LMF-aware Web services for accessing lexical resources

Yoshihiko Hayashi (Osaka University, Japan)

In collaboration with:
Bora Savas, Chiharu Narawa (Osaka University, Japan)
Monica Monachini, Claudia Soria, Nicoletta Calzolari
(ILC-CNR, Pisa, Italy)
Executive summary

• International standards for language resource management, worked out by ISO, can be effectively utilized in implementing standardized language Web services, in particular for accessing lexical resources.

• As a proof of concept example, a Web service for accessing WordNet-type semantic lexicons is described.

• Key ingredients:
  • *lexical markup framework (LMF)*: a standardized framework for modeling lexical resources
  • *representational state transfer (REST)*: a Web service architecture
WordNet-type semantic lexicons

• A number of wordnets have been developed after Princeton WordNet (PWN) for languages other than English, including Japanese WordNet (WN-Ja).
• PWN is a large lexical database of English based on “relational semantics.”
  • Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.
  • Synsets are interlinked by means of conceptual-semantic and lexical relations.
Relational structure in WordNet

- `hyponym_of`:
  - `{motor vehicle#1, automotive vehicle#1}`
  - `{car#2, railcar#1, railway car#1, railroad car#1}`

- `meronym_of`:
  - `{automobile engine#1}`
  - `{car window#1}`
Lexical Markup Framework (LMF)

- LMF is a standardized framework (ISO 24613:2008) for *modeling a wide range of lexical resources* (lexicons).
- *Wordnet-LMF* is a version of LMF, especially dedicated to the modeling of WordNet-type lexicons.
Basic ideas

- Accessing to a lexicon is accomplished by query-driven extraction and presentation of the relevant portions of the lexicon (sub-lexicon).
- A sub-lexicon can be specified by a REST URI, if we can devise a set of URI patterns to represent possible sub-lexicons.
- The resulting sub-lexicon can be rendered as a Wordnet-LMF/XML document.
URI patterns

http://[server]/[lexicon]/[LMF_path]/[LMF_attribute]/[qterm]/

structural constraints

http://[server]/[lexicon]/[LMF_path]/?[LMF_attribute]=[value]/

?[attr/dir]= [value]

further constraints and/or directives for improving usability can be specified

Examples:
http://server/pwn/LexicalEntry/Lemma/writtenForm/book/?partOfSpeech=v&get_synsets

http://server/pwn/Synset/Definition/?gloss=*born in Ireland*
Directives

- **get_synsets**: to collect all associated synsets for the obtained lexical entry
- **get_synset_by_index=number**: to retrieve the only synset associated with the obtained lexical entry with the designated sense number
- **get_synsets_by_relation=relation**: to collect all linked synsets from the obtained source synset with the designated conceptual relation type
<LexicalResource>
  <GlobalInformation label="/lmf/pwn/LexicalEntry/Lemma/writtenForm/book/?partOfSpeech=v&get_synsets"/>
  <Lexicon languageCoding="ISO 639-3" label="English Wordnet 3.0" language="en" owner="PRINCETON" version="3.0">
    <LexicalEntry id="w148381"/>
    <Synset id="eng-09-00678777-v" baseConcept="3">
      <Definition gloss="engage for a performance">
        <Statement example="Her agent had booked her for several concerts in Tokyo"/>
      </Definition>
    </Synset>
    <Synset id="eng-09-02498320-v" baseConcept="3">
      <Definition gloss="arrange for and reserve (something for someone else) in advance">
        <Statement example="reserve me a seat on a flight; 'The agent booked tickets to the show for the whole family'; 'please hold a table at Maxim's"/>
      </Definition>
    </Synset>
    <Synset id="eng-09-02498136-v" baseConcept="3">
      <Definition gloss="record a charge in a police register">
        <Statement example="The policeman booked her when she tried to solicit a man"/>
      </Definition>
    </Synset>
    <Synset id="eng-09-02599754-v" baseConcept="3">
      <Definition gloss="register in a hotel Booker">
        <SynsetRelations/>
      </Definition>
    </Synset>
  </Lexicon>
</LexicalResource>
Wordnet-LMF/XML to HTML via Style conversion

• LMF/XML, primarily for machine consumption, can be readily converted to HTML for human consumption (navigation).

• By adding &html to a URI, the resulting XML is accompanied by a predefined XSLT stylesheet.

• Web browsers can perform accordingly defined style conversion.

http://server/pwn/LexicalEntry/Lemma/writtenForm/book/?partOfSpeech=v&get_synsets&html
A proposal to revise Wordnet-LMF to accommodate bilingual semantic lexicons

The EDR Electronic Dictionary is thought of as a bilingual semantic lexicon whose information structure can be modeled as WordNet-type.

concept node (3bc999) can be represented as a kind of synset: {銀行, バンク, bank, bnk., bk.}
Synset/Definition should be bilingual, more generally, multilingual

Synset
id=edr_cph-3bcbd2-x

Lexical Entry
(E: Ireland)

Definition
lang=en
gloss="a country called Ireland"

Lexical Entry
(J: アイルランド)

Definition
lang=ja
gloss="アイルランドという国"
Resulting Wordnet-LMF/XML document:

EDR example

```
- <LexicalResource>
  - <GlobalInformation label="/lmf/edr/Synset/Definition/?gloss=a%20country%20called%20Ireland"/>
  - <Lexicon languageCoding="ISO 639-3" label="EDR" language="en ja" owner="NICT" version="1.0">
    - <LexicalEntry id="edr_je_1411">
      <Lemma writtenForm="アイルランド" partOfSpeech="JN2" pronunciation="アイルランド"/>
      <Sense index="1" id="x_1411_JN2" synset="edr_cph-3bcbd2-x"/>
    </LexicalEntry>
    + <LexicalEntry id="edr_je_37912"></LexicalEntry>
    + <LexicalEntry id="edr_je_1412"></LexicalEntry>
    + <LexicalEntry id="edr_ej_5552"></LexicalEntry>
    + <LexicalEntry id="edr_ej_5754"></LexicalEntry>
    + <LexicalEntry id="edr_ej_8362"></LexicalEntry>
    + <LexicalEntry id="edr_ej_9251"></LexicalEntry>
    + <LexicalEntry id="edr_ej_9262"></LexicalEntry>
    + <LexicalEntry id="edr_ej_9335"></LexicalEntry>
    - <Synset id="edr_cph-3bcbd2-x" baseConcept="undef">
      <Definition lang="en" gloss="a country called Ireland"/>
      <Definition lang="ja" gloss="アイルランドという国"/>
    </Synset>
  </Lexicon>
</LexicalResource>
```
Concluding remarks

• LMF, an International standards for modeling lexicons, can be effectively utilized in implementing standardized lexicon access Web services; however slight modifications were required in order to accommodate an innately bilingual semantic lexicon, EDR.

• Our next steps include:
  • incorporate *more wordnets* to further attest the applicability of the LMF standard
  • *Web-servicize other types of lexical resources*: e.g. bilingual dictionaries with the same architecture
  • implement an *RDF endpoint for Lexical Linked Data*
    • requiring standardized RDF representation of LMF
Thank you very much!

Acknowledgments:

• Research collaboration
  • ILC-CNR, Italy
  • NICT Language Grid Project, Japan
• Grant support
  • The Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan
References

- Princeton WordNet: http://wordnet.princeton.edu/
- EDR: http://www2.nict.go.jp/r/r312/EDR/index.html
- LMF: http://www.lexicalmarkupframework.org/
- Wordnet-LMF: http://wordnetlmf.sourceforge.net/

- Representative publication:
- Project web site: http://chiron.lang.osaka-u.ac.jp/scope/