Knowledge management issues in the workflow of translation memory systems

Timmy Oumai Wang (Imperial)
Mark Shuttleworth (UCL)
Introduction

• Inspiration of the project: producing a new kind of translation resource, or improving on old ones.
• Now the aim is to find a way to supplement what TMSs already do by providing them with a new kind of translation resource based on the semantic power of knowledge management.
Commercial TMSs: the state of the art

- The technology is now 20 years old.
- Still essentially the same technology: provides suggestions based on string matching rather than some measure of semantic similarity.
- This means that the two sentences
  "I live in a house & I live in a skyscraper"
may not be recognised as similar.
More sophisticated approaches

- Déjà Vu ‘Assemble from portions’ and memoQ ‘Assemble from fragments’.
- Supplementing TM hits by the use of on-line MT.
- Similis: grammatical pattern matching.
- Research currently being conducted on methods of semantic matching.
- This is the context in which this present work is being carried out.
Knowledge management and knowledge management systems

- Knowledge management: a generic concept that refers to the process of creating, sharing and applying knowledge (Stevens et al. 2010:131-132).
- Knowledge in the KM context.
- A knowledge management system is an information system that supports or enables activities of managing knowledge (Alavi & Leidner, 2001).
- KMSs should serve the general objectives of knowledge management.
TMS as a type of KMS

• Most CAT tools are designed for two purposes:
  i. Improving translation quality
  ii. Improving translation efficiency

• A TMS can be seen as a type of KMS that aims to serve a translation purpose.
Three Categories of Knowledge involved in the workflow of a TMS

1. The knowledge that is manipulated directly by the TMS (e.g. translation memory files)
2. The knowledge that is used within the TMS to enhance its performance (e.g. linguistic data, ontologies, etc.)
3. The knowledge that is used by translators to employ translation suggestions (e.g. translator’s competences)
Workflow of TMSs as a type of KMS

- Using the Knowledge Spiral Model (Nonaka & Takeuchi, 1995) to describe the activities in a TMS workflow.

![Nonaka's Knowledge Spiral Model](image-url)
Workflow of TMSs as a type of KMS

Two types of Knowledge:

1. Explicit knowledge: ‘formal and systematic information, such as quantifiable data, codified procedures, [and] universal principles’.

2. Tacit knowledge: hard to express and is embedded in individual experiences in forms such as insights, intuitions and hunches.
Workflow of TMSs as a type of KMS

- Knowledge Capture
- Knowledge Codification
- Knowledge Application
- Knowledge Creation
Current KM Bottlenecks in TMS

• Low efficiency of using TM files

• XML-based formats such as TMX or XLIFF do not store semantic information or other more descriptive features of translation units.
The transfer and conversion of knowledge in the workflow of a TMS seen as a type of KMS

<table>
<thead>
<tr>
<th>Knowledge Capture</th>
<th>Tacit knowledge to Explicit Knowledge</th>
<th>The first category of knowledge is usually involved.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Codification</td>
<td>Tacit knowledge to Explicit Knowledge</td>
<td>No knowledge is necessarily involved</td>
</tr>
<tr>
<td>Knowledge Application</td>
<td>Explicit Knowledge to Tacit Knowledge</td>
<td>The second category of knowledge can be involved</td>
</tr>
<tr>
<td>Knowledge Creation</td>
<td>Tacit Knowledge to Explicit Knowledge</td>
<td>The third category of knowledge is involved</td>
</tr>
</tbody>
</table>
A New Development for TMSs

- New resources that can be exploited for a second category of knowledge: WordNet, ConceptNet, Wikipedia.

- Initial stage of experimentation:
  1. Three corpora based on different text types (each 230,000 words)
  2. Terrier as the information retrieval platform
conceptIndexBean = config.getString("concept_index.bean");

IConceptIndex index = (IConceptIndex) context.getBean("conceptIndexBean");
logger.info("size of source index: " + index.size());
IConceptExtractor esaExtractor = index.getConceptExtractor();

TextDocument docA = new TextDocument("text_a");
docA.setText("content", Language.EN, "Neither pilot saw the other aircraft in time to take effective avoiding action");

TextDocument docB = new TextDocument("text_b");
docB.setText("content", Language.EN, "Tomorrow, Japan is over ");
//config.getString("text_b");
First type of translation suggestion

- Semantically Similar Suggestions:

<table>
<thead>
<tr>
<th>Query</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>The cause of the braking loss could not be positively established.</td>
<td>The cause of the failure could not be assessed.</td>
</tr>
</tbody>
</table>
Second type of translation suggestion

- Conceptual explanation or clarification:

<table>
<thead>
<tr>
<th>Query</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark energy will continue to push galaxies ever faster away until they fade completely from view.</td>
<td>But if the dark energy density decreases and matter becomes dominant again, our cosmic horizon will grow, revealing more of the universe.</td>
</tr>
</tbody>
</table>
Conclusion
Thank you very much!

Timmy Oumai Wang
oumai.wang10@imperial.ac.uk

Mark Shuttleworth
m.shuttleworth@ucl.ac.uk