WNBuilder

1. BASIC INFORMATION
    1. Tool name: *WNBuilder (WordNetBuilder)*
    2. Overview and purpose of the tool:

WNBuilder is a configurable graphical interface, click controlled, by means of which a lexicographer has access to the language resources necessary in building an interlingually-aligned wordneti, i.e. a bilingual dictionary, a dictionary of synonyms and an explanatory dictionary.

3. A short description of the algorithm

The interface ensures the following main functions:

- synsets definition (sense assignment to the literals of the synonymy series and gloss attachment) and their mapping onto the interlingual index (ILI) via a set of user defined equivalence relations;

- importing relations from the source wordnet (PWN) into the target wordnet;

- validation functions: the most useful functions are: validating the syntax of the created synsets, search for sense assignment conflicts, duplicated literals in a synset, dangling nodes or relations, missing synsets, etc.

2. TECHNICAL INFORMATION
    1. Software dependencies and system requirements

WNBuilder requires Jscript, ActiveState Perl and MSXML. It runs under InternetExplorer v. 6 to 8.

2. Installation

To install WNBuilder, run the file install.bat.

3. Execution instructions

Given a txt file containing a list of ILIs (of the synsets to be implemented), WNBuilder builds the XML concepts file. Then this is opened in the interface.

Clicking any ILI code will turn its color into red (signifying a visited ILI record, not yet implemented) and in the upper right frame there will appear:

- the English synset (from the the source wordnet) (and its associated gloss) which is mapped onto the respective ILI record;

- a list of translation equivalents for the words in the English synset. The translation equivalents are taken from a bilingual dictionary. By selecting (clicking) one translation equivalent in this list, the interface will display the following information:

- the definitions of the selected translation (in the left low frame; they are extracted on the fly from the explanatory dictionary);

- all the synonymy sets which the selected translation belongs to (in the right low frame; they are extracted on the fly from the dictionary of synonyms). Each literal in a synonymy set is linked to a headword entry in the explanatory dictionary so that the lexicographer has the possibility to see all the definitions for each word in the current synonymy set.

With this information displayed in a friendly format, the lexicographer has to answer four main questions and make decisions that in the end would result in a target language synset, mapped to the starting ILI-record:

1. *which are the best equivalents for the literals in the selected English synset; the lexicographer has the possibility to add new translation equivalents.*
2. *which of the synonymic sets best fits the English synset. The lexicographer can add or delete words from each of the synonym set, or can create his/her own synonym set if a relevant one is not present in the dictionary of synonyms;*
3. *which of the definitions (if different) of the translation and its synonyms best fits the English gloss;*
4. *which is the interlingual relation between the English synset and the Romanian synset under construction; the interface gives the lexicographer the possibility to select among a set of interlingual relations.*

After the lexicographer completed one or more target synsets and mapped them on the ILI records (via interlingual relations) s/he may launch the syntactic validation functions of WNBuilder. These validation functions are also automatically launch each time the work of the lexicographer is saved. The color of the ILI codes in the upper left frame linked to the synsets that were completed and passed the validation tests turns green (signifying an already implemented ILI record). The ILI codes may be ordered according to their colors so that the unvisited or not yet implemented ILI records come to the top of the upper left frame (the blue and red codes). The completed synsets (name stamped) may be at any moment saved in XML format. If errors are still present in the generated semantic sub-network, they are recorded into a separate file for the subsequent correction.

4. Input/Output data formats
   a. Input data formats

The txt file containing the list of ILIs must be UTF8 encoded.

b. Output data formats

The implemented synsets are saved in XML format.

5. Integration with external tools

WNBuilder is fully self-contained.

3. CONTENT INFORMATION
   1. a test input file

The input files are: the xml format of the source wordnet and a txt file containing a list of ILIs to be implemented in the target wordnet.

2. the output file

The output file is in xml format and contains the implemented synsets in the target language.

3. approximation of the time necessary to process the test input file.

The implementation of synsets is done manually.

4. ADMINISTRATIVE INFORMATION

1. Contact

Dan Tufiş, "Mihai Drăgănescu" Research Institute for Artificial Intelligence of the Romanian Academy, tufis@racai.ro

5. RELEVANT REFERENCES AND OTHER INFORMATION

Dan Tufis and Eduard Barbu. *A Methodology and Associated Tools for Building Interlingual Wordnets*. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), pp. 1067-1070, Lisbon, Portugal, May 2004. ELRA - European Language Resources Association. ISBN 2-9517408-1-6.

Dan Tufis, Eduard Barbu, Verginica Barbu-Mititelu, Radu Ion, and Luigi Bozianu. *The Romanian Wordnet*. In Dan Tufis (ed.), Romanian Journal on Information Science and Technology. Special Issue on BalkaNet, volume 7, pp. 105-122. Romanian Academy, April 2004. ISSN 1453-8245.