

DanNet

1. BASIC INFORMATION

1.1 Lexicon type

Wordnet

1.2 Representation of the lexicon

Two files: csv format and OWL format + README file

1.3 Character encoding

The characters have been encoded in UTF8

2. ADMINISTRATIVE INFORMATION

2.1 Contact person

Name: Bolette Sandford Pedersen

E-mail: bspedersen@hum.ku.dk

2.2 Copyright statement and information on IPR

Open Source, Princeton license

3. TECHNICAL INFORMATION

3.1 Data structure of an entry

A synset contains a list of lemma(s), a gloss, and a list of relations and features. The information is organized in the following synset files:

synsets.rdf: The synsets are declared in this file.

glossary.rdf: The gloss of each synset.

words.rdf: Words and their lexical form.

wordsenses.rdf: Connects synsets and words.

synset_attributes.rdf: features that describe the synset further (such as connotation ('positive' or 'negative') and 'sex' ('male' or 'female')).

register.rdf: Information about the register of the word senses, e.g. 'slang' or 'sj.' (i.e. 'old-fashioned').

3.2 Lexicon size

58,716 nouns, 8195 verbs and 3680 adjectives. Amounts to 65,000 synsets, 5,000 links to Princeton Wordnet (core wordnet)

4. CONTENT INFORMATION

4.1 The natural language(s) of the lexicon

Danish

4.2 Entry Type

Synonym set (Synset)

4.3 Attributes

The key attributes are the semantic relations and features. The resource contains 32 relation types and seven feature types.

4.4 Coverage of the lexicon

The resource covers 58,716 nouns, 8195 verbs and 3680 adjectives. It represents the basic vocabulary of modern Danish.

4.5 Intended application of the lexicon

Intended applications are advanced applications such as Information Retrieval, Question Answering and Machine Translation.

4.6 Reliability (automatically/manually constructed)

The resource is semi-automatically constructed. 2% has been manually validated.

5. RELEVANT REFERENCES AND OTHER INFORMATION

Pedersen, B.S, S. Nimb, J. Asmussen, N. Sørensen, L. Trap-Jensen, H. Lorentzen (2009). DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. Language Resources and Evaluation, Computational Linguistics Series. 10.1007/s10579-009-9092-1
[http:// wordnet.dk](http://wordnet.dk)