

PWN 3.0-2.0 Concept Mapping

Table of Contents

1	BASIC INFORMATION.....	2
1.1	Resource composition.....	2
1.2	Representation of the resource (flat files, database, markup).....	2
1.3	Character encoding.....	2
2	ADMINISTRATIVE INFORMATION.....	2
2.1	Contact person (name, address, affiliation, position, telephone, fax, e-mail).....	2
2.2	Delivery medium (if relevant; description of the content of each piece of medium).....	2
2.3	Copyright statement and information on IPR.....	3
3	TECHNICAL INFORMATION.....	3
3.1	Directories and files.....	3
3.2	Data structure of an entry.....	3
3.3	Resource size (no. of rules, MB occupied on disk).....	3
4	CONTENT INFORMATION.....	3
4.1	Type of the resource.....	3
4.2	The natural language(s) for the resource is applicable (if language dependent).....	3
4.3	Domain(s)/register(s) of the corpus.....	4
4.4	Annotations in the corpus (if an annotated corpus).....	4
4.5	Intended application of the resource.....	4
4.6	Reliability of the annotations (automatically/manually assigned) – if any.....	4
5	REFERENCES.....	4

1 BASIC INFORMATION

1.1 Resource composition

PWN 3.0-2.0 Concept Mapping is a resource containing the correspondences between Princeton WordNet (Fellbaum, 1998) concepts / synsets for versions 3.0 and 2.0. These concepts / synsets are encoded as Inter Lingual Indexes (ILIs), which are 8-digit numbers (e.g.: 00001740).

1.2 Representation of the resource (flat files, database, markup)

This resource contains 4 files, each corresponding to the existing Part-of-Speech (POS) categories (nouns, verbs, adjectives, adverbs). Each file contains raw text.

1.3 Character encoding

Character Encoding is ANSI.

2 ADMINISTRATIVE INFORMATION

2.1 Contact person (name, address, affiliation, position, telephone, fax, e-mail)

NAME: Dan TUFİŞ / Dan ŞTEFĂNESCU

ADDRESS: Calea 13 Septembrie, No. 13, CASA ACADEMIEI, Bucharest 050711, ROMANIA

AFFILIATION: Research Institute for Artificial Intelligence, Romanian Academy

POSITION: Director / Researcher

TELEPHONE: +40213 188 103

FAX: +40213 188 142

E-MAIL: tufis@racai.ro / danstef@racai.ro

2.2 Delivery medium (if relevant; description of the content of each piece of medium)

This resource can be downloaded from any of the links below:

<http://nlptools.racai.ro/nlptools/index.php?page=pwn3to2>

<http://www.racai.ro/ResearchActivity/WebServicesandResources/WordNetMappings/tabid/138/Default.aspx>

2.3 Copyright statement and information on IPR

This is a Free resource.

3 TECHNICAL INFORMATION

3.1 Directories and files

Folder *PWN_3.0-2.0_Concept_Mapping* has 4 files, each containing PWN 3.0 to PWN 2.0 concept mapping for the existing POS categories:

- adj.txt
- adv.txt
- noun.txt
- verb.txt

3.2 Data structure of an entry

An entry has the following structure:

ILI_PWN_3.0 <tab> ILI_PWN_2.0

An ILI is the Inter Lingual Index encoding a unique PWN concept / synset.

3.3 Resource size (no. of rules, MB occupied on disk)

Resource size is detailed by the following table:

POS Category	File Name	Disk Space	Concepts mapped (lines)
Adjectives	adj.txt	330.7 KB	17,828
Adverbs	adv.txt	66.6 KB	3,594
Nouns	noun.txt	1.4 MB	79,647
Verbs	verb.txt	250.3 KB	13,492
Total		2MB	114,561

Table 1: Resource size for all POS categories

4 CONTENT INFORMATION

4.1 Type of the resource

This is a language-dependent resource. It works for ENGLISH PWN.

4.2 The natural language(s) for the resource is applicable (if language dependent)

This resource refers to ENGLISH PWN concepts / synsets.

4.3 Domain(s)/register(s) of the corpus

This resource falls into General Domain.

4.4 Annotations in the corpus (if an annotated corpus)

No annotation is present.

4.5 Intended application of the resource

Update tools that work for PWN 2.0

4.6 Reliability of the annotations (automatically/manually assigned) – if any

This resource has been created using both automatic and manual operations. After the automatic step, which involved finding identical definitions, similar hyper/hypo-nyms, similar sense numbers for the literals, etc., this resource was compared with a similar one, automatically developed by the UPC NLP Research Group¹. All common mappings were considered correct and the differences were manually analyzed by linguists. During this analysis, the incorrect mappings were removed (Tufiş et al., 2011).

5 REFERENCES

Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Tufiş, D., Ion, R., Mititelu, V., Irimia, E., Ştefănescu, D., Mihăilă, C. (2011). Extending and completing the Ro-WordNet lexical ontology by eliminating the existing semantic conflicts and by validating the differential semantics model based on Ro-WordNet. Academic report (in Romanian). Bucharest, Romania.

¹ http://nlp.lsi.upc.edu/web/index.php?option=com_content&task=view&id=21&Itemid=59