Combining language process metadata with linguistic linked data

Felix Sasaki
DFKI / W3C fellow
<html> ...
<p>The <span translate="no">LIDER</span> project ...</p> ...</html>
Opportunity

- Regard marked up documents with translation process metadata as a corpus
- Convert to NIF linked data representation
- Benefits: data integration of language process metadata with linguistic linked data
<html> ...
<p>The <span translate="no">LIDER</span> project ...</p> ...
</html>

<http://example.com/service?input=mydoc&char=35,40> nif:beginIndex "35";
nif:endIndex "40";
nif:referenceContext <http://example.com/service?input=mydoc&char=0,140>;
nif:wasConvertedFrom
a nif:RFC5147String;
itsrdf:translate "no".
<html> ...

<p>The <span translate="no">LIDER</span> project ...

<http://example.com/service?input=mydoc&char=35,40> nif:beginIndex "35";
nif:endIndex "40";
nif:referenceContext <http://example.com/service?input=mydoc&char=0,140>;
nif:wasConvertedFrom <http://example.com/service?input=mydoc&xpath=/html/body/p[1]/span[1] >;
a nif:RFC5147String;
itsrdf:translate "no";
a lemon:LexicalEntry ; ...

• How to represent process information in RDF
  – ITS 2.0 ontology (to be finalized)
• Conversion of existing (XML) vocabularies
  – Representing terminology information (e.g. TBX) in RDF
• Integration with further non linguistic (non process) information
• Provide access to LLD for non linked data people
  – See qLabel example
Workflow steps:
1) User clicks on “change language to X” in Web page
2) Web page calls qLabel library with target language X as parameter
3) qLabel resolves linked data identifier available as ITS 2.0 “Text Analysis”
4) qLabel retrieves cross-lingual information available for the identifier
5) User sees the Web page in the changed language
Potential synergies

• W3C Annotations workshop
  http://www.w3.org/2014/04/annotation/

• Upcoming hooks to link into heterogeneous types of Web content
  – HTML documents, ePub, multimedia, ...

• Similar discussions like in LIDER
  – E.g. licenses for (linked) annotations

• A wide range of usage scenarios beyond translation + linguistic analysis
  – Try to avoid competing solutions
Combining language process metadata with linguistic linked data

Felix Sasaki
DFKI / W3C fellow