CombiLex 1.0

Sources

CombiLex is based on the following sources:
1. Wordlist of the Dutch Language 2005, source file 2.0
2. e-Lex 1.0
3. Reference Lexicon for Dutch 2.0 (RBN 2.0)

Spelling

- The Wordlist of the Dutch Language is in accordance with the current spelling rules (2005)
- e-Lex is not in accordance with the current spelling rules
- RBN is not in accordance with the current spelling rules

Number of words

CombiLex contains over 213.000 unique lemmas.
CombiLex contains over 442.000 unique lemmas and word forms.

Two files

CombiLex consists of two files:

1. **combilex.txt**
   This is a tab-separated file in which the first column contains the lemmas, and any following columns contain corresponding word forms. These word forms are not listed in a particular order. Word forms of homonyms are put together: e.g. the lemma *bal* is followed by both the plural *bals* and the plural *ballen*.

2. **combilex_list.txt**
   This is a "flat" version of combilex.txt: all lemmas and word forms are simply listed in one column (so one entry per row).

Selection of the content

The lemmata and word forms in CombiLex are taken from the following lexicon fields:
- Wordlist of the Dutch Language: the fields *lem05*, *flex105*, *flex205*, *flex305*, *flex405*, *flex505*;
- e-Lex: the fields *lem* en *wordf*;
- Reference Lexicon for Dutch: the information is derived from different tables:

  o Adjectives: the fields *id_form*, *mor_comp*, *mor_super*
  o Verbs: the field *id_form*
  o Substantive: the fields *id_form*, *morpho_plurform1*, *morpho_plurform2*
  o Function words and adverbs: the field *id_form*

The word information in the Wordlist of the Dutch Language contains codes which are attached to the words themselves. Those codes were converted (e.g. codes for letters with an accent) or removed.

Remarks

For bugs, suggestions or other remarks concerning this product, please contact the servicedesk of the TST-Centrale: servicedesk@tst-centrale.org.