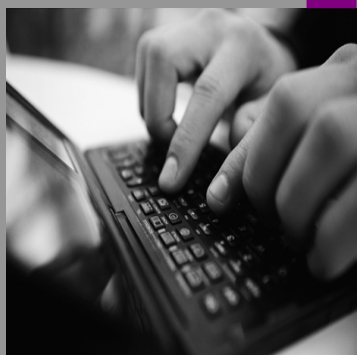

How-to Guide

SAP NetWeaver '04

How To... Configure TREX 6.1 for Efficient Indexing

Document Version 1.00 – January 2005

Applicable Releases:

SAP NetWeaver '04

Content

1 About This Document

This document contains recommendations for configuring Search and Classification (TREX) 6.1 for efficient indexing. The recommendations have the following aims:

- Fast initial indexing of large data sets
- Fast updating of indexes
- Fast index replication in distributed TREX systems

Target group of document

Technology consultants

Constraints

Indexing can take place with or without the TREX queue server. This document is targeted only at scenarios in which indexing takes place using the TREX queue server.

2 Indexing Configuration

Both the configuration of the application to which TREX is connected and the configuration of TREX itself are important to the performance of indexing. The parameters below are significant on TREX-side.

- Queue parameters

  The queue parameters control the interaction between the queue server and the index server. They determine when the queue server transmits documents to the index server and when the index server begins the indexing process. In a distributed TREX system the queue parameters also control when the updated indexes are replicated.
- Delta index

  The delta index is a separate index that TREX generates in addition to the main index. Activating the delta index at the right time speeds up the index update.

The best configuration for these parameters depends on the scenario in which you are using TREX. The individual scenarios differ in the following ways:

- The types of data to be indexed (whether the data is structured or unstructured)
- The initial amount of data to be indexed
- The amount of data to be indexed during productive operation
- How fast you want an updated index to be available for searching

The sections below describe which settings we recommend for individual scenarios. The appendix contains information on how you set up the configuration.

2.1 TREX as Part of Knowledge Management (KM)

If you are using TREX as part of Knowledge Management (KM), unstructured data in the form of text documents (Word documents, PDF documents, and so on) needs to be indexed. Usually there are very large document collections (more than 100,000 documents) to be indexed initially. The initial indexing run should be as efficient as possible.

After the initial indexing run there may be ongoing changes to the document collection: Documents may be added, changed, or deleted. The indexes should be updated as promptly as possible and replicated if necessary so that the changes are available for searching as quickly as possible.

2.1.1 Initial Indexing

Configure TREX as follows before beginning the initial indexing run:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**`Count`**	
Schedule Time	–	
Schedule Max Documents	**`10000`**	
Transmit Bulk Size	**`10000`**	
Max. Size of Transmit Bulk	**`1073741824`** (= 100 MB)	
Synchronize Bulk Size	**`10`**	
Initial Indexing Mode	**`On`**	
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	**`Off`**	

Delta index

Not activated. This is the standard configuration.

2.1.2 Ongoing Index Updates

Change the configuration of TREX as follows after the initial indexing run:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**`Time`**	
Schedule Time	**`All-01:00`**	Documents are transferred to the index server every hour.
Schedule Max Documents	–	
Transmit Bulk Size	**`10000`**	
Max. Size of Transmit Bulk	**`1073741824`** (= 100 MB)	
Synchronize Bulk Size	**`1`**	
Initial Indexing Mode	**`Off`**	

Parameters	Recommended Value	Comment
Merge Time (Delta Index)	**All-08:00**	The delta index is integrated into the main index every 8 hours.
Replication Time	–	
Replicate After Synchronize	Distributed TREX system: **On** Non-distributed TREX system: **Off**	

Delta index

Activated.

Parameter	Recommended Value	Comment
Threshold	**500 MB**	

2.2 TREX and Master Data Management/Search Engine Service (MDM/SES)

If you are using TREX with Master Data Management/Search Engine Service (MDM/SES), indexing and searching takes place using TREX business objects. Business objects are structured data such as product data or business partner data.

Note the following: If an index might possibly exceed 10 million business objects, we recommend implementing a distributed TREX system with one master and two slave hosts.

2.2.1 Initial Indexing

Configure TREX as follows before beginning the initial indexing run:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**Both**	
Schedule Time	**All-0:30**	Documents are transferred to the index server every 30 minutes.
Schedule Max Documents	**10000**	
Transmit Bulk Size	**100000**	
Max. Size of Transmit Bulk	**100000000**	
Synchronize Bulk Size	**1**	
Initial Indexing Mode	On	This value is set automatically by MDM. You do not need to set this value manually.
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	**Off**	

Delta index

Not activated. This is the standard configuration.

2.2.2 Ongoing Index Updates

Change the configuration of TREX as follows after the initial indexing run:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**`Both`**	
Schedule Time	**`All-0:01`**	Documents are transferred to the index server every minute.
Schedule Max Documents	**`10000`**	
Transmit Bulk Size	**`100000`**	
Max. Size of Transmit Bulk	**`100000000`**	
Synchronize Bulk Size	**`1`**	
Initial Indexing Mode	On	This value is set automatically by MDM. You do not need to set this value manually.
Merge Time (Delta Index)	**`All(0)`** or **`Sun(6)`**	The delta index is integrated every night at midnight or every Sunday at 6am. Choose a time when the system load is light.
Replication Time	–	
Replicate After Synchronize	Distributed TREX System: **`On`** Non-distributed TREX system: **`Off`**	

Delta index

Activated.

Parameter	Recommended Value	Comment
Threshold	**`1000000`** or **`1 GB`**	You can specify the number of documents in the index (**`1000000`**) or the size of the index in GB (**`1 GB`**).

2.3 TREX and SAP Business Information Warehouse (SAP BW)

If you are using TREX with SAP Business Information Warehouse (SAP BW), configure TREX as follows:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**Both**	
Schedule Time	**All-0:01**	Documents are transferred to the index server every minute.
Schedule Max Documents	**10000**	
Transmit Bulk Size	**6000**	
Max. Size of Transmit Bulk	**10000**	
Synchronize Bulk Size	**1**	
Initial Indexing Mode	**Off**	
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	**Off**	

Delta index

Not activated. This is the standard configuration.

2.4 TREX and SAP Knowledge Warehouse (SAP KW)

If you are using TREX with SAP Knowledge Warehouse (SAP KW), configure TREX as follows:

2.4.1 Initial Indexing

Configure TREX as follows before beginning the initial indexing run:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**Both**	
Schedule Time	**All-0:30**	Documents are transferred to the index server every 30 minutes.
Schedule Max Documents	**10000**	
Transmit Bulk Size	**6000**	
Max. Size of Transmit Bulk	**10000**	
Synchronize Bulk Size	**1**	
Initial Indexing Mode	**Off**	
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	**Off**	

Delta index

Not activated. This is the standard configuration.

2.4.2 Ongoing Index Updates

Change the configuration of TREX as follows after the initial indexing run:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**Both**	
Schedule Time	**All-0:01**	Documents are transferred to the index server every minute.
Schedule Max Documents	**10000**	
Transmit Bulk Size	**6000**	
Max. Size of Transmit Bulk	**10000**	
Synchronize Bulk Size	**1**	
Initial Indexing Mode	**Off**	
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	**Off**	

Delta index

Not activated. This is the standard configuration.

2.5 TREX and mySAP Human Resources e-Recruiting

If you are using TREX with mySAP Human Resources e-Recruiting, configure TREX as follows:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	`Both`	
Schedule Time	`All-0:01`	Documents are transferred to the index server every minute.
Schedule Max Documents	`10000`	
Transmit Bulk Size	`6000`	
Max. Size of Transmit Bulk	`10000`	
Synchronize Bulk Size	`1`	
Initial Indexing Mode	`Off`	
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	`Off`	

Delta index

Not activated. This is the standard configuration.

2.6 TREX and mySAP Product Lifecycle Management (mySAP PLM)

If you are using TREX with mySAP Product Lifecycle Management (mySAP PLM), configure TREX as follows:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	`Both`	
Schedule Time	`All-0:01`	Documents are transferred to the index server every minute.
Schedule Max Documents	`10000`	
Transmit Bulk Size	`6000`	
Max. Size of Transmit Bulk	`10000`	
Synchronize Bulk Size	`1`	
Initial Indexing Mode	`Off`	
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	`Off`	

Delta index

Not activated. This is the standard configuration.

2.7 TREX and mySAP Customer Relationship Management (mySAP CRM)

If you are using TREX with mySAP Customer Relationship Management (mySAP CRM), configure TREX as follows:

Queue parameters

Parameters	Recommended Value	Comment
Schedule Type	**Both**	
Schedule Time	**All-0:01**	Documents are transferred to the index server every minute.
Schedule Max Documents	**10000**	
Transmit Bulk Size	**6000**	
Max. Size of Transmit Bulk	**10000**	
Synchronize Bulk Size	**1**	
Initial Indexing Mode	**Off**	
Merge Time (Delta Index)	–	
Replication Time	–	
Replicate After Synchronize	**Off**	

Delta index

Not activated. This is the standard configuration.

3 Appendix

3.1 Configuring Queue Parameters

You can use the following tools to configure the queue parameters:

Tool	Path
TREX Admin Tool (Stand-Alone)	*Queue Admin* → *Queue Parameters*
TREX monitor in the portal	*System Administration* → *Monitoring* → *Knowledge Management* → *TREX Monitor* → *Edit Queue Parameters*
From Release 6.20: TREX Admin Tool in SAP R/3	*Transaction TREXADMIN* → *Queue Server Administration* → *Set Queue Parameters*

For more detailed information on the tools and the significance of the queue parameters, see **`help.sap.com/nw04`** → *SAP NetWeaver* → *Information Integration* → *Knowledge Management* → *Administration Guide* → *Technical Operations in Detail* → *TREX Admin Tools* in the SAP Library.

3.2 Activating the Delta Index

You can use the following tool to configure the delta index:

Tool	Path
TREX Admin Tool (Stand-Alone)	*Index Admin* → *Index Info* → *Index Parameters* → *Configure Delta Index*

For more detailed information on the delta index, see **`help.sap.com/nw04`** → *SAP NetWeaver* → *Information Integration* → *Knowledge Management* → *Administration Guide* → *Technical Operations in Detail* → *Delta Index Configuration* in the SAP Library.

http://www.sdn.sap.com/irj/sdn/howtoguides