Understanding Full Text Search in BMC Remedy AR System v7

Bob Poulos
Lead Product Developer
BMC Software
Agenda

› What is Full Text Search?
› Architecture
› Installation and Licensing
› Configuration
› Indexing
› Searching
› Server Groups
› Logging and Troubleshooting
› Tips and Tricks
› Demonstration
› Look to the Future
What is Full Text Search?

Full Text Search is an add-on product that allows for the indexing and searching of data not easily searchable by the database.

Full Text Search enables you to easily search through BMC® Remedy® Action Request System® (AR System®) to find existing information.

- Find existing solutions to related problems
- Search for specific instances of a phrase or key words

Full Text Search solves many problems that users encounter when performing database searches, including:

- Searching long text and attachment fields
- Improving search performance by searching large volumes of data with the aid of an index
- Performing case-insensitive searches
- Ranking search results according to their relevance to the search criteria
v6 Indexing and Searching Architecture

AR System Server

Full Text Indexing Process

arftp.lst

arftinp.lst

Full Text Indexes

Database
The search engine provided with AR System v7 is SearchServer version 5.4 from Hummingbird.

- SearchServer is the engine also used in the BMC® Remedy® Knowledge Management product.

- It is available on all platforms except SuSe Linux.

- The search engine works for all supported languages.

How does Full Text Search differ from the BMC Remedy Knowledge Management product?

- The Full Text Search capabilities of AR System are focused on locating information within AR System and its related items.

- BMC Remedy Knowledge Management is a complete solution that indexes multiple sources of information, including AR System.
Installation

› Hummingbird SearchServer is installed as a separate product using the vendor’s installation.
  – No search engine product files are included in the AR System installation.
  – A global key is provided to install the search engine; however the Full Text Search product must still be licensed to activate the functionality.

› BMC recommends that you install the search engine before installing or upgrading AR System.
  – Linkage of products is performed automatically by AR System if the search engine is installed.
  – The search engine can be installed at a later time, but manual configuration work is required.
Licensing is user based.
  – Fixed and floating full text user licenses

“New” v7 licenses must be used to activate Full Text Search.
  – BMC Remedy AR FTS Fixed
  – BMC Remedy AR FTS Floating

Existing licenses of the old types must be exchanged.
  – The Statement of Direction found on http://supportweb.remedy.com details the process for exchanging existing licenses for new ones.
All indexing is performed by the AR Server process.
- New Full Text Indexer queue type with the potential for multiple threads

Indexing is performed in the background.
- Tasks are recorded in a database table named ft_pending during form entry processing.
- Indexing will begin immediately if the Full Text Indexer threads are not busy.

A new feature in v7 is the ability to index attachment content.
- All common formats supported: MS Word, Excel, PDF, HTML, XML, plain text, etc.

Each field selected for indexing has its own index in the index collection.
- The index for a field consists of multiple files with different file extensions.
A form entry with a full text indexed field is created, modified, or deleted.

- Task record added to ft_pending table within the original database transaction.

The full text dispatcher thread awakes to read and queue the pending task after the original transaction completes.

A full text indexing thread processes the task and updates the index via the search engine interface.

- When the indexing completes, the pending record is removed from the database.
Attachment Indexing

› Contents of common file formats are now indexable.
  – PDF, MS Word, MS Excel, MS PowerPoint, HTML, XML, plain text, etc.
› Attachment fields are selected for indexing individually by the administrator, not by the entire attachment pool.
› File names are indexed along with the file contents for backward compatibility.
› When attempting to index file types not supported by the search engine, the only indication of the failure is written to the full text indexer log file if logging is turned on.
  – This is considered an expected failure.
  – This policy avoids cluttering the error log.
Searching

- Search qualifications that include full text indexed fields will employ the full text index if the user has a license.
  - If the user does not have a full text license, the search will be performed with whatever capability the database can provide.
- Full Text Search provides significant search performance improvement over database table scans.
- When using wildcards, there are three ways to include them:
  - End user direct input
  - Client Query By Example (QBE) settings and default
  - Server setting
A special display only field with reserved field ID 178 can be used. If placed on a form, the field can be used as a shortcut to generate a Query By Example “OR” search for all full text indexed fields on the form.

- Example: Entering Remedy in the field generates ‘Worklog’ LIKE “Remedy” OR ‘Description’ LIKE “Remedy”

Use caution when labeling the field so users do not think all fields on the form are searched.

The server will return an error if the field is used and the user does not have a full text license.
Field 178 should not be made available to pre-v7 clients.

- Older clients will not include this field in the qualification sent to the server.
  - This can produce unintentional unqualified queries.
- Hide the field for older clients with client-side workflow.
  - Execute on Window Open
  - Run If: $VERSION$ < “7”
    - Blank character before the 7 is intentional
  - Use a Change Field action to change the visibility of the field to ‘Hidden.’
Attachment Searching

- Attachment fields do not offer a standard QBE option.
- Qualifications can be built using the Advanced Search Bar.
  - The field name menu can be used to populate the search bar with the field name.
  - Clicking on the attachment field to populate the search bar with the field name is not a supported feature.
- Form search field can be used as a clever method to get a QBE search for an attachment field.
  - Only when there are no other indexed fields on the form
Server Group Considerations

- The definition of the full text indexing operation in a server group has changed from v6.
  - In v6, the server owning the operation performed all the indexing.
- In v7, each server performs the indexing tasks that are generated on it.
- Owning the operation in a server group now means which server will take over the pending tasks of a server determined to be delinquent.
  - Pending tasks are transferred by updating the pending records in the ft_pending table to transfer them from one server to another.
Logging and Troubleshooting

› **Indexing activity can be logged.**
  - Default file name is arftindx.log.
  - It is useful for tracking search engine interaction.
  - Indexing commands given to the search engine are logged.
  - Also logs full text related configuration information when the AR Server is started.

› **Searching activity can be logged.**
  - Shares the SQL log with database logging
  - Log entries prefixed with “FTS:”
  - Searching commands given to the search engine are logged.
Logging Example

Oftindx.log - Notepad

<FTI> <Queue: Admin> /* Tue Aug 01 2005 12:06:57.3200 */ FT Indexer Trace Log -- ON
<FTI> <Queue: Admin> /* Tue Aug 01 2005 12:07:12.2110 */ Configured disable full text indexer value
<FTI> <Queue: Admin> /* Tue Aug 01 2005 12:07:12.2110 */ Configured collection directory is C:\Program Files
<FTI> <Queue: Admin> /* Tue Aug 01 2005 12:07:12.2110 */ Configured configuration directory is C:\Program Files
<FTI> <Queue: Admin> /* Tue Aug 01 2005 12:07:12.2110 */ Configured temp directory is C:\Program Files
<FTI> <Queue: Admin> /* Tue Aug 01 2005 12:07:12.2110 */ Configured recovery interval is 60 minutes
<FTI> <Queue: Admin> /* Tue Aug 01 2005 12:07:12.2110 */ Configured optimization threshold is 1000
<FTI> <Queue: FTDIspatch> /* Tue Aug 01 2005 12:07:12.3810 */ Recognizing field Long description from file
<FTI> <Queue: FTDIspatch> /* Tue Aug 01 2005 12:07:14.6950 */ Staged 0 pending entries for indexing (limit = 0)
<FTI> <Queue: FTDIspatch> /* Tue Aug 01 2005 12:13:12.4490 */ Staged 1 pending entry for indexing (limit = 0)
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:13:12.4740 */ Processing entry started [seq number 1] for field
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:13:12.4740 */ Initializing full text session
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:13:12.6590 */ Set up indexing session options
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:13:12.6590 */ SET CHECK_TEXT_STATUS 'FALSE'
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:13:12.6690 */ SET CHARACTER_SET 'ISO_8859_1'
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:13:12.6690 */ SET TABLE_LANGUAGE 'ENGLISH'
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:13:14.3720 */ Processing entry completed (seq number 1) for field
<FTI> <Queue: FTDIspatch> /* Tue Aug 01 2005 12:15:06.0020 */ Staged 1 pending entry for indexing (limit = 0)
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:15:06.0020 */ Processing entry started [seq number 2] for field
<FTI> <Queue: FTIIndexer> /* Tue Aug 01 2005 12:15:06.0020 */ UPDATE T61C9 SET updateTime = 1154459705.4690
<FTI> <Queue: FTDIspatch> /* Tue Aug 01 2005 12:15:06.2730 */ Processing entry completed (seq number 2)

arsql.log - Notepad

12:11:19.8970 */ FTS: SELECT RELEVANCE('v2:4') AS REL, EID FROM T61C9 WHERE (((FT_TEXT CONTAINS
12:11:20.0670 */ FTS: Fetched 106 records
12:11:20.1080 */ BEGIN TRANSACTION
12:11:20.1080 */ DELETE FROM AROchar1688
Tips and Tricks

› To re-index a single field, simply deselect it for indexing, save the change, select it for indexing, and save the change.

› Before importing data into an empty form, deselect all indexed fields and then select them again after the importing completes.
  – Bulk indexing an entire field is more efficient than processing each entry as a separate indexing task.

› When adding the Full Text Search feature to an existing installation, review the list of fields already selected for Full Text Search by turning on Indexer logging, restarting the AR Server, and viewing the configuration information at the start of the log file.
  – There exists a high probability that “sleeper” fields will be found that are not desirable for indexing.
  – Deselect any fields listed that do not add full text searching value.
Summary of Benefits

› At the application level, Full Text Search enables you to find existing solutions to related problems.

› Search performance is improved for data not indexable by the database.
  – Directly help the user performing the search and indirectly help other users by avoiding database table scans that can lock up tables.

› Attachment contents are searchable.

› Full Text Search provides case-insensitive searching if the database does not.

› Form search field can be used to make qualification specification easier.
Look to the Future

AR System will continue to improve its ability to search for information contained within AR System itself.

› Google-like search
  – Search on all the fields in a form
  – Search on all the fields in a group of forms or all forms

› Categorization of results

› Highlighting of search terms in the result set

› Presenting context around the search term

› Using search technology to provide searching for AR System object definitions in the development environment
References

› AR System 7.0 Configuring guide
  – Chapter 8, Using full text search

› BMC Remedy Action Request System 7.0 Release Notes
  – Full Text Search section starting on page 54.

› Hummingbird SearchServer

› BMC Remedy Knowledge Management
  – www.bmc.com/products/products_services_detail/0,0,19052_19426_38115233,00.html