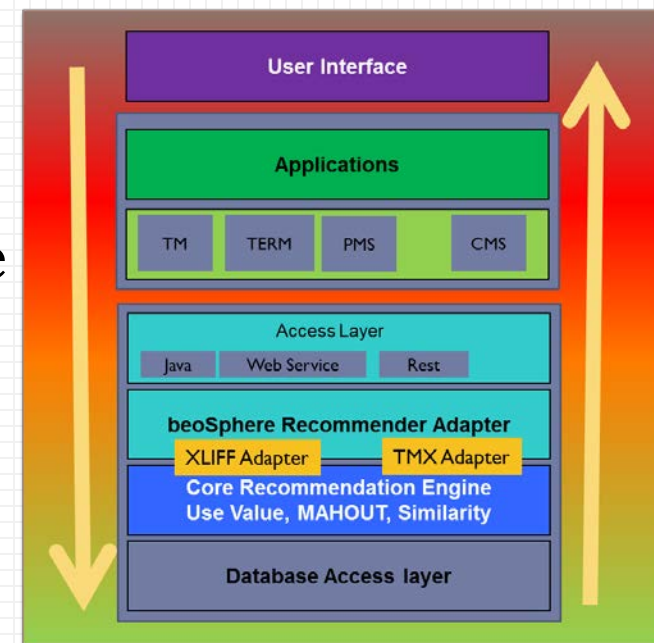


# Recommender Systems as part of Localization Project Management with XLIFF

Prof. Dr. Klemens Waldhör  
Heartsome Europe GmbH &  
FOM University of Applied Science

FEISGILTT 2013  
11 June 2013  
London, GB



# Presentation overview

---

<b>Segment</b>	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names are property of their rightful owners / copyright.	All trademarks used, as well as trade and company names are the property of their rightful owners/authors.	All used brands as well as trade names and company names are a property of her lawful owners / originators.
<b>Rank</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>4</b>	<b>5</b>	<b>3</b>

- ✓ Why Recommendation for translation?
- ✓ Some Mathematics and Model
  - ✓ The Specificity Index Model
  - ✓ Recommendation implementation
- ✓ Examples
  - ✓ Integration in TMX and XLIFF
- ✓ Summary

# Why recommendation?

---

- ▶ Recommendation is a proven technology in many domains
  - ▶ Books - Amazon, travel, car buying, ...
- ▶ Recommendation as a tool for
  - ▶ Deciding between multiple options
    - ▶ Multi dimensionality
  - ▶ Establishing rank orders
  - ▶ Quality control
- ▶ Recommendation gives rationality on a decision
  - ▶ Reflection of user preferences
    - ▶ Some more “objectives measures” (?)
  - ▶ Explanation capability
  - ▶ Decision support

# Application areas in localization

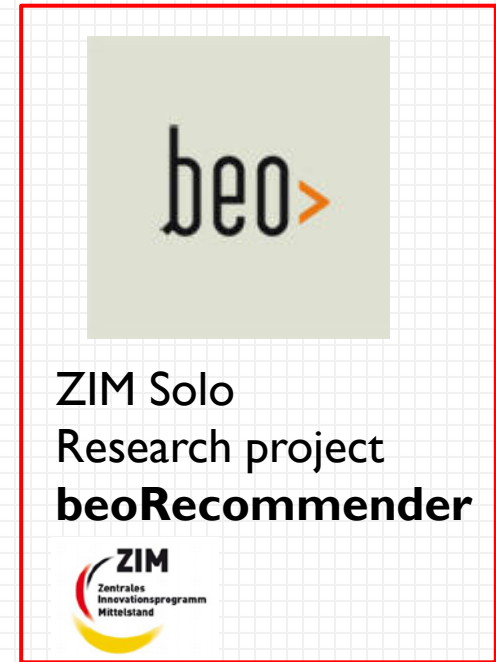
---

- ▶ **Translation Project management**
  - ▶ Selection of translators
  - ▶ Distribution of translators across projects
  - ▶ Tool recommendation based
- ▶ **Translation process**
  - ▶ Selection of TMs, Terminology databases
- ▶ **TM**
  - ▶ Selection and ordering of TM matches
- ▶ **Terminology**
  - ▶ Selection and ordering of term matches
- ▶ ...

# Application areas in localization

---

- ▶ Translation Project management
  - ▶ Selection of translators
  - ▶ Distribution of translators across projects
  - ▶ Tool recommendation based
- ▶ Translation process
  - ▶ Selection of TMs, Terminology databases
- ▶ TM
  - ▶ Selection and ordering of TM matches
- ▶ Terminology
  - ▶ Selection and ordering of term matches
- ▶ ...



# Why recommendation in translation tools?

---

- ▶ **Huge TM databases**
  - ▶ E.g. TAUS TM
  - ▶ “Giga” TM memories
  - ▶ Many matching results possible depending on used fuzzy match quality
    - ▶ How should the project manager / translator decide which ones to choose?
- ▶ **Define selection criteria based on**
  - ▶ Similarity
  - ▶ Multiple attributes
    - ▶ Date, domain, number of translations, directionality...
  - ▶ Customer preferences

# The basic model

- ▶ **Given**
  - ▶ Set of entries
    - ▶ e.g. TM matches, translators, projects, ...
  - ▶ Set of attributes
    - ▶ Attribute classes
      - E.g. origin of translation
  - ▶ Weighting of attributes and classes
- ▶ **Apply a recommendation model**
  - ▶ Recommender models
    - ▶ Content based filtering
    - ▶ Collaborative filtering
  - ▶ Use value analysis

**Use Value Analysis**

	A	B	C	D	E	F	G	H
<b>Translator</b>			Translator A		Translator B		Translator C	
<b>Criteria/Weight</b>			reasonable		expensive		average	
Word cost	25%	3	0,75	1	0,25	2	0,5	
Quality	40%	1	0,4	3	1,2	2	0,8	
Delivery date adherence	25%	1	0,25	2	0,5	3	0,75	
Availability	10%	3	0,3	3	0,3	1	0,1	
$\Sigma$ /Gesamtnutzwert	100%		1,7		2,25		2,15	
		1 bad	3		1		2	
		2 neutral						
		3 very good						



# Recommendation Technology

- ▶ Collaborative Filtering
  - ▶ Recommendations based **behavior of similar users**
- ▶ Content Based Filtering
  - ▶ Recommendations based **properties of similar items**
- ▶ Use Value Analysis
  - ▶ Simple method for **evaluating alternatives** based on attributes, criteria and weights for alternatives

**Übersetzung für Kunden mit unterschiedlichen Dokumentformaten**

Kunde vs Format	Word	PDF	Excel	HTML	XML
A GmbH		5 (3)	4 (2)		7 (2)
B AG	1	6 (3)	3 (2)		
...					
C GmbH		5 (4)	5 (1)		5 (1)
...					
K Inc	1			1	

z.B. Zellenwerte zeigen, dass eine Bewertung für das Format vorgenommen wurde  
 Werte in () stellen die einzelnen Schritte im Empfehlungsalgorithmus dar  
 Fragestellung: Wie hoch ist die Relevanz des Items i (z.B. pdf) für den Kunden C

- Firma für die Empfehlung gesucht wird und für bestimmte Items Beurteilungen abgegeben hat
- Ähnliche Firmenelemente auf den durchgeführten Beurteilungen der Items (2)
- Bewertungen eines Items durch ähnliche Firmen (3)
- Mögliche Empfehlung für Firma C für Item (4)

**Übersetzung für Kunden mit unterschiedlichen Dokumentformaten**

Kunde vs Format	Word	PDF	Excel	HTML	XML
A GmbH		5 (3)	4 (2)		7 (2)
B AG	1	6 (3)	3 (2)		
...					
C GmbH		5 (4)	5 (1)		5 (1)
...					
K Inc	1			1	

z.B. Zellenwerte zeigen, dass eine Bewertung für das Format vorgenommen wurde  
 Werte in () stellen die einzelnen Schritte im Empfehlungsalgorithmus dar  
 Fragestellung: Wie hoch ist die Relevanz des Items i (z.B. pdf) für den Kunden C

- Firma für die Empfehlung gesucht wird und für bestimmte Items Beurteilungen abgegeben hat (1)
- Ähnliche Empfehlungselemente auf den durchgeführten Beurteilungen der Items (2)
- Items, die von den ähnlichen Firmen (hoch) wertet wurden (3)
- Mögliche Empfehlung für Firma C für Item (4)

$$R(I_y, U_x) = \frac{\sum_{j=1..n} S(I_y, I_j) * R(I_j, U_x)}{\sum_{j=1..n} S(I_y, I_j)}$$





# Some mathematics for similarity / distance measures

## ► Similarity measures

Measure	Formula	Type	Comment
Cosinus	$\frac{\sum_{i=1..n} v_i * w_i}{\sqrt{\sum_{i=1}^n v_i^2} * \sqrt{\sum_{i=1}^n w_i^2}}$	Sim	The cosine measure calculates the angle between two vectors. Range -1 .. 1
Pearson correlation Coefficient	$\frac{\sum_{i=1..n} ((v_i - \bar{v}) * (w_i - \bar{w}))}{\sqrt{\sum_{i=1..n} (v_i - \bar{v})^2} * \sqrt{\sum_{i=1..n} (w_i - \bar{w})^2}}$	Sim	Here, the attribute values are corrected by the mean values (Statistics) Range -1 .. 1
Overlap- Coefficient	$\frac{\sum_{i=1..n} \min(v_i, w_i)}{\min(\sum_{i=1}^n v_i, \sum_{i=1}^n w_i)}$	Sim	Measure of the mutual agreement
Dice- Coefficient	$\frac{2 * \sum_{i=1..n} v_i * w_i}{\sum_{i=1}^n v_i^2 + \sum_{i=1}^n w_i^2}$	Sim	Correspondence of the same elements
Jaccard-Coefficient	$\frac{\sum_{i=1..n} (v_i * w_i)}{\sum_{i=1}^n v_i + \sum_{i=1}^n w_i - \sum_{i=1..n} (v_i * w_i)}$	Sim	Range 0 .. 1
Euclidian Distance	$\sqrt{\sum_{i=1}^n (v_i - w_i)^2}$	Dist	Distance between vectors

# Attributes of a match for recommendation

Term	Abbreviation	Comment
Source segment	SS	The sentence to be translated, the source-language segment
Target Segment	TS	Translation
Source language	SL	Source language of the document
Target language	TL	Language of the translation
Results 1 .. n	$M_i$	Match for the source segment SS
Output language segment of the hit ( $i = 1..K$ )	$M_i (SS)$	The $i_{th}$ matching translation of the k hits for the source segment
Results matching quality	$Q_i$	The match quality 0 ... 100
j is target-language segment of the hit ( $i = 1..meters$ )	$M_i (TS_j)$	A source segment hit may have several 1..m translations for the target language
Number of translations of the hit i	$N_i$	Number of translations for the hit
Number of non-target-language translations of the source segment	$M_i$	Number of translations to $MA_i(SS)$ for all languages. An output segment may also include translations into other languages.
Number of possible translations back to $MA_{i(j)TS, TL}$ in the source language	$R_{ij}$	Each target language segment has at least one output segment (otherwise it would be no hit). R now counts all segments for a target segment segments into the source language



# The Translation Specificity Index Model

The Translation Specificity Index  $TMSP_{ij}$  for a match  $i$  and translation  $j$

$$TMSP_{ij} = Q_n \cdot \frac{1}{N_i} + Q_M \cdot \left(1 - \frac{1}{M_i}\right) + Q_R \cdot \frac{1}{R_{ij}}$$

and

with  $Q_n$  as weight  $0 \leq Q_n \leq 1$ ,

$Q_M$  as weight  $0 \leq Q_M \leq 1$ ,

$Q_R$  as weight  $0 \leq Q_R \leq 1$

and  $Q_n + Q_M + Q_R = 1$

The Translation Specificity Index  $TMSP_{ij}$  for a match  $i$

$$TMSP_i = \sum_{j=0}^m TMSP_{ij}$$

The Translation Specificity Index  $TMSP_{ij}$  for a segment:

$$TMSP = \sum_{i=0}^n TMSP_i$$

# Some Attributes of a match useful for recommendation

---

<b>Attribute Name</b> ↓	<b>Class</b>	<b>Comment</b>
changenumber	ADMIN	Number of changes in the translation during the lifetime of the segment
missing attributes	ADMIN	The higher the number of missing attribute values for segment, the worse the entry will be judged.
database	ADMIN	Database name of the match
document	ADMIN	Uniform naming convention required for the document where the translation comes from
document type	ADMIN	Document type
domain	ADMIN	Domain of the segment
keyword	ADMIN	Multiple keywords associated with the match
productkey	ADMIN	Product key
application	ADMIN	Application
business area	ADMIN	Business area
creation date	ADMIN	Days (in terms of age) or any data format
change date	ADMIN	Last change date of the match
project	ADMIN	Project Reference
usage number	ADMIN	How often was the translation used?



# Attributes of a match for recommendation

---

Attribute Name <sup>1</sup>	Class	Comment
translationNumber	LING	How many translations for a source segment? Relevant for fuzzy matches
source target ratio	LING	Comparing source and target length
identicalTranslations	LING	How many translation are identical with this translation, that is how many source segments link to this translation
translationLenght	LING	Short / medium / long? Words better than number of characters?
different formats	LING	Difference in format between source and target segment
style	LING	This refers to properties such as inconsistent upper / lower case usage, multiple blank characters, using numbers, ...
term usage	LING	Terminology used in source / translation
matchtype	LING	MT, TM, TERM, ...
different translations	LING	Similarity of this translation with the other translations of the same segment
author	PERSON	author
reviewer	PERSON	reviewer
translator	PERSON	translator
review status	QUAL	Audited / unaudited / in examination
match quality	QUAL	Irrelevant for exact match
quality translator	QUAL	Rated quality of translator

# Applying the specificity model - example matches and attributes

Segment	Translation	Source segment	Translation	Database	Match type	Translation quality	Date	Useage Count	Translator	Quality Translator	Domain
1	1	Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum ihrer rechtmäßigen	All the brands as well as trade and company names used are property of their lawful	translationwork	TM	2	12.05.2010	5	Mayer	1	law
1	2	Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum ihrer rechtmäßigen	All the brands as well as trade and company names used are property of their lawful	translationwork	TM	2	13.06.2011	4	Müller	2	law
1	3	Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum ihrer rechtmäßigen	All the brands as well as trade and company names used are property of their lawful	marketing	TM	1	15.08.2010	2	Kerner	3	marketing
1	4	Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum ihrer rechtmäßigen	All the brands as well as trade and company names used are property of their lawful proprietor/creator.	translationwork	TM	1	22.03.209	4	Sorder	2	general
1	5	Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum ihrer rechtmäßigen	All the brands as well as trade and company names are property of their rightful owners / copyright.	Google TM	MT	3	-	1	Google	4	law
1	6	Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum ihrer rechtmäßigen	All trademarks used, as well as trade and company names are the property of their rightful	Microsoft MT	MT	3	-	1	Microsoft	4	law
1	7	Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum ihrer rechtmäßigen	All used brands as well as trade names and company names are a property of her lawful owners /	Prompt MT	MT	3	-	1	Prompt	5	general



# Comparing matching translations

Segment	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/creator.	All the brands as well as trade and company names are property of their rightful owners / copyright.	All trademarks used, as well as trade and company names are the property of their rightful owners/authors.	All used brands as well as trade names and company names are a property of her lawful owners / originators.	Mean	Rank of translation
All the brands as well as trade and company names used are property of their lawful proprietor/originator.	100	100	100	96	75	64	76	85	1
All the brands as well as trade and company names used are property of their lawful proprietor/originator.	100	100	100	96	75	64	76	85	1
All the brands as well as trade and company names used are property of their lawful proprietor/originator.	100	100	100	96	75	64	76	85	1
All the brands as well as trade and company names used are property of their lawful proprietor/creator.	96	96	96	100	74	67	73	84	2
All the brands as well as trade and company names are property of their rightful owners / copyright.	75	75	75	74	100	76	75	75	4
All trademarks used, as well as trade and company names are the property of their rightful owners/authors.	64	64	64	67	76	100	64	67	5
All used brands as well as trade names and company names are a property of her lawful owners / originators.	76	76	76	73	75	64	100	73	3
<b>Mean</b>	<b>85</b>	<b>85</b>	<b>85</b>	<b>84</b>	<b>75</b>	<b>67</b>	<b>73</b>	<b>79</b>	

# Ranking translations – use value analysis

---

<b>Segment</b>	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/originator.	All the brands as well as trade and company names used are property of their lawful proprietor/creator.	All the brands as well as trade and company names are property of their rightful owners / copyright.	All trademarks used, as well as trade and company names are the property of their rightful owners/authors.	All used brands as well as trade names and company names are a property of her lawful owners / originators.
<b>Rank</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>4</b>	<b>5</b>	<b>3</b>



# Why use and integrate into XLIFF?

---

- ▶ XLIFF is a standard
- ▶ All information in one place = file
- ▶ XLIFF provides attributes – properties
- ▶ XLIFF 2.0 can now be extended with new modules
- ▶ Translators / project managers require information about the rationality behind decisions taken during the translation
- ▶ XLIFF used within beo
  - ▶ Experience with XLIFF
- ▶ Combination with TMX

# TMX enhanced with recommendation attributes

```
<tu datatype="112" usagecount="0" changeid="11d63a436a360_klemens" creationid="11d63a436a360_klemens">
  <!-- Beorec spezific attributes -->
  <prop type="BeoRec::Domain">translation</prop>
  <prop type="BeoRec::Project">openTMS</prop>
  <prop type="BeoRec::Keyword">openTMS</prop>
  <prop type="BeoRec::Keyword">FOLT</prop>
  <prop type="BeoRec::Application">translation</prop>
  <!-- End beorec spezific attributes -->
  <tuv xml:lang="de" ...>
    <!-- Beorec spezific attributes -->
    <prop type="BeoRec::Author">Klemens</prop>
    <prop type="BeoRec::Source">>true</prop>
    <prop type="BeoRec::Document">http://www.folt.org/</prop>
    <prop type="BeoRec::DocumentType">html</prop>
    <!-- Ende beorec spezific attributes -->
    <seg>Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-
  proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.</seg>
  </tuv>
  <tuv xml:lang="en" ...>
    <!-- Beorec spezific attributes -->
    <prop type="BeoRec::TranslationType">human</prop>
    <prop type="BeoRec::Translator">Otto</prop>
    <prop type="BeoRec::Quality">95</prop>
    <prop type="BeoRec::Reviewer">Kurt</prop>
    <prop type="BeoRec::Source">>false</prop>
    <prop type="BeoRec::UsageNumber">10</prop>
    <prop type="BeoRec::CreationDate">1325868064390</prop>
    <prop type="BeoRec::ChangeDate">1325868078999</prop>
    <!-- Ende beorec spezific attributes -->
    <seg>Basic objectives of FOLT are the support of standardised exchange formats, non-proprietary software
  and trialling new translation technologies and methods.</seg>
  </tuv>
```

Overall attributes

Translation specific attributes

Translation specific attributes

# XLIFF Match with attributes

```
<trans-unit ..">
  <source>Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.</source>
  <alt-trans match-quality="100">
    <source xml:lang="de">Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.</source>
    <target xml:lang="en">Basic objectives of FOLT are the support of standardised exchange formats, non-proprietary software and trialling new translation technologies and methods.</target>
    <prop-group name="TARGET-MOL-f4a13751-aaa7-4a46-8461-fad178c2f089">
      <prop prop-value="BeoRec::UsageNumber">10</prop>
      <prop prop-value="BeoRec::Translator">Otto</prop>
      <prop prop-value="BeoRec::Source">>false</prop>
      <prop prop-value="BeoRec::ChangeDate">1325868078999</prop>
      <prop prop-value="BeoRec::TranslationType">human</prop>
      <prop prop-value="BeoRec::Quality">95</prop>
      <prop prop-value="BeoRec::CreationDate">1325868064390</prop>
      <prop prop-value="BeoRec::Reviewer">Kurt</prop>
    </prop-group>
    <prop-group name="SOURCE-MOLaf21c829-ffec-4d1e-81ac-705ba1911271">
      <prop prop-value="BeoRec::Document">http://www.folt.org/</prop>
      <prop prop-value="BeoRec::Source">>true</prop>
      <prop prop-value="BeoRec::DocumentType">html</prop>
      <prop prop-value="BeoRec::Author">Klemens</prop>
    </prop-group>
    <prop-group name="MULTI-7fa64837-7b54-489c-a50e-1311345a9371">
      <prop prop-value="BeoRec::Keyword">openTMS</prop>
      <prop prop-value="BeoRec::Keyword">FOLT</prop>
      <prop prop-value="BeoRec::Project">openTMS</prop>
      <prop prop-value="BeoRec::Domain">translation</prop>
      <prop prop-value="BeoRec::Application">translation</prop>
    </prop-group>
    <prop-group name="useValue">
      <prop name="useValue">15.240000000000002</prop>
    </prop-group>
  </alt-trans>
```

Explanations can be given for each match in combination with the criteria matrix

# XLIFF Match with attributes

```
<trans-unit ...">
  <source>Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.</source>
  <alt-trans match-quality="100">
    <source xml:lang="de">Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.</source>
    <target xml:lang="en">Basic objectives of FOLT are the support of standardised exchange formats, non-proprietary software and trialling new translation technologies and methods.</target>
    <prop-group name="TARGET-MOL-f4a13751-aaa7-4a46-8461-fad178c2f089">
      <prop prop-value="BeoRec::UsageNumber">10</prop>
      <prop prop-value="BeoRec::Translator">Otto</prop>
      <prop prop-value="BeoRec::Source">false</prop>
      <prop prop-value="BeoRec::ChangeDate">1325868078999</prop>
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
    </prop-group>
    <prop-group name="SOURCE-MO:
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
    </prop-group>
    <prop-group name="MULTI-7fa
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::
      <prop prop-value="BeoRec::Domain">translation</prop>
      <prop prop-value="BeoRec::Application">translation</prop>
    </prop-group>
    <prop-group name="useValue">
      <prop name="useValue">15.240000000000002</prop>
    </prop-group>
  </alt-trans>
```

```
<prop-group name="useValue">
  <prop name="useValue">15.240000000000002</prop>
</prop-group>
```

This **useValue** will now be used to sort the alt-trans matches and present them to the translator in their specific order, will overwrite the similarity value

# Weighting values – criteria matrix – Administration related measures

```
<criteriaMatrix name="runFilter">
  <criteria name="ADMIN" id="a3c7a556-477d-4aa5-b23f-319b0a1d092f">
    <attribute name="BeoRec::ProductKey" weight="7" javascriptTransformer="0" />
  </attribute>
  <attribute name="BeoRec::Application" weight="7" javascriptTransformer="0" />
  </attribute>
  <attribute name="BeoRec::BusinessArea" weight="7" javascriptTransformer="0" />
  </attribute>
  <attribute name="BeoRec::Domain" weight="7" javascriptTransformer="retval = 0;if (BeoRecDomain == "translation") retval
= 4; retval;" />
  </attribute>
  <attribute name="BeoRec::ChangeDate"
  weight="7"
  javascriptTransformer="retval=0;diff=CurrentTimeMillis-
BeoRecChangeDate;diff=diff/(1000*3600*24);if (diff < 7) retval=5;else if (diff < 30) retval=3;else
if (diff < 365) retval=1;retval;" />
  </attribute>
  <attribute name="BeoRec::CreationDate" weight="7" javascriptTransformer="retval=0;diff=CurrentTimeMillis-
BeoRecCreationDate;diff=diff/(1000*3600*24);if (diff < 7) retval=5;else if (diff < 30) retval=3;else if (diff < 365)
retval=1;retval;" />
  </attribute>
  ...
  <attribute name="BeoRec::DocumentType" id="527cc182-8aa0-4458-8161-4c046a17275e" weight="7" javascriptTransformer="0" />
  </attribute>
</criteria>
```

JavaScript is used to compute weights; the rationality behind this is that values must be standardised in some way

# Weighting values – criteria matrix – Linguistic and person related measures

```
<criteria name="LING" >
  <attribute name="BeoRec::TranslationType" id="6bc40361-94e1-4619-b199-3533fef26ef5" weight="0"
  javascriptTransformer="retval=0;if (BeoRecTranslationType == "human") retval=5; if (BeoRecTranslationType
  == "mt") retval=1; retval;" />
  </attribute>
  <attribute name="BeoRec::SourceTargetRatio" weight="5" javascriptTransformer="retval=0;if
  ((BeoRecSourceTargetRatio > 0.80) && (BeoRecSourceTargetRatio < 1.20)) retval=1;if
  ((BeoRecSourceTargetRatio > 0.90) && (BeoRecSourceTargetRatio < 1.10)) retval=2;if
  ((BeoRecSourceTargetRatio > 0.95) && (BeoRecSourceTargetRatio < 1.05)) retval=4;retval;" />
  </attribute>
  ...
</criteria>
```

```
<criteria name="PERSON" >
  <attribute name="BeoRec::Translator" weight="50" javascriptTransformer="retval=5;if (BeoRecTranslator
  == "MicrosoftMT") retval=2; if (BeoRecTranslator == "googleMT") retval=1;if (BeoRecTranslator == "Otto")
  retval=5;if (BeoRecTranslator == "Carla") retval=5;if (BeoRecTranslator == "Susi") retval=5;if
  (BeoRecTranslator == "Karl") retval=5;retval;" />
  </attribute>
  <attribute name="BeoRec::Reviewer" id="74398ee5-31c4-4dfa-b024-6906176fa7b2" weight="40"
  javascriptTransformer="retval = 0;if (BeoRecReviewer == "Kurt") retval = 5;if (BeoRecReviewer == "Maria")
  retval = 2;if (BeoRecReviewer == "Fritz") retval = 3;if (BeoRecReviewer == "none") retval = 0;retval;" />
  </attribute>
  <attribute name="BeoRec::Author" id="2aa7cfbe-7ed9-4800-a30c-822381b6e255" weight="10"
  javascriptTransformer="retval = 0;if (BeoRecAuthor == "Kurt") retval = 5;if (BeoRecAuthor == "Maria")
  retval = 2;if (BeoRecAuthor == "Fritz") retval = 3;if (BeoRecAuthor == "Klemens") retval = 7;retval;" />
  </attribute>
</criteria>
```

# Weighting values – criteria matrix – Quality measures

---

```
<criteria name="QUAL" id="423bef75-a2c1-4eb4-9bc0-d1da004337ec">
  <attribute name="BeoRec::ReviewStatus" id="2093aa5c-e590-4fb1-af7c-be2fc6f14e11" weight="10"
  javascriptTransformer="0" />
  </attribute>
  <attribute name="BeoRec::TranslatorQuality" id="74af8d5f-d1c6-4dc1-aabb-842cebdfd89f" weight="10"
  javascriptTransformer="0" />
  </attribute>
  <attribute name="BeoRec::MatchQuality" id="872bbbda-0057-4621-baa7-dcb8f4c0b578" weight="80"
  javascriptTransformer="0" />
  </attribute>
</criteria>
```

# XLIFF Match with attributes

```
<alt-trans match-quality="100" id="af21c829-ffec-4d1e-81ac-705ba1911271" xml:space="preserve"
origin="d:\eclipse\workspace\beosphereRecommender\test\beorectest\beorec.tmx">
  <source xml:lang="de">Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-
  proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.</source>
  <target xml:lang="en">Basic objectives of FOLT are the support of standardised exchange formats, non-proprietary
  software and trialling new translation technologies and methods.</target>
  <prop-group name="useValue">
    <prop name="useValue">15.240000000000002</prop>
  </prop-group>
</alt-trans>
```

Best recommended match

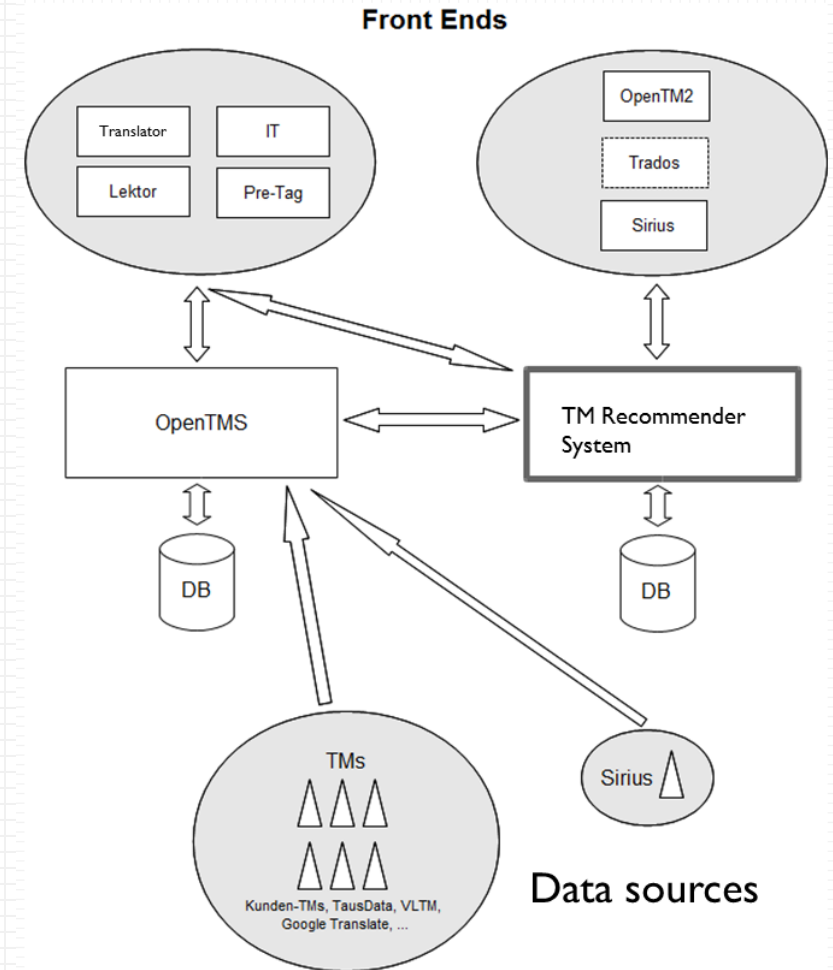
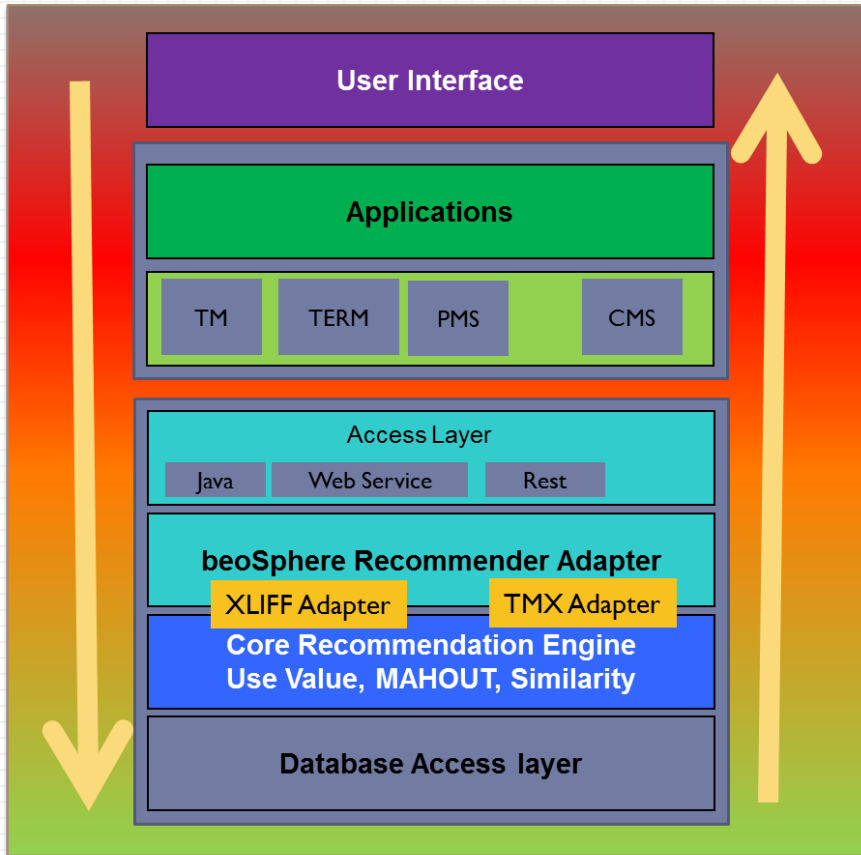
```
<alt-trans match-quality="92">
  <source>Ein Ziel von FOLT ist die Unterstützung von standardisierten Austauschformaten, nicht-proprietärer Software und
  die Erprobung neuer Übersetzungstechnologien und -methoden.</source>
  <target>One objective of FOLT is the support of standardised exchange formats, non-proprietary software and trialling new
  translation technologies and methods.</target>
  ...
  <prop-group name="useValue">
    <prop name="useValue">7.17</prop>
  </prop-group>
</alt-trans>
```

Note difference in match quality

```
<alt-trans match-quality="100">
  <source xml:lang="de">Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-
  proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.</source>
  <target xml:lang="en">The main objectives of FOLT are the support of standardized exchange formats, non-proprietary
  software and the testing of new translation technologies and methods.</target>
  ...
  </prop-group>
  <prop-group name="useValue">
    <prop name="useValue">6.67</prop>
  </prop-group>
</alt-trans>
```



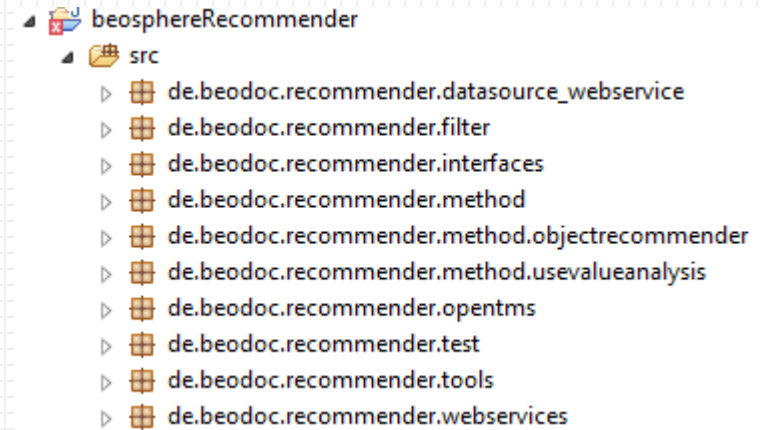
# beoRecommender architecture



# beoRecommender implementation

---

- ▶ Java and JavaScript
- ▶ Uses openTMS core libraries
- ▶ Modules
  - ▶ Use Value Analysis module
  - ▶ Similarity Measure module
  - ▶ XLIFF and TMX extension module
  - ▶ Filter module
  - ▶ Web Service module
  - ▶ Database module



# Summary

---

- XLIFF is useful as it can be easily extended
- Recommendation is a simple and efficient method for a rational decision which translation to use
  - Helps translator and project manager
  - Quality improvement
  - Traceability
  - Explanation for customer
- Depending on recommender method chosen easy to integrate into translation tool applications
  - Even simple methods like use value analysis show good results

---

For more information, please contact:

[klemens.waldhoer@heartsome.de](mailto:klemens.waldhoer@heartsome.de)

