement ont été traités. Cette opération dépend de **IdxItemMinBatchSize**. Généralement, un maximum de 256 éléments peut encore être en attente d'indexation. Net Search Extender s'exécute.

Remarque : Le processus de l'indexeur ne s'arrête pas immédiatement même si vous exécutez un arrêt forcé. L'arrêt d'une instance en cours peut prendre un certain temps.

-useconfig <fichier.ini>;<fichier.ini>

Facultatif.

Permet d'utiliser d'autres fichiers de configuration .ini dans un autre chemin. Plusieurs fichiers .ini peuvent être indiqués. Ils sont chargés dans l'ordre dans lequel ils ont été indiqués.

-usesettings "<option=valeur>;<option=valeur>"

Facultatif.

Active le remplacement direct et temporaire des valeurs de configuration persistantes de l'indexeur stockées dans le fichier .ini. Vous pouvez ajouter un certain nombre de paires <option=valeur> séparées par le délimiteur du système d'exploitation. Les paires de valeurs d'option doivent être placées entre guillemets. Les options de configuration que vous répertoriez (<option>) ne sont pas sensibles à la casse. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de l'indexeur.

-exportconfig <fichier.ini>

Facultatif.

Exporte la configuration de l'indexeur dans le fichier .ini indiqué dans le fichier <fichier .ini>.

-listconfig

Facultatif.

Affiche toutes les options de configuration et les valeurs associées définies pour la procédure d'exécution de l'indexeur en cours. Aucune autre opération n'est effectuée.

-verbose

Facultatif.

Affiche des informations détaillées sur le traitement. La définition de l'option -verbose est utile au début pour obtenir des informations détaillées sur l'ensemble des traitements. Si l'option -verbose est spécifiée, les fichiers journaux contiennent également des informations supplémentaires.

-showsqli

Facultatif.

Affiche toutes les commandes SQL DB2 exécutées par le système. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution

de tous les traitements système DB2. Si cet argument est indiqué, toutes les instructions SQL sont également ajoutées au fichier journal.

-trace

Facultatif.

Fournit des informations plus détaillées que l'option `-verbose` en affichant des données supplémentaires dans les fichiers journaux pour effectuer l'identification des erreurs. Ces informations sont destinées uniquement au service de support logiciel IBM et requièrent l'utilisation du fichier de propriété de trace.

Important : Ne spécifiez pas l'argument de ligne de commande `-trace` en association avec `-shutdown`.

-simulate

Facultatif.

Exécute le processus de l'indexeur sans mettre à jour les bases de données. Net Search Extender ne s'exécute pas. Cet argument est utile pour vérifier le traitement des fichiers et la configuration.

-quiet

Aucune étape de traitement n'est affichée sur la console.

Vous devez surveiller en permanence le processus `afuIndexer` :

- Vérifiez les codes de retour du processus. Pour plus de détails, voir «Vérification du code retour du processus de l'indexeur», à la page 36.
- Vérifiez l'état de l'indexation consignée à la fois sur la console et dans les fichiers journaux. Pour plus de détails, voir «Surveillance de l'indexation via IBM Content Collector Text Search Support», à la page 46. Toutes les informations inscrites dans les fichiers journaux sont munies d'un préfixe prenant en charge l'évaluation automatique dans une certaine mesure.
- Consultez les erreurs de gestion de documents consignées dans la tables des tâches terminées (IDXRC >= 100) de façon régulière. Pour plus de détails, voir «Vérification du statut des éléments indexés», à la page 37.
- Traitez à nouveau les éléments associés à un IDXRC de 200 (erreur grave) en exécutant `afuIndexer` avec le paramètre `-severeerrorsmode` de façon régulière. Pour plus de détails, voir «Eléments ayant généré une erreur grave», à la page 106.

Tâches associées :

«Débogage du traitement d'un élément», à la page 108

«Réindexation de documents», à la page 103

«Débogage du traitement d'un élément», à la page 108

«Réindexation de documents», à la page 103

Référence associée :

«Informations de date/heure créées pendant l'exécution du processus de l'indexeur», à la page 53

«Fichiers journaux créés pendant le processus de l'indexeur», à la page 48

«Informations de date/heure créées pendant l'exécution du processus de l'indexeur», à la page 53

«Fichiers journaux créés pendant le processus de l'indexeur», à la page 48

Vérification du code retour du processus de l'indexeur

Le code retour du processus de l'indexeur présente le résultat du processus de l'indexeur. Si le processus de l'indexeur est automatiquement exécuté dans une planification, le code retour du processus doit être vérifié avant le début du processus de l'indexeur suivant.

IBM Content Collector fournit un exemple de script qui répertorie tous les codes retour de processus système de l'indexeur possibles requérant l'interaction de l'utilisateur avant que le processus de l'indexeur soit à nouveau appelé.

Vous pouvez utiliser l'exemple de script comme base de reconnaissance de situations d'erreurs et de détermination de l'action à prendre. Les codes d'erreur renvoyés sont regroupés comme suit.

Tableau 3. Codes retour du processus système de l'indexeur

Valeur de retour	Description
0	Indexation terminée sans erreur. Aucune action n'est requise avant l'exécution du processus d'indexeur planifié suivant.
1	Aucun élément à traiter n'a été détecté. Soit le processus est redémarré après un temps de veille défini, soit aucune action n'est prise tant que le processus de l'indexeur planifié suivant n'a pas été exécuté.
> 1	Une erreur a été détectée et doit être vérifiée avant que le planificateur du processus de l'indexeur ne soit appelé à nouveau. Il s'agit d'erreurs graves causées, par exemple, par une configuration système non valide due à l'appel incorrect du processus de l'indexeur ou à des problèmes de traitement de la base de données DB2. Remarque : Ces erreurs ne sont pas liées au traitement réel d'un élément. Pour obtenir une liste de ces erreurs, voir «Codes retour de l'index stockés dans la tables des tâches terminées», à la page 37.

Pour évaluer le code retour du processus de l'indexeur s'il est automatiquement exécuté dans une planification :

1. Définissez votre traitement d'erreurs en adaptant l'exemple de script `runAfuIndexer` dans le sous-répertoire `samples`.
2. Exécutez le script :
 - Sous UNIX et Linux, vous pouvez utiliser :
`runAfuIndexer.sh`

- Sous Windows, utilisez :
runAfuIndexer.cmd

Vérification du statut des éléments indexés

Vous devez régulièrement vérifier le statut de tous les éléments indexés pour déterminer quels sont ceux qui n'ont pas été indexés avec succès car ces derniers ne peuvent pas être recherchés et sont introuvables dans le référentiel.

Après traitement, l'élément est ajouté à la table des tâches terminées avec une valeur IDXRC. Les éléments contenus dans la table des tâches terminées sont regroupés comme suit.

- Les éléments ayant été indexés avec succès ont une valeur IDXRC de 0.
- Les éléments ayant été indexés avec un avertissement, ont une valeur IDXRC comprise entre 1 et 99. Ils sont indexés avec une chaîne de notification d'avertissement.
- Les éléments ayant produit une erreur et qui n'ont pas été indexés ont une valeur IDXRC comprise entre 100 et 199.
- Les éléments ayant produit une erreur grave et qui n'ont pas été indexés ont une valeur IDXRC de 200.

Vérifiez le statut d'indexation des éléments en consultant la colonne IDXRC de la table de base de données des tâches terminées pour déterminer le nombre d'éléments ayant provoqué un avertissement, une erreur ou une erreur grave et étant probablement la cause de ces avertissements et erreurs. Les noms de la table des tâches ouvertes et de la table des tâches terminées sont toujours affichés lorsque le processus afuIndexer ou afuCheck est exécuté. Le nom de la table des tâches terminées s'affiche également une fois le processus afuIndexer terminé, ou peut également figurer dans le fichier journal.

1. Demandez le nombre d'éléments avec avertissements, erreurs ou erreurs graves.
 - Pour tous les avertissements, entrez :
db2 "select * from <tableOfCompletedTasks> où IDXRC >= 1 et idxrc < 100"
 - Pour toutes les erreurs, entrez :
db2 "select * from <tableOfCompletedTasks> où IDXRC >= 100 et idxrc < 200"
 - Pour toutes les erreurs graves, entrez :
db2 "select * from <tableOfCompletedTasks> où IDXRC = 200"
 - Pour les valeurs IDXRC de tous les éléments, entrez :
db2 "select IDXRC, count(*) from <tableOfCompletedTasks> group by IDXRC"
2. Réindexez les éléments avec avertissements ou erreurs.

Codes retour de l'index stockés dans la tables des tâches terminées

Tous les éléments traités sont ajoutés à la table des tâches terminées avec une valeur IDXRC indiquant si l'élément a été indexé avec succès.

Une valeur IDXRC de 0 signifie que l'élément a été traité et indexé avec succès.

Une valeur IDXRC comprise entre 1 et 99 indique un avertissement. Un avertissement signifie que l'élément a été traité et qu'il a été correctement ajouté à l'index mais qu'il n'a peut-être pas été totalement indexé. Certaines informations peuvent être manquantes en fonction de la valeur IDXRC.

Une valeur IDXRC comprise entre 100 et 199 indique une erreur. Une erreur signifie que l'élément a été traité mais qu'il n'a PAS été ajouté à l'index en raison d'un problème qui est indiqué par la valeur IDXRC.

Une valeur IDXRC de 200 indique une erreur grave. Une erreur grave signifie que l'élément n'a pas pu être traité et qu'il n'a pas pu être ajouté à l'index car le processus de travail s'est arrêté de manière inattendue.

Le terme "document" dans les descriptions ci-après désigne les pièces jointes imbriquées dans des messages électroniques et les documents du système de fichiers et de SharePoint.

Les valeurs IDXRC 10 à 19 correspondent aux valeurs de TIEFLAG utilisées dans IBM CommonStore et décrivent quelles parties d'un document sont indexées pour chaque valeur de TIEFLAG. Les valeurs 10, 11, 12, 19 et n'apparaissent pas dans cette table, car même si elles ont été réservées en tant que valeurs de TIEFLAG par IBM CommonStore, elles n'ont jamais été affectées.

Tableau 4. Codes retour de l'indexation pour la réindexation des éléments

IDXRC	Description des codes retour
0	L'élément a été indexé. Action de l'utilisateur : Aucune action requise.
1	L'élément a été indexé mais un ou plusieurs de ses documents n'ont pas été intégralement indexés car leur contenu a été tronqué en raison d'une quantité de mémoire insuffisante. Action de l'utilisateur : Essayez d'augmenter la valeur de l'option de configuration TxtCnvTargetBufferSizeInMB et de retraiter les éléments.
2	L'élément a été indexé mais un ou plusieurs de ses documents n'ont pas été indexés car ils n'ont pas pu être soumis à l'analyse syntaxique des Filtres Oracle Outside In Technology. Il est également possible que ces documents n'aient pas été considérées comme des documents de texte normal car ils contenaient une quantité de caractères étendus (point de code > 0x7F) supérieure à celle définie par l'option de configuration TxtCnvMaxExtendedCharactersRatio . Action de l'utilisateur : Essayez de diminuer la valeur de l'option de configuration TxtCnvMaxExtendedCharactersRatio et de retraiter les éléments.
3	L'élément a été indexé mais un ou plusieurs de ses documents ou de leur contenu n'ont pas été intégralement indexés car ils possèdent une page de codes qui n'est pas prise en charge. Action de l'utilisateur : La page de codes utilisée dans les éléments n'est pas prise en charge, ou la spécification de la page de codes n'est pas standard et ne peut donc pas être reconnue. Contactez le support logiciel IBM.
4	L'élément a été indexé mais un ou plusieurs de ses documents n'ont pas été indexés car ils n'ont pas pu être traités par les Filtres Oracle Outside In Technology. Aucune autre tentative n'a été effectuée pour examiner ou réindexer ces documents car l'option de configuration TxtCnvPlainTextHandlingForUnsupportedFormat correspond à 0, ce qui signifie qu'aucune tentative n'a été effectuée pour traiter le document en tant que document de texte normal. Action de l'utilisateur : Tant que TxtCnvPlainTextHandlingForUnsupportedFormat est défini sur 0, aucune conversion de document n'a lieu. Si cette option est définie sur 1, une tentative de traitement du document aura lieu en tant que document texte normal.

Tableau 4. Codes retour de l'indexation pour la réindexation des éléments (suite)

IDXRC	Description des codes retour
5	<p>L'élément a été indexé mais un ou plusieurs de ses documents n'ont pas été indexés car ils n'ont pas pu être gérés par les Filtres Oracle Outside In Technology. Il est également possible que ces documents n'aient pas été considérés comme des documents de texte normal car ils contenaient au moins un caractère de contrôle (point de code < 0x6F) qui indique un contenu non textuel.</p> <p>Action de l'utilisateur : Ces documents ne seront jamais convertis car ils contiennent des caractères de contrôle.</p>
6	<p>L'élément a été indexé mais un ou plusieurs de ses documents n'ont pas été indexés car ces documents possèdent une taille de fichier de 0 octets.</p> <p>Action de l'utilisateur : Ces documents ne seront jamais convertis en texte car ils sont vides.</p>
7	<p>L'élément a été indexé mais un ou plusieurs documents n'ont pas été indexés car les Filtres Oracle Outside In Technology ont généré une erreur grave lors de l'analyse syntaxique.</p> <p>Action de l'utilisateur : Le document peut être corrompu ou autrement défectueux, ce qui conduit à une erreur grave. Le document peut être envoyé au support Oracle pour une analyse approfondie. Contactez le support logiciel IBM.</p>
9	<p>L'élément a été indexé mais les pièces jointes imbriquées n'ont pas été traitées et leur contenu n'a pas été indexé car les pièces jointes imbriquées ont provoqué un arrêt inattendu du processus de travail lors d'une exécution précédente de l'indexeur et l'option de configuration IdxProcessSevereErrors a été activée. Pour plus de détails, voir «Eléments ayant généré une erreur grave», à la page 106.</p> <p>Action de l'utilisateur : Les éléments peuvent être de nouveau traités à l'aide du mode de débogage de l'indexeur afin d'identifier la raison de l'échec du traitement des pièces jointes. Les documents peuvent ensuite être envoyés au support Oracle pour une analyse approfondie. Contactez le support logiciel IBM.</p>
13	<p>Correspond à la valeur originale TIEFLAG 13, qui signifie que le type d'élément n'a pas été précédemment indexé par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Une valeur TIEFLAG de 13 indique que seuls les attributs de documents tels que Objet, A, ou DE, qui sont dénotés par l'indicateur IBM CommonStore CS_ATTR et les attributs IBM Content Manager, tels que CSORIGINATOR ou CSLDOrigDB, qui sont dénotés par l'indicateur IBM CommonStore CS_CMATTR, ont été ajoutés à l'index. Le corps du document, ainsi que les éventuelles pièces jointes, n'ont pas été ajoutées à l'index.</p> <p>Action de l'utilisateur : envisagez de réindexer ces documents, car l'indexeur IBM Content Collector pour la recherche de texte traite les documents de façon plus détaillée que l'indexeur rapide ou que l'indexeur IBM Content Manager standard ne sont capables de le faire.</p>
14	<p>Correspond à la valeur originale TIEFLAG 14, qui signifie que le type d'élément n'a pas été précédemment indexé par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Une valeur TIEFLAG de 14 indique que seul le corps de document, qui est dénoté par l'indicateur IBM CommonStore CS_BODY, et les pièces jointes dénotées par l'indicateur IBM CommonStore CS_ATTACH, ont été ajoutés à l'index. Aucun attribut de document ni attribut IBM Content Manager n'est ajouté à l'index.</p> <p>Action de l'utilisateur : envisagez de réindexer ces documents, car l'indexeur IBM Content Collector pour la recherche de texte traite les documents de façon plus détaillée que l'indexeur rapide ou que l'indexeur IBM Content Manager standard ne sont capables de le faire.</p>

Tableau 4. Codes retour de l'indexation pour la réindexation des éléments (suite)

IDXRC	Description des codes retour
15	<p>Correspond à la valeur originale TIEFLAG 15, qui signifie que le type d'élément n'a pas été précédemment indexé par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Une valeur TIEFLAG de 15 indique que seuls les attributs de documents tels que Objet, A, ou DE, qui sont dénotés par l'indicateur IBM CommonStore CS_ATTR, le corps de document dénoté par l'indicateur IBM CommonStore CS_BODY, ainsi que les attributs IBM Content Manager, tels que CSORIGINATOR ou CSLDOrigDB, qui sont dénotés par l'indicateur IBM CommonStore CS_CMATTR, ont été ajoutés à l'index. Aucune pièce jointe n'est ajoutée à l'index.</p> <p>Action de l'utilisateur : envisagez de réindexer ces documents, car l'indexeur IBM Content Collector pour la recherche de texte traite les documents de façon plus détaillée que l'indexeur rapide ou que l'indexeur IBM Content Manager standard ne sont capables de le faire.</p>
16	<p>Correspond à la valeur originale TIEFLAG 16, qui signifie que le type d'élément n'a pas été précédemment indexé par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Une valeur TIEFLAG de 16 indique que seuls les attributs de documents tels que Objet, A, ou DE, qui sont dénotés par l'indicateur IBM CommonStore CS_ATTR, et le corps de document dénoté par l'indicateur IBM CommonStore CS_BODY, ont été ajoutés à l'index. Aucune pièce jointe et aucun attribut IBM Content Manager ne sont ajoutés à l'index.</p> <p>Action de l'utilisateur : envisagez de réindexer ces documents, car l'indexeur IBM Content Collector pour la recherche de texte traite les documents de façon plus détaillée que l'indexeur rapide ou que l'indexeur IBM Content Manager standard ne sont capables de le faire.</p>
17	<p>Correspond à la valeur originale TIEFLAG 17, qui signifie que le type d'élément n'a pas été précédemment indexé par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Une valeur TIEFLAG de 17 indique que seuls les attributs de documents tels que Objet, A, ou DE, qui sont dénotés par l'indicateur IBM CommonStore CS_ATTR, le corps de document dénoté par l'indicateur IBM CommonStore CS_BODY, ainsi que les pièces jointes, dénotées par l'indicateur IBM CommonStore CS_ATTACH, ont été ajoutés à l'index. Aucune occurrence de IBM Content Manager n'est ajoutée à l'index.</p> <p>Action de l'utilisateur : envisagez de réindexer ces documents, car l'indexeur IBM Content Collector pour la recherche de texte traite les documents de façon plus détaillée que l'indexeur rapide ou que l'indexeur IBM Content Manager standard ne sont capables de le faire.</p>
18	<p>Correspond à la valeur originale TIEFLAG 18, qui signifie que le type d'élément n'a pas été précédemment indexé par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Une valeur TIEFLAG de 18 indique que le document complet, y compris les attributs de document tels que Corps, Objet, A ou DE, les attributs IBM Content Manager tels que CSORIGINATOR ou CSLDOrigDB, ainsi que les pièces jointes, ont été ajoutés à l'index.</p> <p>Action de l'utilisateur : envisagez de réindexer ces documents, car l'indexeur IBM Content Collector pour la recherche de texte traite les documents de façon plus détaillée que l'indexeur rapide ou que l'indexeur IBM Content Manager standard ne sont capables de le faire.</p>
100	<p>L'élément n'a pas été indexé car Net Search Extender a signalé un code de motif.</p> <p>Action de l'utilisateur : Une erreur Net Search Extender inattendue s'est produite. Contactez le support logiciel IBM.</p>

Tableau 4. Codes retour de l'indexation pour la réindexation des éléments (suite)

IDXRC	Description des codes retour
101	<p>L'élément n'a pas été indexé car le composant Net Search Extender a signalé un code d'erreur 101 pour l'élément.</p> <p>Action de l'utilisateur : Une erreur Net Search Extender inattendue s'est produite. Contactez le support logiciel IBM.</p>
102	<p>L'élément n'a pas été indexé car le composant Net Search Extender a signalé un code d'erreur 102 pour l'élément.</p> <p>Action de l'utilisateur : Une erreur Net Search Extender inattendue s'est produite. Contactez le support logiciel IBM.</p>
103	<p>L'élément n'a pas été indexé car le composant Net Search Extender a signalé un code d'erreur 103 pour l'élément.</p> <p>Action de l'utilisateur : L'opération ne peut pas être exécutée car le moteur de recherche de texte a détecté que la structure d'index ou les ensembles de fichiers d'index sont corrompus. Contactez le support logiciel IBM.</p>
104	<p>L'élément n'a pas été indexé car le composant Net Search Extender a signalé un code d'erreur 104 ou 105 pour l'élément.</p> <p>Action de l'utilisateur : Une erreur Net Search Extender inattendue s'est produite. Contactez le support logiciel IBM.</p>
105	<p>L'élément n'a pas été indexé car le composant Net Search Extender a signalé un code d'erreur qui ne correspond pas à l'un des codes ci-dessus pour l'élément.</p> <p>Action de l'utilisateur : Contactez le support logiciel IBM.</p>
110	<p>L'élément n'a pas été indexé car son délai de téléchargement à partir du gestionnaire de ressources a dépassé le délai indiqué.</p> <p>Action de l'utilisateur : Vérifiez que les opérations du gestionnaire de ressources IBM Content Manager sont correctes et augmentez la valeur de l'option de configuration GenRmConnectionTimeOutSecs.</p>
111	<p>L'élément n'a pas été indexé car le délai de traitement des Filtres Oracle Outside In Technology a été dépassé.</p> <p>Action de l'utilisateur : Essayez d'augmenter la valeur de l'option de configuration TxtCnvTimeOutInSecs et de retraiter les éléments.</p>
120	<p>L'élément n'a pas été indexé car une erreur s'est produite lors de l'extraction de l'élément du gestionnaire de ressources.</p> <p>Action de l'utilisateur : Consultez les messages d'erreur dans le fichier journal du gestionnaire de ressources IBM Content Manager.</p>
121	<p>L'élément n'a pas été indexé car un élément de 0 octet a été extrait du gestionnaire de ressources. Cet événement indique généralement une erreur du gestionnaire de ressources car tous les éléments archivés par IBM Content Collector possèdent une taille supérieure à 0 octets.</p> <p>Notez que si la taille d'un document dans un type d'élément qui utilise le modèle de données pour les documents IBM Connections, est de zéro octet, le document est traité sans donner lieu à un IDRXC de 121. Dans ce cas, un IDXRC de 6 est affecté au document et la chaîne de recherche <code>IcmFceError: IcmEmptyFile</code> est indexée avec le document.</p> <p>Action de l'utilisateur : Consultez les messages d'erreur dans le fichier journal du gestionnaire de ressources IBM Content Manager.</p>

Tableau 4. Codes retour de l'indexation pour la réindexation des éléments (suite)

IDXRC	Description des codes retour
122	<p>L'élément n'a pas été indexé car le composant de données de l'élément BRI qui a été extrait du gestionnaire de ressources IBM Content Manager est de zéro octet. Cela indique généralement une erreur du gestionnaire de ressources IBM Content Manager ou une erreur d'archivage car tous les éléments archivés par IBM Content Collector ou par IBM CommonStore ont toujours une taille de données supérieure à zéro octet.</p> <p>Action de l'utilisateur : Consultez les messages d'erreur dans le fichier journal du gestionnaire de ressources IBM Content Manager.</p>
123	<p>L'élément n'a pas été indexé car l'élément n'inclut aucun composant de document de contenu. Cette erreur peut se produire si l'élément est un élément de dossier. Les éléments de dossier ne peuvent pas être indexés car ils n'incluent aucun contenu.</p>
130	<p>L'élément n'a pas été indexé car le contenu textuel d'un élément dépasse la taille de mémoire tampon de transfert de la fonction définie par l'utilisateur AFUFetchFile.</p> <p>Action de l'utilisateur : Reportez-vous à «Modification de la taille de la mémoire tampon de transfert par défaut de la fonction définie par l'utilisateur AFUFetchFile», à la page 87.</p>
140	<p>L'élément n'a pas été indexé car il n'inclut aucune valeur TIEREF valide.</p> <p>Action de l'utilisateur : Exécutez l'outil de l'indexeur appelé afuIndexTool avec les arguments de validation de réparation -opValidate, -deepvalidation et -repairTierref. Lorsque l'argument -repairTierref est spécifié pendant la validation, une nouvelle valeur TIEREF est calculée pour l'élément sans valeur TIEREF et cette valeur est enregistrée dans la table de composant de type d'élément dans IBM Content Manager. Pour plus de détails, voir «Validation d'index et opérations de réparation», à la page 116.</p>
180	<p>L'élément a été traité précédemment par l'indexeur rapide mais n'a pas pu être traité correctement et n'a donc pas été ajouté à l'index de texte intégral. En conséquence, il s'est vu attribuer une valeur TIEFLAG supérieure à 10 000 par l'indexeur rapide. Pour prendre en charge la réindexation par l'indexeur IBM Content Collector de recherche de texte, l'élément a été ajouté à la table des tâches terminées la valeur IDXRC de 180 lui a été attribuée.</p> <p>Action de l'utilisateur : Retraitez les éléments en exécutant l'outil afuIndexer avec -reindexrc 180.</p>
190	<p>L'élément n'a pas été indexé car il lui manque une valeur d'attribut IBM Content Manager obligatoire (ICCMailboxID, CSORGINATOR ou CSLDOrigDB par exemple).</p> <p>Action de l'utilisateur : Vérifiez que ces attributs d'éléments sont correctement extraits et écrits dans IBM Content Manager dans vos routes de tâches IBM Content Collector.</p>
199	<p>L'élément n'a pas été indexé car une erreur s'est produite lors du traitement de l'élément et elle ne correspond pas à l'une des erreurs ci-dessus.</p> <p>Action de l'utilisateur : L'erreur est grave et l'élément ne peut pas être réparé par l'un des outils de réparation de l'indexeur. Cette erreur peut également se produire après l'exécution d'une opération de validation indiquant une erreur grave.</p> <p>Retraitez les éléments en exécutant l'outil afuIndexer avec -reindexrc 199. Si le problème persiste, contactez le support logiciel IBM.</p>

Tableau 4. Codes retour de l'indexation pour la réindexation des éléments (suite)

IDXRC	Description des codes retour
200	<p>L'élément n'a pas été indexé car il était en cours de traitement lors de l'arrêt ou de l'échec du processus de travail de l'indexeur. L'élément est peut-être la cause de cette erreur inattendue. L'élément n'est pas retraité.</p> <p>Action de l'utilisateur : Consultez «Eléments ayant généré une erreur grave», à la page 106.</p>

Indexation des types d'élément non créés sous Content Collector

Vous pouvez utiliser l'indexeur IBM Content Collector pour la recherche de texte en vue d'indexer les types d'élément qui n'ont pas été créés à l'aide de Content Collector.

Il se peut que ces types d'élément aient été créés manuellement dans IBM Content Manager pour l'archivage dans IBM CommonStore ou qu'ils aient été créés pour l'archivage dans Content Collector.

Indexation des éléments qui ont été archivés à l'aide de IBM CommonStore

Utilisez l'indexeur IBM Content Collector de recherche de texte pour indexer les éléments de types d'élément qui ont été archivés à l'aide de IBM CommonStore.

Grâce à ses performances nettement améliorées et d'autres avantages de traitement, vous pouvez utiliser l'indexeur Content Collector de recherche de texte afin d'indexer de manière régulière les nouveaux éléments archivés au lieu de l'indexeur rapide ou de l'indexeur IBM Content Manager standard avec l'exit utilisateur IBM Text Search disponible dans IBM CommonStore.

Les éléments prérequis suivants s'appliquent :

1. Les types d'élément IBM CommonStore pour les modèles de données suivants peuvent être indexés à l'aide de l'indexeur IBM Content Collector de recherche de texte :
 - BUNDLED et le type d'archivage Entier
 - DOCUMENT (GENERIC_MULTIDOC et GENERIC_MULTIPART) et le type d'archivage Entier et Pièce jointe.
2. Les types d'élément IBM CommonStore pour le modèle de données BUNDLED ne peuvent être traités que si l'attribut CSLDOrigDB, CSLDOrigUser ou CSORIGINATOR est défini pour le type d'élément. CSLDOrigDB et CSLDOrigUser sont utilisés dans les types d'élément CommonStore pour Lotus Domino. CSORIGINATOR est utilisé dans les types d'élément CommonStore pour Exchange Server.
3. Les types d'élément IBM CommonStore pour le modèle de données DOCUMENT ne peuvent être traités que si l'attribut CSLDOrigDB, CSLDOrigUser ou CSORIGINATOR est défini pour le type d'élément, ou si la variable d'environnement AFU_INDEXER_USE_IT_CS_GENERIC est définie sur 1 ou sur true. Pour plus de détails sur la variable d'environnement, voir «Variables d'environnement de l'indexeur de texte», à la page 63.

Si aucun de ces attributs n'est défini et si la variable d'environnement n'est pas définie, un message d'erreur semblable au suivant s'affiche :

The data model of the 'xxx' item type is of type 'UNKNOWN'.

Pour plus d'informations, voir «Impossible d'activer les types d'élément de courrier électronique IBM CommonStore via afuEnableItemType», à la page 151.

4. IBM Content Collector Text Search Support doit être installé sur le serveur IBM Content Manager. Aucun autre composant IBM Content Collector ne doit être installé.

A faire : IBM Content Manager utilise des types MIME au lieu de types de contenu pour définir comment afficher le contenu archivé. Avant d'utiliser la recherche de texte Content Collector, vous devez mapper les types de contenu pour les extensions de fichier que vous avez utilisées dans IBM CommonStore avec les types MIME correspondants dans IBM Content Manager. Pour activer l'indexation en texte intégral de ces types MIME, vous devez également vérifier que ces types MIME sont définis comme activés pour la recherche de texte. Si l'indexeur détecte un élément dans le type d'élément et que ses composants ne permettent pas la recherche de texte, l'élément n'est pas traité et est ajouté à la table des tâches terminées avec une valeur IDXRC de 140.

Avant de commencer l'indexation des éléments dans Content Collector qui ont été archivés à l'aide de IBM CommonStore, vous devez activer les types d'élément CommonStore pour l'indexation dans Content Collector.

Après l'activation d'un type d'élément CommonStore, vous pouvez toujours archiver dans le type d'élément à l'aide de IBM CommonStore mais vous ne pouvez pas indexer les éléments à l'aide de l'indexeur rapide ou de l'indexeur IBM Content Manager standard. Vous ne pouvez indexer les éléments qu'à l'aide de l'indexeur de recherche de texte dans IBM Content Collector.

Pour indexer des éléments dans un type d'élément IBM CommonStore à l'aide de l'indexeur de recherche de texte :

1. Arrêtez l'archivage dans le type d'élément IBM CommonStore.
2. Exécutez l'indexeur rapide ou l'indexeur IBM Content Manager standard avec l'exit utilisateur IBM Text Search pour effectuer le traitement de tous les éléments qui ont été mis en file d'attente pour indexation.
3. Activez le type d'élément IBM CommonStore pour l'indexation à l'aide de l'indexeur Content Collector de recherche de texte.
4. Créez une configuration d'index pour le type d'élément à l'aide d'afuConfig.
5. Pour utiliser les valeurs de la colonne TIEFLAG dans IBM Content Manager dans les applications telles que Email Search ou eDiscovery Manager pour déterminer si l'indexation a réussi :
 - a. Définissez **IdxUpdateItemTypeTable** sur 1 dans le fichier de configuration de l'indexeur nouvellement créé afin de mettre à jour la colonne TIEFLAG avec les valeurs IDXRC de l'indexeur Content Collector,
Notez que si vous souhaitez mettre à jour la colonne TIEFLAG avec les anciennes valeurs TIEFLAG compatibles avec l'indexeur rapide (par exemple, avec des valeurs supérieures à 10 000 indiquant une situation d'erreur), vous devez contacter le support logiciel IBM.
 - b. Exécutez `afuIndexTool -opExportToItemType` pour exporter les valeurs IDXRC qui ont été créées lorsque le type d'élément a été activé pour la colonne TIEFLAG IBM Content Manager.
6. Redémarrez l'archivage dans le type d'élément dans IBM CommonStore.
7. Exécutez `afuIndexer` avec `-reindexrc 180` une fois pour tenter de réindexer tous les éléments qui n'ont pas pu être indexé par l'indexeur rapide.
8. Planifiez l'exécution du script `afuIndexer` sur le type d'élément activé.

Indexation des éléments dans un type d'élément créé sous IBM Content Manager

Pour indexer les éléments d'un type d'élément que vous avez créé manuellement dans IBM Content Manager à l'aide de l'indexeur d'IBM Content Collector pour la recherche de texte, vous devez d'abord créer un index documentaire par le biais du fichier modèle approprié.

Lorsque vous créez un type d'élément manuellement dans IBM Content Manager et que vous souhaitez indexer les éléments par l'intermédiaire de l'indexeur de recherche de texte, vous ne pouvez pas activer le type d'élément pour l'indexation et commencer ensuite le processus d'indexation. Si un type d'élément n'est pas créé via IBM Content Collector, un index documentaire n'est généralement pas créé automatiquement pour ce type d'élément. Lorsque vous démarrez le processus d'indexation, le message d'erreur de configuration suivant peut indiquer qu'un index est manquant :

```
A configuration or setup error has occurred with the following reason:  
The Net Search Extender '<indexpath>/NODE0000/IX#####.mox' model file could not be loaded
```

où ##### correspond au numéro d'index.

Ne créez pas l'index vous-même à l'aide des outils Net Search Indexer. Utilisez l'outil `afuRecreateIndex` qui peut être utilisé à la fois pour recréer un index corrompu et créer un index initial sans tenir compte du fait que l'index n'existe pas. `afuRecreateIndex` effectuera les opérations Net Search Extender nécessaires pour créer un nouvel index documentaire.

Lorsque vous créez l'index à l'aide de l'option `afuRecreateIndex`, n'oubliez pas d'ajouter le fichier modèle à utiliser en indiquant l'argument de ligne de commande `-modelfile`.

Le fichier modèle contient des définitions de zone indiquant les données de document (attributs et métadonnées de document) ou les attributs IBM Content Manager à inclure lors de la génération de l'index du type d'élément. N'oubliez pas que les définitions d'attribut d'un fichier modèle doivent correspondre aux définitions d'attribut du fichier de configuration de l'indexeur correspondant que vous créez pour le type d'élément.

Si vous oubliez d'indiquer un nouveau fichier modèle correct, `afuRecreateIndex` tente de réutiliser le fichier modèle par défaut associé à un index inexistant et l'opération échoue.

Pour indexer les éléments d'un type d'élément créé manuellement dans IBM Content Manager, procédez comme suit.

1. Activez le type d'élément IBM Content Manager pour l'indexation à l'aide de l'indexeur Content Collector pour la recherche de texte.
2. Définissez une configuration d'index pour le type d'élément activé.
3. Utilisez l'outil `afuRecreateIndex` pour créer un index.
Rappelez-vous d'indiquer un fichier modèle. Ne tenez pas compte du message indiquant qu'un index Net Search Extender existant ne peut pas être trouvé.
4. Planifiez les éléments d'indexation du type d'élément.

Surveillance de l'indexation via IBM Content Collector Text Search Support

Vous pouvez surveiller le processus et les performances d'indexation.

IBM Content Collector fournit les méthodes suivantes pour accumuler et exploiter les informations de traitement à des fins de surveillance :

Consignation

La consignation est activée à l'aide de l'option de configuration **IdxUseLogging**. Si la consignation est activée, des fichiers journaux (journaux d'événements, journaux système et fichiers journaux de trace) sont créés pour tous les types.

Chaque ligne des fichiers journaux contient un horodatage, un niveau de gravité, un ID d'unité d'exécution et un préfixe d'opération. Les informations de ces fichiers journaux peuvent être extraites à des fins de surveillance, par exemple à l'aide de l'outil grep.

Vous pouvez filtrer les informations que vous souhaitez surveiller en recherchant des préfixes d'opération spécifiques, par exemple vous pouvez extraire toutes les informations de traitement des éléments des fichiers journaux en recherchant les entrées dotées du préfixe **ITEM:** ou extraire les statistiques de traitement globales en recherchant les entrées dotées du préfixe **STATISTICS:**.

Si l'argument de ligne de commande **-verbose** est indiqué en plus de la consignation lorsque l'un des outils de l'indexeur (**afuIndexer**, **afuCreateIndex**, **afuConfig**, **afuEnableItemType** ou **afuCheck**) est exécuté, des informations de traitement supplémentaires sont affichées sur la console ou ajoutés aux fichiers journaux.

Affichage

Vous pouvez surveiller directement le processus de l'indexeur à l'écran en activant les options de configuration **IdxDisplayProgressIndicator**, **IdxDisplayPerformanceMonitor** et **IdxDisplayNSEMonitor**. L'affichage visuel peut être utile, notamment lors de la phase de configuration du système pour vérifier que le processus de l'indexeur s'exécute correctement.

Si l'argument de ligne de commande **-verbose** est indiqué et que l'option de configuration **IdxDisplayProgressIndicator** est activée, des informations de progression détaillées sont affichées lors de l'exécution du processus de l'indexeur.

Fonction de date/heure

Si la fonction de date/heure est activée en définissant l'option de configuration de l'indexation **IdxUseTiming**, des informations de date/heure supplémentaires s'affichent lors de l'exécution du processus de l'indexeur.

Si les fonctions de date/heure et de consignation sont activées, les informations de date/heure sont affichées mais également consignées dans les fichiers journaux.

En outre, si l'option de date/heure est activée, un fichier de date/heure est créé avec des valeurs de date/heure spécifiques pour chaque élément. Le fichier de date/heure possède un format CSV (Comma Separated Values) qui facilite l'extraction d'informations à des fins de surveillance.

Si vous souhaitez surveiller et mesurer des données de performances, activez **IdxDisplayPerformanceMonitor**. Les données de performances sont affichées et consignées dans les fichiers journaux (avec le préfixeTime:).

Sélection des informations à surveiller

Il est recommandé de toujours activer la consignment lors du traitement de l'indexeur en définissant l'option de configuration **IdxUseLogging** sur 1. Toutefois, la fonction de consignment n'est pas indispensable car les informations relatives au traitement des éléments d'un type sont également consignées dans la table des tâches terminées. La seule différence est que la table des tâches terminées ne contient pas de détails sur les erreurs. Vous pouvez activer la consignment lorsque vous configurez le composant de prise en charge de la recherche de texte.

Les options de configuration à utiliser pour surveiller le traitement diffèrent selon que vous configurez toujours le système ou que vous l'exécutez en mode de production. Les paramètres suivants sont recommandés.

1. Pour surveiller les informations de traitement pendant la configuration du composant de prise en charge de la recherche de texte :
 - a. Activez les options de configuration suivantes :
 - 1) **IdxUseLogging**
 - 2) **IdxDisplayProgressIndicator**
 - 3) **IdxDisplayPerformanceMonitor**
 - 4) **IdxDisplayNSEMonitor**
 - 5) **UdfUseLogging**
 - b. Utilisez l'argument de traitement de l'indexeur `-verbose`.
2. Pour surveiller des informations de traitement une fois l'environnement de l'indexeur correctement configuré :
 - a. N'utilisez pas l'argument de traitement de l'indexeur `-verbose`
 - b. Si vous ne souhaitez pas recevoir de commentaires en retour ou que le processus de l'indexeur est planifié pour s'exécuter en mode de traitement par lots inattendu, vous pouvez désactiver les options de configuration suivantes :
 - 1) **IdxDisplayProgressIndicator**
 - 2) **IdxDisplayPerformanceMonitor**
 - 3) **IdxDisplayNSEMonitor**
 - c. Si vous souhaitez recevoir des commentaires en retour de temps en temps, tout particulièrement si vous exécutez le processus de l'indexeur en mode continu, à l'inverse du mode de traitement par lots, il est recommandé d'activer :
 - 1) **IdxDisplayNSEMonitor**

L'option **IdxDisplayProgressIndicator** est automatiquement activée si vous exécutez le processus de l'indexeur en mode continu.
 - d. Il est recommandé de désactiver **UdfUseLogging** pour que le fichier journal de la fonction définie par l'utilisateur ne devienne pas trop gros.
3. Pour surveiller des données de performances, activez :
 - a. **IdxUseTiming**
 - b. **IdxDisplayPerformanceMonitor**

Le cas échéant, vous pouvez modifier les options de configuration qui déterminent les intervalles de ces surveillances. Les options sont les suivantes : **IdxDisplayProgressDelaySecs**, **IdxDisplayPerformanceMonitorDelaySecs** et **IdxDisplayNseMonitorDelaySecs**.

Fichiers journaux créés pendant le processus de l'indexeur

Vous pouvez activer la consignation en associant l'option de configuration **IdxUseLogging** à la valeur true.

Si la consignation est activée dans la configuration de l'indexeur, des fichiers journaux sont créés dans le répertoire défini par l'option de configuration **GenLogDirectory**. Le nom des fichiers journaux varie en fonction du paramètre de l'option de configuration **GenLogFileNameTimeStamp**.

Le format général des noms de fichier journal est `<itemTypeName><GenLogFileNameTimeStamp>.<type>.[<generation>].log`. Les fichiers journaux suivants sont générés :

- Fichiers journaux avec tous les messages associés au traitement des éléments et des erreurs.
- Fichiers journaux de trace. Recommandé uniquement sur demande du support logiciel IBM.

Si l'argument de ligne de commande-verbose est indiqué lorsque vous appelez l'un des outils de l'indexeur, un fichier journal détaillé doté du suffixe verbose.log et contenant les messages système et les messages détaillés est généré avec les fichiers journaux système.

En outre, le processus de travail crée son propre jeu de fichiers journaux en utilisant les mêmes noms de fichier que les fichiers journaux de l'indexeur, mais en y ajoutant l'infixe worker. Si un processus de travail est défaillant, un nouvel ensemble de fichiers journaux comportant l'infixe worker1 est créé, par exemple, après la première défaillance.

Si l'argument de ligne de commande -verbose n'est pas indiqué, seuls les fichiers journaux système sont créés lorsque la consignation est activée.

Si l'indexeur est exécuté en mode de traitement par lots, un seul jeu de fichiers journaux est créé pour chaque indexeur exécuté contenant toutes les informations de consignation. Cela signifie que, en mode de traitement par lots, les fichiers journaux ne sont pas tronqués ou limités en taille et que leur taille peut augmenter.

Si d'autre part l'indexeur est exécuté en mode continu, un marqueur de génération est automatiquement ajouté aux noms des fichiers journaux. Le marqueur de génération est indiqué par un caractère dièse (#) suivi d'un nombre. Lorsqu'un fichier journal atteint une taille de 128 Mo, il est fermé et la consignation se poursuit dans un nouveau fichier journal avec un nouveau numéro. La date et l'horodatage de tous les fichiers journaux sont toutefois identiques. Par exemple, avant l'arrêt du traitement, les fichiers journaux suivants doivent avoir été générés :

```
TESTIT_20100607_1339.system.#0.log
TESTIT_20100607_1339.system.#1.log
TESTIT_20100607_1339.system.#2.log
TESTIT_20100607_1339.system.#3.log
TESTIT_20100607_1339.system.#4.log
```

Notez que le nombre de fichiers journaux générés utilisant ce schéma est limité à 16. Après 16 fichiers journaux, les anciens fichiers journaux sont automatiquement supprimés. Définissez **IdxLogItemIds** sur 0 pour supprimer la journalisation du statut de traitement pour chaque élément contenu dans le fichier journal. Seuls les éléments qui génèrent des erreurs de traitement seront consignés.

Tous les fichiers journaux de l'indexeur contiennent les informations de journal suivantes. Chaque ligne du fichier journal contient un horodatage, un niveau de gravité, un ID unité d'exécution et un préfixe d'opération. Si un mot de passe a été spécifié sur la ligne de commande, il n'est pas affiché en clair.

- Le nom de l'outil, par exemple `afuIndexer`, ainsi que les arguments de ligne de commande utilisés
- L'environnement d'exploitation
- L'environnement d'exécution Java
- Les versions de DB2 et de Net Search Extender
- Les variables d'environnement définies, ainsi que leurs valeurs
- Les options de configuration par défaut de l'indexeur, ainsi que leurs valeurs
- Les paramètres de configuration modifiés
- Les attributs définis pour le type d'élément
- Les filtres des types de fichier définis
- Les statistiques de performances
- Les éléments qui ont été traités avec leurs valeurs `IDXRC`
- Le résultat du processus de mise à jour d'index
- Des statistiques générales et des informations sur la planification

Le fichier journal de la fonction définie par l'utilisateur `AFUFetchFile` contient les informations suivantes :

- L'ID de compilation et la version de la fonction définie par l'utilisateur
- L'environnement de la fonction définie par l'utilisateur
- La configuration de la fonction définie par l'utilisateur
- Des sections distinctes pour chaque élément, contenant l'ID d'élément, le nom du fichier XML, le contenu XML tronqué, l'horodatage, l'ID de processus, la taille du fichier XML et le temps de traitement et des fonctions UDF

Les préfixes d'opération suivants sont utilisés dans les fichiers journaux :

- `ITEM` : Informations sur le traitement d'un élément
- `NSE` : Informations sur le traitement Net Search Extender
- `TIME` : Informations sur la date/heure
- `STATISTICS` : Informations statistiques
- `CFG` : Informations sur les paramètres de configuration
- `SYS` : Informations sur le système d'exploitation
- `CMD` : Informations sur les commandes externes lancées
- `ERROR` : Informations liées aux erreurs de traitement

Les préfixes supplémentaires suivants sont utilisés si les arguments de ligne de commande `-verbose` ou `-showsql` sont indiqués lors de l'appel de l'outil `afuIndexer` :

- `SQL` : Informations sur les opérations SQL
- `RES` : Résultats des opérations dans les bases de données SQL

- RM : Informations sur le fonctionnement du gestionnaire de ressources IBM Content Manager
- LS : Informations sur le fonctionnement du serveur de bibliothèques IBM Content Manager

Les informations des fichiers journaux peuvent être extraites à des fins de surveillance, par exemple toutes les informations de traitement des éléments (entrées dotées du préfixe ITEM) ou statistiques de traitement (entrées dotées du préfixe STATISTICS).

Fichiers de trace créés lors de l'exécution des outils de l'indexeur

Si l'argument de ligne de commande `-trace` est indiqué lorsque l'un des outils d'indexeur (`afuConfig`, `afuIndexer`, `afuCheck`, `afuEnableItemType` et `afuRecreateIndex`) est appelé, des fichiers de trace sont créés.

Il inclut le répertoire dans lequel les fichiers de trace sont copiés. La valeur par défaut est le répertoire en cours.

Remarque :

- Un processus de travail dans `afuIndexer` crée son propre jeu de fichiers de trace en utilisant les mêmes noms de fichier que les fichiers journaux de l'indexeur mais en y ajoutant l'infixe `worker`.
- Vérifiez que le répertoire en cours (répertoire dans lequel vous exécutez l'outil) est accessible en écriture.
- Sauf instruction contraire, il est recommandé de ne pas modifier les paramètres du fichier des propriétés de la fonction de trace.

L'exécution d'un outil de l'indexeur avec l'argument `-trace` requiert l'utilisation d'un fichier des propriétés de la fonction de trace dans le sous-répertoire `cfg` fourni lors de l'installation du composant de recherche de texte. Ce fichier de propriétés de trace est accessible via la variable d'environnement `AFU_INDEXER_TRACE_PROPERTIES` qui le définit. Cette variable est automatiquement définie lorsque le composant de support de recherche de texte est installé.

Pour plus d'informations, voir [Contact avec le service de support logiciel IBM](#).

Statistiques créées pendant la procédure de l'indexeur

Les statistiques de traitement sont collectées à chaque exécution du processus de l'indexeur.

Si la consignation est activée à l'aide de l'option de configuration **IdxUseLogging** et qu'un ou plusieurs fichiers journaux sont créés, les statistiques de traitement sont consignées dans ces fichiers journaux et peuvent être extraites pour un traitement ultérieur. Toutes les informations statistiques consignées possèdent le préfixe `STATISTICS`.

Si l'option de configuration **IdxDisplayPerformanceMonitor** est également activée, des données de performances supplémentaires, telles que le taux de traitement des éléments, sont consignées avec le préfixe `STATISTICS`.

Les statistiques suivantes sont collectées et consignées dans les fichiers journaux :

Tableau 5. Statistiques collectées pendant l'exécution d'un processus de l'indexeur

Type de statistiques	Description
ITEMS_SCANNED	Nombre d'éléments extraits de la table de base de données des tâches ouvertes et envoyé au processus de travail.
ITEMS_BYTES	Nombre d'octets chargés pour les éléments du gestionnaire de ressources IBM Content Manager
ITEMS_MISSED	Nombre d'éléments de la table de base de données de tâches ouvertes n'ayant pu être envoyés au processus de travail car le processus de travail s'est terminé de manière inattendue et a été redémarré automatiquement.
ITEMS_CHECKED	Nombre d'éléments de la table de base de données des tâches ouvertes qui ont été vérifiés par l'outil de réparation et qui disposent d'une structure de fichier CSN valide
WORKER_ITEMS_WITH_NO_ERRORS	Nombre d'éléments correctement traités par le processus de travail sans génération d'avertissements ou d'erreurs
WORKER_ITEMS_WITH_WARNINGS	Nombre d'éléments qui ont été correctement traités par le processus de travail mais qui ont généré un avertissement lié au traitement (valeur IDXRC comprise entre 1 et 99)
WORKER_ITEMS_WITH_ERRORS	Nombre d'éléments qui n'ont pas pu être complètement traités par le processus de travail en raison d'une erreur de traitement (valeur IDXRC comprise entre 100 et 199)
WORKER_ITEMS_WITH_SEVERE_ERRORS	Nombre d'éléments qui n'ont pas pu être traités par le processus de travail en raison d'une erreur grave de traitement (valeur IDXRC de 200)
WORKER_ATTACHMENTS_LOADED	Nombre de pièces jointes dédoublonnées chargées depuis Content Manager Resource Manager
ATTACHMENTS_BYTES	Nombre d'octets des pièces jointes dédoublonnées chargées depuis Content Manager Resource Manager
NSE_ITEMS_TO_INDEX	Nombre d'éléments envoyés à Net Search Extender pour indexation. Il s'agit de la somme de WORKER_ITEMS_WITH_NO_ERRORS et de WORKER_ITEMS_WITH_WARNINGS.
NSE_ITEMS_WITH_NO_ERRORS	Nombre d'éléments indexés par Net Search Extender
NSE_ITEMS_WITH_ERRORS	Nombre d'éléments qui n'ont pas pu être indexés par Net Search Extender.

Tableau 5. Statistiques collectées pendant l'exécution d'un processus de l'indexeur (suite)

Type de statistiques	Description
NSE_BYTES	Nombre d'octets envoyés à Net Search Extender sous forme de contenu textuel dans des fichiers XML
WORKER_INSTANCES_LAUNCHED	Nombre de processus de travail démarrés par le processus de l'indexeur
NSE_INSTANCES_LAUNCHED	Nombre d'instances Net Search Extender démarrées par le processus de l'indexeur
PEAK_XML_SIZE	Plus grand fichier XML (en octets) créé lors de l'indexation
PEAK_ITEM_SIZE	Plus grand élément (en octets) traité lors de l'indexation
PEAK_ATTACHMENT_SIZE	Plus grande pièce jointe (en octets) traitée lors de l'indexation

Remarque : Des incohérences dans les statistiques globales peuvent apparaître dans les situations suivantes :

- Si le processus de travail s'arrête brutalement, la valeur de WORKER_ITEMS_WITH_NO_ERRORS peut être inférieure à la valeur ITEMS_SCANNED.
- Si les éléments ne sont pas traités correctement en raison du dépassement du délai d'attente, ils ne sont pas envoyés à Net Search Extender pour indexation. Cela signifie que la valeur de NSE_ITEMS_TO_INDEX peut être inférieure à la valeur de WORKER_ITEMS_WITH_NO_ERRORS.
- Si le processus de travail s'arrête brutalement, le nombre réel d'octets chargés à partir du gestionnaire de ressources IBM Content Manager peut être supérieur à la valeur de ITEMS_BYTES.
- Si des éléments n'ont pas été traités correctement, le nombre réel d'octets traités par Net Search Extender et la fonction définie par l'utilisateur AFUFetchFile peut être inférieur à la valeur de NSE_BYTES.
- WORKER_ATTACHMENTS_LOADED et ATTACHMENTS_BYTES n'affichent que des nombres positifs si le type d'élément utilise le modèle de données composé avec des pièces jointes dédoublonnées.
- La valeur de ITEMS_MISSED est la différence de la valeur de ITEMS_SCANNED moins la somme des valeurs de NSE_ITEMS_TO_INDEX, WORKER_ITEMS_WITH_ERRORS et WORKER_ITEMS_WITH_SEVERE_ERRORS.
- La valeur ITEMS_CHECKED n'affiche un nombre positif que si l'outil afuRepairCSN est exécuté au préalable et si l'outil a détecté des éléments à vérifier.
- La valeur de WORKER_INSTANCES_LAUNCHED est généralement de 1, sauf si le processus de travail s'est arrêté de manière inattendue et a été redémarré automatiquement par l'indexeur.
- La valeur de NSE_INSTANCES_LAUNCHED est généralement de 1 en mode de traitement par lots et peut correspondre à n'importe quel nombre si le processus de l'indexeur est exécuté en mode continu.
- Les éléments comptabilisés dans ITEMS_MISSED seront traités lors de l'exécution suivante de afuIndexer, en mode par lots ou lors de l'itération suivante si afuIndexer est exécuté en mode continu.

- Lorsque l'indexeur s'exécute en mode spécial, utilisé pour traiter les éléments avec des erreurs graves (-severeerrorsmode), la valeur de ITEMS_SCANNED peut ne pas être exacte.

Informations de date/heure créées pendant l'exécution du processus de l'indexeur

La fonction de date/heure peut être activée en associant l'option de configuration **IdxUseTiming** à la valeur true.

Des informations de date/heure générales sont consignées dans le fichier journal avec le préfixe TIME. Un fichier de date/heure contenant des valeurs spécifiques pour chaque élément est copié dans un fichier au format CSV (Comma Separated Values) dont les données peuvent être extraites à des fins de contrôle.

Le nom du fichier possède le même nom de base que tous les autres fichiers journaux d'IBM Content Collector. Seule l'extension de fichier est `timing.csv`.

La première ligne du fichier CSV contient les noms de colonne séparés par des virgules. Toutes les lignes suivantes du fichier contiennent les informations de date/heure pour un élément.

Dans le fichier CSV, la largeur minimale de chaque zone de date/heure est définie par l'option **GenTimingFormatFieldWidth**.

La première colonne répertorie l'ID élément et les colonnes suivantes contiennent les heures de traitement. Les opérations suivantes sont mesurées en millisecondes :

Tableau 6. Opérations pour lesquelles des informations de date/heure sont collectées

Opération	Description
DATE	Date à laquelle le type d'élément a été traité
TIME	Heure à laquelle le type d'élément a été traité
TASK	Durée de traitement globale de l'élément
RM_RETRIEVAL	Temps passé à extraire et à charger l'élément et les pièces jointes associées du gestionnaire de ressources IBM Content Manager
LS_RETRIEVAL	Temps passé à extraire et à charger les attributs de l'élément à partir du serveur de bibliothèques IBM Content Manager
XML_GENERATION	Temps passé à générer la structure XML de l'élément
XML_PARSING	Temps passé à analyser l'attribut de changement de zone IBM Content Manager (ICCVaryingFields)
XML_FILESAVE	Temps passé à enregistrer le fichier XML. Si une transformation XML est spécifiée par l'activation de l'option de configuration XmlTransformationXslFile , le temps passé au cours de la transformation XML est inclus dans cette opération.
DOC_PARSER	Temps passé à exécuter l'analyse syntaxique de documents, y compris les Filtres Oracle Outside In Technology
DOC_FILEDUMP	Temps passé à vider le contenu de l'élément d'origine dans le répertoire de vidage si l'option de configuration GenDump est activée.

Opérations et paramètres avancés

La procédure permettant d'activer la recherche de texte dans des documents archivés dans IBM Content Manager après l'installation du composant de prise en charge de la recherche de texte et la création ou l'activation de types d'élément inclut des étapes comportant des paramètres prédéfinis par défaut. Si vous souhaitez ajouter des attributs aux fichiers modèles par défaut, modifier des variables d'environnement ou ajuster des valeurs par défaut de configuration de l'indexation ou en savoir plus sur l'exécution de l'indexation en mode débogage ou sur les options de réindexation disponibles, consultez les rubriques d'opérations et de paramètres avancés.

Préparation à l'indexation

Avant de lancer l'indexation, vérifiez que l'environnement a été installé et configuré correctement.

Avant de lancer l'indexation des documents dans un type d'élément où des recherches de texte peuvent être effectuées, procédez comme suit.

1. La base de données définie pour l'archivage dans IBM Content Manager doit être une base de données UTF-8. Si la base de données n'a pas été créée dans ce format, vous devez la télécharger, la recréer, puis recharger les données à l'aide des commandes `db2look` et `db2move`.
2. Vérifiez que le répertoire de travail et le répertoire du fichier d'index Net Search Extender se trouvent sur le même volume. Si les répertoires se trouvent sur des volumes différents, une grande quantité de données doit être copiée à la fin de chaque mise à jour d'index ; de plus, la totalité de l'index doit être copiée lors de sa réorganisation. Si les fichiers se trouvent sur le même volume, ils sont simplement renommés au lieu d'être copiés.

Si un index Net Search Extender a été créé en utilisant des répertoires différents, la modification peut être effectuée à l'aide de la commande `db2text alter index`.

3. Le répertoire de stockage de l'index Net Search Extender doit avoir au moins deux fois plus d'espace disponible que la taille d'index prévue. L'espace total du système de fichiers doit être trois fois plus élevé que la taille d'index prévue. La taille de l'index varie en fonction de l'utilisateur mais dans le cadre de la planification, nous supposons que la taille de l'index doit représenter au moins 15 % de la taille totale de tous les courriers électroniques indexés.

Par exemple, si 1 To du courrier doit être indexé, la taille d'index peut atteindre 150 Go. L'espace total du système de fichiers doit être au moins 450 Go.

L'indexeur de recherche de texte vérifie la taille des fichiers de l'indexeur et si l'espace disponible est insuffisant, le traitement est arrêté et un message s'affiche.

4. Configurez les fonctions DB2 et les procédures sous AIX, Linux ou Solaris pour qu'elles s'exécutent avec l'ID du propriétaire de l'instance, pas avec un ID isolé. Modifiez l'ID isolé en définissant le propriétaire à l'emplacement `/home/db2inst1/sqllib/adm/.fenced`.

Fichier modèle

Le fichier modèle est un fichier de configuration qui contient des définitions de zone indiquant les données de document (attributs et métadonnées de document) ou les attributs IBM Content Manager à inclure lors de la génération de l'index du type d'élément.

Le fichier modèle est accessible lorsque le type d'élément est créé. Cette opération est effectuée implicitement par les outils de configuration Content Collector ou lors de l'exécution de la configuration initiale. Le fichier de modèle utilisé dépend du modèle de données du type d'élément.

Les fichiers de modèle répertoriés dans le tableau se trouvent dans le sous-répertoire `model` dans lequel vous avez installé le composant pour la prise en charge de la recherche de texte dans IBM Content Manager.

Tableau 7. Fichiers modèle utilisés par IBM Content Collector

Type d'élément	Fichier modèle	Version du produit
Type d'élément de courrier électronique Content Collector groupé	<code>répertoire_TextSearchSupport/model/afu_mail_entire.xml</code>	Créé à l'aide de l'IBM Content Collector V2.1.0
Type d'élément de courrier électronique Content Collector composé	<code>répertoire_TextSearchSupport/model/afu_mail_entire_cmp.xml</code>	Créé à l'aide d'IBM Content Collector V2.1.1 et V2.2
Type d'élément de courrier électronique Content Collector composé	<code>répertoire_TextSearchSupport/model/afu_mail_entire_cmp2.xml</code>	Créé à l'aide d'IBM Content Collector V3
Types d'élément de courrier électronique IBM CommonStore pour le modèle de données BUNDLED et le type d'archivage ENTIRE (BRI)	<code>répertoire_TextSearchSupport/model/cs_mail_entire.xml</code>	Créé à l'aide d'IBM CommonStore V8.4
Types d'élément de système de fichiers	<code>répertoire_TextSearchSupport/model/afu_generic.xml</code>	Toutes les versions d'IBM Content Collector
Types d'élément Content Collector créés pour des données IBM Connections	<code>répertoire_TextSearchSupport/model/afu_conn.xml</code>	Créé à l'aide d'IBM Content Collector V3
Types d'élément Microsoft SharePoint	<code>répertoire_TextSearchSupport/model/afu_generic.xml</code>	Créé avant IBM Content Collector V3
Types d'élément Microsoft SharePoint avec des pièces jointes intégrées	<code>répertoire_TextSearchSupport/model/afu_sp.xml</code>	Créé à l'aide d'IBM Content Collector V3

Les définitions d'attribut d'un fichier modèle doivent correspondre aux définitions d'attribut du fichier de configuration de l'indexeur du type d'élément.

Une fois que l'index documentaire a été initialisé à l'aide d'un fichier modèle particulier, la définition d'attribut de l'index ne peut plus être modifiée. Cela se produit lorsque le type d'élément est créé ou lorsque l'index documentaire est recréé.

Vous ne pouvez rechercher des attributs Content Collector dans les documents archivés qu'après l'indexation des documents et de leurs attributs. En revanche, vous pouvez rechercher des attributs IBM Content Manager dans les documents archivés avant l'indexation. Ce processus peut toutefois être très long car les attributs n'ont pas été indexés.

Ajout d'attributs au fichier modèle

Pour inclure des informations supplémentaires provenant des zones disponibles dans des documents ou stockés dans les attributs IBM Content Manager de l'index, vous devez ajouter les définitions de zone correspondantes au fichier modèle.

Le fichier modèle fournit le lien vers le moteur d'index de IBM Content Manager. Il est enregistré lorsque vous créez le type d'élément dans lequel le texte peut être recherché.

Si vous apportez des modifications au fichier modèle, il est vivement conseillé de ne PAS supprimer les entrées du fichier modèle ajoutées par Content Collector. N'ajoutez que de nouvelles entrées au fichier modèle.

Pour ajouter de nouveaux attributs, procédez comme suit.

1. Ouvrez le fichier modèle dans le sous-répertoire `model` dans lequel vous avez installé le composant pour la prise en charge de la recherche de texte dans IBM Content Manager et ajoutez le bloc de définition suivant :

```
<XMLFieldDefinition
  name =
  exclude="NO"
  locator=
/>
```

XMLFieldDefinition

Ajoute une nouvelle définition de zone

name

Nom à attribuer à la zone supplémentaire du fichier modèle

exclude

Définit une condition d'exclusion. Comme vous souhaitez inclure la zone supplémentaire au lieu de l'exclure, `exclude` doit être associé à `"NO"`.

locator

Chemin XML que le moteur d'indexation utilise pour localiser des informations d'indexation dans sa structure XML interne.

Remarque : Tous les chemins doivent commencer par le préfixe défini pour le fichier modèle. Le préfixe dépend du modèle de données sélectionné pour le type d'élément.

2. Enregistrez vos modifications et fermez le fichier.

Important : Lorsqu'un type d'élément est créé, un index documentaire complet est initialisé à l'aide du fichier modèle. La définition d'attribut de l'index ne peut plus être modifiée une fois le type d'élément créé.

Exemple pour un type d'élément composé Content Collector

```
<XMLFieldDefinition  
name = "Record number"  
exclude="NO"  
locator = "/icc_document/icc_email/icc_content/record_nr" />
```

Dans cet exemple, un attribut appelé **Record number** est ajouté. Les chemins de recherche qui peuvent être utilisés pour localiser ces informations sont `icc_document`, `icc_email`, `icc_content` et `record_nr` (les mêmes que dans le fichier de configuration du type d'élément).

Ajout d'attributs au fichier de configuration du type d'élément

Pour inclure des informations supplémentaires provenant des zones disponibles dans des documents ou des attributs IBM Content Manager de l'index, vous devez ajouter les définitions de zone correspondantes au fichier de configuration du type d'élément.

D'autres attributs IBM Content Manager définis par l'utilisateur peuvent être définis pour les composants enfant d'un type d'élément ou pour le composant racine d'un type d'élément. Vous ne pouvez pas définir un attribut à la fois pour le composant enfant et le composant racine d'un type d'élément avec des noms identiques. Dans ce cas, l'indexeur sélectionne les attributs définis pour le composant enfant du type d'élément et l'attribut du composant racine n'est pas accessible.

Les définitions d'attribut (les localisateurs `moddef` ou le chemin dans la structure XML utilisé pour l'indexation) définies pour les attributs du fichier modèle doivent également être définies et doivent correspondre aux définitions d'attribut dans le fichier de configuration de l'indexeur. L'indexeur vérifie que les définitions d'attribut des deux fichiers sont concordantes et émet une erreur si ce n'est pas le cas.

Il est conseillé de ne pas supprimer les entrées ajoutées par IBM Content Collector au fichier de configuration. Ajoutez uniquement les définitions d'attribut aux fichiers de configuration que vous avez ajoutés au fichier modèle pour le type d'élément.

Pour ajouter des définitions de zone, ouvrez le fichier de configuration de l'indexeur pour le type d'élément que vous avez créé :

1. Ajoutez les définitions de zone. Ajoutez les définitions de zone du document à la section commençant par `[Attributes]` et les définitions d'attribut IBM Content Manager à la section commençant par `[CMAttributes]`.

Important :

- Les zones ajoutées à la section `[Attributes]` du fichier de configuration de l'indexeur ne sont pas sensibles à la casse.
- Les attributs ajoutés à la section `[CMAttributes]` du fichier de configuration de l'indexeur sont sensibles à la casse. Cela signifie que la casse doit correspondre à celle utilisée dans la définition d'attribut du type d'élément IBM Content Manager.

La syntaxe de ces deux sections est la suivante :

```
attribute_name = moddef_locator
```

où

attribute_name

Indique le nom de la zone du document ou de l'attribut IBM Content Manager dont les informations doivent être extraites.

Si les données de l'attribut IBM Content Manager sont de type CLOB ou BLOB, le nom de la zone du client doit avoir pour préfixe CLOB ou BLOB, par exemple,

```
clob:ICCEJRcpAddrs = /document/enveloppe/recipient_addresses
```

ou

```
blob:ICCVaryingFields = /icc_document/icc_journal
```

Les attributs IBM Content Manager de type String, Integer ou VARCHAR peuvent être également utilisés et ne requièrent pas de préfixe. Les autres types de données ne peuvent pas être utilisés.

moddef_locator

Répertorie le chemin de la structure XML utilisé pour l'indexation par le moteur d'indexation. Vous devez indiquer le même `moddef_locator` que celui indiqué pour `locator` dans le fichier modèle du type d'élément. Le préfixe doit correspondre à la valeur indiquée pour le paramètre de configuration de l'indexeur `XmlModdefLocatorBase`.

2. Enregistrez vos modifications et fermez le fichier.

Exemple d'une définition de zone de document

[Attributes]

```
Record_number = /icc_document/icc_email/icc_content/record_nr
```

Dans cet exemple, l'indexeur recherche des informations à indexer dans les chemins suivants de sa structure XML interne : `icc_document`, `icc_email`, `icc_content` et `record_nr`. Le nom de l'attribut à partir duquel les informations sont extraites est **Record_number**.

Attributs du type d'élément

Il existe des fichiers de configuration de l'indexeur par défaut pour tous les types d'élément qui sont créés et utilisés dans IBM Content Collector. Les fichiers contiennent les attributs par défaut stockés dans l'index.

Lorsque vous créez un fichier de configuration de l'indexeur pour un type d'élément à l'aide du processus `afuConfig`, le contenu de la configuration du type d'élément par défaut est copié dans un nouveau fichier de configuration pour ce type d'élément.

Les fichiers de configuration de type d'élément par défaut suivants existent pour chaque modèle de données de type d'élément dans le répertoire dans lequel vous avez installé Content Collector Text Search Support dans IBM Content Manager.

Tableau 8. Fichiers modèles utilisés IBM Content Collector

Type d'élément	Fichier modèle	Version du produit
Type d'élément de courrier électronique Content Collector groupé	<i>TextSearchSupportDir/afu_attributes.ini</i>	Créé à l'aide de l'IBM Content Collector V2.1.0
Type d'élément de courrier électronique Content Collector composé	<i>TextSearchSupportDir/afu_attributes_cmp.ini</i>	Créé à l'aide d'IBM Content Collector V2.1.1 et V2.2
Type d'élément de courrier électronique Content Collector composé	<i>TextSearchSupportDir/afu_attributes_cmp2.ini</i>	Créé à l'aide d'IBM Content Collector V3
Types d'élément de courrier électronique IBM CommonStore pour le modèle de données BUNDLED et le type d'archivage ENTIRE (BRI)	<i>TextSearchSupportDir/afu_attributes_cs.ini</i>	Créé à l'aide d'IBM CommonStore V8.4
Types d'élément de système de fichiers	<i>TextSearchSupportDir/afu_attributes_generic.ini</i>	Toutes les versions d'IBM Content Collector
Types d'élément Content Collector créés pour des données IBM Connections	<i>TextSearchSupportDir/afu_attributes_conn.ini</i>	Créé à l'aide d'IBM Content Collector V3
Types d'élément Microsoft SharePoint	<i>TextSearchSupportDir/afu_attributes_generic.ini</i>	Créé avant IBM Content Collector V3

Tableau 8. Fichiers modèles utilisés IBM Content Collector (suite)

Type d'élément	Fichier modèle	Version du produit
Types d'élément Microsoft SharePoint avec des pièces jointes intégrées	<i>TextSearchSupportDir/afu_attributes_sp.ini</i>	Créé à l'aide d'IBM Content Collector V3

Les attributs des fichiers de configuration par défaut qui sont encadrés par un double trait de soulignement sont des attributs IBM Content Collector intégrés. Ils incluent :

- __ITEMID__ : ID de l'élément réel.
- __@INDEXDATE__ : Date à laquelle le fichier XML est lu et traité par le moteur d'indexation Net Search Extender.
- __@RECEIVED_DATE__ : Date à laquelle le document a été placé dans la boîte aux lettres.

Le format de date est contrôlé par le paramètre de configuration **XmlReceivedDateAndTime**. La valeur par défaut du paramètre de configuration correspond à 0, qui est corrélé à un format de date AAAAMMJJhh, ce qui est correct pour un type d'élément composé Content Collector. La rétrocompatibilité est prise en charge.

Pour un type d'élément composé (Content Collector et CommonStore), la valeur de l'option de configuration **XmlReceivedDateAndTime** doit être changée pour 1, ce qui correspond au format de date AAAAMMJJhhmmss.

- __ATTACHMENT__ : Contenu textuel d'une pièce jointe imbriquée. Un élément peut inclure plusieurs pièces jointes.
- __@CONTENT_NAME__ : Nom de fichier contenant le contenu autre que du contenu de courrier électronique.
- __@ATTACHMENT_NAME__ : Nom du fichier d'une pièce jointe imbriquée. Un élément peut inclure plusieurs pièces jointes.
- __@CONTENT_NAME__ : Nom de fichier contenant du contenu autre que du contenu de courrier électronique.
- __CONTENT__ : Représentation textuelle d'un contenu autre que du contenu de courrier électronique, c'est-à-dire du contenu Microsoft SharePoint, IBM Connections et de document de système de fichiers.
- __ENCRYPTED__ : Le contenu est chiffré et ne peut pas être indexé.
- __PRIVATE__ : Le contenu est privé et ne peut pas être indexé.
- __PARTID__ : ID de partie de document (ID de ressource IBM Content Manager) de la partie de document.

Le symbole arobas (@) devant un noeud dans une valeur de définition d'attribut (moddef_locator) indique que ce qui suit n'est pas un nouveau noeud mais un attribut du noeud précédent dans la structure XML générée. Par exemple,

__@INDEXDATE__ = /icc_document/@icc_index_date

entraîne la structure XML suivante :

```
<icc_document icc_index_date="20120402">
```

Autrement, sans le symbole arobas,
__@INDEXDATE__ = /icc_document/icc_index_date

entraîne la structure XML suivante :

```
<icc_document>  
<icc_index_date>20120402</icc_index_date>  
</icc_document>
```

Le symbole arobas (@) dans le nom d'attribut (par exemple : __@INDEXDATE__) a pour seule fonction d'indiquer qu'il est recommandé d'utiliser cet attribut avec un symbole arobas (@) devant un noeud dans une définition d'attribut.

Vous pouvez modifier les définitions d'attribut Content Collector par défaut. Toutefois, vous devez être vigilant lors de la modification de la définition d'attribut BCC. Normalement, toutes les valeurs d'attribut qui sont extraites à partir d'un e-mail et qui sont indexées sont consultables. Toutefois, dans le cas de l'attribut BCC, vous devez être attentifs à ce que le contenu BCC ne soit pas mappé avec une section consultable par tous les utilisateurs, car cela peut permettre aux utilisateurs de rechercher d'autres utilisateurs BCC sans que cela soit souhaitable.

La manière dont les informations BCC sont mappées dans Content Collector est influencée par la manière dont le dédoublement est effectué et diffère entre Microsoft Exchange et Lotus Notes. Lorsqu'un courrier électronique contient des informations BCC, le dédoublement génère une copie supplémentaire à stocker pour l'émetteur. Ceci signifie qu'il y aura une copie pour l'expéditeur et une autre pour le journal et tous les destinataires.

Dans Microsoft Exchange, le message ne contient pas d'informations BCC. Par conséquent, seule la copie destinée à l'expéditeur contient de telles informations. La section BCC est vide dans toutes les copies de destinataire, ainsi que dans la copie du journal. Dans la copie du journal, seule l'enveloppe inclut les informations BCC.

Dans Lotus Notes, en revanche, les informations BCC sont disponibles dans toutes les copies. Chaque copie de destinataire inclut uniquement les destinataires respectifs, tandis que les copies de journal et de l'expéditeur incluent tous les destinataires. Il est donc essentiel que ces informations ne figurent jamais dans la section consultable par les utilisateurs.

Si vous modifiez le mappage d'attribut BCC et ne souhaitez pas que les utilisateurs puissent rechercher d'autres utilisateurs BCC, vous devez désactiver le dédoublement de manière à ce que, chaque fois qu'un courrier électronique comporte une liste BCC, une copie du courrier électronique soit stockée pour l'expéditeur, le journal et chaque destinataire. Vous pouvez désactiver le dédoublement en ajoutant l'élément BlindCopyTo à la liste d'éléments supplémentaires utilisée pour calculer le hachage de dédoublement dans les rubriques des routes de tâches respectives. Toutefois, vous devez garder à l'esprit que cela peut entraîner d'importants besoins en espace supplémentaire, car cette pratique est courante pour permettre l'envoi de courriers électroniques avec de grandes listes de distribution à tous les destinataires spécifiés en BCC.

Utilisation de TIEFLAG lors de l'indexation dans IBM Content Collector

TIEFLAG est le nom prédéfini d'une colonne dans le serveur de bibliothèque IBM Content Manager.

A l'origine, la valeur TIEFLAG déterminait les parties d'un document qui étaient indexées par Net Search Extender et elle était évaluée par l'exécutif utilisateur d'IBM Text Search. La valeur TIEFLAG est généralement comprise entre 10 et 19. Tous les éléments d'un type d'élément reçoivent la même valeur TIEFLAG. Cette dernière dépend de l'option de configuration TEXT_SEARCHABLE qui se trouve dans le fichier de configuration, appelé archint.ini, sur le serveur IBM CommonStore (archpro). Elle peut être vérifiée dans la table de composants du type d'élément dans IBM Content Manager. Dans la plupart des cas, la valeur TIEFLAG est définie sur 17 ou 18. Afin de garantir la compatibilité en amont, IBM Content Collector affecte la valeur TIEFLAG 18 à tous les nouveaux éléments qui sont créés pendant l'archivage.

Toutefois, le composant de recherche de texte IBM Content Collector n'utilise PAS la valeur TIEFLAG pour déterminer les parties d'un élément à ajouter à l'index documentaire. Cela signifie que la valeur réelle de l'option de configuration TEXT_SEARCHABLE stockée dans le fichier de configuration (archint.ini) sur le serveur IBM CommonStore ne sert pas à contrôler les parties d'un document qui sont indexées. Désormais, cette opération est contrôlée par les paramètres de la configuration de l'indexeur.

Dans IBM Content Collector, tous les éléments traités par l'indexeur de recherche de texte sont insérés dans la table des tâches terminées avec un code de résultat appelé IDXRC. IDXRC représente le code de statut affecté à chaque élément afin d'indiquer si l'élément a été correctement indexé.

L'indexeur rapide et l'indexeur IBM Content Manager standard ne surveillent pas le statut d'indexation des éléments traités dans une table distincte et ne définissent pas de valeur IDXRC. L'indexeur rapide s'appuie sur la valeur TIEFLAG pour indiquer si un élément peut être indexé ou non en définissant la valeur TIEFLAG sur une valeur supérieure à 10 000 si l'élément n'a pas pu être indexé.

Lorsque vous activez l'indexation d'un type d'élément dans IBM Content Collector, la table des tâches terminées est complétée avec tous les éléments qui ont été indexés à l'aide de l'indexeur rapide ou de l'indexeur IBM Content Manager standard. Cette opération d'importation est automatique si le type d'élément a été archivé par le biais d'IBM CommonStore. Elle peut également être effectuée manuellement en lançant l'outil `afuIndexTool` avec l'argument de ligne de commande `-opImportToCompleted`.

De plus, sachant que ni l'indexeur rapide ni l'indexeur IBM Content Manager standard n'affectent de valeur IDXRC aux éléments qui ont été indexés, la valeur IDXRC attribuée à ces nouveaux éléments correspond à la valeur TIEFLAG utilisée au cours de l'indexation à l'aide de l'indexeur rapide ou de l'indexeur IBM Content Manager standard. La valeur TIEFLAG est généralement définie sur 17 ou 18, ce qui signifie que la valeur IDXRC de la table des tâches terminées associée aux éléments archivés via CommonStore correspond également à 17 ou 18 après que les valeurs TIEFLAG aient été importées dans le cadre du traitement d'activation du type d'élément.

Si la file d'attente d'indexation de l'indexeur rapide est vide et que la valeur TIEFLAG d'un élément de la table des composants du type d'élément IBM Content Manager est toujours comprise entre 10 et 19, l'élément a été correctement indexé. Dans ce cas, la valeur IDXRC de l'élément de la table des tâches terminées correspond à la valeur TIEFLAG d'origine (valeur comprise entre 10 et 19, généralement 17 ou 18).

Cependant, si la valeur TIEFLAG d'un élément figurant dans la table des composants du type d'élément IBM Content Manager est supérieure à 10 000, l'élément n'a pas pu être traité correctement par l'indexeur rapide. L'indexeur rapide attribue généralement la valeur TIEFLAG 10 017 ou 10 018. Dans l'indexeur Content Collector, l'opération d'importation affecte à ces éléments une nouvelle valeur IDXRC égale à 180 (ERR_ITEMTYPE_TIEFLAG_ERROR) dans la table des tâches terminées. Cette valeur IDXRC 180 indique que les éléments n'ont pas été correctement indexés par l'indexeur rapide ou l'indexeur IBM Content Manager standard. Par conséquent, une fois l'importation automatique terminée, la table des tâches terminées affiche le statut de tous les éléments associés au type d'élément avec une valeur IDXRC valide.

Vous devez réindexer tous les éléments qui possèdent une valeur IDXRC 180 (ITEMTYPE_TIEFLAG_ERROR) à l'aide de l'indexeur de recherche de texte Content Collector, car ce dernier produit de meilleurs résultats d'indexation que l'indexeur rapide, augmentant ainsi les chances de traiter et d'indexer correctement les documents qui ont échoué précédemment. Utilisez l'argument de ligne de commande `-reindexrc 180` pour réindexer ces éléments.

Si vous employez une autre application pour connaître le statut des éléments indexés ou récupérer les documents, par exemple eDiscovery Manager, vous pouvez configurer l'indexeur Content Collector pour ajouter la valeur IDXRC de chaque élément à la colonne TIEFLAG du type d'élément dans IBM Content Manager. Pour configurer ainsi l'indexeur Content Collector, définissez l'option de configuration **IdxUpdateItemTypeTable** sur 1. Lorsque la table **IdxUpdateItemTypeTable** est définie sur 1, le processus d'indexation ajoute uniquement l'élément traité à la table des tâches terminées avec une valeur IDXRC, et ajoute également cette valeur IDXRC à la colonne TIEFLAG de l'élément dans la table des composants du type d'élément IBM Content Manager. Pour plus d'informations sur l'option de configuration, voir **IdxUpdateItemTypeTable**.

Variables d'environnement de l'indexeur de texte

Des variables d'environnement peuvent être définies pour contrôler l'exécution de l'indexeur lors de la recherche de texte dans différents environnements.

Les variables d'environnement suivantes sont prises en charge par les différents outils d'indexation :

AFU_INDEXER_RMI_PORT

Le processus de l'indexeur communique avec le processus de travail et cette communication requiert un numéro de port disponible.

Le numéro de port par défaut utilisé par le processus de l'indexeur est 4242. Si vous rencontrez des problèmes de communication entre l'indexeur et le processus de travail, sélectionnez un autre numéro de port qui est disponible en utilisant **AFU_INDEXER_RMI_PORT** et en l'affectant à un numéro de port unique ou en utilisant l'argument `-port` lorsque le processus de l'indexeur est démarré.

Bien que le processus de l'indexeur ne puisse s'exécuter que sur un seul type d'élément à la fois, plusieurs types d'élément peuvent être traités simultanément. Cependant, pour ce faire, chaque instance de l'indexeur doit utiliser un port unique ; soit en utilisant l'argument de ligne de commande `-port` lorsque le processus de l'indexeur est démarré, soit en définissant la variable d'environnement **AFU_INDEXER_RMI_PORT**.

AFU_INDEXER_PROMPT_RUN

L'indexeur détermine si une instance est en cours d'exécution pour un type d'élément donné en créant un fichier dans le système de fichiers. Ce mécanisme permet d'éviter les exécutions multiples pour le même type d'élément. Si l'indexeur est redémarré pour le même type d'élément, l'outil signale qu'il a identifié le fichier indiquant qu'une autre instance est en cours d'exécution. Un message est généré avec le nom du fichier détecté et le code retour du processus, 100.

En règle générale, ce fichier de contrôle est supprimé dès que l'indexeur termine son traitement. Toutefois, si la procédure de l'indexeur se termine de manière incorrecte ou est interrompue, le fichier risque d'être conservé sur le système de fichiers et doit être supprimé manuellement avant la réexécution de l'indexeur pour le type d'élément.

Si **AFU_INDEXER_PROMPT_RUN** correspond à 1, le processus de l'indexeur ne s'arrête pas en générant un code retour 100 lorsqu'il détecte qu'un processus de l'indexeur est déjà en cours d'exécution pour le même type d'élément mais il demande à l'utilisateur s'il doit continuer et ignorer le code retour. Naturellement, cela empêche le traitement par lots mais peut être utile lors de la phase de configuration du processus de l'indexeur.

AFU_INDEXER_PROMPT_TOOL

AFU_INDEXER_PROMPT_TOOL est pris en charge lors de l'exécution de `afuIndexTool`. Le traitement par défaut lorsque `afuIndexTool` est exécuté consiste à demander à l'utilisateur une confirmation au cours d'une opération de l'outil d'index. Ceci empêche l'exécution de `afuIndexTool` en mode de traitement par lots.

Si **AFU_INDEXER_PROMPT_TOOL** est défini sur 0, l'outil d'index ne demande pas à l'utilisateur mais suppose que la réponse à toutes les invites est Oui.

Utilisez ce paramètre avec précaution.

AFU_INDEXER_PROMPT_RECREATE

Le traitement par défaut de l'outil `afuRecreateIndex` consiste à demander à l'utilisateur une confirmation au cours du processus de recréation d'index. Ceci empêche l'exécution de `afuRecreateIndex` en mode de traitement par lots.

Si **AFU_INDEXER_PROMPT_RECREATE** est défini sur 0, l'outil de recréation d'index ne demande pas à l'utilisateur mais suppose que la réponse à toutes les invites est Yes.

Utilisez ce paramètre avec précaution.

AFU_INDEXER_XML_VIEWER

Lorsque l'indexeur s'exécute en mode débogage et traite les documents un par un, les fichiers XML générés pour chaque élément peuvent être affichés si **AFU_INDEXER_XML_VIEWER** correspond à une application capable d'afficher des fichiers XML. L'indexeur vous invite ensuite à indiquer s'il doit afficher le fichier XML et exécuter l'application indiquée avec le nom du fichier XML si l'utilisateur a décidé d'afficher le fichier.

Par exemple, sous Windows, définissez la variable d'environnement sur `C:\Mozilla\FireFox3\firefox.exe` pour utiliser le navigateur Web Mozilla Firefox en tant qu'afficheur de fichier XML.

Remarque : Sous UNIX ou Linux, il est conseillé de configurer l'application de l'afficheur XML pour utiliser une fenêtre distincte pour l'affichage du contenu XML ; par exemple, le navigateur Mozilla Firefox.

AFU_FETCHFILE_CONFIG_PATH

Généralement, le fichier de configuration de la fonction définie par l'utilisateur AFUFetchFile automatiquement créé par l'indexeur pour la recherche de texte se trouve dans le répertoire destiné aux fichiers temporaires, souvent indiqué par les variables d'environnement TMP ou TEMP sous Windows, et dans le répertoire /tmp sous UNIX ou Linux. En indiquant **AFU_FETCHFILE_CONFIG_PATH**, vous pouvez utiliser un répertoire de votre choix pour stocker le fichier de configuration de la fonction définie par l'utilisateur.

Pour plus d'informations, voir «Modifications du répertoire par défaut et des paramètres de nom de fichier de la configuration de la fonction définie par l'utilisateur AFUFetchFile», à la page 161

AFU_INDEXER_TRACE_PROPERTIES

Cette variable d'environnement détermine le nom du fichier de propriétés de la fonction de trace si la fonction de trace est activée à l'aide de l'option de ligne de commande `-trace`.

Si la fonction de trace est activée et que cette variable d'environnement est définie, le processus de l'indexeur recherche un fichier dont le nom correspond à celui indiqué par la variable d'environnement.

Si la fonction de trace est activée et que cette variable d'environnement n'est PAS définie, le processus de l'indexeur recherche le fichier de propriétés de la fonction de trace par défaut, `afuIndexerTrace.properties`, dans le répertoire en cours. Si ce fichier par défaut est introuvable, un message d'erreur s'affiche et le processus de l'indexeur s'arrête.

Remarque : Cette variable est automatiquement définie pendant l'installation de la fonction Content Collector Text Search Support dans tous les fichiers de script de l'outil de l'indexeur du répertoire `bin`.

AFU_INDEXER_USE_IT_CS_GENERIC

La définition de cette variable d'environnement sur la valeur `true` force le processus de l'indexeur à reconnaître un type d'élément comme étant un type d'élément de courrier électronique avec le modèle de données de document créé à l'aide de IBM CommonStore. La valeur par défaut de **AFU_INDEXER_USE_IT_CS_GENERIC** est `false`.

Normalement, lorsque le processus de l'indexeur tente de reconnaître si un type d'élément est un type d'élément de courrier électronique IBM CommonStore, il vérifie l'un des attributs suivants : `CSLDOriDB`, `CSLDOriUser` ou `CSORIGINATOR`. Si l'un de ces attributs est défini pour le type d'élément, ce dernier est considéré comme un type d'élément de courrier électronique IBM CommonStore.

Si ni `CSLDOriDB`, `CSLDOriUser`, ni `CSORIGINATOR` n'existe, en définissant **AFU_INDEXER_USE_IT_CS_GENERIC** sur `true`, le processus de l'indexeur suppose automatiquement que le type d'élément est un type d'élément de courrier électronique IBM CommonStore.

Utilisez ce paramètre avec précaution.

Modification de la configuration par défaut de l'indexeur

Vous pouvez utiliser plusieurs fichiers de configuration lors du traitement d'un type d'élément, ignorer des fichiers lors du traitement d'un document et ajuster les valeurs par défaut de la plupart des options de configuration.

Utilisation de plusieurs fichiers de configuration

Lorsque `afuConfig` est lancé, il crée une configuration par défaut pour le type d'élément. Cette configuration par défaut est suffisante pour exécuter l'indexation mais n'est pas nécessairement assez précise pour obtenir les meilleures performances système et de capacité de traitement. Cela signifie que vous pouvez être amené à modifier certaines des valeurs par défaut.

Pour éviter d'avoir à apporter les mêmes modifications dans plusieurs fichiers de configuration, vous devez distribuer la configuration de l'indexeur d'un type d'élément entre plusieurs fichiers de configuration. Par exemple, vous pouvez disposer d'un fichier contenant toutes les options de type d'élément génériques qui sont partagées par tous les types d'élément (par exemple, les chemins de fichier pour les fichiers XML et les fichiers journaux) et un autre fichier avec les définitions de filtre de fichier.

Pour créer plusieurs fichiers de configuration :

1. Copiez le fichier de configuration par défaut créé pour un type d'élément lors de l'exécution de `afuConfig`.
2. Vous ne pouvez pas partager les configurations de l'indexeur entre les types d'éléments qui utilisent des modèles de données différents car il pourrait y avoir un conflit entre les paramètres de définition [Attributes] et [CMAAttributes]. Il peut également y avoir un conflit si vous utilisez les options de configuration commençant par `Xml`, telles que `XmlTransformationXslFile` ou `XmlAttributeRequiresNseDateConversion`.
3. Editez ce fichier pour qu'il contienne différents types d'options de configuration ; par exemple, ajoutez les options génériques à un fichier, les options que vous souhaitez surveiller à un autre fichier et celles que vous souhaitez modifier à un autre fichier encore.

Si des options obligatoires sont introuvables dans l'un des fichiers de configuration, leurs valeurs par défaut sont utilisées.

Vous pouvez également créer votre propre fichier de configuration. Si vous créez intégralement une configuration, il est recommandé de faire débiter le fichier de configuration par la ligne :

```
[AFU_V1]
```

Vous pouvez ignorer les deux lignes suivantes qui apparaissent également dans l'exemple ci-après car elles ne sont utilisées qu'à des fins de référence :

```
CreationDate = 2009/07/21 13:55  
CreatedForItemType = TEST
```

Toutes les options de configuration que vous ajoutez doivent être précédées de [Settings].

Pour spécifier plusieurs fichiers de configuration lorsque `afuIndexer` est lancé, utilisez l'argument de ligne de commande :

```
-useconfig <.ini file1>;<.ini file2>
```

Les noms de fichiers de configuration sont séparés par des points-virgules (;) sous Windows et des deux-points (:) sous UNIX et LINUX.

Fichiers ignorés lors du traitement de documents

Un jeu de types de fichiers par défaut utilisé pour le filtrage lors du traitement des pièces jointes des courriers électroniques ou des documents SharePoint et File System est fourni par Content Collector dans un fichier appelé `afu_filetypefilter.ini`. Ce fichier représente la liste noire, c'est-à-dire la liste des extensions de fichier à ignorer lors du traitement, et est automatiquement copié dans la configuration par défaut lors de l'exécution de `afuConfig`.

Si vous souhaitez ajouter des extensions de fichier à ignorer pendant le traitement des documents, vous pouvez modifier ce fichier avant de créer le fichier de configuration de l'indexeur ou modifier directement la configuration de l'indexeur ultérieurement. Les filtres de type de fichier peuvent être modifiés après la création du type d'élément.

Une autre méthode conseillée de gestion des fichiers à ignorer lors du traitement consiste à supprimer le jeu de types de fichiers de la configuration par défaut et de conserver un fichier de configuration distinct avec cette liste noire d'extensions de fichier partagée entre tous les types d'élément. Le fichier de configuration inclut **GenFileTypeFilterIsBlackList** = true sous [SETTINGS] et répertorie les extensions de fichier à ignorer sous [FILETYPEFILTER].

Vous pouvez également conserver une liste blanche des extensions de fichier représentant uniquement les types de fichiers que vous souhaitez traiter. Si **GenFileTypeFilterIsBlackList** est désactivé, la liste des extensions de fichier est considérée comme étant la liste blanche.

Modification des valeurs des options de configuration par défaut

Les valeurs des options de configuration du fichier de configuration créé après l'exécution de `afuConfig` sont permanentes. Si vous devez modifier des valeurs, vous pouvez le faire de deux manières :

- Modifiez directement le fichier de configuration avant l'appel de `afuIndexer`.
- Transmettez les valeurs des options de configuration à l'aide de l'argument de ligne de commande `-usesettings` lors de l'appel de `afuIndexer`. Les valeurs de configuration transmises à l'aide de cet argument de ligne de commande remplacent provisoirement les valeurs des options de configuration lors de l'indexation.

Si la consignation est activée, toutes les options de configuration indiquant celles qui ont été modifiées et celles qui ne disposent plus de valeurs par défaut sont consignées.

Options de configuration de l'indexeur

Le fichier de configuration de l'indexeur défini pour un type d'élément contient toutes les options de configuration nécessaires pour indexer les éléments du type d'élément.

Remarque :

- Les noms des options de configuration ne font pas de distinction entre les majuscules et les minuscules.
- Si la valeur de l'option est de type booléen, elle doit correspondre à 0 (désactivée) ou à 1 (activée), ou à true ou à false.

- Si la valeur de l'option est de type chaîne et que la valeur est vide, laissez la chaîne vide. N'indiquez pas de chaîne vide à l'aide de guillemets.
- Si la valeur de l'option est de type entier, vous pouvez indiquer la valeur sous la forme d'un nombre ou d'un nombre doté d'un suffixe. Le suffixe est un facteur multiplié par une valeur. Les suffixes suivants sont pris en charge :
 - k = * 1000
 - K = * 1024
 - m = * 1000 * 1000
 - M = * 1024 * 1024
 - g = * 1000 * 1000 * 1000
 - G = * 1024 * 1024 * 1024

Par exemple : 1g = 1000m = 1000 000k = 1000 000 000

Vous devez arrondir tous les nombres. Par exemple, vous ne pouvez pas utiliser 1,5m mais seulement 1m.

Options de configuration du processus d'indexation générique

Les options de configuration GenCommandXXX suivantes peuvent être utilisées pour ajouter des étapes de traitement définies par l'utilisateur juste avant ou juste après l'indexation Net Search Extender ou la réorganisation de l'index Net Search Extender. L'insertion de commandes définies par l'utilisateur lors des étapes de traitement clés permet de contrôler les options de traitement de l'index ou de sauvegarder les données pertinentes.

Remarque : Sous Windows, si la commande utilisateur est un fichier de commandes .BAT ou .CMD, la commande doit inclure `cmd /c`. Par exemple, `cmd /c macommande.bat`.

Toutes les options de configuration GenCommandXXX peuvent utiliser les variables suivantes, qui sont remplacées par leurs valeurs réelles dans les chaînes de la commande.

- %ITEMTYPE% : Nom du type d'élément
- %ITEMTYPEID% : Identificateur du type d'élément
- %INDEXNAME% : Nom de l'index Net Search Extender
- %INDEXID% : Identificateur de l'index Net Search Extender
- %INDEXDIR% : Répertoire de l'index Net Search Extender
- %INDEXWORKDIR% : Répertoire de travail de l'index Net Search Extender

Ces variables permettent d'utiliser des commandes de sauvegarde qui peuvent être imbriquées dans des processus d'indexation, puis de recevoir les arguments nécessaires à ces commandes pour accéder à des informations propres à l'élément. Par exemple :

```
GenCommandAfterNseIndexing =
  myBackup.exe -itemtype %ITEMTYPE% -indexdirectory %INDEXDIR%
```

ou

```
GenCommandAfterNseIndexing =
  cmd /c myBackup.cmd -itemtype %ITEMTYPE% -indexdirectory %INDEXDIR%
```

GenCommandAfterNseIndexing

Type de valeur : Chaîne

La valeur par défaut est aucune valeur.
Si cette option est définie, la commande utilisateur est directement exécutée après le processus Net Search Update Index.

GenCommandAfterNseReorg

Type de valeur : Chaîne

La valeur par défaut est aucune valeur.
Si cette option est définie, la commande utilisateur est immédiatement exécutée après l'opération de réorganisation de l'index Net Search Extender.

GenCommandBeforeNseIndexing

Type de valeur : Chaîne

La valeur par défaut est aucune valeur.
Si cette option est définie, la commande utilisateur est directement exécutée avant le début du processus Net Search Update Index.

GenCommandBeforeNseReorg

Type de valeur : Chaîne

La valeur par défaut est aucune valeur.
Si cette option est définie, la commande utilisateur est exécutée juste avant le début de l'opération de réorganisation de l'index Net Search Extender.

Les options de configuration génériques restantes incluent :

GenCommandErrorRcAborts

Type de valeur : Booléen

Valeur par défaut : 0

Si cette option correspond à 0, l'indexeur continue le traitement même si une commande utilisateur de l'une des options GenCommandXXX se termine avec un code retour différent de 0.

Lorsque l'option correspond à 1, le processus de l'indexeur s'arrête si l'une des commandes utilisateur s'exécute en générant un code retour différent de zéro.

Le processus de l'indexeur s'arrête en générant le code retour 9, qui indique qu'une erreur s'est produite pour l'une des commandes utilisateur GenCommandXXX.

Remarque : Si une erreur s'est produite pendant l'exécution de l'une des commandes GenCommandXXX et que le paramètre **GenCommandErrorRcAborts** a pour valeur 1, les fichiers XML temporaires ne sont pas supprimés. Une fois l'erreur corrigée, vous devez supprimer les fichiers manuellement. Gardez en mémoire que, à cause de cette erreur, le traitement de l'élément ne s'est pas terminé et que la table des tâches complète n'est pas mise à jour.

GenRmRetryAllErrors

Type de valeur : Booléen

Valeur par défaut : 1

En général, l'indexeur tente uniquement d'extraire une nouvelle fois les éléments qui ont généré une erreur liée au dépassement du délai d'attente.

Si **GenRmRetryAllErrors** correspond à 1, toutes les erreurs renvoyées par le gestionnaire de ressources pour les éléments sont traitées comme s'il s'agissait d'erreurs liées au dépassement du délai d'attente et le processus de l'indexeur tente d'extraire à nouveau ces éléments à sa prochaine exécution. Le nombre de tentatives d'extraction des éléments par l'indexeur est défini par l'option **GenMaxNumberOfRetriesForRmErrors**.

Si l'option **GenRmRetryAllErrors** correspond à 0, l'indexeur tente uniquement d'extraire à nouveau les éléments qui ont généré une erreur liée au

dépassement du délai d'attente. Les éléments qui ont généré d'autres erreurs dans le gestionnaire de ressources ne sont pas une nouvelle fois extraits mais ils sont ajoutés à la table des tâches terminées avec une erreur. Les éléments qui n'ont pas pu être extrait du gestionnaire de ressources sont également ajoutés au fichier journal.

GenDump

Type de valeur : Booléen

Valeur par défaut : 0

Si cette option correspond à 1, les documents originaux extraits du gestionnaire de ressources sont stockés dans le répertoire de vidage défini par l'option **GenDumpDirectory**. Si vous ne parvenez pas à traiter certains documents et que vous avez besoin d'un support pour analyser les documents, associez cette valeur à 1.

GenDumpDirectory

Type de valeur : Chaîne

Valeur par défaut : \$(TMP)/afu/dump

Répertoire dans lequel les documents extraits du du gestionnaire de ressources sont placés si **GenDump** correspond à 1. Pour vous aider à différencier les fichiers, les noms de fichier du répertoire de vidage contiennent l'ID de l'élément.

Remarque :

- Les opérations de vidage générant des erreurs ne sont pas consignées ou gérées. Le répertoire de vidage n'est pas non plus nettoyé par le processeur de l'indexeur. Il revient à l'utilisateur de supprimer les fichiers du répertoire de vidage et de redéfinir l'option de vidage après le débogage.
- Il est recommandé que la valeur de l'option **GenDumpDirectory** et **GenWorkingDirectoryForItemTypes** soit identique pour toutes les configurations de type d'élément.

GenFileTypeFilterIsBlackList

Type de valeur : Booléen

Valeur par défaut : 1

Détermine les pièces jointes ou les documents qui doivent faire l'objet d'une analyse syntaxique à l'aide des Filtres Oracle Outside In Technology et ceux qui doivent être ignorés. Si cette option correspond à 1, la liste d'extensions de nom de fichier indiquée à la section FileTypeFiles dans la configuration est considérée comme la liste des entrées refusées. Cela signifie que tous les documents dont les extensions de nom de fichier correspondent à ces entrées sont ignorés.

Si cette option correspond à 0, la liste d'extensions de nom de fichier indiquée à la section FileTypeFiles est considérée comme la liste des entrées acceptées. Cela signifie que seuls les documents dont l'extension de nom de fichier est indiquée à la section FileTypeFiles sont traités. Tous les autres documents sont ignorés.

GenLogDirectory

Type de valeur : Chaîne

Valeur par défaut : \$(TMP)/afu/log

Répertoire dans lequel les fichiers journaux sont placés si la consignment est activée via l'option **IdxUseLogging** associée 1.

Remarque : Seuls les caractères alphanumériques ci-après peuvent être utilisés lors la définition du nom de répertoire.

- Tous les caractères alphabétiques (A à Z et a à z) et les nombres compris entre 0 et 9
- Sous UNIX et Linux : /_-.@#
- Sous Windows : \:~_-.@#\$ et le caractère espace

La longueur maximale de la chaîne est 255 caractères.

Restriction : La valeur de l'option **GenXmlDirectory** et **GenLogDirectory** DOIT être identique pour toutes les configurations de type d'élément.

GenLogFileNameTimeStamp

Type de valeur : Chaîne

Valeur par défaut : _aaaaMMjj_HHmm

Les fichiers journaux utilisent des horodatages dans leurs noms. Il est recommandé de conserver l'horodatage par défaut des noms de fichier afin qu'il y ait un nouveau jeu de fichiers journaux pour chaque exécution du processus de l'indexeur.

Si vous n'indiquez pas de valeur pour **GenLogFileNameTimeStamp**, tous les fichiers journaux possèdent le même nom et sont remplacés lors de la prochaine exécution de l'indexeur pour le même type d'élément.

Si vous souhaitez modifier l'horodatage dans le nom des fichiers journaux, utilisez les lettres ci-après. Toutes les autres lettres sont traitées telles quelles.

Remarque : Seuls les caractères suivants sont pris en charge :

- Sous UNIX et Linux : /_-.@#
- Sous Windows : \:~_-.@#\$ et le caractère espace

Tableau 9. Lettres à utiliser pour les horodatages des noms de fichier

Lettre	Date ou heure	Exemple
y	Année	1997 ; 97
M	Mois de l'année	July ; Jul ; 07
w	Semaine de l'année	27
W	Semaine du mois	2
D	Jour de l'année	189
d	Jour du mois	31
F	Jour de la semaine dans un mois	2
E	Jour de la semaine	Tuesday ; Tue
a	Marqueur am/pm	(après-midi)
H	Heure de la journée (0-23)	0
k	Heure de la journée (1-24)	24
K	Heure au format am/pm (0-11)	0
h	Heure au format am/pm (1-12)	12
m	Minute dans une heure	30
s	Seconde dans une minute	57
S	Milliseconde	978
z	Fuseau horaire	PST; GMT-08:00

GenRmConnectionTimeOutSecs

Type de valeur : Entier

Valeur par défaut : 60

Délai d'attente pour extraire les éléments du gestionnaire de ressources. Si le délai nécessaire à l'extraction d'un élément du gestionnaire de ressources dépasse la valeur indiquée pour l'option **GenRmConnectionTimeOutSecs**, le système considère que l'extraction a généré une erreur. Cette erreur peut entraîner une nouvelle tentative de traitement de l'élément lors de la prochaine exécution de l'indexeur ou l'ajout de l'élément à la table des éléments terminés avec le message d'erreur.

GenMaxNumberOfRetriesForRmTasks

Type de valeur : Entier

Valeur par défaut : 3

Nombre de tentatives effectuées pour extraire des éléments renvoyés avec une erreur liée au dépassement du délai et, si l'option **GenRmRetryAllErrors** correspond à 1, qui ont généré une autre erreur du gestionnaire de ressources. Le nombre maximal de tentatives de traitement est 10.

GenWorkingDirectoryForItemTypes

Type de valeur : Chaîne

Valeur par défaut : \$(TMP)/afu

Ce répertoire est nécessaire pour tous les fichiers de travail. Un sous-répertoire est automatiquement créé pour chaque type d'élément. Les Filtres Oracle Outside In Technology utilisent également ce répertoire de travail. Veillez à utiliser un système de fichiers hautes performances pour ce répertoire.

Remarque : Il est recommandé que la valeur de l'option **GenDumpDirectory** et **GenWorkingDirectoryForItemTypes** soit identique pour toutes les configurations de type d'élément.

GenXmlDirectory

Type de valeur : Chaîne

Valeur par défaut : \$(TMP)/afu/xml

Un seul fichier XML est généré pour chaque élément. Les fichiers XML sont des fichiers intermédiaires qui sont automatiquement supprimés après l'indexation sauf si l'option de configuration

XmlDeleteFilesAfterNseProcessing correspond à 0. Vérifiez qu'il y a suffisamment d'espace dans ce répertoire pour tous les fichiers XML.

Remarque : Seuls les caractères alphanumériques suivants peuvent être utilisés lors la définition du nom de répertoire :

- Tous les caractères alphabétiques (A à Z et a à z) et les nombres compris entre 0 et 9
- Sous UNIX et Linux : / _ . @ #
- Sous Windows : \ : ~ _ . @ # \$ et le caractère espace

La longueur maximale de la chaîne est 255 caractères.

Le répertoire XML est partagé par tous les processus de l'indexeur exécutés simultanément pour plusieurs types d'élément et peut donc devenir trop volumineux.

Restriction : La valeur de l'option **GenXmlDirectory** et **GenLogDirectory** DOIT être identique pour toutes les configurations de type d'élément.

GenXmlDirectoryMinFreeSizeRequiredInMB

Type de valeur : Entier

Valeur par défaut : 512

Lorsque le processus de l'indexeur vérifie le système de fichiers indiqué, le nombre de Mo défini par l'option de configuration **GenXmlDirectory** doit être au moins disponible dans le répertoire XML.

GenNseIndexDirectoryMinFreeSizeRequiredInMB

Type de valeur : Entier

Valeur par défaut : 512

Lorsque le processus de l'indexeur vérifie le système de fichiers indiqué, le nombre de Mo défini doit être au moins disponible dans le répertoire d'index Net Search Extender.

GenDatabaseTransactionSize

Type de valeur : Entier

Valeur par défaut : 100 000

La valeur de **GenDatabaseTransactionSize** détermine le nombre maximal de ligne qui seront insérées ou mises à jour dans une table spécifique au cours d'une opération de base de données avant l'émission d'une instruction SQL COMMIT. Le nombre requis dépend de la taille du journal des transactions IBM DB2. Si vous rencontrez des erreurs SQL pendant des opérations d'insertion ou de mise à jour, des escalades de verrou dans DB2 par exemple, réduisez ce nombre. Vous pouvez également augmenter la taille du journal des transactions IBM DB2.

Options de configuration du processus d'index**IdxMaxNumberOfSubsequentRmErrors**

Type de valeur : Entier

Valeur par défaut : 100

La valeur minimale est 10 et la valeur maximale est 9999. Cette option indique le nombre maximal d'erreurs d'extraction du gestionnaire de ressources qui sont autorisées sur une ligne sans tentatives d'extraction réussies entre ces erreurs d'extraction. Cette option est utile pour déterminer si le gestionnaire de ressources n'est pas en cours d'exécution et requiert l'intervention de l'administrateur pour éviter un traitement illimité d'un lot volumineux sans élément à indexer.

IdxUseTiming

Type de valeur : Booléen

Valeur par défaut : 0

Active le délai et la création de fichiers de date/heure au format CSV dans le répertoire de consignation.

Remarque : Si l'option de configuration **IdxDisplayPerformanceMonitor** est activée, le taux de traitement de l'indexeur, c'est-à-dire le nombre d'éléments traités par seconde, est affiché, quelle que soit la valeur de l'option **IdxUseTiming**.

IdxUseLogging

Type de valeur : Booléen

Valeur par défaut : 1

Active la consignation et crée les fichiers journaux dans le répertoire de consignation indiqué par l'option de configuration **GenLogDirectory**. Par

défaut, les fichiers journaux contiennent uniquement es informations importantes, par exemple les éléments qui ont été correctement traités, les éléments qui ont été traités en générant une erreur et les données de configuration système.

Remarque : Si l'argument de ligne de commande `-verbose` est indiqué lorsque la commande `afuIndexer` est appelée, des informations complémentaires sont ajoutées au fichier journal si la consignation est activée.

IdxLogItemIds

Type de valeur : Booléen

Valeur par défaut : 1

Si la consignation est activée, les ID de tous les éléments sont consignés deux fois par défaut, une fois lorsque les éléments sont envoyés aux travaux et une autre fois lorsqu'ils sont traités. Cela signifie que deux lignes sont consignées dans le fichier journal pour chaque élément et que cette procédure peut avoir une incidence sur la taille du fichier journal, notamment pour les lots très volumineux.

Comme les informations d'état d'un élément sont également stockées dans la table des tâches terminées, la consignation de l'état du traitement de chaque élément dans le fichier journal peut être désactivée en associant l'option **IdxLogItemIds** à 0. Seuls les éléments qui génèrent des erreurs de traitement seront consignés.

Il est conseillé de désactiver cette option lorsque l'indexeur s'exécute en mode continu pour réduire la quantité de données ajoutées aux fichiers journaux.

IdxItemMaxBatchSize

Type de valeur : Entier

Valeur par défaut : 100

Nombre de documents indexés au cours d'une exécution de l'indexation.

La valeur de **IdxItemMaxBatchSize** est ajustée automatiquement par l'indexeur si nécessaire (pour plus de détails, voir «Contraintes lors de l'indexation en mode de traitement par lots», à la page 88). Si vous exécutez le processus de l'indexeur quotidiennement en mode de traitement par lots, vérifiez que votre taux moyen d'archivage correspond à la taille de lot maximale. Si tel est le cas, la taille de lot par défaut est adaptée et tous les documents nouvellement archivés sont inclus. Si votre taux d'archivage moyen dépasse la taille de lot maximale, augmentez la taille de lot afin de permettre au moteur d'indexation d'indexer plusieurs documents dans un seul lot.

Si vous devez traiter un grand nombre de commandes en attente, vous devrez peut-être ajuster la taille de lot maximale. Cette opération doit être effectuée en prenant en compte les considérations suivantes :

- Chaque exécution de lot se divise en deux phases :
 - La phase d'indexation et de filtrage des documents, qui produit les fichiers XML. L'effort de traitement est proportionnel à la taille du lot. Le seul facteur limitatif peut être l'espace disponible et les restrictions d'accès si le nombre de fichiers est très élevé, par exemple s'il y en a plus de 500 000.
 - La phase d'indexation, qui indexe le contenu des fichiers XML. Cette phase est elle-même divisée en deux phases:
 - La phase d'indexation du contenu. L'effort et le temps de traitement au cours de cette phase est proportionnel à la taille du lot.
 - La phase de fusion, au cours de laquelle le contenu des fichiers XML indexés est inclus dans un index secondaire. L'effort de traitement au cours de cette phase augmente avec la taille de l'index. Chaque lot de documents dont la taille est à peu près équivalente demande un temps

de fusionnement plus long, ce qui augmente le temps d'indexation global. Une réorganisation des index (lorsque les index secondaire et primaire sont fusionnés) est automatiquement planifiée après l'indexation d'1,5 million de documents. Une réorganisation d'index nécessite un temps de traitement supplémentaire, qui augmente également avec la taille globale des index. Pour cette raison, aucun index doit être supérieur à 200 ou 300 Go, car la réorganisation de l'index serait trop longue. Une fois qu'un index a été réorganisé, la mise à jour suivante de l'index ne nécessite pas de phase de fusion, puisque l'index est vide. Pour cette raison, le premier processus de traitement par lots après une réorganisation d'index voit sa taille tripler automatiquement, afin de générer un débit beaucoup plus élevé.

| **Important :** Il convient de surveiller les fichiers journaux de l'indexeur IBM
| Content Collector pour vérifier les débits moyens et les temps d'exécution des
| lots individuels. Le processus de l'indexeur extrait les valeurs pertinentes à
| partir de toutes les sources d'informations disponibles. Vous ne devriez en
| principe pas avoir à surveiller les tables d'événements de Net Search Extender
| séparément. Une taille de lot maximale de plus de 300 000 n'est pas
| recommandée. Si votre taux d'archivage moyen dépasse 300 000, vous devez
| exécuter l'indexeur plusieurs fois.

IdxItemMinBatchSize

Type de valeur : Entier

Valeur par défaut : 0

Si la valeur 0 est définie, le processus de l'indexeur détermine automatiquement la taille du lot d'éléments en attente de traitement conformément à la valeur **IdxDaysConsideredForMinBatchSize**. Dès que la taille minimale du lot est atteinte, les éléments du lot sont traités. Si la taille minimale n'est pas atteinte, l'indexeur s'arrête en générant un code retour indiquant qu'il n'y a pas de documents à traiter.

Si vous souhaitez quand même exécuter l'indexation même si le lot ne contient que quelques éléments, associez l'option à 1.

IdxAutoIndexingEveryNDays

Type de valeur : Entier

La valeur par défaut est 7 (jours). Les valeurs admises sont comprises entre 0 et 30.

IdxAutoIndexingEveryNDays affecte uniquement le processus de l'indexeur si `afuIndexer` s'exécute en mode continu.

Le fait de définir **IdxAutoIndexingEveryNDays** sur une valeur supérieure à 0 jour permet de s'assurer que tous les éléments archivés lors de cette période sont indexés quelles que soient les spécifications de taille de lot. Cela permet de rechercher le contenu archivé plus rapidement, sans avoir à attendre que la taille maximale de lot soit atteinte pour que Net Search Extender démarre et que le contenu de l'élément soit indexé. Si Net Search Extender est démarré car le nombre de jours défini dans cette option est dépassé, un message est enregistré dans la console et dans le fichier journal de l'indexeur.

Le fait de définir cette option sur 0 signifie que l'indexation n'a pas eu lieu et que Net Search Extender est uniquement démarré lorsque la taille maximale de lot est atteinte.

Remarque : Le temps effectif est toujours supérieur de quelques secondes ou de quelques minutes à $N \text{ jours} * 24 \text{ heures}$.

IdxDaysConsideredForMinBatchSize

Type de valeur : Entier

Valeur par défaut : 10 (jours)

Nombre de jours à utiliser pour calculer le nombre moyen d'éléments ajoutés au type d'élément par jour. Cette moyenne est utilisée pour calculer la taille minimale du lot.

Pour plus d'informations sur l'utilisation de ces options, voir «Contraintes lors de l'indexation en mode de traitement par lots», à la page 88.

IdxReorgBatchSize

Type de valeur : Entier

Valeur par défaut : 1,5 million

Nombre d'éléments qui ont été indexés avant l'exécution automatique d'une réorganisation d'index.

IdxMaxNumberOfUpdatesWithoutReorg

Type de valeur : Entier

La valeur par défaut est de 10 millions.

Limite supérieure absolue d'éléments qui peuvent être indexés et au-delà de laquelle le traitement s'arrête jusqu'à la réorganisation de l'index. Dès que le nombre d'éléments ajoutés à l'index a atteint la valeur indiquée dans cette option et qu'aucune opération d'indexation n'a été exécutée jusqu'à présent, l'indexeur refuse d'indexer d'autres éléments et arrête le traitement en générant une erreur. L'opération de réorganisation d'index doit être lancée manuellement en exécutant la commande `afuIndexer` avec l'argument de ligne de commande `-nseindexreorg`.

IdxReorgTimeBlockIntervalBegin

Type de valeur : Chaîne au format horaire HH:mm (heures du jour 0:23)

Par défaut, aucune valeur n'est indiquée.

Cette option de configuration est utilisée avec **IdxReorgTimeBlockIntervalEnd** pour définir l'intervalle pendant lequel aucune réorganisation d'index ne peut être exécutée. Cette option définit le début de l'intervalle. Comme l'opération de réorganisation de l'index prend un certain temps et bloque la recherche dans l'index pendant une période donnée, il est conseillé de définir un intervalle pendant lequel la réorganisation d'index ne peut pas être exécutée, par exemple pendant les principales heures d'activités. Par défaut, il n'y a pas d'heure de début définie.

IdxReorgTimeBlockIntervalEnd

Type de valeur : Chaîne au format horaire HH:mm (heures du jour 0-23)

Par défaut, aucune valeur n'est indiquée.

Cette option de configuration est utilisée avec **IdxReorgTimeBlockIntervalBegin** pour définir l'intervalle pendant lequel aucune réorganisation d'index ne peut être exécutée. Cette option définit la fin de l'intervalle. Par défaut, il n'y a pas d'heure de fin définie.

IdxDelayDocumentProcessingInDays

Type de valeur : Entier

Valeur par défaut : 0

Si la valeur indiquée est 0, l'élément est traité immédiatement. La définition d'un délai de quelques jours peut être utile pour éviter le traitement redondant d'éléments identiques ajoutés au type d'élément via des demandes d'archive soumises par différents utilisateurs pendant cette période. Si le traitement de ces éléments est repoussé, vous aurez plus de chance de n'avoir à traiter ces

éléments qu'une seule fois. L'inconvénient est que ces documents ne peuvent pas être recherchés immédiatement car ils ne sont ajoutés à l'index qu'après le nombre de jours indiqués.

IdxProcessSevereErrors

Type de valeur : Booléen

Valeur par défaut : 0

Par défaut, les éléments susceptibles d'avoir provoqué l'arrêt inattendu du processus de travail pendant le traitement seront déplacés dans la table des tâches terminées avec une IDXRC de 200, ce qui indique une erreur grave. Si `IdxProcessSevereErrors` est défini sur 1, les éléments susceptibles d'avoir provoqué l'arrêt inattendu du processus de travail ne sont pas déplacés dans la table des tâches terminées avec une valeur IDXRC de 200, mais seront traités à nouveau sans les pièces jointes imbriquées lors de l'exécution suivante de l'indexeur. Les pièces jointes imbriquées ne sont pas traitées car la raison la plus courante pour laquelle le processus de travail s'arrête de manière inattendue est liée aux Filtres Oracle Outside In Technology qui sont utilisées lors du traitement des pièces jointes imbriquées.

Seulement si les éléments continuent de provoquer l'arrêt inattendu du processus de travail alors que les pièces jointes n'ont pas été traitées, l'élément sera déplacé dans la table des tâches terminées avec une valeur IDXRC de 200 lors de l'exécution suivante de l'indexeur.

Pour plus de détails, voir «Éléments ayant généré une erreur grave», à la page 106.

IdxNumberOfParallelIndexerTasks

Type de valeur : Entier

Valeur par défaut : 0

Si l'option est associée à 0, le système détermine automatiquement le nombre d'unités d'exécution approprié en fonction de la quantité d'UC système disponible. Le nombre maximal est 32.

Utilisez le paramètre par défaut si le système où l'indexeur s'exécute peut utiliser la majorité des ressources système disponibles pour la charge de l'indexeur, par exemple, lors de l'exécution de l'indexeur pendant les périodes de faibles charges.

Configurez le nombre d'unités d'exécution du traitement si le système peut uniquement utiliser une quantité limitée des ressources système (notamment les ressources d'UC) pour la charge de l'indexeur.

En général, le nombre d'unités d'exécution de l'indexeur doit correspondre au nombre de coeurs de traitement d'UC que la charge de l'indexeur doit utiliser. Par exemple, si seulement 60 % des ressources d'UC peuvent être utilisées par l'indexeur, allouez un nombre d'unités d'exécution correspondant à 60 % des coeurs de traitement d'UC.

Le système doit toujours être surveillé afin d'assurer des temps de réponse appropriés pour tous les services proposés. La diminution du nombre de tâches d'indexation parallèles libère des ressources système pour d'autres charges. Plus le nombre de tâches est grand, plus l'impact sur la consommation de mémoire est important.

Important : L'augmentation du nombre d'unités centrales affecte uniquement le traitement des documents lors de la création des fichiers XML. L'indexation qui suit la phase de création des fichiers XML repose sur une unité d'exécution simple et ne peut pas être accélérée en ajoutant davantage d'unités centrales. Si vous rencontrez de nombreuses situations où la mémoire du processus de travail dans l'indexeur est saturée, réduisez le nombre de tâches parallèles, ou spécifiez plusieurs processus de travail reposant chacun sur une seule unité

d'exécution. Pour plus de détails, voir «Utilisation de plusieurs processus à unités d'exécution multiples durant le traitement et l'indexation de documents», à la page 89.

IdxDisplayProgressIndicator

Type de valeur : Booléen

Valeur par défaut : 1

Si cette option est défini sur 1, un indicateur de progression graphique est affiché à intervalles réguliers pour toutes les opérations plus longues. Le délai séparant ces intervalles est défini dans l'option de configuration

IdxDisplayProgressDelaySecs et peut être modifié. La valeur par défaut entraîne l'affichage de la progression toutes les 30 secondes.

Il est conseillé de désactiver cette option si le processus de l'indexeur s'exécute dans un environnement dans lequel la sortie de la console est redirigée vers un fichier.

Lorsque `afuIndexer` est exécuté en mode détaillé, l'indicateur de progression graphique est remplacé par un indicateur de progression textuel.

Lorsque `afuIndexer` est exécuté en mode continu, l'indicateur de progression graphique est remplacé par une surveillance d'état globale qui affiche les statistiques du processus, les tailles de lots et le temps écoulé depuis le démarrage du processus de l'indexeur. Cette option est automatiquement activée lorsque l'indexeur est exécuté en mode continu.

IdxDisplayProgressDelaySecs

Type de valeur : Entier

Valeur par défaut : 30 secondes

Intervalle pendant lequel la progression du traitement est affichée. La valeur minimale est de 5 secondes et la valeur maximale de 600 secondes.

IdxDisplayPerformanceMonitor

Type de valeur : Booléen

Valeur par défaut : 1

Si cette option correspond à 1, le taux de traitement indiquant le nombre d'éléments traités par seconde est affiché à intervalles réguliers pour les opérations plus longues. Le délai séparant ces intervalles est défini dans l'option de configuration **IdxDisplayPerformanceMonitorDelaySecs** et peut être modifié. La valeur par défaut est un affichage toutes les 60 secondes. Le taux de traitement est également consigné dans le fichier journal sous la forme d'informations statistiques, si la consignation est activée. L'analyseur de performances s'exécute indépendamment du paramètre de l'option

IdxUseTiming.

IdxDisplayPerformanceMonitorDelaySecs

Type de valeur : Entier

Valeur par défaut : 60 secondes

Intervalle pendant lequel le taux de traitement des performances de surveillance des éléments est affiché. La valeur minimale est de 5 secondes et la valeur maximale de 600 secondes.

Le moniteur de performances affiche le nombre d'éléments qui ont été traités au taux de traitement d'éléments par seconde, notamment le taux actuel et le taux maximum. Le moniteur de performances affiche ces valeurs à un intervalle configurable défini à l'aide de **IdxDisplayPerformanceMonitorDelaySecs**.

IdxDisplayNseMonitor

Type de valeur : Booléen

Valeur par défaut : 0

IdxDisplayNseMonitor affiche l'état du processus de mise à jour de l'index Net Search Extender à intervalles réguliers en indiquant le nombre d'éléments ajoutés, mis à jour ou supprimés par Net Search Extender.

Si cette option correspond à 1, la progression des opérations de mise à jour de l'index Net Search Extender est affichée à intervalles réguliers après le démarrage de Net Search Extender. Le délai séparant ces intervalles est défini dans l'option de configuration **IdxDisplayNseMonitorDelaySecs** et peut être modifié. L'intervalle par défaut est 300 secondes (toutes les 5 minutes).

IdxDisplayNseMonitorDelaySecs

Type de valeur : Entier

Valeur par défaut : 300 secondes (5 minutes)

Intervalle pendant lequel les opérations de mise à jour de l'index Net Search Extender sont affichées. La valeur minimale est 60 secondes et la valeur maximale est 3600 secondes (1 heure).

IdxWorkerMaxMemorySizeInMb

Type de valeur : Entier

Valeur par défaut : 1024 Mo

Taille maximale de la mémoire de segments de processus de travail de la machine virtuelle Java. Si vous constatez que les processus de travail n'ont pas suffisamment de mémoire et qu'ils s'arrêtent, augmentez progressivement la taille définie pour l'option **IdxWorkerMaxMemorySizeInMb**. La taille maximale de la mémoire de segments de processus de travail de la machine virtuelle Java est 3072 Mo.

IdxUpdateItemTypeTable

Type de valeur : Booléen

Valeur par défaut : 0

Si **IdxUpdateItemTypeTable** est défini sur 1, le processus de l'indexeur n'ajoutera pas seulement l'élément traité à la table des tâches terminées avec une valeur IDXRC, mais ajoutera également la valeur IDXRC à la colonne TIEFLAG de l'élément dans la table de composant de type d'élément IBM Content Manager.

Cela peut être utilisé si les applications, telles que eDiscovery Manager, souhaitent évaluer la valeur TIEFLAG en tant qu'indicateur pour déterminer si un élément a été indexé avec succès ou non.

Si vous ne souhaitez pas utiliser la valeur TIEFLAG, ou ne pas utiliser des applications qui utilisent la valeur TIEFLAG, vous devez définir cette option sur 0 afin d'enregistrer le temps système de traitement supplémentaire nécessaire lors de la mise à jour de la valeur TIEFLAG.

XmlModdefLocatorBase

Type de valeur : Chaîne

La valeur par défaut est /icc_document pour les types d'élément du modèle de données composé.

Si le modèle de données du type d'élément est le modèle intégré, la valeur doit être remplacée par /document.

Cette valeur doit correspondre au préfixe des localisateurs moddef utilisés dans le fichier modèle associé à la configuration de l'indexeur du type d'élément.

Remarque : Seuls les caractères alphanumériques suivants peuvent être utilisés dans une chaîne de localisateur moddef.

- Tous les caractères alphabétiques (A à Z et a à z) et les nombres compris entre 0 et 9
- Les caractères /_@ (où @ a une signification spéciale)

XmlReceivedDateAndTime

Type de valeur : Booléen

Valeur par défaut : 0

Si cette option correspond à 0, l'attribut du type d'élément `received_date` contient uniquement la date et l'heure. Il s'agit du format de date par défaut pour l'attribut `received_date` dans des types d'élément composés. Il s'agit du format de date recommandé pour tous les nouveaux types d'élément.

Si cette option correspond à 1, l'attribut du type d'élément `received_date` contient la date et l'heure précise. Il s'agit du format de date par défaut pour l'attribut `received_date` des types d'élément intégrés.

Options de configuration de l'analyse syntaxique des documents

Les options suivantes ont une incidence directe sur l'analyse syntaxique des documents. Elles ont également une incidence sur l'analyse syntaxique des documents effectués par les Filtres Oracle Outside In Technology. Ces filtres sont appelés pour toutes les pièces jointes imbriquées des messages électroniques et les documents du système de fichiers et de SharePoint. Le terme "document" est utilisé pour désigner à la fois les pièces jointes des messages électroniques et les documents du système de fichiers et de SharePoint.

MsgMapAllUnsupportedCodepagesToDefault

Type de valeur : Booléen

Valeur par défaut : 1

Si cette option correspond à 1, tous les documents qui possèdent une page de codes non prise en charge ou non autorisée répertoriée par l'option **MsgCodepageMapTable** sont traités comme s'ils possédaient la page de codes indiquée par l'option **MsgDefaultCodePageForNonUnicodeFiles** et sont convertis. Cette option de configuration est uniquement appropriée lors du traitement de documents Microsoft Exchange.

MsgDefaultCodePageForNonUnicodeFiles

Type de valeur : Entier

Page de codes par défaut : 1252

Indique la page de codes à utiliser par défaut si un document Exchange possède une page de codes qui n'est pas prise en charge. Cette option est nécessaire uniquement si l'option **MsgCodepageMapTable** ou **MsgMapAllUnsupportedCodepagesToDefault** est également définie. Si cette option est omise, la page de codes de l'environnement local en cours est utilisée comme page de codes par défaut.

Cette option de configuration est uniquement appropriée lors du traitement de documents Microsoft Exchange.

MsgCodepageMapTable

Type de valeur : Chaîne

La valeur par défaut est aucune valeur.

Indique une série de pages de codes (séparées par des espaces) qui ne sont pas prises en charge par l'indexeur. Si le document utilise l'une de ces pages de codes, il est converti comme s'il possédait la page de codes indiquée par l'option **MsgDefaultCodePageForNonUnicodeFiles**. Cette option permet de gérer une liste de pages de codes qui peuvent être converties dans la page de codes indiquée par l'option **MsgDefaultCodePageForNonUnicodeFiles** car leurs jeux de

caractères sont pratiquement identiques à ceux de la page de codes indiquée par l'option **MsgDefaultCodePageForNonUnicodeFiles**.

Remarque : Toutes les pages de codes indiquées doivent être des valeurs numériques. Seuls les CCSID IBM peuvent être utilisés, pas les pages de codes ISO. Par exemple : iso-8859-1 = CCSID 1252.

Cette option de configuration est uniquement appropriée lors du traitement de documents Microsoft Exchange.

MsgCodepageIgnoreTable

Type de valeur : Chaîne

Valeur par défaut : 42

Indique une série de pages de codes (séparées par des espaces) qui doivent être ignorées par l'indexeur. Si le documents utilise l'une de ces pages de codes, il n'est pas converti. Cela permet d'ignorer des pages de codes telles que Windows 42, qui est une page de codes de symboles qui n'est pas prise en charge par l'indexeur et pour laquelle une conversion en contenu textuel serait utile.

Remarque : Toutes les pages de codes répertoriées doivent être constituées de valeurs numériques uniquement. Seuls les CCSID IBM peuvent être utilisés, pas les pages de codes ISO. Par exemple, iso-8859-1 est CCSID 1252.

Cette option de configuration est uniquement appropriée lors du traitement de documents Microsoft Exchange.

TxtCnvEnabled

Type de valeur : Booléen

Valeur par défaut : 1

Si cette option est définie sur 1, la conversion de document est activée.

Si **TxtCnvEnabled** est défini sur 0, les documents ne sont pas convertis dans leur représentation textuelle requise pour l'activation de l'indexation. Cette option de configuration doit uniquement être définie sur 0 si vous ne souhaitez pas indexer des pièces jointes de courriers électroniques lorsque vous utilisez des types d'élément de courriers électronique. Ne définissez pas cette option de configuration sur 0 lorsque vous utilisez des types d'élément Microsoft SharePoint et Système de fichiers.

TxtCnvPlainTextHandlingForUnsupportedFormat

Type de valeur : Booléen

Valeur par défaut : 1

Si cette option correspond à 1, le processus tente de traiter tous les documents dont le format n'est pas pris en charge en tant que documents de texte normal. Pour plus d'informations, reportez-vous à la rubrique relative au traitement des pièces jointes imbriquées dans les messages électroniques et des documents.

TxtCnvAutoCodepageDetection

Type de valeur : Entier

Valeur par défaut : 1

Si cette option correspond à 1, le processus de l'indexeur tente de détecter automatiquement la page de codes des documents qui n'ont pas pu être analysés par les Filtres Oracle Outside In Technology. En fonction du niveau de confiance établi lors de la détection de la page de codes, le document est traité ou associé à une erreur indiquant que la détection automatique de la page de

codes a échoué. Pour plus d'informations, reportez-vous à la rubrique relative au traitement des pièces jointes imbriquées dans les messages électroniques et des documents.

TxtCnvAutoCodepageDetectionMinConfidence

Type de valeur : Entier

Valeur par défaut : 70

La détection automatique de la page de codes est considérée comme réussie si le niveau minimal de confiance atteint est supérieur ou égal à la valeur définie pour l'option **TxtCnvAutoCodepageDetectionMinConfidence**. Si le niveau de confiance est supérieur au niveau défini, le document est converti à partir de sa page de codes et traité en tant que document de texte normal. Le niveau minimal de confiance (entre 1 et 100 %) est calculé si l'option **TxtCnvAutoCodepageDetection** correspond à 1. Pour plus d'informations, reportez-vous à la rubrique relative au traitement des pièces jointes imbriquées dans les messages et les documents.

TxtCnvDefaultCodepageForPlainText

Type de valeur : Entier

Valeur par défaut : 0

Si **TxtCnvDefaultCodepageForPlainText** n'est pas indiqué ou est défini sur 0, tous les documents pour lesquels aucune page de codes n'a pu être détectée sont considérés comme étant en texte en clair et indexés en tant que texte en clair.

Si la valeur définie dans **TxtCnvDefaultCodepageForPlainText** est différente de 0, tous les documents qui ne sont pas traités autrement sont considérés comme étant codés dans la page de codes spécifiée et convertis à partir de cette page de codes pour traitement.

TxtCnvMaxExtendedCharactersRatio

Type de valeur : Entier

Valeur par défaut : 30

Pourcentage de caractères étendus (point de code > 0x7F) qu'un document peut contenir. Ce pourcentage est calculé si la détection automatique de la page de codes échoue ou est désactivée. Si le rapport est dépassé, la valeur IDXRC du document est définie sur 2. Pour plus d'informations, voir «Recherche de chaînes d'erreur», à la page 114.

TxtCnvTargetBufferSizeInMB

Type de valeur : Entier

Valeur par défaut : 10

Quantité maximale d'espace pour chaque unité d'exécution que la section de texte extraite d'un document analysé peut remplir. Si le contenu textuel pur du document ne peut pas être stocké dans cette mémoire tampon, il est tronqué et marqué comme tel. La valeur minimale est 1 Mo pour chaque unité d'exécution et la valeur maximale est 256 Mo.

Ne choisissez pas une valeur trop élevée car une mémoire tampon cible de conversion est créée pour chaque tâche (unité d'exécution) parallèle qui est exécutée par la tâche de travail de l'indexeur. Cela signifie que la quantité de mémoire requise pour le processus est le produit du nombre de tâches (unités d'exécution) parallèles fois la taille de la mémoire tampon de conversion spécifiée.

A faire : La valeur de cette option de configuration doit être inférieure à la taille de la mémoire tampon de transfert de la fonction définie par l'utilisateur. La mémoire tampon de transfert de la fonction définie par l'utilisateur est

utilisée lorsque la fonction définie par l'utilisateur AFUFetchFile charge les fichiers XML temporaires dans la mémoire (qui est la mémoire tampon de transfert) pour un accès par Net Search Extender. La taille par défaut de la mémoire tampon de transfert de la fonction définie par l'utilisateur est de 32 Mo. Pour plus d'informations sur la manière de modifier la taille de la mémoire tampon de transfert de la fonction définie par l'utilisateur, voir «Modification de la taille de la mémoire tampon de transfert par défaut de la fonction définie par l'utilisateur AFUFetchFile», à la page 87.

TxtCnvTimeOutInSecs

Type de valeur : Entier

Valeur par défaut : 120

Délai, en secondes, pendant lequel le document doit être analysé. La valeur minimale est de 30 secondes et la valeur maximale est de 3600 secondes (1 heure). Une fois le délai écoulé, le document est marqué et associé à une erreur de dépassement de délai.

Options de configuration secondaires de l'indexeur

Certaines options de configuration sont uniquement utilisées à des fins de rétrocompatibilité et sont toujours configurées avec une valeur par défaut. Comme pour toutes les options de configuration, vous pouvez modifier la valeur par défaut si vous le souhaitez. Cependant, ces options de configuration sont uniquement exportées ou répertoriées comme partie intégrante de la configuration si la valeur n'est plus la valeur par défaut du système.

Les options de configuration les moins importantes sont les suivantes :

GenTimingFormatFieldWidth

Type de valeur : Entier

Valeur par défaut : 8

Largeur maximale de chaque zone de date/heure (colonne) dans le fichier de date/heure au format CSV.

IdxIgnoreUnknownCmAttributes

Type de valeur : Booléen

Valeur par défaut : 1

Un attribut IBM Content Manager inconnu est un attribut répertorié dans la configuration de l'indexeur de recherche de texte mais il n'est pas défini dans le type d'élément IBM Content Manager cible. Si cette option correspond à 1 et qu'un attribut IBM Content Manager est détecté, le traitement continue sans générer d'avertissement. Si elle correspond à 0, le traitement s'arrête en générant une erreur.

IdxMaxNumberOfErrorsDisplayed

Type de valeur : Entier

Valeur par défaut : 0

Indique le nombre maximal de messages d'erreur affichés pendant le traitement de l'indexeur. Si le nombre de messages d'erreur dépasse la valeur indiquée, un avertissement est émis pour signaler que les erreurs suivantes ne seront plus affichées. Toutefois, le traitement de l'indexeur continue de consigner toutes les erreurs dans le fichier journal si la consignation est activée. Si **IdxMaxNumberOfErrorsDisplayed** correspond à 0, toutes les erreurs sont affichées.

Si **IdxUseLogging** est défini sur 0 (la consignation est désactivée), toutes les

erreurs sont toujours affichées, ce qui signifie que **IdxMaxNumberOfErrorsDisplayed** est ignorée, quelle que soit sa valeur.

IdxDisplayTimeStamp

Type de valeur : Booléen

Valeur par défaut : 0

Si cette option correspond à 1, tous les messages d'informations sont précédés d'un horodatage.

L'option est automatiquement activée lorsque le processus de l'indexeur est exécuté en mode continu.

IdxWorkerMinMemorySizeInMb

Type de valeur : Entier

Valeur par défaut : 1024 Mo

Taille minimale de la mémoire de segments de processus de travail de la machine virtuelle Java.

Si vous définissez la valeur sur 0, la machine virtuelle Java détermine automatiquement une valeur minimale.

XmlDeleteFilesAfterNseProcessing

Type de valeur : Booléen

Valeur par défaut : 1

Si cette option correspond à 1, les fichiers XML sont automatiquement supprimés après l'exécution du processus Net Search Extender.

Si cette option correspond à 0, les fichiers ne sont pas automatiquement supprimés. Vous devez les supprimer manuellement. Les fichiers XML se trouvent dans le répertoire indiqué par l'option **GenXmlDirectory**.

Remarque : Le répertoire XML est partagé par l'ensemble des processus de l'indexeur pour tous les types d'élément. Vous devez donc supprimer les fichiers XML avec prudence si des instances de `afuIndexer` sont en cours d'exécution au même moment pour d'autres types d'élément. Si aucune autre instance de `afuIndexer` n'est en cours d'exécution, l'arborescence du répertoire complète spécifiée dans l'option de configuration **GenXmlDirectory** peut être supprimée. Le processus de l'indexeur recrée automatiquement la structure du répertoire si nécessaire.

XmlKeepFilesCausingNseErrors

Type de valeur : Booléen

Valeur par défaut : 0

En général, si l'option **XmlDeleteFilesAfterNseProcessing** correspond à 1, tous les fichiers XML sont automatiquement supprimés. Si

XmlKeepFilesCausingNseErrors correspond à 1, les fichiers XML qui ont généré l'erreur Net Search Extender ne sont pas supprimés.

Si l'option **XmlKeepFilesCausingNseErrors** correspond à 0 et que l'option **XmlDeleteFilesAfterNseProcessing** correspond à 1, tous les fichiers XML sont supprimés, y compris ceux à l'origine de l'erreur Net Search Extender.

XmlInsertAttachmentProcessingWarnings

Type de valeur : Booléen

Valeur par défaut : 1

Si un avertissement (mais pas une erreur) est émis pendant le traitement des documents (pièce jointe imbriquée d'un message électronique ou document du système de fichiers ou de SharePoint) et que l'option correspond à 1, une chaîne d'avertissement `IcmFceWarning: <message d'avertissement>` est indexée avec le document.

Si l'option correspond à 0, la chaîne d'avertissement `IcmFceWarning: <a warning message>` n'est pas indexée avec le document.
Pour connaître la liste des chaînes d'avertissement stockées avec l'élément, reportez-vous à la rubrique relative à la recherche des chaînes d'erreur.

XmlTransformationXslFile

Type de valeur : Chaîne

Par défaut, aucune valeur n'est indiquée.

Indique le nom d'un fichier XSL qui décrit la transformation XML appliquée au document XML généré avant de l'écrire sur le disque. Vous pouvez spécifier un nom de fichier avant ou sans le chemin d'accès au fichier. Si vous ne spécifiez pas de chemin d'accès, le fichier est recherché dans le répertoire en cours et dans le sous-répertoire `cfg` du répertoire défini par la variable d'environnement `AFU_INDEXER_ROOT`.

Cette option n'est généralement requise que pour les types d'élément Microsoft SharePoint auquel cas **XmlTransformationXslFile** pointe automatiquement vers le fichier XSL `afu_sp.xsl` fourni par le composant Content Collector Text Search Support.

XmlAttributeRequiresNseDateConversion

Type de valeur : Chaîne

Par défaut, aucune valeur n'est indiquée.

Spécifie un attribut dont la valeur date/heure doit être convertie d'une représentation ISO en une représentation convenant à l'indexation par Net Search Extender et permettra ainsi la recherche de la valeur de l'attribut.

Cette option est nécessaire uniquement pour les types d'éléments IBM Connections, auquel cas **XmlAttributeRequiresNseDateConversion** est automatiquement défini sur `icc_date`.

TxCnvGeneralErrorAborts

Type de valeur : Booléen

Valeur par défaut : 0

Si cette option correspond à 1 et qu'une erreur grave se produit lors du traitement des Filtres Oracle Outside In Technology, l'élément correspondant est associé au code d'erreur 199 dans la colonne `IDXRC` de la table des tâches terminées et aucune tentative n'est effectuée pour traiter l'élément. Les pièces jointes ou les documents restants associés à l'élément ne sont pas traités.

Si cette valeur correspond à 0 et qu'une erreur se produit pendant le traitement du document, l'erreur est ignorée et le document est indexé avec une erreur qui peut faire l'objet d'une recherche. Tous les documents qui ont été associés à une erreur peuvent ainsi être identifiés. Pour obtenir la liste des codes d'erreur renvoyés, reportez-vous à la rubrique relative aux codes retour de l'index.

TxCnvOutsideInExtendedTest

Type de valeur : Booléen

Valeur par défaut : 1

Si cette option correspond à 1, les Filtres Oracle Outside In Technology exécutent un test étendu pour identifier le type de document. Elle prend en charge une détection plus précise du type de document mais cette analyse étendue requiert davantage de ressources d'UC.

TxCnvOutsideInFilterGroups

Type de valeur : Booléen

Valeur par défaut : 1

Utilisez les groupes de documents Filtres Oracle Outside In Technology pour filtrer des types de fichiers spécifiques. Ces documents sont filtrés et ne sont

pas traités car leur contenu ne peut pas faire l'objet d'une indexation (par exemple, des fichiers graphiques ou binaires).

MimeGeneralErrorAborts

Depuis IBM Content Collector V3, cette option de configuration n'est plus prise en charge. Si elle est trouvée dans un fichier de configuration, elle sera ignorée.

MimeTreatHtmlAsAttachment

Depuis IBM Content Collector V3, cette option de configuration n'est plus prise en charge. Si elle est trouvée dans un fichier de configuration, elle sera ignorée.

Options de configuration liées à la fonction définie par l'utilisateur AFUFetchFile

La fonction définie par l'utilisateur AFUFetchFile comporte sa propre configuration qui est automatiquement créée par le processus de l'indexeur. La configuration de la fonction définie par l'utilisateur utilise les options de configuration afuIndexer suivantes :

- **UdfMaxNumberOfWarnings**
- **UdfUseLogging**
- **UdfAutoConfigReloadInMin**
- **GenXmlDirectory**
- **GenLogDirectory**

Lorsque la valeur de l'une de ces options de configuration est modifiée, la configuration de la fonction définie par l'utilisateur est automatiquement mise à jour. Une fois la configuration de la fonction définie par l'utilisateur mise à jour, l'utilisateur en est informé par un message. Selon la valeur de l'option de configuration **UdfAutoConfigReloadInMin**, l'utilisateur est invité à redémarrer DB2 pour que les modifications soient prises en compte ou à attendre le temps que la fonction définie par l'utilisateur recharge la configuration.

Une fois la configuration mise à jour, le processus de l'indexeur se termine avec le code retour de processus 4. Le processus de l'indexeur doit être redémarré manuellement.

Les options de configuration AFUFetchFile suivantes associées à la fonction définie par l'utilisateur sont partagées entre toutes les configurations de type d'élément :

UdfMaxNumberOfWarnings

Type de valeur : Entier

Valeur par défaut : 100 La valeur maximale est 9999.

Si le nombre d'avertissements émis pendant l'exécution de l'index Net Search Extender est supérieur à la valeur indiquée, UDF renvoie une erreur grave à Net Search Extender et Net Search Extender arrête le traitement.

Cela signifie que les éléments qui ont déjà été traités en entrée par Net Search Extender ne sont pas supprimés de la table des tâches ouvertes et ajoutés à la table des tâches supprimées. Tous les éléments traités pendant cette exécution sont une nouvelle fois traités lors de la prochaine exécution de la commande afuIndexer.

UdfUseLogging

Type de valeur : Booléen

Valeur par défaut : 0

Si cette option correspond à 1, la fonction définie par l'utilisateur AFUFetchFile

ajoute des informations de consignation au fichier `AfuFetchFile.log` dans le répertoire de consignation indiqué par l'option de configuration **GenLogDirectory**.

UdfAutoConfigReloadInMin

Type de valeur : Entier

Valeur par défaut : 0

Nombre de minutes avant que la fonction définie par l'utilisateur `AFUFetchFile` ne recharge automatiquement sa configuration. La configuration se trouve soit dans le sous-répertoire temporaire par défaut du répertoire d'installation IBM Content Manager, soit dans le chemin de fichier défini par la variable d'environnement `AFU_FETCHFILE_CONFIG_PATH`. Une valeur par défaut de 0 seconde indique que la configuration n'est pas automatiquement rechargée. Comme la fonction définie par l'utilisateur `AFUFetchFile` est lancée après DB2 et que Net Search Extender est exécuté pour la première fois, les modifications de configuration apportées à la fonction définie par l'utilisateur requièrent un redémarrage de DB2. Le rechargement forcé et automatique de la configuration à intervalles réguliers permet d'éviter le redémarrage de DB2.

Le rechargement automatique de la configuration permet des modifications et tests répétés de l'une des options de configuration suivantes sans avoir à redémarrer DB2 :

- **UdfMaxNumberOfWarnings**
- **UdfUseLogging**
- **GenXmlDirectory**
- **GenLogDirectory**

Dès que la configuration est stable et ne nécessite plus de modifications, le rechargement automatique doit être désactivé en définissant **UdfAutoConfigReloadInMin** sur 0.

Si vous devez reconfigurer la taille de la mémoire tampon de transfert de la fonction définie par l'utilisateur `AFUFetchFile` car la taille de la mémoire tampon est trop faible pour le contenu textuel extrait d'un élément, reportez-vous à la rubrique «Modification de la taille de la mémoire tampon de transfert par défaut de la fonction définie par l'utilisateur `AFUFetchFile`» pour obtenir des informations complémentaires.

Modification de la taille de la mémoire tampon de transfert par défaut de la fonction définie par l'utilisateur `AFUFetchFile`

La fonction définie par l'utilisateur `AFUFetchFile` charge en mémoire les fichiers XML temporaires pour permettre à Net Search Extender d'y accéder.

La taille de cette mémoire tampon correspond à 32 Mo (valeur par défaut) lorsque la fonction définie par l'utilisateur `AFUFetchFile` est enregistrée pour le type d'élément et que celui-ci est activé pour être traité par l'indexeur de recherche de texte. Cette valeur est généralement suffisante. Toutefois, il est possible que la taille soit trop faible si le contenu textuel extrait de l'élément est supérieur à 32 Mo.

Le processus `afuIndexer` vérifie que l'espace de la mémoire tampon de transfert est suffisant pour chaque fichier XML. Si l'espace est insuffisant, l'élément n'est plus traité. L'élément est placé dans la table des tâches terminées et une valeur `IDXRC` de 130 lui est affectée.

Si cela se produit sur de nombreux éléments, il est recommandé d'augmenter la taille de la mémoire tampon de transfert en redéfinissant la fonction définie par l'utilisateur pour prendre en charge des données volumineuses.

La taille par défaut de la mémoire tampon de transfert définie par l'utilisateur est définie lors de l'activation du type d'élément. Si un type d'élément est activé manuellement à l'aide de l'outil `afuEnableItem`, vous pouvez modifier la taille par défaut à l'aide de l'argument de ligne de commande `-udftransferbuffersize`. Pour plus d'informations, voir «Activation de types d'éléments qui n'ont pas été créés dans Content Collector», à la page 16.

Pour vérifier la taille de la mémoire tampon définie par l'utilisateur en cours `AFUFetchFile`, exécutez l'instruction SQL suivante :

```
DB2 SELECT LENGTH FROM SYSCAT.FUNCPARMS WHERE  
      FUNCNAME = 'AFUFETCHFILE' AND ROWTYPE = 'C'
```

Ceci permet d'afficher la taille de la mémoire tampon en octets.

Pour redéfinir la taille de la mémoire tampon de transfert de la fonction définie par l'utilisateur `AFUFetchFile` :

1. Modifiez le script de définition de données `DB2 AFUFetchFile.ddl` dans le sous-répertoire `bin` de l'installation de Content Collector Text Search Support. Remplacez la valeur 32 Mo de l'objet `CLOB` par une valeur appropriée.
2. Exécutez le script avec la commande suivante :

```
db2 -tvf AFUFetchFile.ddl
```

Le script abandonne la fonction définie par l'utilisateur et la redéfinit avec la nouvelle taille maximale des données de renvoi.

Contraintes lors de l'indexation en mode de traitement par lots

L'exécution du processus d'indexeur dans une planification et sur un lot d'éléments est influencée par les limitations qui sont définies pour la taille du lot.

Si l'option de configuration `IdxItemMinBatchSize` correspond à 0, le processus d'indexation tente de déterminer automatiquement le nombre minimal de documents qui doivent se trouver dans la table des tâches ouvertes avant le lancement de l'indexation. Si le nombre de documents de la table des tâches ouvertes est inférieur à la valeur calculée, le processus de l'indexeur s'arrête et génère un code retour indiquant qu'il n'y a pas d'élément à traiter actuellement. Si le nombre de documents de la table des tâches ouvertes est supérieur à la valeur calculée, l'indexeur commence à traiter ces éléments immédiatement.

La formule utilisée pour calculer automatiquement le seuil auquel le processus commence à traiter les éléments est la suivante :

Moyenne arithmétique du nombre d'éléments ajoutés au type d'élément au cours des N derniers jours, où N est la valeur de l'option de configuration `IdxDaysConsideredForMinBatchSize`.

La valeur par défaut (nombre de jours) est 10.

Le calcul de cette formule est basé sur le contenu de la table de base de données du type d'élément IBM Content Collector et peut durer si la table de la base de données ne dispose pas d'un index et qu'une analyse complète de la table est requise. Pour éviter une analyse complète de la table, il est conseillé de créer un

index de base de données pour la table de base de données de type d'élément afin d'accélérer les calculs suivants de cette formule.

Pour créer un index de table de base de données, vous devez connaître le nom de la table de la base de données du type d'élément sous-jacent. Ce nom de table est affiché en regard de l'intitulé Item type component id table name lorsque l'outil afuIndexer ou afuCheck est appelé pour le type d'élément concerné. Appelez afuIndexer ou afuCheck à l'aide de l'option de ligne de commande -verbose.

Par exemple, les informations suivantes s'affichent lorsque l'un des outils est appelé :

```
Item type component id table name.....: ICMUT01363001
```

Using this database table name you can create a database index using the following DB2 command:

```
DB2 CREATE INDEX <schema>.<indexname> ON  
<schema>.<tablename> ("CREATETS" ASC) ALLOW REVERSE SCANS;
```

Si l'option de configuration **IdxItemMinBatchSize** est définie sur une valeur supérieure à 0, l'indexeur utilise la valeur indiquée comme taille minimale de lot. Si le nombre de documents figurant dans la table des tâches ouvertes est inférieur à la valeur indiquée, le processus de l'indexeur s'arrête et génère un code retour indiquant qu'il n'y a pas d'élément à traiter actuellement. Si le nombre de documents figurant dans la table des tâches ouvertes est supérieur ou égal à la valeur indiquée, l'indexeur commence à traiter ces éléments immédiatement.

Si **IdxItemMinBatchSize** est défini sur 1, l'indexation s'exécute même si le lot ne contient qu'un seul élément.

Remarque : L'argument de ligne de commande `-minbatchsize` remplace provisoirement la valeur de l'option de configuration **IdxItemMinBatchSize** et désactive le calcul automatique de la taille du lot d'éléments.

Le processus afuIndexer demande à Net Search Extender de lancer l'indexation lorsque les conditions suivantes sont remplies :

- Le nombre d'éléments à traiter dépasse la taille de lot minimale définie
- Il n'y a plus d'éléments dans la table des tâches ouvertes
- Le nombre d'éléments lus dans la table des tâches ouvertes correspond à la taille maximale du lot définie par l'option de configuration **IdxItemMaxBatchSize** ou à la taille maximale du lot transmise via l'argument `-maxbatchsize` lors de l'appel de afuIndexer.

Si une réorganisation d'index a été effectuée lors de la dernière opération de mise à jour, la taille maximale du lot est automatiquement réajustée. La taille de lot maximale est triplée lors de la première exécution faisant suite à une réorganisation de l'index.

Utilisation de plusieurs processus à unités d'exécution multiples durant le traitement et l'indexation de documents

Cette information s'applique à IBM Content Collector V3.0.0.1 et versions supérieures : le processus afuIndexer peut être exécuté à l'aide de processus de travail à unités d'exécution multiples ou de plusieurs processus de travail à unité d'exécution simple. Lors du traitement par plusieurs processus de traitement à unité d'exécution simple, chaque processus de travail concentre toutes les

ressources de traitement sur un seul document à la fois, ce qui permet d'augmenter la puissance de traitement et d'indexation des documents. Lors du traitement au moyen d'unités d'exécution multiples, les ressources de traitement des documents sont partagées entre plusieurs documents au sein du même processus.

Si des processus de traitement à unités d'exécution multiples sont utilisés par opposition à plusieurs processus de travail à unité d'exécution simple, `afuIndexer` utilise deux processus Java, le processus de l'indexeur principal et un processus de traitement. Pour garantir la robustesse 24 x 7 des performances, un processus de traitement distinct est utilisé pour interagir avec l'instance `Filtres Oracle Outside In Technology` dont l'appel a lieu pendant le traitement de document en vue d'extraire le contenu textuel. L'utilisation d'un processus de traitement distinct est explicitement recommandée par `Filtres Oracle Outside In Technology`, car la qualité des documents traités est inconnue et les documents de format incorrect peuvent provoquer un comportement inattendu lors du traitement. En utilisant l'option de configuration `IdxNumberOfParallelIndexerTasks`, ce processus de traitement distinct peut être configuré pour créer une ou plusieurs unités d'exécution capables de gérer plusieurs documents en parallèle et de les préparer pour l'indexation par `Net Search Extender`. Dans la plupart des cas, le traitement et l'indexation des documents aboutit. Toutefois, il peut arriver qu'un document contenant une pièce jointe incorrecte ou endommagée ne puisse pas être traité par l'instance `Filtres Oracle Outside In Technology` et que cette pièce jointe entraîne l'échec du processus de traitement. Le processus de l'indexeur principal détecte cette erreur et démarre automatiquement une autre instance du processus de traitement afin de poursuivre et terminer le lot.

Lorsque l'argument de ligne de commande `-usemultiprocessing` est utilisé, `afuIndexer` crée plusieurs processus de traitement comportant chacun une unité d'exécution au lieu de créer plusieurs unités d'exécution comportant un processus de traitement unique.

Important : Sur le système d'exploitation AIX, la création de plusieurs processus de traitement est l'option par défaut. Sur tous les autres systèmes d'exploitation, l'option par défaut consiste à créer plusieurs unités d'exécution au sein d'un seul processus de traitement. Si vous souhaitez créer plusieurs processus de traitement, vous devez le spécifier manuellement en utilisant l'argument de ligne de commande `- usemultiprocessing` lors de l'exécution de `afuIndexer`.

L'avantage de plusieurs processus de travail à unité d'exécution unique est que les ressources de traitement des documents ne sont pas partagées par un grand nombre de documents au sein du même processus et que les incidents qui se produisent dans le contexte du traitement multiprocessus et de `Filtres Oracle Outside In Technology` peuvent être évités.

Si l'un des processus de traitement s'arrête de façon inattendue en raison d'une pièce jointe incorrecte ou endommagée, ou si la pièce jointe endommagée ne peut pas être traitée par l'instance `Filtres Oracle Outside In Technology`, un autre processus de traitement démarre automatiquement et poursuit le traitement du lot.

L'inconvénient de l'utilisation de processus de traitement multiples est que la mémoire virtuelle requise sur la machine augmente en même temps que le nombre de processus de traitement. De plus, comme le planificateur du système d'exploitation est responsable de l'attribution du temps processeur à chaque processus de traitement, la distribution des documents qui seront traités par ces

processus de traitement ne peut pas être contrôlé précisément par les principaux processus de l'indexeur. La répartition des documents entre les processus de travail peut donc être inégale.

L'option de configuration **IdxNumberOfParallelIndexerTasks** continue de contrôler le traitement parallèle. Dans le cas d'un traitement faisant intervenir plusieurs processus de traitement à unité d'exécution simple (par opposition au traitement multitâche), le nombre de processus de travail démarrés est déterminé.

Le passage d'un processus de traitement à unités d'exécution multiples à l'utilisation de processus de traitement à unité d'exécution simple produit les effets suivants :

- Le nombre de processus de travail parallèles qui est calculé automatiquement si l'option de configuration **IdxNumberOfParallelIndexerTasks** est définie sur 0 dépend du nombre d'unités centrales dont est équipé le système. Un processus de traitement est démarré pour chaque unité centrale.
Si des processus de traitement à unités d'exécution multiples sont utilisés, le nombre d'unités d'exécution qui est calculé automatiquement si l'option de configuration **IdxNumberOfParallelIndexerTasks** est définie sur 0 dépend également du nombre d'unités centrales dont est équipé le système. Dans ce cas toutefois, deux unités d'exécution sont créées dans le processus de traitement pour chaque unité centrale.
- Le nombre maximal de processus de traitement automatiquement calculé si l'option de configuration **IdxNumberOfParallelIndexerTasks** est définie sur 0 est limité à 8, même si le nombre d'unités centrales dans la machine est plus élevé. Si vous souhaitez exécuter plus de 8 processus de traitement parallèles, vous devez désactiver le calcul automatique du nombre de tâches parallèles et définir **IdxNumberOfParallelIndexerTasks** sur la valeur souhaitée. Assurez-vous que votre ordinateur possède assez de mémoire virtuelle (espace de permutation) pour fournir une mémoire suffisante pour tous ces processus.
- La quantité de mémoire utilisée par chaque processus de traitement reste déterminée par les options de configuration **IdxWorkerMaxMemorySizeInMb** et **IdxWorkerMinMemorySizeInMb**. Assurez-vous que votre ordinateur possède assez de mémoire virtuelle (espace de permutation) pour fournir une mémoire suffisante pour tous les processus en fonction de vos paramètres de configuration. Pour plus de détails sur l'utilisation de ces options de configuration, voir «Options de configuration de l'indexeur», à la page 67. La taille de mémoire allouée à chaque processus de traitement via l'option de configuration **IdxWorkerMinMemorySizeInMb** doit être une valeur faible, telle que 256. La quantité maximale de mémoire allouée à chaque processus de traitement à l'aide de l'option de configuration **IdxWorkerMaxMemorySizeInMb** ne doit pas dépasser 1024.
- Si l'argument de ligne de commande `-trace` est spécifié, un fichier de trace distinct est créé pour chaque processus de traitement. Les fichiers de trace des processus de traitement sont reconnaissables à leur infixe `worker`, à l'ID de processus de travail et à un nombre qui indique l'instance du processus de traitement pour l'ID du processus de travail. Chaque fois qu'une nouvelle instance de processus de travail est créée en raison de l'arrêt inattendu du processus, un nouveau numéro d'instance est utilisé.
- Le compteur statistique `WORKER_INSTANCES_LAUNCHED` affiche toujours le nombre de processus de travail qui ont été démarrés par le processus de l'indexeur, mais ce nombre n'est plus égal à 1 par défaut, ce qui indique un seul processus de travail. Au lieu de cela, ce nombre dépend désormais du nombre

de processus de travail qui sont démarrés. Si un processus de travail à unités d'exécution multiples est utilisé, chaque processus de travail démarré est également comptabilisé.

- Si vous spécifiez le paramètre `-verbose` ou `-showsql` pendant l'indexation, il se peut que vous voyiez s'afficher le même message plusieurs fois, car il est généré séparément par chaque processus de travail. Par exemple, le message qui signale qu'un arrêt a été notifié sera affiché sur la console pour chaque processus de travail séparément si l'argument de ligne de commande `-verbose` est spécifié. En revanche, chaque processus de travail utilise son propre nom ; à titre d'exemple, le message peut commencer par le préfixe `worker1-1`. De cette manière, chaque message peut être relié au processus de travail correspondant.
- Les modes de traitement par lot et en continu utilisent plusieurs processus de travail si le multitraitement est activé.
- Le mode d'erreur grave et le mode de débogage n'effectuent pas de traitement parallèle et ne sont pas affectés par le traitement multiprocessus.

Exécution de `afuIndexer` en mode continu

Le processus de l'indexeur `afuIndexer` peut être configuré pour s'exécuter dans une planification sur une taille de lot d'éléments définie ou pour s'exécuter en mode continu jusqu'à interruption par une commande d'arrêt explicite.

L'exécution du processus de l'indexeur dans une planification sur un lot d'éléments est le mode de traitement par défaut.

Pour exécuter le processus de l'indexeur en mode continu :

1. Exécutez `afuIndexer` avec l'argument supplémentaire `-continuousmode`. Le processus de l'indexeur s'exécute avec des itérations sans fin jusqu'à l'émission d'une commande d'arrêt.

Dans chaque itération, le processus de l'indexeur recherche, dans la table des tâches ouvertes, de nouveaux éléments à indexer et les traite. Si la table des tâches ouvertes ne contient pas de nouveaux éléments lors d'une itération, le processus de l'indexeur attend environ une minute avant de démarrer automatiquement une autre itération.

Le comportement de l'indexeur change lorsque le processus de l'indexeur est démarré en mode continu :

- L'option de configuration `IdxDisplayTimeStamp` est automatiquement activée.
- `IdxDisplayProgressIndicator` est également automatiquement activée. Cela signifie que la surveillance de la progression est remplacée par une surveillance globale du statut qui affiche :
 - l'horodatage actuel
 - les statistiques actuelles de l'indexeur
 - la taille actuel du lot : le nombre d'éléments déjà traités et attendant d'être indexés par Net Search Extender. Cela correspond au nombre de lignes contenues dans la table de journal Net Search Extender.
 - la taille minimale du lot
 - la taille maximale du lot
 - le temps écoulé depuis le démarrage du processus de l'indexeur
- Les noms des fichiers journaux générés incluent un marqueur de génération montrant la date et l'heure de génération de chaque fichier journal. Ces fichiers sont limités en taille et en nombre.

- Si un processus de travail s'arrête, il est automatiquement redémarré sans arrêter le processus de l'indexeur. Si le processus de l'indexeur s'exécute en mode continu, cette opération peut se produire jusqu'à 1 000 fois avant que le processus de l'indexeur ne s'arrête. Par contre, si le processus de l'indexeur est exécuté en mode de traitement par lots, le redémarrage du processus de travail est automatiquement limité à 100 fois.
- La taille maximale de lot définie dans l'option de configuration **IdxItemMaxBatchSize** est recalculée après chaque phase d'indexation Net Search Extender de chaque lot. Le recalcul de la taille maximale du lot est régulièrement nécessaire car l'index peut avoir été réorganisé lors de l'une des nombreuses phases d'indexation se produisant lors des exécutions de `afuIndexer` en mode continu.
Si vous indiquez la taille maximale de lot à l'aide de l'argument `-maxbatchsize` de ligne de commande `afuIndexer`, elle n'est pas recalculée. Il est donc recommandé de ne pas utiliser l'argument de ligne de commande `-maxbatchsize` lorsque l'indexeur s'exécute en mode continu.
- La taille minimale de lot définie dans l'option de configuration **IdxItemMinBatchSize** a un autre effet sur le traitement de l'indexeur lorsque `afuIndexer` s'exécute en mode continu :
 - La taille minimale de lot est uniquement appropriée lorsque `afuIndexer` est arrêté à l'aide de l'argument `-shutdown complete`. Net Search Extender s'exécute si la taille de lot actuelle est supérieure ou égale à la taille de lot minimale calculée.
 - La taille minimale de lot est automatiquement recalculée toutes les 24 heures lorsque **IdxItemMinBatchSize** est défini sur 0, et la valeur calculée est dépendante de la charge moyenne de l'élément lors d'une période définie de N jours.
Il est recommandé de ne pas utiliser l'argument de ligne de commande `-minbatchsize` pour définir une valeur, ce qui empêche la taille minimale de lot d'être automatiquement recalculée.
- L'option de configuration **IdxAutoIndexingEveryNDays** permet de démarrer Net Search Extender quelles que soient les spécifications de taille de lot. La valeur par défaut de **IdxAutoIndexingEveryNDays** est 7. Si la charge d'archivage chute brusquement, **IdxAutoIndexingEveryNDays** assure que tout le contenu archivé est indexé, même si les tailles minimale et maximale de lot calculées qui démarrent généralement Net Search Extender ont été atteintes. Si **IdxAutoIndexingEveryNDays** est défini sur 0, Net Search Extender démarre uniquement lorsque la taille maximale de lot est atteinte, ce qui signifie que, si la charge d'archivage est faible, des éléments peuvent ne pas être indexés pour un certain temps.
- Les arguments de ligne de commande `-continuousmode` et `-debug` ne peuvent pas être utilisés simultanément.
- Les argument de ligne de commande `-continuousmode` et `-nseindexreorg` ne peuvent pas être utilisés simultanément.
- Si le processus de l'indexeur est arrêté alors qu'il s'exécute en mode continu, le code 102 n'est pas renvoyé, ce qui signifie qu'il s'est interrompu à la suite d'une demande d'arrêt. En revanche, le code 0 ou 1 est renvoyé afin d'indiquer si des éléments ont été traités ou non.
- Les éléments ayant produit une erreur ou une erreur irrémédiable (par exemple, si le traitement est arrêté de manière inattendue) sont déplacés vers la table des tâches terminées lors de l'itération de traitement suivante après l'exécution de Net Search Extender. Cela correspond à deux exécutions consécutives de l'indexeur en mode de traitement par lots.

2. Arrêtez le processus à l'aide de l'un des arguments suivants :

Option	Description
afuIndexer -shutdown complete	Le traitement dépend de la taille de lot actuelle : <ul style="list-style-type: none"> • Si la taille de lot actuelle est supérieure ou égale à la taille minimale de lot, l'indexeur met fin au traitement de tous les éléments dans le lot en cours et indexe ces éléments à l'aide de Net Search Extender avant l'arrêt du traitement. • Si la taille de lot actuelle est inférieure à la taille minimale de lot, l'indexeur arrête le traitement dès que possible. Net Search Extender n'est pas appelé.
afuIndexer -shutdown force	Quelle que soit la taille de lot actuelle, l'indexeur arrête immédiatement le traitement. Net Search Extender n'est pas appelé.

Traitement des courriers électroniques sans pièces jointes

Si vous utilisez la mise en évidence de la casse dans le contenu des courriers électroniques uniquement et non dans le contenu des pièces jointes, vous pouvez configurer le processus de l'indexeur pour omettre le traitement et l'indexation du contenu des pièces jointes.

Pour traiter des courriers électroniques sans leurs pièces jointes :

Traitez le contenu des courriers électroniques en ignorant le contenu des pièces jointes imbriquées en définissant l'une des options de configuration suivantes :

Option	Description
<ul style="list-style-type: none"> • GenFileTypeFilterIsBlackList = 0 • La section [filetypefilter] du fichier de configuration de l'indexeur est vide. 	<ul style="list-style-type: none"> • L'élément de courrier électronique est indexé mais toutes les pièces jointes imbriquées sont ignorées. Une valeur IDXRC de 0 est affectée à l'élément de la table des tâches terminées. • Si XmlInsertAttachmentProcessingWarnings = 1, la chaîne de notification IcmFceWarning: IcmConfigFilteringFile est indexée avec l'élément. • Si XmlInsertAttachmentProcessingWarnings = 0, aucune chaîne de notification n'est indexée.

Option	Description
TxtCnvEnabled = 0	<ul style="list-style-type: none"> L'élément de courrier électronique est indexé mais toutes les pièces jointes imbriquées sont ignorées. Une valeur IDXRC de 0 est affectée à l'élément de la table des tâches terminées. Si XmlInsertAttachmentProcessingWarnings = 1, la chaîne de notification <code>IcmFceWarning:IcmConversionDisabled</code> est indexée avec l'élément. Si XmlInsertAttachmentProcessingWarnings = 0, aucune chaîne de notification n'est indexée.

Traitement simultané de plusieurs types d'élément

Le processus de l'indexeur de recherche de texte ne peut être exécuté que sur un seul type d'élément à la fois mais plusieurs types d'élément peuvent être traités simultanément.

Cependant, pour traiter plusieurs types d'élément à la fois, chaque instance du processus de l'indexeur doit utiliser un port unique et libre ;

Pour garantir un port unique à chaque instance :

Définissez une valeur de port distincte pour chaque instance `afuIndexer` :

Option	Description
Argument de processus	<p>Exécutez</p> <pre>afuIndexer -port <numéro_port></pre> <p>La valeur transmise via l'argument <code>-port</code> écrase la valeur définie dans la variable d'environnement AFU_INDEXER_RMI_PORT.</p>
Variable d'environnement	Définissez AFU_INDEXER_RMI_PORT pour chaque instance avant de démarrer <code>afuIndexer</code> .

Traitement des pièces jointes imbriquées dans des courriers électroniques et des documents

Si vous souhaitez indexer des documents pour la recherche de texte, vous devez d'abord extraire le contenu textuel des documents.

Les Filtres Oracle Outside In Technology sont utilisés pour extraire des informations textuelles de toutes les pièces jointes imbriquées, du contenu IBM Connections et des documents Système de fichiers et Microsoft SharePoint.

Le terme "document" est utilisé dans cette rubrique pour désigner à la fois les pièces jointes imbriquées dans des messages électroniques et les documents du Système de fichiers et de Microsoft SharePoint.

Certains documents du gestionnaire de ressources IBM Content Manager ne contiennent pas d'informations textuelles qui doivent être indexées pour activer la recherche de texte. Cela signifie que les documents qui ne contiennent pas d'informations textuelles doivent être identifiés et filtrés avant leur indexation par Net Search Extender. Cette opération peut être effectuée de l'une des deux manières suivantes :

- Par extension de nom de fichier :

Une liste d'extensions de nom de fichier peut être configurée pour le traitement des documents. En fonction du paramètre de l'option de configuration **GenFileTypeFilterIsBlackList**, la liste des extensions de nom de fichier est considérée comme la liste des entrées rejetées ou des entrées acceptées.

Si l'option de configuration correspond à 1, la liste des extensions de nom de fichier indiquée à la section [FileTypeFilter] dans la configuration représente la liste des entrées rejetées. Cela signifie que tous les documents du gestionnaire de ressources IBM Content Manager dont les extensions de nom de fichier figurent dans la liste sont ignorés et ne sont pas traités par les filtres d'extraction de texte.

Si l'option correspond à 0, la liste des extensions de nom de fichier figurant dans la section [FileTypeFilter] est considérée comme la liste des entrées acceptées. Cela signifie que seuls les documents du gestionnaire de ressources IBM Content Manager dont l'extension de nom de fichier est indiquée à la section FileTypeFiles sont extraits.

Le jeu de types de fichier utilisé pour le filtrage doit être stocké dans une section distincte du fichier de configuration .ini appelée [FileTypeFilter]. Pour chaque type ou extension de fichier, une option distincte appelée extension = doit être indiquée. Par exemple.

```
[FileTypeFilter]
extension = .JPE
extension = .JPEG
extension = .JPG
extension = .PJP
extension = .PJPEG
...
```

Le modèle de fichier filtre de type de fichier inclus dans le composant de recherche de texte est une liste des entrées rejetées.

- Par type de fichier des Filtres Oracle Outside In Technology :

Si l'option de configuration **TxtCnvOutsideInFilterGroups** est activée, tous les documents identifiés par les Filtres Oracle Outside In Technology membres de l'un des groupes de types de fichier suivants sont filtrés. En d'autres termes, ces documents ne sont pas traités pour l'extraction de texte :

- BMP, TIFF, PCX, GIF, EPSTIFF, CCITTGRP3, MACPICT2, WPG, WINDOWSMETA, LOTUSPIC, MACPICT1, AMIDRAW, TARGA, GEMIMG, OS2DIB, WINDOWSICON, WINDOWSCURSOR, MICROGRAFX, MACPAINT, CORELDRAW2, CORELDRAW3, HPGL, HARVARDDOS3, HARVARDDOS2, HARVARDDOS3PRS, FREELANCE, WPG2, CGM
- MultiMedia
- Graphic2
- Bitmap
- Le groupe Filtres Oracle Outside In Technology *Reserved*
- Le groupe Filtres Oracle Outside In Technology *OEM*
- Le groupe Filtres Oracle Outside In Technology *Other*

Important : Si un document est au format texte, il ne dispose pas d'informations sur son codage. Le traitement correct d'un tel document ne peut être garanti, notamment si le codage du document se base sur un jeu de caractères non latin. LeFiltres Oracle Outside In Technology peut ne pas reconnaître le codage du document. Dans ce cas, l'indexeur Content Collector tente de déterminer la page de code en appliquant plusieurs heuristiques. Toutefois, cette méthode peut ne pas réussir et entraîner une mauvaise indexation, ce qui pourrait empêcher la recherche de contenu du document.

La conversion de document avec Filtres Oracle Outside In Technology est influencée par l'activation ou la désactivation de l'option de configuration **TxtCnvEnable** :

Tableau 10. La conversion de document avec Filtres Oracle Outside In Technology lorsque le format de document est identifié ou qu'aucune conversion n'est demandée.

Option de configuration	Résultat
TxtCnvEnable = 1 ou n'est pas spécifié	<ul style="list-style-type: none"> Le document est converti. L'article est ajouté au tableau des tâches terminées avec IDXRC = 0.
TxtCnvEnable = 0	<ul style="list-style-type: none"> Aucune conversion de document. La chaîne de notification IcmFceWarning: IcmConversionDisabled est ajoutée au contenu textuel indexé pour l'article.

Lorsque **TxtCnvEnable** est activé, l'option de configuration suivante s'active :

- **TxtCnvPlainTextHandlingForUnsupportedFormat**
- **TxtCnvAutoCodepageDetection**
- **TxtCnvAutoCodepageDetectionMinConfidence**
- **TxtCnvMaxExtendedCharactersRatio**
- **TxtCnvDefaultCodepageForPlainText**

Les valeurs de ces options n'ont aucune signification lorsque **TxtCnvEnable** est désactivé.

Le tableau suivant montre ce qui se passe lorsque le format du document ne peut pas être identifié par Filtres Oracle Outside In Technology ou lorsque le document ne peut pas être traité.

Tableau 11. Conversion de document lorsque Filtres Oracle Outside In Technology ne parvient pas à identifier le format du document

Option de configuration	Résultat
<ul style="list-style-type: none"> • TxtCnvPlainTextHandlingForUnsupportedFormat = 0 	<ul style="list-style-type: none"> Aucune conversion de document. L'article est ajouté au tableau des tâches terminées avec IDXRC = 4, (WARN_FORMAT_NOT_SUPPORTED) La chaîne de notification IcmFceError: InsoErrorNoFilterAvailable est ajoutée au contenu textuel indexé pour l'article.

Tableau 11. Conversion de document lorsque *Filtres Oracle Outside In Technology* ne parvient pas à identifier le format du document (suite)

Option de configuration	Résultat
<ul style="list-style-type: none"> • TxtCnvPlainTextHandlingForUnsupportedFormat = 1 • TxtCnvAutoCodepageDetection = 1 • TxtCnvDefaultCodepageForPlainText = 0 	<ul style="list-style-type: none"> • Le document est traité et une tentative de reconnaissance du codage de caractères (page de codes) du document est exécutée. • Si le niveau de confiance avec lequel la page de codes a été détecté est égal ou supérieur à la valeur indiquée dans TxtCnvAutoCodepageDetectionMinConfidence : <ul style="list-style-type: none"> – Le document est converti. – L'article est ajouté au tableau des tâches terminées avec IDXRC = 0. – La chaîne de notification IcmFceWarning: IcmConversionAutoCodePage est ajoutée au contenu textuel indexé pour l'article. • Si le niveau de confiance avec lequel la page de codes a été détecté est inférieur à la valeur indiquée dans TxtCnvAutoCodepageDetectionMinConfidence : <ul style="list-style-type: none"> – L'article est traité comme si TxtCnvAutoCodepageDetection était défini sur 0. <p>Important : Les résultats du test montrent que la valeur par défaut de TxtCnvAutoCodepageDetectionMinConfidence est optimale et vous encouragent à ne pas la modifier. Toutefois, si vous souhaitez vous assurer qu'aucun document n'est converti incorrectement, augmentez la valeur à environ 90.</p>

Tableau 11. Conversion de document lorsque Filtrage Oracle Outside In Technology ne parvient pas à identifier le format du document (suite)

Option de configuration	Résultat
<ul style="list-style-type: none"> • TxtCnvPlainTextHandlingForUnsupportedFormat = 1 • TxtCnvAutoCodepageDetection = 0 • TxtCnvDefaultCodepageForPlainText = 0 	<ul style="list-style-type: none"> • Si la taille du fichier est 0 : <ul style="list-style-type: none"> - Aucune conversion de document. - L'article est ajouté au tableau des tâches terminées avec IDXRC = 6. - La chaîne de notification IcmFceError: IcmEmptyFile est ajoutée au contenu textuel indexé pour l'article. • Si la taille du fichier est supérieure à 0 : <ul style="list-style-type: none"> - Le document est traité et tous les caractères binaires (point de code < 0x05) et caractères étendus (point de code > 0x7f) du document sont comptés. - Si des caractères binaires sont trouvés : <ul style="list-style-type: none"> - Aucune conversion de document. - L'article est ajouté au tableau des tâches terminées avec IDXRC = 5, (WARN_BINARY) - La chaîne de notification IcmFceError: InsoErrorNoFilterAvailable est ajoutée au contenu textuel indexé pour l'article. - Si aucun caractère binaire n'est trouvé : <ul style="list-style-type: none"> - Si des caractères étendus sont trouvés et que le pourcentage est inférieur à la valeur dans TxtCnvMaxExtendedCharactersRatio : <ul style="list-style-type: none"> • Le document est converti. • L'article est ajouté au tableau des tâches terminées avec IDXRC = 0. • La chaîne de notification IcmFceWarning: IcmConversionPlainText est ajoutée au contenu textuel indexé pour l'article. - Si des caractères étendus sont trouvés et que le pourcentage est supérieur à la valeur dans TxtCnvMaxExtendedCharactersRatio : <ul style="list-style-type: none"> • Aucune conversion de document. • L'article est ajouté au tableau des tâches terminées avec DXRC = 2. • La chaîne de notification IcmFceError: IcmConversionFailedTooManyExtendedChars est ajoutée au contenu textuel indexé pour l'article.

Tableau 11. Conversion de document lorsque *Filtres Oracle Outside In Technology* ne parvient pas à identifier le format du document (suite)

Option de configuration	Résultat
<ul style="list-style-type: none"> • TxtCnvPlainTextHandlingForUnsupportedFormat = 1 • TxtCnvAutoCodepageDetection = 0 ou 1 • TxtCnvMaxExtendedCharactersRatio = 100 • TxtCnvDefaultCodepageForPlainText = <CCSID> 	<ul style="list-style-type: none"> • Si TxtCnvAutoCodepageDetection est désactivé ou que la détection de la page de codes échoue : <ul style="list-style-type: none"> – Si le document ne contient pas de caractères binaires : <ul style="list-style-type: none"> - Le document est converti à l'aide de la page de codes indiquée dans TxtCnvDefaultCodepageForPlainText. - L'article est ajouté au tableau des tâches terminées avec IDXRC = 0. - La chaîne de notification IcmFceWarning: IcmConversionDefaultCodePage est ajoutée au contenu textuel indexé pour l'article. – Si la page de codes indiquée dans TxtCnvDefaultCodepageForPlainText n'est pas prise en charge : <ul style="list-style-type: none"> - Aucune conversion n'a lieu. - L'article est ajouté au tableau des tâches terminées avec IDXRC = 3. - La chaîne de notification IcmFceError: IcmConversionFailedCodepageNotSupported est ajoutée au contenu textuel indexé pour l'article. – Si la page de codes indiquée dans TxtCnvDefaultCodepageForPlainText est prise en charge : <ul style="list-style-type: none"> - Le document est converti à l'aide de la page de codes indiquée. - L'article est ajouté au tableau des tâches terminées avec IDXRC = 0. - La chaîne de notification IcmFceWarning: IcmConversionDefaultCodePage est ajoutée au contenu textuel indexé pour l'article. <p>Important :</p> <ul style="list-style-type: none"> • La configuration de TxtCnvDefaultCodepageForPlainText sur une page de codes par défaut est uniquement conseillée si les pièces jointes de texte en clair ont probablement toutes le même codage de caractères. Ce paramètre n'est pas conseillé si vous pensez disposer de documents de texte en clair avec différents codage de caractères. • Si vous définissez TxtCnvDefaultCodepageForPlainText sur une page de codes par défaut, vous devez définir TxtCnvMaxExtendedCharactersRatio sur 100.

Exemple de script pour l'évaluation des codes retour du processus de l'indexeur

IBM Content Collector fournit un exemple de script à utiliser pour évaluer le code retour du processus de l'indexeur.

Les exemples de scripts des différentes plateformes sont disponibles dans le sous-répertoire `samples` dans lequel IBM Content Collector Text Search Support est installé sur le serveur IBM Content Manager.

Le script répertorie les codes d'erreur pouvant être renvoyés par le processus système de l'indexeur. Si vous décidez d'utiliser l'exemple de script, exécutez le script en ajoutant les étapes de l'action à prendre pour chaque erreur ou pour un groupe d'erreurs. Seuls les codes retour 0 et 1 ne nécessitent pas d'action utilisateur.

Exemple de script UNIX structuré de la manière suivante :

```
# This sample script file can be used for embedding the afuIndexer script
# and branch according to the return code ("$?")
#
# *****
AFU_INDEXER_RC_NO_ERROR=0
AFU_INDEXER_RC_NO_ACTION_PERFORMED=1
AFU_INDEXER_RC_CONFIG_ERROR=2
AFU_INDEXER_RC_LOGGING_ERROR=3
AFU_INDEXER_RC_INDEXING_ERROR=4
AFU_INDEXER_RC_NSE_ERROR=5
AFU_INDEXER_RC_SQL_ERROR=6
AFU_INDEXER_RC_IO_ERROR=7
AFU_INDEXER_RC_GENERAL_EXCEPTION=8
AFU_INDEXER_RC_RMI_ERROR=9
AFU_INDEXER_RC_OUT_OF_DISK_SPACE=10
AFU_INDEXER_RC_OUT_OF_MEMORY=11
AFU_INDEXER_RC_USER_COMMAND_ERROR=50
AFU_INDEXER_RC_TOO_MANY_ERRORS=60
AFU_INDEXER_RC_ALREADY_RUNNING=100
AFU_INDEXER_RC_NOT_RUNNING=101
AFU_INDEXER_RC_SHUTDOWN=102
AFU_INDEXER_RC_GENERAL_COMMAND_LINE_ERROR=200

# launch the indexer
afuIndexer $*
AFU_INDEXER_RC=$?

# echo afuIndexer RC: $AFU_INDEXER_RC

# determine action depending on return code
if [[ $AFU_INDEXER_RC >= $AFU_INDEXER_RC_GENERAL_COMMAND_LINE_ERROR ]] ; then
echo IBM afuIndexer was launched with incorrect command line arguments
# ...

elif [[ $AFU_INDEXER_RC == $AFU_INDEXER_RC_NOT_RUNNING ]] ; then
echo IBM afuIndexer terminated with: NOT_RUNNING ($AFU_INDEXER_RC_NOT_RUNNING)
# ...

elif [[ $AFU_INDEXER_RC == $AFU_INDEXER_RC_ALREADY_RUNNING ]] ; then
echo IBM afuIndexer terminated with: ALREADY_RUNNING
($AFU_INDEXER_RC_ALREADY_RUNNING)
# ...

elif [[ $AFU_INDEXER_RC == $AFU_INDEXER_RC_TOO_MANY_ERRORS ]] ; then
echo IBM afuIndexer terminated with: TOO_MANY_ERRORS
($AFU_INDEXER_RC_TOO_MANY_ERRORS)
# ...
...

```

Les codes suivants sont renvoyés par le processus de l'indexeur une fois le processus terminé. Pour plus d'informations sur le code du processus système renvoyé, reportez-vous également aux fichiers journaux.

Tableau 12. Codes retour du processus de l'indexeur de recherche de texte

Code retour	Description
0	Le processus d'indexation a abouti. Aucune erreur n'a été générée.
1	Aucun élément à traiter n'a été détecté. Cette valeur est renvoyée si aucun document n'a été archivé depuis la dernière indexation du type d'élément, si l'option IdxDelayDocumentprocessingInDays correspond à un nombre de jours qui ne se sont pas encore écoulés ou si la taille minimale du lot déclenchant le processus d'indexation et définie par l'option IdxItemsMinBatchSize n'est pas atteinte.
2	Une erreur grave s'est produite en raison d'une erreur de paramétrage ou de configuration de la procédure d'indexation.
3	Une erreur grave s'est produite lors de la consignation.
4	Une erreur grave s'est produite lors du traitement des documents avant l'appel du composant Net Search Extender.
5	Une erreur grave du composant Net Search Extender a été détectée.
6	Une erreur grave liée au traitement de DB2 a été détectée.
7	Une erreur grave liée à l'accès au fichier (chemin ou droits d'accès au fichier) a été détectée.
8	Une erreur de traitement grave générale qui ne correspond pas à l'une des erreurs ci-dessus a été détectée.
9	Une erreur de communication grave a été détectée. Le processus de l'indexeur n'a pas pu configurer les communications le processus.
10	Le processus de l'indexeur a détecté que l'espace du système de fichiers était insuffisant.
11	Le processus de l'indexeur ou le processus de travail a détecté une situation de mémoire saturée.
50	L'une des commandes définies par l'utilisateur et exécutables avant ou après le composant Net Search Extender a renvoyé un code différent de 0.
60	Trop d'erreurs se sont produites. Le nombre d'erreurs de récupération IBM Content Manager Resource Manager suivantes s'étant produites a dépassé la limite indiquée dans l'option de configuration IdxMaxNumberOfSubsequentRmErrors .
100	Le processus de l'indexeur est déjà en cours d'exécution pour le type d'élément indiqué.
101	L'indexeur a été exécuté avec un argument de ligne de commande shutdown mais aucune instance en cours d'exécution n'a été détectée pour le type d'élément indiqué.
102	Le processus de l'indexeur a été arrêté à la suite d'une demande d'arrêt.
200	Le processus de l'indexeur a été appelé en utilisant une syntaxe incorrecte pour la ligne de commande.
201	Le processus de l'indexeur a été appelé avec l'argument -copyright. Aucune autre opération n'a été effectuée.
202	Le processus de l'indexeur a été appelé avec un argument inconnu.
203	Un argument obligatoire n'a pas été indiqué lors de l'appel du processus de l'indexeur.
204	Un argument nécessitant une valeur de type chaîne a été transmis sans valeur au processus de l'indexeur.
205	Un argument nécessitant une valeur de type entier a été transmis sans valeur au processus de l'indexeur.

Tableau 12. Codes retour du processus de l'indexeur de recherche de texte (suite)

Code retour	Description
207	Le processus de l'indexeur a été appelé plusieurs fois avec le même argument.
208	La valeur d'un argument qui doit être numérique n'est pas une valeur numérique.
210	Le processus de l'indexeur a été appelé avec des arguments incompatibles.
211	L'indexeur a été appelé avant des arguments manquants.

Réindexation de documents

Utilisez les arguments de réindexation du processus de l'indexeur lorsque vous exécutez à nouveau l'indexation pour tenter de retraiter des éléments qui avaient été exécutés sur des erreurs spécifiques ou indexés à un moment particulier avant que des modifications ou des adaptations importantes aient été apportées au système.

Une fois vos erreurs regroupées en erreurs d'avertissement et en erreurs plus importantes, vous pouvez choisir entre différentes options de réindexation que vous pouvez communiquer à `afuIndexer` lors de la prochaine exécution planifiée du processus.

A faire : La réindexation d'éléments ayant provoqué des avertissements ou des erreurs est uniquement effective si des configurations d'index ont été modifiées ou corrigées ; par exemple, si des délais d'attente ont été ajustés ou de nouvelles mises à jour installées (es filtres de document supplémentaires ont été ajoutés, des filtres existants ont été améliorés ou une nouvelle version de Net Search Extender a été installée). Les erreurs graves se produisent généralement si le processus de l'indexeur s'arrête de manière inattendue lors du traitement d'un élément en raison de son altération ou de sa structure incorrecte. Ces éléments doivent être analysés et probablement corrigés avant qu'une nouvelle tentative d'indexation soit effectuée. Des erreurs graves peuvent également se produire si le processus est à court de mémoire.

Les options de réindexation comprennent :

- La réindexation des éléments individuels. Seuls les éléments dont les ID sont répertoriés sont traités à nouveau.
- La réindexation des éléments disposant d'une valeur IDXRC spécifique. Pour obtenir une liste de toutes les valeurs stockées dans la colonne IDXRC de la table des tâches terminées, voir «Codes retour de l'index stockés dans la tables des tâches terminées», à la page 37.
- La réindexation de tous les éléments ayant une valeur d'avertissement IDXRC comme groupe. Cette opération sert de raccourci vers la plage d'avertissement IDXRC.
- La réindexation de tous les éléments ayant une valeur d'erreur IDXRC comme groupe. Cette opération sert de raccourci vers la plage d'erreur IDXRC.
- La réindexation de tous les éléments ayant une valeur IDXRC de 200 comme groupe. Cette opération sert de raccourci vers toutes les erreurs graves. Reportez-vous à «Opérations à effectuer en cas d'arrêt brutal du processus de l'indexeur», à la page 105.
- La réindexation de tous les éléments ayant été indexés avec une chaîne d'erreur ou d'avertissement de recherche.

- La réindexation de tous les éléments ayant été indexés avec une chaîne arbitraire.

La réindexation nécessite toujours la recherche des éléments qui remplissent les conditions de réindexation demandées dans l'ensemble de la table de base de données des tâches terminées. Elle peut prendre un certain temps en fonction du nombre de documents déjà indexés. Une analyse complète de la table est nécessaire sauf si des index de base de données supplémentaires sont créés manuellement.

Si vous indiquez l'un des arguments de réindexation, les éléments correspondant aux critères de réindexation sont ajoutés à la table des tâches ouvertes, prêts pour le traitement. Ils ne sont pas supprimés de la table des tâches terminées. Ces entrées de la table des tâches terminées sont écrasées une fois les éléments traités.

Il n'existe cependant aucune garantie que ces éléments soient traités pendant la même procédure de l'indexeur que celle qui les avait identifiés pour réindexation. Ils peuvent être traités lors d'une exécution ultérieure de l'indexeur. Si les critères de réindexation ne correspondent à aucun élément de la table des tâches terminées, un avertissement est émis pour indiquer qu'aucun élément correspondant aux critères de réindexation n'est trouvé mais que l'indexation continue.

Le nouveau traitement d'éléments comprend :

- La réindexation d'éléments avec des avertissements et des erreurs :

Option	Description
Éléments individuels	Exécutez <code>afuIndexer -reindexitems <ID_élément>;<ID_élément></code>
Tous les éléments avec avertissements	Exécutez <code>afuIndexer -reindexwarnings</code>
Tous les éléments avec erreurs	Exécutez <code>afuIndexer -reindexerrors</code>
Tous les éléments avec erreurs graves	Exécutez <code>afuIndexer -reindexsevereerrors</code>
Tous les éléments avec une valeur IDXRC spécifique	Exécutez <code>afuIndexer -reindexitems <IDXRC></code>
Tous les éléments qui ont été indexés avec une chaîne de notification d'avertissement qui commence par <code>IcmFceWarning</code>:	Exécutez <code>afuIndexer -reindexsearchwarnings</code> Pour plus d'informations sur les chaînes de notification d'avertissement, voir «Recherche de chaînes d'erreur», à la page 114.
Tous les éléments qui ont été indexés avec une chaîne de notification d'erreur qui commence par <code>IcmFceError</code>:	Exécutez <code>afuIndexer -reindexsearcherrors</code> Pour plus d'informations sur les chaînes de notification d'erreur indexées, voir «Recherche de chaînes d'erreur», à la page 114.
Tous les éléments qui ont été indexés avec une chaîne de recherche arbitraire	Exécutez <code>afuIndexer -reindexsearch terme_recherché</code>

- La réindexation d'éléments indexés lors d'une période spécifique après laquelle des modifications ont été apportées à l'environnement de traitement qui influencent le traitement du document et le résultat de l'indexation. Exécutez `afuIndexer -reindexdate <aaaammjj>`

où `<aaaammjj>` correspond à la date d'archivage de l'élément.

Tous les éléments dont la date d'archivage est <aaaammjj> ou ayant une date ultérieure sont réindexés.

Référence associée :

«Codes retour de l'index stockés dans la tables des tâches terminées», à la page 37

«Codes retour de l'index stockés dans la tables des tâches terminées», à la page 37

Opérations à effectuer en cas d'arrêt brutal du processus de l'indexeur

Si le processus de l'indexeur s'arrête de manière inattendue avant le traitement de tous les éléments du lot, ou si les touches CTRL-BREAK sont activées lors du traitement, le processus de l'indexeur doit être redémarré manuellement.

Pour s'assurer une performance 24H/24H, le processus de l'indexeur utilise un processus de travail distinct qui traite tous les documents, même si certains d'entre eux sont dans un format incorrect et, en conséquence, détruit la structure de mémoire du processus ou provoque l'arrêt inattendu du processus. Ce processus de travail utilise Filtres Oracle Outside In Technology pour convertir des documents en leur contenu textuel. L'utilisation d'un processus de travail distinct est explicitement recommandée par Filtres Oracle Outside In Technology car la qualité des documents traités est inconnue et les documents de format incorrect peuvent provoquer un comportement inattendu du processus.

Remarque : Si le processus de travail de l'indexeur s'arrête de manière inattendue sous UNIX ou Linux, le système d'exploitation crée généralement un fichier central. Par défaut, le processus de l'indexeur supprime automatiquement le fichier central. Pour supprimer la création d'un fichier central, émettez la commande `ulimit -c 0` dans le shell dans lequel le processus de l'indexeur est démarré. Pour plus d'informations, consultez le manuel du système d'exploitation.

Pour empêcher plusieurs instances du même type d'élément, l'indexeur crée un fichier pour chaque type d'élément dans le système de fichiers dès que le processus de l'indexeur est démarré. En règle générale, ce fichier de contrôle est supprimé dès que le traitement est terminé. Toutefois, si le processus de l'indexeur s'arrête de manière inattendue ou est interrompu, il se peut que le fichier ne soit pas supprimé. Si l'indexeur est à nouveau démarré pour le même type d'élément, le processus de l'indexeur affiche un message indiquant qu'une instance est déjà en cours d'exécution. Le message inclut le nom du fichier de contrôle et le code retour du processus, 100.

La variable d'environnement `AFU_INDEXER_PROMPT_RUN` a une influence sur ce comportement. Si elle est définie sur 1, le processus de l'indexeur ne s'arrête pas avec un code retour 100 lorsqu'il détecte qu'un processus de l'indexeur est déjà en cours d'exécution pour le même type d'élément mais il demande à l'utilisateur s'il doit continuer et ignorer le code retour. Naturellement, l'invite envoyée à l'utilisateur place le processus en attente.

Si le processus de l'indexeur s'arrête de manière inattendue ou si les touches CTRL-BREAK sont activées lors du traitement :

1. Supprimez le fichier de contrôle de l'instance apparaissant dans le message d'erreur affiché par le processus de l'indexeur.
2. Exécutez de nouveau `afuIndexer` avec les mêmes arguments qu'avant l'interruption du processus.

Éléments ayant généré une erreur grave

Le traitement et l'indexation des éléments aboutissent dans la plupart des cas. Toutefois, il peut arriver qu'un élément contenant une pièce jointe incorrecte ou endommagée ne puisse pas être traité par les Filtres Oracle Outside In Technology et que cette pièce jointe entraîne l'échec du processus de traitement. Cette situation est considérée comme une erreur grave.

Pour pallier ce problème, le processus de l'indexeur lance un nouveau processus de traitement afin de poursuivre le traitement des documents restants dans le lot d'indexation. La prochaine fois que le processus de l'indexeur est exécuté, il identifie les documents à l'origine de l'échec et émet le message suivant :

```
afuIndexer: Warning:
Un élément de la table des tâches ouvertes a un statut inattendu de
'1' ('IN_PROGRESS').
Cet élément a entraîné l'arrêt du processus de l'indexeur de manière inattendue
et il ne sera pas traité à nouveau.
Au lieu de cela, il sera déplacé dans la table des tâches terminées avec un
code IDXRC de 200 (erreur grave) :
A1001001xxxxxxxxxxxxxxxxxxxx-V1
```

Si l'option de configuration **IdxProcessSevereErrors** est définie sur 0, les éléments qui étaient en cours de traitement au moment de l'arrêt inopiné du processus et qui se trouvent toujours dans un état IN_PROGRESS (en cours) sont déplacés immédiatement de la table des tâches ouvertes vers la table des tâches terminées avec une valeur IDXRC de 200. Cette valeur indique qu'une erreur grave s'est produite pendant le traitement de l'élément. Ces éléments ne sont pas traités de nouveau.

La valeur par défaut du paramètre **IdxProcessSevereErrors** est 0. Pour tenter de retraiter les éléments pour lesquels une erreur grave s'est produite, voir «Retraitement des éléments ayant généré une erreur grave en mode spécial».

Si l'option de configuration **IdxProcessSevereErrors** est définie sur 1, les éléments dont l'état est IN_PROGRESS sont traités de nouveau sans leurs pièces jointes. L'élément peut probablement être retraité correctement car les pièces jointes qui étaient à l'origine de l'erreur grave ne sont plus traitées.

Pour identifier ultérieurement les éléments qui ont été indexés sans leurs pièces jointes, la chaîne de recherche `IcmFceError: IcmConversionSuppressed` est indexée à l'élément et ce dernier est ajouté à la table des tâches terminées avec une valeur IDXRC de 9.

Vous pouvez essayer de réindexer ces éléments avec leurs pièces jointes à l'aide des fonctions correspondantes de l'indexeur `afuIndexer`, par exemple, en utilisant l'option de ligne de commande `-reindexrc 9` après qu'une version plus récente ou corrigée des Filtres Oracle Outside In Technology ait été installée.

Retraitement des éléments ayant généré une erreur grave en mode spécial

La raison la plus probable de l'échec du traitement des éléments et de l'arrêt inattendu du processus de traitement est que les pièces jointes n'ont pas pu être traitées par les Filtres Oracle Outside In Technology. Le manque de mémoire du processus de traitement peut également expliquer le problème. L'élément ou ses pièces jointes peuvent être très volumineux ou un trop grand nombre de tâches (unités d'exécution) s'exécutent en parallèle.

Vous pouvez exécuter `afuIndexer` en mode spécial dans lequel les éléments sont traités à nouveau à l'aide des paramètres de configuration optimisés pour la gestion des situations d'erreur, mais non optimisés pour les performances ou un débit élevé.

Vous pouvez activer ce mode spécial en exécutant `afuIndexer` avec `-severeerrorsmode`. Lorsque l'indexeur `afuIndexer` s'exécute avec `-severeerrorsmode`, il retraite uniquement les éléments qui ont été traités dans une exécution d'indexation précédente et qui sont associés à une valeur `IDXRC` de 200 (erreur grave) ou 9. Une valeur 9 indique que les éléments ont été traités sans leurs pièces jointes, car une erreur grave s'est produite et l'option de configuration **`IdxProcessSevereErrors`** a été activée.

Lorsque `-severeerrorsmode` est sélectionné, l'archivage vers le type d'élément peut poursuivre. Il est inutile d'arrêter cet archivage, car le processus d'indexation fait appel à une autre table temporaire des tâches ouvertes qui est automatiquement alimentée avec les éléments ayant généré une erreur grave.

Restriction : Vous ne pouvez pas exécuter `afuIndexer` simultanément dans ce mode spécial et dans le mode de traitement classique (en lots ou en mode continu) sur le même type d'élément.

Lorsque l'argument de ligne de commande `-severeerrorsmode` est spécifié, l'indexeur `afuIndexer` fonctionne en mode spécial :

- Les paramètres de configuration sont optimisés pour la prise en charge du traitement d'un élément. La vitesse importe peu dans ce mode de traitement :
 - Un seul élément peut être traité à la fois. Cela correspond à l'utilisation de l'option de configuration **`IdxNumberOfParallelIndexerTasks`** définie sur 1.
 - La fonction de vidage est activée.
 - Les paramètres horaires sont adaptés.

A faire : Lorsque le traitement est réalisé en mode d'unité d'exécution unique et que les valeurs horaires internes sont particulièrement bien adaptées à la gestion des éléments avec erreurs graves, le traitement de ces éléments peut prendre beaucoup plus de temps.

- Le processus de l'indexeur fait appel à une autre table temporaire des tâches ouvertes. En conséquence, ce mode de traitement spécial n'a aucune incidence sur l'archivage dans le type d'élément qui utilise la table des tâches ouvertes et peut donc se poursuivre en parallèle.

Tous les éléments ayant provoqué une erreur grave lors de précédentes exécutions d'indexation sont à nouveau traités. La façon dont les éléments sont traités lors de l'indexation en mode spécial dépend également de l'option de configuration **`IdxProcessSevereErrors`** qui détermine la valeur `IDXRC` affectée à chacun de ces éléments :

- Si le traitement de l'élément n'aboutit pas dans ce mode d'indexation spécial et que le paramètre **`IdxProcessSevereErrors`** est défini sur 0, l'élément est déplacé vers la table des tâches terminées avec une valeur `IDXRC` égale à 200. Cet élément ne peut pas être traité par l'indexeur et doit être fourni au service de support logiciel IBM pour une analyse plus approfondie.
- Si le traitement de l'élément n'aboutit pas dans ce mode d'indexation spécial et que le paramètre **`IdxProcessSevereErrors`** est défini sur 1, l'élément est de nouveau traité sans ses pièces jointes.

Si le traitement aboutit cette fois, l'élément est ajouté à la table des tâches terminées avec une valeur IDXRC de 9 et la chaîne de recherche IcmFceError: IcmConversionSuppressed est indexée à l'élément.

Si le traitement échoue encore, l'élément est ajouté à la table des tâches terminées avec une valeur IDXRC de 200. L'élément ne peut pas être traité par l'indexeur et il doit être fourni au service de support logiciel IBM pour une analyse plus approfondie.

Cela signifie que `-severeerrorsmode` retraits les éléments dont la valeur IDXRC est égale à 200 et ceux dont la valeur IDXRC est égale à 9 à partir de la table des tâches terminées.

- Si un élément provoque un nouvel échec des processus de traitement, un nouveau processus de traitement démarre automatiquement.
- L'indexeur s'exécute en mode spécial jusqu'à ce que tous les éléments ayant généré des erreurs graves au cours de processus d'indexation précédents soient traités de nouveau. Autrement dit, le processus s'arrête automatiquement lorsque la table temporaire des tâches ouvertes alimentée avec les éléments qui ont généré des erreurs graves est de nouveau vide. L'indexeur peut être arrêté prématurément, comme dans les autres modes d'indexation, en exécutant la commande `-shutdown`.
- Lorsque l'indexeur s'exécute en mode spécial, il se peut que le nombre d'éléments analysés (ITEMS_SCANNED), signalé dans les statistiques créées pendant l'exécution de l'indexeur, ne soit pas exact.

Important : Ne confondez pas l'argument d'indexation `-severeerrorsmode` avec l'argument `-reindexsevereerrors` :

- Lorsque vous spécifiez le mode `-severeerrorsmode`, l'indexeur retraits uniquement les éléments qui ont généré une erreur grave et qui ont été ajoutés à la table des tâches terminées avec une valeur IDXRC de 200 ou de 9. Ces éléments sont retraités dans un mode spécial qui fait appel à une table des tâches ouvertes différente et qui traite un seul élément à la fois.
- Lorsque vous spécifiez le paramètre `-reindexsevereerrors`, l'indexeur réindexe les éléments dont la valeur IDXRC est égale à 200 en ajoutant les éléments à la table des tâches ouvertes et en les retraitant au cours de l'exécution de l'indexeur ou au cours d'une exécution ultérieure.

Débogage du traitement d'un élément

Lorsque le processus de l'indexeur est exécuté en mode débogage, l'indexeur passe dans un mode de traitement qui prend en charge un examen approfondi des documents traités.

Comme l'option de débogage requiert l'intervention de l'utilisateur, `afuIndexer` ne peut pas être exécuté en mode de traitement par lots. Le mode de traitement avec débogage a une incidence sur le processus de l'indexeur :

- L'indexeur s'exécute avec un processus unique. Aucun processus de travail distinct n'est démarré.
- Le travail de l'indexeur s'exécute en mode de traitement unique et non en mode parallèle. Cela signifie que les éléments sont traités un par un.
- Les options de configuration suivantes sont automatiquement modifiées :
 - L'option de vidage est activée.
 - La consignation est activée. Les noms du fichier journal n'incluent pas d'horodatage, ce qui signifie que toutes les données du journal sont

consignées dans un même fichier distinct des autres journaux et sont remplacées à la prochaine session de débogage.

- Le délai de traitement de chaque élément n'est pas enregistré car le temps de traitement est désactivé.
- Les fichiers XML temporaires ne sont pas supprimés après l'indexation. Vous pouvez ensuite supprimer ces fichiers manuellement dans le répertoire indiqué par l'option de configuration **GenXmlDirectory**.
- Les délais d'expiration sont augmentés et les valeurs horaires internes sont adaptées.
- Aucune commande de sauvegarde ou toute autre commande configurée n'est exécutée.
- Si la variable d'environnement AFU_INDEXER_XML_VIEWER est définie, l'utilisateur doit, pour chaque élément, indiquer si les fichiers XML temporaires doivent être affichés. Pour plus d'informations, reportez-vous à la description de la variable d'environnement de l'indexeur AFU_INDEXER_XML_VIEWER.
- Des commentaires supplémentaires sont ajoutés au contenu des fichiers XML qui comprennent la valeur IDXRC et des noms d'attribut. Ces commentaires peuvent vous aider à comprendre de quelle manière l'indexeur a traité le document d'origine. Ils sont ignorés par Net Search Extender et ne sont pas ajoutés à l'index documentaire complet.
- L'utilisateur doit, pour chaque élément, confirmer qu'il doit être ajouté à l'index documentaire Net Search Extender, c'est-à-dire qu'il doit être indexé.
- Le mode débogage passe aux effets des arguments de la ligne de commande **-verbose**.
- Les exceptions et les erreurs sont affichées dans l'ensemble du contexte. Une trace de pile peut être incluse.
- Comme les éléments indiqués sont automatiquement réindexés lorsque l'outil afuIndexer est exécuté en mode débogage, aucun argument de réindexation ne peut être utilisé.
- Les arguments de ligne de commande -debug et -nseindexreorg ne peuvent pas être utilisés simultanément.
- Les arguments de ligne de commande -debug et -continuousmode ne peuvent pas être utilisés simultanément.

Pour déboguer un ou plusieurs éléments, vous devez tout d'abord définir la variables d'environnement AFU_INDEXER_XML_VIEWER puis exécutez l'indexeur avec l'argument de ligne de commande de débogage pour chaque élément :

```
afuIndexer -debug itemid
```

ou exécutez l'indexeur avec l'argument de ligne de commande de débogage pour un ensemble d'éléments :

```
afuIndexer -debug itemid1;itemid2;itemid3; ...
```

Référence associée :

«Variables d'environnement de l'indexeur de texte», à la page 63

«Variables d'environnement de l'indexeur de texte», à la page 63

Extraction des éléments du gestionnaire de ressources IBM Content Manager (fonction de cliché)

Pour déboguer des éléments qui ne peuvent pas être traités correctement par le processus de l'indexeur, les documents originaux et leurs pièces jointes peuvent être nécessaires pour la détermination de la cause, tout particulièrement si le service de support logiciel IBM est demandé.

Les documents originaux sont stockés dans le gestionnaire de ressources IBM Content Manager.

Pour accéder à ces documents :

Extrayez les documents d'origine du gestionnaire de ressources IBM Content Manager vers un disque local :

Option	Description
Définition de l'option de configuration GenDump sur la valeur True	Les éléments archivés et leurs pièces jointes sont exportés lorsque afuIndexer est exécuté. L'argument de ligne de commande -debug n'est pas obligatoire.
Utilisation de l'argument -debug	<p>1. Exécutez</p> <pre>afuIndexer -debug <ID_élément>;<ID_élément></pre> <p>où <ID_élément> correspond à l'élément ne pouvant être traité correctement.</p> <p>Si vous exécutez afuIndexer avec l'argument -debug, l'option de configuration GenDump est automatiquement activée.</p>
Utilisation de l'argument -reindexitems	<p>1. Activez l'option de configuration GenDump ou utilisez l'argument de l'indexeur -usesettings "GenDump=1".</p> <p>2. Exécutez</p> <pre>afuIndexer -reindexitems <ID_élément>;<ID_élément></pre> <p>où <ID_élément> correspond à l'élément ne pouvant être traité correctement.</p> <p>3. Désactivez l'option de configuration GenDump.</p>

Remarque :

- Les documents originaux sont placés dans le répertoire indiqué par l'option de configuration **GenDumpDirectory**. Le répertoire par défaut est \$(TMP)/afu/dump.
- Les fichiers de vidage utilisent la convention de désignation suivante pour le modèle de données Bundled utilisé par IBM Content Collector et IBM CommonStore :

afu#A1001001A09I22B25936C30988-V1.bin

où

A1001001A09I22B25936C30988-V1

Est l'ID d'élément archivé

- Les fichiers de vidage utilisent la convention de désignation suivante pour le modèle de données Compound utilisé par IBM Content Collector :

afu#A1001001A09I22B25936C30988-V1#
A1001001A09I22B25937J28960-V1#
Report_2009.pdf

où

A1001001A09I22B25936C30988-V1

Est l'ID d'élément archivé

A1001001A09I22B25937J28960-V1

Est l'ID de la pièce jointe intégrée

Report_2009.pdf

Est le nom de fichier de la pièce jointe

- Les fichiers de vidage utilisent la convention de désignation suivante pour les modèles de données de document Multi_doc et Multi_part utilisés par IBM CommonStore :

afu#A1001001A09I22B25936C30988-V1#

1@300#

A1001001A09I22B25937J28960-V1#

Report_2009.pdf

où

A1001001A09I22B25936C30988-V1

Est l'ID d'élément archivé

1 Est l'ID de ressource IBM Content Manager de la partie

300 Est l'ID de la table des parties de composants dans laquelle la partie a été stockée

A1001001A09I22B25937J28960-V1

Est l'ID de la partie

Report_2009.pdf

Est le nom de fichier de la partie du document

Arrêt prématuré, ou à un moment spécifique, d'un processus de l'indexeur IBM Content Collector

Si vous devez arrêter prématurément un processus en cours pour un type d'élément, par exemple à un moment spécifique ou après un intervalle de temps spécifique, vous pouvez émettre une demande d'arrêt sur un processus de l'indexeur en cours ou sur un processus de l'indexeur s'exécutant en mode continu.

Pour arrêter le processus, démarrez une seconde instance du processus afuIndexer pour le même type d'élément et incluez l'argument de ligne de commande d'arrêt. Il existe deux modes d'arrêts différents :

- Complet
- Forcé

Pour un arrêt complet, démarrez un nouveau processus afuIndexer à l'aide de l'argument de ligne de commande suivant :

```
-shutdown complete
```

Cette instance du processus afuIndexer prend en charge les tâches suivantes :

- Arrêt de la lecture d'éléments supplémentaires à partir de la table des tâches ouvertes
- Fin du traitement de tous les éléments pour toutes les tâches actives
- Exécution de Net Search Extender pour mettre à jour l'index documentaire avec les éléments traités jusqu'à présent si la taille actuelle du lot dépasse la taille de lot minimale.

Cela signifie que l'outil de l'indexeur termine le traitement du type d'élément comme si la tâche en cours de traitement avait atteint la limite de taille de lot maximale.

Pour un arrêt forcé, démarrez un nouveau processus `afuIndexer` à l'aide de l'argument de ligne de commande suivant :

```
-shutdown force
```

Cette instance du processus `afuIndexer` prend en charge les tâches suivantes :

- Arrêt de la lecture d'éléments supplémentaires à partir de la table des tâches ouvertes
- Arrêt du traitement des éléments pour toutes les tâches actives
- Fin sans l'exécution de Net Search Extender pour mettre à jour l'index documentaire.

Cela signifie que tous les travaux exécutés pour le type d'élément lors de l'exécution de cet indexeur sont perdus et seront réexécutés à la prochaine exécution de l'indexeur.

Avertissement : Si vous prévoyez d'arrêter un processus d'indexeur pour un type d'élément en cours d'exécution à un moment spécifique et que vous émettez une demande `-shutdown complete` (par exemple, via un travail cron), rappelez-vous qu'il peut s'écouler beaucoup de temps entre le moment où l'indexation s'arrête et celui où la demande d'arrêt est reconnue par le processus de l'indexeur en cours d'exécution :

- L'exécution de toutes les tâches actives peut prendre du temps (jusqu'à quelques minutes) selon les performances du système, le nombre de tâches actives avec des documents en cours et le paramétrage de l'option de configuration **`TxtCnvTimeOutInSecs`**.
- L'appel Net Search Extender suivant pour mettre à jour l'index documentaire prend également du temps. En fonction du nombre d'éléments à traiter, cet appel peut prendre un temps considérable, parfois plusieurs heures.
- La commande d'arrêt n'a pas d'impact sur l'exécution du processus de mise à jour de Net Search Extender. Cela signifie que si une commande d'arrêt est émise après le démarrage de Net Search Extender, Net Search Extender poursuit le traitement du lot. La demande d'arrêt n'est exécutée qu'après la fin du traitement du lot par Net Search Extender.

Si le processus `afuIndexer` est arrêté alors qu'il s'exécutait en mode de traitement par lots, le code 102 est renvoyé, ce qui signifie que le processus de l'indexeur s'est interrompu à la suite d'une demande d'arrêt. Si le processus `afuIndexer` est arrêté alors qu'il s'exécutait en mode continu, le code d'erreur 0 s'affiche pour indiquer que le traitement a abouti, ou un autre code d'erreur s'affiche pour indiquer qu'aucun élément n'a été traité ou qu'une erreur s'est produite.

Réorganisation d'index

La multiplication des mises à jour apportées à un index Net Search Extender et l'accumulation des données d'index entraînent une fragmentation des structures d'index internes et requièrent davantage d'opérations d'entrées-sorties pour les structures de données non contiguës. Les performances de la recherche d'index sont donc altérées.

Pour améliorer les performances de la recherche d'index, vous devez réorganiser l'index régulièrement. Le processus `afuIndexer` tente de déterminer quand

l'organisation d'index doit être lancée automatiquement en utilisant des paramètres de configuration qui permettent de suivre le nombre de documents indexés par le moteur d'indexation Net Search Extender pour chaque type d'élément.

L'existence et la date de lancement de la procédure de réorganisation d'index sont déterminés par les valeurs des options de configuration suivantes :

- **IdxReorgBatchSize:**

Si le nombre de documents indexés par Net Search Extender lors des exécutions précédentes de l'indexeur dépasse le nombre d'éléments définis dans l'option de configuration **IdxReorgBatchSize**, une tentative de réorganisation de l'index est effectuée.

La valeur par défaut de l'option **IdxReorgBatchSize** est 1,5 million d'éléments.

- **IdxMaxNumberOfUpdatesWithoutReorg :**

La valeur de l'option **IdxMaxNumberOfUpdatesWithoutReorg** permet d'éviter que le nombre d'éléments ajoutés à l'index n'augmente indéfiniment sans qu'une réorganisation d'index ne soit effectuée. Cela peut se produire si l'indexeur doit s'exécuter pendant la période de blocage définie par les options

IdxReorgTimeBlockIntervalBegin et **IdxReorgTimeBlockIntervalEnd**.

Dès que le nombre d'éléments ajoutés à l'index atteint la valeur indiquée dans l'option **IdxMaxNumberOfUpdatesWithoutReorg** et qu'aucune opération d'indexation n'a été exécutée jusqu'à présent, l'indexeur refuse d'indexer d'autres éléments et arrête le traitement en générant une erreur. L'opération de réorganisation d'index doit être lancée manuellement en exécutant la commande `afuIndexer` avec l'argument de ligne de commande `-nseindexreorg`.

- **GenNseIndexDirectoryMinFreeSizeRequiredInMB :**

Avant une tentative de réorganisation, le processus de l'indexeur vérifie s'il y a suffisamment d'espace disque. La quantité minimale qui doit être disponible dans les répertoires Net Search Extender est déterminée par la valeur de **GenNseIndexDirectoryMinFreeSizeRequiredInMB**.

S'il y a suffisamment d'espace disque dans ces répertoires d'index, une tentative de réorganisation d'index est effectuée. S'il n'y a pas suffisamment d'espace disque, l'index ne peut pas être réorganisé. Un message d'erreur est affiché et consigné.

- **IdxReorgTimeBlockIntervalBegin** et **IdxReorgTimeBlockIntervalEnd :**

Ces options de configuration peuvent être utilisées pour empêcher une réorganisation d'index pendant une période donnée. Comme l'opération de réorganisation d'index prend un certain temps et bloque la recherche dans l'index pendant une période donnée, il peut être utile d'empêcher la réorganisation de l'index, par exemple pendant les principales heures d'activités. En définissant ces options, vous pouvez indiquer une période pendant laquelle aucune réorganisation d'index ne peut être exécutée.

Si ces options sont indiquées, l'opération de réorganisation est effectuée uniquement si le délai de traitement en cours n'est pas compris dans l'intervalle indiqué par les deux options. Si le délai de traitement est compris dans cet intervalle, un avertissement est affiché et consigné et l'opération de réorganisation est repoussée jusqu'à la prochaine exécution de l'indexeur.

Si ces options ne sont pas indiquées, l'opération de réorganisation n'est pas bloquée pendant une période donnée.

Si l'argument de ligne de commande `-nseindexreorg` de l'outil `afuIndexer` est indiqué, une réorganisation d'index est effectuée lors de l'opération de mise à jour de l'index, quelles que soient la valeur de **IdxReorgBatchSize** ou les options

définissant une période de blocage. Seule la valeur de l'option **GenNseIndexDirectoryMinFreeSizeRequiredInMB** est vérifiée et peut empêcher la réorganisation d'index.

Remarque : Si une réorganisation a été effectuée lors de la dernière opération de mise à jour de l'index, la taille maximale du lot est automatiquement ajustée (augmentée).

Recherche de chaînes d'erreur

Outre la valeur stockée dans la colonne IDXRC de la table des tâches terminées qui indique si l'indexation a abouti ou a eu pour résultat une erreur ou un avertissement, une chaîne de notification est indexée avec l'élément si un avertissement ou une erreur a été détecté.

La chaîne de notification générée à la suite d'un avertissement commence par **IcmFceWarning:** suivi d'un message d'avertissement, par exemple **IcmFceWarning: IcmConfigFilteringFile**. Une chaîne de notification d'avertissement est indexée avec un élément si l'option de configuration **XmlInsertAttachmentProcessingWarnings** est définie sur 1, qui est le paramètre recommandé.

La chaîne de notification générée à la suite d'une erreur de document commence par **IcmFceError:** suivi d'un message d'erreur, par exemple **IcmFceError: IcmConversionFailedTooManyExtendedChars**. Les chaînes de notification d'erreur sont toujours indexées avec un élément.

Ces chaînes d'avertissement et d'erreur étaient également générées par l'exit utilisateur de recherche de texte dans les versions antérieures de Content Collector et de CommonStore et étaient stockées avec l'élément dans l'index.

Vous pouvez utiliser ces chaînes de message dans une requête de recherche pour identifier les documents indexés associés à ces chaînes et les réindexer, par exemple, si de nouveaux filtres de document ont été installés pour résoudre des erreurs de conversion antérieures.

Les chaînes de message d'avertissement suivantes sont indexées avec un élément et peuvent faire l'objet d'une recherche :

Tableau 13. Chaînes d'avertissement de l'indexeur

Chaîne de notification	Code retour	Description
IcmFceWarning: IcmConfigFilteringFile	IDXRC = 0	Le document, ou l'une des pièces jointes imbriquées de l'élément, a été filtré en utilisant la spécification de filtrage des types de fichier ou l'option de configuration TxtCnv0OutsideInFilterGroups a été activée.
IcmFceWarning: IcmConversionPlainText	IDXRC = 0	La détection automatique de la page de codes ou l'analyse des caractères étendus a indiqué que le document, ou l'une des pièces jointes imbriquées de l'élément, est un document de texte normal.
IcmFceWarning: IcmConversionAutoCodePage	IDXRC = 0	Le document a été soumis à la détection automatique de pages de codes.
IcmFceWarning: IcmConversionFailedNoText	IDXRC = 0	L'une des pièces jointes, ou un document, ne contient pas de texte. Il s'agit, par exemple, d'un fichier graphique.
IcmFceWarning: IcmConversionDefaultCodePage	IDXRC = 0	Le codage de document était supposé être dans la page de codes par défaut et converti.

Tableau 13. Chaînes d'avertissement de l'indexeur (suite)

Chaîne de notification	Code retour	Description
IcmFceWarning: IcmConversionFailedProtected	IDXRC = 0	Le document, ou l'une des pièces jointes imbriquées de l'élément, n'a pas pu être géré par les Filtres Oracle Outside In Technology car il était protégé ou chiffré.
IcmFceWarning: IcmUnhandledOLEObject	IDXRC = 0	L'élément a été indexé mais il contenait des objets OLE imbriqués qui ne pouvaient pas être traités.
IcmFceWarning: IcmUnhandledEmbeddedMsg	IDXRC = 0	Cette chaîne de notification n'est plus utilisée comme contenu textuel des fichiers MSG imbriqués, et même les fichiers MSG imbriqués de manière récursive peuvent désormais être indexés. Cela signifie que la chaîne de notification IcmFceWarning: IcmUnhandledEmbeddedMsg ne peut pas être recherchée. Action utilisateur : Recherchez tous les éléments avec des fichiers MSG imbriqués qui ont été indexés avec la chaîne de notification IcmFceWarning: IcmUnhandledEmbeddedMsg, puis réindexez-les.
IcmFceWarning: IcmConversionDisabled	IDXRC = 0	Le document, ou l'une des pièces jointes imbriquées de l'élément, n'est pas traité car l'option de configuration TxtCnvEnabled est désactivé.
IcmFceWarning: IcmDocIsEncrypted	IDXRC = 0	Les courriers électroniques Exchange, Domino ou SMTP/MIME n'ont pas pu être traités ou indexés car le contenu du corps du courrier est chiffré.
IcmFceWarning: IcmDocIsPrivate	IDXRC = 0	Les courriers électroniques Exchange, Domino ou SMTP/MIME n'ont pas pu être traités ou indexés car le courrier est privé.
V3.0.0.2: IcmFceWarning: IcmEmptyBody	IDXRC = 0	Le document e-mail contient un corps vide. Un corps vide est un corps contenant des données textuelles qui ne peuvent pas être indexées, par exemple un corps sans caractères ou contenant seulement des espaces (caractères avec points de code 0x01 à 0x20).

Les chaînes de message d'erreur suivantes sont indexées avec un élément et peuvent faire l'objet d'une recherche :

Tableau 14. Chaînes d'erreur de l'indexeur

Chaîne de notification	Code retour	Description
IcmFceError: InsoErrorNotEnoughMemory	IDXRC = 1	Le document, ou l'une des pièces jointes imbriquées de cet élément, n'a pas pu être intégralement traité en raison des limitations de la mémoire et il n'a été que partiellement indexé.
IcmFceError: IcmConversionFailedTooManyExtendedChars	IDXRC = 2	Le document, ou l'une des pièces jointes imbriquées de l'élément, n'a pas pu être géré par les Filtres Oracle Outside In Technology. L'option de configuration TxtCnvPlainTextHandlingForUnsupportedFormat a été activée mais le rapport de caractères étendus (caractères > 0x7F) est supérieur à la valeur indiquée par l'option de configuration TxtCnvMaxExtendedCharactersRatio .
IcmFceError: IcmConversionFailedCodepageNotSupported	IDXRC =3	Le codage du document était spécifié pour être dans une page de codes par défaut qui n'est pas prise en charge.

Tableau 14. Chaînes d'erreur de l'indexeur (suite)

Chaîne de notification	Code retour	Description
IcmFceError: InsoErrorNoFilterAvailable	IDXRC = 4	TxtCnvPlainTextHandlingForUnsupportedFormat = 0 Le document, ou l'une des pièces jointes imbriquées de l'élément, n'a pas pu être géré par les Filtres Oracle Outside In Technology qui ne prennent pas en charge le format du document.
IcmFceError: InsoErrorNoFilterAvailable	IDXRC = 5	TxtCnvPlainTextHandlingForUnsupportedFormat = 1 Le document, ou l'une des pièces jointes imbriquées de l'élément, n'a pas pu être géré par les Filtres Oracle Outside In Technology et même si l'option de configuration TxtCnvPlainTextHandlingForUnsupportedFormat a été activée, les caractères reconnus dans le document ou la pièce jointe indiquent que le contenu n'est pas un texte (binaire).
IcmFceError: IcmEmptyFile	IDXRC = 6	Le document, ou l'une des pièces jointes imbriquées de cet élément, n'a pas pu être traité car le fichier est vide (taille de 0 octet).
IcmFceError: InsoErrorConvertingAttachment	IDXRC = 7	Le document, ou l'une des pièces jointes imbriquées de l'élément, n'a pas pu être géré par les Filtres Oracle Outside In Technology en raison d'une erreur générale.
IcmFceError: IcmConversionSuppressed	IDXRC = 9	Le document ou l'une des pièces jointes imbriquées de l'élément, n'a pas pu être traité en raison d'une erreur grave. Le document a été traité sans les pièces jointes imbriquées. Le contenu de l'élément a été indexé. Toutefois, le contenu de toutes les pièces jointes imbriquées de l'élément n'a pas été indexé.

Validation d'index et opérations de réparation

L'indexeur de l'outil d'index de recherche de texte `afuIndexTool` propose des opérations d'index particulièrement utiles pour la vérification des incohérences ou la mise à jour des tables de la base de données d'index en vue de prendre en charge la fonction TIEFLAG d'IBM CommonStore.

L'outil `afuIndexTool` doit être appelé avec une opération. Toutefois, il ne peut effectuer qu'une opération à la fois sur l'index. Les opérations disponibles sont les suivantes :

- **-opValidate** : vérifie la présence d'incohérences d'indexation. Cette opération doit être réalisée sur l'ensemble des types d'élément CommonStore.
- **-opImportToCompleted** : ajoute tous les éléments indexés de type CommonStore à la table de base de données d'index Content Collector des tâches terminées. La valeur IDXRC affectée à chaque élément correspond à la valeur TIEFLAG de l'élément. En règle générale, cette opération d'importation est effectuée par l'intermédiaire du paramètre `afuEnableItemtype` sur les types d'élément ayant été archivés à l'aide d'IBM CommonStore.
- **-opExportToItemtype** : exporte les valeurs IDXRC qui ont été ajoutées à la table de base de données d'index des tâches terminées pour chaque élément (via l'opération `opImportToCompleted` au cours de l'exécution de l'outil `afuIndexTool`, ou au cours de l'indexation si le paramètre de configuration **IdxUpdateItemtypeTable** est défini sur True ou 1) vers la colonne TIEFLAG de la

table de composant du type d'élément dans IBM Content Manager. Ce traitement est généralement exécuté par l'outil `afuIndexer`.

- **-opRecreateTriggers** : recrée les déclencheurs de la table de base de données. Pour que l'indexation démarre, cette opération ne doit être réalisée que si elle est explicitement demandée.
- **-opDb2Hints** : affiche les paramètres DB2 à utiliser pour l'indexation.

Remarque : Pour toutes les opérations, à l'exception de `opDb2Hints`, effectuez les étapes suivantes avant de démarrer l'outil `afuIndexTool` sur un type d'élément :

1. Arrêtez l'archivage de tous les types d'élément.
2. Arrêtez `afuIndexer` avant d'exécuter `afuIndexTool`.
3. Vérifiez que tous les éléments du type d'élément ont été traités. `afuIndexTool` vérifie si la table des tâches ouvertes est vide avant de lancer une opération. La table des tâches ouvertes doit être vide (ne contient aucune ligne) pour des raisons techniques. Si cette table contient des lignes ou des éléments, `afuIndexTool` génère un message d'erreur indiquant que l'indexeur `afuIndexer` doit d'abord être exécuté.

Vous pouvez autrement supprimer manuellement toutes les lignes de la table des tâches ouvertes. La validation d'index identifie les éléments supprimés comme éléments manquants car il n'existe aucune entrée pour eux dans la table des tâches ouvertes ni dans la table des tâches terminées, et les ajoute à la table des tâches terminées avec un `IDXRC` de 199 si une option de réparation a été spécifiée.

Pour effectuer des opérations de l'outil d'indexation, exécutez le script suivant :

- Sous UNIX et Linux, exécutez :
`afuIndexTool.sh`
- Sous Windows :
`afuIndexTool.cmd`

Remarque : Lorsque vous exécutez `afuIndexTool` :

- Tous les arguments de ligne de commande font une distinction entre les majuscules et les minuscules.
- Les arguments de ligne de commande peuvent apparaître dans n'importe quel ordre.
- Utilisez les délimiteurs suivants pour les arguments qui peuvent transmettre plusieurs valeurs :
 - Sous Windows, un point-virgule (;)
 - Sous UNIX et Linux, deux-points (:)
- Le nom du type d'élément ne fait pas de distinction entre les majuscules et les minuscules.
- Les options de configuration de l'indexeur transmises via un argument ne sont pas sensibles à la casse.

Pour exécuter `afuIndexTool`, tenez compte des restrictions suivantes lors de la transmission des arguments :

- Vous ne devez transmettre qu'un seul argument d'opération. L'argument `-opValidate`, `-opRecreateTriggers`, `-opImportToCompleted`, `-opExportToItemtype` ou `opDb2Hints` est requis.
- L'argument `-useNseQueries`, `-useNseMemoryCopy` ou `-useNseTableCopy` est requis lorsque `-opValidate` est transmis.

- Les arguments facultatifs suivants ne peuvent être transmis qu'à l'aide de l'argument `-opValidate` :

```
-force
-deepvalidation
-repairAsOpen
-repairAsCompleted
-repairTieref
-begin <aaaa-mm-jj>
-end <aaaa-mm-jj>
-fulltieref
```

Le type d'élément et le nom de la base de données du serveur de bibliothèque dans laquelle est stocké le type d'élément doivent être transmis en tant qu'arguments obligatoires au processus d'indexation. Si vous exécutez l'outil `afuIndexTool` sans indiquer d'argument, ou si vous l'exécutez avec l'argument `-?`, les éléments suivants sont affichés :

```
afuIndexTool  IBM Content Collector V2.2.0.2
               Indexer for Text Search Tool for Validation and Repair
               Copyright IBM Corporation 2009, 2011
```

```
-----
afuIndexTool: Syntaxe :
  [-?|-h|-copyright]
  -it <type_élément>
  -database <base_de-données>
  [-user <ID_utilisateur|@variable_environnement|:nom_de_fichier>]
  [-password <mot_de_passe|@variable_environnement|:nom_de_fichier>]
  [-schema <schéma>]
  [-opValidate]
  -opRecreateTriggers|
  -opImportToCompleted|
  -opExportToItemtype|
  -opDb2Hints]
  [-deepvalidation]
  [-force]
  [-repairAsOpen]
  [-repairAsCompleted]
  [-repairTieref]
  [-useNseQueries|
  -useNseMemoryCopy|
  -useNseTableCopy]
  [-begin <aaaa-mm-jj>]
  [-end <aaaa-mm-jj>]
  [-fulltieref]
  [-batchsize <nombre_max_éléments_vérifiés_par_lots>]
  [-useconfig <fichier.ini>;<...>]
  [-usesettings <option=valeur>;<...>]
  [-exportconfig <fichier.ini>]
  [-listconfig]
  [-verbose] [-showsqli] [-trace]
  [-simulate]
  [-quiet]
```

Où :

- ? Répertorie l'ensemble des arguments que vous transmettez à l'outil d'indexation `afuIndexTool`.
- h Répertorie l'ensemble des arguments que vous transmettez à l'outil d'indexation `afuIndexTool`.
- copyright**
Affiche la notice de copyright et les informations de version.

-it *<type_élément>*
Obligatoire.

Type d'élément Content Manager (collection de documents). Les noms de type d'élément ne sont pas sensibles à la casse.

-database *<base_de_données>*
Obligatoire.

Nom de la base de données du serveur de bibliothèques où le type d'élément est stocké. Le nom de la base de données peut être indiqué dans deux formats qui déterminent comment l'outil de l'indexeur se connecte à la base de données. La connexion à la base de données est mise en oeuvre à l'aide de JDBC (Java Data Base Connectivity) qui définit comment un client Java peut accéder à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Pour utiliser une connexion JDBC de type 4, le nom de la base de données de l'argument `-database` doit inclure un nom de serveur, un numéro de port et un nom de base de données au format `<serveurbasededonnées>:<numérodeport>/<nombasededonnées>`, par exemple `-database localhost:50000/icm1sdb`. Il est recommandé d'utiliser JDBC type 4 comme méthode d'accès à la base de données mais ce mécanisme requiert la spécification complète de l'ID utilisateur et le mot de passe.

Pour utiliser une connexion JDBC type 2, la spécification du nom de la base de données de l'argument `-database` doit correspondre à un nom de connexion DB2 existant. En d'autres termes, le nom de la base de données doit être catalogué dans DB2, par exemple `-database icm1sdb`. Comme JDBC type 2 utilise une connexion à une base de données existante, il est inutile d'indiquer un ID utilisateur et un mot de passe. Toutefois, une version 32 bits native du pilote JDBC type 2 DB2 (bibliothèque partagée) doit être disponible et définie dans le chemin d'accès aux classes pour l'utilisation de la connexion JDBC type 2.

-user *<ID_utilisateur|@variable_environnement|:nom_de_fichier>*
Facultatif.

Indiquez un ID utilisateur. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un ID utilisateur de trois manières :

1. Indiquez l'ID utilisateur directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle l'ID utilisateur est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_USERID=user8
```

```
afuIndexTool -user @INDEXER_USERID
```

3. Indiquez le nom d'un fichier texte contenant l'ID utilisateur, par exemple,

```
afuIndexTool -user :myUserID.txt
```

Le fichier texte doit contenir une seule ligne avec l'ID utilisateur (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer l'ID utilisateur dans toutes les informations d'état des processus.

L'utilisateur doit disposer de droits d'accès à IBM Content Manager.

-password <mot_de_passe|@variable_enviromnement|:nom_fichier>

Facultatif.

Mot de passe de l'ID utilisateur doté des droits administrateur de la base de données. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un mot de passe de trois manières :

1. Indiquez le mot de passe directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle le mot de passe est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,
export INDEXER_PASSWORD=12345678

```
afuIndexTool -password @INDEXER_PASSWORD
```

3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,
afuIndexTool -password :myPassword.txt

Le fichier texte doit contenir une seule ligne avec le mot de passe (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer le mot de passe dans toutes les informations d'état des processus.

Les mots de passe ne sont pas consignés dans un format de texte normal lisible.

-schema <schéma>

Facultatif.

Nom du schéma de la base de données s'il est différent du schéma de la base de données en cours.

-tablespace <espace_de_table>

Facultatif.

Permet d'indiquer quel espace de table utiliser pour les tables utilisées par le processus de l'indexeur. Si cet argument n'est pas transmis, l'espace de table DB2 par défaut est utilisé.

-opValidate

Un argument d'opération est requis.

Il vérifie les incohérences d'indexation. L'opération vérifie que tous les éléments sont enregistrés dans la table des tâches terminées avec une valeur IDXRC et qu'ils sont présents dans l'index.

L'opération de validation est conseillée pour tous les types d'élément d'IBM CommonStore qui ont été activés pour l'indexation dans Content Collector afin de garantir l'intégrité des entrées d'index par rapport aux entrées de la table des tâches terminées. Cette opération n'est pas nécessaire pour les types d'élément qui ont été créés à l'aide de Content Collector compte tenu que tous les nouveaux éléments sont automatiquement suivis et surveillés dans les tables de l'indexeur Content Collector.

L'argument de l'opération de validation interagit avec les arguments suivants :

- -force (facultatif)
- -deepvalidation (facultatif)
- -repairAsOpen ou -repairAsCompleted ou -repairTiered (facultatif)
- -useNseQueries ou -useNseMemoryCopy ou -useNseTableCopy (obligatoire)

- -begin <aaaa-mm-jj> et -end <aaaa-mm-jj> (facultatif)
- -fulltieref avec -repairTieref (facultatif)

-opRecreateTriggers

Un argument d'opération est requis.

Il recrée les déclencheurs requis pour compléter la table des tâches ouvertes. Cette opération ne doit être effectuée que si afuIndexer signale l'erreur suivante :

The configuration option IdxUpdateItemTypeTable is enabled but the database triggers currently defined for the 'xxx' item type do not support updating the item type component table in Content Manager.

You must run afuIndexTool with the -opRecreateTriggers command-line argument first to correct this.

Cet argument est obligatoire si l'option de configuration

IdxUpdateItemTypeTable est activée pour un type d'élément qui a déjà été traité par Content Collector avant que la version 2.2 groupe de correctifs 1 ne soit installée.

Important : L'archivage dans le type d'élément doit être arrêté pendant la recréation des déclencheurs.

-opImportToCompleted

Un argument d'opération est requis.

Il alimente la table des tâches terminées à l'aide des éléments du type d'élément et leur attribue une valeur IDXRC selon qu'ils ont été précédemment indexés par l'indexeur rapide ou l'indexeur IBM Content Manager standard avec l'exit utilisateur IBM Text Search.

Ce traitement est généralement effectué par afuEnableItemtype si l'archivage dans les types d'élément a été réalisé par IBM CommonStore. Toutefois, si le type d'élément a été activé à l'aide du paramètre afuEnableItemtype avant IBM Content Collector V2.2.0.1, vous pouvez utiliser -opImportToCompleted pour renseigner la table des tâches terminées avec tous les éléments et leur attribuer les valeurs IDXRC correspondantes.

La valeur IDXRC qui est attribuée à chaque élément dépend de la valeur TIEFLAG qui a été utilisée dans les types d'élément au cours de l'indexation via l'indexeur rapide ou l'indexeur IBM Content Manager standard :

- Si la valeur TIEFLAG est comprise entre 10 et 19, la valeur IDXRC de l'élément correspondra à la valeur TIEFLAG d'origine (la valeur IDXRC sera comprise entre 10 et 19).
- Si la valeur TIEFLAG est supérieure à 10000, la valeur IDXRC affectée à l'élément sera égale à 180 (ERR_ITEMTYPE_TIEFLAG_ERROR), ce qui signifie que l'élément n'a pas encore été correctement indexé.

Après l'importation, vous devez réindexer les éléments avec un IDXRC de 180 à l'aide de l'indexeur Content Collector de recherche de texte ultérieure pour tenter de traiter correctement ces éléments et les ajouter à l'index documentaire. Pour plus de détails, voir «Utilisation de TIEFLAG lors de l'indexation dans IBM Content Collector», à la page 61.

-opExportToItemtype

Un argument d'opération est requis.

Exporte les valeurs IDXRC qui ont été ajoutées à la table de base de données d'index des tâches terminées pour chaque élément (via l'opération

opImportToCompleted au cours de l'exécution de l'outil afuIndexTool, ou au cours de l'indexation si l'option de configuration **IdxUpdateItemTypeTable** est activée) vers la colonne TIEFLAG de la table de composant du type d'élément dans IBM Content Manager. Ce traitement est généralement effectué automatiquement par l'outil afuIndexer, mais peut également être requis si l'option de configuration **IdxUpdateItemTypeTable** n'a pas été initialement paramétrée ou si vous avez indexés des éléments à l'aide de l'indexeur rapide ou une version de IBM Content Collector antérieure à 2.2.0.1 et que vous souhaitez continuer à utiliser la colonne TIEFLAG en ajoutant les valeurs IDXRC à cette colonne.

Ainsi, la colonne TIEFLAG indique la valeur IDXRC de chaque élément et les valeurs IDXRC de tous les éléments peuvent être recherchées dans la colonne TIEFLAG dans IBM Content Manager par une autre application, par exemple eDiscovery Manager.

-opDb2Hints

Un argument d'opération est requis.

Il affiche les paramètres DB2 à utiliser pour l'indexation. Les mêmes paramètres s'affichent lorsque le type d'élément est créé au cours de la configuration initiale ou à l'aide des outils de configuration de Content Collector.

-deepvalidation

Facultatif.

Cet argument ne peut être utilisé qu'avec l'argument d'opération -opValidate.

-deepvalidation vérifie l'intégrité de l'index. Il inclut la vérification si tous les éléments de la table des tâches terminées avec une valeur IDXRC inférieure à 100 ont été réellement indexés, à savoir, si les éléments se trouvent réellement dans l'index.

Remarque : Cette opération est considérablement plus longue que le traitement de la validation par défaut effectuée lorsque l'argument -opValidate seul est appelé.

-force

Facultatif.

Cet argument ne peut être utilisé qu'avec l'argument d'opération -opValidate.

Il applique le traitement de validation même si les conditions requises ne sont pas remplies, par exemple, si l'index devait être réorganisé.

-repairAsOpen

Facultatif. Si aucune action de réparation n'est mentionnée, l'outil afuIndexTool se contente de répertorier les erreurs et de les ajouter au fichier de commande DB2 ayant été généré.

Cet argument ne peut être utilisé qu'avec l'argument d'opération -opValidate.

Il ajoute tous les éléments ayant été identifiés comme incohérents à la table des tâches ouvertes. Ces éléments seront réindexés lors de la prochaine exécution de l'indexeur afuIndexer.

-repairAsCompleted

Facultatif. Si aucune action de réparation n'est mentionnée, l'outil afuIndexTool se contente de répertorier les erreurs et de les ajouter au fichier de commande DB2 ayant été généré.

Cet argument ne peut être utilisé qu'avec l'argument d'opération -opValidate.

Il ajoute tous les éléments ayant été identifiés comme incohérents à la table des tâches terminées avec une valeur IDXRC de 199. Vous pouvez tenter de retraiter ces éléments ultérieurement par l'intermédiaire de la fonction de réindexation de l'indexeur `afuIndexer`.

-repairTieref

Facultatif. Si aucune action de réparation n'est mentionnée, l'outil `afuIndexTool` se contente de répertorier les erreurs et de les ajouter au fichier de commande DB2 ayant été généré.

Cet argument ne peut être utilisé qu'avec l'argument d'opération `-opValidate`.

La valeur TIEREF est normalement générée par IBM Content Manager, automatiquement attribuée à chaque utilisateur et requise par la fonction définie par l'utilisateur `AFUFetchFile` pour charger les données XML d'un élément. Ceci correspond à l'utilitaire `GenTie` utilisé par l'indexeur rapide.

L'argument `-repairTieref` vérifie que tous les éléments disposent d'une valeur dans la colonne TIEREF de la table de composant d'IBM Content Manager.

Lorsque l'indexeur `afuIndexer` est exécuté, le processus requiert une valeur TIEREF pour chaque élément. Si un élément ne comporte pas de valeur TIEREF valide, l'élément ne peut pas être indexé et il est ajouté à la table des tâches terminées avec une valeur IDXRC de 140.

Lorsque l'argument `-repairTieref` est mentionné au cours de la validation, une nouvelle valeur est calculée pour les éléments dépourvus de valeur TIEREF et cette valeur est stockée dans la table de composant du type d'élément d'IBM Content Manager. Notez que cette nouvelle valeur TIEREF peut être utilisée par l'indexeur Content Collector uniquement. La valeur TIEREF générée par `afuIndexTool` ne convient PAS au traitement par l'indexeur rapide.

Conseil : Pour vérifier de manière approfondie les éléments ne possédant pas de valeurs TIEREF, vous devez toujours spécifier l'option de ligne de commande `-deepvalidation` et `-repairAsCompleted` avec `-repairTieref`.

Une fois l'opération de réparation terminée, tous les éléments comportant une valeur IDXRC égale à 140 peuvent être réindexés.

-useNseQueries

Une opération d'accès à l'index est requise.

Cet argument ne peut être utilisé qu'avec l'argument d'opération `-opValidate`.

Il effectue la vérification des éléments de l'index à l'aide des requêtes standard Net Search Extender. Il s'agit de la méthode privilégiée et celle-ci ne nécessite aucune ressource supplémentaire.

-useNseTableCopy

Une opération d'accès à l'index est requise.

Cet argument ne peut être utilisé qu'avec l'argument d'opération `-opValidate`.

Il effectue la vérification des éléments de l'index en copiant une structure de données Net Search Extender interne dans une table DB2 temporaire. Cette méthode nécessite des ressources DB2 supplémentaires.

-useNseMemoryCopy

Une opération d'accès à l'index est requise.

Cet argument ne peut être utilisé qu'avec l'argument d'opération `-opValidate`.

Il effectue la vérification des éléments de l'index en réalisant une copie en mémoire d'une structure de données Net Search Extender interne. Cette méthode risque d'échouer en raison de ressources mémoire insuffisantes si l'index est trop volumineux.

-begin <aaaa-mm-jj>

Facultatif.

Cet argument ne peut être utilisé qu'avec l'argument d'opération `-opValidate`. Il définit la date à partir de laquelle vérifier les éléments. Par défaut, tous les éléments à partir de 2000-01-01 sont vérifiés.

-end <aaaa-mm-jj>

Facultatif.

Cet argument ne peut être utilisé qu'avec l'argument d'opération `-opValidate`. Il définit la date jusqu'à laquelle vérifier les éléments. Par défaut, tous les éléments jusqu'à la date en cours sont vérifiés.

-fulltieref

Facultatif.

Si aucune action de réparation n'est mentionnée, l'outil `afuIndexTool` se contente de répertorier les erreurs et de les ajouter au fichier de commande DB2 ayant été généré.

Cet argument ne peut être utilisé qu'avec l'argument d'opération `-opValidate` et l'argument `-repairTieref`.

Lorsque l'argument `-repairTieref` est mentionné au cours de la validation, une nouvelle valeur de TIEREF est calculée pour les éléments dépourvus de valeur TIEREF. Cette valeur est enregistrée dans la table de composant de type d'élément dans IBM Content Manager et peut être utilisée par l'indexeur Content Collector pour la recherche de texte. Par défaut, la valeur TIEREF générée par `afuIndexTool` ne convient pas au traitement par l'indexeur rapide ni par IBM eMail Search for CommonStore. Toutefois, si vous indiquez `-fulltieref` en même temps que l'argument `-repairTieref`, la valeur TIEREF générée est également adaptée au traitement par IBM eMail Search for CommonStore ou par d'autres applications qui nécessitent pas l'utilisation de la fonction UDF `ICMFetchFilter`.

Il est à noter que lorsque vous indiquez `-fulltieref`, le calcul complet de l'information TIEREF est beaucoup plus coûteux et demande beaucoup plus de temps que le calcul de l'information TIEREF abrégée, qui est suffisante pour l'indexation avec l'indexeur Content Collector pour la recherche de texte.

Une fois l'opération de réparation terminée, tous les éléments comportant une valeur IDXRRC égale à 140 peuvent être réindexés.

-batchsize <nombre_max_éléments_vérifiés_par_lots>

Facultatif.

Cet argument indique le nombre d'éléments à réparer par transaction. Par défaut, le nombre d'éléments traités par transaction est déterminé par l'option **GenDatabaseTransactionSize** dans le fichier de configuration de l'indexeur associé au type d'élément. La valeur par défaut est 100 000.

Lorsque vous indiquez cet argument de ligne de commande, vous remplacez la valeur de **GenDatabaseTransactionSize**. Le nombre requis varie en fonction de la taille du journal des transactions IBM DB2. Toutefois, si vous détectez des erreurs SQL lors de l'insertion d'index ou de la mise à jour d'opérations, telles que l'escalade de verrous dans DB2, réduisez cette valeur jusqu'à 10 000.

-useconfig <fichier.ini>;<fichier.ini>

Facultatif.

Permet d'utiliser d'autres fichiers de configuration .ini dans un autre chemin. Plusieurs fichiers .ini peuvent être indiqués. Ils sont chargés dans l'ordre dans lequel ils ont été indiqués.

-usettings "<option=valeur>;<option=valeur>"

Facultatif.

Active le remplacement direct et temporaire des valeurs de configuration persistantes de l'indexeur stockées dans le fichier .ini. Vous pouvez ajouter un certain nombre de paires <option=valeur> séparées par le délimiteur du système d'exploitation. Les paires de valeurs d'option doivent être placées entre guillemets. Les options de configuration que vous répertoriez (<option>) ne sont pas sensibles à la casse. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de l'indexeur.

-exportconfig <fichier.ini>

Facultatif.

Exporte la configuration de l'indexeur dans le fichier .ini indiqué dans le fichier <fichier .ini>.

-listconfig

Facultatif.

Affiche toutes les options de configuration et les valeurs associées définies pour la procédure d'exécution de l'indexeur en cours. Aucune autre opération n'est effectuée.

-verbose

Facultatif.

Affiche des informations détaillées sur le traitement. La définition de l'option -verbose est utile au début pour obtenir des informations détaillées sur l'ensemble des traitements. Si l'option -verbose est spécifiée, les fichiers journaux contiennent également des informations supplémentaires.

-showsql

Facultatif.

Affiche toutes les commandes SQL DB2 exécutées par le système. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de tous les traitements système DB2. Si cet argument est indiqué, toutes les instructions SQL sont également ajoutées au fichier journal.

-trace

Facultatif.

Fournit des informations plus détaillées que l'option -verbose en affichant des données supplémentaires dans les fichiers journaux pour effectuer l'identification des erreurs. Ces informations sont destinées uniquement au service de support logiciel IBM et requièrent l'utilisation du fichier de propriété de trace.

-simulate

Facultatif.

Exécute le processus de l'indexeur sans mettre à jour les bases de données. Net Search Extender ne s'exécute pas. Cet argument est utile pour vérifier le traitement des fichiers et la configuration.

-quiet

Facultatif

Aucune étape de traitement n'est affichée sur la console.

Critères de validation d'index et étapes de traitement

Les opérations de validation d'index ne peuvent être exécutées sur un index que si certains critères sont réunis.

En outre, vous devez spécifier la méthode à utiliser pour accéder aux éléments de l'index afin de procéder au traitement de validation. La méthode d'accès à l'index est un argument obligatoire qui doit être transmis lors de l'appel de l'opération de validation.

Vous pouvez également spécifier les opérations de réparation à exécuter sur les éléments erronés. Il existe trois arguments d'opérations de réparation. Ces trois arguments peuvent être transmis séparément ou simultanément lors de l'appel de l'opération de validation.

Depuis IBM Content Collector V3, la validation de l'index est exécutée en plusieurs tâches (unités d'exécution) parallèles. Ceci permet d'améliorer considérablement la vitesse de traitement lors de la validation de l'index. Le nombre de tâches de validation de l'index parallèles est également contrôlé par l'option de configuration **IdxNumberOfParallelIndexerTasks**.

Critères de validation

Lorsque `afuIndexTool` est exécuté avec l'argument `-opValidate`, les critères suivants sont vérifiés avant que l'opération de validation ne commence :

- Si la table des tâches ouvertes n'est pas vide, le processus de validation de l'index s'arrête immédiatement. Un message d'erreur indiquant que `afuIndexer` doit être démarré pour pouvoir commencer la validation s'affiche. Tous les éléments de la table des tâches ouvertes doivent être traités et la table des tâches ouvertes doit être vide.
- Si le nombre d'éléments de l'index Net Search Extender est de zéro, cela signifie que l'index est vide, le processus de validation de l'index s'arrête immédiatement et un message d'erreur indiquant que l'index ne contient aucun élément s'affiche. Dans ce cas, vous devez exécuter `afuRecreateIndex`.
- Si l'index n'a pas été récemment réorganisé et qu'il doit l'être, le processus de validation de l'index s'arrête et un message indiquant que l'index doit tout d'abord être réorganisé s'affiche, sauf si `-opValidate` est appelé avec `-force`.
- Si la table des tâches terminées est vide et qu'aucun élément n'a été indexé, le processus de validation s'arrête immédiatement sauf si `-opValidate` est appelé avec `-force`.
- Si le nombre d'éléments de la table des tâches terminées est égal au nombre d'éléments du type d'élément, l'outil suppose que tous les éléments ont été correctement traités et le processus de validation est arrêté sauf si `-opValidate` est appelé avec `-force` ou `-deepvalidation`.
- Si le nombre d'éléments de la table des tâches terminées est supérieur au nombre d'éléments du type d'élément, le processus de validation est exécuté sur les éléments du type d'élément qui figurent également dans la table des tâches terminées et ignore les éléments de la table des tâches terminées qui ont été supprimés du type d'élément. Ces entrées peuvent être supprimées de la table des tâches terminées à chaque nouvelle exécution de `afuIndexer`.

Méthodes d'accès à l'index

Pour valider le contenu de l'index, vous devez sélectionner la méthode à utiliser pour accéder aux éléments indexés dans l'index Net Search Extender, puis transmettre cette méthode en tant qu'argument lors de l'appel de l'opération de validation. Vous pouvez choisir d'utiliser l'une des trois méthodes d'accès suivantes :

1. `-useNseMemoryCopy` : Vérifie les éléments de l'index en effectuant une copie en mémoire de la structure de données Net Search Extender interne. La structure de données interne est chargée une fois en mémoire, ce qui peut prendre du temps, mais tout accès ultérieure est alors plus rapide que l'émission de requêtes.

Important : Cette méthode peut échouer en raison de ressources de mémoire insuffisantes si les index sont volumineux. Dans ce cas, choisissez une autre méthode d'accès.

2. `-useNseTableCopy` : Vérifie les éléments de l'index en créant une copie de la structure de données Net Search Extender interne dans une table DB2 temporaire. La création d'une copie de la structure de données Net Search Extender interne peut prendre du temps, mais tout accès ultérieur est alors plus rapide que l'émission de requêtes.

Important : Cette méthode requiert des ressources DB2 supplémentaires. A la fin du processus de validation, ces ressources sont libérées automatiquement.

3. `-useNseQueries` : Vérifie les éléments de l'index à l'aide de requêtes Net Search Extender standard. N'utilisez cette méthode que si vous avez une bonne raison d'utiliser les autres options car elle ne requiert aucune ressource supplémentaire.

Période de validation

Si aucune date de début et de fin n'a été spécifiée, tous les éléments du type de données sont vérifiés afin de déterminer le premier élément ajouté (par année, mois et jour). Pour s'assurer que tous les éléments sont vérifiés, vous ne devez pas définir de date de début et de fin.

Si vous avez indiqué une date de début et de fin sur la ligne de commande, l'intervalle de temps est limité aux dates de début et de fin spécifiées. La fenêtre de temps spécifiée et utilisée doit toujours être complète, c'est-à-dire que vous devez définir une date de début et une date de fin.

Indiqué dans le fichier journal par :

```
Looking for items archived in year: 2000 ... Looking for items
archived in year/month: 2009/01 ... Checking items for day ...
Checking items for hour ... Checking items for minute
```

Une fois les informations d'année, de mois et de jour de début déterminées, les éléments sont validés sur une base quotidienne, c'est-à-dire que tous les éléments qui ont été ajoutés au type d'élément un jour donné sont identifiés, et ces éléments sont ensuite validés. Si le nombre d'éléments ajoutés au type d'élément un jour donné dépasse la taille de lot indiquée, la fenêtre de temps passe d'un jour à une heure, ce qui signifie que tous les éléments ajoutés au type d'élément pendant une heure donnée sont validés. Si le nombre d'éléments ajoutés pendant cette heure dépasse la taille de lot indiquée, la fenêtre de temps est encore réduite à une minute, et tous les éléments ajoutés au type d'élément pendant cette minute sont validés.

Etapes de traitement de validation

Le processus de validation vérifie toujours un lot d'éléments qui ont été ajoutés au type d'élément au cours d'une même période, c'est-à-dire le même jour, la même heure ou la même minute.

Indiqué dans le fichier journal par :

```
Processing items matching: xxx...
```

Le processus de validation commence par déterminer le nombre d'éléments qui ont été ajoutés au type d'élément au cours de la période donnée.

Indiqué dans le fichier journal par :

```
RES: Number of items found: ...
```

Le processus de validation recherche ensuite les éléments par cardinalité dans la table des tâches terminées.

Indiqué dans le fichier journal par :

```
Searching in table of completed tasks for pattern ... Expecting xxx items... SQL: SELECT COUNT(*) FROM AFU1FTICOMPxxxxx WHERE ITEMID LIKE ? || '%'.
```

Les critères de cardinalité sont les suivants :

- Si le nombre d'éléments trouvés dans la table des tâches terminées ne correspond pas au nombre attendu d'éléments ajoutés au type d'élément au cours de la période donnée, une erreur est renvoyée et tous les éléments de la période sont vérifiés individuellement.
- Si le nombre d'éléments trouvés dans la table des tâches terminées correspond au nombre attendu d'éléments ajoutés au type d'élément au cours de la période donnée, la période suivante est calculée et le lot d'éléments suivant est validé sauf si -deepvalidation est spécifié. Si -deepvalidation n'est pas spécifié, seule la cardinalité des ensembles est vérifiée.

Les étapes suivantes sont effectuées pour valider des éléments individuels si le nombre d'éléments ne correspond pas ou si -deepvalidation est spécifié :

- Si l'élément se trouve dans la table des tâches terminées avec une valeur IDXRC supérieure ou égale à 100, cela signifie que l'élément n'a PAS été ajouté à l'index de texte intégral. L'élément est considéré comme valide.
- Si l'élément se trouve dans la table des tâches terminées avec une valeur IDXRC inférieure à 100, l'élément doit être ajouté à l'index de texte intégral.
 - Si -deepvalidation n'a PAS été spécifié sur la ligne de commande, l'élément est considéré comme valide.
 - Si -deepvalidation a été spécifié sur la ligne de commande, le processus de validation recherche l'élément dans l'index de texte intégral.
- Si l'élément est introuvable dans la table des tâches terminées, ou si -deepvalidation a été spécifié sur la ligne de commande, un avertissement est renvoyé, et le processus de validation vérifie si l'élément se trouve dans l'index de texte intégral.

Indiqué dans le fichier journal par :

```
"WARNING x items were expected in the table of completed tasks for the given time interval. However, only x items were listed. A validation process will now be started to detect which items were not processed by the indexer"
```

Etapes de traitement validation avancées

Le traitement de validation des éléments dans l'index de texte intégral Net Search Extender est effectué via l'émission d'une requête Net Search Extender ou l'analyse du contenu de la structure de données Net Search Extender interne qui peut être chargée en mémoire ou dans une table DB2 en fonction de l'argument de méthode d'accès à l'index transmis.

Indiqué dans le fichier journal par :

- Using a Net Search Extender query: SQL: SELECT ITEMID FROM ICMUT01260001 WHERE CONTAINS(TIÉREF, '(SECTIONS ("itemid")' ...)
- Utilisation d'une requête DB2 : SQL: SELECT * FROM AFU1xxx WHERE COMPKEY = ? ...)

Les étapes suivantes sont effectuées pendant le traitement de validation avancé :

- Si l'élément se trouve dans la table des tâches terminées, mais pas dans l'index de texte intégral Net Search Extender, un message d'erreur est consigné dans le journal et l'élément est traité en fonction de l'option d'argument de réparation transmis sur la ligne de commande.

Indiqué dans le fichier journal par :

SEVERE: Item xxx could not be found in the index but was found in the table of completed tasks

- Si l'élément est introuvable dans la table des tâches terminées et dans l'index de texte intégral Net Search Extender, un message d'erreur est consigné dans le journal et l'élément est traité en fonction de l'option d'argument de réparation transmis sur la ligne de commande.

Indiqué dans le fichier journal par :

SEVERE: Item xxx could neither be found in the table of completed tasks nor in the index

- Si l'élément est introuvable dans la table des tâches terminées mais qu'il se trouve dans l'index de texte intégral Net Search Extender, un message d'erreur est consigné dans le journal et l'élément est traité en fonction de l'option d'argument de réparation transmis sur la ligne de commande.

Indiqué dans le fichier journal par :

WARNING: Item xxx could not be found in the table of completed tasks but was found in the index

Ceci est considéré comme une erreur car l'élément a été ajouté à l'index mais il ne figure pas dans la table des tâches terminées.

Etapes de traitement de réparation pendant la validation

Une fois que le processus de validation a déterminé si un élément se trouve dans la table des tâches terminées et dans l'index, l'action de réparation qui est exécutée sur un élément erroné est déterminée par les options d'argument de réparation qui ont été transmis avec l'opération de validation (validation standard et avancée). Si aucune action de réparation n'est spécifiée, afuIndexTool répertorie simplement les erreurs et crée une entrée dans le fichier de commande SQL DB2.

Les options d'action de réparation suivantes sont disponibles :

- **-repairAsOpen** : ajoute tous les éléments identifiés comme incohérents à la table des tâches ouvertes. Ces éléments seront indexés lors de l'exécution suivante de afuIndexer.

Si `-repairAsOpen` n'est PAS transmis en tant qu'argument, `afuIndexTool` crée un fichier de commande SQL DB2 dans le répertoire de l'instance de travail avec les commandes SQL DB2 qui auraient été exécutées si `-repairAsOpen` avait été transmis en tant qu'argument. Ce fichier répertorie les commandes SQL qui renseignent la table des tâches ouvertes et qui peuvent être exécutées en entrant `db2 [-a] -f <commandFile>`.

- **-repairAsCompleted** : ajoute tous les éléments identifiés comme incohérents à la table des tâches terminées avec la valeur IDXRC de 199. Le retraitement de ces éléments peut être envisagé à une étape ultérieure à l'aide de la fonction de réindexation dans `afuIndexer`.

Si `-repairAsCompleted` n'est PAS transmis en tant qu'argument, `afuIndexTool` crée un fichier de commande SQL DB2 dans le répertoire d'instance de travail avec les commandes DB2 qui auraient été exécutées si `-repairAsCompleted` avait été transmis en tant qu'argument. Ce fichier répertorie les commandes SQL qui renseignent la table des tâches terminées et qui peuvent être exécutées en entrant `db2 [-a] -f <commandFile>`.

- **-repairTieref** : si un élément ne possède pas de valeur TIEREF valide, il ne peut généralement pas être indexé et est ajouté à la table des tâches terminées avec une valeur IDXRC de 140. Lorsque `-repairTieref` est spécifié pendant la validation, une nouvelle valeur TIEREF est calculée pour les éléments sans valeur TIEREF, cette valeur est enregistrée dans la table de composant de type d'élément dans IBM Content Manager et peut être utilisée par l'indexeur Content Collector. La valeur TIEREF générée ne convient PAS au traitement par l'indexeur rapide.

Utilisez cette option pour vérifier que tous les éléments possèdent une valeur TIEREF valide et, si ce n'est pas le cas, pour calculer une valeur valide. Pour vérifier de manière approfondie les éléments ne possédant pas de valeurs TIEREF, vous devez toujours spécifier l'option de ligne de commande `-deepvalidation` et `-repairAsCompleted` avec `-repairTieref`. Une fois l'opération de réparation exécutée, vous devez exécuter l'outil `afuIndexer` pour réindexer tous les éléments avec une valeur IDXRC de 140.

Si `-repairTieref` n'est PAS transmis en tant qu'argument, `afuIndexTool` crée un fichier de commande SQL DB2 dans le répertoire d'instance de travail avec les commandes DB2 qui auraient été exécutées si `-repairTieref` avait été transmis en tant qu'argument. Ce fichier répertorie les commandes SQL qui recréent la valeur TIEREF et qui peuvent être exécutées en entrant `db2 [-a] -f <commandFile>`.

Exécution de l'outil de contrôle de la configuration

Le processus de l'indexeur de recherche de texte `afuCheck` valide l'installation et la configuration de l'indexeur.

La procédure de vérification de la configuration recherche une configuration par défaut correspondant au nom du type d'élément dans le répertoire en cours. Si elle la trouve, elle la charge. Si aucun fichier `.ini` par défaut n'est détecté ou qu'aucun fichier `.ini` n'est transmis en tant qu'argument, la procédure de vérification de la configuration affiche un message d'erreur et arrête le traitement.

Elle recherche toutes les bibliothèques dépendantes et vérifie que le type d'élément possède un fichier de configuration `.ini` valide. Elle vérifie notamment la configuration de Net Search Extender, compare les paramètres de configuration par rapport au modèle de fichier Net Search Extender et détermine si la fonction définie par l'utilisateur `AFUFetchFile` est correctement installée.

afuCheck ne corrige aucun élément de la configuration. Il affiche uniquement les éléments détectés.

Remarque : Ce processus de validation est automatiquement effectué chaque fois que les outils afuIndexer et afuRecreateIndex sont exécutés. Cela signifie que le processus afuCheck n'a pas à être appelé explicitement avant l'exécution des processus afuIndexer et afuRecreateIndex.

Pour vérifier la configuration d'un type d'élément, procédez comme suit.

- Sous UNIX et Linux, exécutez
afuCheck.sh
- Sous Windows, exécutez
afuCheck.cmd

Remarque : Lorsque vous exécutez le processus afuCheck :

- Tous les arguments de ligne de commande font une distinction entre les majuscules et les minuscules.
- Les arguments de ligne de commande peuvent apparaître dans n'importe quel ordre.
- Utilisez les délimiteurs suivants pour les arguments qui peuvent transmettre plusieurs valeurs :
 - Sous Windows, un point-virgule (;)
 - Sous UNIX et Linux, deux-points (:)
- Le nom du type d'élément ne fait pas de distinction entre les majuscules et les minuscules.
- Les options de configuration de l'indexeur transmises via un argument ne sont pas sensibles à la casse.

Si vous exécutez AfuCheck et que vous n'indiquez pas d'argument ou que vous entrez la commande suivie de `-?`, les données suivantes s'affichent :

```
afuCheck: Syntaxe :
[-?|-h|-copyright]
-it <itemtype>
-database <database>
[-user <userID|@environmentVariable|:filename>]
[-password <password|@environmentVariable|:filename>]
[-schema <schema>]
[-useconfig <.ini file>;<.ini file>]
[-usesettings "<option=value>;<...>"]
[-exportconfig <fichier.ini>]
[-listconfig]
[-verbose] [-showsql] [-trace]
[-quiet]
```

où :

- ? Répertorie tous les arguments que vous pouvez transmettre à l'outil afuCheck.
- h Répertorie tous les arguments que vous pouvez transmettre à l'outil afuCheck.
- copyright Affiche la notice de copyright et les informations de version.

-it <typed'élément>
Obligatoire.

Type d'élément Content Manager (collection de documents) que vous souhaitez indexer.

Important : Si votre type d'élément est un type d'élément composé de courrier électronique, n'entrez pas le type d'élément de pièce jointe. Les types d'élément de pièce jointe ne peuvent pas être traités seuls mais sont traités automatiquement comme partie du type d'élément de courrier électronique.

-database <basededonnées>

Obligatoire.

Nom de la base de données du serveur de bibliothèques où le type d'élément est stocké. Le nom de la base de données peut être indiqué dans deux formats qui déterminent comment l'outil de l'indexeur se connecte à la base de données. La connexion à la base de données est mise en oeuvre à l'aide de JDBC (Java Data Base Connectivity) qui définit comment un client Java peut accéder à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Pour utiliser une connexion JDBC de type 4, le nom de la base de données de l'argument -database doit inclure un nom de serveur, un numéro de port et un nom de base de données au format <serveurbasededonnées>:<numérodeport>/<nombasededonnées>, par exemple -database localhost:50000/icm1sdb. Il est recommandé d'utiliser JDBC type 4 comme méthode d'accès à la base de données mais ce mécanisme requiert la spécification complète de l'ID utilisateur et le mot de passe.

Pour utiliser une connexion JDBC type 2, la spécification du nom de la base de données de l'argument -database doit correspondre à un nom de connexion DB2 existant. En d'autres termes, le nom de la base de données doit être catalogué dans DB2, par exemple -database icm1sdb. Comme JDBC type 2 utilise une connexion à une base de données existante, il est inutile d'indiquer un ID utilisateur et un mot de passe. Toutefois, une version 32 bits native du pilote JDBC type 2 DB2 (bibliothèque partagée) doit être disponible et définie dans le chemin d'accès aux classes pour l'utilisation de la connexion JDBC type 2.

-user <userID|@environmentVariable|:filename>

Facultatif

Indiquez un ID utilisateur. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un ID utilisateur de trois manières :

1. Indiquez l'ID utilisateur directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle l'ID utilisateur est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,
export INDEXER_USERID=user8

afuCheck -user @INDEXER_USERID
3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,
afuCheck -user :myUserID.txt

Le fichier texte doit contenir une seule ligne avec l'ID utilisateur (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer l'ID utilisateur dans toutes les informations d'état des processus.

L'utilisateur doit disposer de droits d'accès à IBM Content Manager.

-password <userID|@environmentVariable|:filename>

Facultatif

Mot de passe de l'ID utilisateur doté des droits administrateur de la base de données. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un mot de passe de trois manières :

1. Indiquez le mot de passe directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle le mot de passe est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_PASSWORD=12345678
```

```
afuCheck -password @INDEXER_PASSWORD
```

3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuCheck -password :myPassword.txt
```

Le fichier texte doit contenir une seule ligne avec le mot de passe (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer le mot de passe dans toutes les informations d'état des processus.

Les mots de passe ne sont pas consignés dans un format de texte normal lisible.

-schema <schéma>

Facultatif.

Nom du schéma de la base de données s'il est différent du schéma de la base de données en cours ou que l'option de configuration **GenJdbcType** correspond à 4.

-useconfig <.ini file>;<.ini file>

Facultatif.

Permet d'utiliser d'autres fichiers de configuration .ini dans un autre chemin. Plusieurs fichiers .ini peuvent être indiqués. Ils sont chargés dans l'ordre dans lequel ils ont été indiqués.

-usettings "<option=value>;<option=value>"

Facultatif.

Active le remplacement direct des valeurs de configuration de l'indexeur stockées dans le fichier .ini. Vous pouvez ajouter un nombre illimité de paires <option=valeur> séparées par le délimiteur de votre système d'exploitation. Les paires de valeurs d'option doivent être placées entre guillemets. Les options de configuration que vous affichez (<option>) ne font pas de distinction entre les majuscules et les minuscules. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de l'indexeur.

-exportconfig <fichier.ini>

Facultatif.

Exporte la configuration de l'indexeur dans le fichier .ini indiqué dans le fichier <fichier .ini>.

-listconfig

Facultatif.

Affiche toutes les options de configuration et les valeurs associées définies pour l'exécution de l'indexeur en cours.

-verbose

Facultatif.

Affiche des informations détaillées sur le traitement. L'utilisation de l'option **-verbose** est utile au début pour obtenir des informations détaillées sur l'ensemble des traitements. Si **-verbose** est indiqué, les fichiers journaux contiennent également des informations supplémentaires.

-showsql

Facultatif.

Affiche toutes les commandes SQL DB2 exécutées par le système. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de tous les traitements système DB2. Si cet argument est indiqué, toutes les instructions SQL sont également ajoutées au fichier journal.

-trace

Facultatif.

Fournit des informations plus détaillées que l'option **-verbose** en affichant des données supplémentaires dans les fichiers journaux pour effectuer l'identification des erreurs. Cette opération est obligatoire pour le service de support logiciel IBM uniquement et requiert l'utilisation du fichier des propriétés de la fonction de trace.

-quiet

Aucune étape de traitement n'est affichée sur la console.

Recréation d'un index

Dans la plupart des cas, vous devez uniquement vérifier que le type d'élément peut faire l'objet de recherches de texte, c'est-à-dire qu'il a été activé pour l'indexation et que vous n'avez pas à créer l'index de texte. Toutefois, si la configuration de l'index Net Search Extender est incorrecte ou si l'index est corrompu et qu'il ne peut pas être restauré à partir d'une sauvegarde, l'index doit être recréé.

Le processus de recréation de l'index ne peut être lancé que s'il existe un fichier de configuration de l'indexeur pour le type d'élément. Si le processus de recréation de l'index est appelé sans fichier de configuration, il recherche une configuration par défaut pour le type d'élément dans le répertoire en cours et le charge, le cas échéant. Si aucun fichier de configuration n'est trouvé ou si le fichier de configuration est transmis dans un argument au processus de recréation de l'index, le processus affiche un message d'erreur et le processus s'arrête.

La recréation de l'index lui-même prend peu de temps. Toutefois, le remplissage de l'index est une procédure très lourde, notamment si le nombre de documents à réindexer est élevé. Ne créez pas d'index sauf si vous êtes absolument certain que l'index a été altéré et qu'une recréation de l'index est indispensable.

Important : Avant le lancement du processus de recréation, vous devez arrêter les services d'application Web IBM Content Collector. L'archivage dans le type d'élément doit être arrêté.

Par défaut, la recréation de l'index est exécutée en mode interactif. Cela signifie que l'utilisateur est invité, par exemple, à confirmer si la recréation de l'index doit réellement être effectuée et si l'archivage dans ce type d'élément a été arrêté. Pour

exécuter le processus de recréation de l'index en mode de traitement par lots sans demander la confirmation de l'utilisateur, affectez la valeur 0 à la variable d'environnement **AFU_INDEXER_PROMPT_RECREATE**.

La recréation de l'index (auparavant appelée fonction `RecreateIndex`) comprend les opérations suivantes :

- Suppression de l'index documentaire Net Search Extender
- Création d'un index documentaire Net Search Extender vide
- Création d'une table de base de données vide pour les tâches terminées
- Création d'une table de base de données vide pour les tâches ouvertes
- Réinitialisation des données de surveillance pour le type d'élément dans la table des tâches de surveillance
- Suppression des anciens déclencheurs et création de déclencheurs dans la table des tâches ouvertes
- Insertion de tous les ID élément du type d'élément dans la table des tâches ouvertes

Avertissement :

- Si vous souhaitez modifier des définitions de zone d'attribut pour le type d'élément existant, effectuez ces modifications dans le fichier modèle du type d'élément et le fichier de configuration du type d'élément avant de recréer l'index.
- Le processus `afuRecreateIndex` crée uniquement l'index ; il ne le remplit pas. Pour remplir l'index, vous devez exécuter `afuIndexer` qui, en fonction de la taille de l'index d'origine, peut être une procédure très longue.
- Par défaut, le processus de recréation de l'index utilise le fichier modèle associé à l'index Net Search Extender lorsque l'index est à nouveau créé. Cela signifie que, par défaut, `afuRecreateIndex` crée une copie du fichier modèle Net Search Extender avant de supprimer l'index, puis il utilise cette copie pour créer le nouvel index Net Search Extender. Si vous souhaitez ou si vous devez utiliser un autre fichier modèle lors de la recréation de l'index, exécutez `afuRecreateIndex` avec l'argument de ligne de commande `-model file` et indiquez un fichier modèle différent.

Pour recréer l'index, procédez comme suit.

- Sous UNIX et Linux, exécutez
`afuRecreateIndex.sh`
- Sous Windows, exécutez
`afuRecreateIndex.cmd`

Remarque : Lorsque vous exécutez le processus `afuRecreateIndex` :

- Tous les arguments de ligne de commande font une distinction entre les majuscules et les minuscules.
- Les arguments de ligne de commande peuvent apparaître dans n'importe quel ordre.
- Utilisez les délimiteurs suivants pour les arguments qui peuvent transmettre plusieurs valeurs :
 - Sous Windows, un point-virgule (;)
 - Sous UNIX et Linux, deux-points (:)
- Le nom du type d'élément ne fait pas de distinction entre les majuscules et les minuscules.

- Les options de configuration de l'indexeur transmises via un argument ne sont pas sensibles à la casse.

Si vous exécutez `afuRecreateIndex` et que vous n'indiquez pas d'argument ou que vous entrez la commande suivie de `-?`, les données suivantes s'affichent :

```
afuRecreateIndex : Syntaxe :
[-?|-h|-copyright]
-it <type_élément>
-database <base_de_données>
[-user <ID_utilisateur|@variable_environnement|:nom_fichier>]
[-password <mot_de_passe|@variable_environnement|:nom_fichier>]
[-schema <schéma>]
[-indexdirectory <répertoire>]
[-workdirectory <répertoire>]
[-modelfile <nom_fichier>]
[-tablespace <espace_table>]
[-useconfig <fichier.ini>;<...>]
[-usesettings "<option=valeur>;<...>"]
[-exportconfig <fichier.ini>]
[-listconfig]
[-verbose] [-showsql] [-trace]
[-simulate]
[-quiet]
```

où :

-? Répertorie tous les arguments que vous pouvez transmettre à l'outil `afuRecreateIndex`.

-h Répertorie tous les arguments que vous pouvez transmettre à l'outil `afuRecreateIndex`.

-copyright

Affiche la notice de copyright et les informations de version.

-it <type_élément>

Obligatoire.

Type d'élément IBM Content Manager (collection de documents) que vous souhaitez indexer. Les noms de type d'élément des modèles de données intégrés et composé ne font pas de distinction entre les majuscules et les minuscules.

Important : Si votre type d'élément est un type d'élément composé de courrier électronique, n'entrez pas le type d'élément de pièce jointe. Les types d'élément de pièce jointe ne peuvent pas être traités seuls mais sont traités automatiquement comme partie du type d'élément de courrier électronique.

-database <base_de_données>

Obligatoire.

Nom de la base de données du serveur de bibliothèques où le type d'élément est stocké. Le nom de la base de données peut être indiqué dans deux formats qui déterminent comment l'outil de l'indexeur se connecte à la base de données. La connexion à la base de données est mise en oeuvre à l'aide de JDBC (Java Data Base Connectivity) qui définit comment un client Java peut accéder à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Pour utiliser une connexion JDBC de type 4, le nom de la base de données de l'argument `-database` doit inclure un nom de serveur, un numéro de port et un nom de base de données au format `<serveurbasededonnées>:<numérodeport>/<nombasededonnées>`, par exemple `-database localhost:50000/icm1sdb`. Il est

recommandé d'utiliser JDBC type 4 comme méthode d'accès à la base de données mais ce mécanisme requiert la spécification complète de l'ID utilisateur et le mot de passe.

Pour utiliser une connexion JDBC type 2, la spécification du nom de la base de données de l'argument `-database` doit correspondre à un nom de connexion DB2 existant. En d'autres termes, le nom de la base de données doit être catalogué dans DB2, par exemple `-database icm1sdb`. Comme JDBC type 2 utilise une connexion à une base de données existante, il est inutile d'indiquer un ID utilisateur et un mot de passe. Toutefois, une version 32 bits native du pilote JDBC type 2 DB2 (bibliothèque partagée) doit être disponible et définie dans le chemin d'accès aux classes pour l'utilisation de la connexion JDBC type 2.

-user *<ID_utilisateur|@variable_environnement|:nom_de_fichier>*
Facultatif

Indiquez un ID utilisateur. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques. Vous pouvez entrer un ID utilisateur de trois manières :

1. Indiquez l'ID utilisateur directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle l'ID utilisateur est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_USERID=user8
```



```
afuRecreateIndex -user @INDEXER_USERID
```
3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuRecreateIndex -user :myUserID.txt
```

Le fichier texte doit contenir une seule ligne avec l'ID utilisateur (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer l'ID utilisateur dans toutes les informations d'état des processus.

L'utilisateur doit disposer des droits nécessaires pour supprimer et mettre à jour les index de recherche de texte. Pour connaître les droits utilisateur requis pour supprimer et mettre à jour les index de recherche de texte, voir <http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html> pour DB2 Net Search Extender version 9.5 et <http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html> pour DB2 Net Search Extender version 9.7.

-password *<mot_de_passe|@variable_environnement|:nom_de_fichier>*
Facultatif

Mot de passe de l'ID utilisateur doté des droits administrateur de la base de données. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques. Vous pouvez entrer un mot de passe de trois manières :

1. Indiquez le mot de passe directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle le mot de passe est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_PASSWORD=12345678
```

```
afuRecreateIndex -password @INDEXER_PASSWORD
```

3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,
afuRecreateIndex -password :myPassword.txt

Le fichier texte doit contenir une seule ligne avec le mot de passe (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer le mot de passe dans toutes les informations d'état des processus.

Les mots de passe ne sont pas consignés dans un format de texte normal lisible.

-schema <*schéma*>

Facultatif.

Nom du schéma de la base de données s'il est différent du schéma de la base de données en cours ou que l'option de configuration **GenJdbcType** correspond à 4.

-indexdirectory <*répertoire*>

Facultatif.

Chemin d'accès absolu du répertoire sur le serveur de bibliothèques Content Manager sur lequel Net Search Extender doit créer l'index documentaire pour le type d'élément.

Permet la recréation de l'index dans un répertoire différent. Si cet argument n'est pas transmis, l'index est recréé dans le même répertoire que celui utilisé pour stocker l'ancien index.

-workdirectory <*répertoire*>

Facultatif.

Chemin d'accès absolu du répertoire de travail sur le serveur de bibliothèques Content Manager requis par Net Search Extender pour les fichiers d'index temporaires.

Permet d'indiquer un répertoire différent de celui que le processus d'index va utiliser comme répertoire de travail. Si cet argument n'est pas transmis, le répertoire de travail est identique au répertoire de l'index.

Remarque : Il est recommandé de placer le répertoire de l'index et le répertoire de travail sur le même volume.

-modelfile <*nom_fichier*>

Facultatif.

Permet de recréer l'index du type d'élément à l'aide du fichier modèle que vous avez modifié.

Si aucun fichier n'est spécifié, le modèle qui a été utilisé pour l'index Net Search Extender existant est utilisé pour recréer l'index. Cela signifie qu'une copie du fichier modèle utilisé pour créer l'index est générée avant de supprimer l'index et que cette copie est utilisée pour créer le nouvel index Net Search Extender.

Indiquez le fichier modèle avec un chemin d'accès absolu.

-tablespace <*espace_table*>

Facultatif.

Permet d'indiquer quel espace de table utiliser lors de la création de l'index. Si cet argument n'est pas transmis, l'espace de table DB2 par défaut est utilisé.

-useconfig <fichier.ini>;<fichier.ini>

Facultatif.

Permet d'utiliser d'autres fichiers de configuration .ini dans un autre chemin. Plusieurs fichiers .ini peuvent être indiqués. Ils sont chargés dans l'ordre dans lequel ils ont été indiqués.

-usettings "<option=valeur>;<option=valeur>"

Facultatif.

Active le remplacement direct et temporaire des valeurs de configuration de l'indexeur stockées dans le fichier .ini. Vous pouvez ajouter un nombre illimité de paires <option=valeur> séparées par le délimiteur de votre système d'exploitation. Les paires de valeurs d'option doivent être placées entre guillemets. Les options de configuration que vous répertoriez (<option>) ne sont pas sensibles à la casse. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de l'indexeur.

-exportconfig <fichier.ini>

Facultatif.

Exporte la configuration de l'indexeur dans le fichier .ini indiqué dans le fichier <fichier .ini>.

-listconfig

Facultatif.

Affiche toutes les options de configuration et les valeurs associées définies pour l'exécution de l'indexeur en cours.

-verbose

Facultatif.

Affiche des informations détaillées sur le traitement. L'utilisation de l'option **-verbose** est utile au début pour obtenir des informations détaillées sur l'ensemble des traitements. Si l'option **-verbose** est spécifiée, les fichiers journaux contiennent également des informations supplémentaires.

-showsqli

Facultatif.

Affiche toutes les commandes SQL DB2 exécutées par le système. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de tous les traitements système DB2. Si cet argument est indiqué, toutes les instructions SQL sont également ajoutées au fichier journal.

-trace

Facultatif.

Fournit des informations plus détaillées que l'option **-verbose** en affichant des données supplémentaires dans les fichiers journaux pour effectuer l'identification des erreurs. Ces informations sont destinées uniquement au service de support logiciel IBM et requièrent l'utilisation du fichier de propriété de trace.

-simulate

Facultatif.

Exécute le processus de création de l'index sans mettre à jour les bases de données. Net Search Extender n'est pas appelé. Utilisez ce paramètre pour valider la configuration d'index.

-quiet

Aucune étape de traitement n'est affichée sur la console.

Correction de la structure du fichier CSN dans des courriers électroniques Lotus Domino MIME

Tous les courriers électroniques Internet (documents MIME) avec des objets intégrés codés qui ont été archivés à partir d'un serveur de messagerie Lotus Domino à l'aide d'une version IBM Content Collector comprise entre les versions 2.1.1.1 LA006 et 2.1.1.3 LA006 n'ont pas été correctement archivés et la structure du fichier CSN qui a été enregistrée dans le référentiel n'est pas valide. Ces documents n'ont pas été indexés correctement. Ils doivent être retraités.

Pour identifier les documents qui peuvent être affectés et qui peuvent nécessiter une réindexation, utilisez l'outil `afuRepairCSN`. Cet outil doit être exécuté sur TOUS les types d'élément contenant des documents de courrier électronique Lotus Domino qui ont été archivés à l'aide d'une version IBM Content Collector Server comprise entre les versions V2.1.1.1 LA006 et V2.1.1.3 LA006.

`afuRepairCSN` tente de déterminer l'intervalle de temps pendant lequel les documents ont pu être affectés en demandant les dates d'application d'un groupe de correctifs spécifique. Chaque document du type d'élément pouvant être affecté est ajouté à la table des tâches ouvertes de l'indexeur avec un marqueur spécifique. Si vous exécutez l'outil de réparation en mode de traitement par lots, l'intervalle de temps doit être spécifié sur la ligne de commande car le traitement commence immédiatement sans demande de confirmation de l'utilisateur. En mode de traitement par lots, vous devez inclure l'option de ligne de commande `-RUNWITHOUTPROMPT` par mesure de précaution.

Lorsque `afuIndexer` s'exécute à nouveau, le processus de l'indexeur vérifie si l'un des documents de courrier électronique marqués a une structure de fichier CSN non valide. Pour cette vérification, l'indexeur télécharge le fichier CSN du gestionnaire de ressources IBM Content Manager.

Seulement si une structure de fichier non valide est identifiée, le document de courrier électronique est retraité et ajouté à l'index de texte intégral Net Search Extender.

Si le processus de l'indexeur ne peut pas identifier si la structure de fichier n'est pas valide, le document de courrier électronique n'est pas retraité et est supprimé de la table des tâches ouvertes de l'indexeur.

Si `afuIndexer` est exécuté en mode de traitement par lots, l'exécution de plusieurs lots peut être nécessaire pour traiter la table des tâches ouvertes en fonction du nombre d'éléments dans le type d'élément. Si `afuIndexer` est exécuté en mode continu, il traite les éléments de la table des tâches ouvertes pendant chaque itération jusqu'à ce que tous les éléments aient été traités.

L'archivage dans le type d'élément peut se poursuivre pendant l'exécution de l'outil de réparation. Toutefois, `afuRepairCSN` et `afuIndexer` ne peuvent pas s'exécuter en même temps sur le même type d'élément.

Pour démarrer l'outil de réparation pour un type d'élément :

- Sous UNIX et Linux, exécutez :
`afuRepairCSN.sh`

- Sous Windows :
afuRepairCSN.cmd

Remarque : Lorsque vous exécutez le processus afuRepairCSN :

- Tous les arguments de ligne de commande sont sensibles à la casse.
- Les arguments de ligne de commande peuvent apparaître dans n'importe quel ordre.
- Utilisez les délimiteurs suivants pour les arguments qui peuvent transmettre plusieurs valeurs :
 - Sous Windows, un point-virgule (;)
 - Sous UNIX et Linux, deux-points (:)
- Le nom du type d'élément ne fait pas de distinction entre les majuscules et les minuscules.
- Les options de configuration de l'indexeur transmises via un argument ne sont pas sensibles à la casse.

Le type d'élément et le nom de la base de données du serveur de bibliothèques dans laquelle le type d'élément est stocké doit être transmis sous la forme d'arguments obligatoires au processus d'indexation.

Si vous exécutez afuRepairCSN sans indiquer d'argument, ou si vous exécutez afuRepairCSN avec l'argument `-?`, les éléments suivants sont affichés :

```
afuRepairCSN : Syntaxe :
[-?|-h|-copyright]
-it <itemtype>
-database <database>
[-user <userID|@environmentVariable|:filename>]
[-password <password|@environmentVariable|:filename>]
[-schema <schema>]
[-useconfig <fichier.ini>;<...>]
[-usesettings "<option=value>;<...>"]
[-exportconfig <fichier.ini>]
[-listconfig]
[-begin <yyyy-mm-dd> -end <yyyy-mm-dd> -RUNWITHOUTPROMPT]
[-verbose] [-showsql] [-trace]
[-simulate]
[-quiet]
```

où

-? Répertorie tous les arguments que vous pouvez transmettre à l'outil de l'indexeur afuRepairCSN.

-h Répertorie tous les arguments que vous pouvez transmettre à l'outil de l'indexeur afuRepairCSN.

-copyright

Affiche la notice de copyright et les informations de version.

-it <typed'élément>
Obligatoire.

Type d'élément Content Manager (collection de documents). Les noms de type d'élément des modèles de données intégré et composé ne font pas de distinction entre les majuscules et les minuscules.

Important : Si votre type d'élément est un type d'élément composé de courrier électronique, n'entrez pas le type d'élément de pièce jointe. Les types d'élément

de pièce jointe ne peuvent pas être traités seuls mais sont traités automatiquement comme partie du type d'élément de courrier électronique.

-database <basededonnées>

Obligatoire.

Nom de la base de données du serveur de bibliothèques où le type d'élément est stocké. Le nom de la base de données peut être indiqué dans deux formats qui déterminent comment l'outil de l'indexeur se connecte à la base de données. La connexion à la base de données est mise en oeuvre à l'aide de JDBC (Java Data Base Connectivity) qui définit comment un client Java peut accéder à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Pour utiliser une connexion JDBC de type 4, le nom de la base de données de l'argument -database doit inclure un nom de serveur, un numéro de port et un nom de base de données au format <serveurbasededonnées>:<numérodeport>/<nombasededonnées>, par exemple -database localhost:50000/icm1sdb. Il est recommandé d'utiliser JDBC type 4 comme méthode d'accès à la base de données mais ce mécanisme requiert la spécification complète de l'ID utilisateur et le mot de passe.

Pour utiliser une connexion JDBC type 2, la spécification du nom de la base de données de l'argument -database doit correspondre à un nom de connexion DB2 existant. En d'autres termes, le nom de la base de données doit être catalogué dans DB2, par exemple -database icm1sdb. Comme JDBC type 2 utilise une connexion à une base de données existante, il est inutile d'indiquer un ID utilisateur et un mot de passe. Toutefois, une version 32 bits native du pilote JDBC type 2 DB2 (bibliothèque partagée) doit être disponible et définie dans le chemin d'accès aux classes pour l'utilisation de la connexion JDBC type 2.

-user <userID|@environmentVariable|:filename>

Facultatif.

Indiquez un ID utilisateur. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un ID utilisateur de trois manières :

1. Indiquez l'ID utilisateur directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle l'ID utilisateur est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,
export INDEXER_USERID=user8
afuRepairCSN -user @INDEXER_USERID
3. Indiquez le nom d'un fichier texte qui contient l'ID utilisateur, par exemple,
afuRepairCSN -user :myUserID.txt

Le fichier texte doit contenir une seule ligne avec l'ID utilisateur (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer l'ID utilisateur dans toutes les informations d'état des processus.

L'utilisateur doit disposer des droits nécessaires pour mettre à jour les index de recherche de texte. Pour connaître les droits utilisateur requis pour mettre à jour les index de recherche de texte, voir <http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=/>

com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html pour DB2 Net Search Extender version 9.5 et <http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp?topic=/com.ibm.db2.luw.admin.nse.topics.doc/doc/c0052101.html> pour DB2 Net Search Extender version 9.7.

-password *<password|@environmentVariable|:filename>*
Facultatif.

Mot de passe de l'ID utilisateur doté des droits administrateur de la base de données. Obligatoire uniquement s'il n'existe pas de connexion à la base de données du serveur de bibliothèques ou qu'une connexion JDBC type 4 est utilisée pour établir une connexion. Vous pouvez entrer un mot de passe de trois manières :

1. Indiquez le mot de passe directement sur la ligne de commande.
2. Indiquez le nom d'une variable d'environnement à partir de laquelle le mot de passe est lu. Le nom de la variable d'environnement doit posséder le préfixe arobas (@), par exemple,

```
export INDEXER_PASSWORD=12345678
```

```
afuRepairCSN -password @INDEXER_PASSWORD
```

3. Indiquez le nom d'un fichier texte contenant le mot de passe ; par exemple,

```
afuRepairCSN -password :myPassword.txt
```

Le fichier texte doit contenir une seule ligne avec le mot de passe (sans guillemets ni espaces).

L'utilisation d'une variable d'environnement ou d'un fichier présente l'avantage de masquer le mot de passe dans toutes les informations d'état des processus.

Les mots de passe ne sont pas consignés dans un format de texte normal lisible.

-schema *<schéma>*
Facultatif.

Nom du schéma de la base de données s'il est différent du schéma de la base de données en cours.

-useconfig *<.ini file>;<.ini file>*
Facultatif.

Permet d'utiliser d'autres fichiers de configuration .ini dans un autre chemin. Plusieurs fichiers .ini peuvent être indiqués. Ils sont chargés dans l'ordre dans lequel ils ont été indiqués.

-usettings *"<option=value>;<option=value>"*
Facultatif.

Active le remplacement direct et temporaire des valeurs de configuration persistantes de l'indexeur stockées dans le fichier .ini. Vous pouvez ajouter un certain nombre de paires *<option=valeur>* séparées par le délimiteur du système d'exploitation. Les paires de valeurs d'option doivent être placées entre guillemets. Les options de configuration que vous répertoriez (*<option>*) ne sont pas sensibles à la casse. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de l'indexeur.

-exportconfig *<fichier.ini>*
Facultatif.

Exporte la configuration de l'indexeur dans le fichier .ini indiqué dans le fichier *<fichier.ini>*.

-listconfig

Facultatif.

Affiche toutes les options de configuration et les valeurs associées définies pour la procédure d'exécution de l'indexeur en cours. Aucune autre opération n'est effectuée.

-begin <yyyy-mm-dd>

Facultatif.

Définit la fenêtre de temps de début des éléments à vérifier. La valeur par défaut consiste à vérifier tous les éléments du type d'élément à partir de 2000-01-01.

-end <yyyy-mm-dd>

Facultatif.

Définit la fenêtre de temps de fin des éléments à vérifier. La valeur par défaut consiste à vérifier tous les éléments du type d'élément jusqu'à la date actuelle.

-RUNWITHOUTPROMPT

Facultatif.

L'outil n'affiche pas d'invite de confirmation pour vous aider à identifier les documents de courrier électronique du type d'élément pouvant contenir une structure non valide.

-verbose

Facultatif.

Affiche des informations détaillées sur le traitement. La définition de l'option -verbose est utile au début pour obtenir des informations détaillées sur l'ensemble des traitements. Si -verbose est indiqué, les fichiers journaux contiennent également des informations supplémentaires.

-showsqli

Facultatif.

Affiche toutes les commandes SQL DB2 exécutées par le système. L'utilisation de cet argument est utile au début pour vérifier la configuration et l'exécution de tous les traitements système DB2. Si cet argument est indiqué, toutes les instructions SQL sont également ajoutées au fichier journal.

-trace

Facultatif.

Fournit des informations plus détaillées que l'option -verbose en affichant des données supplémentaires dans les fichiers journaux pour effectuer l'identification des erreurs. Cette opération est obligatoire pour le service de support logiciel IBM uniquement et requiert l'utilisation du fichier des propriétés de la fonction de trace.

Important : Ne spécifiez pas l'argument de ligne de commande -trace en association avec -shutdown.

-simulate

Facultatif.

Exécute le processus de l'indexeur sans mettre à jour les bases de données. Net Search Extender ne s'exécute pas. Cet argument est utile pour vérifier le traitement des fichiers et la configuration.

-quiet

Aucune étape de traitement n'est affichée sur la console.

Sauvegarde des fichiers d'index

Les fichiers d'index peuvent être altérés pendant les étapes finales de fusion ou de réorganisation du processus d'indexation si les processus db2text ou cteprupx sont arrêtés brutalement.

L'index peut également être altéré si des fichiers ne peuvent pas être générés en raison d'une quantité d'espace disque insuffisante sur le système de fichiers. Il est donc indispensable de mettre au point une stratégie pour restaurer une sauvegarde donnée, puis mettre à jour cette sauvegarde avec les documents indexés avant la sauvegarde.

Si vous ne sauvegardez pas l'index, vous devrez recréer l'index, ce qui est une tâche coûteuse et lourde.

Le traitement de la sauvegarde peut être inclus dans le traitement de l'indexeur à l'aide des options de configuration GenCommandXXX. Cette procédure peut être utile si les étapes de traitement de la sauvegarde définies doivent s'exécuter uniquement dans des données Net Search Extender et que la sauvegarde des données d'index ne fait pas partie d'une solution de sauvegarde sur disque exécutée à intervalles réguliers.

A l'aide de ces options, vous pouvez ajouter des étapes de traitement définies par l'utilisateur directement avant l'indexation ou le début de la réorganisation de l'index ou immédiatement après l'indexation ou la réorganisation.

Si vous rencontrez des problèmes lors de la sauvegarde des fichiers d'index et que vous souhaitez vérifier si les fichiers d'index sont cohérents, vous pouvez utiliser l'outil afuIndexTool pour valider l'index. Pour plus de détails, consultez la rubrique «Validation d'index et opérations de réparation», à la page 116.

Présentation de la table de base de données pour les tables de tâches de l'indexeur

Il y a deux tables de tâches de l'indexeur de base de données, une pour les tâches ouvertes, l'autre pour les tâches terminées.

Les deux tables sont créées pour chaque type d'élément. Le contenu de ces tables de base de données peut être utilisé pour référence, en plus des arguments de ligne de commande et des options de configuration disponibles avec le processus de l'indexeur.

La table des tâches ouvertes contient tous les éléments qui n'ont pas été traités et indexés. Le nom de la table de base de données des tâches ouvertes est composé de AFU1FTIOPEN<item type id> où <item type id> correspond à l'ID du type d'élément numérique. Si nécessaire, entrez des zéros comme préfixe pour que l'ID ait 5 chiffres.

Le nom se présente sous la forme suivante :

Column name	Type schema	Type name	Length	Scale	Nulls
ITEMID	SYSIBM	CHARACTER	26	0	No
VERSIONID	SYSIBM	SMALLINT	2	0	No
ADDDATE	SYSIBM	TIMESTAMP	10	0	Yes
OP	SYSIBM	SMALLINT	2	0	Yes
IDXSTATUS	SYSIBM	SMALLINT	2	0	Yes

Le nom de la table de base de données des tâches terminées est composé de AFU1FTICOMP<item type id> où <item type id> correspond également à l'ID du type d'élément numérique et doit avoir 5 chiffres. La table de base de données des tâches terminées contient tous les éléments qui ont été indexés ou ignorés. Le nom se présente sous la forme suivante :

Column name	Type schema	Type name	Length	Scale	Nulls
ITEMID	SYSIBM	CHARACTER	26	0	No
VERSIONID	SYSIBM	SMALLINT	2	0	No
ADDDATE	SYSIBM	TIMESTAMP	10	0	Yes
IDXDATE	SYSIBM	TIMESTAMP	10	0	Yes
IDXRC	SYSIBM	SMALLINT	2	0	Yes

Pour obtenir une liste des valeurs IDXRC disponibles, voir «Codes retour de l'index stockés dans la tables des tâches terminées», à la page 37.

Suppression d'un type d'élément

Avant de supprimer un type d'élément existant, exécutez l'outil de vérification de l'indexeur pour déterminer les noms de table de base de données dont vous avez besoin pour supprimer les tables de base de données et l'index documentaire.

Pour supprimer un type d'élément, procédez comme suit.

- Exécutez l'outil afuCheck avec l'argument `-verbose`. Ce mode affiche les données suivantes :
 - Le nom de la table de base de données pour les tâches ouvertes (sous **Table of open tasks**)
 - Le nom de la table de base de données pour les tâches terminées (sous **Table of complete tasks**)
 - Le nom de l'index documentaire Net Search Extender (sous **NSE Properties > nom d'index NSE**)
- Supprimez la table des tâches ouvertes pour le type d'élément à l'aide de la commande `DB2 DROP TABLE` et du nom de la table des tâches ouvertes.
- Supprimez la table des tâches terminées pour le type d'élément à l'aide de la commande `DB2 DROP TABLE` et du nom de la table des tâches terminées.
- Supprimez l'index documentaire Net Search Extender à l'aide de la commande `db2text drop index`, comme indiqué dans la documentation de Net Search Extender.

Messages d'erreur de l'indexeur de recherche de texte

La liste ci-après répertorie les messages d'erreur émis par l'indexeur, ainsi qu'une description de l'action à effectuer.

1 Une instance de l'indexeur est déjà en cours d'exécution pour ce type d'élément ; ce qu'indique le fichier "{0}". Impossible de démarrer une seconde instance.

Action utilisateur : Cette situation peut se produire si le processus de l'indexeur a été arrêté de manière inattendue ou a été interrompu.

Vérifiez que le processus de l'indexeur n'est plus en cours d'exécution. Une fois le processus arrêté, supprimez le fichier indiqué et relancez l'indexeur.

2 Un problème de communication s'est produit : le processus de l'indexeur ne peut pas communiquer avec le processus agent.

Action utilisateur : Sélectionnez un autre numéro de port en indiquant une autre valeur pour la variable d'environnement `AFU_INDEXER_RMI_PORT` ou en utilisant l'argument du processus de l'indexeur `-port`. Vérifiez que le numéro de port est libre et qu'il n'est pas utilisé par une autre instance du processus de l'indexeur ou une autre application.

3 Le fichier de configuration Net Search Extender actuel ne contient pas les valeurs de configuration recommandées.

Action utilisateur : Il est conseillé de configurer Net Search Extender à l'aide des paramètres recommandés. Pour plus d'informations sur les paramètres de configuration et les valeurs qu'ils nécessitent, voir «Configuration Net Search Extender incorrecte», à la page 150. Pour utiliser de nouveaux paramètres, l'index Net Search Extender doit être recréé à l'aide de l'outil afuRecreate.

4 Le type d'élément ne peut pas être activé car il contient toujours des éléments qui n'ont pas été indexés par Net Search Extender. La table de journal Net Search Extender doit être vide pour que le type d'élément puisse être activé pour une utilisation avec l'indexeur.

Action utilisateur : Désactivez l'archivage du type d'élément, puis indexez les éléments qui n'ont pas été traités à l'aide de l'indexeur rapide ou de l'indexeur Content Manager standard associé à IBM Text Search. Le type d'élément peut ensuite être activé afin d'être traité par l'indexeur de recherche de texte.

5 Impossible de demander le service d'indexation de texte Net Search Extender.

Action utilisateur : Vérifiez l'installation de Net Search Extender. Les commandes Net Search Extender doivent être disponibles et définies dans la variable PATH et disposer des droits nécessaires pour s'exécuter.

Vérifiez que vous êtes connecté en tant que propriétaire de l'instance et que vous disposez des autorisations nécessaires pour exécuter la commande db2text dans la session actuelle.

6 Impossible d'identifier la fonction définie par l'utilisateur Net Search Extender enregistrée pour ce type d'élément. Vérifiez l'installation de la fonction Net Search Extender définie par l'utilisateur.

Action utilisateur : Vérifiez que la fonction définie par l'utilisateur AFUFetchFile a été copiée dans le sous-répertoire des fonctions DB2.

7 Impossible de déterminer la section [{0}] dans le fichier de configuration Net Search Extender "{1}".

Action utilisateur : Vérifiez que le fichier de configuration de l'index Net Search Extender est valide conformément à la documentation du composant Net Search Extender.

8 La valeur de l'option de configuration de l'index Net Search Extender {0} n'est pas la valeur recommandée pour une indexation correcte (valeur actuelle : {1}). L'index Net Search Extender doit être recréé à l'aide d'une configuration d'index dans laquelle la valeur de l'option de configuration {0} est définie sur {2}

Action utilisateur : Vous devez remplacer la valeur de l'option de configuration de l'index par la valeur recommandée dans le fichier modèle de configuration de l'index Net Search Extender. Le fichier modèle de configuration de l'index Net Search Extender s'appelle cteixcfg.ini et se trouve dans le répertoire d'installation de Net Search Extender (généralement .../sql11ib/db2ext).

Pour plus d'informations sur les paramètres de configuration et les valeurs qu'ils nécessitent, voir «Configuration Net Search Extender incorrecte», à la page 150.

L'index Net Search Extender doit être recréé à l'aide de l'outil afuRecreateIndex.

Le modèle de configuration de l'index Net Search Extender est automatiquement copié du répertoire d'installation de Net Search Extender dans le nouveau répertoire d'index texte pendant la procédure de recréation de l'index de texte.

9 L'option de configuration d'index Net Search Extender {0} n'est pas définie. L'index Net Search Extender doit être recréé à l'aide d'une configuration d'index dans laquelle l'option de configuration {0} est définie et configurée sur {1}.

Action utilisateur : Vous devez remplacer la valeur de l'option de configuration de l'index par la valeur recommandée dans le fichier modèle de configuration de l'index Net Search Extender. Le fichier modèle de configuration de l'index Net Search Extender s'appelle cteixcfg.ini et se trouve dans le répertoire d'installation de Net Search Extender (généralement .../sql11ib/db2ext). L'index Net Search Extender doit être recréé à l'aide de l'outil afuRecreateIndex.

Le modèle de configuration de l'index Net Search Extender est automatiquement copié du répertoire d'installation de Net Search Extender dans le nouveau répertoire d'index texte pendant la procédure de recréation de l'index de texte.

10 Impossible de charger le fichier de configuration Net Search Extender "{0}" pour une vérification.

Action utilisateur : Vérifiez les droits d'accès au fichier et déterminez si le fichier est disponible.

11 Impossible de démarrer le service d'indexation de texte Net Search Extender.

Action utilisateur : Vérifiez l'installation de Net Search Extender. Les commandes Net Search Extender doivent être disponibles et définies dans la variable PATH.

Vérifiez que vous êtes connecté en tant que propriétaire de l'instance et que vous disposez des autorisations nécessaires pour exécuter la commande db2text dans la session actuelle.

12 Impossible de déterminer le statut de Net Search Extender en raison d'une sortie inattendue : "{0}".

Action utilisateur : Vérifiez l'installation de Net Search Extender. Les commandes Net Search Extender doivent être disponibles et définies dans la variable PATH.

Vérifiez que vous êtes connecté en tant que propriétaire de l'instance et que vous disposez des autorisations nécessaires pour exécuter la commande db2text dans la session actuelle.

13 Le processus Net Search Extender n'a pas produit de messages d'état ("{0}").

Action utilisateur : Vérifiez l'installation de Net Search Extender. Les commandes Net Search Extender doivent être disponibles et définies dans la variable PATH.

14 Impossible de démarrer le moteur de recherche Net Search Extender : Raison : {0}

Action utilisateur : Vérifiez l'installation de Net Search Extender. Les commandes Net Search Extender doivent être disponibles et définies dans la variable PATH.

Vérifiez que vous êtes connecté en tant que propriétaire de l'instance et que vous disposez des autorisations nécessaires pour exécuter la commande db2text dans la session actuelle.

15 Impossible de charger le fichier modèle Net Search Extender "{0}".

Action utilisateur : Vérifiez les droits d'accès au fichier et déterminez si le fichier est disponible et valide conformément à la documentation du composant Net Search Extender. Si cela ne résout pas le problème, recréez l'index.

16 Impossible d'analyser le fichier modèle Net Search Extender "{0}".

Action utilisateur : Vérifiez les droits d'accès au fichier et déterminez si le fichier est disponible et valide conformément à la documentation du composant Net Search Extender.

17 Impossible de localiser le répertoire d'installation DB2 : la variable d'environnement "{0}" n'est pas définie.

Action utilisateur : Associez la variable d'environnement à la valeur appropriée et réexécutez la commande.

18 Impossible de localiser le fichier de configuration Net Search Extender ("{0}") dans le répertoire d'installation DB2 ("{1}") défini dans la variable d'environnement "{2}".

Action utilisateur : Vérifiez que la variable d'environnement indiquée correspond au répertoire d'installation de DB2. Le fichier de configuration Net Search Extender est recherché dans tous les sous-répertoires du répertoire indiqué.

19 Aucun index de recherche de texte Net Search Extender trouvé pour le type d'élément indiqué ("{0}"). Ce type d'élément ne peut pas être traité.

Action utilisateur : Vérifiez que le type d'élément approprié a été indiqué lors de l'appel du processus de l'indexeur et qu'il a été activé afin d'être traité par l'indexeur de recherche de texte.

20 Une réorganisation de l'index Net Search Extender aurait du être exécutée ! Le nombre d'éléments mis à jour depuis la dernière réorganisation d'index Net Search Extender ({0}) dépasse le nombre indiqué dans l'option configuration "{1}". Plus aucune mise à jour d'index n'est possible jusqu'à l'exécution d'une réorganisation d'index.

Action utilisateur : Le processus de l'indexeur ne peut pas être réexécuté sans l'argument de ligne de commande -nseindexreorg et la taille de lot 1.

21 Un élément sans valeur a été trouvé pour un attribut Content Manager défini.

Action utilisateur : Si l'un des attributs suivants est défini pour un type d'élément, le processus d'indexeur attend une valeur pour l'attribut dans chaque élément qu'il traite :

- ICCMailboxID
- CSORGINATOR (lors de l'accès à un élément de ressource intégré CommonStore)
- CSLDOrigDB (lors de l'accès à un élément de ressource intégré CommonStore)

Vérifiez votre configuration de route de tâches. Si ces valeurs sont manquantes, la route de tâches ne parvient pas à extraire les valeurs de ces attributs.

22 Impossible de démarrer le processus afuIndexer.

Explication : Le message d'erreur suivant s'affiche lorsque le processus d'indexeur afuIndexer est démarré :

```
afuIndexer: An error occurred:
A configuration or setup problem occurred:
The version of the Net Search Extender UDF
registered for this item type is not
supported.
Version '[V2]' is required.
Verify that the installation of the Net
Search Extender UDF AFUFetchFile is correct
```

Ce message indique qu'une ancienne version de la fonction définie par l'utilisateur (UDF) AFUFetchFile est toujours installée sur votre système. Cela peut se produire si des problèmes sont rencontrés lorsque Content Collector Text Search Support est installé.

Action utilisateur : Vous devez mettre à jour la fonction définie par l'utilisateur (UDF) AFUFetchFile manuellement :

1. Arrêtez DB2.
2. Copiez le fichier AfuFetchFile.dll du sous-répertoire lib du répertoire d'installation de l'indexeur (par défaut : /opt/IBM/ICCIndexer/lib) vers le sous-répertoire fonction du répertoire d'instance DB2 (par défaut : \$DB2HOME/fonction).
3. Redémarrez DB2.
4. Redémarrez le processus afuIndexer.

23 Un élément de la table des tâches ouvertes a un statut inattendue de 1 (IN_PROGRESS).

Explication : Cet élément a entraîné l'arrêt du processus de l'indexeur de manière inattendue lors de l'exécution précédente de l'indexeur. L'élément n'est pas retiré. Au lieu de cela, il est déplacé dans la table des tâches terminées avec le code IDXRC 200 (erreur grave). Lors de l'exécution suivante du processus de l'indexeur, il détecte cet élément et le considère comme étant la cause de l'arrêt du processus de travail.

Action utilisateur : La manière dont les éléments avec une erreur grave sont traités par l'indexeur est influencée par la valeur de l'option de configuration **IdxProcessSevereErrors**. Pour plus de détails sur l'utilisation de l'option **IdxProcessSevereErrors**, voir «Éléments ayant généré une erreur grave», à la page 106.

24 Un type d'élément spécifié ne peut pas être identifié correctement par le processus de l'indexeur.

Explication : Le message d'erreur suivant s'affiche lorsque le processus d'indexeur afuIndexer est démarré :

```
The AFU1FTIOPENxxxxx database table is required
but does not exist.
Check if the item type has been
enabled for processing by the indexer.
```

Ce message indique que le type d'élément n'est pas reconnu par l'indexeur pour la recherche de texte.

Action utilisateur : Exécutez l'outil afuEnableItemType pour le type d'élément spécifié afin d'activer le type d'élément pour l'indexation dans IBM Content Collector, en d'autres termes, pour créer les tables des tâches ouvertes et la table des tâches terminées pour le type d'élément.

Identification et résolution des problèmes de l'indexeur de recherche de texte

Les rubriques ci-après vous aident à résoudre les problèmes courants qui peuvent se produire lors de l'utilisation de la prise en charge de la recherche de texte dans IBM Content Manager.

Configuration Net Search Extender incorrecte

La prise en charge de la recherche de texte IBM Content Collector exige que tous les index documentaires gérés par Net Search Extender soient créés avec des paramètres de configuration d'index spécifiques. Si ces paramètres ne sont pas correctement configurés, il se peut que l'ensemble de résultats de la recherche de texte ne soit pas complet ou qu'il soit vide.

Symptômes

Des paramètres incorrects peuvent également conduire à une dégradation significative des performances ainsi qu'à une perte des données. Il est donc essentiel que la configuration de l'index soit correcte pour chaque index documentaire Net Search Extender de chaque type d'élément.

Les options de configuration suivantes sont essentielles car elles contrôlent la manière dont le texte est indexé et si ce texte peut à nouveau être trouvé :

- **RESPECTCASE**
- **TREATNUMBERSASWORDS**
- **BLOCKMODE**

Ces valeurs sont contrôlées automatiquement par l'indexeur pour la recherche de texte. Si elles ne sont pas correctement configurées dans la configuration de l'index documentaire Net Search Extender, l'indexeur émet un ou plusieurs messages d'erreur avec la clause de protection suivante :

```
A configuration or setup problem occurred:  
The Net Search Extender index configuration is not as recommended and cannot be  
used with IBM Content Collector.  
The Net Search Extender index must be re-created using a corrected  
Net Search Extender index configuration.  
Modify the Net Search Extender index configuration template file 'cteixcfg.ini'  
in the Net Search Extender installation directory and then re-create the index.  
Net Search Extender will then copy the index configuration template file to the  
new index.
```

Causes

Vous ne pouvez pas modifier la configuration de l'index Net Search Extender une fois l'index documentaire créé. De même, Content Collector ne peut pas corriger les paramètres de configuration de l'index Net Search Extender tant que l'indexeur ne les a pas utilisés.

Cette restriction existe car, lorsqu'un type d'élément est créé à l'aide de l'assistant de configuration initial IBM Content Collector ou de l'outil de configuration IBM Content Collector, il est créé sur le serveur IBM Content Collector et non sur la machine sur laquelle IBM Content Manager et les applications Net Search Extender sont généralement exécutés.

Cela signifie que les outils servant à la création du type d'élément ne peuvent pas accéder au fichier de modèle de configuration appelé `cteixcfg.ini`. Les paramètres de configuration peuvent uniquement être vérifiés une fois l'index documentaire

créé et l'indexeur exécuté sur le type d'élément. A ce point du processus, il est cependant trop tard pour modifier un paramètre car l'index est déjà créé en fonction des paramètres dans l'exemple de fichier de configuration.

Résolution de l'incident

Si un index documentaire Net Search Extender est créé avec des paramètres de configuration incorrects, l'index complet doit être recréé et régénéré. Pour corriger ce problème, vous devez procéder comme suit.

1. Modifiez les valeurs des options de configuration de l'index dans le fichier de modèle de configuration de l'index Net Search Extender pour les valeurs recommandées. Le fichier de modèle de configuration de l'index Net Search Extender appelé `cteixcfg.ini` se trouve dans le répertoire d'installation de Net Search Extender (généralement `.../sql1lib/db2ext`).

2. Recréez l'index Net Search Extender à l'aide de l'outil `afuRecreateIndex`.

Tous les éléments du type d'élément sont à nouveau traités à la prochaine exécution de `afuIndexer`. Selon la taille du type d'élément, le nouveau traitement de tous les éléments peut prendre un certain temps.

L'outil de l'indexeur ne parvient pas à traiter le type d'élément

Les outils de l'indexeur (`afuIndexer`, `afuCreateIndex`, `afuConfig` ou `afuCheck`) ne parviennent pas à traiter le type d'élément transmis.

Symptômes

L'erreur suivante est rapportée :

A configuration or setup error has occurred with the following reason:

The required database table 'AFU1FTIOPENxxx' does not exist. Check if this item type has been enabled for processing by the indexer

Causes

Cette erreur est due au fait que l'outil de l'indexeur a été appelé avec un type d'élément qui n'est pas activé pour le traitement par l'outil.

Résolution de l'incident

Vérifiez que le type d'élément entré est correct et qu'il a été activé pour le traitement par l'indexeur de recherche de texte.

Si votre type d'élément est un type d'élément composé de courrier électronique, n'entrez pas le nom du type d'élément de pièce jointe. Les types d'élément de pièce jointe ne peuvent pas être traités seuls mais sont traités automatiquement comme partie du type d'élément de courrier électronique.

Impossible d'activer les types d'élément de courrier électronique IBM CommonStore via `afuEnableItemType`

L'indexation des types d'élément pourvus d'un modèle de données de document ne peut pas être activée à l'aide de l'outil d'indexation `afuEnableItemType` d'IBM Content Collector.

Symptômes

Si vous activez un type d'élément IBM Content Manager avec un modèle de données de document via `afuEnableItemType`, le message d'erreur ci-après s'affiche.

The data model of the 'xxx' item type is of type 'UNKNOWN'.

This data model is not supported by the indexer.

The 'xxx' item type cannot be enabled to be used by the indexer

Causes

Lorsqu'un type d'élément est créé manuellement afin d'être utilisé dans IBM CommonStore, les attributs spécifiques, tels que CSMailboxid, ne sont pas obligatoires dans la définition du type d'élément. Par conséquent, lorsque ces types d'élément ne sont pas utilisés dans IBM Content Collector, l'indexeur n'est pas en mesure de déterminer si les types d'élément prennent en charge ou non l'indexation et il arrête le traitement immédiatement.

Résolution de l'incident

Afin de traiter ces types d'élément, la variable d'environnement **AFU_INDEXER_USE_IT_CS_GENERIC** doit être définie pour l'indexeur de tous les outils de recherche de texte. Si la variable d'environnement **AFU_INDEXER_USE_IT_CS_GENERIC** est définie sur 1 ou true, l'indexeur des outils de recherche de texte reconnaît le type d'élément comme étant un type d'élément de courrier électronique IBM CommonStore générique et permet son traitement.

Action de l'utilisateur : Pour plus de détails sur l'utilisation de la variable d'environnement **AFU_INDEXER_USE_IT_CS_GENERIC**, voir «Variables d'environnement de l'indexeur de texte», à la page 63.

Impossible de charger une bibliothèque DB2 native

La bibliothèque DB2 native ne peut pas être chargée.

Symptômes

Le bibliothèque DB2 db2jcc2 ne peut pas être chargée en raison d'une incompatibilité de niveau entre DB2 et le support de recherche de texte IBM Content Collector. L'erreur suivante est rapportée :

```
Connecting to database.....: icmnlbdb
com.ibm.db2.jcc.a.SqlException: [jcc][10389][12245][3.52.95] Failure in loading
native library db2jcc2, java.lang.UnsatisfiedLinkError: db2jcc2 (A file or dir
ectory in the path name does not exist.):  ERRORCODE=-4472, SQLSTATE=null
    at com.ibm.db2.jcc.a.ad.a(ad.java:650)
```

...

```
at com.ibm.afu.ftsidx.createconfig.Main.main(Main.java:101)
```

```
-----
afuConfig: An error occurred:
An SQL problem occurred:
[jcc][10389][12245][3.52.95] Failure in loading native library
db2jcc2, java.lang.UnsatisfiedLinkError: db2jcc2 (A file or
directory in the path name does not exist.):  ERRORCODE=-4472,
SQLSTATE=null
```

Causes

Une non-concordance de niveau binaire existe entre Content Collector DB2 et Text Search Support sur le serveur. De plus, les paramètres LIBPATH sont peut-être incorrects.

Résolution de l'incident

Le support de recherche de texte Content Collector est configuré en mode 32 bits. La connexion à DB2 doit correspondre à ce niveau binaire. En l'absence de correspondance, vous recevez un message d'erreur lors du chargement de la bibliothèque db2jcc2. Ce problème peut être résolu de deux façons.

1. Recommandé : installez la version 32 bits de DB2. Du fait que Content Collector Text Search Support est une applications 32 bits, elle s'aligne correctement avec la version 32 bits de DB2 et vous ne rencontrerez plus cette erreur.
2. Utilisez un pilote JDBC de type 4 et faites pointer le support de recherche de texte vers les bibliothèques appropriées. Consultez le lien suivant pour plus d'informations sur l'utilisation d'une installation 32 bits de DB2 sur un système d'exploitation 64 bits : Nouvelle prise en charge 32 bits et 64 bits des serveurs DB2.
3. Vérifiez les paramètres de la liste de contrôle d'accès (LIBPATH). Les informations de chemin des bibliothèques 32 bits doivent précéder les informations de chemin des bibliothèques 64 bits. Par exemple :

```
LIBPATH=/opt/IBM/db2cmv8/lib:/usr/lib:/lib:/home/db2inst1/sqllib/lib32:/home/db2inst1/sqllib/lib64:/home/db2inst1/sqllib/lib:...
```

Impossible de localiser le répertoire d'installation DB2

Symptômes

Lorsque l'outil de l'indexeur `afuRecreateIndex` est exécuté, le message d'erreur suivant est rapporté :

```
The DB2 installation directory could not be located: The environment variable DB2HOME n'est pas défini.
```

Causes

La variable d'environnement `DB2HOME` n'était probablement pas définie.

Résolution de l'incident

Pour exécuter l'outil d'indexation `afuRecreateIndex`, vous devez définir la variable d'environnement `DB2HOME` si ce n'est pas déjà fait.

Cette variable d'environnement est utilisée par `afuRecreateIndex` pour déterminer les paramètres de configuration du modèle Net Search Extender et doit pointer vers le répertoire d'installation DB2 (par exemple, `C:\Program Files\IBM\sqllib` sous Windows ou `/home/db2inst1/sqllib` sous UNIX).

Vous devez définir cette variable d'environnement de manière permanente sur toutes les plateformes.

Escalade de verrous DB2

DB2 procède à l'escalade de verrous interne afin de réduire au minimum le nombre de verrous.

Symptômes

DB2 signale qu'il effectue une escalade de verrous au cours des opérations d'indexation sur la table de base de données des tâches terminées ("`AFU1FTICOMPxxxx`") ainsi que sur celle des tâches ouvertes ("`AFU1FTIOPENxxxx`").

Causes

La cause de cette erreur est que l'option de configuration de l'indexeur appelée `GenDatabaseTransactionSize` a été définie sur une valeur trop élevée.

Résolution de l'incident

Réduisez la valeur de l'option **GenDatabaseTransactionSize** en la définissant, par exemple, sur 50 000 ou 20 000.

Index introuvable après sa recréation

La recréation d'un index avec le même nom et au même emplacement que l'index supprimé provoque une erreur qui indique que l'index est introuvable.

Symptômes

Si vous envoyez une demande à un index Net Search Extender à l'aide de CONTAINS et qu'elle aboutit, que vous supprimez l'index, que vous le recréez avec le même nom d'index et que vous envoyez à nouveau cette demande, l'erreur suivante est rapportée :

```
CTE0198 No corresponding text index
```

Causes

Cette erreur est due au fait que l'ancien emplacement de l'index est toujours en mémoire cache. DB2 conserve les résultats de toutes les sous-requêtes internes, notamment l'emplacement de l'index, dans sa mémoire cache de requête pour éviter d'avoir à régénérer et exécuter des analyses potentiellement longues.

Résolution de l'incident

Nettoyez la mémoire cache de requête à l'aide de la commande suivante :

```
db2 flush package cache dynamic
```

Net Search Extender s'arrête de manière inattendue à l'émission de COMMIT

Le processus de l'indexeur s'arrête de manière inattendue car Net Search Extender a détecté une erreur fatale.

Symptômes

Si la fonction définie par l'utilisateur AFUFetchFile du processus de l'indexeur ne peut pas traiter les fichiers XML, elle envoie une erreur fatale à Net Search Extender qui exécute un arrêt immédiat sans émettre de commande COMMIT.

L'erreur suivante est rapportée :

```
Net Search Extender indexing stopped abnormally (CTE0005_INDEX_COMMIT was not found)
```

Causes

La fonction définie par l'utilisateur AFUFetchFile ne peut pas lire le fichier XML créé par le processus de l'indexeur contenant le contenu textuel de l'élément à indexer.

Résolution de l'incident

Pour savoir pourquoi le fichier XML n'a pas été lu :

- Activez la fonction de trace et de consignation AFUFetchFile en définissant l'option de configuration **UdfUseLogging** sur 1.
- Redémarrez DB2.
- Vérifiez la valeur de l'option de configuration **UdfMaxNumberOfWarnings** et comparez-la avec le nombre réel d'avertissements rapportés. Si le nombre d'avertissements émis est supérieur à la valeur indiquée, Net Search Extender arrête le traitement.

- Vérifiez que le répertoire défini dans l'option de configuration **GenXmlDirectory** existe et est accessible et qu'il dispose de suffisamment d'espace pour le traitement.

Net Search Extender échoue lorsque le processus de l'indexeur est en cours d'exécution

Symptômes

Lorsque l'outil de l'indexeur afuIndexer est exécuté, le message d'erreur suivant est rapporté :

CTE0103 Une erreur interne s'est produite. Emplacement : "72", "1".

Causes

Cette erreur signifie que le Groupe de correctifs 3 doit être appliqué pour DB2 version 9.7.

Résolution de l'incident

Appliquez le Groupe de correctifs 3 pour DB2 version 9.7 avant de réexécuter l'outil de l'indexeur afuIndexer. afuIndexer traitera à nouveau le lot d'éléments.

Recherche d'un document introuvable

Si vous ne parvenez pas à retrouver un document au cours d'une recherche, vérifiez s'il a été correctement indexé.

Symptômes

Si vous ne parvenez pas à retrouver un document au cours d'une recherche, vérifiez s'il a été correctement traité et vérifiez la façon dont il a été ajouté à l'index documentaire.

Résolution de l'incident

Pour rechercher un document introuvable :

1. Déterminez l'ID d'élément de l'élément manquant dans la liste des résultats de la recherche en consultant les fichiers journaux.

Dans le cas d'un document Lotus Notes, par exemple, une autre solution consiste à vérifier les propriétés du document :

- Sélectionnez le document dans votre boîte de réception.
- Sélectionnez les propriétés du document et recherchez la zone IBMAfuRestoreInfo. L'ID d'élément est la chaîne d'une longueur de 26 caractères qui commence par A100100.

2. Identifiez la table de base de données des tâches terminées utilisée par le processus de l'indexeur associé au type d'élément.

Exécutez afuIndexer ou afuCheck. Le nom de la table de base de données des tâches ouvertes et celui de la base de données des tâches terminées, ainsi que d'autres propriétés relatives au type d'élément s'affichent.

3. Vérifiez si l'élément a été traité par l'indexeur à l'aide de la requête SQL DB2 suivante :

```
select * from <table of complete task name> where itemid = "<itemid>"
```

- Si la ligne de l'élément est trouvée, passez à l'étape 4, à la page 156
- Si aucune ligne n'est trouvée dans la table, l'indexeur n'a pas encore traité l'élément. Assurez-vous que l'élément se trouve dans la table des tâches ouvertes en exécutant la requête suivante :

```
select * from <table of open task name> where itemid = "<itemid>"
```

- Si la ligne est détectée dans la table des tâches ouvertes, afuIndexer traitera l'élément lors de sa prochaine exécution.
Si IDXSTATUS est défini sur 0, l'élément n'a pas encore été traité et n'a pas été indexé. Dans ce cas, afuIndexer traitera l'élément numéro un lors de la prochaine exécution en fonction de la taille de lot et du nombre d'éléments présents dans la table des tâches ouvertes.
Si IDXSTATUS est un nombre négatif, afuIndexer a tenté de traiter l'élément mais n'a pas pu l'extraire du gestionnaire de ressources IBM Content Manager. Le nombre de tentatives d'extraction d'un document à partir du gestionnaire de ressources dépend de la valeur de l'option de configuration **GenRmRetryAllErrors**. Pour plus d'informations sur l'option de configuration, voir «Options de configuration de l'indexeur», à la page 67.
- Si aucune ligne n'est détectée dans la table, le document n'a pas été archivé correctement ou l'indexation du type d'élément n'a pas été correctement activée.
Ce problème ne doit pas se produire. Exécutez afuIndexTool avec l'argument -opValidate pour vérifier les incohérences d'indexation, et tentez de les corriger.

4. Vérifiez si l'élément a été ajouté à l'index.

Si l'élément figure dans la table des tâches terminées, examinez la valeur dans la colonne IDXRC de l'élément :

- Une valeur 0 indique que l'élément a été ajouté correctement à l'index.
- Une valeur comprise entre 1 et 99 indique que le traitement n'a pas abouti, ce qui a provoqué des avertissements informant que l'élément a peut-être été indexé partiellement. Passez à l'étape 5.
- Une valeur supérieure à 100 indique une erreur de traitement. Cet élément n'est pas ajouté à l'index. En fonction de la valeur IDXRC, vous pouvez essayer de réindexer l'élément une fois la cause de l'erreur résolue.

Pour plus de détails, voir «Codes retour de l'index stockés dans la tables des tâches terminées», à la page 37.

5. Identifiez les modalités d'indexation de l'index

Effectuez les étapes ci-après et ce, que la valeur IDXRC indique que l'élément a été ajouté à l'index en entier ou partiellement :

- a. Exécutez afuIndexer en mode débogage sur l'élément à l'aide de l'argument -debug <id_élément>). Pour plus d'informations sur l'indexation en mode débogage et sur l'affichage des résultats dans un fichier XML, voir «Débogage du traitement d'un élément», à la page 108.
- b. Assurez-vous que le document XML cible contient le texte correct du document source. Si le contenu XML n'est pas complet, contrôlez l'attribut et les sections de l'attribut CM dans le fichier de configuration de l'indexeur.
- c. Veillez à ce qu'une requête de recherche ait été émise pour le document :
 - Localisez le fichier de trace de l'application Web appelé afu_webapplication_trace_0.log. Le chemin d'accès par défaut est C:\Program Files\IBM\ContentCollector\ctms\Log.
 - Recherchez la requête de recherche dans le fichier de trace en effectuant une recherche sur le terme RAL Query. La requête suivante illustre un exemple de requête de recherche :

```
[/iccmemailcmp1d42[(contains-text(@TIEREF,
"(((ATTRIBUTE 'received_date' between 2008093021 and 2010100422))) &
(SECTIONS('mailboxid')
(((('e870f0cbc84b04620cab0e868b9c9c8d'))))")=1]]]
```

Assurez-vous que les termes recherchés dans la requête sont également présents dans le document XML généré.

- Si le document reste introuvable, contactez le service de support logiciel IBM.

Le processus de l'indexeur s'arrête car la fonction définie par l'utilisateur AFUFetch n'est pas accessible

La fonction définie par l'utilisateur AFUFetchFile utilisée par Net Search Extender doit être enregistrée dans DB2.

Symptômes

Le processus de l'indexeur s'arrête et le message suivant s'affiche :

```
The Net Search Extender UDF registered for this item type cannot be identified.
```

```
Verify the correct installation of the Net Search Extender UDF {0}
```

Causes

Cela est dû au fait que la fonction définie par l'utilisateur AFUFetchFile n'est pas accessible car elle n'a pas été enregistrée dans DB2.

Résolution de l'incident

- Vérifiez que la bibliothèque de la fonction définie par l'utilisateur AFUFetchFile existe. Lorsque Content Collector Text Search Support est installé, la bibliothèque de la fonction définie par l'utilisateur AFUFetchFile est copiée, par exemple sous Windows, dans le répertoire SQLLIB\FUNCTION\. Si vous n'y trouvez pas le fichier AfuFetchFile.dll, copiez-le du répertoire IBM\Indexer\lib\ vers votre répertoire SQLLIB\FUNCTION.
- Vérifiez que la fonction définie par l'utilisateur AFUFetchFile est enregistrée dans le Centre de contrôle de DB2 sous *nom_serveur_bibliothèque* > **objets_application** > **Fonctions définies par l'utilisateur**.
- Connectez-vous au serveur IBM Content Manager en tant que propriétaire de l'instance DB2.

- Exécutez

```
DB2 SELECT FUNCNAME FROM SYSCAT.FUNCTIONS WHERE FUNCNAME LIKE 'AFU%'
```

Vous obtenez une liste qui, si elle n'est pas vide, indique que la fonction définie par l'utilisateur est enregistrée.

- Si la fonction définie par l'utilisateur n'est pas enregistrée, exécutez le script suivant dans le sous-répertoire bin du répertoire d'installation de Content Collector Text Search Support :

```
db2 -tvf AfuFetchFile.ddl
```

Une fois le script exécuté, lorsque le processus de l'indexeur démarre, il vérifie automatiquement que la fonction définie par l'utilisateur fonctionne correctement.

- Pour vérifier que la fonction définie par l'utilisateur fonctionne, exécutez :

```
db2 "values ( AFUFETCHFILE ( '_AFU_' ) )"
```

La fonction définie par l'utilisateur est enregistrée et s'exécute correctement si vous visualisez une entrée de table similaire à l'entrée suivante :

```
AfuFetchFile: [V2] Build Id: 2.2.0.0-1 [Compile TOD: Sep 21 2010 15:50:42]  
Copyright (C) 2009 by IBM Corporation
```

- Pour obtenir l'identificateur de l'index pour un type d'élément donné, exécutez la requête :

```
db2 select EVENTTABLENAME, LOGTABLENAME, INDEXIDENTIFIER from DB2EXT.TTEXTINDEXES
WHERE INDEXNAME = upper('<IndexName>')
```

où les parties suivantes constituent le nom de l'index du type d'élément donné :
"ICMUU" + item type component id (5 digits) + "001TIE"

- Pour obtenir le nom de la fonction, exécutez la requête :

```
db2 select FUNCTIONNAME from DB2EXT.TTEXTCOLUMNS WHERE
INDEXIDENTIFIER = '<IndexIdentifier>'
```

Si le résultat n'est pas AFUFETCHFILE, exécutez :

```
db2 UPDATE DB2EXT.TTEXTCOLUMNS SET FUNCTIONNAME = 'AFUFETCHFILE' WHERE
INDEXIDENTIFIER = '<IndexIdentifier>'
```

Obtention du numéro de port DB2

Vous devez connaître le nom de la base de données DB2 et le numéro de port de connexion pour configurer une connexion entre le client d'administration de système et la base de données DB2 distante.

Symptômes

Pour obtenir le numéro de port de connexion DB2 :

Résolution de l'incident

1. Exécutez la commande suivante :
`db2 get dbm cfg`
2. Dans la sortie de la commande, recherchez TCP/IP-Servicename (SVCENAME) = XXX où XXX correspond au nom du service.
3. A l'aide du nom de service, entrez les commandes suivantes :
 - Sous UNIX dans /etc/services :

```
cat /etc/services | grep XXX
```

 - Sous Windows dans <WINDOWSINSTALLDIR>\system32\drivers\etc\services :

```
type c:\WINDOWS\system32\drivers\etc\services | find XXX
```

La sortie ressemble à l'exemple suivant :

- Pour UNIX : db2c_db2inst1 50000/tcp
- Pour Windows : db2c_DB2 50000/tcp

Le numéro de port DB2 est 50000.

Utilisation d'instructions SQL longues ou complexes

Les instructions SQL très longues ou très complexes provoquent un message d'erreur SQL0101N.

Symptômes

Le message d'erreur SQL010N s'affiche dans le fichier journal et l'instruction SQL n'est pas traitée.

Causes

Ce problème se produit car la taille de pile de l'instruction DB2 est trop petite.

Identification de l'incident

Vérifiez si l'instruction SQL :

- dépasse une limite système serveur ou une limite système de source de données en longueur ou en complexité.
- viole une autre limite spécifique de source de données.

Résolution de l'incident

Suivez les étapes de la section SQL troubleshooting technote.

Utilisation d'une configuration JDBC incorrecte pour les processus de l'indexeur

L'erreur SQL SQL1224N est provoquée par une configuration JDBC incorrecte.

Symptômes

La connexion à la base de données du serveur de bibliothèque dans laquelle des types d'élément sont stockés est mise en oeuvre à l'aide de JDBC (Java Data Base Connectivity) qui définit comment un client Java accède à une base de données. JDBC fournit différents types de connexion et IBM Content Collector prend en charge JDBC type 2 et JDBC type 4.

Si une configuration JDBC incorrecte est utilisée, le processus de l'indexeur s'arrête et l'erreur SQL est rapportée :

```
afuIndexer[worker]: An error occurred:  
An SQL problem occurred:  
[jcc][4038][12241][3.53.95] T2LUW exception: SQL1224N The  
database manager is not able to accept new requests, has  
terminated all requests in progress, or has terminated your  
particular request due to an error or a force interrupt.  
SQLSTATE=55032  
ERRORCODE=-1224, SQLSTATE=55032
```

Résolution de l'incident

Passez à une connexion JDBC de type 4 ou augmentez le nombre maximal de connexions client dans DB2. Le paramètre de configuration de cette limite s'appelle **max_connections**. Pour plus d'informations, reportez-vous à la documentation DB2.

Si cette erreur se produit dans le contexte

```
Cannot clone database connection
```

, cela signifie que le nombre maximal de connexions client a été dépassé. Le nombre de connexions client JDBC de type 2 est limité dans DB2. Réduisez le nombre de tâches parallèles.

Chargement impossible du pilote JDBC de type 2

Le processus afuIndexer ne peut pas être démarré. Un message d'erreur SQL s'affiche, indiquant qu'une bibliothèque native n'a pas pu être chargée.

Symptômes

Dès lors que le processus afuIndexer est démarré, le message d'erreur suivant apparaît :

```
[jcc][10389][12245][3.52.95] Failure in loading native library  
db2jcct2, java.lang.UnsatisfiedLinkError: db2jcct2 (A file or  
directory in the path name does not exist.): ERRORCODE=-4472,  
SQLSTATE=null
```

Causes

Ce problème peut provenir d'un chemin de recherche de bibliothèque incorrect sur un système d'exploitation 64 bits.

Résolution de l'incident

Vérifiez les paramètres LIBPATH de l'indexeur de recherche de texte. Le chemin des versions 32 bits des bibliothèques partagées doit précéder celui des versions 64 bits des bibliothèques partagées. Par exemple :

```
LIBPATH=...:/home/db2inst1/sqllib/lib32:/home/db2inst1/sqllib/lib64:  
/home/db2inst1/sqllib/lib:...
```

Les éléments dont le contenu textuel est supérieur à 1 Mo ne sont pas indexés lors de l'utilisation de DB2 version 9.5.

Si vous utilisez DB2 Universal Database version 9.5 et l'indexeur de recherche de texte, les éléments dont le contenu textuel est supérieur à 1 Mo peuvent ne pas être indexés.

Symptômes

Aucune donnée n'est transmise à Net Search Extender pour être indexée.

Résolution de l'incident

Procédez comme suit avant d'appeler DB2 et de démarrer l'indexation. Les éléments non indexés doivent être réindexés :

- Définissez la variable d'environnement DB2 DB2_RESTRICT_DDF sur true.
- Incluez cette variable dans DB2ENVLIST sous UNIX.
- Définissez cette variable en tant que variable d'environnement système sous Windows.

Le processus de l'indexeur s'arrête car l'index est introuvable

Le processus de l'indexeur afuIndexer s'arrête et un message d'erreur s'affiche pour indiquer que l'index n'existe pas.

Symptômes

Lorsque le processus afuIndexer de l'indexeur tente d'exécuter la commande de mise à jour de l'index Net Search Extender, l'erreur suivante se produit :

```
CTE0132 The text index "ADMINISTRATOR"."<nom_index NSE>" n'existe pas.
```

Notez que le nom de l'index est composé du nom actuel du schéma, d'un point et du nom de l'index.

Causes

Il manque très certainement une spécification de schéma sur la ligne de commande lorsque le processus afuIndexer est exécuté.

Résolution de l'incident

Redémarrez le processus afuIndexer en utilisant la valeur correcte pour l'argument de ligne de commande -schema.

Pour déterminer le schéma par défaut de votre connexion actuelle, utilisez la commande DB2 suivante :

```
db2 select current_schema from sysibm.sysdummy1
```

Pour déterminer le schéma approprié pour un type d'élément et son index de texte, utilisez la commande DB2 suivante :

```
db2 select indexschema from db2ext.ttextindexes where indexname = '<nom_index NSE>'
```

où <nom_index NSE> correspond au nom de l'index figurant dans le message d'erreur.

Modifications du répertoire par défaut et des paramètres de nom de fichier de la configuration de la fonction définie par l'utilisateur AFUFetchFile

Le processus de l'indexeur crée et gère automatiquement un fichier de configuration binaire appelé AfuFetchFile.config, qui est utilisé par la fonction définie par l'utilisateur AFUFetchFile.

Symptômes

L'emplacement par défaut de ce fichier de configuration est le répertoire \$(TMP). Par exemple, /tmp/afufetchfile.config sous AIX et Linux, et C:\Documents and Settings\...\AfuFetchFile.config sous Windows. Sous Windows, vous devez définir la variable d'environnement AFU_FETCHFILE_CONFIG_PATH selon l'emplacement du chemin par défaut.

Si la fonction définie par l'utilisateur exécutée par DB2 possède une valeur **TMP** ou **TEMP** différente de celle du shell dans lequel l'indexeur de recherche de texte est exécuté, l'outil d'indexation émet un message d'erreur. Dans ce cas, vous devez utiliser la variable d'environnement **AFU_FETCHFILE_CONFIG_PATH** pour qu'elle pointe vers le répertoire dans lequel le fichier de configuration de la fonction définie par l'utilisateur doit être écrit.

L'outil afuIndexer signale l'erreur suivante :

```
afuIndexer: Error:
```

```
A configuration or setup error has occurred with the following reason:
```

```
The path to the UDF configuration file is defined inconsistently:  
The value of the environment variable <ENVVAR> that is used in  
the DB2 UDF environment points to the UDF configuration  
file at '<XXX>'. However, in the environment in which the  
indexer is running the same environment variable points to  
a UDF configuration file in a different directory at '<YYY>'.  
You should always use AFU_FETCHFILE_CONFIG_PATH as the unique  
system environment variable for the directory in which to store the  
UDF configuration file.
```

L'indexeur a détecté que le contexte dans lequel l'outil d'indexation s'exécute et le contexte dans lequel la fonction définie par l'utilisateur est exécutée via DB2 possèdent des valeurs de variable d'environnement <ENVVAR> différentes. L'emplacement du répertoire dans lequel le fichier de configuration de la fonction définie par l'utilisateur doit être écrit est ambigu.

Résolution de l'incident

Pour éviter toute confusion, utilisez la variable d'environnement AFU_FETCHFILE_CONFIG_PATH pour qu'elle pointe vers le répertoire dans lequel le fichier de configuration de la fonction définie par l'utilisateur est stocké.

Important :

- Si vous utilisez la variable d'environnement AFU_FETCHFILE_CONFIG_PATH pour accéder au fichier de configuration de la fonction définie par l'utilisateur, assurez-vous d'avoir défini la variable d'environnement avec les paramètres système sur le panneau de configuration Windows. (**Panneau de configuration > Système > Avancé > Variables d'environnement > Ajouter**). La simple définition de la variable d'environnement à l'aide de la commande SET sur le shell de ligne de commande ne suffit pas et ne fonctionnera pas dans un contexte DB2.
- Il est recommandé d'utiliser la variable d'environnement AFU_FETCHFILE_CONFIG_PATH si vous exécutez et contrôlez le processus d'indexation via un service de terminal ou une session à distance Windows. Dans un tel environnement, il se peut que les variables d'environnement TMP et TEMP ne pointent pas vers le répertoire défini dans les paramètres système (par exemple : C:\DOCUME~1\icadmin\LOCALS~1\Temp) mais qu'elles pointent vers un sous-répertoire du répertoire d'origine (par exemple : C:\DOCUME~1\icadmin\LOCALS~1\Temp\1). De ce fait, l'indexeur utilise une autre valeur de variable d'environnement comparée à DB2 et produit une erreur CTE0100.
- Le répertoire par défaut de ce fichier de configuration sous UNIX ou Linux reste le répertoire /tmp. Cependant, une nouvelle option vous permet également d'utiliser un répertoire de votre choix pour stocker le fichier de configuration en définissant la variable d'environnement AFU_FETCHFILE_CONFIG_PATH et en lui attribuant une valeur. Par exemple :

```
export AFU_FETCHFILE_CONFIG_PATH=/IccIndexerData/UdfConfig
```

La variable d'environnement doit être définie à la fois dans le contexte dans lequel l'indexeur s'exécute et à l'emplacement où DB2 est démarré. Notez également que si vous utilisez la variable d'environnement AFU_FETCHFILE_CONFIG_PATH pour accéder au fichier de configuration de la fonction définie par l'utilisateur, vous devez inclure AFU_FETCHFILE_CONFIG_PATH à DB2ENVLIST pour que la variable d'environnement soit accessible dans l'environnement DB2.

Erreur lors de l'accès aux fichiers XML dans la fonction définie par l'utilisateur AFUFetchFile

Le processus de l'indexeur ne parvient pas à accéder aux fichiers XML contenant le contenu de document à indexer.

Symptômes

La fonction définie par l'utilisateur AFUFetchFile ne parvient pas à ouvrir les fichiers XML créés lors du traitement du document, ce qui génère le message d'erreur suivant :

```
afuIndexer: An error occurred:
CTE0100 A DB2 operation failed. DB2 information: "01H99"
"[IBM][CLI Driver][DB2/AIX64] SQL0462W Routine "*ETCHFILE"
(specific name "") has returned a warning SQLSTATE, with diagnostic
text "XML: ERR: 22: /path/ICC211_XML/01d1/afu#A100100B2".
SQLSTATE=01H99 "
```

Causes

Ce message d'erreur peut être dû à des droits d'accès de fichier inadéquats. Par exemple, le fichier XML répertorié dans le message d'erreur ne dispose que de droits en lecture/écriture :

```
-rw----- 1 db2inst1 db2iadml 7775 Feb 23 19:03 afu#A1001001A10B19C35634F05123-V1#.xml
```

Identification de l'incident

Cette situation se produit si les fichiers XML et le répertoire XML sont créés à l'aide d'un paramètre umask très restrictif. En conséquence de quoi, la fonction définie par l'utilisateur AFUFetchFile exécutée par DB2 en tant qu'utilisateur isolé ne peut pas accéder aux fichiers XML créés par le processus de l'indexeur exécuté en tant que propriétaire de l'instance. Le propriétaire du processus db2fmp détermine quels sont les droits nécessaires pour les fichiers XML.

Résolution de l'incident

1. Consultez les droits d'accès du propriétaire et de l'ID de groupe pour la fonction définie par l'utilisateur afuFetchFile dans le fichier de contrôle DB2 `~/sql1lib/adm/.fenced`. La fonction AFUFetchFile nécessite les droits d'accès de l'utilisateur isolé et non du propriétaire de l'instance.
2. Si les droits d'accès des fichiers de l'index Net Search Extender ne sont pas identiques à ceux des fichiers XML (400), modifiez-les. Le propriétaire de l'instance et l'utilisateur isolé doivent partager le même groupe principal. Un umask de 077 pour le propriétaire de l'instance doit être modifié en umask d'au moins 027.
3. Une fois les droits modifiés, supprimez les fichiers XML et la structure de répertoire définis dans l'option de configuration **GenXmlDirectory**.
4. Réexécutez le processus de l'indexeur.

Erreur lors de l'accès aux fichiers XML dans la fonction définie par l'utilisateur AFUFetchFile si l'ID d'élément est constitué de points d'interrogation

Le processus de l'indexeur ne parvient pas à accéder aux fichiers XML contenant le document à indexer lorsque l'ID d'élément est constitué de points d'interrogation (?).

Symptômes

La fonction définie par l'utilisateur AFUFetchFile ne parvient pas à ouvrir les fichiers XML créés lors du traitement du document, ce qui génère le message d'erreur suivant :

```
afuIndexer: An error occurred:  
CTE0100 A DB2 operation failed. DB2 information: "01H99"  
"[IBM][CLI Driver][DB2/AIX64] SQL0462W Routine "*ETCHFILE"  
(specific name "") has returned a warning SQLSTATE, with diagnostic  
text "XML: ERR: 22: /path/ICC211_XML/01d1/afu#?????????".  
SQLSTATE=01H99 ".
```

Causes

Ce message d'erreur est dû au fait que la valeur TIEREF de l'élément initial est vide. L'élément doit posséder une valeur TIEREF valide pour pouvoir être traité correctement.

Identification de l'incident

Cette situation survient si l'élément en cours de traitement ne possède pas de valeur TIEREF, autrement dit si la valeur contenue dans la colonne TIEREF de

l'élément est vide. Cela se produit parfois si le type d'élément a déjà été traité par l'indexeur rapide et s'il a été stocké dans un état incorrect.

Résolution de l'incident

Pour attribuer une valeur TIEREF à tous les types d'élément qui n'en ont pas, exécutez l'outil `afuIndexTool` avec l'argument `-opValidate -repairTierref -deepvalidation`.

Le processus de travail de l'indexeur s'arrête et indique une situation de mémoire saturée.

Le processus de travail de l'indexeur s'arrête parfois et suppose qu'il n'y a pas assez de mémoire disponible pour poursuivre le traitement.

Symptômes

Cette situation n'est pas due à la taille du lot qui serait trop importante pour permettre un autre traitement.

Causes

La raison de cette erreur peut être l'une des suivantes :

- Un élément contient une pièce jointe de taille volumineuse qui ne peut pas être traitée.
- Trop de tâches parallèles ont été spécifiées.
- Les paramètres de mémoire configurés sont trop élevés.

Résolution de l'incident

Ce problème n'est pas grave et n'a pas d'impact sur le traitement global de l'indexeur. Malheureusement, cette situation et ses conséquences sont inévitables dans Java.

Pour cette raison, le processus de l'indexeur continue le traitement des autres éléments du lot. Les éléments n'ayant pu être traités en raison de ce problème de mémoire sont stockés dans la table des tâches terminées et peuvent être traités à nouveau une fois les options de configuration appropriées modifiées.

Partie 2. Indexation dans IBM FileNet P8

Indexation dans IBM FileNet P8

IBM FileNet P8 propose deux composants de Content Search Engine, IBM Legacy Content Search Engine et IBM Content Search Services, que vous pouvez exécuter en parallèle pour l'indexation et la recherche.

Cependant, Content Collector ne prend pas en charge l'indexation simultanée avec IBM Content Search Services et IBM Legacy Content Search Engine pour un seul conteneur d'objets. Les conteneurs d'objets qui sont utilisés pour l'archivage avec Content Collector doivent être configurés pour utiliser IBM Legacy Content Search Engine ou IBM Content Search Services. Par conséquent, si votre installation d'IBM FileNet P8 Content Engine inclut les deux composants de Content Search Engine et que vous souhaitez utiliser IBM Content Search Services au lieu de IBM Legacy Content Search Engine pour l'indexation des documents archivés avec Content Collector, créez un nouveau conteneur d'objets et configurez-le pour l'indexation avec IBM Content Search Services uniquement.

Les conteneurs d'objets pour l'indexation avec IBM Legacy Content Search Engine doivent être configurés avec les classes de document ICCMail2 et ICCMailSearch2. Les conteneurs d'objets pour l'indexation avec IBM Content Search Services doivent être configurés avec la classe de document ICCMail3.

Important : Bien que FileNet P8 prenne en charge la migration des index IBM Legacy Content Search Engine à IBM Content Search Services, vous ne pouvez pas utiliser ce processus de migration pour les conteneurs d'objets qui contiennent des courriers électroniques archivés avec Content Collector dans les classes de document ICCMail2 et ICCMailSearch2.

Nouvelles fonctions d'indexation de IBM FileNet P8

IBM Content Collector prend en charge l'indexation dans IBM FileNet P8 à l'aide d'IBM Legacy Content Search Engine et IBM Content Search Services.

Fonctions d'indexation ajoutées dans IBM Content Collector V3.0

IBM Content Collector P8 Content Search Services Support

IBM Content Collector P8 Content Search Services Support est un plug-in du constructeur de document facultatif dans IBM Content Search Services pour un prétraitement personnalisé de tous les documents archivés à l'aide d'IBM Content Collector autre que des documents du système de fichiers.

Fonctions d'indexation ajoutées à IBM Content Collector V3.0.0.1

Nouvel attribut de courrier électronique Lotus Notes

L'attribut Lotus Notes suivant a été ajouté au modèle de configuration de messagerie Content Search Services Support :

```
$INetPrincipal = /icc_document/icc_email/icc_from
```

Dans les versions antérieures d'IBM Content Collector, seul le principal de l'e-mail était indexé sous forme canonique (Principal = /icc_document/icc_email/icc_from). L'adresse électronique ou Internet du propriétaire principal, par exemple jdoe@company.com, n'était pas été indexée du fait que l'attribut était manquant dans les fichiers de configuration de la messagerie électronique.

Désormais, le propriétaire de l'e-mail, l'adresse e-mail du propriétaire et l'adresse électronique de la personne déléguée qui a effectivement envoyé le courrier électronique sont indexées sous `icc_from`.

Pour pouvoir rechercher l'adresse SMTP du propriétaire de boîte aux lettres dans le courrier électronique de délégation, l'index recherché doit avoir été créé en utilisant l'attribut `$INetPrincipal`. Seuls les index créés via IBM Content Collector V3.0.0.1 et versions ultérieures contiennent cet attribut. Les index créés avant IBM Content Collector V3.0.0.1 ne contiennent pas cette information dans l'index. Si vous souhaitez rechercher l'adresse SMTP de la boîte aux lettres du propriétaire dans un courrier électronique de délégation contenu dans des index anciens, vous devez ajouter cette information au magasin d'objets IBM FileNet P8, puis réindexer tous les documents de messagerie.

Fonctions d'indexation ajoutées à IBM Content Collector V3.0.0.2

Recherche de courrier électronique avec un corps vide dans l'index

Vous pouvez interroger l'index pour rechercher tous les messages électroniques dont le corps est vide.

Indexation avec IBM Legacy Content Search Engine

IBM Legacy Content Search Engine peut être utilisé avec IBM FileNet P8 pour indexer les documents et activer la fonction de recherche.

Pour trouver des objets qui ont été archivés dans un conteneur d'objets IBM FileNet P8 par la recherche de mots ou d'expressions intégrés, le contenu de ces objets doit être indexé. L'indexation pour l'extraction de contenu (CBR) est le processus de création de ces index en texte intégral. Pour vous assurer de l'indexation correcte des documents qui sont archivés à l'aide d' IBM Content Collector, vous devez activer le conteneur d'objets en conséquence.

Concepts d'indexation dans IBM Legacy Content Search Engine

Utilisez IBM Legacy Content Search Engine en tant que le moteur de recherche de contenu pour écrire et lire les informations d'index pour les objets IBM FileNet Content Engine afin que les utilisateurs puissent effectuer des recherches en texte intégral sur ces objets.

Avec IBM Legacy Content Search Engine, les index en texte intégral sont également appelés collectes Autonomy ou Verity. Les collectes sont conservées dans les zones d'index. Une zone d'index est associée à un conteneur d'objets spécifique. Plusieurs zones d'index peuvent être utilisées pour stocker des données pour un seul conteneur d'objets, mais chaque zone d'index peut contenir des données pour un seul conteneur d'objets.

Un index ne peut être ouvert que par un seul serveur d'index IBM Legacy Content Search Engine à la fois. Content Engine affecte les index aux serveurs d'index disponibles de sorte que la charge de travail soit à peu près la même pour tous les serveurs.

Chaque fois qu'un objet est créé, mis à jour ou supprimé dans un conteneur d'objets qui a été activé pour l'indexation en texte intégral à l'aide d'IBM Legacy Content Search Engine, une demande d'index est démarrée. Les demandes d'index sont regroupées pour former un index par lots. Une partition est créée pour

chaque lot d'indexation que Content Engine envoie au serveur d'index IBM Legacy Content Search Engine. Les demandes d'index peuvent être surveillées pour identifier les incidents d'indexation à l'aide de l'IBM FileNet Enterprise Manager.

Activation de l'indexation en texte intégral avec IBM Legacy Content Search Engine

Pour activer l'extraction basée sur le contenu des documents qui ont été archivés dans un conteneur d'objets FileNet P8 en utilisant IBM Content Collector, vous devez configurer FileNet P8 Content Search Engine en conséquence.

La procédure générale d'activation d'un conteneur d'objets pour la recherche à l'aide d'IBM Content Collector est la suivante. Les étapes à suivre sont liées à la version de FileNet P8 utilisée.

Pour activer un conteneur d'objets pour la recherche avec IBM Content Collector :

1. Activez le conteneur d'objets pour l'indexation en texte intégral avec IBM(r) Legacy Content Search Engine. Cette procédure s'applique à IBM FileNet P8 Version 4.5.1 et ultérieure.
2. Vérifiez votre configuration IBM Legacy Content Search Engine.
 - «Vérification de votre configuration IBM Legacy Content Search Engine pour les conteneurs d'objets qui ne sont pas partitionnés par date», à la page 178
 - «Vérification de votre configuration IBM Legacy Content Search Engine pour les conteneurs d'objets partitionnés par date», à la page 177Cette procédure s'applique à IBM FileNet P8 Version 4.5.1 et ultérieure.

Activation des conteneurs d'objets pour l'indexation en texte intégral avec IBM Legacy Content Search Engine

Configurez l'extraction de contenu des courriers électroniques qui ont été archivés à l'aide d'IBM Content Collector, IBM FileNet P8 Content Engine Version 4.5.1 et ultérieure, et IBM Legacy Content Search Engine.

Conditions requises :

- Une installation d'IBM FileNet P8 Content Engine Version 4.5.1 ou version ultérieure et de Content Search Engine avec IBM Legacy Content Search Engine
- Un conteneur d'objets FileNet P8 vide utilisant la zone de stockage de fichier comme conteneur de contenu par défaut. Si le conteneur d'objets possède déjà une zone d'index, reportez-vous à la procédure de la rubrique «Remplacement d'un index», à la page 182. Pour plus d'informations sur la définition d'un conteneur d'objets, consultez la documentation IBM FileNet P8.

Lorsque vous créez un nouveau conteneur d'objets, assurez-vous que seul l'additif **Base Content Engine Extension** est installé. Vous pouvez pour cela sélectionner **Configuration minimale**. L'installation d'additifs crée plusieurs classes de documents et index de bases de données qui ne sont pas requis par Content Collector, ce qui affecte la performance.

Pour activer un conteneur d'objets pour Content Collector :

1. Accédez au dossier ctms du répertoire d'installation de Content Collector.
2. Sélectionnez l'ensemble de styles approprié et extrayez le contenu du fichier .zip dans le répertoire stylesets de FileNet P8 Content Search Engine. Les répertoires de styles se trouvent à l'emplacement suivant.

Tableau 15. Emplacement du répertoire *stylesets*

Système d'exploitation	Chemin
AIX	/opt/verity/data/stylesets
Windows	<i>répertoire_installation_FileNet</i> \FileNet\ ContentEngine\Verity\data\stylesets\ Où <i>FileNetInstallDir</i> est le chemin d'installation de FileNet P8.

Ces ensembles de style sont disponibles pour FileNet P8 Version 4.5.1 ou ultérieure. Pour plus d'informations, voir la rubrique relative aux ensembles de styles.

ICC_FileSystem_PushAPI_p8cse_4.5.1

Pour les conteneurs d'objets partitionnées par date qui sont configurés avec un paramètre régional autre que l'anglais de base

ICC_FileSystem_PushAPI_p8cse_4.5.1_english

Pour les conteneurs d'objets partitionnées par date qui sont configurés avec le paramètre régional anglais de base

ICC_FileSystem_PushAPI_p8cse_4.5.1_no_partition

Pour les conteneurs d'objets non partitionnées par date qui sont configurés avec un paramètre régional autre que l'anglais de base

3. Vérifiez que le répertoire de l'ensemble de styles contient le fichier *style.ufl*. Si ce fichier est manquant, copiez une version à partir du répertoire *FileNet_fileSystem_PushAPI*.
4. Dans le tableau de bord Autonomy K2, sélectionnez **Collections > Manage Style Sets > Import** pour importer cet ensemble de styles.
 - a. Dans la zone **Alias de l'ensemble de styles**, entrez le nom de l'ensemble de styles.
 - b. Dans la zone **Gateway Type**, entrez : File System Gateway.
 - c. Dans la zone **Source Path**, entrez le chemin d'accès au répertoire *stylesets*.
 - d. Cliquez sur **Importer**.
 - e. Retournez à la page d'accueil du tableau de bord d'Autonomy K2 et suivez la procédure indiquée. Ces étapes sont nécessaires pour régénérer la configuration Autonomy et mettre en oeuvre vos modifications.
5. Pour un nouveau conteneur d'objets, vous devez définir une valeur de paramètres régionaux CBR avant de pouvoir créer une zone d'index. Dans IBM FileNet Enterprise Manager, accédez à l'onglet CBR (FileNet P8 Version 4.5) ou à l'onglet CBR Verity (FileNet P8 Version 5) de la feuille des propriétés du conteneur d'objets. Pour obtenir une liste exhaustive des valeurs de paramètres régionaux CBR valide, reportez-vous à la documentation sur Autonomy K2.

Conseil : En général, vous pouvez utiliser le paramètre régional de version multilingue (*uni*) avec l'ensemble de styles *ICC_FileSystem_PushAPI_p8cse_4.5.1* ou *ICC_FileSystem_PushAPI_p8cse_4.5.1_no_partition*. Toutefois, pour les environnements à haut débit ou si vous souhaitez limiter la charge de traitement de l'indexation de contexte, utilisez l'anglais de base (*english*) car ce paramètre régional offre un avantage de performances par rapport au paramètre régional en plusieurs langues. Le paramètre régional anglais de base assure un traitement suffisant du texte à l'aide de l'alphabet latin.

Toutefois, le traitement des jeux de caractères à deux octets n'est pas possible avec cet paramètre régional. Utilisez l'ensemble de stylesICC_FileSystem_PushAPI_p8cse_4.5.1_english avec le paramètre régional anglais de base.

6. Si vous souhaitez inclure des attributs personnalisés dans l'index, adaptez le fichier `style.xml` à cet effet. Le fichier `style.xml` contient une définition pour une zone appelée `icc_custom_metadata`. Par défaut, toutes les métadonnées définies par l'utilisateur y sont stockées. Pour que chaque attribut personnalisé soit indexé séparément dans sa propre zone et que les recherches soient activées dans cette zone spécifique, ajoutez un élément `<preserve xmltag="xxx" >` pour chaque attribut.

L'ensemble de styles contient un exemple de fichier de mots vides nommé `style.stp.disabled`. Ce fichier est désactivé par défaut. Pour activer le fichier de mots vides, supprimez le suffixe `.disabled` du nom de fichier. Vous pouvez ajouter une liste de mots spécifiques à la langue à ignorer lors de l'indexation. Les mots doivent être séparés par un retour chariot ou un saut de ligne.

Etant donné que les mots vides sont spécifiques à la langue, l'exemple fourni offre les optimisations minimales qui ne dépendent pas de la langue utilisée, en ignorant :

- Les mots contenant plus de 40 caractères
- Les jetons entiers comportant plus de 8 chiffres

Lorsque vous activez le fichier de mots vides, archivez un petit ensemble de données de courrier électronique représentatives dans un environnement de test avant de le déployer dans un environnement de production pour vous assurer que le fichier de mots vides est compatible avec l'usage spécifique que vous en fait.

Important : Si vous modifiez l'un des fichiers de style utilisés pour créer un index, vous devez remplacer l'index pour un conteneur d'objets existant. La réindexation des objets n'intégrera pas les modifications du fichier de style dans un index existant.

7. Dans IBM FileNet Enterprise Manager, créez une zone d'index pour le conteneur d'objets Content Collector à l'aide de l'ensemble de styles Content Collector.

Important : Seules les zones d'index créées à l'aide du type de modèle sélectionné doivent exister pour le conteneur d'objets que vous souhaitez utiliser pour l'archivage de courriers électroniques avec IBM Content Collector. Si les zones d'index du conteneur d'objets ont été configurées avec un autre modèle, vous devez nettoyer les collectes de l'ancienne zone d'index, tel que décrit dans la rubrique relative au remplacement d'un index.

8. Testez l'index.
 - a. Ajoutez un fichier texte au conteneur d'objets en utilisant IBM FileNetWorkplace ou IBM FileNet Enterprise Manager.
 - b. Attendez que le fichier soit indexé. Cette opération peut prendre environ 10 minutes.
 - c. Utilisez la fonction Test de recherche du tableau de bord Autonomy K2 sur votre serveur de recherche pour rechercher le fichier. Le fichier doit s'afficher dans les résultats de la recherche.

9. Installez le modèle de données IBM CC Compliance pour le conteneur d'objets en exécutant l'outil ICCComplianceInstaller. Voir la rubrique sur la configuration d'un référentiel IBM FileNet P8 en mode console pour plus d'informations.
10. Retournez à IBM FileNet Enterprise Manager et vérifiez que le modèle de données est installé avec les paramètres appropriés. Dans l'onglet CBR (FileNet P8 Version 4.5) ou l'onglet CBR Verity (FileNet P8 Version 5) de la feuille de propriétés du conteneur d'objets, vérifiez l'intervalle définie pour la propriété de date ICCMailDate.

Pour configurer les paramètres de partition, sélectionnez **Modifier les paramètres de partitionnement de la collecte** et définissez un intervalle pour la propriété de date ICCMailDate.

Veillez à utiliser une valeur appropriée pour l'intervalle de partitionnement. Reportez-vous à l'exemple de calcul des intervalles de partitionnement pour plus d'informations.

Important : Si vous modifiez l'intervalle pour un conteneur d'objets qui n'est pas vide, vous devez réindexer ce conteneur pour que les nouveaux paramètres prennent effet pour les courriers électroniques déjà archivés.

11. FileNet P8 Version 5.1 : pour faciliter l'identification des incidents d'indexation d'objet, configurez chaque conteneur d'objets de sorte à propager les erreurs d'indexation. Dans l'onglet Général de la feuille de propriétés du conteneur d'objets, vérifiez que la valeur du **Niveau d'enregistrement des incidents d'indexation** est définie sur **Propager à la source**. Utilisez IBM FileNet Enterprise Manager à cette fin. Corrigez le paramètre si nécessaire. Les incidents propagés sont enregistrés dans la propriété Codes d'incident d'indexation de chaque objet.

Le conteneur d'objets est maintenant configuré pour l'archivage avec Content Collector et l'indexation en texte intégral avec IBM Legacy Content Search Engine.

Vérifiez maintenant votre configuration IBM Legacy Content Search Engine.

Information associée :

 Centre de documentation d'IBM FileNet P8 Version 5.1

Modification de l'ensemble de styles pour un conteneur d'objets :

Modifiez le comportement de l'index et du serveur de recherche en modifiant l'ensemble de styles pour un conteneur d'objets.

Pour modifier l'ensemble de styles pour un conteneur d'objets comportant déjà une ou plusieurs zones d'index :

1. Accédez au dossier ctms du répertoire d'installation de Content Collector.
2. Sélectionnez l'ensemble de styles approprié et extrayez le contenu du fichier .zip dans le répertoire stylesets de FileNet P8 Content Search Engine. Les répertoires de styles se trouvent à l'emplacement suivant.

Tableau 16. Emplacement du répertoire stylesets

Système d'exploitation	Chemin
AIX	/opt/verity/data/stylesets

Tableau 16. Emplacement du répertoire *stylesets* (suite)

Systeme d'exploitation	Chemin
Windows	<i>répertoire_installation_FileNet</i> \FileNet\ ContentEngine\Verity\data\stylesets\ Où <i>FileNetInstallDir</i> est le chemin d'installation de FileNet P8.

3. Dans le tableau de bord Autonomy K2, sélectionnez **Collections > Manage Style Sets > Import** pour importer cet ensemble de styles.
 - a. Dans la zone **Alias de l'ensemble de styles**, entrez le nom de l'ensemble de styles.
 - b. Dans la zone **Gateway Type**, entrez : File System Gateway.
 - c. Dans la zone **Source Path**, entrez le chemin d'accès au répertoire *stylesets*.
 - d. Cliquez sur **Importer**.
 - e. Retournez à la page d'accueil du tableau de bord d'Autonomy K2 et suivez la procédure indiquée. Ces étapes sont nécessaires pour régénérer la configuration Autonomy et mettre en oeuvre vos modifications.
 4. Si vous souhaitez inclure des attributs personnalisés dans l'index, adaptez le fichier *style.xml* à cet effet. Le fichier *style.xml* contient une définition pour une zone appelée *icc_custom_metadata*. Par défaut, toutes les métadonnées définies par l'utilisateur y sont stockées. Pour que chaque attribut personnalisé soit indexé séparément dans sa propre zone et que les recherches soient activées dans cette zone spécifique, ajoutez un élément `<preserve xml tag="xxx" >` pour chaque attribut.
Veillez à copier n'importe quel fichier de mots vides précédemment configuré dans le répertoire d'ensemble de styles.
 5. Affectez l'ensemble de styles à toutes les zones d'index dans le conteneur d'objets utilisé pour l'archivage avec Content Collector.
 - a. Dans IBM FileNet Enterprise Manager, développez le noeud du conteneur d'objets et sélectionnez le dossier Zones d'index.
 - b. Cliquez avec le bouton droit sur la zone d'index et cliquez sur **Propriétés**
 - c. Sous l'onglet Propriétés, entrez l'alias de l'ensemble de styles approprié pour la propriété Type de modèle.
 6. Réindexez chaque classe de base pour laquelle l'extraction de contenu est activée avec des sous-classes pour remplacer les collectes existantes par les nouvelles collectes qui utilisent les nouveaux ensembles ou les ensembles de styles Content Collector modifiés.
- Important :** Seules les zones d'index créées à l'aide du type de modèle sélectionné doivent exister pour le conteneur d'objets que vous souhaitez utiliser pour l'archivage de courriers électroniques avec IBM Content Collector.
7. Surveillez les tâches d'indexation pour vérifier qu'elles aboutissent.

Ensembles de styles IBM Legacy Content Search Engine :

IBM Content Collector fournit les ensembles de styles IBM Legacy Content Search Engine (également appelés ensembles de styles Verity) de quatre scénarios différents. Les ensembles de styles sont installés dans le sous-dossier *ctms* du répertoire d'installation Content Collector.

ICC_FileSystem_PushAPI_p8cse_4.5.1

Cet ensemble de styles convient à une utilisation avec IBM FileNet P8 Content Search Engine version 4.5.1 ou versions ultérieures. Il peut être utilisé avec n'importe quel paramètre régional. Le comportement de chaque paramètre régional est tel que décrit dans la documentation de Verity. Il n'y a pas de configuration spéciale pour les paramètres régionaux spécifiques.

Utilisez cet ensemble de styles pour les conteneurs d'objets partitionnés par date sauf si l'un des conteneurs est configuré avec le paramètre régional de base anglais. Les requêtes de plages de dates sont exécutées par rapport à la propriété ICCMailDate de l'objet ICCMailSearch2. Les mappages d'archive par défaut pour la recherche de courriers électroniques IBM Content Collector sont configurés en conséquence.

Pour utiliser l'ensemble de styles, suivez les instructions dans les rubriques «Activation des conteneurs d'objets pour l'indexation en texte intégral avec IBM Legacy Content Search Engine», à la page 169 ou «Modification de l'ensemble de styles pour un conteneur d'objets», à la page 172. Utilisez les noms suivants :

- Le nom du fichier de package est ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1.zip.
- Le nom du dossier et le nom de l'ensemble de styles par défaut dans le dossier d'ensembles de styles Verity sont ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1.

ICC_FileSystem_PushAPI_p8cse_4.5.1_english

Cet ensemble de styles convient à une utilisation avec IBM FileNet P8 Content Search Engine version 4.5.1 ou versions ultérieures, si le conteneur d'objets FileNet P8 est partitionné par date et configuré avec le paramètre régional de base anglais. Grâce à cet ensemble de styles, les utilisateurs peuvent rechercher des adresses électroniques au format myname@my_company.com. Cet ensemble de styles convient uniquement au paramètre régional de base anglais et ne doit s'appliquer à aucun autre paramètre régional.

Les recherches de dates sont exécutées par rapport à la propriété ICCMailDate de l'objet ICCMailSearch2. Le mappage d'archive par défaut pour la recherche de courriers électroniques IBM Content Collector est configuré en conséquence.

Pour utiliser l'ensemble de styles, suivez les rubriques «Activation des conteneurs d'objets pour l'indexation en texte intégral avec IBM Legacy Content Search Engine», à la page 169 ou «Modification de l'ensemble de styles pour un conteneur d'objets», à la page 172. Utilisez les noms suivants :

- Le nom du fichier de package est ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1_english.zip.
- Le nom du dossier et le nom de l'ensemble de styles par défaut dans le dossier d'ensembles de styles Verity sont ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1_english.

Cet ensemble de styles est identique à l'ensemble de styles ICC_FileSystem_PushAPI_2.1_p8cse_4.5.1, sauf qu'il est doté d'un autre fichier style.lex. Ce fichier permet d'effectuer des recherches de chaînes contenant le signe arobase (@).

ICC_FileSystem_PushAPI_p8cse_4.5.1_no_partition

Cet ensemble de styles convient à une utilisation avec IBM FileNet P8 Content Search Engine version 4.5.1 ou versions ultérieures. Il peut être utilisé avec n'importe quel paramètre régional. Le comportement de chaque paramètre régional est tel que décrit dans la documentation de Verity. Il n'y a pas de configuration spéciale pour les paramètres régionaux spécifiques.

Utilisez cet ensemble de styles uniquement si votre conteneur d'objets P8 Content Engine n'est pas partitionné par date et si vous souhaitez utiliser la propriété XIT `icc_received_date` pour les requêtes de plage de dates. Vos mappages d'archives pour la recherche de courriers électroniques IBM Content Collector doivent être configurés en conséquence.

```
<field nm="EMAIL_DATE" type="DATE" partitionkey="true">
  <search format="yyyyMMddHH">icc_received_date</search>
  <attr>ICCMailDate</attr>
</field>
```

Si par la suite, vous définissez le partitionnement de date pour le conteneur d'objets, n'utilisez pas cet ensemble de styles. Dans ce cas, utilisez l'ensemble de styles `ICC_FileSystem_PushAPI_p8cse_4.5.1` et réindexez votre collecte.

Pour utiliser l'ensemble de styles, suivez les instructions dans les rubriques «Activation des conteneurs d'objets pour l'indexation en texte intégral avec IBM Legacy Content Search Engine», à la page 169 ou «Modification de l'ensemble de styles pour un conteneur d'objets», à la page 172. Utilisez les noms suivants :

- Le nom du fichier de package
`estICC_FileSystem_PushAPI_p8cse_4.5.1_no_partition.zip`.
- Le nom du dossier et le nom de l'ensemble de styles par défaut dans le dossier d'ensembles de styles Verity sont
`ICC_FileSystem_PushAPI_p8cse_4.5.1_no_partition`.

Si votre conteneur d'objets est configuré avec les paramètres régionaux anglais de base et que vous souhaitez pouvoir rechercher des chaînes contenant le signe arobase (@), copiez le fichier `style.lex` inclus dans l'ensemble de styles `ICC_FileSystem_PushAPI_p8cse_4.5.1_english` dans l'ensemble de styles `ICC_FileSystem_PushAPI_p8cse_4.5.1_no_partition`.

Tâches associées :

«Remplacement d'un index», à la page 182

Options de configuration pour l'indexation dans des environnements en langue mixte :

L'identification de la langue peut s'avérer difficile lors de l'indexation dans des environnements où plusieurs locales sont utilisées. Le filtre d'identification de la langue est désormais configuré par défaut dans les ensembles de styles IBM Legacy Content Search Engine (Verity) livrés avec IBM Content Collector.

Le filtre d'identification de la langue est défini par `flt_lang` dans une instruction `postformat` dans le fichier `style.uni` de l'ensemble de styles. Dans certains environnements, la désactivation de cette option peut entraîner le blocage de l'indexeur IBM Legacy Content Search Engine et est, par conséquent, non recommandée.

Verity fournit les solutions suivantes :

- Appliquez la langue dans le fichier d'insertion en vrac (BIF).

Si vous définissez l'option `-bifmime` dans les filtres de reconnaissance du fichier `style.uni`, vous pouvez attribuer une langue en définissant une zone `VLANG` dans le `BIF`, qui remplace l'identification de langue `flt_lang`. Lorsque `VLANG` est défini sur `annuler`, un document est analysé avec le marqueur sémantique générique. En définissant une zone `VLANG`, vous pouvez définir la langue jusqu'au niveau du document, mais vous devez changer les données du `BIF` soumises à Verity K2.

- Modifiez le comportement de détection dans les paramètres régionaux `uni`.
Dans le fichier `langlist.cfg` du répertoire `k2/common/langid/`, vous pouvez définir l'ensemble de langues possibles, la langue par défaut et le seuil de détection. Cette modification affecte le comportement de l'environnement des paramètres régionaux pour toutes les collectes utilisant les paramètres régionaux `uni` dans cette installation K2. Toutefois, si le problème principal est la détection des documents (par exemple, si les documents sont en chinois, mais que du contenu en particulier désactive la détection), cette option peut être applicable.
- Modifiez le marquage sémantique des langues spécifiques dans les paramètres régionaux `uni`.
Plutôt que de modifier l'identification de la langue, vous pouvez modifier le marquage sémantique qu'utilise une langue donnée. Dans le fichier `uni.cfg` du répertoire `k2/common/uni/`, ajoutez l'instruction `marqueur sémantique: builtin` à l'entrée de cette langue afin que le marqueur sémantique intégré soit utilisé. Ceci peut désactiver certaines fonctions linguistiques, ce qui se produit également si vous désactivez l'identification de la langue `flt_lang`. Cette modification affecte le comportement de l'environnement des paramètres régionaux pour toutes les collectes utilisant les paramètres régionaux `uni` dans cette installation K2.
- Indiquez un autre fichier `langlist.cfg` pour la collecte.
Cela équivaut à modifier le comportement de détection dans les paramètres régionaux `uni`, mais prend effet uniquement au niveau de la collecte. Lorsque vous affectez `flt_lang`, vous pouvez indiquer un autre fichier de liste de langues avec le paramètre `-config`. Le nom par défaut du fichier de liste de langues est `langlist.cfg`. Pour plus d'informations, consultez la section relative à la configuration du filtre de l'identification de la langue dans *Verity Collection Reference*. La documentation Verity est disponible sur le site Web des services Autonomy.
- Appliquez le paramètre de langue `annuler` au niveau de la collecte.
Vous pouvez appliquer l'utilisation d'un paramètre de langue spécifique pour tous les documents à l'aide du paramètre `-lang` lorsque vous attribuez `flt_lang`. Cela équivaut à désactiver `flt_lang` sans ignorer la détection de jeu de caractères.

Important : Ces solutions se trouvent dans la portée de configuration de Verity, elles ne sont pas spécifiques à Content Collector. Par conséquent, Content Collector ne conserve pas les ensembles de styles spéciaux pour la configuration des paramètres régionaux de Verity. Consultez la documentation de Verity pour plus d'informations sur la configuration de votre environnement Verity pour un paramètre régional donné. Si nécessaire, contactez le service de support logiciel IBM (pour IBM FileNet P8 avec IBM Legacy Content Search Engine) pour plus de détails sur les solutions décrites.

Information associée :

 Site Web des services Autonomy

Définition de la taille de la partition d'index LCSE pour le traitement de documents volumineux :

Lorsque des documents contenant de très grandes quantités de texte sont traités par IBM Content Collector, les collectes IBM Legacy Content Search Engine (LCSE) peuvent s'avérer corrompues si la taille de la partition d'index maximum n'est pas correctement définie.

La collecte s'endommage lorsque le serveur d'index tente de l'optimiser. Au cours de l'optimisation, les petits fichiers de partition d'index sont fusionnés aux grands. Le nombre maximal d'objets indexés placés dans une partition est spécifié par le paramètre **merge_maxdocs** dans le fichier de style `style.plc`. Si la taille de la partition fusionnée dépasse 2 Go, l'optimisation échoue et la collecte est endommagée. Toutefois, la collecte peut encore être interrogée mais si elle est encore modifiée, elle devient trop endommagée pour être ouverte.

Un message d'erreur similaire au message suivant indique une telle condition :

```
com.filenet.api.exception.EngineRuntimeException:  
FNRCR0022E: RETRIEVE_CANT_OPEN_COLLECTION: Impossible d'ouvrir la collecte Verity  
"K20SLD_Document_20110921132134_E7EB5AC43E264DEEA1E5EFF480B6710",  
erreur : collecté en panne après son ouverture. (10)
```

Lorsque l'optimisation échoue, vous devez réindexer la collecte. Vous risquez de ne pas savoir que l'optimisation échoue à moins que vous remarquiez de longs délais d'optimisation ou des fichiers temporaires dans la collecte. Cependant, dès que l'erreur `RETRIEVE_CANT_OPEN_COLLECTION` se produit, la collecte ne peut pas être réindexée. Dans ce cas, vous devrez contacter le support logiciel IBM pour résoudre le problème.

Pour éviter ce problème, évaluez la taille maximale des partitions de la collecte et ajustez la taille de la partition d'index maximale en conséquence.

Exemple de procédure pour l'estimation de la taille :

1. Indexez 10 000 éléments dans une collecte.
2. Exécutez une requête d'optimisation pour la collecte.
3. Vérifiez la taille du fichier `.did` dans le répertoire des pièces pour la collecte.
4. Multipliez cette valeur par 6,5 pour obtenir la taille maximale estimée d'une partition.
5. Si la taille maximale estimée dépasse 1 Go, réduisez la valeur du paramètre **merge_maxdocs**.

Pour plus de détails sur la taille de la partition d'index, reportez-vous à la référence de l'ensemble de styles `P8CSE50` dans la documentation FileNet P8.

Vérification de votre configuration IBM Legacy Content Search Engine pour les conteneurs d'objets partitionnés par date

Après avoir installé IBM Content Collector Server et exécuté l'assistant de configuration initiale, vérifiez votre configuration IBM Legacy Content Search Engine à l'aide de IBM FileNet Enterprise Manager (FEM).

Pour vérifier la configuration des conteneurs d'objets partitionnés par date :

1. Créez deux documents, de classe `ICCMailSearch2`, contenant des fichiers texte. Les fichiers texte doivent posséder l'extension de fichier `.xml` et contenir l'instruction XML suivante : `<icc_content>textetest</icc_content>`. Par exemple, `searchtest1.xml` avec `<icc_content>test partition`

one</icc_content> et searchtest2.xml avec <icc_content>test partition two </icc_content>. Si la classe de documents ICCMailSearch2 n'existe pas dans le conteneur d'objets, vous devez réexécuter l'assistant de configuration initiale afin d'activer le conteneur d'objets pour l'archivage des courriers électroniques à l'aide d'IBM Content Collector.

2. Importez les fichiers en tant que nouveaux documents P8 de type ICCMailSearch2. Veillez à attacher chaque fichier en tant qu'élément de contenu, puis attribuez une valeur de date/heure à la propriété ICCMailDate de chaque fichier. Par exemple, affectez à la propriété ICCMailDate du premier document la valeur "4/20/2008 9:53:00 AM", en omettant les guillemets. Affectez à la propriété ICCMailDate du second document la valeur "1/10/2009 8:01:00 AM", en omettant les guillemets. Vérifiez que les deux dates sont comprises dans deux partitions de collecte différentes, en fonction de la configuration des partitions de collecte du conteneur d'objets.
3. Attendez que l'intervalle du répartiteur de l'extraction de contenu (CBR) soit écoulé (la valeur par défaut est 180 secondes). Ensuite, vérifiez si des collectes de zones d'index ont été créées pour la zone d'index de conteneur d'objets. Si aucune collecte de zones d'index n'a été créée, recherchez des erreurs éventuelles dans les fichiers journaux du moteur de contenu.
4. Utilisez le tableau de bord Autonomy K2 pour afficher les collectes que vous venez de créer. Il devrait y avoir deux collectes de documents. Patientez un instant pour que les index s'affichent.
5. Ouvrez la fonction Test de recherche dans le tableau de bord Autonomy K2 sur le serveur de recherche puis exécutez la requête ci-après à l'aide de l'analyseur syntaxique de requêtes en langage Verity Query :

```
textetest <dans> icc_content
```

où *textetest* représente le texte que vous avez spécifié dans l'un de vos fichiers texte .xml. FileNet P8 Content Search Engine devrait trouver les documents de test.

Information associée :

 Centre de documentation d'IBM FileNet P8 Version 5.1

Vérification de votre configuration IBM Legacy Content Search Engine pour les conteneurs d'objets qui ne sont pas partitionnés par date

Après avoir installé IBM Content Collector Server et exécuté l'assistant de configuration initiale, vérifiez votre configuration IBM Legacy Content Search Engine des conteneurs d'objets qui ne sont pas partitionnés par date à l'aide de FileNet Enterprise Manager (FEM).

Pour vérifier la configuration des conteneurs d'objets qui ne sont pas partitionnés par date :

1. Créez un document de classe ICCMailSearch2 contenant un fichier texte. Le fichier texte doit posséder l'extension .xml (par exemple searchtest.xml) et doit contenir l'instruction XML suivante :

```
<icc_content>test</icc_content>
```

Si la classe de documents ICCMailSearch2 n'existe pas dans le conteneur d'objets, vous devez réexécuter l'assistant de configuration initiale d'IBM Content Collector afin d'activer le conteneur d'objets pour l'archivage des courriers électroniques à l'aide d'IBM Content Collector.

2. Attendez que l'intervalle de répartiteur du module CBR soit écoulé (la valeur par défaut est de 180 secondes). Ensuite, vérifiez si une collecte de zones

d'index a été créée pour la zone d'index de conteneur d'objets. Si aucune zone d'index n'a été créée, recherchez des erreurs éventuelles dans les fichiers journaux du moteur de contenu.

3. Utilisez le tableau de bord Autonomy K2 pour afficher la collecte que vous venez de créer. Il doit y avoir un seul index de document. Patientez un instant avant que l'index ne s'affiche.
4. Ouvrez la fonction Test de recherche dans le tableau de bord Autonomy K2 sur le serveur de recherche, puis exécutez la requête ci-après à l'aide de l'analyseur syntaxique de requêtes en langage Verity Query :

```
test <in> icc_content
```

FileNet P8 Content Search Engine devrait trouver le document de test.

Identification des incidents d'indexation lors de l'indexation avec IBM Legacy Content Search Engine

Déterminez les objets ayant provoqué des incidents d'indexation et étant probablement la cause pour laquelle ces objets n'ont pu être indexés. Les objets qui n'ont pas été indexés ne peuvent pas être recherchés et sont introuvables dans le référentiel.

Les incidents d'indexation sont des incidents d'indexation des objets ou d'indexation des index.

Les incidents d'indexation d'objets empêchent l'indexation correcte du texte pour un objet particulier. Ces incidents peuvent être associés au document lui-même ou au système. Les incidents liés au document sont dus à la nature du texte. Par exemple, l'objet peut ne pas être indexé car le document est protégé par un mot de passe. Pour les incidents liés au document, la demande d'index n'est pas automatiquement relancée. Les incidents liés au système sont dus à l'environnement système en cours. Par exemple, une demande d'index peut échouer car aucun serveur d'index n'est actuellement disponible pour traiter la demande. Pour les incidents liés au système, la demande est relancée jusqu'à ce qu'elle aboutisse (car vous pouvez corriger les problèmes d'environnement système) ou jusqu'à ce que vous la supprimiez.

Un incident de tâche d'index n'est pas lié à une demande d'index ou à un objet particulier mais à un ensemble de demandes d'index.

Pour plus d'informations sur l'identification et la résolution des incidents liés au système ou aux tâches d'index, consultez les sections correspondantes dans la documentation FileNet P8.

Important : Les étapes de traitement suivantes ne s'appliquent pas aux versions d'IBM FileNet P8 antérieures à la Version 5.1. Pour identifier les incidents liés au document dans les versions antérieures de FileNet P8, consultez les journaux du serveur Content Engine pour les erreurs d'indexation.

Pour identifier les incidents d'indexation liés au document :

1. Pour rechercher des objets présentant des incidents d'indexation, utilisez la requête suivante afin de parcourir le conteneur d'objets.

```
SELECT d.This, d.CmIndexingFailureCode FROM doc_class d WHERE  
d.CmIndexingFailureCode > 0
```

Remplacez *doc_class* par une classe de document appropriée.

- ICCMailSearch2 pour le courrier électronique (la version indexable et interrogeable du courrier électronique, pas le courrier électronique en cours)
- ICCFileInstance2 pour les fichiers
- ICCSharepointInstance2 pour les documents Microsoft SharePoint

Cette requête parcourt la table DocVersion, qui correspond à la date de base de données pour les objets Document.

Vous pouvez paramétrer cette requête dans l'affichage simple du générateur de requête. Dans ce cas, sélectionnez les paramètres suivants :

Zone	Valeur
Sélectionner à partir de la table	ICCMailSearch2, ICCFileInstance2, ou ICCSharepointInstance2
Sélectionner des colonnes	*
Critères > Colonne	Incident d'indexation
Critères > Condition	Est supérieur à
Critères > Valeur	0

Les requêtes sur la classe ICCMailSearch2 renvoient les objets XIT qui n'ont pas pu être indexés et non les documents électroniques réels. Chacune de ces entrées dans la liste de résultats a une propriété ICCMailReference. Cette propriété renvoie au courrier électronique réel, stocké dans la classe de document ICCMail2.

Les requêtes sur la propriété CmIndexingFailureCode peuvent entraîner une analyse de table de base de données, ce qui prend beaucoup de temps lorsqu'elle est effectuée sur une table volumineuse. Un index de base de données pour la propriété CmIndexingFailureCode peut améliorer les performances de ces requêtes car le gestionnaire de base de données accède aux lignes de qualification et renvoie les données de la colonne demandée directement via l'analyse de l'index. Envisagez donc de demander à l'administrateur de bases de données d'ajouter un index à la colonne index_failure_code dans la table DocVersion.

Pour limiter le nombre de résultats, ajoutez des critères de filtrage à votre recherche, tels que la date de création du document, ou recherchez un code incident spécifique. Dans ce cas, vous devez également créer les index de base de données appropriées pour éviter des analyses de tables. Pour plus d'informations, consultez la rubrique sur la création des index de bases de données dans la documentation FileNet P8.

2. Dans la liste des résultats, affichez la colonne **CmIndexingFailureCodes** pour déterminer le type d'incident d'indexation pour l'objet. Le code d'incident d'indexation pour chaque document traité est une valeur binaire qui correspond exactement à l'une des descriptions de codes d'échec dans la liste des codes d'incident d'indexation, ou est une combinaison d'un ou plusieurs codes d'incident dans laquelle la valeur de chaque code d'incident est logiquement combinés à l'aide de l'OR de l'opérateur. Par exemple, un code d'incident d'indexation 161 est une combinaison des codes d'incident 1, 32 et 128.

Consultez la liste des codes d'incidents dans la documentation FileNet P8 pour trouver la condition ou l'événement à l'origine de l'erreur et savoir ce que vous pouvez faire pour gérer cette erreur ou empêcher une répétition de l'incident. Vous pouvez trouver cette liste dans la rubrique relative à l'onglet Général de la feuille de propriétés du conteneur d'objets.

Information associée :

 Centre de documentation d'IBM FileNet P8 Version 5.1

Réindexation de documents

Indexez à nouveau les documents pour tenter de retraiter les objets exécutés sur des erreurs spécifiques ou indexés au cours d'une période donnée avant que des modifications ou des adaptations importantes aient été apportées au système.

Après avoir regroupé vos erreurs dans des erreurs d'avertissement et des erreurs plus graves, vous pouvez choisir entre plusieurs options de réindexation afin de recréer les informations d'index pour les objets.

A faire : La réindexation d'objets ayant provoqué des avertissements ou des erreurs n'est efficace que si des configurations d'index ont été modifiées ou corrigées, par exemple, si des délais d'attente ont été ajustés ou de nouvelles mises à jour installés, si des filtres de document supplémentaires ont été ajoutés, les filtres existants ont été améliorés ou une nouvelle version du moteur de recherche de contenu a été installée. Les erreurs graves se produisent généralement si le processus d'indexation est arrêté de manière inattendue lors du traitement d'un objet car l'objet a été endommagé ou altéré de façon inattendue. Ces objets doivent être analysés et peut-être corrigés avant qu'une nouvelle tentative d'indexation soit effectuée. Des erreurs graves peuvent également se produire si le processus est à court de mémoire.

Les options de réindexation comprennent :

- Réindexation d'objets individuels
- Réindexation d'objets en tant que groupe

La réindexation nécessite toujours la recherche dans une table de base de données DocVersion pour les objets qui remplissent les conditions de réindexation demandées et peut prendre un temps considérable en fonction du nombre de documents déjà indexés. Une analyse complète de la table est nécessaire sauf si vous créez manuellement des index de base de données supplémentaires.

Utilisez Enterprise Manager pour rechercher les objets présentant des erreurs d'indexation, comme décrit dans la rubrique relative à l'identification des incidents d'indexation. Réindexez les objets en fonction des résultats de recherche :

- Pour réindexer un objet individuel, cliquez avec le bouton droit sur l'objet dans la liste des résultats et cliquez sur **Toutes les tâches > Index pour Content Search**.
- Pour réindexer un petit ensemble de documents sélectionnés manuellement, sélectionnez les objets dans la liste des résultats. Cliquez avec le bouton droit de la souris et sélectionnez **Opérations de multisélection**. Sur l'onglet Action de la fenêtre Opérations en vrac de Content Engine, sélectionnez **Index pour Content Search**. Cliquez sur **OK** pour soumettre les objets sélectionnés à la réindexation.
- Pour réindexer tous les documents qui répondent à des critères spécifiques, limitez votre requête à un code incident spécifique, par exemple. Sur l'onglet Action de Content Engine Query Builder, sélectionnez **Index pour Content Search**. Cliquez sur **OK** pour rechercher et soumettre automatiquement tous les résultats à la réindexation.
- Vous pouvez sinon écrire votre propre outil d'identification et de réindexation des objets avec des incidents d'indexation.

Remplacement d'un index

Remplacez un index lorsque vous avez modifié l'ensemble de styles pour des index IBM Legacy Content Search Engine, lorsque vous souhaitez ajouter d'autres attributs à l'index ou lorsque vous pensez que l'index actuel est endommagé.

Pour remplacer l'index des classes sélectionnées :

1. Démarrez IBM FileNet Enterprise Manager (FEM).
2. Désactivez l'extraction de contenu (CBR) pour toutes les classes de document. En règle générale, l'extraction de contenu n'est activée que pour ces classes, selon le modèle de données installé IBM Content Collector :
 - ICCMail3 (utilisé avec IBM Content Search Services uniquement)
 - ICCMailSearch2 (utilisé avec IBM Legacy Content Search Engine uniquement)
 - ICCFileInstance2
 - ICCSharepoint2

Si vous avez de classes de document personnalisées pour lesquelles l'extraction de contenu est activée, désactivez également l'extraction de contenu pour celles-ci.

- a. Dans l'arborescence Enterprise Manager, développez le noeud du conteneur d'objets et cliquez avec le bouton droit sur la classe de document.
 - b. Sélectionnez **Propriétés** dans le menu.
 - c. Dans l'onglet Général de la feuille de propriétés, décochez la case **Extraction du contenu activée**.
3. Réindexez les classes.
 - Dans IBM FileNet P8 Version 4.5, validez la tâche qui est automatiquement créée dans le gestionnaire des tâches d'index.
 - a. Cliquez avec le bouton droit sur le conteneur d'objets et sélectionnez **Toutes les tâches > Gestion des tâches d'index**.
 - b. Cliquez sur **Afficher la définition de classe Modifications activées par CBR** pour ouvrir la fenêtre Liste des définitions de classe modifiées.
 - c. Sélectionnez les définitions de classe que vous souhaitez réindexer.
 - d. Cliquez sur **Créer un travail d'index**.
 - Dans IBM FileNet P8 Version 5, cliquez avec le bouton droit sur la classe de document racine et sélectionnez **Toutes les tâches > Classe d'index pour Content Search (inclure les sous-classes)**.
 4. Surveillez la tâche d'index. Cliquez avec le bouton droit sur le conteneur d'objets et sélectionnez **Toutes les tâches > Gestion des tâches d'index**. Le statut de la tâche d'index doit finalement afficher <TERMINATED NORMALLY>. Les zones d'index existantes ne doivent plus afficher les index associés. Si des index existent toujours, vérifiez vos journaux FileNet Content Engine Server.
 5. Si l'extraction de contenu est activée pour le conteneur d'objets avec IBM Legacy Content Search Engine, vérifiez la configuration de la zone d'index pour vous assurer que le fichier de style Content Collector est utilisé à la place de FileNet_FileSystem_PushAPI (norme pour les configurations d'extraction de contenu d'IBM FileNet Content Engine).
 6. Activez à nouveau l'extraction de contenu pour toutes les classes de document pour lesquelles vous avez désactivé l'extraction de contenu à l'étape 2.
 7. Réindexez les classes.

Important : Lorsque vous réindexez la classe de document ICCMail3, assurez-vous de ne pas propager la configuration d'extraction de contenu pour

la sous-classe ICCMail Pre Index 3. Par conséquent, sélectionnez **Toutes les tâches > Classe d'index for Content Search** pour la classe de document ICCMail3.

8. Surveillez la tâche d'index. Cliquez avec le bouton droit sur le conteneur d'objets et sélectionnez **Toutes les tâches > Gestion des tâches d'index**. Le statut de la tâche d'index doit finalement afficher <TERMINATED NORMALLY>.
9. Vérifiez les zones d'index. Au moins un index doit être associé à chaque zone d'index.

Référence associée :

«Ensembles de styles IBM Legacy Content Search Engine», à la page 173

Contrôle de l'indexation avec IBM Legacy Content Search Engine

Vous pouvez améliorer les performances d'indexation en surveillant le processus d'indexation et en ajustant les paramètres de configuration d'un serveur IBM Legacy Content Search Engine en conséquence.

Surveillez les informations suivantes :

- Les fichiers journaux IBM Legacy Content Search Engine
- Les fichiers journaux Content Engine, notamment le fichier `p8_server_error.log`
- Les tables de base de données spécifiques

Fichiers journaux d'IBM Legacy Content Search Engine

IBM Legacy Content Search Engine génère des fichiers journaux pour chaque type de serveur K2. Pour les problèmes d'indexation, vérifiez les journaux du serveur d'index K2 associé à la zone d'index du conteneur d'objets utilisé pour l'archivage avec IBM Content Collector. Pour afficher les fichiers journaux, accédez au tableau de bord Autonomy K2. Dans **Affichage système**, sélectionnez le serveur d'index K2 respectif et cliquez sur **Afficher les journaux**.

IBM Legacy Content Search Engine écrit les informations de consignation concernant le serveur d'index K2 dans le fichier `status.log`, à l'emplacement suivant :

```
LCSE_installdir/Data/services/indexserver_name/log/status.log
```

où `LCSE_installdir` est le répertoire d'installation de IBM Legacy Content Search Engine et `indexserver_name` est le nom unique du serveur d'index K2 pour lequel les informations de consignation sont écrites.

Fichiers journaux d'IBM FileNet Content Engine

Content Engine consigne les erreurs d'indexation dans le fichier `p8_server_error.log` par défaut. Le serveur Content Engine fournit la consignation de trace dans le fichier `p8_server_trace.log` si vous activez le sous-système CBR pour le traçage. Pour plus de détails, consultez la section sur les fichiers journaux Content Engine dans la documentation FileNet P8.

Pour surveiller le rendement d'indexation, activez la fonction de trace CBR récapitulative. Vous pouvez ensuite analyser le fichier `p8_server_trace.log` pour savoir combien d'objets ont été envoyés au serveur IBM Legacy Content Search Engine pour l'indexation, comme dans l'exemple ci-après.

```

for filename in $(ls -r /opt/IBM/WebSphere/AppServer/profiles/AppSrv01/FileNet/server1/*trace*)
do
  grep "Dispatching" $filename > /qaTools/results/$HONAME.ce2cbrDispatch.$(basename $filename)
  grep "Deleted" $filename > /qaTools/results/$HONAME.ce2cbrDelUpdates.$(basename $filename)
  grep "NumberOfObjects" $filename > /qaTools/results/$HONAME.ce2cbrIndexing.$(basename $filename)
  grep "Dispatching" $filename >> /qaTools/results/$HONAME.ce2cbrDispatch.txt
  grep "Deleted" $filename >> /qaTools/results/$HONAME.ce2cbrDelUpdates.txt
  grep "NumberOfObjects" $filename >> /qaTools/results/$HONAME.ce2cbrIndexing.txt
done

```

Importez les résultats dans une feuille de calcul et créez des graphiques des données pour avoir une idée du débit d'indexation.

Contrôle des tables de base de données

Utilisez les requêtes de base de données pour contrôler régulièrement ces tables :

- IndexRequests
- ContentQueue

En guise de meilleure pratique, exécutez les requêtes sur ces tables dans une boucle et collectez la sortie dans un fichier journal.

Exemples de requêtes de base de données pour contrôler les tables de base de données

Utilisez les requêtes de base de données comme exemples de requêtes afin de contrôler la charge d'indexation pour les scénarios à volume élevé.

Ces exemples de requêtes de base de données fournissent les informations suivantes :

- Le nombre d'objets de document qui ont été archivés dans la classe ICCMail3.

```
-x db2 "select count(*) as Number_of_email_instances from ClassDefinition CD, Docversion D où D.object_class_id = CD.object_id etsymbolic_name='ICCMail3' for fetch only "
```
- Le nombre d'instances de chaque objet. Le nombre d'instances indique le nombre de doublons d'un document qui ont été archivés. Initialement, chaque document possède une seule instance. Pour chaque doublon, Content Collector crée une autre instance dans la classe d'objet personnalisé ICCMailInstance3.

```
-x db2 "select count(*) as Number_of_email_instances from ClassDefinition CD, G générique où G.object_class_id = CD.object_id et symbolic_name='ICCMailInstance2' for fetch only"
```
- Le nombre d'objets XIT qui ont été archivés dans la classe ICCMail2. Le nombre d'instances indique le nombre de doublons d'un document qui ont été archivés. Initialement, chaque document possède une seule instance. Pour chaque doublon, Content Collector crée une autre instance dans la classe d'objet personnalisé ICCMailInstance2.

```
-x db2 "select count(*) as Number_of_email_instances from ClassDefinition CD, Docversion D où D.object_class_id = CD.object_id et symbolic_name='ICCMailSearch2' for fetch only"
```
- Le nombre d'annotations qu'il reste à consolider pour l'indexation. Le nombre d'éléments affiche le nombre de doublons de courriers électroniques qui doivent encore être ajoutés à la classe XIT pour l'indexation.

```
db2 -x "select count(*) from ClassDefinition CD, Annotation A où A.object_class_id = CD.object_id et symbolic_name='ICCMailSearchUpdateAnnotation' for fetch only"
```
- Le nombre d'objets qui doivent toujours être indexés. Il s'agit de l'accumulation d'indexation.

```
db2 -x "select count(*) from indexrequests for fetch only"
```

- Le nombre d'objets (par exemple, courrier électronique, pièces jointes ou fichiers) qui ont été archivés dans le conteneur d'objets mais n'ont pas encore été déplacés vers leur emplacement permanent, tels que la zone de stockage fixe, ou qui ont été marqués pour suppression par le gestionnaire d'expiration mais n'ont pas encore été supprimés de la zone de stockage.

```
db2 -x "select count(*) from contentqueue for fetch only"
```

Important : Si vous prévoyez d'utiliser des requêtes plus complexes, vous devez également créer les index de base de données appropriés. Pour les propriétés système, demandez à votre administrateur de base de données de créer l'index approprié. Pour plus de détails, consultez la rubrique sur la création d'index de base de données dans la section sur l'extraction du contenu avec IBM Legacy Content Search Engine dans la documentation FileNet P8.

Indexation avec IBM Content Search Services

IBM Content Search Services est un composant de serveur évolutif qui peut être utilisé avec IBM FileNet P8 pour indexer les documents et activer la fonction de recherche.

IBM Content Search Services est une nouvelle approche de l'indexation en texte intégral optimisée pour les documents du système source pris en charge par IBM Content Collector.

IBM Content Search Services assure le traitement complet du document, y compris l'extraction de texte dans des formats binaires populaires, une vaste prise en charge du codage et le traitement de la langue.

IBM Content Collector P8 Content Search Services Support est un plug-in du constructeur de document dans IBM Content Search Services pour un prétraitement personnalisé des documents archivés à l'aide d'IBM Content Collector et pour améliorer les performances lors de l'indexation. IBM Content Collector P8 Content Search Services Support extrait le contenu du document depuis le référentiel IBM FileNet P8 et le traite avant qu'il ne soit envoyé via le pipeline de traitement de texte et d'indexation IBM Content Search Services ordinaire.

Concepts d'indexation dans IBM Content Search Services

IBM Content Search Services est configuré comme un serveur IBM Content Search Services sur IBM FileNet Content Engine.

Le serveur IBM Content Search Services contient les informations de connexion et de configuration pour un seul serveur IBM Content Search Services et est associé à un site. Tous les conteneurs d'objets du même site peuvent utiliser le serveur pour créer et interroger des index.

Selon la taille et la configuration de votre système, vous aurez besoin d'une seule ou de plusieurs instances du serveur IBM Content Search Services.

Dans une configuration monoserveur, une instance unique du serveur IBM Content Search Services effectue à la fois les tâches d'indexation et de recherche pour Content Engine. Dans une configuration à plusieurs serveurs, les instances de serveur sont affectés des rôles différents afin de répartir la charge pour les tâches d'indexation et de recherche.

IBM Content Search Services appelle IBM Content Collector P8 Content Search Services Support avec un ensemble d'ID de documents. Content Search Services Support prétraite chaque document Content Collector autre que les documents du système de fichiers de la manière suivante :

- Crée un modèle XML vide pour chaque ID de document
- Extrait le document archivé du référentiel et ajoute des éléments tels que l'objet et le corps au modèle XML
- Extrait les valeurs d'attribut du document et ajoute ces valeurs au modèle XML
- Les données des pièces jointes sont traitées par le service d'extraction de texte IBM Content Search Services et les données textuelles extraites sont ajoutées au modèle XML
- Le contenu XML généré est indexé par IBM Content Search Services. Si des avertissements ou des erreurs se produisent au cours de ce processus, ces messages sont liés au document indexé.

Un index appartient à une zone d'index et une zone d'index appartient à un conteneur d'objets.

Un index ne peut être ouvert que par un seul serveur d'index IBM Content Search Services à la fois. Content Engine affecte les index aux serveurs d'index disponibles de sorte que la charge de travail soit à peu près la même pour tous les serveurs.

Chaque fois qu'un objet est créé, mis à jour ou supprimé dans un conteneur d'objets qui a été activé pour l'indexation en texte intégral à l'aide d'IBM Content Search Services, une demande d'index est démarrée. Les demandes d'index font partie d'une tâche d'indexation qui est configurée pour s'exécuter dans les plages horaires définies. Les demandes d'index peuvent être surveillées pour identifier les incidents d'indexation à l'aide de l'IBM FileNet Enterprise Manager.

Le répartiteur du sous-système Content Engine regroupe une demande d'index avec d'autres demandes d'index pour former un index par lots. L'index cible est le même pour toutes les demandes d'index dans un lot.

Activation de l'indexation en texte intégral avec IBM Content Search Services

Pour activer l'indexation en texte intégral des documents qui sont archivés par IBM Content Collector et pouvoir accéder ultérieurement à ces documents, vous devez configurer votre système en conséquence.

Pour activer un conteneur d'objets pour la recherche avec IBM Content Collector :

1. Activez le conteneur d'objets pour l'indexation en texte intégral.
2. Installez IBM Content Collector P8 Content Search Services Support sur plusieurs serveurs IBM Content Search Services et configurez toutes les instances de ces serveurs IBM Content Search Services.
3. Vérifiez votre configuration Content Search Engine.

Activation des conteneurs d'objets pour l'indexation en texte intégral avec IBM Content Search Services

Configurez l'extraction de contenu des courriers électroniques qui ont été archivés à l'aide d'IBM Content Collector, IBM FileNet P8 Content Engine 5.1 et version ultérieure, et IBM Content Search Services.

Conditions requises :

- Un environnement IBM FileNet P8 Version 5.1 installé et configuré avec IBM FileNet Content Engine Version 5.1 et IBM FileNet Content Search Engine Client version 5.1 avec IBM Content Search Services comme option de moteur de recherche de contenu. Pour plus d'informations, consultez la section relative à l'installation et à la configuration IBM Content Search Services dans la documentation FileNet P8.

Important :

Si vous effectuez une mise à niveau à partir de IBM FileNet Content Search Engine Client version 5.0 vers IBM FileNet Content Search Engine Client version 5.1, vous devez modifier manuellement le paramètre suivant dans le fichier `mimetypes.xml` situé dans le dossier `<ContentSearchServices_Installation>/config`. Le chemin par défaut est `opt/IBM/ContentSearchServices/CSS_Server/config/mimetypes.xml`.

Modification :

```
<MappedMimetype Name="text/xml">
  <Extension>.xml</Extension>
</MappedMimetype>
```

en :

```
<MappedMimetype Name="text/plain">
  <Extension>.xml</Extension>
</MappedMimetype>
```

Cette action n'est pas requise s'il s'agit d'une première installation de IBM FileNet Content Search Engine Client version 5.1.

- Un conteneur d'objets FileNet P8 vide utilisant la zone de stockage de fichier comme conteneur de contenu par défaut.

Important : Vous ne pouvez pas utiliser un conteneur d'objets dans lequel la classe de document ICCMail2 est installée.

Pour plus d'informations sur la création d'un nouveau conteneur d'objets, consultez la section relative à l'administration de Content Engine dans la documentation FileNet P8.

Lorsque vous créez un nouveau conteneur d'objets, assurez-vous que seul l'additif **Base Content Engine Extension** est installé. Vous pouvez pour cela sélectionner **Configuration minimale**. L'installation d'additifs crée plusieurs classes de documents et index de bases de données qui ne sont pas requis par Content Collector, ce qui affecte la performance.

Pour activer un conteneur d'objets pour Content Collector :

1. Suivez la procédure suivante dans IBM FileNet Enterprise Manager.
 - a. Accédez à l'onglet CBR IBM Search de la feuille de propriétés du conteneur d'objets et définissez la langue d'index. Sélectionnez un code de langue autre que `auto`, de préférence le code de la langue dans laquelle la plupart des documents seront écrits. Assurez-vous que **IBM Content Search Services activé** n'est pas sélectionné.
 - b. Créez une zone d'index pour le conteneur d'objets Content Collector.
 - c. Retournez à l'onglet CBR IBM Search de la feuille de propriété du conteneur d'objets, sélectionnez **IBM Content Search Services activé** et cliquez sur **OK**.

2. Installez le modèle de données IBM Content Collector pour le conteneur d'objets. Sur la machine serveur Content Collector, exécutez l'outil de configuration Content Collector pour l'activation du référentiel comme décrit dans «Configuration d'un référentiel IBM FileNet P8». Passez ensuite à l'étape suivante de la procédure.
3. Retournez à IBM FileNet Enterprise Manager et vérifiez que le modèle de données est installé avec les paramètres appropriés.
 - a. Dans l'onglet Général de la feuille de propriétés du conteneur d'objets, **Propager à la source** doit être défini pour le **Niveau d'enregistrement des incidents d'indexation**.
 - b. Dans l'onglet CBR IBM Search de la feuille de propriétés du conteneur d'objets, vérifiez l'intervalle défini pour la propriété de date ICCMailDate.

Important : Si vous modifiez l'intervalle pour un conteneur d'objets qui n'est pas vide, vous devez réindexer ce conteneur pour que le paramètre prenne effet pour les courriers électroniques déjà archivés.

Le conteneur d'objets est maintenant configuré pour l'archivage avec Content Collector et l'indexation en texte intégral avec IBM Content Search Services.

Installez maintenant IBM Content Collector P8 Content Search Services Support.

Information associée :

 Centre de documentation d'IBM FileNet P8 Version 5.1

Configuration d'un référentiel IBM FileNet P8 :

Tous les documents archivés à l'aide d'IBM Content Collector sont stockés dans un magasin d'objets d'IBM FileNet P8.

Lorsque vous utilisez un conteneur d'objets activé pour l'extraction basée sur le contenu (CBR), le conteneur doit être dédié à l'archivage avec IBM Content Collector car vous ne pouvez utiliser qu'une seule configuration CBR pour un conteneur d'objets. Il doit être vide pour être configuré pour IBM Content Collector et ne peut pas être utilisé à d'autres fins.

Il n'est pas nécessaire de configurer différents magasins d'objets pour chaque système source dans IBM FileNet P8. Par exemple, vous pouvez configurer un conteneur d'objets pour le courrier électronique, le système de fichiers et les documents SharePoint. Cette approche de conteneur d'objets unique présente des avantages par rapport au dédoublement car, dans un conteneur d'objets unique, les fichiers en double provenant de sources multiples (pièces jointes de courrier électronique et documents de système de fichiers ou SharePoint) peuvent être détectés, réduisant ainsi les besoins en stockage.

Lorsque vous activez la recherche de texte pour vos classes de document, IBM Content Collector utilise une configuration d'indexation personnalisée pour permettre l'indexation à granularité fine des documents.

Vous pouvez configurer un conteneur d'objets en mode interface graphique ou en mode console.

Installation d'IBM Content Collector P8 Content Search Services Support

IBM Content Collector P8 Content Search Services Support est un plug-in de constructeur de document facultatif dans IBM Content Search Services pour un

prétraitement personnalisé de tous les documents archivés à l'aide d'IBM Content Collector autres que les documents de système de fichiers.

Avant de commencer :

- Vérifiez que IBM FileNet Content Search Engine Client version 5.1 est installé.
- Dans tous les environnements UNIX et sous Linux et Linux on System z, assurez-vous que J2SE 1.4 ou version ultérieure est installé. IBM Content Collector P8 Content Search Services Support requiert que l'environnement d'exécution Java soit installé sur le serveur.
- Dans tous les environnements UNIX et sous Linux et Linux on System z, assurez-vous que la variable d'environnement PATH contient java (version 1.4 ou ultérieure).
- Dans tous les environnements UNIX, il est recommandé d'utiliser les mêmes données d'identification pour installer le client IBM FileNet Content Engine, IBM Content Search Services et IBM Content Collector P8 Content Search Services Support.

Vous pouvez installer IBM Content Collector P8 Content Search Services Support en mode interface graphique, en mode console ou en mode silencieux.

Lorsque vous installez IBM Content Collector P8 Content Search Services Support, vous pouvez sélectionner les documents de système source IBM Content Collector suivants pour le prétraitement :

- Documents Courrier électronique
- Documents Microsoft SharePoint
- Documents IBM Connections

Sélectionnez les mêmes documents source que ceux sélectionnés lorsque vous avez configuré le magasin d'objets pour IBM Content Collector à l'aide de l'outil de configuration Content Collector pour l'activation de référentiel.

Il n'est pas nécessaire de prétraiter les documents de système de fichiers par Content Search Services Support avant qu'ils ne soient envoyés aux serveurs d'index IBM Content Search Services pour l'indexation.

Content Search Services Support est exécuté sur le serveur IBM FileNet Content Search Engine sur lequel est installé IBM Content Search Services.

Content Search Services Support peut être installé sur Windows, les machines UNIX, Solaris, Linux et Linux on System z.

Le programme d'installation de Content Search Services Support installe Content Search Services Support dans un dossier qui ne se trouve pas l'arborescence de dossiers IBM Content Search Services et déploie ensuite ces fichiers dans le dossier d'instance d'installation de IBM Content Search Services.

L'utilisation d'un dossier d'installation distinct permet à IBM Content Collector de contrôler les fichiers Content Search Services Support, indépendamment des toutes les instances IBM Content Search Services installées.

Avant d'installer Content Search Services Support, une copie de sauvegarde de la configuration existante est réalisée. Ce fichier est appelé `constructors.xml.ICC.backup` et il se trouve dans le répertoire `repInstallContentSearchServices\config\`.

A partir de Content Collector V3.0, les options de configuration Content Search Services Support qui sont nécessaires pour indexer les éléments à l'aide de IBM Content Search Services sont stockées dans la base de données de configuration IBM Content Collector et peuvent être définies et modifiées à l'aide du Gestionnaire de configuration IBM Content Collector . Le fichier `constructors.xml` ne contient plus les options de configuration, il contient désormais le lien vers la base de données de configuration dans laquelle la configuration est stockée. Avant Content Collector V3.0, toutes les options de configuration Content Search Services Support été stockées directement dans le fichier `constructors.xml`.

Installation de Content Search Services Support en mode interface graphique :

Procédez comme suit pour installer IBM Content Collector P8 Content Search Services Support sur le serveur IBM FileNet Content Search Engine sur lequel est installé IBM Content Search Services.

Pour installer Content Search Services Support à l'aide de l'assistant d'installation :

1. Arrêtez l'instance IBM Content Search Services.
2. Insérez le CD-ROM du produit sur le serveur IBM FileNet Content Search Engine sur lequel est installé IBM Content Search Services. Vous pouvez également télécharger et extraire le module d'installation approprié.
3. Démarrez le programme d'installation :
 - Dans les environnements Linux, Linux on System z et UNIX, exécutez **install.bin** qui se trouve dans le répertoire `/P8_CSS_Support/système_exploitation.système_exploitation` correspond au nom du système d'exploitation, par exemple, Linux.
 - Sous Windows, exécutez **install.exe** qui se trouve dans le répertoire `\\P8_CSS_Support\WIN`.
4. Répondez aux invites restantes à mesure que le programme d'installation vous guide tout au long du processus.
 - a. Indiquez le dossier d'installation dans lequel est installé Content Search Services Support ou utilisez le dossier par défaut.

Le chemin par défaut est `C:\Program Files\IBM\ContentCollector_CSS_Support` pour Windows et `/opt/IBM/ContentCollector_CSS_Support` dans les environnements UNIX, Linux et Linux on System z.
 - b. Spécifiez le chemin d'accès complet du dossier contenant l'instance IBM Content Search Services sur le serveur.

Le chemin par défaut est `C:\Program Files\IBM\Content Search Services\CSS Server` pour Windows et `/opt/IBM/ContentSearchServices/CSS_Server` dans les environnements UNIX, Linux et Linux on System z.

Une fois que le programme d'installation a localisé l'instance d'installation IBM Content Search Services, Content Search Services Support est déployé dans le dossier `\plugins` de l'arborescence de dossiers de l'instance d'installation IBM Content Search Services.

Le programme d'installation de Content Search Services Support ajoute la configuration de chaque source de document sélectionnée au fichier de configuration IBM Content Search Services `<repInstallContentSearchServices>\config\constructors.xml`. La configuration de Content Search Services Support est divisée en sections distinctes, une pour chaque source de document sélectionnée.

Le fichier de configuration contient le nom du préprocesseur du document, la classe de document, les informations de chemin de la base de données de configuration Content Collector qui stocke les données de configuration et le nom symbolique.

Le nom symbolique indique les classes de document qui sont traitées par Content Search Services Support.

Il est défini par IBM Content Collector lorsque le magasin d'objets est configuré pour IBM Content Search Services. Vous ne devez pas modifier la valeur insérée par IBM Content Collector.

Si vous utilisez des noms de classe de document personnalisée, vous devez les ajouter manuellement au fichier `<repInstallContentSearchServices>\config\constructors.xml` dans chaque instance de IBM Content Search Services sur le serveur IBM FileNet Content Search Engine. Les noms de classe de document doivent être séparés par une virgule, par exemple,

```
SymbolicClassName=ICCMail3,ICCMailCustomClassName,CustomerEmailJournaling
```

Vous pouvez utiliser des expressions régulières pour indiquer des noms de classe de document lorsque vous spécifiez **SymbolicClassName**. Le nom du magasin d'objets peut être ajouté en préfixe aux noms de classe de document pour limiter la configuration des classes de document spécifiques à un seul magasin d'objets. Par exemple,

```
SymbolicClassName=EmailArchivingOS1:ICCMail3_Group_[0-9_]+,JournalArchivingOS2:ICCMailJournals_[0-9_]+
```

A partir de Content Search Services Support pour Content Collector V3, la configuration de Content Search Services Support est stockée dans la base de données de configuration Content Collector, à laquelle vous pouvez accéder via le Gestionnaire de configuration IBM Content Collector.

Si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support V3 et que vous avez modifié la configuration du fichier `constructors.xml` que vous souhaitez conserver, par exemple, vous avez adapté les options de configuration ou ajouté des attributs personnalisés, vous devez ajouter ces modifications manuellement après l'installation à l'aide du Gestionnaire de configuration. Vous pouvez copier vos modifications depuis le fichier de configuration de sauvegarde `repInstallContentSearchServices\config\constructors.xml.ICC.backup` qui est créé lorsque vous installez Content Search Services Support pour Content Collector.

Il est recommandé de stocker la configuration de Content Search Services Support dans la base de données Content Collector. Cependant, si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support et que ne souhaitez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez remplacer le fichier `constructors.xml` récemment installé par la copie de sauvegarde qui a été créée pendant l'installation. Notez que vous devez remplacer le fichier de configuration sur toutes les instances IBM Content Search Services. De plus, si vous avez utilisé l'option de configuration ***DumpDocuments** à des fins de débogage dans votre installation antérieure, vous devez modifier le fichier et remplacer l'option par **DumpDocumentIds**.

Important : Le fichier `constructors.xml` est un fichier XML avec une structure imbriquée qui ne peut en aucun cas être modifiée, par exemple, aucun renvoi à la

ligne n'est autorisé. Etant donné que le fichier n'est pas strictement un fichier XML, vous ne pouvez pas l'ouvrir et le modifier à l'aide d'un éditeur XML.

Si vous installez Content Search Services Support pour la première fois et ne voulez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez installer Content Search Services Support et utiliser des valeurs factices lors de la spécification des données d'identification du service d'application Web. Si vous devez configurer les options de Content Search Services Support, vous devez le faire directement dans le fichier <répInstallContentSearchServices>\config\constructors.xml. Pour plus d'informations sur la procédure à suivre, contactez le service de support logiciel IBM.

Installation de Content Search Services Support en mode console :

Pour installer IBM Content Collector P8 Content Search Services Support en mode console, suivez la procédure ci-après.

Pour installer Content Search Services Support en mode console sur le serveur Content Search Engine sur lequel est installé IBM Content Search Services :

1. Arrêtez l'instance IBM Content Search Services.
2. Ouvrez une fenêtre d'invite de commande et démarrez le programme d'installation en mode console :
 - Dans les environnements Linux, Linux on System z et UNIX, exécutez :

```
install.bin -i console
```

qui se trouve dans le répertoire /P8_CSS_Support/système_exploitation. *système_exploitation* correspond au nom du système d'exploitation, par exemple, Linux.
 - Sous Windows, exécutez :

```
install.exe -i console
```

qui se trouve dans le répertoire \\P8_CSS_Support\WIN.
3. Suivez les instructions fournies dans la fenêtre d'invite de commande.
 - a. Indiquez un dossier d'installation dans lequel les fichiers IBM Content Collector P8 Content Search Services Support sont installés ou utilisez le dossier par défaut.

Le chemin par défaut est C:\Program Files\IBM\ContentCollector_CSS_Support pour Windows et /opt/IBM/ContentCollector_CSS_Support dans les environnements UNIX et Linux.
 - b. Spécifiez le chemin d'accès complet du dossier contenant une instance IBM Content Search Services sur le serveur.

Le chemin par défaut est C:\Program Files\IBM\Content Search Services\CSS Server pour Windows et /opt/IBM/ContentSearchServices/CSS_Server dans les environnements UNIX et Linux.

Une fois que le programme d'installation a localisé l'instance d'installation IBM Content Search Services, Content Search Services Support est déployé dans le dossier \plugins de l'arborescence de dossiers de l'instance d'installation IBM Content Search Services.

Le programme d'installation de Content Search Services Support ajoute la configuration de chaque source de document sélectionnée au fichier de configuration IBM Content Search Services <répInstallContentSearchServices>\

config\constructors.xml. La configuration de Content Search Services Support est divisée en sections distinctes, une pour chaque source de document sélectionnée.

Le fichier de configuration contient le nom du préprocesseur du document, la classe de document, les informations de chemin de la base de données de configuration Content Collector qui stocke les données de configuration et le nom symbolique.

Le nom symbolique indique les classes de document qui sont traitées par Content Search Services Support.

Il est défini par IBM Content Collector lorsque le magasin d'objets est configuré pour IBM Content Search Services. Vous ne devez pas modifier la valeur insérée par IBM Content Collector.

Si vous utilisez des noms de classe de document personnalisée, vous devez les ajouter manuellement au fichier <repInstallContentSearchServices>\config\constructors.xml dans chaque instance de IBM Content Search Services sur le serveur IBM FileNet Content Search Engine. Les noms de classe de document doivent être séparés par une virgule, par exemple,

```
SymbolicClassName=ICCMail3,ICCMailCustomClassName,CustomerEmailJournaling
```

Vous pouvez utiliser des expressions régulières pour indiquer des noms de classe de document lorsque vous spécifiez **SymbolicClassName**. Le nom du magasin d'objets peut être ajouté en préfixe aux noms de classe de document pour limiter la configuration des classes de document spécifiques à un seul magasin d'objets. Par exemple,

```
SymbolicClassName=EmailArchiving0S1:ICCMail3_Group_[0-9_]+,JournalArchiving0S2:ICCMailJournals_[0-9_]+
```

A partir de Content Search Services Support pour Content Collector V3, la configuration de Content Search Services Support est stockée dans la base de données de configuration Content Collector, à laquelle vous pouvez accéder via le Gestionnaire de configuration IBM Content Collector.

Si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support V3 et que vous avez modifié la configuration du fichier constructors.xml que vous souhaitez conserver, par exemple, vous avez adapté les options de configuration ou ajouté des attributs personnalisés, vous devez ajouter ces modifications manuellement après l'installation à l'aide du Gestionnaire de configuration. Vous pouvez copier vos modifications depuis le fichier de configuration de sauvegarde *repInstallContentSearchServices\config\constructors.xml.ICC.backup* qui est créé lorsque vous installez Content Search Services Support pour Content Collector.

Il est recommandé de stocker la configuration de Content Search Services Support dans la base de données Content Collector. Cependant, si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support et que ne souhaitez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez remplacer le fichier constructors.xml récemment installé par la copie de sauvegarde qui a été créée pendant l'installation. Notez que vous devez remplacer le fichier de configuration sur toutes les instances IBM Content Search Services. De plus, si vous avez utilisé l'option de configuration ***DumpDocuments** à des fins de débogage dans votre installation antérieure, vous devez modifier le fichier et remplacer l'option par **DumpDocumentIds**.

Important : Le fichier `constructors.xml` est un fichier XML avec une structure imbriquée qui ne peut en aucun cas être modifiée, par exemple, aucun renvoi à la ligne n'est autorisé. Etant donné que le fichier n'est pas strictement un fichier XML, vous ne pouvez pas l'ouvrir et le modifier à l'aide d'un éditeur XML.

Si vous installez Content Search Services Support pour la première fois et ne voulez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez installer Content Search Services Support et utiliser des valeurs factices lors de la spécification des données d'identification du service d'application Web. Si vous devez configurer les options de Content Search Services Support, vous devez le faire directement dans le fichier `<répInstallContentSearchServices>\config\constructors.xml`. Pour plus d'informations sur la procédure à suivre, contactez le service de support logiciel IBM.

Installation de Content Search Services Support en mode silencieux :

Procédez comme suit pour installer IBM Content Collector P8 Content Search Services Support en mode silencieux sur le serveur IBM FileNet Content Search Engine sur lequel est installé IBM Content Search Services.

Pour installer Content Search Services Support en mode silencieux :

1. Créez un fichier de réponses avec les options d'installation suivantes. Les valeurs sont fournies à titre d'exemple uniquement.

```
#Si la licence a été acceptée
#-----
LICENSE_ACCEPTED=TRUE
#Indiquer le dossier d'installation
#-----
# Pour Windows :
USER_INSTALL_DIR=C:\\Program Files\\IBM\\ContentCollector_CSS_Support
# Pour les environnements UNIX et Linux :
#USER_INSTALL_DIR=/opt/IBM/ContentCollector_CSS_Support
#Obtenir l'entrée utilisateur
#-----
# Indiquer un dossier d'installation dans lequel les fichiers
# IBM Content Collector P8 Content Search Service
# Support sont installés ou utiliser le
# dossier par défaut.
# Indiquer un dossier Content Search Services
#-----
# Pour Windows :
ECTMS_INSTALL_DIRECTORY=C:\\Program Files\\IBM\\Content Search Services\\CSS Server
# Pour les environnements UNIX et Linux :
#ECTMS_INSTALL_DIRECTORY=/opt/IBM/ContentSearchServices/CSS_Server
# Indiquer le chemin d'accès complet du dossier
# qui contient l'instance IBM Content Search Services
# sur le serveur IBM FileNet Content Search
# Engine ou utiliser le dossier par défaut.
# Sélectionner les documents de système source. Définir la
# valeur sur true ou false.
#-----
SOURCE_EMAIL_CONFIGURATION=true
SOURCE_SHAREPOINT_CONFIGURATION=true
SOURCE_CONNECTIONS_CONFIGURATION=true
# Indiquer le nom d'hôte et le port de la machine
# qui exécute les applications Web IBM Content
# Collector.
#-----
ICC_HOST_NAME=<aHostname>
ICC_PORT=>aPortNumber
```

2. Enregistrez le fichier de réponses sur votre disque.
3. Arrêtez l'instance IBM Content Search Services.

4. Insérez le CD-ROM du produit sur le serveur IBM FileNet Content Search Engine sur lequel est installé IBM Content Search Services. Vous pouvez également télécharger et extraire le module d'installation approprié.
5. Ouvrez une fenêtre d'invite de commande et démarrez le programme d'installation en mode silencieux :

- Dans les environnements Linux, Linux on System z et UNIX, exécutez :
`install.bin -i SILENT -f <chemin_complet_fichier_réponses>`

qui se trouve dans le répertoire `/P8_CSS_Support/système_exploitation`. `système_exploitation` correspond au nom du système d'exploitation, par exemple, AIX, Linux ou Linux on System z.

- Sous Windows, exécutez :
`install.exe -i SILENT -f <chemin_complet_fichier_réponses>`

qui se trouve dans le répertoire `\\P8_CSS_Support\WIN`.

`<chemin_complet_au_fichier_réponses>` est le chemin d'accès complet au fichier de réponses, par exemple `c:\temp\monFichierRéponses.txt`.

Important : Indiquez SILENT en majuscules.

Le programme d'installation de Content Search Services Support ajoute la configuration de chaque source de document sélectionnée au fichier de configuration IBM Content Search Services `<répInstallContentSearchServices>\config\constructors.xml`. La configuration de Content Search Services Support est divisée en sections distinctes, une pour chaque source de document sélectionnée.

Le fichier de configuration contient le nom du préprocesseur du document, la classe de document, les informations de chemin de la base de données de configuration Content Collector qui stocke les données de configuration et le nom symbolique.

Le nom symbolique indique les classes de document qui sont traitées par Content Search Services Support.

Il est défini par IBM Content Collector lorsque le magasin d'objets est configuré pour IBM Content Search Services. Vous ne devez pas modifier la valeur insérée par IBM Content Collector.

Si vous utilisez des noms de classe de document personnalisée, vous devez les ajouter manuellement au fichier `<répInstallContentSearchServices>\config\constructors.xml` dans chaque instance de IBM Content Search Services sur le serveur IBM FileNet Content Search Engine. Les noms de classe de document doivent être séparés par une virgule, par exemple,

`SymbolicClassName=ICCMail3,ICCMailCustomClassName,CustomerEmailJournaling`

Vous pouvez utiliser des expressions régulières pour indiquer des noms de classe de document lorsque vous spécifiez **SymbolicClassName**. Le nom du magasin d'objets peut être ajouté en préfixe aux noms de classe de document pour limiter la configuration des classes de document spécifiques à un seul magasin d'objets. Par exemple,

`SymbolicClassName=EmailArchivingOS1:ICCMail3_Group_[0-9_]+,JournalArchivingOS2:ICCMailJournalS_[0-9_]+`

A partir de Content Search Services Support pour Content Collector V3, la configuration de Content Search Services Support est stockée dans la base de

données de configuration Content Collector, à laquelle vous pouvez accéder via le Gestionnaire de configuration IBM Content Collector.

Si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support V3 et que vous avez modifié la configuration du fichier `constructors.xml` que vous souhaitez conserver, par exemple, vous avez adapté les options de configuration ou ajouté des attributs personnalisés, vous devez ajouter ces modifications manuellement après l'installation à l'aide du Gestionnaire de configuration. Vous pouvez copier vos modifications depuis le fichier de configuration de sauvegarde `répInstallContentSearchServices\config\constructors.xml.ICC.backup` qui est créé lorsque vous installez Content Search Services Support pour Content Collector.

Il est recommandé de stocker la configuration de Content Search Services Support dans la base de données Content Collector. Cependant, si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support et que ne souhaitez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez remplacer le fichier `constructors.xml` récemment installé par la copie de sauvegarde qui a été créée pendant l'installation. Notez que vous devez remplacer le fichier de configuration sur toutes les instances IBM Content Search Services. De plus, si vous avez utilisé l'option de configuration ***DumpDocuments** à des fins de débogage dans votre installation antérieure, vous devez modifier le fichier et remplacer l'option par **DumpDocumentIds**.

Important : Le fichier `constructors.xml` est un fichier XML avec une structure imbriquée qui ne peut en aucun cas être modifiée, par exemple, aucun renvoi à la ligne n'est autorisé. Étant donné que le fichier n'est pas strictement un fichier XML, vous ne pouvez pas l'ouvrir et le modifier à l'aide d'un éditeur XML.

Si vous installez Content Search Services Support pour la première fois et ne voulez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez installer Content Search Services Support et utiliser des valeurs factices lors de la spécification des données d'identification du service d'application Web. Si vous devez configurer les options de Content Search Services Support, vous devez le faire directement dans le fichier `<répInstallContentSearchServices>\config\constructors.xml`. Pour plus d'informations sur la procédure à suivre, contactez le service de support logiciel IBM.

Installation de Content Search Services Support sur plusieurs instances de IBM Content Search Services :

Si plusieurs instances de IBM Content Search Services sont installées sur le serveur IBM FileNet Content Search Engine, Content Search Services Support doit être installé sur chaque instance.

Procédez comme suit pour installer IBM Content Collector P8 Content Search Services Support sur chaque instance de IBM Content Search Services sur le serveur IBM FileNet Content Search Engine.

Exécutez l'assistant d'installation de Content Search Services Support pour chaque instance sur le serveur, ou exécutez l'assistant d'installation une fois pour une instance, puis exécutez le script fourni pour toutes les autres instances sur le serveur.

Option	Description
<p>Exécutez l'assistant d'installation pour toutes les instances</p>	<p>Installez Content Search Services Support sur chaque instance IBM Content Search Services du serveur.</p>
<p>Exécutez l'assistant d'installation pour une instance et le script fourni pour toutes les autres instances</p>	<ol style="list-style-type: none"> 1. Installez Content Search Services Support sur la première instance IBM Content Search Services. 2. Exécutez le script de déploiement pour toutes les instances supplémentaires sur le serveur. Le script se trouve dans le répertoire dans lequel vous avez installé Content Search Services Support à l'aide de l'assistant d'installation. Le chemin par défaut est C:\Program Files\IBM\ContentCollector_CSS_Support sous Windows et /opt/IBM/ContentCollector_CSS_Support dans les environnements UNIX et Linux. <p>Arrêtez chaque instance IBM Content Search Services avant d'exécuter le script.</p> <p>Chaque fois que vous exécutez le script de déploiement, vous devez transmettre le chemin du nom de fichier complet à l'instance. Si vous ne transmettez pas le chemin à une instance, le script utilise le chemin d'installation d'instance spécifié lors de l'installation de Content Search Services Support.</p> <ul style="list-style-type: none"> • Sous Windows, exécutez : <pre> deploy.bat "C:\Program Files\IBM\ Content Search Services\ <CSS Server New Instance>\plugins" </pre> • Sur les machines UNIX, Linux et Linux on System z, exécutez : <pre> deploy.sh "/opt/IBM/ContentSearchServices/ <CSS_Server_New_Instance>/plugins" </pre>

Le programme d'installation de Content Search Services Support ajoute la configuration de chaque source de document sélectionnée au fichier de configuration IBM Content Search Services <repInstallContentSearchServices>\config\constructors.xml. La configuration de Content Search Services Support est divisée en sections distinctes, une pour chaque source de document sélectionnée.

Le fichier de configuration contient le nom du préprocesseur du document, la classe de document, les informations de chemin de la base de données de configuration Content Collector qui stocke les données de configuration et le nom symbolique.

Le nom symbolique indique les classes de document qui sont traitées par Content Search Services Support.

Il est défini par IBM Content Collector lorsque le magasin d'objets est configuré pour IBM Content Search Services. Vous ne devez pas modifier la valeur insérée par IBM Content Collector.

Si vous utilisez des noms de classe de document personnalisée, vous devez les ajouter manuellement au fichier `<répInstallContentSearchServices>\config\constructors.xml` dans chaque instance de IBM Content Search Services sur le serveur IBM FileNet Content Search Engine. Les noms de classe de document doivent être séparés par une virgule, par exemple,

```
SymbolicClassName=ICCMail3,ICCMailCustomClassName,CustomerEmailJournaling
```

Vous pouvez utiliser des expressions régulières pour indiquer des noms de classe de document lorsque vous spécifiez **SymbolicClassName**. Le nom du magasin d'objets peut être ajouté en préfixe aux noms de classe de document pour limiter la configuration des classes de document spécifiques à un seul magasin d'objets. Par exemple,

```
SymbolicClassName=EmailArchiving0S1:ICCMail3_Group_[0-9_]+,JournalArchiving0S2:ICCMailJournals_[0-9_]+
```

A partir de Content Search Services Support pour Content Collector V3, la configuration de Content Search Services Support est stockée dans la base de données de configuration Content Collector, à laquelle vous pouvez accéder via le Gestionnaire de configuration IBM Content Collector.

Si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support V3 et que vous avez modifié la configuration du fichier `constructors.xml` que vous souhaitez conserver, par exemple, vous avez adapté les options de configuration ou ajouté des attributs personnalisés, vous devez ajouter ces modifications manuellement après l'installation à l'aide du Gestionnaire de configuration. Vous pouvez copier vos modifications depuis le fichier de configuration de sauvegarde `répInstallContentSearchServices\config\constructors.xml.ICC.backup` qui est créé lorsque vous installez Content Search Services Support pour Content Collector.

Il est recommandé de stocker la configuration de Content Search Services Support dans la base de données Content Collector. Cependant, si vous réalisez une mise à niveau vers Content Search Services Support à partir d'une version antérieure à Content Search Services Support et que ne souhaitez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez remplacer le fichier `constructors.xml` récemment installé par la copie de sauvegarde qui a été créée pendant l'installation. Notez que vous devez remplacer le fichier de configuration sur toutes les instances IBM Content Search Services. De plus, si vous avez utilisé l'option de configuration ***DumpDocuments** à des fins de débogage dans votre installation antérieure, vous devez modifier le fichier et remplacer l'option par **DumpDocumentIds**.

Important : Le fichier `constructors.xml` est un fichier XML avec une structure imbriquée qui ne peut en aucun cas être modifiée, par exemple, aucun renvoi à la ligne n'est autorisé. Etant donné que le fichier n'est pas strictement un fichier XML, vous ne pouvez pas l'ouvrir et le modifier à l'aide d'un éditeur XML.

Si vous installez Content Search Services Support pour la première fois et ne voulez pas stocker les données de configuration de Content Search Services Support dans la base de données de configuration Content Collector, vous pouvez installer Content Search Services Support et utiliser des valeurs factices lors de la spécification des données d'identification du service d'application Web. Si vous

devez configurer les options de Content Search Services Support, vous devez le faire directement dans le fichier <répInstallContentSearchServices>\config\constructors.xml. Pour plus d'informations sur la procédure à suivre, contactez le service de support logiciel IBM.

Configuration du Content Search Services Support

Les options de configuration Content Search Services Support qui sont nécessaires pour indexer les éléments à l'aide de l' IBM Content Search Services sont stockées dans le Gestionnaire de configuration IBM Content Collector.

Les valeurs de configuration sont remplies par IBM Content Collector lorsque le conteneur d'objets est configuré pour IBM Content Search Services et n'a généralement pas besoin d'être modifié ou réinitialisé, sauf si, par exemple, les paramètres changent dans votre environnement IBM FileNet P8.

Si vous souhaitez spécifier des données supplémentaires (attributs et métadonnées de document) ou des attributs IBM FileNet P8 à inclure lors de la génération de l'index, vous devez ajouter ces attributs personnalisés et définitions d'attributs d'archivage à la configuration du document source dans le Gestionnaire de configuration sous **Paramètres généraux > Content Search Services Support** avant de commencer la première indexation. Vous ne pouvez pas modifier les paramètres d'attribut une fois que l'index a été généré.

| Une fonctionnalité de journalisation et de traçage séparée comportant des fichiers
| journaux distincts existe pour IBM Content Search Services et Content Search
| Services Support. Lorsque vous activez la journalisation et du traçage dans le IBM
| FileNet Enterprise Manager pour IBM Content Search Services, ces fonctions sont
| également activées pour Content Search Services Support. Aucune information de
| journalisation pour Content Search Services Support n'est ajoutée aux fichiers de
| journalisation et de trace IBM Content Search Services.

Si vous souhaitez activer les options de journalisation, de trace et de vidage dans Content Search Services Support aux fins de débogage, ou que vous avez besoin d'adresser les fichiers journaux au service de support logiciel IBM, vous devez activer les options respectives dans le Gestionnaire de configuration. La journalisation et le traçage dans Content Search Services Support ne peuvent pas être activés à l'aide de la IBM FileNet Enterprise Manager.

Une fois que vous avez modifié la configuration Content Search Services Support, veuillez à redémarrer IBM Content Search Services pour que vos modifications prennent effet.

Paramètres de journal communs pour Content Search Services Support :

Les options de journal, de trace et de synchronisation pour Content Search Services Support sont communes à tous les préprocesseurs de documents sources configurés et lorsqu'elles sont activées, s'appliquent à tous les documents qui sont prétraités par Content Search Services Support.

| **Important :** L'activation de la journalisation et du traçage dans le IBM FileNet
| Enterprise Manager pour IBM Content Search Services ne permet pas d'activer la
| consignation et le traçage pour Content Search Services Support. Des fichiers
| journaux et de trace distincts existent pour IBM Content Search Services et Content
| Search Services Support.

Vous pouvez activer la journalisation, le traçage et les options de temporisation pour Content Search Services Support dans le Gestionnaire de configuration IBM Content Collector sous **Paramètres généraux > Content Search Services Support**.

Etant donné que la journalisation et le traçage ne peuvent pas être activés uniquement pour les courriers électroniques, les documents IBM Connections ou Microsoft SharePoint, les fichiers journaux peuvent contenir un mélange de tous les documents sources et devenir très volumineux. Pour vous assurer que les fichiers journaux et de trace surveillent les erreurs pour un seul type de document source, activez la journalisation et le traçage uniquement lorsqu'une route de tâches pour un document source est en cours de traitement par Content Search Services Support.

Les paramètres de journal incluent :

- **Niveau de journalisation :**

- **Information :** écrit les informations de journal dans le fichier journal à l'emplacement indiqué.

Les fichiers journaux utilisent des horodatages dans leurs noms. Vous devez conserver l'horodatage par défaut pour les noms de fichier journal afin qu'il y ait un nouveau jeu de fichiers journaux à chaque fois que le Content Search Services Support s'exécute.

- **Trace :** écrit des informations de trace détaillées dans les fichiers journaux. Les noms de fichier journal contenant des informations de trace utilisent également des horodatages dans les noms de fichier.

- **Désactivé :** désactive la consignation et le traçage.

- **Nombre de fichiers de trace :** indique le nombre de fichiers de trace générés. Par défaut, seul un fichier de trace contenant toutes les informations de trace est généré.

- **Taille limite en Mo :** indique la taille maximale d'un fichier de trace en mégaoctets. Dès que le fichier de trace généré atteint la taille spécifiée, un nouveau fichier est créé. La taille par défaut est 0, ce qui signifie qu'il n'y a aucune limite de taille de fichier.

- **Emplacement du fichier journal :** indique le répertoire de consignation dans lequel les fichiers journaux dans Content Search Services Support sont écrits. Le répertoire par défaut est `./log` (Le sous-répertoire du journal dans lequel IBM Content Search Services est installé).

Si le délai est activé, Content Search Services Support écrit des informations temporelles dans les fichiers journaux. Les informations temporelles seront écrites dans les fichiers journaux uniquement si la consignation est activée.

Paramètres de configuration de Content Search Services Support pour les documents de courrier électronique :

La configuration des documents de courrier électronique inclut les paramètres généraux, tels que les paramètres de connexion du référentiel et de vidage, les paramètres par défaut qui incluent l'attribut et les paramètres d'attribut d'archivage, ainsi que les paramètres personnalisés qui peuvent être ajoutés par l'utilisateur.

Les paramètres généraux de configuration pour les documents de courrier électronique incluent :

- **Connexion du référentiel :** indique les informations d'accès pour FileNet P8Content Engine. Content Search Services Support n'utilise qu'une partie des

informations de connexion du référentiel : l'URL Content Engine et les données d'identification de connexion au serveur Content Engine. Cet ID utilisateur de connexion doit disposer d'un accès en lecture valable dans tous le domaine à tous les documents devant être indexés. Vous pouvez sélectionner n'importe quelle connexion du référentiel qui fournit l'URL requises et les propriétés de connexion, quel que soit le conteneur d'objets défini pour la connexion.

Cette valeur est définie par IBM Content Collector lorsque le conteneur d'objets est configuré pour IBM Content Search Services.

- **Paramètres de vidage :**

- **Créer des fichiers de vidage pour tous les documents d'entrée :** génère des fichiers de vidage de tous les documents d'entrée dans le sous-répertoire doc du répertoire de vidage spécifié. Une copie du document d'origine est créée dans le format (binaire) et le codage d'origine ; chaque nom de fichier de vidage contient l'identificateur du document.
- **Créer des fichiers de vidage du fichier XML pour l'indexation :** génère des fichiers de vidage de tous les documents XML générés dans le sous-répertoire xml du répertoire de vidage spécifié. Chaque nom de fichier de vidage contient l'identificateur du document.

Le fichier XML sert d'entrée pour le moteur d'indexation. Activation de ce paramètre :

- Génère une copie du fichier XML qui contient le contenu textuel transmis au moteur d'indexation. Le fichier XML n'indique pas quel contenu a été correctement indexé par IBM Content Search Services. Pour obtenir ces informations, vous devez utiliser le IBM FileNet Enterprise Manager.
- Permet de vérifier la structure dans laquelle le contenu textuel doit être conservé dans l'index. Si vous ajoutez des attributs personnalisés ou modifiez les descriptions de chemin d'attribut existantes, vous pouvez utiliser le fichier XML en vue de vérifier que les descriptions de chemin d'attribut modifiées sont correctement définies.
- **Créer des fichiers de vidage pour le contenu textuel des pièces jointes :** génère des fichiers de vidage de tout le contenu textuel des pièces jointes intégrées dans le sous-répertoire txt du répertoire de vidage spécifié. Une copie du contenu textuel de la pièce jointe intégrée est créée ; chaque nom de fichier de vidage contient l'identificateur du document et le nom de la pièce jointe.
- **Répertoire de vidage :** indique le nom du répertoire dans lequel les fichiers de vidage sont écrits dans les sous-répertoires correspondants. Un paramètre de vidage minimum doit être activé.

Le répertoire par défaut est ./log/dump (Le répertoire de vidage dans le sous-répertoire du journal dans lequel IBM Content Search Services est installé).

Les paramètres par défaut incluent les attributs de courrier électronique et les attributs d'archivage requis par Content Collector. Vous ne devez pas retirer les attributs de courrier électronique qui ont été ajoutés à la configuration de Content Collector.

Les attributs qui sont encadrés par un double trait de soulignement sont des attributs intégrés et s'appliquent à Lotus Domino et à Microsoft Exchange.

Tableau 17. Descriptions des attributs de courrier électronique et des chemins d'accès des attributs

Attributs intégrés	Description du chemin de l'attribut
<u>__ITEMID__</u>	/icc_document/icc_item_id

Tableau 17. Descriptions des attributs de courrier électronique et des chemins d'accès des attributs (suite)

Attributs intégrés	Description du chemin de l'attribut
__INDEXDATE__	/icc_document/@icc_index_date
__PROPERTIES__	/icc_document/icc_doc_properties
__RECEIVED_DATE__	/icc_document/icc_email/@icc_received_date
__SUBJECT__	/icc_document/icc_email/icc_subject
__FROM__	/icc_document/icc_email/icc_from
__TO__	/icc_document/icc_email/icc_recipients/icc_to
__CC__	/icc_document/icc_email/icc_recipients/icc_cc
__BCC_MSG__ (Les BCC Exchange sont mappés à des destinataires)	/icc_document/icc_email/icc_recipients/icc_bcc
__BCC_SMTP__ (Les courriers électroniques SMTP sont mappés à des destinataires)	/icc_document/icc_email/icc_recipients/icc_bcc
__BCC_CSN__ (Les BCC Domino BCC doivent être mappés à un journal)	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses/icc_bcc
__BODY__	/icc_document/icc_email/icc_content/icc_body
__ATTACHMENT__	/icc_document/icc_email/icc_content/icc_attachments/icc_attachment
__ATTACHMENT_NAME__	/icc_document/icc_email/icc_content/icc_attachments/icc_attachment/@name
Attributs Lotus Notes	
Form	/icc_document/icc_doc_properties/icc_form
Principal	/icc_document/icc_email/icc_from
SMTPOriginator	/icc_document/icc_email/icc_from
INetFrom	/icc_document/icc_email/icc_from
INetSendTo	/icc_document/icc_email/icc_recipients/icc_to
INetCopyTo	/icc_document/icc_email/icc_recipients/icc_cc
INetBlindCopyTo	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses/icc_bcc
\$JournalRecipients	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses
\$JournalRecipientsExpanded_1	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses
\$JournalRecipientsExpanded_2	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses
\$JournalRecipientsExpanded_3	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses
\$JournalRecipientsExpanded_4	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses
\$JournalRecipientsExpanded_5	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses
\$JournalRecipientsExpanded_6	/icc_document/icc_journal/icc_addresses/icc_recipient_addresses

Tableau 17. Descriptions des attributs de courrier électronique et des chemins d'accès des attributs (suite)

Attributs intégrés	Description du chemin de l'attribut
\$JournalRecipientsExpanded_7	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
\$JournalRecipientsExpanded_8	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
\$JournalRecipientsExpanded_9	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
\$JournalRecipientsExpanded_10	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_1	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_2	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_3	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_4	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_5	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_6	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_7	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_8	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_9	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses
EMSJournalRecipients_10	/icc_document/icc_journal/icc_addresses/ icc_recipient_addresses

Les attributs d'archivage sont des attributs de courrier électronique stockés dans IBM FileNet P8. Par défaut, les trois attributs d'archivage suivants sont configurés pour les courriers électroniques : ICCMailboxID, ICCExpirationDate, et ICCVaryingFields.

L'attribut ICCMailboxID est requis par Content Search Services Support ; il doit être défini et avoir une valeur. Si un document est traité et qu'il n'a pas de valeur pour ICCMailboxID, Content Search Services Support arrêtera le traitement du document immédiatement et le document ne sera pas indexé.

Tableau 18. Descriptions des attributs de document et des chemins d'attributs

Attribut d'archivage	Description du chemin de l'attribut
ICCMailboxID	/icc_document/icc_mailbox_id
ICCExpirationDate	/icc_document/icc_email/@icc_expiration_date
ICCVaryingFields	/icc_document

Vous pouvez modifier les définitions d'attribut Content Collector par défaut. Toutefois, vous devez être vigilant lors de la modification de la définition d'attribut BCC. Normalement, toutes les valeurs d'attribut qui sont extraites à partir d'un e-mail et qui sont indexées sont consultables. Toutefois, dans le cas de l'attribut BCC, vous devez être attentifs à ce que le contenu BCC ne soit pas mappé avec une section consultable par tous les utilisateurs, car cela peut permettre aux utilisateurs de rechercher d'autres utilisateurs BCC sans que cela soit souhaitable.

La manière dont les informations BCC sont mappées dans Content Collector est influencée par la manière dont le dédoublement est effectué et diffère entre Microsoft Exchange et Lotus Notes. Lorsqu'un courrier électronique contient des informations BCC, le dédoublement génère une copie supplémentaire à stocker pour l'émetteur. Ceci signifie qu'il y aura une copie pour l'expéditeur et une autre pour le journal et tous les destinataires.

Dans Microsoft Exchange, le message ne contient pas d'informations BCC. Par conséquent, seule la copie destinée à l'expéditeur contient de telles informations. La section BCC est vide dans toutes les copies de destinataire, ainsi que dans la copie du journal. Dans la copie du journal, seule l'enveloppe inclut les informations BCC.

Dans Lotus Notes, en revanche, les informations BCC sont disponibles dans toutes les copies. Chaque copie de destinataire inclut uniquement les destinataires respectifs, tandis que les copies de journal et de l'expéditeur incluent tous les destinataires. Il est donc essentiel que ces informations ne figurent jamais dans la section consultable par les utilisateurs.

Si vous modifiez le mappage d'attribut BCC et ne souhaitez pas que les utilisateurs puissent rechercher d'autres utilisateurs BCC, vous devez désactiver le dédoublement de manière à ce que, chaque fois qu'un courrier électronique comporte une liste BCC, une copie du courrier électronique soit stockée pour l'expéditeur, le journal et chaque destinataire. Vous pouvez désactiver le dédoublement en ajoutant l'élément BlindCopyTo à la liste d'éléments supplémentaires utilisée pour calculer le hachage de dédoublement dans les rubriques des routes de tâches respectives. Toutefois, vous devez garder à l'esprit que cela peut entraîner d'importants besoins en espace supplémentaire, car cette pratique est courante pour permettre l'envoi de courriers électroniques avec de grandes listes de distribution à tous les destinataires spécifiés en BCC.

Ajoutez des attributs personnalisés, des attributs d'archivage personnalisés et des paramètres d'options de configuration Content Search Services Support dans l'onglet **Paramètres personnalisés**.

Les attributs personnalisés et des attributs d'archivage personnalisés doivent être ajoutés sous les sections respectives dans le format indiqué dans l'exemple.

Important :

- Les attributs ajoutés à la section [Attributes] du fichier de configuration ne sont pas sensibles à la casse.
- Les attributs ajoutés à la section [ArchiveAttributes] du fichier de configuration sont sensibles à la casse. La casse de l'attribut d'archivage doit correspondre à la casse utilisée pour les attributs dans IBM FileNet P8.

Content Search Services Support utilise les options de configuration supplémentaires suivantes si elles sont définies dans la configuration. Si vous

souhaitez définir ces options, ajoutez-les en-dessous des attributs personnalisés et des attributs d'archivage personnalisés, de la manière suivante :

```
[Settings]
AttachmentTextConversion = 0
DumpDocumentIds = {50F0181B-9FE3-5718-5383-9A5ECAAF35E14} {361B375F-98C8-476D-BBAE-F3AD4778FE81}
```

Important : Les options de configuration ne sont pas sensibles à la casse.

Les options de configuration que vous pouvez définir ont une influence sur le traitement de Content Search Services Support.

AttachmentTextConversion

Type de valeur : Booléen

Valeur par défaut : 1 ou true

Si la valeur est définie sur 1 : Content Search Services Support traitera toutes les pièces jointes intégrées et filtrera le contenu textuel.

DumpDocumentIds

Type de valeur : Chaîne

Valeur par défaut : non définie

En définissant **DumpDocumentIds**, vous pouvez spécifier uniquement un sous-ensemble de documents pour lesquels les fichiers de vidage sont créés dans le répertoire de vidage spécifié. Vous devez répertorier chaque document à l'aide de son ID. Chaque ID de document doit apparaître entre accolades et être séparé par un espace.

Pour plus d'informations sur l'utilisation de cette option de configuration, consultez «Débogage des documents introuvables», à la page 221.

DumpDocumentIds a été renommé. Il a été appelé **DumpDocuments**.

Paramètres de configuration de Content Search Services Support pour les documents IBM Connections :

La configuration des documents IBM Connections inclut les paramètres généraux tels que les paramètres de connexion du référentiel et de vidage, les paramètres par défaut qui incluent l'attribut et les paramètres d'attribut d'archivage ainsi que les paramètres personnalisés qui peuvent être ajoutés par l'utilisateur.

Les paramètres de configuration générale pour les documents IBM Connections incluent :

- **Connexion du référentiel :** indique les informations d'accès pour FileNet P8Content Engine. Content Search Services Support n'utilise qu'une partie des informations de connexion du référentiel : l'URL Content Engine et les données d'identification de connexion au serveur Content Engine. Cet ID utilisateur de connexion doit disposer d'un accès en lecture valable dans tous le domaine à tous les documents devant être indexés. Vous pouvez sélectionner n'importe quelle connexion du référentiel qui fournit l'URL requises et les propriétés de connexion, quel que soit le conteneur d'objets défini pour la connexion.
Cette valeur est définie par IBM Content Collector lorsque le conteneur d'objets est configuré pour IBM Content Search Services.
- **Paramètres de vidage :**
 - **Créer des fichiers de vidage pour tous les documents d'entrée :** génère des fichiers de vidage de tous les documents d'entrée dans le sous-répertoire doc du répertoire de vidage spécifié. Une copie du document d'origine est créée dans le format (binaire) et le codage d'origine ; chaque nom de fichier de vidage contient l'identificateur du document.

- **Créer des fichiers de vidage du fichier XML pour l'indexation** : génère des fichiers de vidage de tous les documents XML générés dans le sous-répertoire xml du répertoire de vidage spécifié. Chaque nom de fichier de vidage contient l'identificateur du document.

Le fichier XML sert d'entrée pour le moteur d'indexation. Activation de ce paramètre :

- Génère une copie du fichier XML qui contient le contenu textuel transmis au moteur d'indexation. Le fichier XML n'indique pas quel contenu a été correctement indexé par IBM Content Search Services. Pour obtenir ces informations, vous devez utiliser le IBM FileNet Enterprise Manager.
- Permet de vérifier la structure dans laquelle le contenu textuel doit être conservé dans l'index. Si vous ajoutez des attributs personnalisés ou modifiez les descriptions de chemin d'attribut existantes, vous pouvez utiliser le fichier XML en vue de vérifier que les descriptions de chemin d'attribut modifiées sont correctement définies.

- **Créer des fichiers de vidage pour le contenu textuel des pièces jointes** : génère des fichiers de vidage de tout le contenu textuel des pièces jointes intégrées dans le sous-répertoire txt du répertoire de vidage spécifié. Une copie du contenu textuel de la pièce jointe intégrée est créée ; chaque nom de fichier de vidage contient l'identificateur du document et le nom de la pièce jointe.

- **Répertoire de vidage** : indique le nom du répertoire dans lequel les fichiers de vidage sont écrits dans les sous-répertoires correspondants. Un paramètre de vidage minimum doit être activé.

Le répertoire par défaut est ./log/dump (Le répertoire de vidage dans le sous-répertoire du journal dans lequel IBM Content Search Services est installé).

Les paramètres par défaut incluent les attributs de documents IBM Connections requis par Content Collector. Vous ne devez pas retirer les attributs qui ont été ajoutés à la configuration de Content Collector. Ajoutez uniquement de nouvelles entrées aux attributs existants sur l'onglet **Paramètres personnalisés**.

Tableau 19. Descriptions des attributs de document et des chemins d'attributs IBM Connections

Attribut	Description du chemin de l'attribut
__ITEMID__	/icc_document/icc_item_id
__INDEXDATE__	/icc_document/@icc_index_date
__PROPERTIES__	/icc_document/icc_doc_properties
Pour les parties XML	
__XMLPART__	/icc_document/icc_connections/icc_part
__XMLPART_NAME__	/icc_document/icc_connections/icc_part/@name
__XMLPART_MIMETYPE__	/icc_document/icc_connections/icc_part/@mimetype
Pour tous les autres documents et toutes les autres parties de document	
__CONTENT__	/icc_document/icc_connections/icc_attachments/icc_attachment/icc_part
__CONTENT_NAME__	/icc_document/icc_connections/icc_attachments/icc_attachment/icc_part/@name
__CONTENT_MIMETYPE__	/icc_document/icc_connections/icc_attachments/icc_attachment/icc_part/@mimetype

Les attributs d'archivage répertorient les attributs de documents IBM Connections stockés dans IBM FileNet P8.

Tableau 20. Attributs d'archivage des documents IBM Connections

Attribut d'archivage	Description du chemin de l'attribut
ICCApplicationName	/icc_document/icc_application_name

Ajoutez des attributs personnalisés, des attributs d'archivage personnalisée et des paramètres d'options de configuration Content Search Services Support dans l'onglet **Paramètres personnalisés**.

Les attributs personnalisés et des attributs d'archivage personnalisés doivent être ajoutés sous les sections respectives dans le format indiqué dans l'exemple.

Important :

- Les attributs ajoutés à la section [Attributes] du fichier de configuration ne sont pas sensibles à la casse.
- Les attributs ajoutés à la section [ArchiveAttributes] du fichier de configuration sont sensibles à la casse. La casse de l'attribut d'archivage doit correspondre à la casse utilisée pour les attributs dans IBM FileNet P8.

Content Search Services Support utilise les options de configuration supplémentaires suivantes si elles sont définies dans la configuration. Si vous souhaitez définir ces options, ajoutez-les en-dessous des attributs personnalisés et des attributs d'archivage personnalisés, de la manière suivante :

```
[Settings]
AttachmentTextConversion = 0
DumpDocumentIds = {50F0181B-9FE3-5718-5383-9A5ECAAF35E14} {361B375F-98C8-476D-BBAE-F3AD4778FE81}
```

Important : Les options de configuration ne sont pas sensibles à la casse.

Les options de configuration que vous pouvez définir ont une influence sur le traitement de Content Search Services Support.

AttachmentTextConversion

Type de valeur : Booléen

Valeur par défaut : 1 ou true

Si la valeur est définie sur 1 : Content Search Services Support traitera toutes les pièces jointes intégrées et filtrera le contenu textuel.

DumpDocumentIds

Type de valeur : Chaîne

Valeur par défaut : non définie

En définissant **DumpDocumentIds**, vous pouvez spécifier uniquement un sous-ensemble de documents pour lesquels les fichiers de vidage sont créés dans le répertoire de vidage spécifié. Vous devez répertorier chaque document à l'aide de son ID. Chaque ID de document doit apparaître entre accolades et être séparé par un espace.

Pour plus d'informations sur l'utilisation de cette option de configuration, consultez «Débogage des documents introuvables», à la page 221.

DumpDocumentIds a été renommé. Il a été appelé **DumpDocuments**.

Paramètres de configuration de Content Search Services Support pour les documents Microsoft SharePoint :

La configuration des documents Microsoft SharePoint inclut les paramètres généraux tels que les paramètres de connexion du référentiel et de vidage, les paramètres par défaut qui incluent l'attribut et les paramètres d'attribut d'archivage ainsi que les paramètres personnalisés qui peuvent être ajoutés par l'utilisateur.

Les paramètres généraux de configuration pour les documents Microsoft SharePoint incluent :

- **Connexion du référentiel** : indique les informations d'accès pour FileNet P8Content Engine. Content Search Services Support n'utilise qu'une partie des informations de connexion du référentiel : l'URL Content Engine et les données d'identification de connexion au serveur Content Engine. Cet ID utilisateur de connexion doit disposer d'un accès en lecture valable dans tous le domaine à tous les documents devant être indexés. Vous pouvez sélectionner n'importe quelle connexion du référentiel qui fournit l'URL requises et les propriétés de connexion, quel que soit le conteneur d'objets défini pour la connexion.

Cette valeur est définie par IBM Content Collector lorsque le conteneur d'objets est configuré pour IBM Content Search Services.

- **Paramètres de vidage** :

- **Créer des fichiers de vidage pour tous les documents d'entrée** : génère des fichiers de vidage de tous les documents d'entrée dans le sous-répertoire doc du répertoire de vidage spécifié. Une copie du document d'origine est créée dans le format (binaire) et le codage d'origine ; chaque nom de fichier de vidage contient l'identificateur du document.
- **Créer des fichiers de vidage du fichier XML pour l'indexation** : génère des fichiers de vidage de tous les documents XML générés dans le sous-répertoire xml du répertoire de vidage spécifié. Chaque nom de fichier de vidage contient l'identificateur du document.

Le fichier XML sert d'entrée pour le moteur d'indexation. Activation de ce paramètre :

- Génère une copie du fichier XML qui contient le contenu textuel transmis au moteur d'indexation. Le fichier XML n'indique pas quel contenu a été correctement indexé par IBM Content Search Services. Pour obtenir ces informations, vous devez utiliser le IBM FileNet Enterprise Manager.
- Permet de vérifier la structure dans laquelle le contenu textuel doit être conservé dans l'index. Si vous ajoutez des attributs personnalisés ou modifiez les descriptions de chemin d'attribut existantes, vous pouvez utiliser le fichier XML en vue de vérifier que les descriptions de chemin d'attribut modifiées sont correctement définies.

- **Créer des fichiers de vidage pour le contenu textuel des pièces jointes** : génère des fichiers de vidage de tout le contenu textuel des pièces jointes intégrées dans le sous-répertoire txt du répertoire de vidage spécifié. Une copie du contenu textuel de la pièce jointe intégrée est créée ; chaque nom de fichier de vidage contient l'identificateur du document et le nom de la pièce jointe.

- **Répertoire de vidage** : indique le nom du répertoire dans lequel les fichiers de vidage sont écrits dans les sous-répertoires correspondants. Un paramètre de vidage minimum doit être activé.

Le répertoire par défaut est ./log/dump (Le répertoire de vidage dans le sous-répertoire du journal dans lequel IBM Content Search Services est installé).

Les paramètres par défaut incluent les attributs de documents Microsoft SharePoint requis par Content Collector. Vous ne devez pas retirer les attributs qui ont été ajoutés à la configuration de Content Collector. Ajoutez uniquement de nouvelles entrées aux attributs existants sur l'onglet **Paramètres personnalisés**.

Tableau 21. Descriptions des attributs de document et des chemins d'attributs Microsoft SharePoint

Attribut	Description du chemin de l'attribut
__ITEMID__	/icc_document/icc_item_id
__INDEXDATE__	/icc_document/@icc_index_date
__PROPERTIES__	/icc_document/icc_doc_properties
Pour tous les documents et toutes les parties de document	
__CONTENT__	/icc_document/icc_sharepoint/icc_content/icc_main
__CONTENT_NAME__	/icc_document/icc_sharepoint/icc_content/icc_main/@name
__PARTID__	/icc_document/icc_sharepoint/icc_content/icc_main/@partid

Il n'y a aucun attribut d'archivage par défaut pour les documents Microsoft SharePoint qui sont requis par Content Collector.

L'option de configuration **XmlTransformationXslFile** spécifie un fichier de transformation XML supplémentaire qui peut être utilisé lors du traitement du document Microsoft SharePoint. A l'aide de ce fichier de transformation XML, un chemin XML différent est appliqué à toutes les pièces jointes intégrées au document Microsoft SharePoint principal. Une pièce jointe intégrée est une partie de document avec un `partid` supérieur à 1. Pour appliquer ce traitement aux documents avec des pièces jointes intégrées, vous devez définir **XmlTransformationXslFile** sur l'onglet **Paramètres personnalisés**.

Ajoutez des attributs personnalisés, des attributs d'archivage personnalisée et des paramètres d'options de configuration Content Search Services Support dans l'onglet **Paramètres personnalisés**.

Les attributs personnalisés et des attributs d'archivage personnalisés doivent être ajoutés sous les sections respectives dans le format indiqué dans l'exemple.

Important :

- Les attributs ajoutés à la section [Attributes] du fichier de configuration ne sont pas sensibles à la casse.
- Les attributs ajoutés à la section [ArchiveAttributes] du fichier de configuration sont sensibles à la casse. La casse de l'attribut d'archivage doit correspondre à la casse utilisée pour les attributs dans IBM FileNet P8.

Content Search Services Support utilise les options de configuration supplémentaires suivantes si elles sont définies dans la configuration. Si vous souhaitez définir ces options, ajoutez-les en-dessous des attributs personnalisés et des attributs d'archivage personnalisés, de la manière suivante :

```
[Settings]
XmlTransformationXslFile = afu_sp.xsl
AttachmentTextConversion = 0
DumpDocumentIds = {50F0181B-9FE3-5718-5383-9A5ECA35E14} {361B375F-98C8-476D-BBAE-F3AD4778FE81}
```

Important : Les options de configuration ne sont pas sensibles à la casse.

Les options de configuration que vous pouvez définir ont une influence sur le traitement de Content Search Services Support.

XmlTransformationXslFile

Type de valeur : Chaîne

Valeur par défaut : non définie

Indique le nom d'un fichier XSL qui décrit la transformation XML appliquée au document XML généré avant qu'il ne soit envoyé à IBM Content Search Services. Vous pouvez spécifier un nom de fichier avec ou sans le chemin d'accès au fichier. Si vous n'indiquez pas de chemin, le fichier est recherché dans le répertoire en cours et dans le sous-répertoire config de l'instance IBM Content Search Services.

Cette option est nécessaire uniquement pour les classes de document Microsoft SharePoint, auquel cas **XmlTransformationXslFile** pointe automatiquement vers le fichier XSL `afu_sp.xsl` fourni par Content Search Services Support.

AttachmentTextConversion

Type de valeur : Booléen

Valeur par défaut : 1 ou true

Si la valeur est définie sur 1 : Content Search Services Support traitera toutes les pièces jointes intégrées et filtrera le contenu textuel.

DumpDocumentIds

Type de valeur : Chaîne

Valeur par défaut : non définie

En définissant **DumpDocumentIds**, vous pouvez spécifier uniquement un sous-ensemble de documents pour lesquels les fichiers de vidage sont créés dans le répertoire de vidage spécifié. Vous devez répertorier chaque document à l'aide de son ID. Chaque ID de document doit apparaître entre accolades et être séparé par un espace.

Pour plus d'informations sur l'utilisation de cette option de configuration, consultez «Débogage des documents introuvables», à la page 221.

DumpDocumentIds a été renommé. Il a été appelé **DumpDocuments**.

Vérification de votre configuration IBM Content Search Services

Une fois que vous avez installé IBM Content Collector Server et exécuté l'assistant de configuration initiale, et après avoir vérifié l'installation et la configuration d'IBM Content Collector P8 Content Search Services Support, vérifiez que l'indexation fonctionne correctement.

Utilisez IBM FileNet Enterprise Manager pour vérifier que les documents peuvent être indexés via IBM Content Search Services.

Pour vérifier que les fichiers et les courriers électroniques sont correctement indexés :

1. Vérifiez que les fichiers sont correctement indexés.
2. Vérifiez que les courriers électroniques sont correctement indexés.

Référence associée :

«Erreurs de traitement de documents Content Search Services Support», à la page 216

«Contrôle de l'indexation avec IBM Content Collector P8 Content Search Services Support», à la page 222

Information associée :

 Centre de documentation d'IBM FileNet P8 Version 5.1

Vérification de l'indexation correcte des fichiers :

Vérifiez que le moteur de recherche de contenu écrit et lit correctement les informations d'index pour les fichiers afin de permettre l'extraction de contenu.

Pour vérifier que les fichiers sont correctement indexés :

1. Activez la consignation du serveur IBM Content Search Services comme décrit dans la section relative à la configuration de la consignation de serveur dans la documentation FileNet P8. En cas d'erreurs, vérifiez les fichiers journaux suivants :
 - Les fichiers journaux IBM Content Search Services
 - Les fichiers journaux Content Engine, notamment les fichiers `p8_server_error.log` et `p8_server_trace.log`
2. Définissez une classe de document pour laquelle l'extraction de contenu est activée.
3. Créer un nouveau document qui contient un fichier texte enrichi, par exemple, un fichier PDF.
4. Contrôlez la file d'attente d'indexation pour vérifier si une demande d'indexation est créée et pour vérifier l'état du document. Dès que le document est indexé, il est retiré sur la file d'attente d'indexation.
5. Vérifiez qu'une collecte a été créée pour la zone d'index de votre conteneur d'objets et affichez les détails de cette collecte. Les propriétés de la zone d'index fournissent des détails sur une collecte, par exemple, sa taille ou le nombre de documents qu'elle contient. La collecte doit contenir un document.
6. Soumettez une requête CBR pour un mot ou une phrase (*searchterm*) dans le document.
 - a. Dans le panneau de navigation, cliquez avec le bouton droit sur le noeud Résultats de la recherche du conteneur d'objets que vous souhaitez rechercher, puis cliquez sur **Nouvelle recherche**.
 - b. Cliquez sur **Affichage > Affichage SQL**. Entrez ensuite la commande SQL à l'aide d'une syntaxe similaire à l'exemple suivant puis soumettez la requête.

```
SELECT d.This FROM Document d
      INNER JOIN ContentSearch c ON d.This = c.QueriedObject
      WHERE CONTAINS(d.*, 'searchterm')
```
7. Si votre document test figure dans la liste des résultats, vous pouvez supprimer le document à l'aide de Enterprise Manager.

Vérifiez que le courrier électronique qui a été archivé à l'aide de IBM Content Collector est traité par IBM Content Collector P8 Content Search Services Support avant qu'il ne soit envoyé via le traitement de texte ordinaire IBM Content Search Services et le pipeline d'indexation.

Vérification de l'indexation correcte des courriers électroniques :

Vérifiez que le moteur de recherche de contenu écrit et lit correctement les informations d'index pour les courriers électroniques afin de permettre l'extraction de contenu.

Pour vérifier que les documents de courriers électroniques sont correctement indexés :

1. Activez la consignation et la création de fichiers de vidage :
 - a. Activez la consignation du serveur IBM Content Search Services comme décrit dans la section relative à la configuration de la consignation de serveur dans FileNet P8.
 - b. Dans le Gestionnaire de configuration IBM Content Collector sous **Paramètres généraux > Content Search Services Support > Paramètres de journal communs**, définissez le niveau de journalisation. Dans **Support pour le courrier électronique**, activez la création de fichiers de vidage des fichiers XML qui sont les entrées pour l'indexation.

En cas d'erreurs au cours du processus de vérification, vérifiez les fichiers journaux suivants :

- Les fichiers journaux IBM Content Search Services
- Les fichiers journaux Content Engine, notamment les fichiers p8_server_error.log et p8_server_trace.log
- Les fichiers journaux IBM Content Collector P8 Content Search Services Support

Content Search Services Support crée des fichiers de vidage de tous les documents XML générés dans le sous-répertoire xml du répertoire de vidage spécifié. Chaque nom de fichier de vidage contient l'identificateur du document. Vous pouvez utiliser le fichier de vidage pour trouver les paramètres pour lequel vous souhaitez effectuer une recherche ultérieurement.

2. Utilisez IBM Content Collector pour archiver un ensemble connu de documents électroniques dans une classe de document ICCMail3.
3. Contrôlez la file d'attente d'indexation pour vérifier si une demande d'indexation est créée et pour vérifier l'état des documents. Dès que les documents sont indexés, ils sont retirés de la file d'attente d'indexation.
4. Vérifiez qu'une collecte a été créée pour la zone d'index de votre conteneur d'objets et affichez les détails de cette collecte. Les propriétés de la zone d'index indiquent le nombre de documents que contient la collecte et les paramètres appropriés pour le partitionnement de date. En fonction de l'intervalle de partition, le premier objet qui est indexé pour une collecte détermine la plage de dates pour l'index.
5. Si le nombre de documents dans la collecte ne correspond pas au nombre de documents archivés, consultez les fichiers journaux et soumettez une requête à partir d'Enterprise Manager pour vérifier les erreurs d'indexation. Sinon, ignorez cette étape. Générez et soumettez une requête en affichage simple. Sélectionnez la classe et, comme critères de recherche, indiquez que la valeur de la colonne **Code d'incident d'indexation** doit être supérieure à la valeur null. Les documents pour lesquels se sont produites des erreurs d'indexation sont répertoriées. Vérifiez les codes incident et suivez la procédure décrite pour l'erreur correspondante.
6. Soumettez une requête à partir d'Enterprise Manager pour rechercher les documents de courrier électronique. Exécutez le générateur de requête dans la vue SQL et utilisez une syntaxe XPath basé sur pour rechercher des documents XML comme dans les exemples suivants. N'oubliez pas que vous pouvez

utiliser les fichiers de vidage IBM Content Collector P8 Content Search Services Support pour déterminer vos termes de recherche :

- Pour rechercher des courriers électroniques où l'adresse de l'expéditeur est user@example.com

```
SELECT d.This, d.Id, cs.ContentSummary FROM ICCMai13 d
INNER JOIN ContentSearch cs ON d.This = cs.QueriedObject
WHERE CONTAINS(d.*, '@xmlxp:'''//icc_from[.contains("user@example.com")]''')
```

- Pour rechercher un mot ou une expression (*searchterm*) contenue n'importe où dans le corps du courrier électronique

```
SELECT d.This, d.Id, cs.ContentSummary FROM ICCMai13 d
INNER JOIN ContentSearch cs ON d.This = cs.QueriedObject
WHERE CONTAINS(d.*, '@xmlxp:'''//icc_body[.contains("searchterm")]''')
```

- Pour rechercher un mot ou une expression (*searchterm*) contenue n'importe où dans une pièce jointe du courrier électronique

```
SELECT d.This, d.Id, cs.ContentSummary FROM ICCMai13 d
INNER JOIN ContentSearch cs ON d.This = cs.QueriedObject
WHERE CONTAINS(d.*, '@xmlxp:'''//icc_content[.contains("searchterm")]''')
```

Un serveur de recherche IBM Content Search Services exécute la requête et les résultats sont affichés dans Enterprise Manager. Dans ce cas, la configuration IBM Content Search Services a été correctement vérifiée.

Si, toutefois, la requête ne renvoie aucun résultat, passez à l'étape 7

7. Vérifiez les conditions suivantes :
 - a. Content Search Services Support est correctement installé et déployé sur les instances IBM Content Search Services.
 - b. Content Search Services Support est configuré dans chaque instance.
8. Vérifiez les fichiers journaux en cas d'erreurs.
9. Redémarrez Content Search Services Support. Soumettez ensuite une demande de réindexation pour les documents de courrier électronique et retournez à l'étape 3, à la page 212.

Etapes de traitement Content Search Services Support

Content Search Services Support traite tous les documents de courrier électronique qui sont archivés dans un référentiel FileNet P8 à l'aide d'IBM Content Collector.

Voici les principales étapes de traitement de document dans Content Search Services Support :

- Le document de courrier électronique archivé est récupéré en parties. Il s'agit généralement de deux parties différentes :
 - Le corps.

Le corps du courrier électronique est identifié en vérifiant si le type MIME du courrier électronique correspond à l'un des types MIME admissibles pour le traitement des courriers électroniques par Content Search Services Support. Les types MIME suivants sont pris en charge :

 - application/iccsn
 - application/x-filenet-filetype-csn
 - application/csn
 - application/vnd.ms-outlook
 - application/x-filenet-filetype-msg
 - application/msoutlook
 - message/rfc822
 - application/iccxit

- application/csbundled
- Une ou plusieurs pièces jointes.
- Le corps du message électronique est envoyé à IBM Content Search Services où les analyseurs syntaxiques de document détectent si le message électronique est un CSN Lotus Notes, un fichier Microsoft MSG ou un courrier électronique au format RFC822 SMTP. Le contenu du document dans les champs comme CORPS et OBJET est renvoyé à Content Search Services Support qui ajoute le contenu à une structure XML intermédiaire.
- Les pièces jointes intégrées sont envoyées aux services de filtre de conversion de document dans IBM Content Search Services. Après le traitement du texte, le contenu textuel est renvoyé à Content Search Services Support qui ajoute le contenu de la structure XML intermédiaire.
- Avant que la structure XML ne soit transmise au moteur d'indexation, Content Search Services Support vérifie que l'attribut d'archivage requis appelé ICCMailboxID a une valeur pour chaque document. Si un document n'a pas de valeur pour cet attribut, il ne sera pas indexé et un message d'erreur sera consigné.

Identification et résolution des incidents d'indexation lors de l'indexation avec IBM Content Search Services

Identifiez et résolvez les incidents d'indexation pour vérifier que les objets sont indexés. Les objets qui n'ont pas été indexés ne peuvent pas être recherchés et sont introuvables dans le référentiel.

Identification des incidents d'indexation lors de l'indexation avec IBM Content Search Services

Déterminez les objets ayant provoqué des incidents d'indexation et étant probablement la cause pour laquelle ces objets n'ont pu être indexés.

Les incidents d'indexation sont des incidents d'indexation des objets ou d'indexation des index.

Les incidents d'indexation d'objets empêchent l'indexation correcte du texte pour un objet particulier. Ces incidents peuvent être associés au document lui-même ou au système. Les incidents liés au document sont dus à la nature du texte. Par exemple, l'objet peut ne pas être indexé car le document est protégé par un mot de passe. Pour les incidents liés au document, la demande d'index n'est pas automatiquement relancée. Les incidents liés au système sont dus à l'environnement système en cours. Par exemple, une demande d'index peut échouer car aucun serveur d'index n'est actuellement disponible pour traiter la demande. Pour les incidents liés au système, la demande est relancée jusqu'à ce qu'elle aboutisse (car vous pouvez corriger les problèmes d'environnement système) ou jusqu'à ce que vous la supprimiez.

Un incident de tâche d'index n'est pas lié à une demande d'index ou à un objet particulier mais à un ensemble de demandes d'index.

Pour plus d'informations sur l'identification et la résolution des incidents liés au système ou aux tâches d'index, consultez les sections correspondantes dans la documentation FileNet P8.

Pour identifier les incidents d'indexation liés au document :

1. Pour rechercher des objets présentant des incidents d'indexation, utilisez la requête suivante afin de parcourir le conteneur d'objets.

```
SELECT d.This, d.CmIndexingFailureCode FROM doc_class d WHERE
d.CmIndexingFailureCode > 0
```

Remplacez *doc_class* tion associée : appropriée :

- ICCMail3 pour les courriers électroniques
- ICCFileInstance2 pour les fichiers
- ICCSharepointInstance2 pour les documents Microsoft SharePoint
- ICCConnectionsInstance pour les éléments IBM Connections

Cette requête parcourt la table DocVersion, qui correspond à la date de base de données pour les objets Document.

Vous pouvez paramétrer cette requête dans l'affichage simple du générateur de requête. Dans ce cas, sélectionnez les paramètres suivants :

Zone	Valeur
Sélectionner à partir de la table	ICCMail3, ICCFileInstance2, ICCSharepointInstance2 ou ICCConnectionsInstance
Sélectionner des colonnes	*
Critères > Colonne	Incident d'indexation
Critères > Condition	Est supérieur à
Critères > Valeur	0

Les requêtes sur la propriété CmIndexingFailureCode peuvent entraîner une analyse de table de base de données, ce qui prend beaucoup de temps lorsqu'elle est effectuée sur une table volumineuse. Un index de base de données pour la propriété CmIndexingFailureCode peut améliorer les performances de ces requêtes car le gestionnaire de base de données accède aux lignes de qualification et renvoie les données de la colonne demandée directement via l'analyse de l'index. Envisagez donc de demander à l'administrateur de bases de données d'ajouter un index à la colonne index_failure_code dans la table DocVersion.

Pour limiter le nombre de résultats, ajoutez des critères de filtrage à votre recherche, tels que la date de création du document, ou recherchez un code incident spécifique. Dans ce cas, vous devez également créer les index de base de données appropriées pour éviter des analyses de tables. Pour plus d'informations, consultez la rubrique sur la création des index de bases de données dans la documentation FileNet P8.

2. Dans la liste des résultats, affichez la colonne **CmIndexingFailureCodes** pour déterminer le type d'incident d'indexation pour l'objet. Consultez la liste des codes d'incident d'indexation pour trouver la condition ou l'événement à l'origine de l'erreur et savoir ce que vous pouvez faire pour gérer cette erreur ou empêcher une répétition de l'incident.

Référence associée :

«Contrôle de l'indexation avec IBM Content Collector P8 Content Search Services Support», à la page 222

Information associée :

 Centre de documentation d'IBM FileNet P8 Version 5.1

Erreurs de traitement de documents Content Search Services Support

Les avertissements et erreurs de traitement des documents sont générées par les services de filtre de conversion des documents dans IBM Content Search Services et par le plug-in du constructeur du document Content Search Services Support.

Le tableau suivant répertorie les erreurs (les ID de message qui se terminent par un E) et les avertissements (les ID de message qui se terminent par un W) de traitement et d'indexation des documents.

Les valeurs numériques sont les codes incident d'indexation (CmIndexingFailureCodes) qui sont ajoutés à la table de base de données DocVersion et peuvent être interrogés pour les documents dans une liste de résultats de recherche.

Le code d'incident d'indexation pour chaque courrier électronique traité est une valeur binaire qui correspond exactement à l'une des descriptions de codes d'échec dans la liste des erreurs de traitement des documents, ou est une combinaison d'un ou plusieurs codes d'incident dans laquelle la valeur de chaque code d'erreur est logiquement combiné à l'aide de l'OR de l'opérateur. Par exemple, un code d'incident d'indexation 161 est une combinaison des codes d'erreur 1, 32 et 128, où les valeurs 1, 32 et 128 donnent une valeur globale de 161.

Les ID des messages sont consignés dans le fichier journal Content Search Services Support si la consignation est activée.

Tableau 22. Erreurs de traitement Content Search Services Support

Valeur	ID du message	Code d'incident et description
1	N/A	<p>Le contenu textuel d'un document n'a pas pu être indexé pour une raison inhabituelle.</p> <p>Cette erreur peut être émise par IBM Content Search Services.</p> <p>Action utilisateur : recherchez les messages d'erreur liés à l'incident qui ont été consignés pour un document dans le journal du serveur FileNet P8.</p>
2	IQQP0012W	<p>TRUNCATED_DOCUMENT:</p> <p>Le contenu textuel d'une ou plusieurs pièces jointes intégrées a été tronqué en raison d'une limitation de taille.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion de documents dans IBM Content Search Services et Content Search Services Support.</p> <p>Action de l'utilisateur : modifiez le paramètre de configuration de IBM Content Search Services maxTextSize. Pour plus d'informations, consultez la section de résolution du troncage de texte extrait d'un document lors de l'indexation dans la documentation FileNet P8.</p>

Tableau 22. Erreurs de traitement Content Search Services Support (suite)

Valeur	ID du message	Code d'incident et description
4	AFU_DOC_ENCRYPTED	<p>AFU_DOC_ENCRYPTED:</p> <p>Le document de courrier électronique est lui-même chiffré.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion des documents dans IBM Content Search Services.</p> <p>Action de l'utilisateur : les documents chiffrés ne peuvent pas être déchiffrés ; leur contenu ne peut donc pas être ajouté à l'index. D'autres propriétés telles que l'objet et la liste des destinataires sont indexées et peuvent être interrogées.</p>
8	IQQP0031E	<p>PASSWORD_PROTECTED_DOCUMENT:</p> <p>Le contenu dans une ou plusieurs pièces jointes intégrées n'a pas été indexé car il a été protégé par un mot de passe ou chiffré.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion des documents dans IBM Content Search Services.</p> <p>Action de l'utilisateur : les documents chiffrés ne peuvent pas être déchiffrés ; leur contenu ne peut donc pas être ajouté à l'index.</p>
16	IQQP0032E	<p>MALFORMED_DOCUMENT:</p> <p>Une ou plusieurs pièces jointes intégrées n'ont pas été indexées car le fichier est corrompu.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion des documents dans IBM Content Search Services.</p> <p>Action de l'utilisateur : le contenu des pièces jointes ne peut pas être ajouté à l'index.</p>
32	IQQP0008W	<p>UNSUPPORTED_TYPE:</p> <p>Le contenu d'une ou plusieurs pièces jointes de courrier électronique n'a pas été indexé car le format de fichier n'est pas pris en charge.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion des documents dans IBM Content Search Services.</p> <p>Action de l'utilisateur : essayez de traiter à nouveau le document de courrier électronique après l'installation de nouveaux filtres de conversion de document.</p>

Tableau 22. Erreurs de traitement Content Search Services Support (suite)

Valeur	ID du message	Code d'incident et description
64	IQQP0027W	<p>NO_TEXT_EXTRACTED:</p> <p>Une ou plusieurs des pièces jointes intégrées ne comportent aucun contenu textuel.</p> <p>Cette erreur peut être émise par Content Search Services Support.</p> <p>Les raisons peuvent être les suivantes :</p> <ul style="list-style-type: none"> • Les zones de contenu dans le corps du courrier électronique n'ont pas pu être traitées. • Les services de filtre de conversion de document dans IBM Content Search Services a renvoyé une erreur de traitement du document. • Le contenu textuel renvoyé par les services de filtre de conversion de document dans IBM Content Search Services n'a pas pu être traité. • Les services de filtre de conversion de document dans IBM Content Search Services ont provoqué une exception. <p>Action de l'utilisateur : vérifiez que le document d'origine n'est pas vide ou endommagé. Essayez de traiter à nouveau le document de courrier électronique après l'installation de nouveaux filtres de conversion de document.</p>
128	IQQP0026E	<p>UNRECOGNIZED_ENCODING:</p> <p>Le codage du texte dans une ou plusieurs pièces jointes intégrées n'a pas pu être reconnu car la quantité de texte est trop faible ou plusieurs codages de texte sont mélangés.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion des documents dans IBM Content Search Services.</p> <p>Action de l'utilisateur : le contenu des pièces jointes ne peut pas être correctement ajouté à l'index.</p>
256	AFU_CNV_DISABLED	<p>AFU_CNV_DISABLED:</p> <p>La conversion du texte dans les pièces jointes a été désactivée.</p> <p>L'option de configuration AttachmentTextConversion a été désactivée (définie sur 0) lors de l'indexation, ce qui signifie que Content Search Services Support ne traite pas les pièces jointes intégrées aux messages électroniques et le contenu textuel de filtre.</p> <p>Cette erreur peut être émise par Content Search Services Support.</p> <p>Action utilisateur : activez AttachmentTextConversion dans le fichier <code>constructors.xml</code> dans le sous-répertoire <code>config</code> où IBM Content Search Services est installé.</p>

Tableau 22. Erreurs de traitement Content Search Services Support (suite)

Valeur	ID du message	Code d'incident et description
512	AFU_DOC_PARTIAL	<p>AFU_DOC_PARTIAL:</p> <p>Le contenu textuel d'une ou plusieurs pièces jointes intégrées a été tronqué en raison d'une limitation de taille.</p> <p>Cette erreur peut être émise par Content Search Services Support si les données XML générées servant d'entrée à IBM Content Search Services dépassent la limite configurée dans IBM Content Search Services.</p> <p>Action de l'utilisateur : modifiez le paramètre de configuration de IBM Content Search Services maxXmlTextSize. Pour plus d'informations, consultez la section de résolution du troncage de texte extrait d'un document lors de l'indexation dans la documentation FileNet P8.</p>
1024	IQQP0028E	<p>MALFORMED_XML_DOCUMENT:</p> <p>Une ou plusieurs pièces jointes intégrées n'ont pas pu être indexées car le format XML du fichier n'est pas valide.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion de documents dans Content Search Services Support et Content Search Services Support.</p> <p>Action de l'utilisateur : cette erreur ne doit pas se produire. Contactez le service de support logiciel IBM.</p>
2048	IQQP0033W	<p>FALLBACK_PARSER_INVOKED:</p> <p>L'analyseur syntaxique du document principal n'a pas réussi à analyser le document. L'analyseur de secours a été appelé. Le contenu peut avoir été partiellement indexé.</p> <p>Cette erreur peut être émise uniquement par les services de filtre de conversion de document dans IBM Content Search Services.</p> <p>Action de l'utilisateur : aucune action n'est requise. L'erreur est à titre d'information uniquement.</p>
4096	IQQP0013E	<p>SIZE_LIMIT_EXCEEDED:</p> <p>La taille du fichier binaire ou XML d'une ou plusieurs pièces jointes est supérieure à la limite prédéfinie par collecte.</p> <p>Cette erreur peut être émise par les services de filtrage de conversion des documents dans IBM Content Search Services.</p> <p>Action de l'utilisateur : modifiez la limitation de taille. Pour plus d'informations, consultez la section de résolution du troncage de texte extrait d'un document lors de l'indexation dans la documentation FileNet P8.</p>

Tâches associées :

«Vérification de votre configuration IBM Content Search Services», à la page 210

Recherche de chaînes d'erreur

Outre l'interrogation les codes d'erreur d'indexation dans la table de base de données DocVersion pour les documents répertoriés dans une liste des résultats de recherche, certains documents sont indexés avec une chaîne de notification indiquant qu'une erreur a été détectée lors du traitement du document.

Vous pouvez utiliser ces chaînes de message dans une requête de recherche de texte IBM Content Search Services pour identifier les documents qui ont été indexés avec ces chaînes de message et vous pouvez ensuite réindexer ces documents, par exemple après qu'un document chiffré est déchiffré, ou après que de nouveaux filtres de document ont été installés pour résoudre des problèmes de conversion de l'ancien document.

Content Search Services Support ajoute les chaînes de recherche d'erreur suivantes à l'index de texte intégral :

Tableau 23. Chaînes de recherche d'erreur ajoutées à l'index

Chaîne de recherche	Description
AFU_DOC_PARTIAL	Indique que le document de courrier électronique n'a été traité que partiellement. Le corps, ou une ou plusieurs des pièces jointes intégrées, a peut-être été tronqué.
AFU_DOC_ENCRYPTED	Indique que le document de courrier électronique lui-même est chiffré.
AFU_CNV_DISABLED	Indiqué que la conversion de texte dans les pièces jointes intégrées à des courriers électroniques a été désactivée lors de l'indexation, ce qui signifie que Content Search Services Support ne traite pas les pièces jointes intégrées aux messages électroniques et ne filtre pas le contenu textuel.
AFU_DOC_PRIVATE	Indique que le document de courrier électronique a été marqué comme étant un document privé.
AFU_DOC_ATT_EMPTY	Indique que le document de courrier électronique a été indexé, mais qu'une ou plusieurs de ses pièces jointes n'ont pas été indexées car leur taille était de zéro octet.
V3.0.0.2 : AFU_DOC_BODY_EMPTY	Indique que le document électronique qui a été indexé contient un corps vide. Un corps vide est un corps contenant des données textuelles qui ne peuvent pas être indexées, par exemple un corps sans caractères ou contenant seulement des espaces (caractères avec points de code 0x01 à 0x20).
AFU_DOC_ERROR	Indique qu'une erreur s'est produite pendant le traitement du document de courrier électronique. Notez que cette chaîne de recherche ne peut pas toujours être ajoutée à l'index. Si une erreur grave se produit lors du traitement du document, par exemple, si le document ne peut pas être extrait, aucune représentation de texte n'est créée pour le document et aucune chaîne de recherche ne peut donc être ajoutée. En conséquence, aucune chaîne de recherche n'est indexée pour ces erreurs graves.

Par exemple, vous pouvez rechercher des documents qui ont été chiffrés à l'aide de la requête suivante :

```
SELECT i.This FROM ICCMail3 i
  INNER JOIN ContentSearch cs ON i.This = cs.QueriedObject
 WHERE CONTAINS (*, '@xmlxp:''//icc_doc_properties[.contains
  ('''AFU_DOC_ENCRYPTED''')'])
```

Débogage des documents introuvables

Si vous ne parvenez pas à trouver des documents dans l'index lorsque vous exécutez une requête de recherche, vous pouvez analyser ces documents pour essayer de trouver la cause de l'erreur en générant des fichiers de vidage de ces documents.

En règle générale, lorsque vous activez les options de configuration de vidage à des fins de débogage, les fichiers de vidage sont créés pour tous les documents dans la file d'attente d'indexation. Comme le nombre de documents dans la file d'attente d'indexation peut être très volumineux, l'activation de l'option de vidage peut conduire à la génération d'un très grand nombre de fichiers de vidage.

En définissant l'option de configuration **DumpDocumentIds**, vous pouvez spécifier uniquement un sous-ensemble des documents pour lesquels les fichiers de vidage sont créés dans le répertoire de vidage spécifié.

Pour limiter le nombre de documents pour lesquels les fichiers de vidage sont créés :

1. Déterminez l'ID des documents que vous espérez trouver dans l'index à l'aide de la file d'attente d'indexation dans IBM FileNet Enterprise Manager.
2. Définissez **DumpDocumentIds** dans la configuration du document source Content Search Services Support dans le Gestionnaire de configuration sous **Paramètres généraux > Content Search Services Support**. Ajoutez **DumpDocumentIds** dans la section [Paramètres] et répertoriez chaque document à l'aide de son ID. Chaque ID de document doit apparaître entre accolades et être séparé par un espace, par exemple,

```
DumpDocumentIds =
{50F0181B-9FE3-5718-5383-9A5ECA35E14} {361B375F-98C8-476D-BBAE-F3AD4778FE81}
```

Le répertoire de vidage et au moins l'un des paramètres de vidage, à savoir **Créer des fichiers de vidage pour tous les documents d'entrée**, **Créer des fichiers de vidage du fichier XML pour l'indexation** ou **Créer des fichiers de vidage pour le contenu textuel des pièces jointes** doivent être spécifiés.

Réindexation de documents

Indexez à nouveau les documents pour tenter de retraiter les objets exécutés sur des erreurs spécifiques ou indexés au cours d'une période donnée avant que des modifications ou des adaptations importantes aient été apportées au système.

Après avoir regroupé vos erreurs dans des erreurs d'avertissement et des erreurs plus graves, vous pouvez choisir entre plusieurs options de réindexation afin de recréer les informations d'index pour les objets.

A faire : La réindexation d'objets ayant provoqué des avertissements ou des erreurs n'est efficace que si des configurations d'index ont été modifiées ou corrigées, par exemple, si des délais d'attente ont été ajustés ou de nouvelles mises à jour installés, si des filtres de document supplémentaires ont été ajoutés, les filtres existants ont été améliorés ou une nouvelle version du moteur de recherche de contenu a été installée. Les erreurs graves se produisent généralement si le processus d'indexation est arrêté de manière inattendue lors du traitement d'un

objet car l'objet a été endommagé ou altéré de façon inattendue. Ces objets doivent être analysés et peut-être corrigés avant qu'une nouvelle tentative d'indexation soit effectuée. Des erreurs graves peuvent également se produire si le processus est à court de mémoire.

Les options de réindexation comprennent :

- Réindexation d'objets individuels
- Réindexation d'objets en tant que groupe

La réindexation nécessite toujours la recherche dans une table de base de données DocVersion pour les objets qui remplissent les conditions de réindexation demandées et peut prendre un temps considérable en fonction du nombre de documents déjà indexés. Une analyse complète de la table est nécessaire sauf si vous créez manuellement des index de base de données supplémentaires.

Utilisez Enterprise Manager pour rechercher les objets présentant des erreurs d'indexation, comme décrit dans la rubrique relative à l'identification des incidents d'indexation. Réindexez les objets en fonction des résultats de recherche :

- Pour réindexer un objet individuel, cliquez avec le bouton droit sur l'objet dans la liste des résultats et cliquez sur **Toutes les tâches > Index pour Content Search**.
- Pour réindexer un petit ensemble de documents sélectionnés manuellement, sélectionnez les objets dans la liste des résultats. Cliquez avec le bouton droit de la souris et sélectionnez **Opérations de multisélection**. Sur l'onglet Action de la fenêtre Opérations en vrac de Content Engine, sélectionnez **Index pour Content Search**. Cliquez sur **OK** pour soumettre les objets sélectionnés à la réindexation.
- Pour réindexer tous les documents qui répondent à des critères spécifiques, limitez votre requête à un code incident spécifique, par exemple. Sur l'onglet Action de Content Engine Query Builder, sélectionnez **Index pour Content Search**. Cliquez sur **OK** pour rechercher et soumettre automatiquement tous les résultats à la réindexation.
- Vous pouvez sinon écrire votre propre outil d'identification et de réindexation des objets avec des incidents d'indexation.

Contrôle de l'indexation avec IBM Content Collector P8 Content Search Services Support

Vous pouvez améliorer les performances d'indexation en surveillant le processus d'indexation et les files d'attente et en ajustant les paramètres de configuration d'un serveur IBM Content Search Services en conséquence.

Surveillez les informations suivantes :

- Les fichiers journaux IBM Content Search Services
- Les fichiers journaux Content Engine, notamment le fichier `p8_server_error.log`
- Les fichiers journaux IBM Content Collector P8 Content Search Services Support
- Les files d'attente d'entrée et de sortie pour l'indexation
- Les tables de base de données spécifiques

Fichiers journaux d'IBM Content Search Services

Activez la consignation du serveur IBM Content Search Services comme décrit dans la section relative à la configuration de la consignation de serveur dans la documentation FileNet P8.

Le serveur IBM Content Search Services génère des informations de consignation au cours du démarrage du serveur, de l'indexation et de la recherche. Par défaut, les fichiers journaux IBM Content Search Services sont écrits dans le répertoire <ContentSearchServicesInstallDir>\log.

Fichiers journaux d'IBM FileNet Content Engine

Content Engine consigne les erreurs d'indexation dans le fichier p8_server_error.log par défaut. Le serveur Content Engine fournit la consignation de trace dans le fichier p8_server_trace.log si vous activez le sous-système CBR pour le traçage. Pour plus de détails, consultez la section sur les fichiers journaux Content Engine dans la documentation FileNet P8.

Pour surveiller le rendement d'indexation, activez la fonction de trace CBR récapitulative. Vous pouvez ensuite analyser le fichier p8_server_trace.log pour savoir combien d'objets ont été envoyés au serveur IBM Content Search Services pour l'indexation, comme dans l'exemple suivant :

```
for filename in $(ls -r /opt/IBM/WebSphere/AppServer/profiles/AppSrv01/FileNet/server1/*trace*)
do
  grep "Dispatching" $filename > /qaTools/results/$HONAME.ce2cbrDispatch.$(basename $filename)
  grep "Deleted" $filename > /qaTools/results/$HONAME.ce2cbrDe1Updates.$(basename $filename)
  grep "NumberOfObjects" $filename > /qaTools/results/$HONAME.ce2cbrIndexing.$(basename $filename)
  grep "Dispatching" $filename >> /qaTools/results/$HONAME.ce2cbrDispatch.txt
  grep "Deleted" $filename >> /qaTools/results/$HONAME.ce2cbrDe1Updates.txt
  grep "NumberOfObjects" $filename >> /qaTools/results/$HONAME.ce2cbrIndexing.txt
done
```

Importez les résultats dans une feuille de calcul et créez des graphiques des données pour avoir une idée du débit d'indexation.

Fichiers journaux de Content Search Services Support

Pour activer les options de consignation, de traçage et de vidage à des fins de débogage dans Content Search Services Support, vous devez définir les options respectives dans le fichier de configuration Content Search Services Support dans le Gestionnaire de configuration IBM Content Collector.

Les options de consignation, de trace et de date/heure de Content Search Services Support sont communes à tous les préprocesseurs de document source configurés et lorsqu'elles sont activées, s'appliquent à tous les documents qui sont prétraités par Content Search Services Support. Les options de vidage peuvent être définies spécifiquement pour chaque préprocesseur de document source.

Consignation

Activez la consignation en sélectionnant un niveau de journalisation des **Informations** ou de la **Trace** dans la fenêtre de configuration des paramètres de journal communs. Si la consignation est activée, Content Search Services Support écrit les informations de consignation dans le répertoire de fichier journal spécifié. Les fichiers journaux utilisent des horodatages dans leurs noms.

Fonction de date/heure

Activez la synchronisation pour écrire des informations temporelles supplémentaires dans les fichiers journaux. Les informations temporelles seront écrites dans les fichiers journaux uniquement si la consignation est activée.

Traçage

Activez le traçage en définissant l'option de configuration d'indexation **Traçage**. Si le traçage est activé, des informations de trace détaillées sont

écrites dans les fichiers journaux. Les noms de fichier journal contenant des informations de trace utilisent également des horodatages dans les noms de fichier.

Indiquez également le nombre de fichiers de trace qui sont générés par la définition de l'option de configuration d'indexation **TraceFileCount** et déterminez la taille maximale d'un fichier de trace en définissant l'option de configuration d'indexation **TraceFileLimitInMB**.

Fichiers de vidage

Activez la création des fichiers de vidage en sélectionnant une ou plusieurs options :

Créer des fichiers de vidage pour tous les documents d'entrée

Content Search Services Support écrit les fichiers de vidage de tous les documents d'entrée dans le sous-répertoire doc du répertoire de vidage spécifié. Une copie du document d'origine est créée dans le format (binaire) et le codage d'origine ; chaque nom de fichier de vidage contient l'identificateur du document.

Créer des fichiers de vidage du fichier XML pour indexation

Content Search Services Support écrit les fichiers de vidage de tous les documents XML générés dans le sous-répertoire xml du répertoire de vidage spécifié. Une copie des documents XML générés est créée ; chaque nom de fichier de vidage contient l'identificateur du document.

Créer des fichiers de vidage pour le contenu textuel des pièces jointes

Content Search Services Support écrit les fichiers de vidage de tout le contenu textuel des pièces jointes imbriquées dans le sous-répertoire txt du répertoire de vidage spécifié. Une copie du contenu textuel de la pièce jointe intégrée est créée ; chaque nom de fichier de vidage contient l'identificateur du document et le nom de la pièce jointe comme retourné par les services de filtre de conversion de document dans IBM Content Search Services.

Spécifiez également le nom du répertoire dans lequel les fichiers de vidage sont écrits dans les sous-répertoires correspondants. Le répertoire par défaut est ./log/dump (Le répertoire de vidage dans le sous-répertoire du journal dans lequel IBM Content Search Services est installé).

Contrôle des files d'attente

Contrôler les files d'attente pendant que les documents sont indexés peut vous aider à identifier les zones d'optimisation en adaptant les paramètres spécifiques dans la configuration de serveur IBM Content Search Services. Vous pouvez également utiliser les informations pour créer un diagramme du débit avec une granularité plus fine concernant ce que fait réellement le serveur IBM Content Search Services.

Pour surveiller des files d'attente, ajoutez l'élément `<monitorQueues>valeur</monitorQueues>` au fichier `<ContentSearchServicesInstallDir>\config\config.xml`. Pour la *valeur*, spécifiez un entier non négatif qui indique la fréquence d'impression (en secondes). Redémarrez ensuite le serveur IBM Content Search Services.

Les informations sur l'état de la file d'attente sont écrites dans un fichier CSV dans le répertoire `<ContentSearchServicesInstallDir>\log`. Le fichier `QueueStatus.csv` fournit les colonnes d'informations ci-après.

- Heure actuelle
- Nombre total de documents traités
- Taille totale des documents traités (en ko)
- Nombre de documents dans la file d'attente d'entrée
- Taille de la file d'attente d'entrée (en octets)
- Nombre de documents dans la file d'attente de sortie
- Taille de la file d'attente de sortie (en octets)
- Nombre de documents en attente de prétraitement
- Nombre de documents en cours de prétraitement
- Nombre de documents en attente d'indexation
- Nombre de documents en cours d'indexation

La taille du fichier QueueStatus.csv continue d'augmenter lorsque la surveillance de la file d'attente est activée. Vous pouvez désactiver la surveillance de la file d'attente en spécifiant une valeur de zéro pour le paramètre **monitorQueues** ou en supprimant la *valeur* `<monitorQueues></monitorQueues>` de l'élément dans le fichier config.xml. N'oubliez pas que toutes les modifications apportées au fichier de configuration nécessitent un redémarrage du serveur IBM Content Search Services.

Important : Le fichier QueueStatus.csv est recréé chaque fois que le serveur est démarré. Si vous souhaitez stocker les informations, créez une copie de sauvegarde du fichier avant de redémarrer le serveur.

Contrôle des tables de base de données

Utilisez les requêtes de base de données pour contrôler régulièrement ces tables :

- IndexRequests
- ContentQueue

En guise de meilleure pratique, exécutez les requêtes sur ces tables dans une boucle et collectez la sortie dans un fichier journal.

Tâches associées :

«Identification des incidents d'indexation lors de l'indexation avec IBM Content Search Services», à la page 214

«Vérification de votre configuration IBM Content Search Services», à la page 210

Exemples de requêtes de base de données pour contrôler les tables de base de données

Utilisez les requêtes de base de données comme exemples de requêtes afin de contrôler la charge d'indexation pour les scénarios à volume élevé.

Ces exemples de requêtes de base de données fournissent les informations suivantes :

- Le nombre d'objets de document qui ont été archivés dans la classe ICCMail3.

```
-x db2 "select count(*) as Number_of_email_instances from ClassDefinition CD, Docversion D où D.object_class_id = CD.object_id etsymbolic_name='ICCMail3' for fetch only "
```
- Le nombre d'instances de chaque objet. Le nombre d'instances indique le nombre de doublons d'un document qui ont été archivés. Initialement, chaque document possède une seule instance. Pour chaque doublon, Content Collector crée une autre instance dans la classe d'objet personnalisé ICCMailInstance3.

```
-x db2 "select count(*) as Number_of_email_instances from ClassDefinition CD,  
G générique où G.object_class_id = CD.object_id et  
symbolic_name='ICCMailInstance3' for fetch only"
```

- Le nombre d'objets qui doivent toujours être indexés. Il s'agit de l'accumulation d'indexation.

```
db2 -x "select count(*) from indexrequests for fetch only"
```

- Le nombre d'objets (par exemple, courrier électronique, pièces jointes ou fichiers) qui ont été archivés dans le conteneur d'objets mais n'ont pas encore été déplacés vers leur emplacement permanent, tels que la zone de stockage fixe, ou qui ont été marqués pour suppression par le gestionnaire d'expiration mais n'ont pas encore été supprimés de la zone de stockage.

```
db2 -x "select count(*) from contentqueue for fetch only"
```

Important : Si vous prévoyez d'utiliser des requêtes plus complexes, vous devez également créer les index de base de données appropriés. Pour les propriétés système, demandez à votre administrateur de base de données de créer l'index approprié. Pour plus de détails, consultez la rubrique sur la création d'index de base de données dans la section sur l'extraction du contenu avec IBM Content Search Services dans la documentation FileNet P8.

Identification et résolution des problèmes liés à Content Search Services Support

Cette section comporte des informations utiles à l'identification et à la résolution des problèmes susceptibles de se produire lorsque vous utilisez IBM Content Collector P8 Content Search Services Support.

En plus des informations disponibles dans cette section, consultez la section relative à l'identification et la résolution des problèmes d'IBM Content Search Services dans le Centre de documentation d'IBM FileNet P8 Version 5.1.

Aucun document traité ou indexé

Bien que FileNet P8 et IBM Content Search Services soient correctement installés, les documents ne sont pas traités à l'aide de Content Search Services Support.

Symptômes

Le fichier journal de Content Search Services Support consigne le message suivant :

```
AFUI0024E Une erreur irrémédiable s'est produite lors du chargement de la configuration,  
laquelle a été extraire du serveur de configuration distant.
```

Consultez le message suivant :

```
AFUI0082E La balise obligatoire 'protocol' n'a pas été trouvée dans la configuration  
distante. Vérifiez qu'une connexion à un référentiel valide est configurée  
pour Content Search dans IBM Content Collector Configuration Manager.
```

Causes

La cause la plus probable de ce message d'erreur est qu'aucune information valide de connexion au référentiel FileNet P8 n'a été spécifiée pour Content Search Services Support dans le Gestionnaire de configuration IBM Content Collector.

Résolution de l'incident

Assurez-vous que la connexion au référentiel FileNet P8 est correctement définie pour chaque document source installé dans le Gestionnaire de configuration IBM Content Collector sous **Paramètres généraux > Content Search Services Support**.

Aucun document indexé

Bien que Content Search Services Support soit installé et configuré, aucun document n'est indexé.

Symptômes

Vous vérifiez l'installation et la configuration de Content Search Services Support en archivant un exemple de document dans la classe de document appropriée à l'aide de la route de tâches d'archivage dans IBM Content Collector.

Si aucune demande d'index n'est faite pour ce document dans Enterprise Manager, il se peut que la configuration de Content Search Services Support ne soit pas correcte pour chaque instance IBM Content Search Services dans laquelle Content Search Services Support est installé.

Résolution de l'incident

1. Vérifiez que chaque instance IBM Content Search Services pour laquelle vous avez installé Content Search Services Support contient le fichier de configuration `constructors.xml` dans le répertoire `<ContentSearchServicesInstanceInstallation>\config\`.
La configuration de Content Search Services Support est divisée en sections distinctes, une pour chaque source de document sélectionné. Le fichier de configuration comporte le nom du préprocesseur de document, la classe de document, les informations relatives au chemin d'accès à la base de données de configuration de Content Collector ainsi que le nom symbolique.
2. Vérifiez que les informations relatives au chemin d'accès à la base de données de Content Collector sont correctes et que le service IBM Content Collector Web Application est correctement configuré et en cours d'exécution.
3. Assurez-vous que des noms symboliques corrects ont été spécifiés.

Le nom symbolique indique les classes de document qui sont traitées par Content Search Services Support. Le nom symbolique est défini par Content Collector lorsque le magasin d'objets est configuré pour IBM Content Search Services. Si vous utilisez des noms de classes de documents personnalisés, vous devez les ajouter manuellement à la configuration.

Partie 3. Annexes

Remarques

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service puisse être utilisé. Tout autre élément fonctionnellement équivalent peut être utilisé, s'il n'enfreint aucun droit d'IBM. Toutefois, l'utilisateur est chargé d'évaluer et de vérifier le fonctionnement d'un produit, d'un programme ou d'un service non IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits mentionnés dans le présent document. La remise de ce document ne vous donne aucun droit de licence sur ces brevets ou demandes de brevet. Si vous désirez recevoir des informations concernant l'acquisition de licences, veuillez en faire la demande par écrit à l'adresse suivante :

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Pour le Canada, veuillez adresser votre courrier à :

IBM Director of Commercial Relations
IBM Canada Ltd.
3600 Steeles Avenue East
Markham, Ontario
L3R 9Z7 Canada

Les informations sur les licences concernant les produits utilisant un jeu de caractères double octet peuvent être obtenues par écrit à l'adresse suivante d'IBM :

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan, Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japon

Le paragraphe suivant ne s'applique pas au Royaume-Uni ni à tout autre pays dans lesquels lesdites clauses ne sont pas conformes à la législation en vigueur :
INTERNATIONAL BUSINESS MACHINES CORPORATION FOURNIT CETTE PUBLICATION «EN L'ETAT» SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE Y COMPRIS, MAIS SANS S'Y LIMITER, LES GARANTIES IMPLICITES DE NON VIOLATION, DE QUALITE MARCHANDE OU D'ADEQUATION A DES FINS PARTICULIERES. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et logiciels décrits dans ce document.

Les références à des sites Web non IBM sont fournies à titre d'information uniquement et n'impliquent en aucun cas une adhésion aux données qu'ils contiennent. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation de sa part, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci), et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

IBM Deutschland GmbH
Department M358
IBM-Allee 1
71139 Ehningen
Allemagne

Ces informations peuvent être soumises à des conditions particulières, prévoyant notamment le paiement d'une redevance.

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions du Livret Contractuel IBM, des Conditions Internationales d'Utilisation de Logiciels IBM ou de tout autre contrat équivalent.

Les données de performance indiquées dans ce document ont été déterminées dans un environnement contrôlé. Par conséquent, les résultats peuvent varier de manière significative selon l'environnement d'exploitation utilisé. Certaines mesures évaluées sur des systèmes en cours de développement ne sont pas garanties sur tous les systèmes disponibles. En outre, elles peuvent résulter d'extrapolations. Les résultats peuvent donc varier. Il incombe aux utilisateurs de ce document de vérifier si ces données sont applicables à leur environnement d'exploitation.

Les informations concernant des produits non IBM ont été obtenues auprès des fournisseurs de ces produits, par l'intermédiaire d'annonces publiques ou via d'autres sources disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Elle ne peut recevoir aucune réclamation concernant des produits non IBM. Toute question concernant les performances de produits non IBM doit être adressée aux fournisseurs de ces produits.

Toute instruction relative aux intentions d'IBM pour ses opérations à venir est susceptible d'être modifiée ou annulée sans préavis, et doit être considérée uniquement comme un objectif.

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des

noms fictifs de personnes, de sociétés, de marques ou de produits à des fins illustratives ou explicatives uniquement. Toute ressemblance avec des noms de personnes, de sociétés ou des données réelles serait purement fortuite.

LICENCE DE COPYRIGHT :

Le présent logiciel contient des exemples de programme d'application en langage source destinés à illustrer les techniques de programmation sur différentes plateformes d'exploitation. Vous avez le droit de copier, de modifier et de distribuer ces exemples de programmes sous quelque forme que ce soit et sans paiement d'aucune redevance à IBM, à des fins de développement, d'utilisation, de vente ou de distribution de programmes d'application conformes aux interfaces de programmation des plateformes pour lesquels ils ont été écrits ou aux interfaces de programmation IBM. Ces exemples de programmes n'ont pas été rigoureusement testés dans toutes les conditions. Par conséquent, IBM ne peut garantir expressément ou implicitement la fiabilité, la maintenabilité ou le fonctionnement de ces programmes.

Marques

IBM, le logo IBM et `ibm.com` sont des marques d'International Business Machines Corporation aux Etats-Unis et/ou dans certains autres pays. Les autres noms de produit et de services peuvent être des marques d'IBM ou appartenir à des tiers. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web "Copyright and trademark information" à l'adresse <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques d'Adobe Systems Incorporated aux Etats-Unis et/ou dans certains autres pays.

Java ainsi que tous les logos et toutes les marques incluant Java sont des marques d'Oracle et/ou de ses sociétés affiliées.

Microsoft, SharePoint et Windows sont des marques de Microsoft Corporation aux Etats-Unis et/ou dans certains autres pays.

Linux est une marque de Linus Torvalds aux Etats-Unis et/ou dans certains autres pays.

UNIX est une marque enregistrée de The Open Group aux Etats-Unis et/ou dans certains autres pays.

Le programme Oracle Outside In Technology ci-inclus fait l'objet d'une licence à utilisation limitée et peut seulement être utilisé conjointement à la présente application.

Les autres noms de produit et de services peuvent être des marques d'IBM ou appartenir à des tiers.

Index

A

afuIndexer
Voir indexeur de recherche de texte

C

Content Search Services Support
 installation 190
 installation en mode console 192
 installation en mode silencieux 194
contrôle de l'indexation 222
contrôle des performances système
 indexeur de recherche de texte 46

E

Ensembles de styles IBM Legacy Content Search Engine 174
Ensembles de styles Verity 174
environnements en langue mixte
 indexation avec IBM Legacy Content Search Engine 175

F

FileNet P8
 contrôle de l'indexation 222
 courrier électronique 212
 Ensembles de styles IBM Legacy Content Search Engine 174
 extraction de contenu 211, 212
 fichiers 211
 IBM Content Search Services 187
 IBM Legacy Content Search Engine 169
 IBM Legacy Content Search Engine Verity Autonomy K2 169
 incidents d'indexation en texte intégral 179, 184, 214, 225
 indexation en texte intégral 169, 174, 187, 211, 212
 modification de l'ensemble de styles 172
 moteur de recherche de contenu P8 169, 187
 surveillance de l'indexation 183
Formulaire de configuration
 prise en charge de la recherche de texte 10

I

IBM Content Collector P8 Content Search Services Support
 concepts d'indexation 185
 description 185
 étapes de traitement 213
 installation 189

IBM Content Collector Text Search Support
 installation 8, 9
IBM Content Search Services
 contrôle de l'indexation 222
 incidents d'indexation en texte intégral 214
 réindexation des objets 181, 221
 requêtes de base de données 225
IBM Legacy Content Search Engine
 concepts d'indexation 168
 ensembles de styles 174
 incidents d'indexation en texte intégral 179
 indexation avec une configuration en langue mixte 175
 modification de l'ensemble de styles 172
 réindexation des objets 181, 221
 requêtes de base de données 184
 surveillance de l'indexation 183
identification et résolution des problèmes Content Search Services Support 226
indexeur de recherche de texte 150
incidents d'indexation 179, 214
indexation
 IBM Legacy Content Search Engine 167, 168
 indexation de recherche de texte
 recherche de chaînes d'erreur 114
indexation en texte intégral
 taille de la partition d'index Autonomy K2 177
 taille de la partition d'index IBM Legacy Content Search Engine 177
 taille de la partition d'index Verity 177
indexeur de recherche de texte
 Voir aussi indexeur de recherche de texte
afuIndexer 25, 90
ajout d'attributs
 au fichier de configuration du type d'élément 57
 au fichier modèle 56
arguments 25
arrêt du traitement 111
arrêt forcé 111
attributs de type d'élément 58
codes retour 37
codes retour du processus 36, 100
collecte de statistiques 50
configuration du type d'élément 20, 66
contraintes 88
débogage 108
exécution de afuIndexer en mode continu 92
exécution de l'outil de vérification de la configuration 130

indexeur de recherche de texte (*suite*)
 extraction des éléments du gestionnaire de ressources IBM Content Manager 110
 Fichiers journaux d' 48
 fichiers modèles 55
 fichiers trace 50
 indexation des courriers électroniques avec des pièces jointes 94
 indexation des types d'élément créés par IBM Content Manager 45
 informations de date/heure 53
 messages d'erreur 146, 147, 148, 149
 opérations à effectuer en cas d'arrêt brutal de l'indexation 105
 opérations avancées 54
 options de configuration 67
 préparation à l'indexation 54
 présentation 3, 6, 11
 présentation des tables de base de données 145
 principales étapes 11
 recherche de chaînes d'erreur 114
 recréation d'un index 134
 réindexation d'éléments 103
 réorganisation d'index 112
 réparation d'index 116
 sauvegarde des fichiers 145
 suppression des types d'élément 146
 surveillance 46
 taille du tampon de transfert 87
 traitement de plusieurs types d'élément 95
 traitement des éléments avec erreurs graves 106, 107
 traitement des pièces jointes 95
 types d'élément CommonStore existants 43
 utilisation de plusieurs processus de traitement à unité d'exécution simple 90
 utilisation de TIEFLAG 62
 validation d'index 116
 variables d'environnement 63
 vérification du statut des éléments indexés 37
installation
 Content Search Services Support 190, 192, 194
 IBM Content Collector P8 Content Search Services Support 189
 IBM Content Collector Text Search Support 8, 9

J

journaux
 indexeur de recherche de texte 48

N

- nouveautés
 - indexation
 - IBM Content Manager 3
 - IBM FileNet P8 167
- nouvelles fonctions
 - indexation
 - IBM Content Manager 3
 - IBM FileNet P8 167

O

- outil de validation d'index
 - critères 126
 - étapes de traitement 126
- outils de configuration
 - activation
 - type d'élément existant à des fins de traitement par l'indexeur pour la recherche de texte 15
 - avant la création d'un type d'élément 13
 - configuration
 - référentiel FileNet P8 188
 - référentiel IBM Content Manager 11

P

- paramètres de configuration
 - indexeur de recherche de texte 67
- prise en charge de la recherche de texte
 - formulaire de configuration 10

R

- recherche de texte dans un référentiel
 - IBM Content Manager
 - Voir* indexeur de recherche de texte

S

- surveillance de l'indexation 183

T

- taille de la partition d'index
 - Autonomy K2 177
 - IBM Legacy Content Search Engine 177
 - Verity 177
- texte, recherche
 - FileNet P8 169, 187, 211, 212
 - installation 8, 9
 - moteur de recherche de contenu 169, 187
- traitement des documents
 - volumineux 177
- traitement des incidents
 - chargement impossible du pilote JDBC de type 2 159

- traitement des incidents liés au
 - aucune configuration de fonction définie par l'utilisateur AFUFetchFile trouvée 161
 - configuration Net Search Extender incorrecte 150
 - contenu supérieur à 1 Mo non indexé 160
 - contenu textuel non indexé 160
 - erreur lors de l'accès aux fichiers XML dans la fonction définie par l'utilisateur AFUFetchFile 162, 163
 - erreur SQL SQL0101N 158
 - erreur SQL SQL1224N 159
 - impossible d'activer l'indexation des types d'élément IBM CommonStore 151
 - impossible de charger la bibliothèque DB2 152
 - Impossible de localiser le répertoire d'installation DB2 153
 - impossible de traiter le type d'élément 151
 - index introuvable 160
 - index introuvable après sa recréation 154
 - la fonction définie par l'utilisateur AFUFetchFile n'est pas enregistrée 157
 - le processus de travail de l'indexeur s'arrête et indique une situation de mémoire saturée 164
 - Net Search Extender arrête CTE0005 154
 - Net Search Extender échoue lorsque le processus de l'indexeur est en cours d'exécution 155
 - Net Search Extender s'arrête anormalement 154
 - Net Search Extender s'arrête sans COMMIT 154
 - obtention du numéro de port DB2 158
 - taille de pile trop petite 158

V

- volume élevé de surveillance 184, 225



Numéro de programme : 5724-V57

SH11-1319-01

