NAME

wn_pl − description of Prolog database files

DESCRIPTION

The files **wn_*.pl** contain the WordNet database in a prolog-readable format. A prolog interface to WordNet is not implemented.

The prolog database is very large and may take many minutes to load into the Prolog workspace. A separate file has been created for each WordNet relation giving the user the ability to load only those parts of the database that they are interested.

See **FILES**, below, for a list of the database files and **wndb**(5WN) and **wninput**(5WN) for detailed descriptions of the various WordNet relations (referred to as *operators* in this manual page).

File Format

Each prolog database file contains information corresponding to the synsets and word senses contained in the WordNet database. In the prolog version of the database, the *synset_id*s (defined below) are used as unique synset identifiers.

Each line of a file contains an operator that corresponds to a WordNet relation. All lines with the same *operator* value are stored in the file **wn_***operator***.pl**.

The general format of a line in a prolog database file is as follows:

*operator***(***field1***,** ... **,***fieldn***).**

Each line contains the name of the operator, followed by a left parenthesis, a comma-separated list of fields, a right parenthesis, and a period. Note there are no spaces, and each line is terminated with a newline character.

Operators

Each WordNet relation is represented in a separate file by *operator* name. Some operators are reflexive (i.e. the "reverse" relation is implicit). So, for example, if **x** is a hypernym of **y**, **y** is necessarily a hyponym of **x**. In the prolog database, reflected pointers are usually implied for semantic relations.

Semantic relations are represented by a pair of *synset_id*s, in which the first *synset_id* is generally the source of the relation and the second is the target. If two pairs *synset_id***,***w_num* are present, the operator represents a lexical relation between word forms.

s(*synset_id***,***w_num***,'***word***',***ss_type***,***sense_number***,***tag_count***).**

A **s** operator is present for every word sense in WordNet. In **wn_s.pl**, *w_num* specifies the word number for *word* in the synset.

sk(*synset_id***,***w_num***,'***sense_key***').**

A **sk** operator is present for every word sense in WordNet. This gives the WordNet sense key for each word sense.

g(*synset_id***,'***gloss***').**

The **g** operator specifies the gloss for a synset.

syntax(*synset_id***,***w_num***,***syntax***).**

The **syntax** operator specifies the syntactic marker for a given word sense if one is specified.

hyp(*synset_id***,***synset_id***).**

The **hyp** operator specifies that the second synset is a hypernym of the first synset. This relation holds for nouns and verbs. The reflexive operator, hyponym, implies that the first synset is a hyponym of the second synset.

ins(*synset_id***,***synset_id***).**
The **ins** operator specifies that the first synset is an instance of the second synset. This relation holds for nouns. The reflexive operator, has_instance, implies that the second synset is an instance of the first synset.

ent(*synset_id***,***synset_id***).**
The **ent** operator specifies that the second synset is an entailment of first synset. This relation only holds for verbs.

sim(*synset_id***,***synset_id***).**
The **sim** operator specifies that the second synset is similar in meaning to the first synset. This means that the second synset is a satellite the first synset, which is the cluster head. This relation only holds for adjective synsets contained in adjective clusters.

mm(*synset_id***,***synset_id***).**
The **mm** operator specifies that the second synset is a member meronym of the first synset. This relation only holds for nouns. The reflexive operator, member holonym, can be implied.

ms(*synset_id***,***synset_id***).**
The **ms** operator specifies that the second synset is a substance meronym of the first synset. This relation only holds for nouns. The reflexive operator, substance holonym, can be implied.

mp(*synset_id***,***synset_id***).**
The **mp** operator specifies that the second synset is a part meronym of the first synset. This relation only holds for nouns. The reflexive operator, part holonym, can be implied.

der(*synset_id***,***synset_id***).**
The **der** operator specifies that there exists a reflexive lexical morphosemantic relation between the first and second synset terms representing derivational morphology.

cls(*synset_id***,***w_num***,***synset_id***,***w_num***,***class_type***).**
The **cls** operator specifies that the first synset has been classified as a member of the class represented by the second synset. Either of the *w_num***'s can be 0, reflecting that the pointer is semantic in the original WordNet database.**

cs(*synset_id***,***synset_id***).**
The **cs** operator specifies that the second synset is a cause of the first synset. This relation only holds for verbs.

vgp(*synset_id***,***w_num***,***synset_id***,***w_num***).**
The **vgp** operator specifies verb synsets that are similar in meaning and should be grouped together when displayed in response to a grouped synset search.

at(*synset_id***,***synset_id***).**
The **at** operator defines the attribute relation between noun and adjective synset pairs in which the adjective is a value of the noun. For each pair, both relations are listed (ie. each *synset_id* is both a source and target).

ant(*synset_id***,***w_num***,***synset_id***,***w_num***).**
The **ant** operator specifies antonymous *word*s. This is a lexical relation that holds for all

syntactic categories. For each antonymous pair, both relations are listed (ie. each *synset_id,w_num* pair is both a source and target word.)

sa(*synset_id***,***w_num***,***synset_id***,***w_num***).**
The **sa** operator specifies that additional information about the first word can be obtained by seeing the second word. This operator is only defined for verbs and adjectives. There is no reflexive relation (ie. it cannot be inferred that the additional information about the second word can be obtained from the first word).

ppl(*synset_id***,***w_num***,***synset_id***,***w_num***).**
The **ppl** operator specifies that the adjective first word is a participle of the verb second word. The reflexive operator can be implied.

per(*synset_id***,***w_num***,***synset_id***,***w_num***).**
The **per** operator specifies two different relations based on the parts of speech involved. If the first word is in an adjective synset, that word pertains to either the noun or adjective second word. If the first word is in an adverb synset, that word is derived from the adjective second word.

fr(*synset_id***,***f_num***,***w_num***).**
The **fr** operator specifies a generic sentence frame for one or all words in a synset. The operator is defined only for verbs.

Field Definitions

A *synset_id* is a nine byte field in which the first byte defines the syntactic category of the synset and the remaining eight bytes are a *synset_offset*, as defined in **wndb**(5WN), indicating the byte offset in the **data.***pos* file that corresponds to the syntactic category.

The syntactic category is encoded as:

1	NOUN
2	VERB
3	ADJECTIVE
4	ADVERB

w_num, if present, indicates which word in the synset is being referred to. Word numbers are assigned to the *word* fields in a synset, from left to right, beginning with 1. When used to represent lexical WordNet relations *w_num* may be 0, indicating that the relation holds for all words in the synset indicated by the preceding *synset_id*. See **wninput**(5WN) for a discussion of semantic and lexical relations.

ss_type is a one character code indicating the synset type:

n	NOUN
v	VERB
a	ADJECTIVE
s	ADJECTIVE SATELLITE
r	ADVERB

sense_number specifies the sense number of the word, within the part of speech encoded in the *synset_id*, in the WordNet database.

word is the ASCII text of the word as entered in the synset by the lexicographer. The text of the word is case sensitive. An adjective *word* is immediately followed by a syntactic marker if one was specified in the lexicographer file.

sense_key specifies the WordNet sense key for a given word sense. See **senseidx**(5WN) for the specifications.

syntax is the syntactic marker for a given adjective sense if one was specified in the input files. See **wninput**(5WN) for a list of the syntactic markers. Note that in the Prolog database, the parentheses are not included.

Each synset has a *gloss* that contains a definition and one or more example sentences.

class_type indicates whether the classification relation represented is topical, usage, or regional, as indicated by the *class_type* of **t**, **u**, or **r**, respectively.

f_num specifies the generic sentence frame number for word *w_num* in the synset indicated by *synset_id*. Note that when *w_num* is **0**, the frame number applies to all words in the synset. If non-zero, the frame applies to that word in the synset.

In WordNet, sense numbers are assigned as described in **wndb**(5WN). *tag_count* is the number of times the sense was tagged in the Semantic Concordances, and **0** if it was not instantiated.

NOTES

Since single forward quotes are used to enclose character strings, single quote characters found in *word* and *gloss* fields are represented as two adjacent single quote characters.

The load time can be greatly reduced by creating "object language" versions of the files, an option that is supported by some implementations, such as Quintus Prolog.

ENVIRONMENT VARIABLES (UNIX)

WNHOME Base directory for WordNet. Default is **/usr/local/WordNet-2.1**.

REGISTRY (WINDOWS)

HKEY_LOCAL_MACHINE\SOFTWARE\WordNet\2.1\WNHome
Base directory for WordNet. Default is **C:\Program Files\WordNet\2.1**.

FILES

All files are in **WNHOME/prolog** on Unix platforms and **WNHome\prolog** on Windows platforms

wn_s.pl	synset pointers
wn_sk.pl	sense keys
wn_syntax.pl	syntactic markers
wn_g.pl	gloss pointers
wn_hyp.pl	hypernym pointers
wn_ins.pl	instance pointers
wn_ent.pl	entailment pointers
wn_sim.pl	similar pointers
wn_mm.pl	member meronym pointers
wn_ms.pl	substance meronym pointers
wn_mp.pl	part meronym pointers
wn_der.pl	derivational morphology pointers
wn_cls.pl	class (domain) pointers
wn_cs.pl	cause pointers
wn_vgp.pl	grouped verb pointers

wn_at.pl	attribute pointers
wn_ant.pl	antonym pointers
wn_sa.pl	see also pointers
wn_ppl.pl	participle pointers
wn_per.pl	pertainym pointers
wn_fr.pl	frame pointers

SEE ALSO

wndb(5WN), **wninput**(5WN), **senseidx**(5WN), **wngroups**(7WN), **wnpkgs**(7WN).