Araya XLIFF und TMX Editor Technical and Revision Manual

Contents

1 Revision History

Product[1]	Rev. No.	Date	Change	Files Involved[2]
XL	12.12	04/03/07	Behaviour of "Ask before Adding Translation to TMX Database" changed. Behaviour depends now on status of source and target segment to be added.	lib/arayaserver.jar docs/liffeditorupdate.pdf docs\xliffug_de.pdf docs\xliffug_en.pdf `araya31032007.zip`
XL / AR	12.10	04/03/07	Added Change User/Password in Dialogue Remove/Create TMX Database Automatic Phrase translation added when a terminology database is chosen in XLIFF Editor Phrase translation activated for message TranslateDocument in XML-RPC Server Documentation update - how to move format information - XML RPC Server messages - Unicode BOM handling (file import)	lib/arayaserver.jar docs/liffeditorupdate.pdf docs\xliffug_de.pdf docs\xliffug_en.pdf \ini\config_simplexml.xml test/simplexml (directory) `araya10032007.zip`
XL	12.8	04/03/07	Removed small bug in terminology extraction Removed a small bug in XML file back conversion New navigation button for "Next segment not to translate" Adding "conversion tool" in Plugin supporting converting files from character set to another character set	lib/external.jar lib/arayaserver.jar docs/liffeditorupdate.pdf docs\xliffug_de.pdf docs\xliffug_en.pdf `araya04032007.zip`
XL	12.7	25/02/07	Latest XERCES Parser (Xerces-J-bin.2.9.0.zip); removes some xml parsing bugs contained in the old version of Xerces	lib/external.jar lib/arayaserver.jar docs/liffeditorupdate.pdf `araya25022007.zip`
XL	12.6	20/02/07	Entity Resolving Option and File added Alignment projects added	lib/arayaserver.jar docs\xliffug_de.pdf docs\xliffug_en.pdf `araya20022007.zip`
XL	12.15	14/04/2007	Bilingual Term extraction is now a stand alone application; new user interface – see bilingual extraction manuals; removed from plug-ins menu Removed bug in updating word index when translating	lib/arayaserver.jar docs\xliffug_de.pdf docs\xliffug_en.pdf docs\biextract_de.pdf docs\biextract_en.pdf BiEdit.exe
XL / AR	12.39	30/10/2004	Main changes	ib/arayaserver.jar external.jar docs\xliffug_de.pdf docs\xliffug_en.pdf docs\biextract_de.pdf docs\biextract_en.pdf

---

[1] XL=Araya XLIFF Editor, TX=Araya TMX Editor, AR = Araya Server
[2] File name in Courier specifies the download file name; path usually http://www.heartsome.de/downloads/<filename>; other file names specify the files referenced in the zip file; assumption is always that the basis directory is the Araya root directory, usually c:/Program File/araya.

2   System Requirements

2.1   Windows

To run Araya Java must be installed. Araya requires a Java version >= 1.5.

The installed version can be checked using the command `java –version` in a command shell. The result looks like that:

```
C:\araya>java -version
java version "1.6.0"
Java(TM) SE Runtime Environment (build 1.6.0-b105)
Java HotSpot(TM) Client VM (build 1.6.0-b105, mixed mode, sharing)
```

Java can be downloaded from Sun's website: www.sun.com

e.g. from

```
http://www.sun.com/download/index.jsp?cat=Java%20%26%20Technologies&tab=3&subcat=Java
```

3   Araya Directory Structure

Araya stores its files in various directories.

The start directory is Araya, e.g. c:/Program Files/Araya.

Subdirectory	Content
bin	Some binary files; mainly for the server version
database	Contains all the database files, esp. database.xml a file containing the created Araya databases
docs	Araya documentation
images	Main Araya images
icon	Some Araya images
ini	Contains all the initialisation files of Araya
lib	The Java libraries and some batch files
log	The log output files
perl-lib	Perl files
skl	The skeleton files fro converted documents
test	Araya test examples
tmp	Any temporary files

4 Configuring Application Class Paths

4.1 Araya XLIFF and TMX Editor

Araya XLIFF and TMX Editor load the required Java jar files at start-up time. The required jar files are defined in the file **classpath** in the Araya installation directory.

Standard Araya path: c:/Program Files/Araya

Full file path: c:/Program Files/Araya/classpath

It looks like that:

```
lib\Win32\swt.jar;lib\arayaserver.jar;lib\external.jar;lib\firebirdsql-
full.jar;lib\hsqldb.jar;lib\mysql-connector-java-5.0.4-
bin.jar;lib\msutil.jar;lib\mssqlserver.jar;lib\msbase.jar;lib\derby.jar;lib
\db2jcc.jar;lib\db2jcc_license_c.jar;lib\axion-1.0-
M2.jar;lib\h2.jar;lib\jtds-1.2.jar
-Xms64M
-Xmx1200M
```

The files in bold are required and should never be removed from the classpath. The other jar files describe the database driver jar files and can be removed or adapted if necessary. See also the chapter about "**2 Database drivers**".

4.2 XML-RPC Server

Running the Araya XML RPC server invokes several batch files located in the lib directory of Araya.

Batch file	Function
start_emx_service.bat	Starts XML RPC Server
stop_emx_service.bat	Stops XML RPC Server
emx_service_install.bat	Installs XML RPC Server
emx_service_uninstall.bat	Unstills XML RPC Server
rpcserver-install.bat	Main batch file to start or replace the XML RPC server. Under normal conditions it is only necessary to start this batch file. It will stop, uninstall, reinstall and start the XML RPC Server again.

In order to load the correct jar files the following lines need to be changed in the file **emx_service_install.bat**.

```
REM
REM Databasedrivers
REM
set
EAGLE_MEMEX_DATABASE=%EAGLE_MEMEX_BIN%\derby.jar;%EAGLE_MEMEX_BIN%\msbase.j
ar;%EAGLE_MEMEX_BIN%\mssqlserver.jar;%EAGLE_MEMEX_BIN%\msutil.jar;%EAGLE_ME
MEX_BIN%\mysql-connector-java-5.0.4-bin.jar;%EAGLE_MEMEX_BIN%\jtds-1.2.jar
```

Just add or remove the required database drivers here.

5 Starting the XMP RPC Server

The Araya XML RPC Server is started by running `rpcserver-install.bat`. This will install the service and start it. If a service is running it is stopped, uninstalled, reinstalled and started again.

The server is started in command shell. The successful output looks like that:

```
C:\araya\lib>REM Install the RPC Server
C:\araya\lib>cd c:\araya\lib
C:\araya\lib>call stop_emx_service.bat
C:\araya\lib>set EAGLE_MEMEX_INSTALLATION_NAME=EaglememexService
C:\araya\lib>net stop EaglememexService
EaglememexService wird beendet.
EaglememexService wurde erfolgreich beendet.
C:\araya\lib>call emx_service_uninstall.bat
C:\araya\lib>set EAGLE_MEMEX_INSTALLATION_NAME=EaglememexService
C:\araya\lib>net stop EaglememexService
C:\araya\lib>araya.exe -uninstall EaglememexService
The service was successfully uninstalled.
C:\araya\lib>call emx_service_install.bat
C:\araya\lib>REM
C:\araya\lib>REM ARAYA XML-RPC Installation Batch File
C:\araya\lib>REM
C:\araya\lib>cd c:\araya\lib
C:\araya\lib>REM
C:\araya\lib>REM SET the JAVA_HOME Variable here
C:\araya\lib>REM
C:\araya\lib>set JAVA_HOME=c:\programme\Java\j2sdk1.4.2_04\jre\bin\server
C:\araya\lib>REM set the JVM Variable etc. correctly by testing the Java installation
C:\araya\lib>call getjavahome.bat
C:\araya\lib>java -Djava.library.path=lib\Win32 -Xmx400M -cp "c:\araya\lib\arayaserver.jar"
com.araya.eaglememex.util.Tools c:/araya/lib/javahome.bat
Determining Java Environement
Java Version:  1.6.0
Java Vendor:   Sun Microsystems Inc.
OS Version:    Windows XP
JAVA Home:     "C:\Programme\Java\jre1.6.0"
JAVA JVM DLL:  "C:\Programme\Java\jre1.6.0\bin\client"
Creating javahomebatchfile: "c:/araya/lib/javahome.bat"
C:\araya\lib>REM this batch file is generated by getjavahome.bat == JAVA_JVM computed
C:\araya\lib>call javahome.bat
C:\araya\lib>SET JAVA_JVM=C:\Programme\Java\jre1.6.0\bin\client
C:\araya\lib>SET JAVA_HOME=C:\Programme\Java\jre1.6.0
C:\araya\lib>REM
C:\araya\lib>set EAGLE_MEMEX_BIN=c:/araya/lib
C:\araya\lib>set EAGLE_MEMEX_LOG=c:/araya/log
C:\araya\lib>REM
C:\araya\lib>REM The main ARAYA jar File
C:\araya\lib>REM
C:\araya\lib>set EAGLE_MEMEX_ARAYA=c:/araya/lib/arayaserver.jar;c:/araya/lib/external.jar
C:\araya\lib>REM
C:\araya\lib>REM External Jar files
C:\araya\lib>REN
C:\araya\lib>set EAGLE_MEMEX_EXTERNAL=c:/araya/lib\transform.jar;c:/araya/lib\xalan-
2.4.0.jar;c:/araya/lib\batik.jar;c:/araya/lib\avalon.jar;c:/araya/lib\xercesImpl.jar;c:/araya/
lib\jdom.jar;c:/araya/lib\xmlParserAPIs.jar;c:/araya/lib\xml-rpc-1.2.jar;c:/araya/lib\xml-
apis.jar;c:/araya/lib\win32\swt.jar
C:\araya\lib>REM
C:\araya\lib>REM Databasedrivers
C:\araya\lib>REM
C:\araya\lib>set
EAGLE_MEMEX_DATABASE=c:/araya/lib\derby.jar;c:/araya/lib\msbase.jar;c:/araya/lib\mssqlserver.j
ar;c:/araya/lib\msutil.jar;c:/araya/lib\mysql-connector-java-5.0.4-bin.jar;c:/araya/lib\jtds-
1.2.jar
C:\araya\lib>REM set EAGLE_MEMEX_DATABASE=
C:\araya\lib>REM
C:\araya\lib>REM SET all the necessary jars
```

```
C:\araya\lib>REM
C:\araya\lib>set
EAGLE_MEMEX_ALL=.;c:/araya/lib/arayaserver.jar;c:/araya/lib/external.jar;c:/araya/lib\transfor
m.jar;c:/araya/lib\xalan-
2.4.0.jar;c:/araya/lib\batik.jar;c:/araya/lib\avalon.jar;c:/araya/lib\xercesImpl.jar;c:/araya/
lib\jdom.jar;c:/araya/lib\xmlParserAPIs.jar;c:/araya/lib\xml-rpc-1.2.jar;c:/araya/lib\xml-
apis.jar;c:/araya/lib\win32\swt.jar;c:/araya/lib\derby.jar;c:/araya/lib\msbase.jar;c:/araya/li
b\mssqlserver.jar;c:/araya/lib\msutil.jar;c:/araya/lib\mysql-connector-java-5.0.4-
bin.jar;c:/araya/lib\jtds-1.2.jar
C:\araya\lib>REM
C:\araya\lib>REM Uninstall existing service
C:\araya\lib>REM
C:\araya\lib>call emx_service_uninstall.bat
C:\araya\lib>set EAGLE_MEMEX_INSTALLATION_NAME=EaglememexService
C:\araya\lib>net stop EaglememexService
C:\araya\lib>araya.exe -uninstall EaglememexService
Error while uninstalling the service: Der angegebene Dienst ist kein installierter Dienst.
C:\araya\lib>REM
C:\araya\lib>REM Install
C:\araya\lib>REM
C:\araya\lib>araya.exe -install EaglememexService
"C:\Programme\Java\jre1.6.0\bin\client\jvm.dll" -jvm_option -Xms32M -Xmx1000m -
Djava.class.path=".;c:/araya/lib/arayaserver.jar;c:/araya/lib/external.jar;c:/araya/lib\transf
orm.jar;c:/araya/lib\xalan-
2.4.0.jar;c:/araya/lib\batik.jar;c:/araya/lib\avalon.jar;c:/araya/lib\xercesImpl.jar;c:/araya/
lib\jdom.jar;c:/araya/lib\xmlParserAPIs.jar;c:/araya/lib\xml-rpc-1.2.jar;c:/araya/lib\xml-
apis.jar;c:/araya/lib\win32\swt.jar;c:/araya/lib\derby.jar;c:/araya/lib\msbase.jar;c:/araya/li
b\mssqlserver.jar;c:/araya/lib\msutil.jar;c:/araya/lib\mysql-connector-java-5.0.4-
bin.jar;c:/araya/lib\jtds-1.2.jar" -jvm_option -Dsystem.drive="c:" -start
com.araya.eaglememex.webserver.EagleMemexServer -out "c:/araya/log\service.out" -err
"c:/araya/log\service.err"
The service was successfully installed.
C:\araya\lib>REM
C:\araya\lib>REM Finished
C:\araya\lib>REM
C:\araya\lib>call start_emx_service.bat
C:\araya\lib>set EAGLE_MEMEX_INSTALLATION_NAME=EaglememexService
C:\araya\lib>net start EaglememexService
EaglememexService wird gestartet.
EaglememexService wurde erfolgreich gestartet.
C:\araya\lib>REM Finished Installing the RPC Server
```

The server produces two log files in the log directory.

Service.log contains standard messages, while **service.err** contains any error message.

Service.log looks like that:

```
ARAYA XML-RPC Server (C) Heartsome Europe GmbH (2007)
ARAYA XML-RPC Server Version: XLFEdit 12.5 - Build 11-02-2007
ARAYA XML-RPC Server License: Heartsome
Start Date: 19.2.2007 19:56:42:280
EagleMemexServer: Read 5 external jars: c:/araya/dbjars/mysql-connector-
java-2.0.14-bin.jar;mysql-connector-java-3.0.2-beta-
bin.jar;c:/araya/dbjars/msutil.jar;c:/araya/dbjars/mssqlserver.jar;c:/araya
/dbjars/msbase.jar;
Adding jar file 1 of 5:"file:/c:/araya/dbjars/mysql-connector-java-2.0.14-
bin.jar"
Adding jar file 2 of 5:"file:/C:/WINDOWS/system32/mysql-connector-java-
3.0.2-beta-bin.jar"
Adding jar file 3 of 5:"file:/c:/araya/dbjars/msutil.jar"
Adding jar file 4 of 5:"file:/c:/araya/dbjars/mssqlserver.jar"
Adding jar file 5 of 5:"file:/c:/araya/dbjars/msbase.jar"
Service Class 1 : $default
Service Class 2 : TranslationTools
Service Class 3 : Utilities
Service Class 4 : Registration
Service Class 6 : Stem
Service Class 7 : TermExtract
Service Class 8 : Terminology
Service Class 9 : PhraseTranslatorService
Service Class 10 : TermUtilities
Listening on port 4040
```

6 XML RPC Server Messages

6.1 Translation Messages

6.1.1 Message TranslateDocument

Parameter	Values	Comment
message	TranslateDocument	
phase	String	Translation phase any combination CONV = convert only SEG = segment TRANS = tm translated PT = phrase translate Example: CONVSEGTRANSPT = Convert + segment + tm translate + phrase translation or BACK for back converting the file
sourcefile	String	Full path of source file
inputtype	String	FILE ... indicates content is in the file name specified STRING ... indicates source file is the string which contains the contents ZIP ... string contents is zipped MIME ... string contents is MIME format
xlifffile	String	Full directory name of the xliff file (output file)
segfile	String	Full directory name of the segmented xliff file (output file)
transfile	String	Full directory name of the translated xliff file (output file)
backfile	String	Full directory name of the back converted file (output file)
sklfile	String	Full directory name of the skeleton file (input/output file)
logfile	String	Full path of log file name
sourcelang	String	Source language
targetlang	String	Source language
sourcecharset	String	Source encoding
targetcharset	String	Target encoding
matchquality	Integer String	match quality
matchmaximum	Integer String	maximum number of matches (-1 for all!)
useTranslateAttribute	yes/no	use translatiuon attribute
breakoncrlf	yes/no	use CR/LF for segment end
paraseg	Yes/no	use paragraph based segmentation
tmxfile	String	output tmx file of translations (not required)
donotresolveentities	yes/no	do ot resolve entities
donotresolveentitiesfile	String	entity resolution file
replclass	yes/no	use replacement classes
useextensions	yes/no	use extensions for file format detection
dbname	String	Database name
user	String	database user name
password	String	database password
dbtype	String	database type
dbhost	String	database host
dbport	Integer String	database port
termdb	String	Terminology database name

termuser	String	Terminology database user name
termpassword	String	Terminology database password
termdbtype	String	Terminology database type
termdbhost	String	Terminology database host
termdbport	Integer String	Terminology database port

Return Vector	Value
0	0 or error
1	TRANSSTRING (or error)
2	String of translated file
3	SKLSTRING
	String of skl file

6.1.2 Message ListTmxLanguages

Parameter	Values	Comment
Message	ListTmxLanguages	
inputfile	String	Input tmx file

Return Vector	Value
0	Success or error
1	String with language list or error message

6.1.3 Message FileConv

Parameter	Values	Comment
Message	FileConv	
inputfile	String	input file to be recoded to another character set
outputfile	String	Output file recoded
fromcharset	String	input fil character set
tocharset		output character set

Return Vector	Value
0	Success or error
1	Success or error message

6.1.4 Message Stemmer

Parameter	Values	Comment
Message	Stemmer	
language	String	input file to be recoded to another character set
word	String	Output file recoded

Return Vector	Value
0	Success or error
1	stemmed word or error message

6.1.5 Message TermTranslate

Parameter	Values	Comment
message	TermTranslate	
inputfile	String	Full path of input file
inputtype	String	FILE ... indicates content is in the file name specified STRING ... indicates source file is the string which contains the contents ZIP ... string contents is zipped MIME ... string contents is MIME format
outputfile	String	Full directory name of the phrase translated file (output file)
logfile	String	Full path of log file name
sourcelan	String	Source language
targetlan	String	Target language
dbname	String	Database name
user	String	database user name
password	String	database password
dbtype	String	database type
dbhost	String	database host
dbport	Integer String	database port

Return Vector	Value
0	0
1	PTTRANSSTRING (or error)
2	String of translated file
3	Input file name
4	Output file name
5	Source language
6	Target language
7	Database name

6.2 Terminology Extraction Messages:

6.2.1 Message TermExtract

Parameter	Values	Comment
message	TermExtract	
inputfile	String	Full path of input file
outputpath	String	Full directory name where output files will be written to Three files are written to the output directory: <inputfilename>-tbx.xmlTBX formatted <inputfilename>-xlf.xml XLIFF formatted <inputfilename>-tmx.xml TMX formatted sumfile"<summaryfile>" Full path name of summary file
termseparator	String	List of characters seperating terms; regular expression syntax! Default: [\s;,\.:\?!\(\)\[\]\\"\+\-\&\\#\>\<\/]
termpattern	String	regular expression for terms to be extracted Default "*"
exclusionlist	String	list of words to be excluded from extraction
inclusionlist	String	list of words to be included by extraction
language	String	Language of file
sorting	String	"alpha", "freq asc", "freq desc" Sorting of terms extracted a) alpha sorting b) frequency ascending c) frequency descending
minwords	String	Minimum number of words of a term extracted in characters
minfreq	String	Minimum frequency of a term extracted in characters
minlen	String	Minimum length of a term extracted in characters
maxwords	Integer String	Maximum number of words of a term extracted in characters
maxfreq	String number	Maximum frequency of a term extracted in characters
maxlen	String number	Maximum length of a term extracted in characters
ignorenumbers	yes/no	Include / exclude numbers from frequency counting
longestmatch	yes/no	Display only longest multi terms
logfile	String	Path for log file
sourcecharset	String	Document character set
breakoncrlf	yes/no/<number>	Number of linefeeds forming a paragraph
paraseg	yes/no	Paragraph based segmentation

Return Vector	Value
0	Success or error
1	Message ?? String

6.2.1.1 Output File Formats:

6.2.1.1.1 TBX formatted

```
<?xml version="1.0" encoding="UTF-8" ?>
<!-- !DOCTYPE martif PUBLIC "ISO 12200:1999A//DTD MARTIF core
(XLTcdV04)//EN" -->
```

```
<martif type="DXLT">
      <martifHeader>
            <fileDesc>
                  <sourceDesc>Terminology Extractor version 1.2.8 from
27.08.2003</sourceDesc>
            </fileDesc>
      </martifHeader>
      <text>
            <body>
                  <termEntry id="2569106212203433582">
                        <langSet xml:lang="de">
                              <tig>
                                    <term>4mIT</term>
                                    <termNote
type="ProcessStatus">unapproved</termNote>
                                    <termNote
type="adminstrativeStatus">TermExtract</termNote>
                                    <termNote type="wordnumber">1</termNote>
                                    <termNote
type="absolutefrequency">1</termNote>
                                    <termNote
type="overallwordsindocument">136</termNote>
                                    <termNote
type="relativefrequency">0,007</termNote>
                                    <termNote
type="relativefrequencyexclusion">0,007</termNote>
                              </tig>
                        </langSet>
                  </termEntry>
                  ...
                              </body>
      </text>
</martif>
```

6.2.1.1.2 *XLIFF formatted*

```
<?xml version="1.0" encoding="UTF-8" ?>
<xliff>
      <file source-language="de">
            <header>
                  <phase-group>
                        <phase phase-name="1" process-name="TermExtraction"
company-name="HSCrossLingual" tool="Terminology Extractor version 1.2.8
from 27.08.2003" date="2005-07-10 17:35:11" job-id="XXXXXX"></phase>
                  </phase-group>
            </header>
            <body>
                  <trans-unit id="1"><source
xml:lang="de">4mIT</source></trans-unit>
                  <trans-unit id="2"><source
xml:lang="de">4mIT_CMS</source></trans-unit>
                  <trans-unit id="3"><source
xml:lang="de">Adresse</source></trans-unit>
                  ...
                              </body>
      </file>
</xliff>
```

6.2.1.1.3 TMX formatted

```
<?xml version="1.0" encoding="UTF-8" ?>
<tmx version="1.3">
            <header
                        creationtool="Araya Term Extract Export Tool"
creationtoolversion="Version v1.0" o-tmf="Terminology Extractor version
1.2.8 from 27.08.2003" creationdate="2005-07-10 17:35:11"
creationid="XXXXXX" adminlang="EN-US" srclang="EN-US"></header>
            <body>
                  <tu creationid="1" changeid="XXXXXX"><tuv
xml:lang="de"><seg>4mIT</seg></tuv></tu>
                  <tu creationid="2" changeid="XXXXXX"><tuv
xml:lang="de"><seg>4mIT_CMS</seg></tuv></tu>
                  <tu creationid="3" changeid="XXXXXX"><tuv
xml:lang="de"><seg>Adresse</seg></tuv></tu>
                  <tu creationid="4" changeid="XXXXXX"><tuv
xml:lang="de"><seg>Als</seg></tuv></tu>
                  ...
            </body>
</tmx>
```

6.2.1.1.4 Summaryfile

Format of line:

```
frequency::sum words in document::sum words over all documents::frequency

per document::frequency over all documents::language::document name::term
```

```
1::136::136::0,007::0,007::de::docbook.xml.xlf::4mIT
1::136::136::0,007::0,007::de::docbook.xml.xlf::4mIT_CMS
1::136::136::0,007::0,007::de::docbook.xml.xlf::Adresse
1::136::136::0,007::0,007::de::docbook.xml.xlf::Als
```

6.2.2 Message SummaryExtract

Parameter	Values	Comment
message	SummaryExtract	
inputfile	String	Full path of input file
tbxfile	String	Output tbx file
tmxfile	String	Output tmx file
xlifffile	String	Output xliff file
summaryfile	String	Output summary file
logfile	String	Full path of log file name

Return Vector	Value
Element 0	Success or error
Element 1	String

7 Database Issues

7.1 Database Drivers

The following is a short description of some database driver jar file. Database druver jar files go into the araya/lib directory, configuration files into the araya/ini directory.

Database Type	Jar Files (go into lib directory)	Araya configuration file (from ini directory)

MS SQL Server	`msbase.jar`
	`mssqlserver.jar`
	`msutil.jar`
	`ojdbc14.jar`
	or alternatively
	`jtds-1.2.jar`
		`db-MsSQL2000.xml`
		`dbterm-MsSQL2000.xml`
		
		`db-JdtsMsSQL2000.xml`

Oracle		`db-Oracle.xml`

MySQL
Depending on version of MySQL
	`mysql-connector-java-2.0.14-bin.jar`
	`mysql-connector-java-3.0.2-beta-bin.jar`
	`mysql-connector-java-5.0.4-bin.jar`
		`db-MySQL.xml`
		`dbterm-MySQL.xml`

H2	`h2.jar`	`db-H2Embedded.xml`
Firebird	`firebirdsql-full.jar`	`db-Firebird.xml` `dbterm-Firebird.xml`

HSQL	`hsqldb.jar`	`db-HSQL.xml`
		`db-HSQLEmbedded.xml`
		`dbterm-HSQL.xml`

Cloudscape	`derby.jar`
	`db2jcc.jar`
	`db2jcc_license_c.jar`
	`derbynet.jar`
		`db-CloudscapeServer.xml`
		`db-Cloudscape.xml`
		`dbterm-CloudscapeServer.xml`
		`dbterm-Cloudscape.xm`

Dbterm* references the configuration files for terminology databases, other the tmx database configuration files.

7.2 Database Servers

7.2.1 HSQL Server

The following batch files are required:

Batch file	Function
hsql_service_install.bat	Install HSQL Server
hsql_service_uninstall.bat	Uninstalls HSQL Server
stop_hsql_service.bat	Stops HSQL Server
start_hsql_service.bat	Starts HSQL Server

For starting the HSQL Server run start_hsql_service.bat, for stopping stop_hsql_service.bat. Please note that it must first be installed

---

hsql_service_install.bat

```
net stop HSQLService
set JAVA_HOME=c:\programme\Java\j2sdk1.4.2_04\jre\bin\server
REM set the JVM Variable etc. correctly by testing the Java installation
call getjavahome.bat
REM this batch file is generated by getjavahome.bat == JAVA_JVM computed
call javahome.bat
REM

araya.exe -uninstall HSQLService
set EAGLE_MEMEX_BIN=c:\Program Files\araya\lib
set DATABASE=c:\Program Files\araya\database\hsql_araya
set HSQLLOG=c:\Program Files\araya\log
%EAGLE_MEMEX_BIN%\araya.exe -install HSQLService ""%JAVA_JVM%\jvm.dll"" -
jvm_option -Xms16m -Xmx128m -
Djava.class.path=".;%EAGLE_MEMEX_BIN%\hsqldb.jar" -jvm_option -
Dsystem.drive="c:" -start org.hsqldb.Server -params -database "%DATABASE%"
-out "%HSQLLOG%\servicehsql.out" -err "%HSQLLOG%\servicehsql.err"
net start HSQLService
```

hsql_service_install.bat

```
net stop HSQLService

araya.exe -uninstall HSQLService
```

start_hsql_service.bat

```
net start HSQLService
```

stop_hsql_service.bat

```
net stop HSQLService
```

Please note that depending on the Java and Araya settings installation the above path definitions must be adapted.

7.2.2 Cloudscape (Derby) as a server

To run Cloudscape as a server it is required to run it in server mode. After installation of Cloudscape run batch file startNetworkServer.bat in a command shell.

```
c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>startNetServer
Der Befehl "startNetServer" ist entweder falsch geschrieben oder
konnte nicht gefunden werden.

c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>startNetworkServer
.bat

c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>set CLOUDSCAPE_INS
TALL=C:\Programme\IBM\Cloudscape_10.0

c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>set CLOUDSCAPE_INS
TALL=C:\Programme\IBM\Cloudscape_10.0

c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>FOR %X in ("C:\Pro
gramme\IBM\Cloudscape_10.0") DO SET CLOUDSCAPE_INSTALL=%~sX

c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>SET CLOUDSCAPE_INS
TALL=C:\PROGRA~1\IBM\CLOUDS~1.0

c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>set CLASSPATH=C:\P
ROGRA~1\IBM\CLOUDS~1.0\lib\derby.jar;C:\PROGRA~1\IBM\CLOUDS~1.0\lib\derbytools.j
ar;C:\PROGRA~1\IBM\CLOUDS~1.0\lib\derbynet.jar;

c:\Programme\IBM\Cloudscape_10.0\frameworks\NetworkServer\bin>java org.apache.de
rby.drda.NetworkServerControl start
Server is ready to accept connections on port 1527.
Connection number: 1.
Connection number: 2.
Connection number: 3.
Connection number: 4.
Connection number: 5.
```

For Shutdown of cloudscape server **stopNetworkServer.bat** must be called.

8 Adding new XML DTDs and the XML Converter

Extending the recognised XML DTDs can be done in two ways:

    a)   manually

    b)   Using the XML converter configuration in Araya

The extension will be described based on the DOCBOOK DTD (docbookx.dtd).

8.1 Adding the DTD to the file catalogue.xml

In a first step all relevant DTDs should be copied into the Araya ini directory (or into a subdirectory there).Next the file **catalogue.xml** must be extended with the new DTD/Entity references.

Example DOCBOOK:

```
<!-- DOCBOOK -->
<dtd publicId="-//OASIS//DTD DocBook XML V4.1.2//EN"
systemId="docbookx.dtd">c:/Araya/ini/docbookx.dtd</dtd>
<dtd publicId="-//OASIS//DTD DocBook XML CALS Table Model V4.1.2//EN"
systemId="calstblx.dtd">c:/Araya/ini/calstblx.dtd</dtd>
<dtd publicId="-//OASIS//DTD XML Exchange Table Model 19990315//EN"
systemId="soextblx.dtd">c:/Araya/ini/soextblx.dtd</dtd>
<dtd publicId="-//OASIS//ELEMENTS DocBook XML Information Pool V4.1.2//EN"
systemId="dbpoolx.mod">c:/Araya/ini/dbpoolx.mod</dtd>
<dtd publicId="-//OASIS//ELEMENTS DocBook XML Document Hierarchy V4.1.2//EN"
systemId="dbhierx.mod">c:/Araya/ini/dbhierx.mod</dtd>
<dtd publicId="-//OASIS//ENTITIES DocBook XML Additional General Entities V4.1.2//EN"
systemId="dbgenent.mod">c:/Araya/ini/dbgenent.mod</dtd>
<dtd publicId="-//OASIS//ENTITIES DocBook XML Notations V4.1.2//EN"
systemId="dbnotnx.mod">c:/Araya/ini/dbnotnx.mod</dtd>
<dtd publicId="-//OASIS//ENTITIES DocBook XML Character Entities V4.1.2//EN"
systemId="dbcentx.mod">c:/Araya/ini/dbcentx.mod</dtd>
<dtd publicId="-//OASIS//ELEMENTS DocBook XML HTML Tables V4.3//EN"
systemId="htmltblx.mod">c:/Araya/ini/htmltblx.mod</dtd>

<!-- ................................................................. -->
<!-- ISO entity sets .................................................. -->

<dtd publicId="ISO 8879:1986//ENTITIES Diacritical Marks//EN//XML" systemId="ent/iso-
dia.ent">c:/Araya/ini/ent/iso-dia.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Numeric and Special Graphic//EN//XML"
systemId="ent/iso-num.ent">c:/Araya/ini/ent/iso-num.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Publishing//EN//XML" systemId="ent/iso-
pub.ent">c:/Araya/ini/ent/iso-pub.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES General Technical//EN//XML" systemId="ent/iso-
tech.ent">c:/Araya/ini/ent/iso-tech.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Added Latin 1//EN//XML" systemId="ent/iso-
lat1.ent">c:/Araya/ini/ent/iso-lat1.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Added Latin 2//EN//XML" systemId="ent/iso-
lat2.ent">c:/Araya/ini/ent/iso-lat2.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Greek Letters//EN//XML" systemId="ent/iso-
grk1.ent">c:/Araya/ini/ent/iso-grk1.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Monotoniko Greek//EN//XML" systemId="ent/iso-
grk2.ent">c:/Araya/ini/ent/iso-grk2.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Greek Symbols//EN//XML" systemId="ent/iso-
grk3.ent">c:/Araya/ini/ent/iso-grk3.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Alternative Greek Symbols//EN//XML" systemId="ent/iso-
grk4.ent">c:/Araya/ini/ent/iso-grk4.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Added Math Symbols: Arrow Relations//EN//XML"
systemId="ent/iso-amsa.ent">c:/Araya/ini/ent/iso-amsa.ent</dtd>
```

```
<dtd publicId="ISO 8879:1986//ENTITIES Added Math Symbols: Binary Operators//EN//XML"
systemId="ent/iso-amsb.ent">c:/Araya/ini/ent/iso-amsb.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Added Math Symbols: Delimiters//EN//XML"
systemId="ent/iso-amsc.ent">c:/Araya/ini/ent/iso-amsc.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Added Math Symbols: Negated Relations//EN//XML"
systemId="ent/iso-amsn.ent">c:/Araya/ini/ent/iso-amsn.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Added Math Symbols: Ordinary//EN//XML"
systemId="ent/iso-amso.ent">c:/Araya/ini/ent/iso-amso.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Added Math Symbols: Relations//EN//XML"
systemId="ent/iso-amsr.ent">c:/Araya/ini/ent/iso-amsr.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Box and Line Drawing//EN//XML" systemId="ent/iso-
box.ent">c:/Araya/ini/ent/iso-box.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Russian Cyrillic//EN//XML" systemId="ent/iso-
cyr1.ent">c:/Araya/ini/ent/iso-cyr1.ent</dtd>
<dtd publicId="ISO 8879:1986//ENTITIES Non-Russian Cyrillic//EN//XML" systemId="ent/iso-
cyr2.ent">c:/Araya/ini/ent/iso-cyr2.ent</dtd>
```

Entry structure:

`<dtd`	
`publicId="-//OASIS//DTD DocBook XML V4.1.2//EN"` `systemId="docbookx.dtd"`	Identifier Identifikation DTD
`>`	
`c:/Araya/ini/docbookx.dtd`	Verweis auf den Ort der DTD
`</dtd>`	

8.2 Identifying the XML Converter Configuration File

As an example the following Docbook XML file is used:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE chapter PUBLIC "-//OASIS//DTD DocBook XML V4.1.2//EN"
"docbookx.dtd">
<chapter id="d10178" lang="de"><title>Installation von
<productname>Araya</productname>
</title><indexterm><primary>Installation</primary></indexterm><section><tit
le>Serverinstallation</title><para>Die Installation erfolgte auf dem Server
<guilabel>Araya 1</guilabel>.</para></section></chapter>
```

Araya uses several strategies to identify the configuration files. The following excerpt from the

catalogue.xml is used as an example:

`<dtd`	
`publicId="-//OASIS//DTD DocBook XML V4.1.2//EN"` `systemId="docbookx.dtd"`	Identifier Identifikation DTD
`>`	
`c:/Araya/ini/docbookx.dtd`	Verweis auf den Ort der ini Datei
`</dtd>`	

a) Araya tries based on the PUBLIC ID to find the corresponding configuration file. Then the name of the matching element ist used. The path is removed (in the example the string docbookx.dtd remains), the ini path, the indicator „config_", the remainder and the extension „.xml" added. In the above case the configuration file is therefore called: c:/Araya/ini/config_docbookx.dtd.xml. If the file does not exist **procedure step b** is executed.

b) Araya tries based on the SYSTEM ID to find the corresponding configuration file. Then the name of the SYSTEM ID is used. The Araya ini path is used as a starter, „config_" added, the SYSTEM ID (Example: docbookx.dtd) and the extension „.xml" attached. In the above case the configuration file gets the name c:/Araya/ini/config_docbookx.dtd.xml. If the file does not exist **procedure step c** is executed.

c) Araya recognises the configuration file from the symbol after DOCTPYE, here therefore chapter. The name of the configuration file therefore is config_chapter.xml (the full pathname: c:/Araya/ini/config_chapter.xml) and must exist in the Araya ini directory.

8.3 Extending the definitions manually

The elements which should be used for segmentation are defined in the file with the name:

„config_elementname.xml"..

The following example demonstartes this for the element „chapter" of the DOCBOOK DTD. As an example of a Docbook XML file the following is used:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE chapter PUBLIC "-//OASIS//DTD DocBook XML V4.1.2//EN"
"docbookx.dtd">
<chapter id="d10178" lang="de"><title>Installation von
<productname>Araya</productname>
</title><indexterm><primary>Installation</primary></indexterm><section><tit
le>Serverinstallation</title><para>Die Installation erfolgte auf dem Server
<guilabel>Araya 1</guilabel>.</para></section></chapter>
```

Example of a configuration file:

```
<?xml version="1.0"?>
<!DOCTYPE ini-file PUBLIC "-//OASIS//DTD DocBook XML V4.1.2//EN"
"configuration.dtd">
<ini-file>
      <tag hard-break="yes">title</tag>
      <tag hard-break="yes">chapter</tag>
      <tag hard-break="yes">primary</tag>
      <tag hard-break="yes">para</tag>
 </ini-file>
```

This defines that the elements title, chapter, primary und para should be used for segmenting.

Further options as follows:

Assigning an abstract xliff type for a given element:

```
<tag ctype="bold" hard-break="no">hervorhebung</tag>
```

Element **hervorhebung** will be assigned the type **bold** in XLIFF.

```
<tag attributes="ueberbegriff;unterbegriff" hard-
break="yes">stichwort</tag>
```

This defines that the attributes **ueberbegriff** and **unterbegriff** of element **stichwort** should be made available as a segment. The attribute value will appear as a Segment and can be translated.

8.4 Changing Converter Settings in Araya XLIFF Editor

Menu Options and XML Converter Configuration supports making adaptations
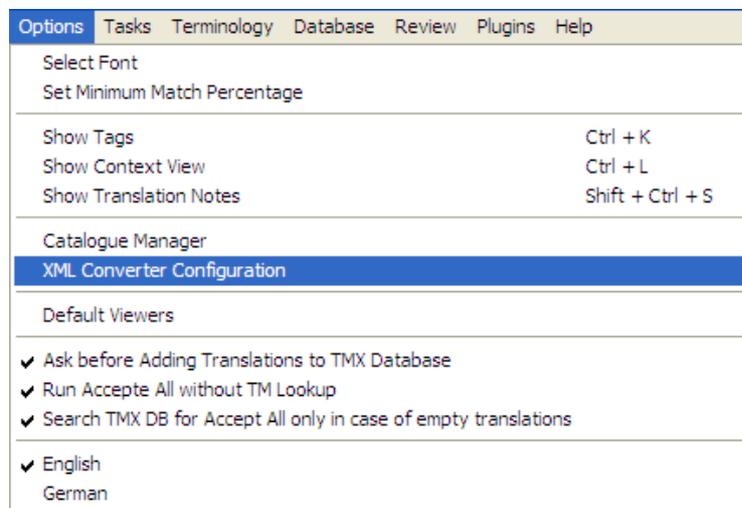
A list of all known configuration files is displayed.
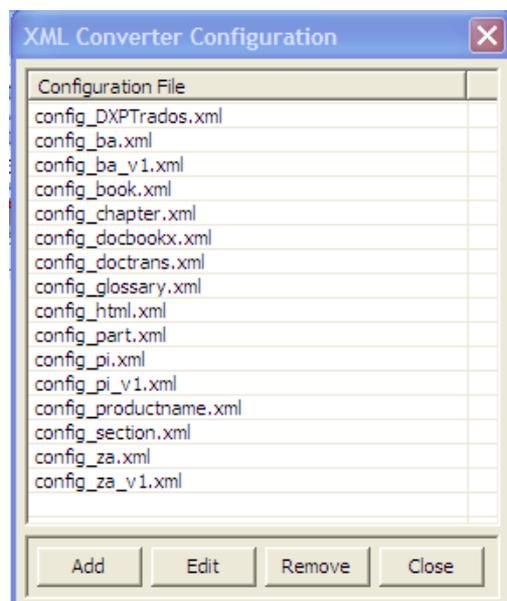
Choosing a configuration file displays all definitions fort his file:
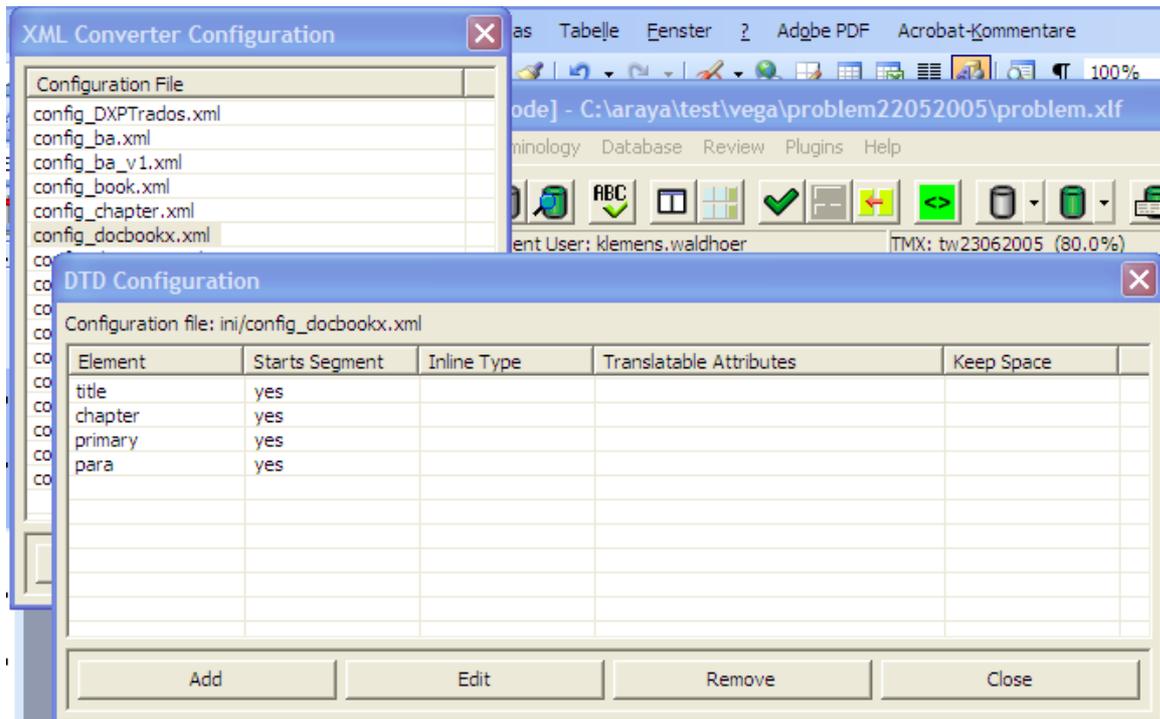
By clicking on an entry it can be changed or a new one can be added.
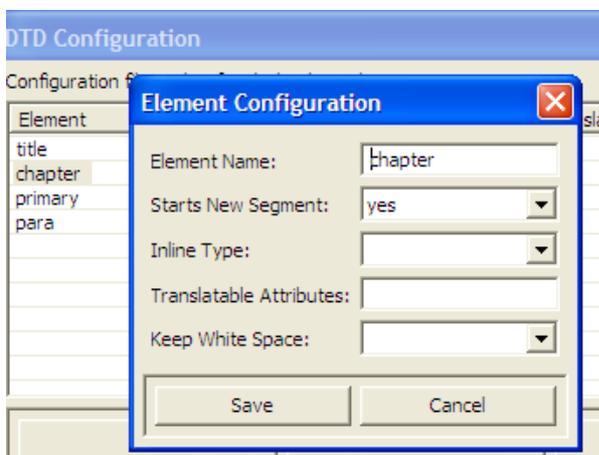
Mit Add kann ein neuer hinzugefügt werden.
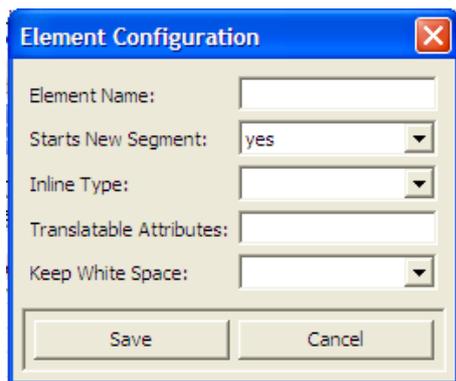
A new configuration file for a DTD can be created by using the Stepps below. Choose a DTD:.
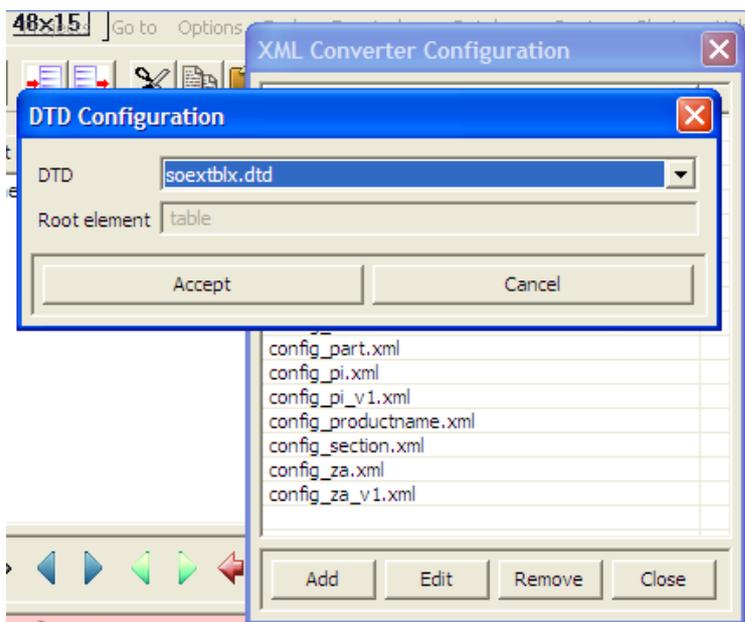
Adding element tgroup as Segment.
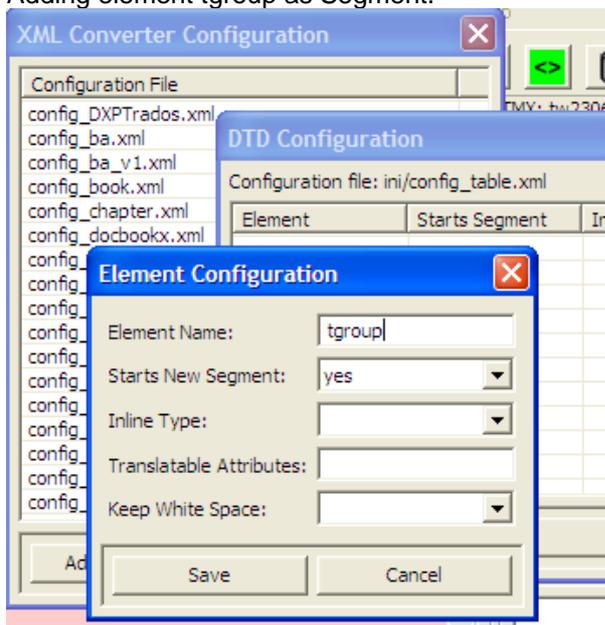
As a result:
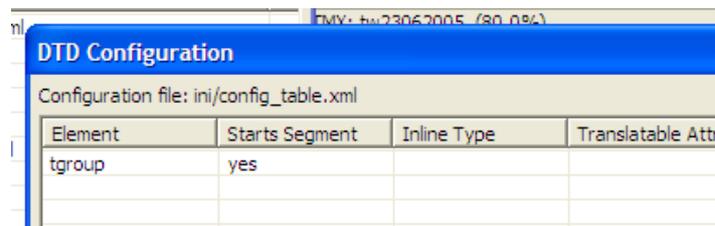
After closing the dialogue one can find the follwoing file in the Araya ini directory: **config_table.xml**

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE ini-file PUBLIC "-//HEARTSOME//Converters 2.0.0//EN"
"configuration.dtd" >
<ini-file><tag hard-break="yes">tgroup</tag>
</ini-file>
```

9   Word based indexing of a TM database

The entries of a TMX database can be automatically word indexed. Word indexing is defined as follows:

a) Stop words are not indexed (except otherwise specified!). The file which contains the stop words is defined in the eaglememex.properties file. Stop words can be defined per language.

b) In addition to the word (full form) the stem is stored too.

c) A segment (sentence) is broken into words- The following characters are used as word separators:

`\s;,.:?!()[]\"+-&#><  (\s` identifies a blank)

d) Numbers are not indexed.

9.1 Stop Word Definitions

Stop words can be defined in a special file. The file name is given in the **eaglememex.properties** file (lib directory):

```
eaglememex.stopwordfile=c:\\Program Files\\Araya\\ini\\stopword.txt
```

The format of the file is like that:

```
de=ein;eine;eines;der;die;das;ich;du;er;sie;es;wir;ihr;sie;dem;des;den;dess
en;deren;ihrer
de=auf;in;im;durch;mit;aus;zu
de=bin;bist;sind;ist
en=a;the;an;I;you;he;she;it;we;they;them;mine;your;her;our;this;that;these;
those
en=on;in;at;with;from;to
en=am;are;is
```

A line of the file consists of a language code (e.g de), followed by the = character, followed by the stop word separted by „;". A language code may appear several times in the file.

9.2 Special Character Definitions

Special characters which should be indexed must be defined in a file. The file name is defined in the **eaglememex.properties** file (lib directory):

```
eaglememex.specialchars.file=c:\\Program
```

```
Files\\Araya\\ini\\specialchars.txt
```

The format of the file is as follows:

```
&#9324
&#9325
&#9326
&#9327
&#9328
&#9329
&#9330
&#9331
...
..
```

Strings starting with **&#** will be converted to their respective character (Unicode). This allows defining Unicode characters with their **decimal value**. All other strings are used as they are. If the string appears in a segment it is considered as a word and word indexed.

10 Unification – Merging TMX Entries and Avoiding Duplicates

A key issue in translation is avoiding duplicate entries in TM databases. Araya uses an **UNIFICATION** approach in order to keep the database as small as possible. Unification here means that a new database (translation) entry is only generated if a new language pair (e.g. DE – EN) cannot be assigned to an existing entry. In the following this is demonstrated in a simple example. Understanding unification is key issue in Araya – otherwise one might be surprised how an TMX entry in the db looks like.

10.1 Unification explained

In the following we assume DE - EN as source and target language.

Segments get identified as follows:

S-<language code>-<n>: <string> - n is the sequence number.

In the examples below segments which result in unifications/merges are coloured.

```
Time 1:

S-DE-1: Haus
S-EN-1: house

After accepted the database will contain these two entries as a translation of the segments. In TMX
term this is now one TU
TU 1:
      TUV: S-DE-1: Haus
      TUV: S-EN-1: house
```

Next another language pair is added.

```
Time 2:

S-DE-2: Heim
S-EN-2: home

After accepting this translation pair the database will contain now two TU entries.
TU 1:
      TUV: S-DE-1: Haus
      TUV: S-EN-1: house
TU 2:
      TUV: S-DE-2: Heim
      TUV: S-EN-2: home
```

Now another pair is added:

```
Time 3:

S-DE-3: Haus
S-EN-3: building
```

After accepting this translation pair the database will contain now two TU entries – but TU 1 is extended with a new translation S-EN-3 – as S-DE-3 is identical to S-DE-1.

```
TU 1:
      TUV: S-DE-1: Haus
      TUV: S-EN-1: house
      TUV: S-EN-3: building
TU 2:
      TUV: S-DE-2: Heim
      TUV: S-EN-2: home
```

Now another translation pair comes in:

```
Time 4:

S-DE-4: Heim
S-EN-4: house
```

Now S-DE-4 is contained as S-DE-2 in TU 2 while S-EN-4 is contained in TU 1. As there is now obvious entry which should be preferred where the translation pair should be added both TU-1 and TU-2 are unified – meaning both entries are merged into one.

The result of this is with TU 2 being removed:
```
TU 1:
      TUV: S-DE-1: Haus
      TUV: S-EN-1: house
      TUV: S-EN-3: building
      TUV: S-DE-2: Heim
      TUV: S-EN-2: home
```

This actually means now that **Haus** can be translated into English as: **house – building – home** (and vice versa), **Heim** also as **house – building – home**, **house** can be translated into German as **Haus – Heim**. And so on.

Although this sounds quite simple for two languages it immediately gets complicated if several languages are involved. Here a language could operate as "pivot language" meaning that – although not really intended – a whole set of entries get merged although before two distinct entries. This can be esp. confusing if several translators translate. The DE-EN translator may be surprised by a unified entry as he never was the source of the merger and never produced double translation. This can be seen by the following example:

```
Time x:

Initial entries

TU 1:
      TUV: S-DE-1: Haus
      TUV: S-EN-1: house
      TUV: S-EN-3: building
      TUV: S-DE-2: Heim
      TUV: S-EN-2: home
      TUV: S-LA-6: domus

TU 2:
      TUV: S-DE-1: Wald
      TUV: S-EN-1: wood
      TUV: S-LA-6: silva
```

Assume now the EN-LA translator makes an error in his translation and adds (but the argument holds for other combinations too!) the following combination:

```
S-EN-11: wood
S-LA-11: domus
```

This results immediately in just one TU, TU 2 being removed.

```
TU 1:
      TUV: S-DE-1: Haus
      TUV: S-EN-1: house
      TUV: S-EN-3: building
      TUV: S-DE-2: Heim
      TUV: S-EN-2: home
      TUV: S-LA-6: domus
      TUV: S-DE-1: Wald
      TUV: S-EN-1: wood
      TUV: S-LA-6: silva
```

The DE – EN translator will be confused the next time he searches for "**Haus**" as he will get now the following EN proposals: **house – building –home – wood**. And the reason was the entries done by the EN-LA translator. One has to add that in some cases the corresponding EN-LA translation pair may be perfectly correct but for DE – EN it may be totally wrong and confusing.

As a consequence translators should be carefully with their translation in order to avoid unexpected translation links.

11 Importing Files into Araya

Araya supports importing and converting various types of input files. While under normal conditions this should not be a problem some issues have to be taken into account when importing / converting Unicode files. Those items are also related to Java, esp. with the way how Sun realised reading UTF-8 files in Java.

11.1 BOM – Byte Order Marks

In order to enable applications to read Unicode files special characters have been to defined the type of Unicode.

Byte **O**rder **M**arks are special characters at the beginning of a Unicode file to indicate whether it is big or little endian, in other words does the high or low order byte come first. These codes also tell whether the encoding is 8, 16 or 32 bit. You can recognise Unicode files by their starting byte order marks, and by the way Unicode-16 files are half zeroes and Unicode-32 files are three-quarters zeros.

UTF BOM (Byte Order Mark) Unicode-encoding Endian Indicators

0xfeff Byte-Order Description

ef bb bf	UTF-8 endian, strictly speaking does not apply, though it uses big-endian most-significant-bytes first representation.
fe ff	UTF-16 for 16-bit internal UCS-2, big endian, Java network order
ff fe	UTF-16 for 16-bit internal UCS-2, little endian, Intel/Microsoft order
00 00 fe ff	UTF-32 for 32-bit internal UCS-4, big-endian, Java network order
ff fe 00 00	UTF-32 for 32-bit internal UCS-4, little endian, Intel/Microsoft order.

The actual Unicode character encoded in all cases is 0xfeff. There are also variants of these encodings that have an implied endian marker.
(from **http://mindprod.com/jgloss/bom.html**)

11.2 Araya File Character Encoding Conversion Tool

As Java esp. does not support reading those BOMs correctly esp. for UTF-8 a conversion tool is supplied in Araya (Plugins menu). Basically this tool converts a given source file in a source character set into another file with a defined target character set.

It has to be noted that this could mean loosing some characters which are not supported in the target character set!
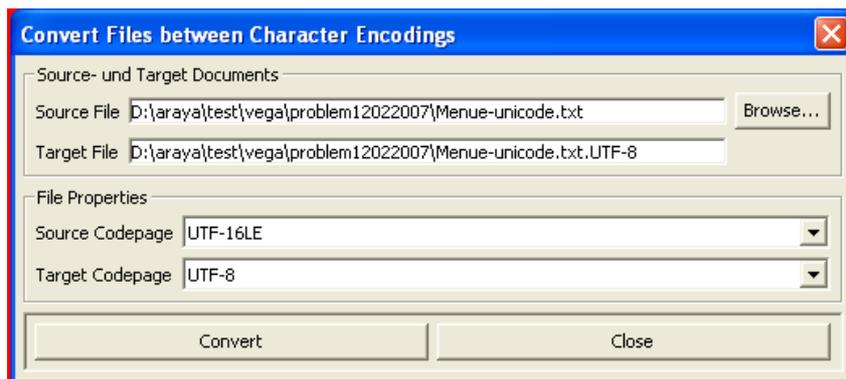
In the case of Unicode encoded files two files are generated.

a) one file as specified in for target file (file name is automatically generated); this file contains the BOM markers

b) another file – target file name plus extension **.nobom** – where the BOMS have been removed from the file.

Example:

Source File as in above:

Menue-unicode.txt	source file	UTF16LE encoded
Menue-unicode.txt.UTF-8	target file	UTF-8 encoded with BOM
Menue-unicode.txt.UTF-8.nobom	target file – no BOM	UTF-8 endoced without BOM

UTF16LE encoded File

```
ÿþd e        e n   f r   e s   r u   i t   z h   n l


 ( > 1 m / m i n )       ( > 1 m / m i n )       ( > 1 m / m i n )        ( >
1 m / m i n )     ( > 1 m / m i n )       ( > 1 m / m i n )       ( > 1 m /
m i n )      ( > 1 m / m i n )
```

UTF-8 encoded File

```
ï»¿de en    fr    es    ru    it    zh    nl
(>1m/min)  (>1m/min)  (>1m/min)  (>1m/min)  (>1m/min)  (>1m/min)
     (>1m/min)  (>1m/min)
```

UTF-8 encoded without BOM

```
de    en    fr    es    ru    it    zh    nl
(>1m/min)  (>1m/min)  (>1m/min)  (>1m/min)  (>1m/min)  (>1m/min)
     (>1m/min)  (>1m/min)
```

11.3 Example Usage

The above can easily be explained when importing a terminology file. Assume you have to import a terminology file as above (UTF16LE encoded File). The file comes from an Excel table and is stored there as:
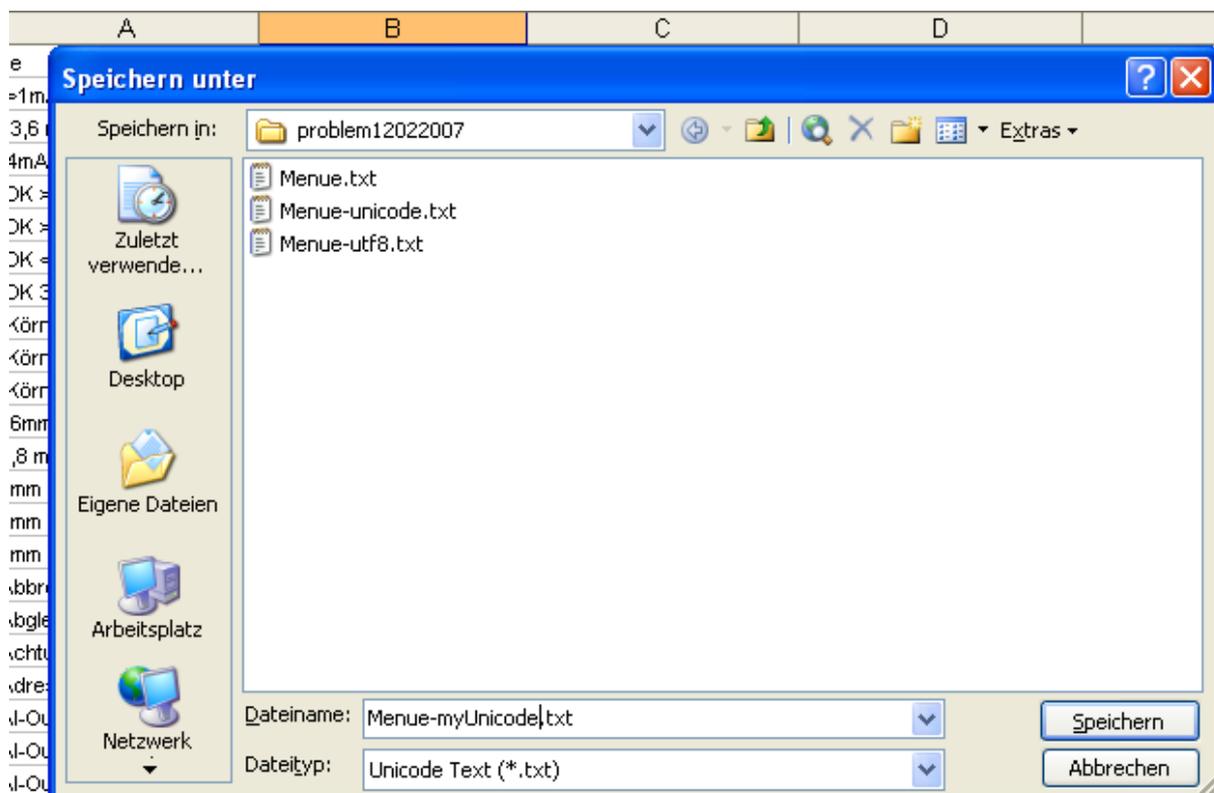
Importing this file will cause problems because of the BOM. Thus first a conversion to UTF-(should be executed.
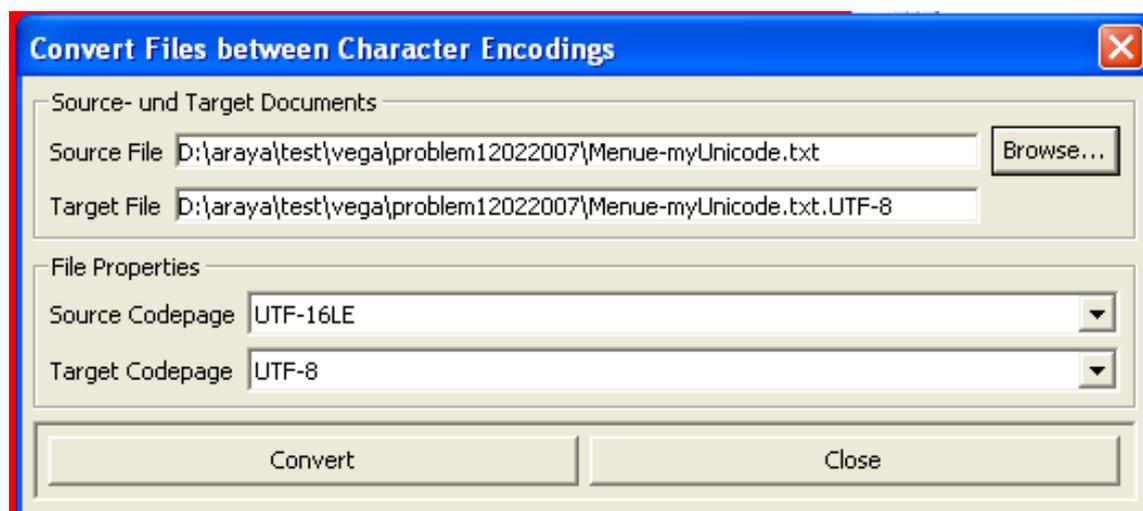
Now we get two files:

Menue-myUnicode.txt.UTF-8 and **Menue-myUnicode.txt.UTF-8.nobom**

In order to import the file correctly we use now Import Terminology File:
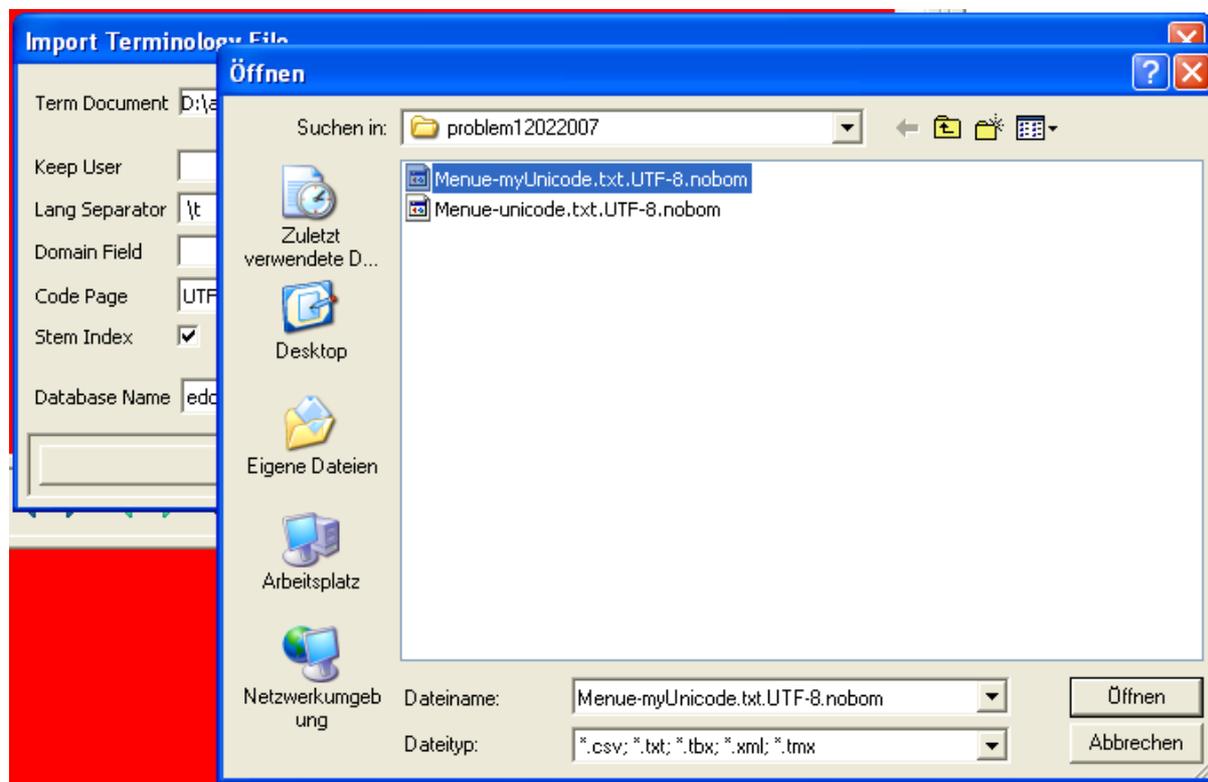
The file gets now correctly imported.

11.4 Recognising incorrect imports because of encoding problems

It is important further to note that the error only applies to the first line read the first 2-4 bytes are the BOM encoding. Thus one can easily recognise this type of error if e.g. language codes do not look correctly. A typical example of incorrect reading can be seen here: The "**?**" in "**?de**" indicates that during reading the first line the BOM was read as part of the language code too. The wrong language code will in most cases appear as the last column of the table.

12 Configuration Files

12.1 Eaglememex.properties File

The **eaglememex.properties** file controls many options of the Araya software system. It is located in

the Araya/lib directory. During the installation the file is rewritten, esp. this means rewriting the

abstract **ARRAYPATHDIRECTORY** pathname to the installation path name. It uses the current Araya

path (taken from the Araya start directory) and replaces all the occurrences of the

ARRAYPATHDIRECTORY with the real path. The original file is backed up.

```
#EagleMemex properties
#Tue Sep 09 17:34:18 ICT 2002

# standard xml rpc error file
eaglememex.rpc.err.file=ARRAYPATHDIRECTORY/log/service.err
# standard xml rpc log file
eaglememex.rpc.log.file=ARRAYPATHDIRECTORY/log/service.out


# path to the ini directory of Araya
eaglememex.ini.dir=ARRAYPATHDIRECTORY/ini/
# path to the temp directory of Araya
eaglememex.tmp.dir=ARRAYPATHDIRECTORY/tmp/
# standard error file
eaglememex.err.file=error.xml
# standard init file for segmentation – obsolete
eaglememex.ini.file=init.xml
# stop word file for Araya – full path
eaglememex.stopwordfile=ARRAYPATHDIRECTORY/ini/stopword.txt
# class path file for Araya for loeading additional database jar files)
# eaglememex.additionaljars=ARRAYPATHDIRECTORY/dbjars/classpath
# path to the log files directory of Araya
eaglememex.log.dir=ARRAYPATHDIRECTORY/log
# path to the test directory of Araya
eaglememex.test.dir=ARRAYPATHDIRECTORY/test
# path to the empty xliff file of Araya (for startup)
eaglememex.test.empty=ARRAYPATHDIRECTORY/test/empty.xlf
# path to the default loag file of Araya
# XLIFF edit creates log based on username and date
eaglememex.log.file=ARRAYPATHDIRECTORY/log/araya.log
# Araya loglevel – influences the amount of log information written
```

```
eaglememex.log.loglevel=0
# Araya adminstrators – overwrites any user restrictions (e.g. permission
to delete databases etc.)
eaglememex.admin.names=
# display next fuzzy special navigation button – either user names or all
for ALL users
# eaglememex.next.fuzzyspecial=
# display the fuzzyspcial button for the specified users
eaglememex.next.fuzzyspecial=all
# path to file which contains the special character encodings
# charachters / strings defined there get coloured (blue)
eaglememex.specialchars.file=ARRAYPATHDIRECTORY/ini/specialchars.txt
# Araya Hompepage URL – general
eaglememex.homepage=http://www.heartsome.de
# Araya Hompepage URL – German
eaglememex.homepage.de=http://www.heartsome.de
# Araya Hompepage URL – English
eaglememex.homepage.en=http://www.heartsome.de
# Araya user documentation – English
eaglememex.userguide.en=ARRAYPATHDIRECTORY/docs/xliffug_en.pdf
# Araya user documentation – German
eaglememex.userguide.de=ARRAYPATHDIRECTORY/docs/xliffug_de.pdf
# display information on tmx entry in match window
eaglememex.Show.TMXDBInfoInTarget=true
# standard language (= replacement language) to be used if no translation
if found for a given language
donotuse.eaglememex.No.TranslateLanguage=en
# issue warning if if replacement language is used
eaglememex.No.TranslateLanguage.Warning=true
eaglememex.No.TranslateLanguage.UseSource=true
# adds empty targe during translation (if does not exist)
eaglememex.addemptytarget=true
# do not update window contents during accept all if true
eaglememex.DoNotUpdateDisplaySegment=true
# sort order for language display in direct TMX Edit
eaglememex.language.sort.order=DE;EN;FR;IT;RU;ES;PT;NL;CS;PL;ZH
# do not show format tag option menu entry
eaglememex.No.Format=false
# Serialise Araya database (internal) when closing XLIFF Editor
eaglememex.database.serialise=yes
# use bulkinsert as default
eaglememex.database.bulkinsert=false
# write tracing information for tmx accesses into log file
eaglememex.trace.tm=true
# CheckAdding defines if new translations should be entered automatically
or asked for adding – this is obsolete now
eaglememex.database.CheckAddingTMXEntry=false
```

```
# Enforce reading TMX db attribute
eaglememex.enforce.attributesreading=false
# Controls saving of translation which match a term entry
eaglememex.database.DoNotSaveTermEntriesInTMXDBSegment=true
# remove multiple eblanks in element tags (<tag   x=" " > --> <tag x=" ">)
eaglememex.xml.doubleblanks=true
# run format checker before Approve All
eaglememex.RunFormatCheckBefore=false
# run translation check before Goto Operations (checks if entries source /
target in database)
eaglememex.RunGotoCheckBefore=false
# when true the operations in tmx edit entry are traced
eaglememex.tmxedit.trace=true


# CheckTMXDBMergeAdding defines if translations should be entered
automatically or asked if it results merged entries
eaglememex.database.CheckTMXDBMergeAdding=false
# CheckTMXDNewAdding defines if  translations should be entered
automatically or asked for new entries (source and target new)
eaglememex.database.CheckTMXDNewAdding=false
# CheckTMXSourceAdding defines if  translations should be entered
automatically or asked for new source entries (target exists)
eaglememex.database.CheckTMXSourceAdding=false
# CheckTMXTargetAdding defines if  translations should be entered
automatically or asked for new target entries (source exists)
eaglememex.database.CheckTMXTargetAdding=false


# Extraction PDF Files
biextract.userguide.en=ARRAYPATHDIRECTORY/docs/biextract_en.pdf
biextract.userguide.de=ARRAYPATHDIRECTORY/docs/biextract_de.pdf
# Bi Tmx Editor PDF Files
bitmx.userguide.en=ARRAYPATHDIRECTORY/docs/bitmx_en.pdf
bitmx.userguide.de=ARRAYPATHDIRECTORY/docs/bitmx_de.pdf
# AutoSearchInTerm - true - automatically phrase translate if true
eaglememex.database.AutoSearchInTerm=true
#End


# defines elements in xml files which sould be translated
eaglememex.translate.plicsmenue=DE,EN,FR,IT,ES,RU,ZH
# replacement language to be used if not in the elements language defined
above
eaglememex.translate.plicsmenue.default=en
# use EN for element in case of cs (example)
eaglememex.translate.plicsmenue.cs=EN


# Default Fonts
# height of font
```

```
eaglememex.font.height=8
# font type
eaglememex.font.name=@Arial Unicode MS
# style of font
eaglememex.font.style=0


eaglememex.val.prop=10
# if a string is less equal this length use 100% search only
eaglememex.val.shortstring=3


# Post process the xml file; applied after the conversion to xml from xliff
# Post process the VEGA file
eaglememex.postprocess.xml.back.zh-
cn=ARRAYPATHDIRECTORY/ini/postprocess.zh-cn.txt
# Post process the xml file for edqm es
eaglememex.postprocess.xml.back.es=ARRAYPATHDIRECTORY/ini/edqmtoent.es.txt
eaglememex.postprocess.xml.back.en=ARRAYPATHDIRECTORY/ini/edqmtoent.es.txt
# Pre process the xml file for edqm es when converting to xliff
eaglememex.preprocess.xml.es=ARRAYPATHDIRECTORY/ini/edqmfroment.es.txt
eaglememex.preprocess.xml.en=ARRAYPATHDIRECTORY/ini/edqmfroment.es.txt


# Default Editor Fuzzy Value
# default fuzzy search percentage = similarity
eaglememex.fuzzy.search=80
# default terminology database
# eaglememex.default.termdatabase=tourismmterm
# eaglememex.default.tmxdatabase=empty


# XLIFF Editor
# XLIFF editor config file - by default replaced by xeditcfg + user name +
.xml
xliffedit.configuration=ARRAYPATHDIRECTORY/lib/xeditcfg.xml
# Navigation stop even if 100% is defined
eaglememex.check100.stop=98


# TMX Editor
# TMX editor config file
tmxedit.configuration=ARRAYPATHDIRECTORY/lib/tmxedcfg.xml
# allowed white black green lightyellow red beige lightblue pink brown
# color for TMX Editor background
tmxedit.cursor.background=black
# color for TMX Editor foreground
tmxedit.cursor.foreground=white


# ASPELL
# path to APELL exe file
aspell.commandline=c:/Programme/aspell/bin/aspell.exe
```

```
# default dictionaries
aspell.dictionary.en=en
aspell.dictionary.de=de
aspell.dictionary.pt=pt


# [align properties]
align.replacementPenalty = 1
align.insertionPenalty = 1
align.deletionPenalty = 1
# [end align section ]


# [match class]
match.tm.maximum=5
# [End]


#[TBX map file]
# tbx (terminology) definition file
mapfile=ARRAYPATHDIRECTORY/lib/tbxmap.dat
termdatabasefile=ARRAYPATHDIRECTORY/ini/openwebterm.txt
# path to irregular form stemmer file
irregularfile=ARRAYPATHDIRECTORY/irregular.xml
# [End TBX map file]


# [Skeleton file]
# defines the size of the skeleton file kept internally
skl.internal.max.size = 1000
# [End Skeleton]


# [Segmentation]
segmentation.enable=false
# default value for number of cr/lf to terminate a segment
segmentation.plaintext.crlf=0
# [End Segmentation]


# [Replacement Class]
# replacement class definition file (starts with the standrad Araya ini
file path!)
replacement.class.file=replclass.xml
# [End Replacement]


# [Tm Database]
debug.level=0
database.type=HSQL
# Path to Araya database directory
database.path=ARRAYPATHDIRECTORY/database/
# Path to Araya database list
database.list=ARRAYPATHDIRECTORY/database/databaseList.xml
```

```
# eaglememx.tm.dbname=translationwork
# eaglememx.tm.user=sa
# eaglememx.tm.password=sa
# eaglememx.tm.dbtype=MsSQL2000
# eaglememx.tm.dbhost=localhost
# eaglememx.tm.dbport=1433
# [End Tm]


# [RPC Service section]
# default port for Araya XML RPC Server
rpc.server.port = 4040
# Conectio string for XML RPC Server
rpc.server.connectstring=http://localhost:4040
# name of service
rpc.server.service.name = $default
# Service names to be loaded
rpc.translation.service.name = TranslationTools
rpc.util.service.name = Utilities
rpc.registration.service.name = Registration
rpc.stem.service.name = Stem
rpc.terminology.service.name= Terminology
rpc.termextract.service.name= TermExtract
rpc.termutil.service.name= TermUtilities
rpc.pt.service.name = PhraseTranslatorService


# [End RPC Service]

#Format Detection Properties
#Fri Sep 12, 2002


# Format.ID[xx]
# name of the format + associated number code
format.id0=plaintext
format.id1=xliff
format.id2=tmx
format.id3=doctrans
format.id4=php
format.id5=xml
format.id6=html
format.id7=mif
format.id8=pdf
format.id9=sgml
format.id10=rtf
format.id11=winword
format.id12=cgi
format.id13=pl
```

```
format.id14=jar
format.id15=js
format.id16=sql
format.id17=_js_
format.id18=sxw
format.id19=sxc


# Format.Pattern
# defines the pattern which allows to determine the document format (start
part refers to format defined above)
plaintext.pattern=.*
html.pattern=.*<html.*>.*
xliff.pattern=.*<\?xml.*\?>.*<xliff.*
tmx.pattern=.*<\?xml.*\?>.*<tmx.*
doctrans.pattern=.*<\?xml.*\?>.*<doctrans.*
php.pattern=.*<\?php.*
xml.pattern=.*<\?xml.*\?>.*
# sgml.pattern=.*<sgml>.*|[^%PDF].*<[^MIF].*>.*
sgml.pattern=.*<sgml>.*|^<.*?>.*?<\.*?>
rtf.pattern=^\{\\rtf.*
mif.pattern=^<MIF.*
pdf.pattern=^%PDF.*
winword.pattern=\xD0\xCF


#End


# [InitFile Section]


# [html section]
# html file description
init.html.default = init.xml
init.php.default = init.xml
# [end html section]


# [xml section]
# defines association with special xml formats
pattern.xml.id1=<doctrans
init.xml.default = init_docbook.xml
init.xml.id1 = init_doctrans.xml
pattern.xml.id2=<svg
init.xml.id2 = init_svg.xml
pattern.xml.id3="ba_v1.dtd"
init.xml.id3 = init_ba.xml
pattern.xml.id4=<TRADOStag
init.xml.id4 = init_trados.xml
pattern.xml.id5=<office:script
init.xml.id5 = init_openoffice.xml
```

```
# [end xml section]


# [sgml section]
# sgml detection
pattern.sgml.id1=<!DOCTYPE BOOK PUBLIC "-//Sun Microsystems//DTD DocBook
V3.0-Based SolBook

init.sgml.default = init_docbook.xml
init.sgml.id1 = init_solbook.xml
# [end sgml section]


# [rtf section]
init.rtf.default = init_rtf.xml
# [end rtf section]



# extension section
ext.jarfiles=ARRAYPATHDIRECTORY/lib/sqlconvert.jar;ARRAYPATHDIRECTORY/lib/c
onverters.jar
# associate file name extensions with format
formatext.id0=js
formatext.id1=cgi
formatext.id2=pl
formatext.id3=jar
formatext.id4=sql
formatext.id5=_js_
formatext.id6=sxw
formatext.id7=sxc


# define the methods to be called for the avove extensions
init.js.id0=init_js.xml
method.js.id0=jscript2xliff.JavaScriptConverter
datatype.js.id0=JavaScript
init.cgi.id1=init_pl.xml
init.cgi.id2=init_pl.xml
method.js.id0=com.araya.converters.javascript.JavaScript
init.jar.id3=init_jar.xml
method.jar.id3=com.araya.converters.javascript.JarFile
init.sql.id4=init_sql.xml
method.sql.id4=com.araya.converters.sql.SQLConverter
init._js_.id5=init_js.xml
method._js_.id5=com.araya.converters.javascript.JavaScript
datatype._js_.id5=JavaScript
init.sxw.id6=init_openoffice.xml
method.sxw.id6=com.araya.eaglememex.OpenOffice.OpenOffice2Xliff
```

```
datatype.sxw.id6=OpenOffice
init.sxc.id7=init_openoffice.xml
method.sxc.id7=com.araya.eaglememex.OpenOffice.OpenOffice2Xliff
datatype.sxc.id7=OpenOffice


# [segmenter section]
init.segmenter.default = init_segmenter.xml
# [end segmenter section]
# [End init file section]


# [language section]
# define the short and long language names
af=Afrikaans
sq=Albanian
ar-iq=Arabic(Iraq)
ar-dz=Arabic(Algeria)
ar-bh=Arabic(Bahrain)
ar-eg=Arabic(Egypt)
ar-jo=Arabic(Jordan)
ar-kw=Arabic(Kuwait)
ar-lb=Arabic(Lebanon)
ar-ly=Arabic(Libya)
ar-ma=Arabic(Morocco)
ar-om=Arabic(Oman)
ar-qa=Arabic(Qatar)
ar-sa=Arabic(Saudi Arabia)
ar-sy=Arabic(Syria)
ar-tn=Arabic(Tunisia)
ar-ae=Arabic(U.A.E.)
ar-ye=Arabic(Yemen)
eu=Basque
bg=Bulgarian
ca=Catalan
zh-hk=Chinese(Hong Kong)
zh-cn=Chinese(People's Republic)
zh-sg=Chinese(Singapore)
zh-tw=Chinese(Taiwan)
hr=Croatian
cs=Czech
da=Danish
nl=Dutch(Standard)
nl-be=Dutch(Belgian)
en=English
en-au=English(Australian)
en-bz=English(Belize)
en-gb=English(British)
```

```
en-ca=English(Canadian)
en-cb=English(Caribbean)
en-ie=English(Ireland)
en-jm=English(Jamaica)
en-nz=English(New Zealand)
en-za=English(South Africa)
en-tt=English(Trinidad)
en-us=English(United States)
et=Estonian
fa=Persian(Farsi)
fo=Faeroese
fi=Finnish
fr-be=French(Belgian)
fr-ca=French(Canadian)
fr-lu=French(Luxembourg)
fr=French(Standard)
fr-ch=French(Swiss)
de-at=German(Austrian)
de-li=German(Liechtenstein)
de-lu=German(Luxembourg)
de=German(Standard)
de-ch=German(Swiss)
el=Greek
he=Hebrew
hu=Hungarian
is=Icelandic
in=Indonesian
it=Italian(Standard)
it-ch=Italian(Swiss)
ja=Japanese
ko=Korean
ko-jb=Korean(Johab)
lv=Latvian
lt=Lithuanian
no=Norwegian
pl=Polish
pt-br=Portuguese(Brazilian)
pt=Portuguese(Standard)
ro=Romanian
ru=Russian
sr=Serbian
sk=Slovak
sl=Slovenian
es-ar=Spanish(Argentina)
es-bo=Spanish(Bolivia)
es-cl=Spanish(Chile)
es-co=Spanish(Colombia)
```

```
es-cr=Spanish(Costa Rica)
es-do=Spanish(Dominican Republic)
es-ec=Spanish(Ecuador)
es-sv=Spanish(El Salvador)
es-hn=Spanish(Honduras)
es-gt=Spanish(Guatemala)
es-mx=Spanish(Mexican)
es-ni=Spanish(Nicaragua)
es-pa=Spanish(Panama)
es-py=Spanish(Paraguay)
es-pe=Spanish(Peru)
es-pr=Spanish(Puerto Rico)
es=Spanish(Spain)
es-uy=Spanish(Uruguay)
es-ve=Spanish(Venezuela)
sv=Swedish
sv-fi=Swedish(Finland)
th=Thai
tr=Turkish
uk=Ukrainian
vi=Vietnamese
# [end language section]
```

12.2 Alignment Properties File

Align.properties

Trans.properties

12.3 Araya XLIFF Editor Configuration File

Example: xeditcfg.klemens.waldhoer.xml

12.4 Araya TMX Editor Configuration File

12.5 Postprocessing File

This file allows to define replacement characters after the back conversion of xliff files which result in xml files that is the original format was xml. The given character sequence is replaced by the hexadecimal value specified.

```
// use decimal values!!!
// Â°F &#2109; hex equals dec 8457
```

```
°F=&#8457;
// µ &#03BC; hex equals dec 956
µ=&#956;
// °C &#2103; hex equals dec 8451
°C=&#8451;
```

°F=℉
// µ &#03BC; hex equals dec 956

13 Database Table Definitions in Araya

13.1 MONO Table Definitions

13.1.1 Table Definitions

```
CREATE TABLE [MONO] (
      [id] [int] IDENTITY (1, 1) NOT NULL ,      7
      [idref] [varchar] (40) ,                   1156244076367
      [idno] [varchar] (20) ,                    1156244076367.de
      [segment] [nvarchar] (255) ,               <seg>Zukunftsvision
Marktplatz Hotelzimmer</seg>
      [language] [varchar] (5) ,                 de
      [locktime] [datetime] NULL ,
      [lockstate] [int] NULL ,
      [lockuser] [varchar] (30) ,
      [readaccess] [int] NULL ,
      [writeaccess] [int] NULL ,
      [tmuser] [varchar] (30) ,
      [groups] [varchar] (30) ,
      [segmentfull] [ntext] ,                    <seg>Zukunftsvision
Marktplatz Hotelzimmer</seg>
      [fdomain] [varchar] (30) ,
      [document] [varchar] (255) ,
      [creationdate] [varchar] (30) ,
      [modificationdate] [varchar] (30) ,
      [creationauthor] [varchar] (30) ,
      [modificationauthor] [varchar] (30) ,
      [replclasses] [int] NULL ,                 0
      [documentid] [varchar] (30) ,
      [attributes] [varchar] (255) ,
      [changedate] [varchar] (100) ,             1156576436452
      [TIMER] [int] NULL                         528164361
```

13.1.2 Indexes

```
CREATE  INDEX [segment] ON [MONO]([segment]
CREATE  INDEX [idref] ON [MONO]([idref], [language]
CREATE  INDEX [fdomain_index] ON [MONO]([fdomain]
CREATE  INDEX [document_index] ON [MONO]([document]        ***
CREATE  INDEX [documentid_index] ON [MONO]([documentid]    ***
CREATE  INDEX [attributes_index] ON [MONO]([attributes]    ***
CREATE  INDEX [monotimerindex] ON [MONO]([TIMER]
CREATE  INDEX [monotimerlanindex] ON [MONO]([TIMER], [language]
```

13.1.3 Commands

```
<search-mono-segment>
      type="FULLTEXTSEARCH">SELECT idref, idno, segment, language,
segmentfull, changedate, creationdate, modificationdate, creationauthor,
modificationauthor FROM MONO WHERE MATCH (segmentfull) AGAINST ? AND
language =?
```

```
        type="equal">SELECT idref, idno, segment, language, segmentfull,
changedate, creationdate, modificationdate, creationauthor,
modificationauthor  FROM MONO WHERE segment = ? AND language =?
        type="idref">SELECT idref, idno, segment, language, segmentfull,
changedate, creationdate, modificationdate, creationauthor,
modificationauthor  FROM MONO WHERE idref = ?
        type="idref_and_lang">SELECT idref, idno, segment, language,
segmentfull, changedate, creationdate, modificationdate, creationauthor,
modificationauthor  FROM MONO WHERE idref = ? AND language IN (?) ORDER BY
language
        type="allmonoentries">SELECT idref, idno, segment, language,
segmentfull, changedate, creationdate, modificationdate, creationauthor,
modificationauthor  FROM MONO WHERE language = ?  AND TIMER > ? ORDER BY
idref
        type="idrefin">SELECT idref, idno, segment, language, segmentfull,
changedate, creationdate, modificationdate, creationauthor,
modificationauthor  FROM MONO WHERE idref IN
<search-all-languages>SELECT DISTINCT language FROM MONO WHERE idref = ?
<search-mono-idno-segment>SELECT idref, idno, segment, language,
segmentfull, changedate, creationdate, modificationdate, creationauthor,
modificationauthor  FROM MONO WHERE idno = ?
<insert-mono-segment>INSERT INTO MONO (idref, idno, segment, language,
segmentfull, replclasses, fdomain, document, documentid, attributes,
creationdate, modificationdate, creationauthor, modificationauthor, tmuser,
changedate, TIMER) VALUES (?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?, ?)
<replace-mono-segment>REPLACE INTO MONO (idref, idno, segment, language,
segmentfull, replclasses, fdomain, document, documentid, attributes,
creationdate, modificationdate, creationauthor, modificationauthor, tmuser,
changedate, TIMER) VALUES (?,?,?,?,?,?,?,?,?,?, ?,?,?,?,?,?, ?)
<updatemonotimer>UPDATE MONO SET TIMER=?  WHERE idref=?
<get-mono>
        type="all">SELECT idref, idno, segment, language, segmentfull,
fdomain, document, documentid, attributes, creationdate, modificationdate,
creationauthor, modificationauthor, tmuser FROM MONO ORDER BY idref
        type="langs">SELECT idref, idno, segment, language, segmentfull,
fdomain, document, documentid, attributes, creationdate, modificationdate,
creationauthor, modificationauthor, tmuser FROM MONO WHERE language IN (?)
ORDER BY idref
        type="all_norepl">SELECT idref, idno, segment, language, segmentfull,
fdomain, document, documentid, attributes, creationdate, modificationdate,
creationauthor, modificationauthor, tmuser FROM MONO WHERE (replclasses =
0) ORDER BY idref
        type="langs_norepl">SELECT idref, idno, segment, language,
segmentfull, fdomain, document, documentid, attributes, creationdate,
modificationdate, creationauthor, modificationauthor, tmuser FROM MONO
WHERE language IN (?) AND (replclasses = 0) ORDER BY idref
<delete-mono>DELETE FROM MONO WHERE idno = ?
<delete-idref>DELETE FROM MONO WHERE idref = ?
<select-attributes>SELECT DISTINCT fdomain, document, documentid,
attributes, creationdate, modificationdate, creationauthor,
modificationauthor, tmuser FROM MONO
<select-id>SELECT DISTINCT idref FROM MONO
<get-multiple-languages>select count(idno), idno, idref from MONO group by
idno, idref having count(idno) > 1
<get-all-segments>SELECT idref, idno, segment, language, segmentfull FROM
MONO
<domain>fdomain
<document>document
<documentid>documentid
<attributes>attributes
<selectid>idref
<CreationDate>creationdate
```

```
<ModificationDate>modificationdate
<CreationAuthor>creationauthor
<ModificationAuthor>modificationauthor
<Tmuser>tmuser
```

13.2 MULTI Table Definition

13.2.1 Table Definition

```
CREATE TABLE [MULTI] (
      [id] [int] IDENTITY (1, 1) NOT NULL ,
      [idref] [varchar] (40) ,
      [srclanguage] [varchar] (5) ,
      [lockstate] [int] NULL ,
      [locktime] [datetime] NULL ,
      [lockuser] [varchar] (30) ,
      [readaccess] [int] NULL ,
      [writeaccess] [int] NULL ,
      [tmuser] [varchar] (30) ,
      [groups] [varchar] (30)
```

13.2.2 Indexes

```
CREATE  INDEX [idref] ON [MULTI]([idref]
```

13.2.3 Commands

```
<insert-multi>INSERT INTO MULTI (idref) VALUES (?)
<replace-multi>REPLACE INTO MULTI (idref) VALUES (?)
<delete-multi>DELETE FROM MULTI WHERE idno = ?
<deletemulti-idref>DELETE FROM MULTI WHERE idref = ?
```

13.3 WORD INDEX Table

13.3.1   Table Definition

```
CREATE TABLE [WORDINDEX] (
     [id] [int] IDENTITY (1, 1) NOT NULL ,      21
     [termidref] [char] (30) ,                  1156244076367
     [language] [char] (5) ,                    de
     [word] [char] (255) ,                      Hotelzimmer
     [type] [int] NULL                          0
```

13.3.2   Indexes

```
CREATE  INDEX [wordindex_word] ON [WORDINDEX]([termidref], [word],
[language], [type]
CREATE  INDEX [wordlan] ON [WORDINDEX]([word], [language]
CREATE  INDEX [wordid] ON [WORDINDEX]([termidref]
CREATE  INDEX [wordind] ON [WORDINDEX]([word]
```

13.3.3   Commands

```
<insert-wordindex>INSERT INTO WORDINDEX (termidref, word, language, type)
VALUES (?, ?, ?, 0)
<insert-stemindex>INSERT INTO WORDINDEX (termidref, word, language, type)
VALUES (?, ?, ?, 1)
<insert-word>INSERT INTO WORDINDEX (termidref, word, language, type) VALUES
(?, ?, ?,?)
<get-word>SELECT DISTINCT termidref, type, language FROM WORDINDEX WHERE
(word IN (VALUES))
<delete-word>DELETE FROM WORDINDEX where language =? AND termidref=?
<delete-word-all>DELETE FROM WORDINDEX where termidref=?
<delete-wordindex>DELETE FROM WORDINDEX
<recreate-wordindex-table>
     step="1">DROP TABLE WORDINDEX
     step="2">CREATE TABLE WORDINDEX (id INT IDENTITY (1, 1) NOT NULL
PRIMARY KEY, termidref CHAR(30), language CHAR(5), word CHAR(255), type
INT)
     step="3">CREATE INDEX wordlan ON WORDINDEX (word, language)
     step="4">CREATE INDEX wordid ON WORDINDEX (termidref)
     step="5">CREATE INDEX wordind ON WORDINDEX (word)
```

13.4 FUZZY Table Definition

13.4.1   Table Definition

```
CREATE TABLE [FUZZY] (
      [id] [int] IDENTITY (1, 1) NOT NULL ,      100
      [KD_ID] [varchar] (40) ,                   10d49496783
      [PARENT_ID] [varchar] (40) ,               10d494969e9
      [LOSON_ID] [varchar] (40) ,                (empty)
      [HISON_ID] [varchar] (40) ,                (empty)
      [language] [varchar] (5) ,                 en
      [data] [text] ,
1155030881346.en|1155031427562.en|1155031427532.en
      [keyvec] [varchar] (255) ,
0|0|1|0|0|0|0|0|1|0|0|1|0|0|4|2|0|0|0|0|0|0|0|0|1|0|0|0|0|1|1|1|0|1|1|0|0|0
|0|0|1|0|0|1|0|0|0|0
      [notrigrams] [int] NULL ,                  17
      [lockstate] [int] NULL ,                   1
      [locktime] [datetime] NULL ,               NULL
      [lockuser] [varchar] (30) ,                NULL
      [readaccess] [int] NULL ,                  NULL
      [writeaccess] [int] NULL ,                 1
      [tmuser] [varchar] (30) ,                  NULL
      [groups] [varchar] (30) ,                  NULL
      [replclasses] [int] NULL ,                 0
      [LEVEL] [int] NULL ,                       0
      [TIMER] [int] NULL                         528164977
```

13.4.2   Indexes

```
CREATE  INDEX [KD_ID_FUZZY] ON [FUZZY]([KD_ID]
CREATE  INDEX [language] ON [FUZZY]([language]
CREATE  INDEX [languagetrigram] ON [FUZZY]([language], [notrigrams]
CREATE  INDEX [timerindex] ON [FUZZY]([TIMER]
```

13.4.3   Commands

```
<search-fuzzy-lang>SELECT language, data, keyvec, notrigrams, writeaccess,
replclasses FROM FUZZY WHERE language = ? ORDER BY id
<new-search-fuzzy-lang>SELECT language, data, keyvec, notrigrams,
writeaccess, replclasses, KD_ID, PARENT_ID, LOSON_ID, HISON_ID, LEVEL FROM
FUZZY WHERE language = ? AND PARENT_ID = '' AND TIMER > ? ORDER BY id
<new-search-fuzzy-childs-lang>SELECT language, data, keyvec, notrigrams,
writeaccess, replclasses, KD_ID, PARENT_ID, LOSON_ID, HISON_ID, LEVEL FROM
FUZZY WHERE language = ? AND notrigrams = ? AND replclasses = ? AND TIMER >
? ORDER BY id
<WkFuzzyIndexReadCommand>SELECT language, data, keyvec, notrigrams,
writeaccess, replclasses, KD_ID, PARENT_ID, LOSON_ID, HISON_ID, LEVEL FROM
FUZZY WHERE language = ? AND TIMER > ? ORDER BY PARENT_ID
<new-insert-fuzzy>INSERT INTO FUZZY (language, data, keyvec, notrigrams,
writeaccess, lockstate, replclasses, KD_ID, PARENT_ID, LOSON_ID, HISON_ID,
LEVEL, TIMER) VALUES (?, ?, ?, ?, 1, 1, ?, ?, ?, ?, ?, ?, ?)
```

```
<new-update-fuzzy>UPDATE FUZZY SET DATA=?, LOSON_ID=?, HISON_ID=?, TIMER=?
WHERE KD_ID=?
<delete-fuzzy-lang>DELETE FROM FUZZY where language =?
<insert-fuzzy>INSERT INTO FUZZY (language, data, keyvec, notrigrams,
writeaccess, lockstate, replclasses, TIMER) VALUES (?, ?, ?, ?, 1, 1, ?, ?)
<get-writeaccess-fuzzy>SELECT writeaccess FROM FUZZY WHERE language = ?
<updatefuzzydata>UPDATE FUZZY SET data=?, TIMER=? WHERE (KD_ID = ?)
<search-fuzzy-data>SELECT data, language, notrigrams, replclasses FROM
FUZZY WHERE (KD_ID = ?)
```

13.5 FUZZYREF Table Definition

13.5.1   Table Definition

```
CREATE TABLE [FUZZYREF] (
      [id] [int] IDENTITY (1, 1) NOT NULL ,     7
      [KD_ID] [varchar] (40) ,                  10d494966f5
      [idref] [varchar] (40) ,                  1156244076367
      [idreflan] [varchar] (40) ,               1156244076367.de
      [sourcelan] [varchar] (5) ,               de
      [targetlan] [varchar] (5) ,               en
      [TIMER] [int] NULL                        528164378
```

13.5.2   Indexes

```
CREATE  INDEX [KD_ID_FUZZYREF_index] ON [FUZZYREF]([KD_ID]
CREATE  INDEX [idref_FUZZYREF] ON [FUZZYREF]([idref]
CREATE  INDEX [ist_FUZZYREF] ON [FUZZYREF]([idref], [sourcelan],
[targetlan]
```

13.5.3   Commands

```
<insertfuzzyref>INSERT INTO FUZZYREF (KD_ID, idref, idreflan, sourcelan,
targetlan, TIMER) VALUES (?, ?, ?, ?, ?, ?)
<selectfuzzyref>SELECT KD_ID, idref, idreflan, sourcelan, targetlan FROM
FUZZYREF WHERE (idref = ?)
<deletefuzzyref>DELETE FROM FUZZYREF WHERE idref = ?
```

13.6 LOCKER Table Defintion

13.6.1   Table Definition

```
CREATE TABLE [LOCKER] (
      [id] [int] IDENTITY (1, 1) NOT NULL ,     1
      [lockstate] [int] NULL ,                  532889676
      [changeuser] [varchar] (100) ,            klemens.waldhoer
      [lastupdatetime] [varchar] (100)          1157048967671
```

13.6.2   Indexes

13.6.3   Commands

```
<get-lockstate>
            step="1">SELECT lockstate, changeuser, lastupdatetime FROM
LOCKER
<lock-locker>
<update-locker><command step="1">UPDATE LOCKER SET lockstate=?,
changeuser=?, lastupdatetime=?
<unlock-table>
```

13.7 *Other Commands*

```
<delete-content-of-tables>
     step="1">TRUNCATE TABLE WORDINDEX
     step="2">TRUNCATE TABLE FUZZY
     step="3">TRUNCATE TABLE FUZZYREF
     step="4">TRUNCATE TABLE MONO
     step="5">TRUNCATE TABLE MULTI
     step="6">TRUNCATE TABLE LOCKER
     step="7">INSERT INTO LOCKER (lockstate, changeuser, lastupdatetime)
VALUES(0, ' ', ' ')


<delete-content-of-tables-1>
     step="1">DELETE FROM WORDINDEX
     step="2">DELETE FROM FUZZY
     step="3">DELETE FROM FUZZYREF
     step="4">DELETE FROM MONO
     step="5">DELETE FROM MULTI
     step="6">DELETE FROM LOCKER


<lock-fuzzy-mono-multi>LOCK TABLES FUZZY WRITE, LOCK TABLES MONO WRITE,
LOCK TABLES MULTI WRITE
```

14 Important TMX Definitions

14.1 Base Document Element

```
<!ELEMENT tmx            (header, body) >
<!ATTLIST tmx
     version             CDATA           #FIXED "1.4" >
```

14.2 Header

```
<!ELEMENT header         (note|prop|ude)* >
<!ATTLIST header
     creationtool        CDATA           #REQUIRED
     creationtoolversion CDATA           #REQUIRED
     segtype             (%segtypes;)    #REQUIRED
     o-tmf               CDATA           #REQUIRED
     adminlang           CDATA           #REQUIRED
     srclang             CDATA           #REQUIRED
     datatype            CDATA           #REQUIRED
     o-encoding          CDATA           #IMPLIED
     creationdate        CDATA           #IMPLIED
     creationid          CDATA           #IMPLIED
     changedate          CDATA           #IMPLIED
     changeid            CDATA           #IMPLIED >
```

14.3 Body

```
<!ELEMENT body           (tu*) >
<!-- No attributes        -->
```

14.4 Note

```
<!ELEMENT note           (#PCDATA) >
<!ATTLIST note
     o-encoding          CDATA           #IMPLIED
     xml:lang            CDATA           #IMPLIED
     lang                CDATA           #IMPLIED >
     <!-- lang is deprecated: use xml:lang -->
```

14.5 Property

```
<!ELEMENT prop           (#PCDATA) >
<!ATTLIST prop
     type                CDATA           #REQUIRED
     xml:lang            CDATA           #IMPLIED
     o-encoding          CDATA           #IMPLIED
     lang                CDATA           #IMPLIED >
     <!-- lang is deprecated: use xml:lang -->
```

14.6 Translation Unit

```
<!ELEMENT tu             ((note|prop)*, tuv+) >
```

```
<!ATTLIST tu
    tuid                 CDATA          #IMPLIED
    o-encoding           CDATA          #IMPLIED
    datatype             CDATA          #IMPLIED
    usagecount           CDATA          #IMPLIED
    lastusagedate        CDATA          #IMPLIED
    creationtool         CDATA          #IMPLIED
    creationtoolversion CDATA           #IMPLIED
    creationdate         CDATA          #IMPLIED
    creationid           CDATA          #IMPLIED
    changedate           CDATA          #IMPLIED
    segtype              (%segtypes;)   #IMPLIED
    changeid             CDATA          #IMPLIED
    o-tmf                CDATA          #IMPLIED
    srclang              CDATA          #IMPLIED >
```

14.7 Translation Unit Variant

```
<!ELEMENT tuv              ((note|prop)*, seg) >
<!ATTLIST tuv
    xml:lang             CDATA          #REQUIRED
    o-encoding           CDATA          #IMPLIED
    datatype             CDATA          #IMPLIED
    usagecount           CDATA          #IMPLIED
    lastusagedate        CDATA          #IMPLIED
    creationtool         CDATA          #IMPLIED
    creationtoolversion CDATA           #IMPLIED
    creationdate         CDATA          #IMPLIED
    creationid           CDATA          #IMPLIED
    changedate           CDATA          #IMPLIED
    o-tmf                CDATA          #IMPLIED
    changeid             CDATA          #IMPLIED
    lang                 CDATA          #IMPLIED >
    <!-- lang is deprecated: use xml:lang -->
```

14.8 Text

```
<!ELEMENT seg              (#PCDATA|bpt|ept|ph|it|hi|ut)* >

<!ENTITY % segtypes     "block|paragraph|sentence|phrase" >
```

14.9 Content Markup

```
<!ELEMENT bpt              (#PCDATA|sub)* >
<!ATTLIST bpt
    i                    CDATA          #REQUIRED
    x                    CDATA          #IMPLIED
    type                 CDATA          #IMPLIED >

<!ELEMENT ept              (#PCDATA|sub)* >
<!ATTLIST ept
    i                    CDATA          #REQUIRED >

<!ELEMENT sub              (#PCDATA|bpt|ept|it|ph|hi|ut)* >
<!ATTLIST sub
    datatype             CDATA          #IMPLIED
    type                 CDATA          #IMPLIED >
```

```
<!ELEMENT it              (#PCDATA|sub)* >
<!ATTLIST it
      pos                 (begin|end)     #REQUIRED
      x                   CDATA           #IMPLIED
      type                CDATA           #IMPLIED >

<!ELEMENT ph              (#PCDATA|sub)* >
<!ATTLIST ph
      x                   CDATA           #IMPLIED
      assoc               CDATA           #IMPLIED
      type                CDATA           #IMPLIED >

<!ELEMENT hi              (#PCDATA|bpt|ept|it|ph|hi|ut)* >
<!ATTLIST hi
      x                   CDATA           #IMPLIED
      type                CDATA           #IMPLIED >


<!-- The <ut> element is deprecated -->

<!ELEMENT ut              (#PCDATA|sub)* >
<!ATTLIST ut
      x                   CDATA           #IMPLIED >
```

14.10 TMX Example File as in Araya

```xml
<?xml version="1.0" encoding="UTF-8"?>
<tmx version="1.4">
      <header creationtool="XLFEdit" creationtoolversion="3.2.1"
srclang="" adminlang="en"  datatype="xml" o-tmf="TmxObject"
segtype="paragraph"/>
            <prop type="entrynumber">1</prop>
      </header>
      <body>
      <tu creationdate="535126438" creationid="10d65683ed1"
changeid="10d65683ed1" changedate="535126438" usagecount="0">
            <prop type="entrynumber">1</prop>
            <prop type="modificationdate">535126438</prop>
            <prop type="modificationauthor">klemens.waldhoer</prop>
            <prop type="creationautor">klemens.waldhoer</prop>
            <prop type="document">test2.tmx</prop>
            <prop type="tmuser">klemens.waldhoer</prop>
            <tuv creationdate="535126438" creationid="10d65683ed1.fr"
changeid="10d65683ed1" changedate="535126438" usagecount="0">
                  <prop type="modificationdate">535126438</prop>
                  <prop type="modificationauthor">klemens.waldhoer</prop>
                  <prop type="creationautor">klemens.waldhoer</prop>
                  <prop type="document">test2.tmx</prop>
                  <prop type="tmuser">klemens.waldhoer</prop>
                  <seg>Projektaire -</seg>
            </tuv>
            <tuv creationdate="535126438" creationid="10d65683ed1.fr"
changeid="10d65683ed1" changedate="535126438" usagecount="0">
                  <prop type="modificationdate">535126438</prop>
                  <prop type="modificationauthor">klemens.waldhoer</prop>
                  <prop type="creationautor">klemens.waldhoer</prop>
                  <prop type="document">test2.tmx</prop>
                  <prop type="tmuser">klemens.waldhoer</prop>
                  <seg>Projektaire</seg>
            </tuv>
            <tuv creationdate="535126438" creationid="10d65683ed1.en"
changeid="10d65683ed1" changedate="535126438" usagecount="0">
                  <prop type="modificationdate">enenen</prop>
                  <prop type="modificationauthor">englisch</prop>
                  <prop type="creationautor">klemens.waldhoer</prop>
                  <prop type="document">test2.tmx</prop>
                  <prop type="tmuser">klemens.waldhoer</prop>
                  <seg>Projects - xxx</seg>
            </tuv>
            <tuv creationdate="535126438" creationid="10d65683ed1.en"
changeid="10d65683ed1" changedate="535126438" usagecount="0">
                  <prop type="modificationdate">535126438</prop>
                  <prop type="modificationauthor">klemens.waldhoer</prop>
                  <prop type="creationautor">klemens.waldhoer</prop>
                  <prop type="document">test2.tmx</prop>
                  <prop type="tmuser">klemens.waldhoer</prop>
                  <seg>Projects -</seg>
            </tuv>
            <tuv creationdate="535126438" creationid="10d65683ed1.en"
changeid="10d65683ed1" changedate="535126438" usagecount="0">
                  <prop type="modificationdate">535126438</prop>
                  <prop type="modificationauthor">klemens.waldhoer</prop>
                  <prop type="creationautor">klemens.waldhoer</prop>
```

```xml
                    <prop type="document">test2.tmx</prop>
                    <prop type="tmuser">klemens.waldhoer</prop>
                    <seg>Projects - vvv</seg>
            </tuv>
            <tuv creationdate="535126438" creationid="10d65683ed1.de"
changeid="10d65683ed1" changedate="535126438" usagecount="0">
                    <prop type="modificationdate">ssssss</prop>
                    <prop type="modificationauthor">deutsch</prop>
                    <prop type="creationautor">klemens.waldhoer</prop>
                    <prop type="document">test2.tmx</prop>
                    <prop type="tmuser">klemens.waldhoer</prop>
                    <seg>Projekte -</seg>
            </tuv>
        </tu>
</body>
</tmx>
```

15 Important XLIFF Elements

15.1 Structural Elements

```
<!ELEMENT xliff       (file)+ >
<!ATTLIST xliff
    version           CDATA                 #FIXED "1.0"
    xml:lang          CDATA                 #IMPLIED
>

<!ELEMENT header      (skl?,phase-group?,(prop-
group|glossary|reference|note|count-group)*) >

<!ELEMENT skl         (internal-file|external-file) >

<!ELEMENT prop-group  (prop)+ >
<!ATTLIST prop-group
    name              CDATA    #IMPLIED
>

<!ELEMENT prop        (#PCDATA) >
<!ATTLIST prop
    prop-type         CDATA    #REQUIRED
    xml:lang          CDATA    #IMPLIED
>

<!ELEMENT phase-group (phase)+ >

<!ELEMENT phase       (note)* >
<!ATTLIST phase
    phase-name        CDATA    #REQUIRED
    process-name      CDATA    #REQUIRED
    company-name      CDATA    #IMPLIED
    tool              CDATA    #IMPLIED
    date              CDATA    #IMPLIED
    job-id            CDATA    #IMPLIED
    contact-name      CDATA    #IMPLIED
    contact-email     CDATA    #IMPLIED
    contact-phone     CDATA    #IMPLIED
>

<!ELEMENT body        (group|trans-unit|bin-unit)* >

<!ELEMENT trans-unit  (source,target?,(count-group|note|context-
group|prop-group|alt-trans)*) >
<!ATTLIST trans-unit
    id                NMTOKEN    #REQUIRED
    approved          (yes|no)    #IMPLIED
    translate         (yes|no)     "yes"
    reformat          (yes|no)     "yes"
    xml:space         (default|preserve) "default"
    datatype          CDATA        #IMPLIED
    ts                CDATA        #IMPLIED
    restype           CDATA        #IMPLIED
    resname           NMTOKEN      #IMPLIED
    extradata         CDATA        #IMPLIED
```

```
    help-id              NMTOKEN     #IMPLIED
    menu                 CDATA       #IMPLIED
    menu-option          CDATA       #IMPLIED
    menu-name            CDATA       #IMPLIED
    coord                CDATA       #IMPLIED
    font                 CDATA       #IMPLIED
    css-style            CDATA       #IMPLIED
    style                NMTOKEN     #IMPLIED
    exstyle              NMTOKEN     #IMPLIED
    size-unit            CDATA       #IMPLIED
    maxwidth             NMTOKEN     #IMPLIED
    minwidth             NMTOKEN     #IMPLIED
    maxheight            NMTOKEN     #IMPLIED
    minheight            NMTOKEN     #IMPLIED
    maxbytes             NMTOKEN     #IMPLIED
    minbytes             NMTOKEN     #IMPLIED
    charclass            CDATA       #IMPLIED
    phase-name           CDATA       #IMPLIED
>
<!-- size-unit: char|byte|pixel|glyph|dlgunit default='pixel' -->

<!ELEMENT source         (%TextContent;)* >
<!ATTLIST source
    xml:lang             CDATA       #IMPLIED
    ts                   CDATA       #IMPLIED
>
<!-- coord = "x;y;cx;cy"
    font= "fontname[;size[;weight]]"
-->

<!ELEMENT target         (%TextContent;)* >
<!ATTLIST target
    state                NMTOKEN     #IMPLIED
    phase-name           NMTOKEN     #IMPLIED
    xml:lang             CDATA       #IMPLIED
    ts                   CDATA       #IMPLIED
    restype              CDATA       #IMPLIED
    resname              NMTOKEN     #IMPLIED
    coord                CDATA       #IMPLIED
    font                 CDATA       #IMPLIED
    css-style            CDATA       #IMPLIED
    style                NMTOKEN     #IMPLIED
    exstyle              NMTOKEN     #IMPLIED
>

<!ELEMENT alt-trans      (source?,target+,(note|context-group|prop-group)*)
>
<!ATTLIST alt-trans
    match-quality        CDATA       #IMPLIED
    tool                 CDATA       #IMPLIED
    crc                  NMTOKEN     #IMPLIED
    xml:lang             CDATA       #IMPLIED
    origin               CDATA       #IMPLIED
    datatype             CDATA       #IMPLIED
    xml:space            (default|preserve) "default"
    ts                   CDATA       #IMPLIED
    restype              CDATA       #IMPLIED
    resname              NMTOKEN     #IMPLIED
    extradata            CDATA       #IMPLIED
    help-id              NMTOKEN     #IMPLIED
    menu                 CDATA       #IMPLIED
    menu-option          CDATA       #IMPLIED
```

```
    menu-name               CDATA       #IMPLIED
    coord                   CDATA       #IMPLIED
    font                    CDATA       #IMPLIED
    css-style               CDATA       #IMPLIED
    style                   NMTOKEN     #IMPLIED
    exstyle                 NMTOKEN     #IMPLIED
>
```

15.2 In-Line Elements

```
<!ELEMENT g               (%TextContent;)* >
<!ATTLIST g
    id                      CDATA       #REQUIRED
    ctype                   CDATA       #IMPLIED
    clone                   (yes|no)    "yes"
    ts                      CDATA       #IMPLIED
>

<!ELEMENT x               EMPTY >
<!ATTLIST x
    id                      CDATA       #REQUIRED
    ctype                   CDATA       #IMPLIED
    clone                   (yes|no)    "yes"
    ts                      CDATA       #IMPLIED
>

<!ELEMENT bx              EMPTY >
<!ATTLIST bx
    id                      CDATA       #REQUIRED
    rid                     NMTOKEN     #IMPLIED
    ctype                   CDATA       #IMPLIED
    clone                   (yes|no)    "yes"
    ts                      CDATA       #IMPLIED
>

<!ELEMENT ex              EMPTY >
<!ATTLIST ex
    id                      CDATA       #REQUIRED
    rid                     NMTOKEN     #IMPLIED
    ts                      CDATA       #IMPLIED
>

<!ELEMENT ph              (%CodeContent;)* >
<!ATTLIST ph
    id                      CDATA       #REQUIRED
    ctype                   CDATA       #IMPLIED
    ts                      CDATA       #IMPLIED
    crc                     NMTOKEN     #IMPLIED
    assoc                   CDATA       #IMPLIED
>

<!ELEMENT bpt             (%CodeContent;)* >
<!ATTLIST bpt
    id                      CDATA       #REQUIRED
    rid                     NMTOKEN     #IMPLIED
    ctype                   CDATA       #IMPLIED
    ts                      CDATA       #IMPLIED
    crc                     NMTOKEN     #IMPLIED
>
```

```
<!ELEMENT ept           (%CodeContent;)* >
<!ATTLIST ept
   id                   CDATA       #REQUIRED
   rid                  NMTOKEN     #IMPLIED
   ts                   CDATA       #IMPLIED
   crc                  NMTOKEN     #IMPLIED
>

<!ELEMENT it            (%CodeContent;)* >
<!ATTLIST it
   id                   CDATA       #REQUIRED
   pos                  (open|close) #REQUIRED
   rid                  NMTOKEN     #IMPLIED
   ctype                CDATA       #IMPLIED
   ts                   CDATA       #IMPLIED
   crc                  NMTOKEN     #IMPLIED
>

<!ELEMENT mrk           (%TextContent;)* >
<!ATTLIST mrk
   mtype                CDATA       #REQUIRED
   mid                  NMTOKEN     #IMPLIED
   comment              CDATA       #IMPLIED
   ts                   CDATA       #IMPLIED
>

<!ELEMENT sub           (%TextContent;)* >
<!ATTLIST sub
   datatype             CDATA       #IMPLIED
   ctype                CDATA       #IMPLIED
>

<!ENTITY % CodeContent "#PCDATA|sub" >
<!ENTITY % TextContent "#PCDATA|g|bpt|ept|ph|it|mrk|x|bx|ex" >

<!ENTITY lt   "&#38;#60;" >
<!ENTITY amp  "&#38;#38;" >
<!ENTITY gt   "&#62;" >
<!ENTITY apos "&#39;" >
<!ENTITY quot "&#34;" >
```