IBM Content Collector
Version 4.0.0.2

Readme

IBM

IBM Content Collector Text Search Support for IBM Content Manager Version 4.0 Fix Pack 2 Readme

Readme documentation for IBM® Content Collector Text Search Support for IBM Content Manager Version 4.0 Fix Pack 2 including installation-related instructions, prerequisites and co-requisites, and information about fixes, changes, and limitations.

Readme file for:
IBM Content Collector Text Search Support for IBM Content Manager

Product/Component Release:
Version 4.0

Update Name:
Fix Pack 2

Fix ID:
4.0.0.2-IBM-ICC-CM8Text-FP002

Publication date:
15 May 2014

Last modified date:
15 May 2014

Online version of the readme file:
http://www.ibm.com/support/docview.wss?uid=swg27041780

Contents

- "Download location"
- "Prerequisites and co-requisites" on page 2
- "Installation information" on page 2
- "Fixes and changes made in version 4.0.0.2" on page 3
- "Documentation" on page 4
- "Notices" on page 5
- "Document change history" on page 7

Download location

Download IBM Content Collector Text Search Support for IBM Content Manager, Version 4.0 Fix Pack 2, from the following location:

http://www.ibm.com/support/docview.wss?uid=swg24037306

The following deliverables are made available for download.

Product/component name:	Platform	File name
IBM Content Collector Text Search Support V4.0 FP002 readme	n/a	`4.0.0.2-IBM-ICC-CM8Text-FP002.pdf`
IBM Content Collector Text Search Support V4.0 FP002 on AIX	AIX	`4.0.0.2-IBM-ICC-CM8Text-AIX-FP002.tar.gz`
IBM Content Collector Text Search Support V4.0 FP002 on Linux	Linux	`4.0.0.2-IBM-ICC-CM8Text-Linux32-FP002.tar.gz`

Product/component name:	Platform	File name
IBM Content Collector Text Search Support V4.0 FP002 on Linux for zSeries	Linux for zSeries	`4.0.0.2-IBM-ICC-CM8Text-Linuxz-FP002.tar.gz`
IBM Content Collector Text Search Support V4.0 FP002 on Solaris	Solaris	`4.0.0.2-IBM-ICC-CM8Text-Solaris-FP002.tar.gz`
IBM Content Collector Text Search Support V4.0 FP002 on Windows	Windows	`4.0.0.2-IBM-ICC-CM8Text-WIN-FP002.zip`

Prerequisites and co-requisites

IBM Content Collector Text Search Support for IBM Content Manager, Version 4.0 Fix Pack 2, can be used with either of the following products:

- IBM CommonStore for Lotus® Domino® V8.4
- IBM CommonStore for Exchange Server V8.4
- IBM Content Collector V2.1.1
- IBM Content Collector V2.2
- IBM Content Collector V3.0
- IBM Content Collector V4.0

The fix pack package does not require a specific version of IBM Content Collector Text Search Support as a prerequisite for the installation.

For information about the IBM Content Collector prerequisites, see the appropriate tab in the detailed system requirements document at http://www.ibm.com/support/docview.wss?uid=swg27039384.

If you work with applications that work with the contents of the TIEFLAG column in the item type component table in IBM Content Manager, such as IBM eDiscovery Manager, contact IBM Software Support for information about which versions of these products can evaluate this value.

Installation information

Apply this fix pack package to install IBM Content Collector Text Search Support for IBM Content Manager Version 4.0 Fix Pack 2.

Prior to installation

> When you upgrade from a previous version or install the IBM Content Collector Text Search Support package for the first time, you might not be able to set the installation directory in the installation wizard. Instead, the default directory for the IBM Content Collector Text Search Support package is used. This default directory is retrieved from the InstallAnywhere registry file when the installation is launched. However, if the installation of the IBM Content Collector Text Search Support package did not complete successfully during an earlier run, for example because it was cancelled towards the end of the process, the default value in the registry file is invalid. It cannot be used when the installation program is run again.

> As a workaround, delete the IBM Content Collector entries in the InstallAnywhere registry file `.com.zerog.registry.xml` manually before you start the installation wizard again. On a Windows system, the file is located at `C:\Program Files\Zero G Registry`. Note that this is a hidden directory. On UNIX-based systems, the file is in the `\var` directory.

Installing

> To upgrade to IBM Content Collector Text Search Support V4.0.0.2:
> - Install the IBM Content Collector Text Search Support V4.0.0.2 package.

Note: When updating the Text Search Support component on IBM Content Manager, check the installation path to make sure that you install it into the same base directory as the earlier version.

For details about the installation and configuration of IBM Content Collector Text Search Support, refer to the IBM Content Collector Indexing Guide.

Fixes and changes made in version 4.0.0.2

For a list of all APARs that were fixed after IBM Content Collector Text Search Support V4.0.0.2 and are included in IBM Content Collector Text Search Support V4.0.0.2, refer to the IBM Content Collector V4.0 Fix Pack 2 download page at http://www.ibm.com/support/docview.wss?uid=swg24037306.

For a list of new features that were added with IBM Content Collector Text Search Support V2.2.0.2 or later, see New indexing features in IBM Content Manager in the IBM Content Collector information center.

New reindexing command-line argument
The command-line argument `-reindexsearchfile <.ini file>` was added to the afuIndexer command-line tool to enable reindexing all documents that were indexed during a previous indexing run and where the search terms that are specified in the INI format file were indexed with the documents. When you run the reindexsearchfile operation, multiple search terms can be specified at once for identifying all documents that need to be reindexed.

The reindexsearchfile command-line argument reads all the search terms in the INI format file and searches for those terms in the full-text index. The INI format file must contain a [reindexsearch] section with the following type of entries:

```
label = search term
```

Label can be an arbitrary textual term and *search term* must be a Net Search Extender term. For example:

```
[reindexsearch]
APAR1234 = SECTIONS("attachment")("IBM")
APAR5678 = DUMMY
```

In this example, all those documents are reindexed in which the term IBM occurs as part of the textual data in the attachments sections as well as those documents in which the term DUMMY occurs anywhere in the documents (the search is not restricted to a particular section).

New text analysis processing for pure plain text attachments
To prevent attachment content from being inadvertently recognized as 7-bit ASCII code during document processing, the document processing algorithms were enhanced by heuristics that ensure that no binary content is indexed.

Sometimes, when attachments are processed by the Oracle Outside In Technology filters, the document attachments are inadvertently recognized as 7-bit ASCII code (plain text files) although they contained Base64 or UUencoded sequences or control characters ((code points < 0x1F) indicating binary content.

To prevent this from happening, the indexer processsing algorithms were enhanced by heuristics that check whether the content that is returned by the Oracle Outside In Technology filters is actually plain text, or whether the returned content contains patterns that indicate binary content, such as control characters (code points < 0x1F), embedded data in UUencode, or Base64 encoding. If embedded binary data is recognized, the attachment is not added to the generated XML input and indexed, but instead is assigned an IDXRC value of 5, indicating binary content, in the table of completed tasks.

New indexer configuration option `IdxIndexItemTypesBeyondWarning`
If the new indexer configuration option `IdxIndexItemTypesBeyondWarning` is defined, indexing to the item types that you specify will continue beyond the critical index size limit.

The maximum size that a Net Search Extender index can have is 200 GB. When this limit is approached during indexing, indexing stops and Net Search Extender displays the following warning stating that the index is approaching the critical size limit and that you should continue indexing to another item type:

```
afuIndexer: Error:
          An indexing error has occurred with the following reason:
          com.ibm.afu.ftsidx.indexer.exceptions.IndexerException: The
          indexing tool has detected that the size of the Net Search Extender
          index is approaching the critical size limit.
          For details, see:
          http://www-01.ibm.com/support/docview.wss?uid=swg21617071

          The current Net Search Extender index size is nnn GB.
          To prevent reaching the index size limit, you should stop
          archiving to this item type now and start archiving to a new item
          type.

          The moment the Net Search Extender index size exceeds the limit of
          200 GB, indexing to the item type is stopped and the remaining
          items in the table of open tasks are not processed.

          If you want to continue indexing to this item type however,
          although you are approaching the permitted index size limit, you
          must add the following configuration option to the indexer
          configuration of the item type:

            IdxIndexItemTypesBeyondWarning = <ItemtTypeName>
```

If you want to continue indexing to the item type although the warning message was displayed, define **IdxIndexItemTypesBeyondWarning** in the indexer configuration file that is used by the item type. Add the option to the file in the following way:

```
IdxIndexItemTypesBeyondWarning = ItemTypeName
```

ItemTypeName is the name of the item type. If you share one indexer configuration file between several item types, you can specify more than one item type separated by blanks. For example:

```
IdxIndexItemTypesBeyondWarning = ItemTypeName1 ItemTypeName2
```

After the configuration option was defined, you can continue indexing to the specified item types. When the ultimate index size limit is reached, indexing is stopped and no more items are indexed to the item type even if items still exist in the table of open tasks. The following message is displayed in this case:

```
afuIndexer: Error:
          An indexing error has occurred with the following reason:
          com.ibm.afu.ftsidx.indexer.exceptions.IndexerException: The
          indexing tool has detected that the size of the Net Search Extender
          index has reached a critical limit:
          For details, see:
          http://www-01.ibm.com/support/docview.wss?uid=swg21617071

          The current Net Search Extender index size is nnn GB.
          This is greater than the permitted limit of 200 GB for a Net Search
          Extender index.

          You cannot continue adding documents to this Net Search Extender
          index.
          Contact IBM Software Support.
```

Documentation

The following documentation is available:

Knowledge Center and updates to Knowledge Center

The IBM Content Collector product documentation including the documentation for IBM Content Collector Text Search Support for IBM Content Manager is available on the web at http://www.ibm.com/support/knowledgecenter/SSAE9L_4.0.0/contentcollector_4.0.0.htm.

For product information that is missing or is incorrect in IBM Content Collector Knowledge Center, see the technote about updates to Knowledge Center for IBM Content Collector V4.0 and the technote about corrections to Knowledge Center for IBM Content Collector V4.0.

PDF documentation

See the IBM Content Collector publication library for links to PDF documentation in English and further languages: http://www.ibm.com/support/docview.wss?uid=swg27037952

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan, Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Deutschland GmbH
Department M358
IBM-Allee 1
71139 Ehningen
Germany

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/us/en/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Microsoft, SharePoint, and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

The Oracle Outside In Technology included herein is subject to a restricted use license and can only be used in conjunction with this application.

Other product and service names might be trademarks of IBM or other companies.

Document change history

Date	Description of change
14 May 2014	initial version