Description

Hindi WordNet

The Hindi WordNet is a system for bringing together different lexical and semantic relations between the Hindi words. It organizes the lexical information in terms of word meanings and can be termed as a lexicon based on psycholinguistic principles. The design of the Hindi WordNet is inspired by the famous English WordNet.

In the Hindi WordNet the words are grouped together according to their similarity of meanings. Two words that can be interchanged in a context are synonymous in that context. For each word there is a synonym set, or synset, in the Hindi WordNet, representing one lexical concept. This is done to remove ambiguity in cases where a single word has multiple meanings. Synsets are the basic building blocks of WordNet. The Hindi WordNet deals with the content words, or open class category of words. Thus, the Hindi WordNet contains the following category of words- Noun, Verb, Adjective and Adverb.

Each entry in the Hindi WordNet consists of following elements

1. **Synset**: It is a set of synonymous words. For example, “विद्यालय, पाठशाला, स्कूल” (vidyaalay, paaThshaalaa, skuul) represents the concept of school as an educational institution. The words in the synset are arranged according to the frequency of usage.

2. **Gloss**: It describes the concept. It consists of two parts:
   
   **Text definition**: It explains the concept denoted by the synset. For example, “वह ठान जहाँ प्राथमिक या माध्यमिक स्तर की औपचारिक शिक्षा दी जाती है” (vah sthaan jahaan praathamik yaa maadhyamik star kii aupachaarik sikshaa dii jaatii hai) explains the concept of school as an educational institution.

   **Example sentence**: It gives the usage of the words in the sentence. Generally, the words in a synset are replaceable in the sentence. For example, “इस विद्यालय में पहली से पाँचवी तक की शिक्षा दी जाती है” (is vidyaalay men pahalii se paanchaviin tak kii shikshaa dii jaatii hai) gives the usage for the words in the synset representing school as an educational institution.

3. **Position in Ontology**: An ontology is a hierarchical organization of concepts, more specifically, a categorization of entities and actions. For each syntactic category namely noun, verb, adjective and adverb, a separate ontological hierarchy is present.

   Each synset is mapped into some place in the ontology. A synset may have multiple parents. The ontology for the synset representing the concept school is shown in figure 1.
Relations in Hindi WordNet

A WordNet is a word sense network. A word sense node in this network is a synset which is regarded as a basic object in the WordNet. Each synset in the Hindi WordNet is linked with other synsets through the well-known lexical and semantic relations of hypernymy, hyponymy, meronymy, troponymy, antonymy, entailment etc. Semantic relations are between synsets and lexical relations are between words. These relations serve to organize the lexical knowledge base.

There are 16 relations in the Hindi WordNet. These relations are described below.

**Hyponymy and Hypernymy** (*is a kind of*): Hypernymy is a semantic relation between two synsets to capture super-set hood. Similarly, hyponymy is a semantic relation between two synsets to capture sub-set hood. The hyponymy relation is transitive and asymmetrical. Hypernymy is the reverse of hyponymy.

**Example:**
बेलपत्र, बेल-पत्र, बेलपत्ती, बिल्वपत्र (bel patra, bel-patra, belpattii, bilvapatr; a leaf of a tree named bela)

==⇒ पत्ता, पात, पर्ण, पत्र, दल (pattaa, paat, parN, patr, dal; leaf)

Here, बेलपत्र (bel patra; a leaf of a tree named bela) is a kind of पत्ता (pattaa; leaf) means पत्ता (pattaa; leaf) is a hyponym and बेलपत्र (bel patra; a leaf of a tree named bela) is the hypernym.

**Meronymy and Holonymy** (*Part-whole relation*): It is a semantic relation between two synsets. If the concepts A and B are related in such a manner that A is one of the constituent of B, then A is the meronym of B and B is the holonym of A. The meronomy relation is transitive and asymmetrical. Holonomy is the reverse of meronomy. It is used to construct a *part-of* hierarchy.

**Example:**
Here, जड़ (jaR; root) is the part of पेड़ (peR; tree), meaning that जड़ (jaR; root) is the meronym of पेड़ (peR; tree) and पेड़ (peR; tree) is the holonym of जड़ (jaR; root).

**Entailment:** Entailment refers to a relationship between two verbs. Any verb A entails B, if the truth of B follows logically from the truth of A. The relation of entailment is unilateral, i.e., it is one way relation.

Example:

खराणटा लेना, नाक बजाना (kharraaTaa lenaa, naak bajaana; snore) 

===> सोना (sonaa; sleep)

**Troponymy:** Troponym denotes a specific manner elaboration of another verb. It shows manner of an action, i.e., X is a troponym of Y if to X is to Y in some manner.

Example:

मुस्कुराना, मुस्कराना, मुस्काना (muskuraanaa, muskaraanaa, muskaanaa; smile) 

===> हाँसना, विहाँसना (hansnaa, vihansnaa laugh)

**Antonymy:** Antonymy is a relation that holds between two words that (in a given context) express opposite meanings. It is a lexical relation as it holds between two words and not the entire synset.

Example:

मोटा, स्थूलकाय (moTaa, sthuulkaay; fat) 

===> पतला, दुबला, दुबला-पतला, छरहरा (patlaa, dublaa, dublaa-patlaa, charharaa; thin)

The words in bold face in the synset are in antonymy relation.

**Gradation:** Gradation is a lexical relation. It represents the intermediate concept between two opposite concepts. Figure 2 shows the gradation relation among three words.

![Figure 2. Gradation relation](image-url)
Causative: In Hindi, there is a convention of forming causation by making morphological change in the base verb. The Causative relation links the causative verbs and the base verbs and show interdependency between them.
Example:

खाना (khaanaa ; eat)

==> खिलाना (khilaanaa; to make someone to eat)

खिलाना (khilaanaa; to make someone to eat) is a causative verb of खाना (khaanaa ; eat).

Cross parts of speech linkage: Following relations are between the synsets of different parts of speech.

Linkages between nominal and verbal concepts

Ability Link: This link specifies the inherited features of a nominal concept. This is a semantic relation.
Example:

मछली, मछली, मछली, मछली मछली (machlii, macchii, matsya, miin, maahii; fish)

==> तैरना, पैरना, पौरना, पौरना, हेलना (tairnaa, pairnaa, pauMrnaa, paurnaa, helanaa; swim)

Capability Link: This link specifies the acquired features of a nominal concept. This is a semantic relation.
Example:

व्यक्ति, मानस, शाख्स, शाख्स, जन (vyakti, maanas, sakhs, jan; person)

==> तैरना, पैरना, पौरना, पौरना, हेलना (tairnaa, pairnaa, pauMrnaa, paurnaa, helanaa; swim)

Function Link: This link specifies the function of a nominal concept. This is a semantic relation.
Example:

अध्यापक, शिक्षक, आचार्य, गुरु, मास्टर (adhyaapak, shikshak, aacaarya, guru, master; teacher)

==> पढाना, शिक्षा देना (paRhaanaa, shikshaa denaa; teach)
Linkage between nominal and adjectival concepts

**Attribute:** This denotes the properties of noun. It is a linkage between noun and an adjective. This is a semantic relation.

Example:

पक्षी,चिड़िया,पंढी,खग,परिदा,विहंग,विहंगम,पखेरु,विहंग (pakshi, ciRiyaa, panchi, khag, parindaa, vihanga, vihangam, pakhru, vihaga; *bird*)

==>

पंखदार,पांखदार,पंखयुक्त (pankhdaar, paankhdaar, pankhyukt; *having wings*)

**Modifies Noun:** Certain adjectives can only modify certain nouns. Such adjectives and nouns are linked in the Hindi WordNet by the relation *Modifies Noun*.

Example:

सुपात्र,सत्पात्र,अच्छा पात्र (supaatra, satpaatra, acchaa paatra, *eligible*)

==>

व्यक्ति,मानस,शख़्स,शख्सा,जन,बंदा,बन्दा (vyakti, maanas, sakhs, jan; *person*)

Linkage between adverbial and verbal concepts

**Modifies Verb:** Certain adverbs can only go with certain verbs. *Modifies Verb* is a relation to show connection between such words.

Example:

कभी,ककसी समय (kabhii, kisii samay; *sometimes*)

==>

काम करना,कायण करना (kaam karnaa, kaarya karnaa; *to work*)

**Derived From:** This relation specifies the root form from which a particular word is derived. This relation can go from noun to adjective or vice versa, noun to verb and adjective to verb and aims to handle derivational morphology. This is a lexical relation.

Example:

क्रिशः,क्रिानुसार,यथाक्रि,मसलमसलेिार,बारी-बारी से,क्रििार (kramashaH, kramaanusaar, yathaakram, silsilevaar, baarii-baarii se, kramvaar; *step by step*)

==>

क्रम,सिलसिला,शृंखला,अनुक्रि,अनुक्रिखर्का (kram, silsilaa, shrinkhalaa, anukram, anukramaNikaa; *series*)
Hindi WordNet is still under construction. In the version 1.0 we have attempted to cover all the common concepts in Hindi. The present status is as follows:

Total number of synset: **28,687**
Total number of unique words: **63,800**

Figure 3. Snapshot web interface for Hindi WordNet (Hindi version)
Figure 4. Snapshot of web interface for Hindi WordNet (Synset Display)

Figure 5. Snapshot of web interface for Hindi WordNet (Relations display)

The Application Programming Interface for the Hindi Wordnet
The Hindi Wordnet data can be accessed by using APIs (Application Programming Interface) written in Java called JHWNL or Java Hindi WordNet Library. These APIs allow searching of synsets containing a particular word and accessing the relations of the synsets. The most important functions in the API are described below:

<table>
<thead>
<tr>
<th>Class : Dictionary</th>
<th>Return Type</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synset</td>
<td>getSynsetAt(POS pos, long synsetId)</td>
<td>Return the Synset with the given SynsetId and POS (Part of Speech).</td>
<td></td>
</tr>
<tr>
<td>IndexWord</td>
<td>getIndexWord(POS pos, String lemma)</td>
<td>Return an IndexWord, which can be used to access all the Synsets with the specified POS containing the lemma as a word.</td>
<td></td>
</tr>
<tr>
<td>IndexWord</td>
<td>lookupIndexWord(POS pos, String lemma)</td>
<td>Returns all Synsets with the specified POS containing the root form of lemma as a word. Morphed forms of words can be supplied to this method.</td>
<td></td>
</tr>
<tr>
<td>Synset</td>
<td>getOntoSynset(long ontoId)</td>
<td>Return the Ontology Hierarchy from the given Ontology-Id.</td>
<td></td>
</tr>
<tr>
<td>IndexWordSet</td>
<td>lookupAllIndexWords(String lemma)</td>
<td>Return a set of IndexWords, with each element in the set corresponding to all POS of the lemma in which synsets are present.</td>
<td></td>
</tr>
<tr>
<td>IndexWordSet</td>
<td>lookupMorphedIndexWord(POS pos, String lemma)</td>
<td>Returns a set of IndexWord for all root forms of the lemma for the specified POS.</td>
<td></td>
</tr>
<tr>
<td>IndexWordSet</td>
<td>lookupAllMorphedIndexWords(String lemma)</td>
<td>Returns a set of IndexWord for all root forms of the lemma for all POS in which synsets are present.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Classes</th>
<th>Class</th>
<th>Return Type</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synset</td>
<td>Pointer[]</td>
<td>getPointers()</td>
<td>Return an array of pointers for this Synset, which can be used to access all relations given in the document earlier. A specific type of relation can be selected by comparing them with the static members of the PointerTarget Class.</td>
<td></td>
</tr>
<tr>
<td>Pointer Target</td>
<td>Pointer Target</td>
<td>getTarget()</td>
<td>Gets the target of this pointer, generally a Synset for semantic relations, and a Word for lexical relations</td>
<td></td>
</tr>
</tbody>
</table>