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

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

<table>
<thead>
<tr>
<th>Criteria/Weight</th>
<th>Translator A</th>
<th>Translator B</th>
<th>Translator C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word cost</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery date adherence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamtnutzwert</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria/Weight</th>
<th>Translator A</th>
<th>Translator B</th>
<th>Translator C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word cost</td>
<td>3</td>
<td>0.75</td>
<td>1</td>
</tr>
<tr>
<td>Quality</td>
<td>low</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>Delivery date adherence</td>
<td></td>
<td>low</td>
<td>medium</td>
</tr>
<tr>
<td>Availability</td>
<td>immediate</td>
<td>immediate</td>
<td>later</td>
</tr>
<tr>
<td>Gesamtnutzwert</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

1: bad
2: neutral
3: very good
Recommendation Technology

- **Collaborative Filtering**
  - Recommendations based on behavior of similar users

- **Content Based Filtering**
  - Recommendations based on properties of similar items

- **Use Value Analysis**
  - Simple method for evaluating alternatives based on attributes, criteria and weights for alternatives
Some mathematics for similarity / distance measures

### Similarity measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Type</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Cosinus                  | \[
\frac{\sum_{i=1}^{n} v_i \cdot w_i}{\sqrt{\sum_{i=1}^{n} v_i^2} \cdot \sqrt{\sum_{i=1}^{n} w_i^2}}
\] | Sim  | The cosine measure calculates the angle between two vectors. Range -1 .. 1                  |
| Pearson correlation Coefficient | \[
\frac{\sum_{i=1}^{n} ((v_i - \bar{v}) \cdot (w_i - \bar{w}))}{\sqrt{\sum_{i=1}^{n} (v_i - \bar{v})^2} \cdot \sqrt{\sum_{i=1}^{n} (w_i - \bar{w})^2}}
\] | 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        | \[
\frac{2 \cdot \sum_{i=1}^{n} v_i \cdot w_i}{\sum_{i=1}^{n} v_i^2 + \sum_{i=1}^{n} w_i^2}
\] | Sim  | Correspondence of the same elements                                                       |
| Jaccard-Coefficient      | \[
\frac{\sum_{i=1}^{n} (v_i \cdot w_i)}{\sum_{i=1}^{n} v_i + \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} (v_i \cdot w_i)}
\] | 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

<table>
<thead>
<tr>
<th>Term</th>
<th>Abbreviation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source segment</td>
<td>SS</td>
<td>The sentence to be translated, the source-language segment</td>
</tr>
<tr>
<td>Target Segment</td>
<td>TS</td>
<td>Translation</td>
</tr>
<tr>
<td>Source language</td>
<td>SL</td>
<td>Source language of the document</td>
</tr>
<tr>
<td>Target language</td>
<td>TL</td>
<td>Language of the translation</td>
</tr>
<tr>
<td>Results 1 .. n</td>
<td>M_i</td>
<td>Match for the source segment SS</td>
</tr>
<tr>
<td>Output language segment of the hit (i = 1..K)</td>
<td>M_i (SS)</td>
<td>The i_th matching translation of the k hits for the source segment</td>
</tr>
<tr>
<td>Results matching quality</td>
<td>Q_i</td>
<td>The match quality 0 … 100</td>
</tr>
<tr>
<td>j is target-language segment of the hit (i = 1. meters)</td>
<td>M_i (TS_j)</td>
<td>A source segment hit may have several 1..m translations for the target language</td>
</tr>
<tr>
<td>Number of translations of the hit i</td>
<td>N_i</td>
<td>Number of translations for the hit</td>
</tr>
<tr>
<td>Number of non-target-language translations of the source segment</td>
<td>M_i</td>
<td>Number of translations to MA_i (SS) for all languages. An output segment may also include translations into other languages.</td>
</tr>
<tr>
<td>Number of possible translations back to MA_i (TS, TL) in the source language</td>
<td>R_ij</td>
<td>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.</td>
</tr>
</tbody>
</table>
The Translation Specificity Index Model

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

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

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

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

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

<table>
<thead>
<tr>
<th>Segment</th>
<th>Translation</th>
<th>Source segment</th>
<th>Translation</th>
<th>Database</th>
<th>Match type</th>
<th>Translation quality</th>
<th>Date</th>
<th>Usage Count</th>
<th>Translator</th>
<th>Quality Translator</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum Ihrer rechtmaßigen</td>
<td>All the brands as well as trade and company names used are property of their lawful</td>
<td>translationwork</td>
<td>TM</td>
<td>2</td>
<td>12.05.2010</td>
<td>5</td>
<td>Mayer</td>
<td>1</td>
<td>law</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum Ihrer rechtmaßigen</td>
<td>All the brands as well as trade and company names used are property of their lawful</td>
<td>translationwork</td>
<td>TM</td>
<td>2</td>
<td>13.06.2011</td>
<td>4</td>
<td>Müller</td>
<td>2</td>
<td>law</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum Ihrer rechtmaßigen</td>
<td>All the brands as well as trade and company names used are property of their lawful</td>
<td>marketing</td>
<td>TM</td>
<td>1</td>
<td>15.08.2010</td>
<td>2</td>
<td>Kerner</td>
<td>3</td>
<td>marketing</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum Ihrer rechtmaßigen</td>
<td>All the brands as well as trade and company names used are property of their rightful owners / copyright.</td>
<td>translationwork</td>
<td>TM</td>
<td>1</td>
<td>22.03.209</td>
<td>4</td>
<td>Sorder</td>
<td>2</td>
<td>general</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum Ihrer rechtmaßigen</td>
<td>All trademarks used, as well as trade and company names are the property of their rightful owners</td>
<td>Google TM</td>
<td>MT</td>
<td>3</td>
<td>-</td>
<td>1</td>
<td>Google</td>
<td>4</td>
<td>law</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum Ihrer rechtmaßigen</td>
<td>All trademarks used, as well as trade and company names are the property of their rightful</td>
<td>Microsoft MT</td>
<td>MT</td>
<td>3</td>
<td>-</td>
<td>1</td>
<td>Microsoft</td>
<td>4</td>
<td>law</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Alle verwendeten Marken sowie Handels- und Firmennamen sind Eigentum Ihrer rechtmaßigen</td>
<td>All used brands as well as trade names and company names are a property of her lawful owners /</td>
<td>Prompt MT</td>
<td>MT</td>
<td>3</td>
<td>-</td>
<td>1</td>
<td>Prompt</td>
<td>5</td>
<td>general</td>
</tr>
</tbody>
</table>
## Comparing matching translations

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

Mean | 85 | 85 | 85 | 84 | 75 | 67 | 73 | 79 |
## Ranking translations – use value analysis

<table>
<thead>
<tr>
<th>Segment</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>All the brands as well as trade and company names used are property of their lawful proprietor/originator.</td>
<td>1</td>
</tr>
<tr>
<td>All the brands as well as trade and company names used are property of their lawful proprietor/originator.</td>
<td>1</td>
</tr>
<tr>
<td>All the brands as well as trade and company names used are property of their lawful proprietor/originator.</td>
<td>1</td>
</tr>
<tr>
<td>All the brands as well as trade and company names used are property of their lawful proprietor/originator.</td>
<td>2</td>
</tr>
<tr>
<td>All trademarks used, as well as trade and company names are the property of their rightful owners/copyright.</td>
<td>4</td>
</tr>
<tr>
<td>All trademarks used, as well as trade and company names are a property of her lawful owners/author</td>
<td>5</td>
</tr>
<tr>
<td>All used brands as well as trade names and company names are property of their rightful owners/originators.</td>
<td>3</td>
</tr>
</tbody>
</table>
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
Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-proprieter Software und die Erprobung neuer Übersetzungstechnologien und -methoden.

Basic objectives of FOLT are the support of standardised exchange formats, non-proprietary software and trialling new translation technologies and methods.
Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht- proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.

Basic objectives of FOLT are the support of standardised exchange formats, non-proprietary software and trialling new translation technologies and methods.
<trans-unit ...
</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>
</alt-trans>

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

```xml
<criteriaMatrix name="runFilter">
  <criteria name="ADMIN" id="a3c7a556-477d-4aa5-b23f-319b0a1d092f">
    <attribute name="BeoRec::ProductKey" weight="7" javaScriptTransformer="0" />
    <attribute name="BeoRec::Application" weight="7" javaScriptTransformer="0" />
    <attribute name="BeoRec::BusinessArea" weight="7" javaScriptTransformer="0" />
    <attribute name="BeoRec::Domain" weight="7" javaScriptTransformer="retval = 0;if (BeoRecDomain == \"translation\") retval = 4; retval;" />
    <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 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 name="BeoRec::DocumentType" id="527cc182-8aa0-4458-8161-4c046a17275e" weight="7" javaScriptTransformer="0" />
  </criteria>
</criteriaMatrix>
```

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

```xml
<criteria name="LING">
    <attribute name="BeoRec::TranslationType" id="6bc40361-94e1-4619-b199-3533f6f26ef5" weight="0"
        javaScriptTransformer="retval=0;if (BeoRecTranslationType == "human") retval=5; if (BeoRecTranslationType == "mt") retval=1; retval;" />
    <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;" />
</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 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 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;" />
</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 name="BeoRec::TranslatorQuality" id="74af8d5f-d1c6-4dc1-aabb-842cebdf89f" weight="10"
    javaScriptTransformer="0" />
  <attribute name="BeoRec::MatchQuality" id="872bbbd4-0057-4621-baa7-dcb8f4c0b578" weight="80"
    javaScriptTransformer="0" />
</criteria>
Wesentliche Ziele von FOLT sind die Unterstützung von standardisierten Austauschformaten, nicht-
proprietärer Software und die Erprobung neuer Übersetzungstechnologien und -methoden.

Basic objectives of FOLT are the support of standardised exchange formats, non-proprietary
software and trialling new translation technologies and methods.

The main objectives of FOLT are the support of standardized exchange formats, non-proprietary
software and the testing of new translation technologies and methods.

Best recommended match

Note difference in match quality
beoRecommender architecture

User Interface

Applications

TM  TERM  PMS  CMS

Access Layer

Java  Web Service  Rest

beoSphere Recommender Adapter

XLIFF Adapter  TMX Adapter

Core Recommendation Engine
Use Value, MAHOUT, Similarity

Database Access layer

Front Ends

Translator  IT  Lektor  Pre-Tag

OpenTMS

TM Recommender System

Data sources

OpenTM2  Trados  Sirius

Sirius

Kunden-TMs, TausData, VLTM, Google Translate, ...
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