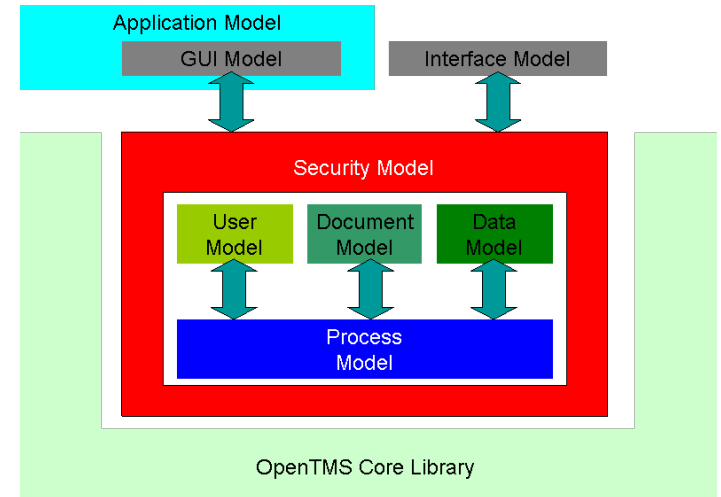




OPEN TMS



Prof. Dr. Klemens Waldhör

Chief Architect

klemens.waldhoer@heartsome.de

The Project

Overview

- Open TMS Goals
- Architecture
- Implementation
- Current Status
- GUIs

The screenshot shows the website for Open TMS (Open Source Translation Memory System). At the top, there are language options for 'Deutsch' and 'English'. The main header features the 'OPEN TMS' logo and the text 'Open Source Translation Memory System'. A navigation menu on the left includes links for 'openTMS', 'Downloads', 'Partner und Sponsoren', 'Ihre Unterstützung', 'Roadmap', and 'Impressum'. The central content area is titled 'Partner und Sponsoren' and lists four companies with their logos, contact information, and icons for a person and a dollar sign. The companies listed are beo, oneword GmbH, eurosript Systems GmbH, and Heartsome Europe GmbH. On the right side, there is a section for 'openSource News' with a list of news items and a 'Weiter' link.

Sprachen Deutsch English

OPEN TMS Open Source Translation Memory System

Startseite

Navigation

- openTMS
- Downloads
- Partner und Sponsoren
- Ihre Unterstützung
- Roadmap
- Impressum

Partner und Sponsoren

beo **beo** Gesellschaft für Sprachen und Technologie mbH
Kupferstraße 36
70565 Stuttgart, Germany
Tel +49 (711) 166 46-0
Fax +49 (711) 166 46-50
<http://www.beo-doc.de>

oneword **oneword** GmbH
Otto-Lilienthal-Str. 36
71034 Böblingen, Germany
Tel +49 (7031) 7149550
Fax +49 (7031) 7149565
<http://www.oneword.de>

euroscript **euroscript** Systems GmbH
Schedestr. 1-3
53113 Bonn, Germany
Tel +49 (228) 36943 0
Fax +49 (228) 36943 22
<http://www.euroscript.de>

Heartsome Europe GmbH **Heartsome Europe GmbH**
Friedrichstr. 17
90574 Roßtal, Germany
Tel +49 (9127) 57 90 01
Fax +49 (9127) 95 11 78
<http://www.heartsome.de>

]project-open[

openSource News

- Münchner Schulnetz wird auf Windows XP umgestellt
- Google zum Android-Versionswirrwarr
- Rudimentärer Open-Source-Treiber für Intels GMA500/Poulsbo

Weiter

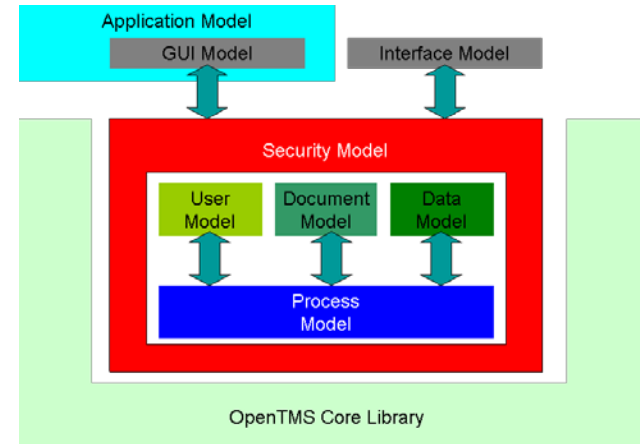
openTMS Sponsors



beo Logisch!



Overview



Free and Open Source Translation
Memory System

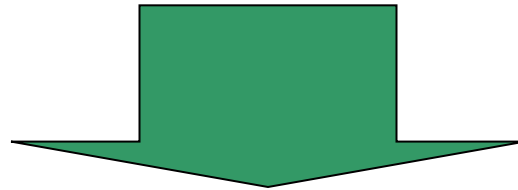
Goals

Three translation memory systems for one and the same process?
Software investments that make translation costs shoot through the roof?
Exchange formats that put the brakes on productivity?
FOLT (Forum Open Language Tools) is concerned with the entire process of producing multilingual documentation. From the creation of the source text to production in foreign languages, we analyze our processes for weaknesses and a lack of standardisation.

Primary objectives:

- Sharing experiences of processes using standard industry software
- Sharing experiences of the use of Open Source software
- Standardisation of interchange formats
- Testing new Open Source technologies and improving existing technologies in the translation market
- Public support for non-proprietary software and software development
- Publication of aims and results

www.folt.de

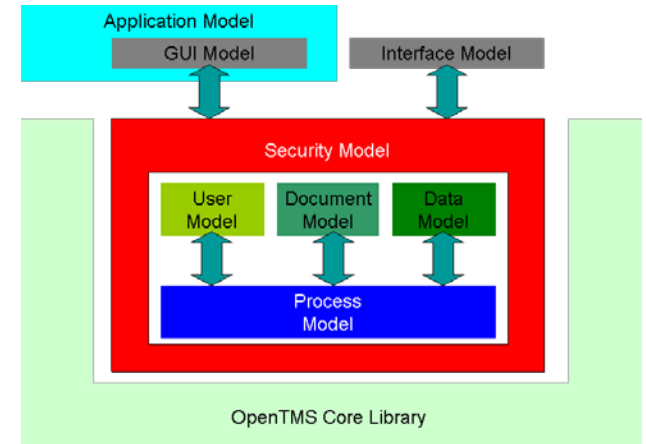


- Development of the OpenSource Translation Memory system **OpenTMS**

OpenTMS Requirements

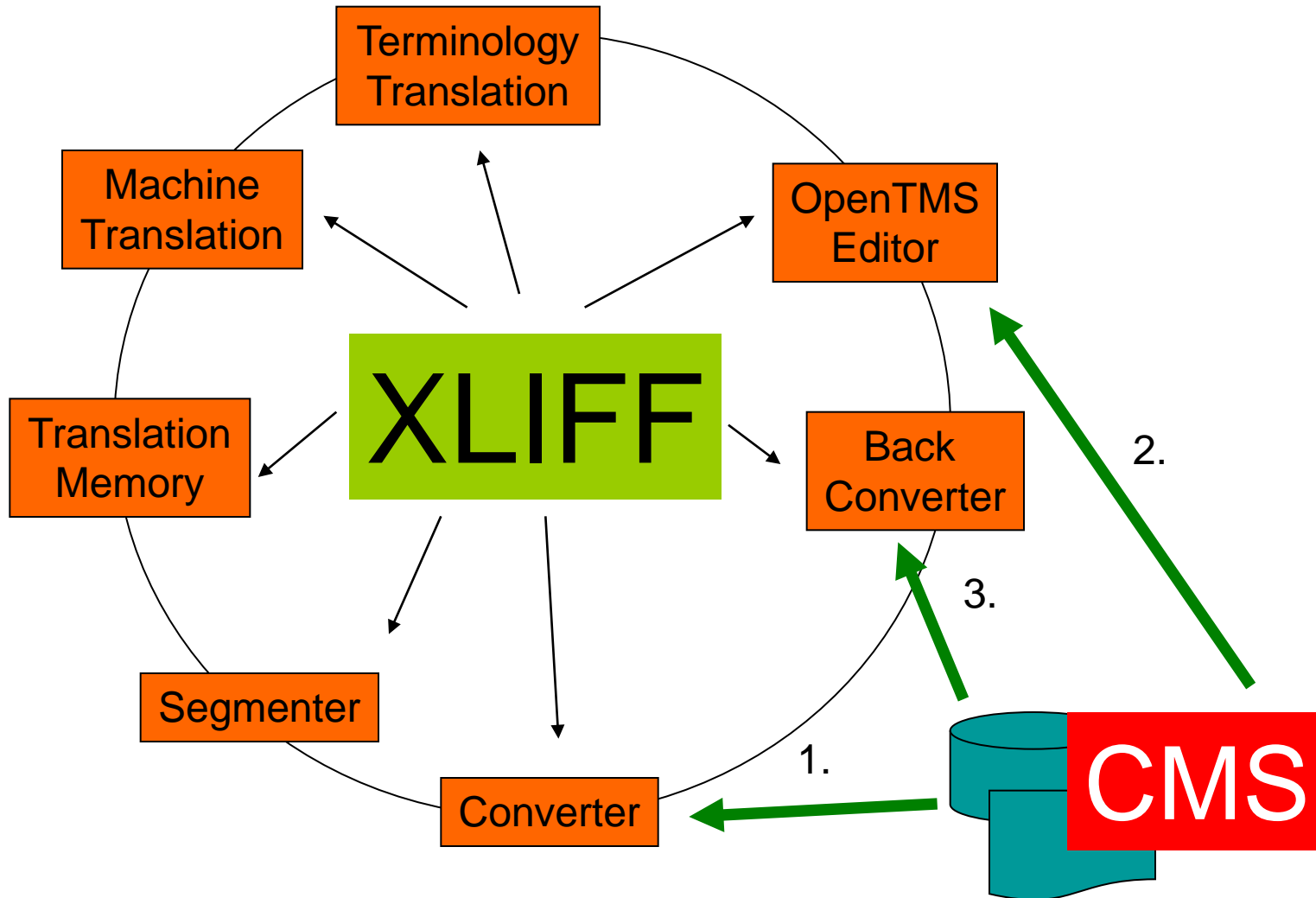
- Architecture
 - Server / Client Architecture
 - Modular software approach
 - Web based
 - Thin client
- Language Modelling
 - High level modelling
 - Extendibility
 - Language independent
 - Standards based
- Data Sources
 - Abstract data source approach
 - Integrated default fuzzy search
 - Integrated translate method
 - Support of all SQL database systems
 - Support of all types of data sources like csv, noSQL, TMX, XLIFF files
 - MT support through databases
- Scalability
 - Single and multi user requirement
 - High performance databases
- Interfaces
 - Interface Layer
 - Integration into CMS
 - Workflow support
- Java
 - Java Coding Standards
 - Java Documentation Standard
 - Delivered as a set of jar files
- Eclipse
 - Development base
 - swt for GUIs
- OS independent
 - Windows, Linux, Mac ...
- Hardware independent

Architecture



The core of openTMS

Example Work Flow




Architecture based on Standards

- XLIFF
- TMX
- TBX
- SRX
- ...

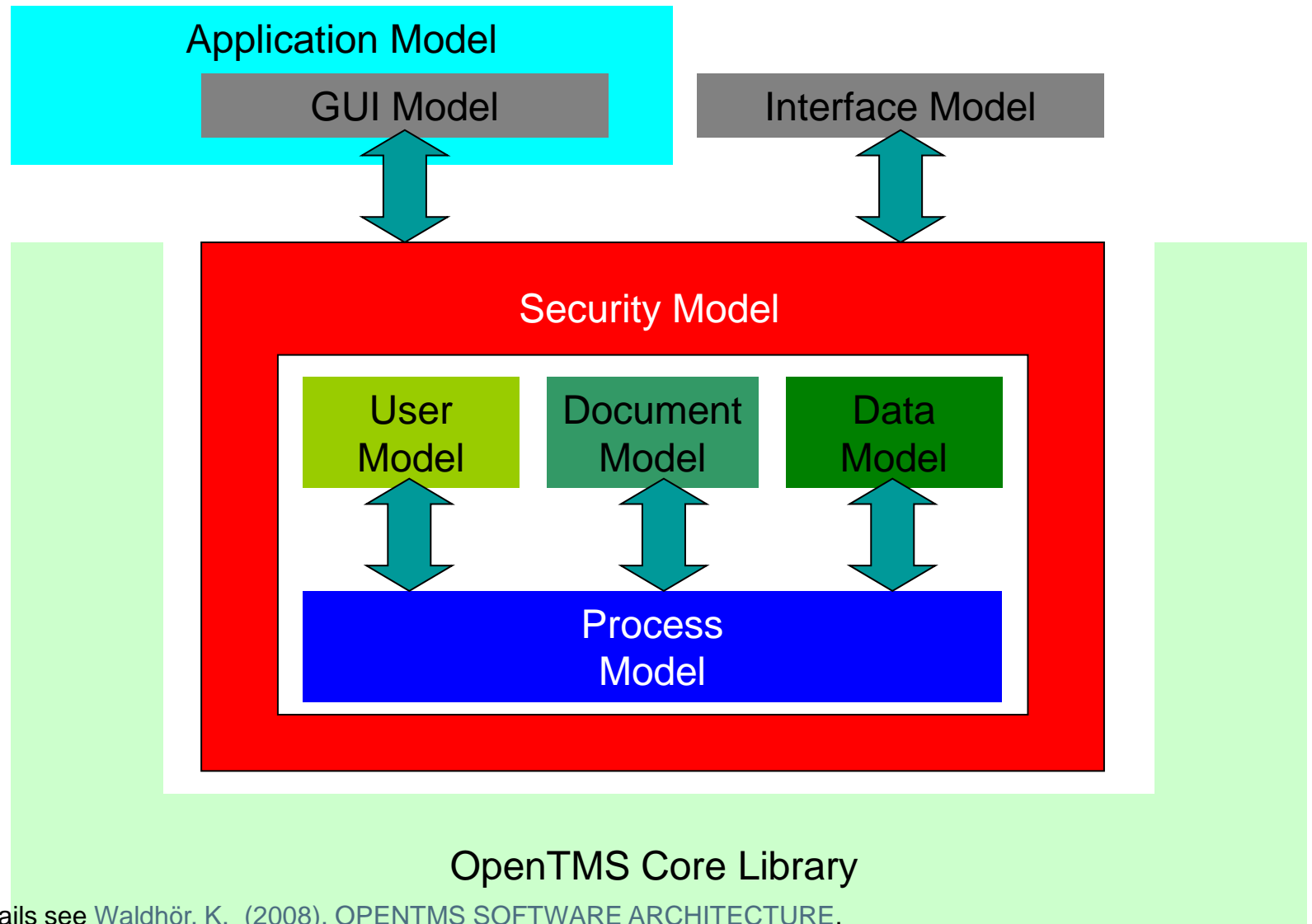
```
<trans-unit id="586">
  <source xml:lang="en">Please wait</source>
  <target xml:lang="de">Bitte warten</target>
</trans-unit>

<tu creationid="4314" changeid="XXXXXX">
  <tuv xml:lang="cz"><seg>Cekej</seg></tuv>
  <tuv xml:lang="de"><seg>Bitte warten</seg></tuv>
  <tuv xml:lang="en"><seg>Please wait</seg></tuv>
  <tuv xml:lang="pl"><seg>Czekaj...</seg></tuv>
  <tuv xml:lang="ru"><seg>ЖДИТЕ</seg></tuv>
</tu>

<termEntry id="7399880175948528441">
  <langSet xml:lang="de">
    <tig>
      <term>Bitte warten</term>
      <termNote type="changeDate">2007-03-06 14:10:59</termNote>
      <termNote type="changeAuthor">araya</termNote>
    </tig>
  </langSet>
  <langSet xml:lang="en">
    <tig>
      <term>Please wait</term>
    </tig>
  </langSet>
</termEntry>
```

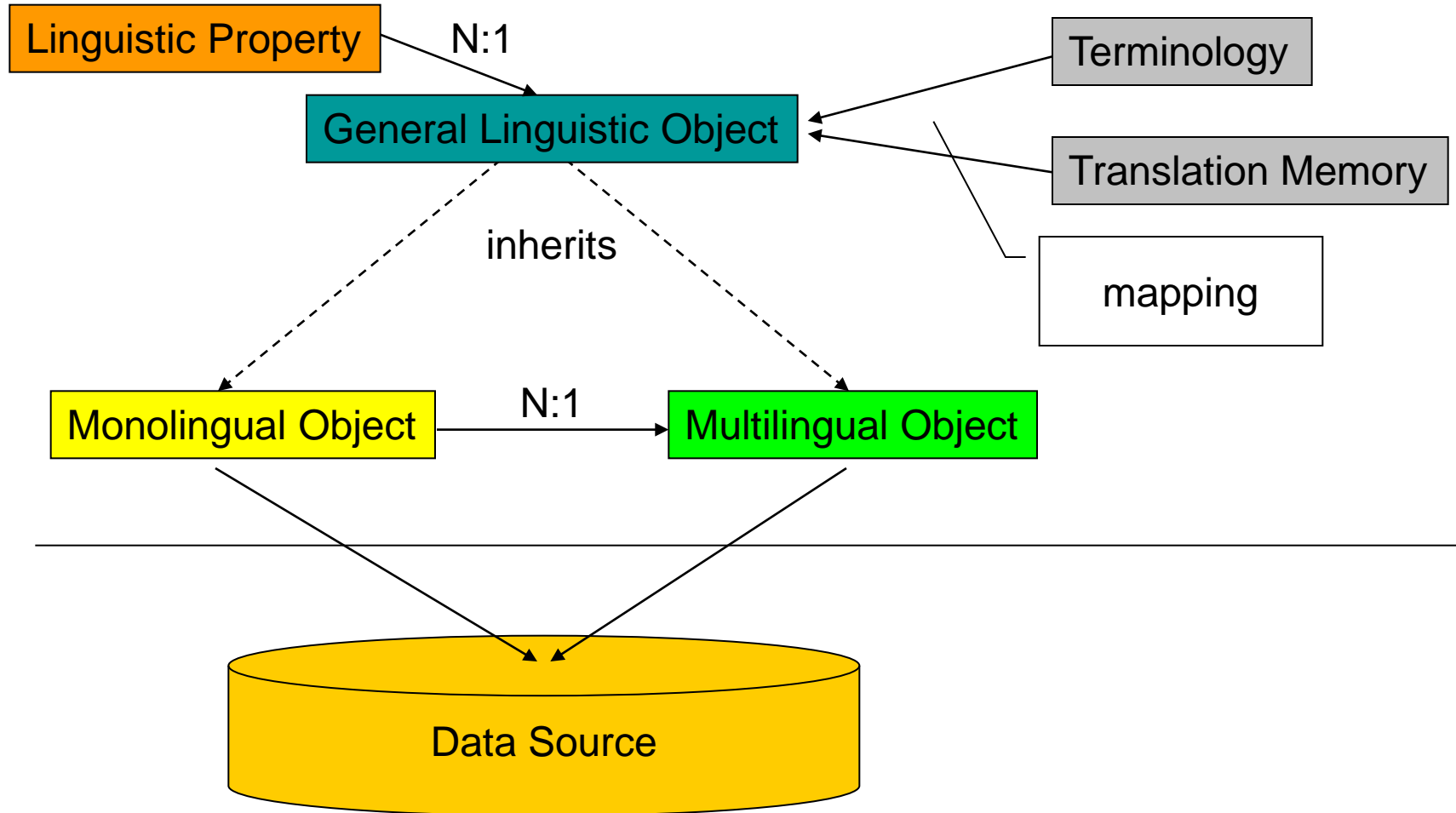
In general  XML

OpenTMS System Architecture

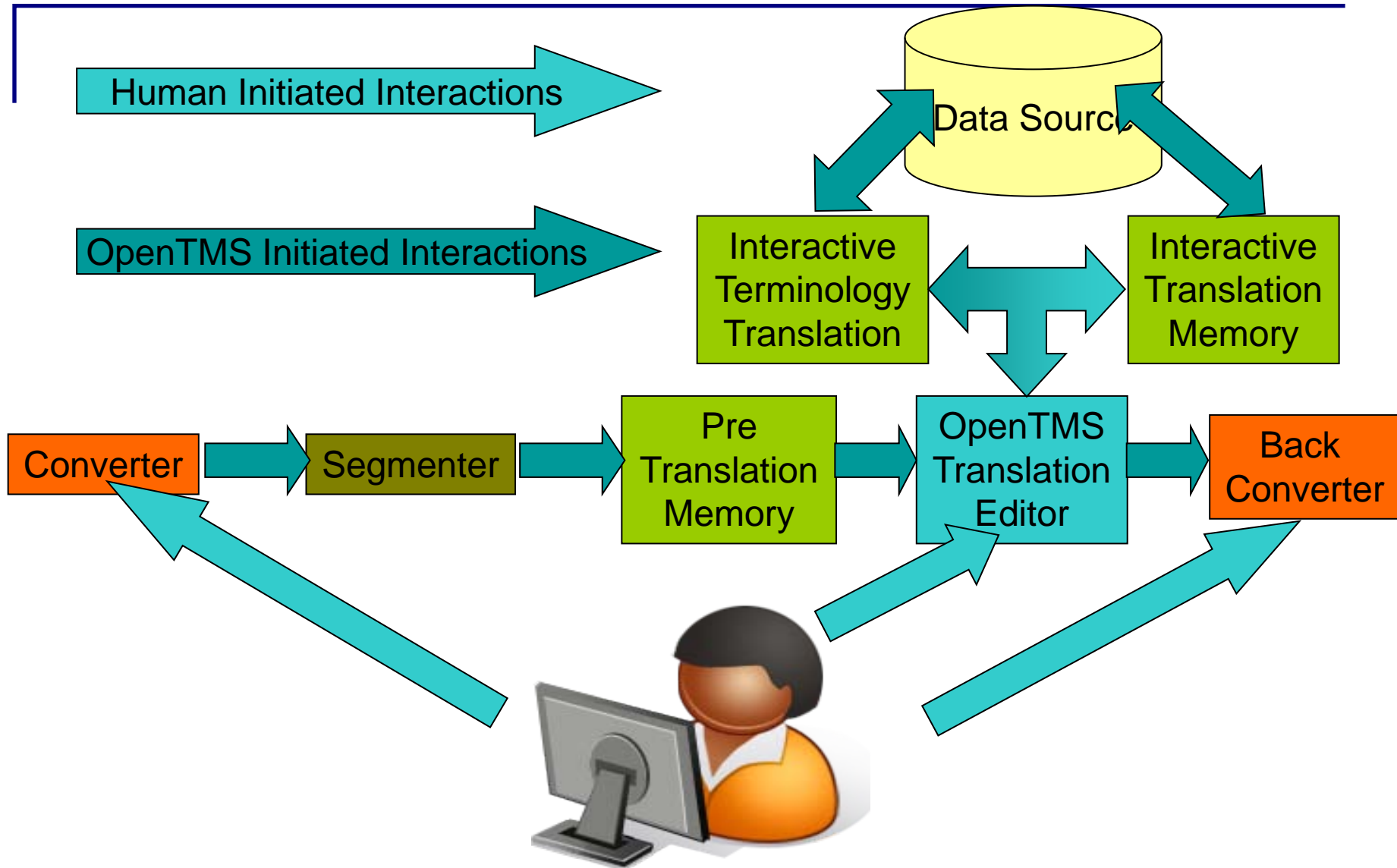


For details see [Waldhör, K. \(2008\). OPENTMS SOFTWARE ARCHITECTURE.](#)

Modelling Language



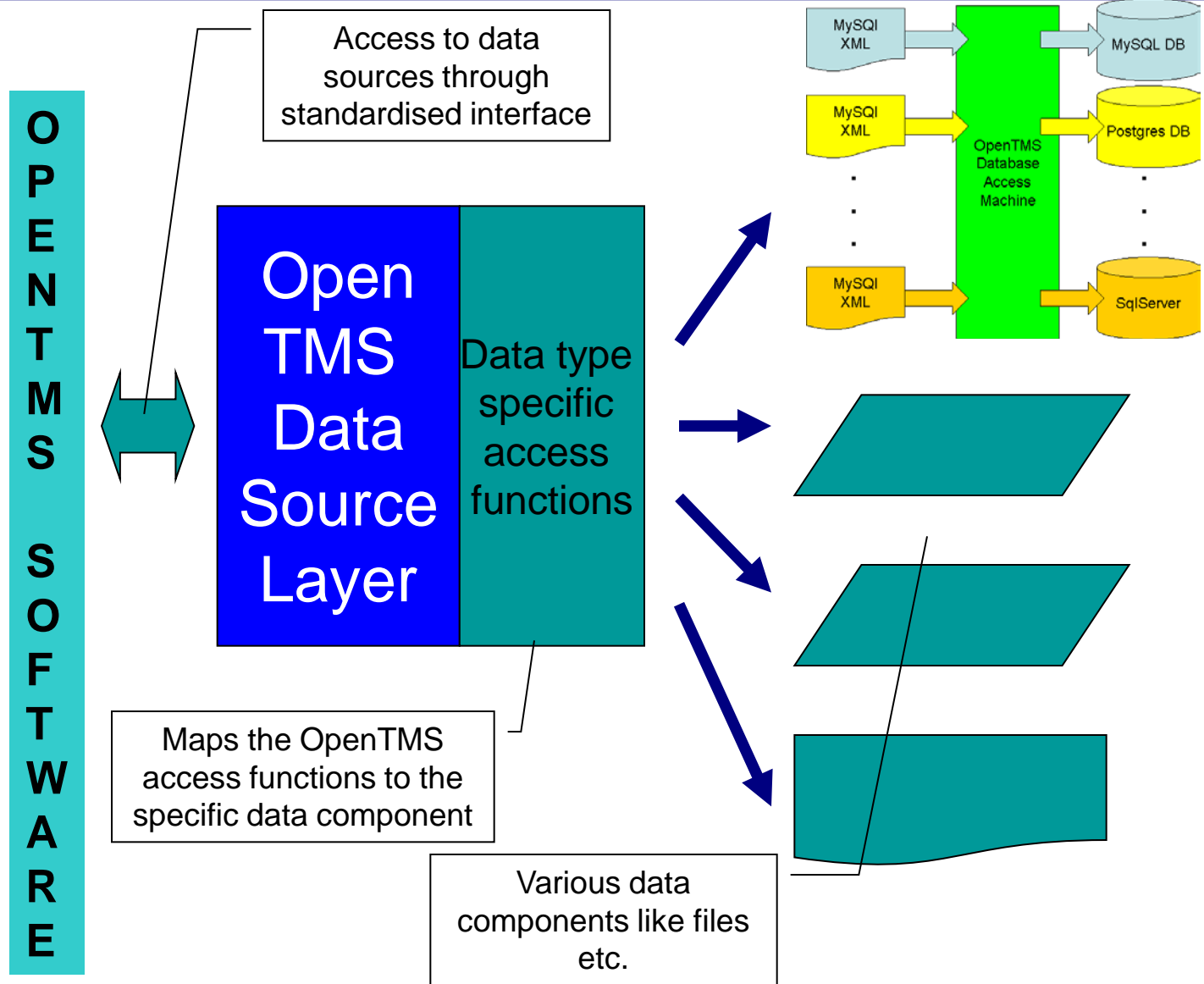
OpenTMS Processes



Data Sources

- Language related data are represented as “data sources”
- Idea
 - Makes the data access interface independent from the data itself
- Not being restricted to SQL databases only
 - Also flat data or xml files
 - TMX, XLIFF files as a data source
 - ...
 - Machine translation (MT) as data source
 - Spread sheets
 - E.g. Excel as terminology lists
 - Object Oriented Databases
 - DMS systems
 - “Web Sites” (http based interfaces)
- Define a common interface for all access functions
 - Allows adaption to individual data source properties
 - e.g. read only data sources like MT

Data Sources



Data Sources

Data Source Status

- Data Source methods defined
 - Are extended depending on needs and requirements
- SQL
 - Access optimisation
 - Hibernate based
- Other OpenSource databases...
 - OODBS
 - DB4O partially implemented for testing purposes
- Other data sources
 - TMX files
 - XLIFF files
 - MT
 - Google & Microsoft Translator

Core Methods

- Create
- Delete
- Import TMX, XLIFF File
- Export TMX, XLIFF File
- Copy between data sources
- Translate
- ...
- Data Source described thru xml based configuration files
- XML-RPC Access
- Servlet Engine Access

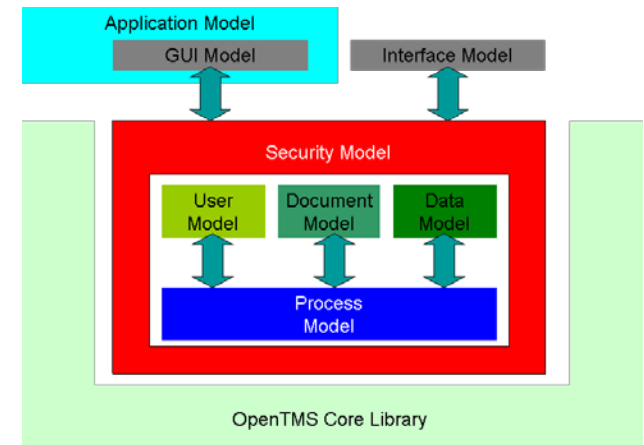
Fuzzy Search – Core Function of TM

- Step 1: Search in KD-TREE
 - Restricts the number of segments to search
 - Finds possible matching (source) segments
- Step 2: Levenshtein Similarity
 - Compare matches from step 1 computing the real similarity
- Step 3: Get source and target MOLs / MUL
 - Create translation (alt-trans)

Translation Core Functions

- Convert (to and from XLIFF)
 - Currently using Araya converters
www.heartsome.de
 - Complex document format like WinWord etc. thru Open Office Converters
- Segmentation
 - Currently thru Araya
- Translate

User Interfaces



Editors and more

Ubuntu VM Distribution

The image shows the Ubuntu desktop environment with the openTMS application menu. The menu items are:

- System postscript (info.txt)
- openTMS
- Start openTMS Server
- Stop openTMS Server
- Terminal

The screenshot shows the openTMS web application running in a Mozilla Firefox browser within a Sun VirtualBox VM. The browser address bar shows `http://localhost/openTMS/OpenTMS.php?`. The application interface includes a navigation menu on the left and a main content area with the following table:

OPTIONEN		PROJEKTDETAILS	
Angemeldet: opentms opentms		Projekt: test Ausgangsdatei: sample.doc	
Willkommen	Quelldatei	<input type="text"/>	<input type="button" value="Speichern"/> <input type="button" value="Browse..."/>
Projekte	Vorbereitung	<input type="text"/>	<input type="button" value="Speichern"/> <input type="button" value="Browse..."/>
Sprachen	Übersetzung		Download
Translation Memory	Zieldatei		<input type="button" value="Speichern"/> <input type="button" value="Browse..."/>
Load XLIFF Editor			Nicht vorhanden
Abmelden			

The application also shows a status bar at the bottom with the text "Done".

Data Source Editor

The screenshot shows the OpenTMS Data Source Editor interface. A yellow callout box labeled "Edit MOL/MOL Properties" points to the "Property Editor" dialog box. Another yellow callout box labeled "Language Specific Segments" points to the "All Ids (2)" list in the main window. A third yellow callout box labeled "Search Functions" points to the search bar and "FUZZY" dropdown in the bottom toolbar. A fourth yellow callout box labeled "Delete & Save Functions" points to the delete and save icons in the bottom toolbar.

Property Editor

ID	Content	Language	O Encoding	Property T...	Type
15	Paegasus			PROP	xtuv-Project
18	hallo	af Afrikaans	EUC-JP	PROP	37ab7f9c-a76a-44c...
22	mynote tuv 1	en	UTF-8	NOTE	
16	keines 3	ar-bh Arabic (Bahrain)		PROP	4f070096-88e1-4a6...
11	19970212T153400Z			CORE	creationdate
14	mynewproperty	ab Abkhazian		PROP	77a779fa-4620-43d...
17	Computing TUV			PROP	xtuv-Domain
20	ffff		Big5-HKSCS	PROP	8c49b402-20b7-40...
12	Klemens Waldhör			CORE	creationid
13	this is a tuv	en		PROP	ytuv-Project
19	dddd	ab Abkhazian		PROP	7f3c9a6a-027a-4b4...
21	mynote tuv			NOTE	

Beo Logisch! docliff Editor

The screenshot shows the docliff Editor 0.4.0 interface. The main window displays a document with XML tags and text. A dialog box titled "Choose Data Source" is open, listing various data sources. The document content includes:


- <sn 105> Radblock-System
- <tn 105>
- <sn 106> Sicherheit
- <tn 106>
- <sn 107> Symbolerklärung
- <tn 107>
- <sn 108> folgende Symbole oder Hinweise warnen vor möglichen Personen- oder Sachschäden
- <tn 108>
- <sn 109> Warnung vor einer Gefahrenstelle
- <tn 109>
- <sn 110> dieses Symbol finden Sie in der Betriebsanleitung bei allen Hinweisen zur Arbeitssicherheit
- <tn 110>
- <sn 111> beachten Sie diese Hinweise stets und verhalten Sie sich besonders aufmerksam und
- <tn 111>
- <sn 112> geben Sie die Sicherheitshinweise an alle Personen weiter, die mit Arbeiten an dem Produkt
- <tn 112>
- <sn 113> beachten Sie darüber hinaus auch stets die allgemeingültigen Sicherheitsvorschriften
- <tn 113>
- <sn 114> Warnung vor gefährlicher elektrischer Spannung
- <tn 114>
- <sn 115> Eine Berührung von unter Spannung stehenden Teilen (z. B. bei angebauten Motoren) vorübergehender Abschaltung der zutreffenden Betriebsspannung (Einspeise-, Betriebs- oder Fremdeinspeisung)
- <tn 115>
- <sn 116> Betriebssicherheit der Anlage in Gefahr
- <tn 116>
- <sn 117> dieses Symbol steht in der Betriebsanleitung bei allen Hinweisen, bei deren Nichtbeachtung Schäden an dem Produkt, an der Anlage oder dem zu transportierenden Gut verursacht werden können
- <tn 117>
- <sn 118> Lagerung
- <tn 118>
- <sn 119> die Radblöcke sind in trockenen Räumen, in denen keine starken Temperaturschwankungen auftreten, z. B. auf einem Holzunterbau, zu lagern
- <tn 119>
- <sn 120> Das Übereinanderstapeln von unverpackten Radblöcken ist nichtzulässig!
- <tn 120>
- <sn 121> Hinweis: Anstrich nicht beschädigen!
- <tn 121>


The "Choose Data Source" dialog box lists the following data sources:


- C:\arayal\test\opentms\tekomp2009\sample.tmx (de.folt.models.datamodel.tmxfile.TmxFileDataSource)
- C:\temp\sample.tmx (de.folt.models.datamodel.tmxfile.TmxFileDataSource)
- Google Translate (de.folt.models.datamodel.googletranslate.GoogleTranslate)
- Microsoft Translate (de.folt.models.datamodel.microsofttranslate.MicrosoftTranslate)
- aopentms (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- fuzzycsv (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- postl (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- postcsv (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- tckmine (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- tcktext (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- tcktextL (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- test (de.folt.models.datamodel.multipledatasource.MultipleDataSource)
- testms (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- testbox (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)
- transworkopentms (de.folt.models.datamodel.sql.OpenTMSSQLDataSource)

The bottom of the editor shows a toolbar with various icons, a status bar with "Go to: 0", "Seg.: 120/122 (5/5) de-en", and "Match Quality: MT #: 1/1 1/1". The bottom right corner displays a message: "The stacking of unwrapped wheel blocks is not allowed!".

Web Xliff Editor







Demonstrator eines webbasierten Xliff-Editors, der die Möglichkeiten der Integration von Werkzeugen wie Translation Memories, Terminologie, maschineller Übersetzung und Anwendung von statistischen Verfahren bei Übersetzungswerkzeugen im Web-Umfeld evaluiert.

sample2.odt.xlf.1298446894254
Goto Segment

Xliff File:
Durchsuchen_
Load Xliff File
Reset

TM Datasources: C:\araya\test\opentms\tekom2009\sample.tmx
Similarity

Term Datasources: C:\araya\test\opentms\tekom2009\sample.tmx
100
de
->
en

#3 2	41830244.EPS	
#4 3	Vorwort	
#5 4	Sie haben ein Produkt der Mustermann AG erworben, das nach den neusten Erkenntnissen der Technik hergestellt worden ist.	
#6 5	Mit dieser Betriebsanleitung möchten wir dem Benutzer zweckdienliche Anweisungen zum sicheren und sachgerechten Arbeiten geben sowie eine leichte Instandhaltung ermöglichen.	
#7 6	Jede Person, die mit Transport, Aufstellung, Inbetriebnahme, Bedienung, Wartung und Reparatur unseres Radblocksystems XXX und dessen Zusatzeinrichtungen beauftragt ist, muß gelesen und verstanden haben:	
#8 7	die Betriebsanleitung	

#	%	Source Segment	Target Segments
1	87	Sie haben ein Produkt von der Mannesmann Dematic AG erworben, das nach den neusten Erkenntnissen der Technik hergestellt worden ist.	1: You have purchased a Mannesmann Dematic product. This equipment was manufactured in accordance with state-of-the-art engineering principles.

© NetVenture Europe & Co KG
 Prototyp: Version 1.0
 Log File: C:\Program Files (x86)\Apache Software Foundation\Tomcat 6.0\HEARTSS.232201184013809.log

Integration Araya XLIFF Editor

ARAYA Translation Editor 15.19 [Edit Mode / Translations Accepted] - C:\araya\test\TekomDemo\docs\sample.doc.xlf

File Edit Projects Go to Options Tasks Terminology Database Review Plugins OpenTMS Plugin Help

Translate XLIFF File using OpenTMS Data Source

- Create OpenTMS Database
- Delete OpenTMS Database
- Import OpenTMS Database
- Export OpenTMS Database
- OpenTMS Website

de->en Current User: klemens

Document Segments

die Sicherheitsvorschriften

Seg.: 9/122 Go to: 9

the security instructions

New Words: 2

Accept Translation Target: 1/1

/ ID: 120d189a1d72_klemens / internalArayaTM

Number of Segments: 122 / Translated Segments: 1 / Approved Segments: 0 / Untranslated Segments: 121 / T: 133234688 F: 123274008 M: 1248657408

Translate XLIFF File using OpenTMS Data Source

Source- und Target Documents

Source file C:\araya\test\TekomDemo\docs\sample.htm Browse...

XLIFF file C:\araya\test\TekomDemo\docs\sample.htm.xlf Browse...

Document Format Information

Determine Document Format 6: html

Source File Properties

Language: de German (Standard) Code Page UTF-8

Conversion options:

Segmentation by Element Segmentation by CR/LF

Number of CR/LF 1 Set Document Properties

Do not resolve Entities

Entities not to resolve File Browse...

Translation Options

TMX/XLIFF File for translating the document Browse...

OpenTMS Databases mpentms

Target Language en English

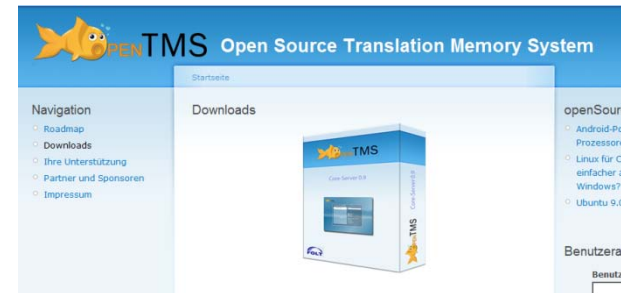
Minimum Match Percentage 70 Maximum Match Count 5

Open File After Conversion Use Translate Attribute

Convert File to XLIFF Format Close

Downloads

- <http://sourceforge.net/projects/open-tms>
- Ubuntu Version
- Windows Version
 - www.heartsome.de/arayatest/opentmsserver.exe
 - Includes data source editor
 - Web GUI and docliff editor
 - Send E-Mail to klemens.waldhoer@heartsome.de
 - As part of the Araya Xliff Editor:
www.heartsome.de/arayatest/araya-freeversion.exe



Contact

Heartsome Europe GmbH
Friedrichstr. 17
D-90574 Roßtal

www.heartsome.de



Prof. Dr. Klemens Waldhör



T: +49 9127 579001

F: +49 9127 951178

klemens.waldhoer@heartsome.de