

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 स्थान 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.

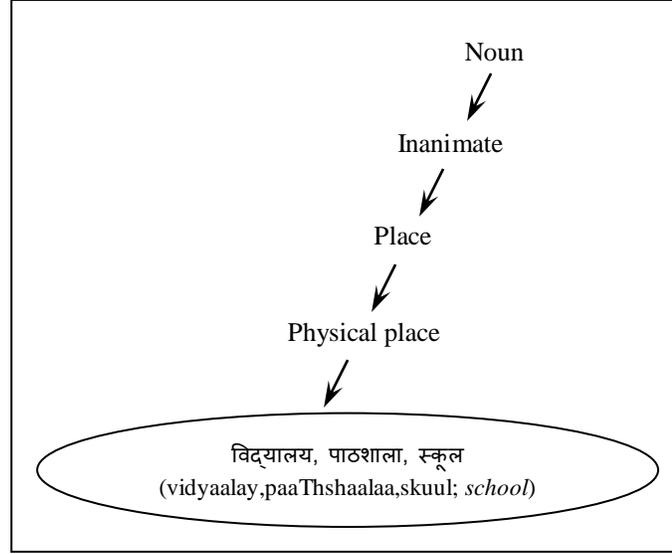


Figure1. Ontology for the synset of *school*

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 *hyponymy*, *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 patr, bel-patr, 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 hypernym and *बेलपत्र* (bel patra; *a leaf of a tree named bela*) is the hyponym.

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 meronymy relation is transitive and asymmetrical. Holonymy is the reverse of meronymy. It is used to construct a *part-of* hierarchy.

Example:

जड़, मूल, सोर (jaR, muul, sor; *root*)

==> पेड़, वृक्षा, पादप, द्रुम, तरु, विटप, रूख, अघ्रिप, अग (peR, vriksh, paadap, drum, taruu, viTap, ruukh, ruuMkh, aghrip, ag; *tree*)

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:

खर्राटा लेना, नाक बजाना (kharraaTaa lenaa, naak bajaanaa; *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:

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

==> हँसना, विहँसना (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.

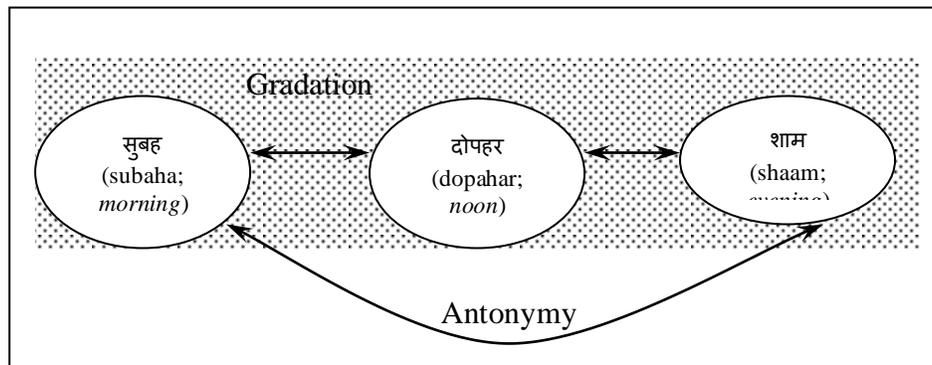


Figure 2. Gradation relation

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, मच्छी, मत्स्य, मीन, माही (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, शख्स, शख्स, 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 (adhyapak, shikshak, aacaarya, guru, master; *teacher*)

==> पढ़ाना, 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, पंछी, खग, परिंदा, विहंग, विहंगम, पखेरू, विहंग (pakshii, ciRiyaa, panchi, khag, parindaa, vihangaa, vihangam, pakheru, vihaga; *bird*)
==> पंखदार, पाँखदार, 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, शख्स, शख्स, 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, किसी समय (kabhi, kisii samay; *sometimes*)
==> काम करना, कार्य 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:

क्रमशः, क्रमानुसार, यथाक्रम, सिलसिलेवार, बारी-बारी से, क्रमवार (kramashaH, kramaanusaaar, yathaakram, silsilevaar, baarii-baarii se, kramvaar; *step by step*)
==> क्रम, सिलसिला, शृंखला, अनुक्रम, अनुक्रमणिका (**kram, silsilaa, shrinkhalaa, anukram, anukramaNikaa; series**)

Current Status Hindi WordNet

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**

हिन्दी शब्दतंत्र Hindi Wordnet

Total unique words:46001 | Total Synsets: 21944 | last updated 28 Jun 2006

Search Hindi Word(हिन्दी शब्द खोजें) खोजें(Search)

KEYBOARD/की-बोर्ड

Space Backspace Reset Collapse

अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ	अः	आः	इः
ए	ऐ	उ	ऊ	ऋ	ॠ	ऌ	ॡ	ऋः	ॠः	ऌः	ॡः	ॠ
क	ख	ग	घ	ङ	च	छ	ज	झ	ञ			
ट	ठ	ड	ढ	ण			त	थ	द	ध	न	
प	फ	ब	भ	म			य	र	ल	व		
श	ष	स							ह	ळ		
ॠ	ॡ	ऋ								ॠ	ॡ	।

उदाहरण Examples सहायता Help प्रतिक्रिया दें Give feedback प्रतिक्रियाएँ देखें Previous feedbacks लॉग देखिए View log

पुराना इंटरफेस Previous interface मराठी शब्दबंध Marathi Wordnet हिन्दी-अंग्रेजी शब्दकोश Hindi-English Dictionary

सी.एफ.आई.एल्टी. मुख्दपृष्ठ CFILT home हिन्दी शब्दतंत्र Hindi Wordnet

Figure 3. Snapshot web interface for Hindi WordNet (Hindi version)

हिन्दी शब्दतंत्र Hindi Wordnet

आम...

Noun(2)

- (R) आम, रसाल, आम, अंब, अम्ब - एक फल जो खाया या चूसा जाता है "तोता पेड़ पर बैठकर आम खा रहा है। शास्त्रों ने आम को इन्द्रासनी फल की संज्ञा दी है"
- (R) आम, आम वृक्ष - एक बड़ा पेड़ जिसके फल खाए या चूसे जाते हैं "आम की लकड़ी का उपयोग साज-सज्जा की वस्तुएँ बनाने में किया जाता है"

Adjective(2)

- (R) सामान्य, आम, साधारण, कामचलाऊ, मामूली, अविशिष्ट, अविशेष, अदिव्य - जिसमें कोई विशेषता न हो या अच्छे से कुछ हल्के दरजे का "यह सामान्य साड़ी है"
- (R) सामूहिक, आम, सार्वजनिक, सामुदायिक, सामान्य - प्रायः सभी व्यक्तियों, अवसरों, अवस्थाओं आदि में पाया जानेवाला या उनसे संबंध रखनेवाला "साक्षरता पर विचार-विमर्श हेतु एक सामूहिक सभा का आयोजन किया गया"

Total unique words: 46001 | Total Synsets: 21944 | last updated 28 Jun 2006

Search Hindi Word(हिन्दी शब्द खोजें): आम

KEYBOARD/की-बोर्ड

उदाहरण Examples	सहायता Help	प्रतिक्रिया दें Give feedback	प्रतिक्रियाएँ देखें Previous feedbacks	लॉग देखिए View log
पुराना इंटरफेस Previous interface	मराठी शब्दबंध Marathi Wordnet	हिन्दी-अंग्रेजी शब्दकोश Hindi-English Dictionary		
सी एफ् आई प्लू टी. मुम्बयपुठ CFILT home		हिन्दी शब्दतंत्र Hindi Wordnet		

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

हिन्दी शब्दतंत्र Hindi Wordnet

आम...

Noun(2)

- (R) आम, रसाल, आम, अंब, अम्ब - एक फल जो खाया या चूसा जाता है "तोता पेड़ पर बैठकर आम खा रहा है। शास्त्रों ने आम को इन्द्रासनी फल की संज्ञा दी है"
 - Ontology Nodes
 - Hyponymy (is a kind of ...)
 - (R) आष फल, आष-फल, आषफल - यह फल जो खाया जाता हो "आम एक आष फल है"
 - (R) फल, फल, प्रसून - वनस्पति में होने वाला गूदे या बीज से भरपूर बीजकोष जो किसी विशिष्ट भ्रतु में फूल आने के बाद उत्पन्न होता है "उसने फल की दुकान से एक किलो आम खरीदा"
 - (R) वनस्पति अंग, वनस्पति अवयव - पेड़-पौधे आदि की एक कार्यात्मक और रचनात्मक इकाई "जङ्गल में फलदायक वनस्पति अंग हैं"
 - (R) वनस्पति शाखा, पेड़-पौधे का शाखा, वनस्पति का शाखा - किसी पेड़-पौधे का कोई भाग "तलाऊड़ आदि वनस्पति शाखा हैं"
 - (R) प्राकृतिक वस्तु, नैसर्गिक वस्तु - वह वस्तु जो प्रकृति से संबंधित हो या प्रकृति द्वारा निर्मित हो "प्राकृतिक वस्तुओं का दुरुपयोग नहीं करना चाहिए"
 - (R) वस्तु, चीज, चीज - वास्तविक या कल्पित सत्ता "इस एक अमूर्त वस्तु है"
 - (R) अस्तित्व, मौजूदगी, मौजूदगी, वजूद, वजूद, संभूति, विद्यमानता, सत्ता, हस्ती - सत्ता का भाव "कभी-कभी हमारे मन में यह प्रश्न उठता है कि क्या ईश्वर का अस्तित्व है"
 - Hyponymy (... is a kind of)
 - Meronymy - Component Object (part of ...)
 - Holonymy - Component Object (is a part of)
 - Holonymy - Stuff Object (is a part of)
- (R) आम, आम वृक्ष - एक बड़ा पेड़ जिसके फल खाए या चूसे जाते हैं "आम की लकड़ी का उपयोग साज-सज्जा की वस्तुएँ बनाने में किया जाता है"

Adjective(2)

- (R) सामान्य, आम, साधारण, कामचलाऊ, मामूली, अविशिष्ट, अविशेष, अदिव्य - जिसमें कोई विशेषता न हो या अच्छे से कुछ हल्के दरजे का "यह सामान्य साड़ी है"
- (R) सामूहिक, आम, सार्वजनिक, सामुदायिक, सामान्य - प्रायः सभी व्यक्तियों, अवसरों, अवस्थाओं आदि में पाया जानेवाला या उनसे संबंध रखनेवाला "साक्षरता पर विचार-विमर्श हेतु एक सामूहिक सभा का आयोजन किया गया"

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:

<i>Class : Dictionary</i>			
<i>Return Type</i>		<i>Name</i>	<i>Description</i>
Synset		getSynsetAt(POS pos, long synsetId)	Return the Synset with the given <i>Synset-Id</i> and <i>POS</i> (Part of Speech).
IndexWord		getIndexWord(POS pos, String lemma)	Return an IndexWord, which can be used to access all the Synsets with the specified <i>POS</i> containing the <i>lemma</i> as a word.
IndexWord		lookupIndexWord(POS pos, String lemma)	Returns all Synsets with the specified <i>POS</i> containing the root form of <i>lemma</i> as a word. Morphed forms of words can be supplied to this method.
Synset		getOntoSynset(long ontoId)	Return the Ontology Hierarchy from the given <i>Ontology-Id</i> .
IndexWordSet		lookupAllIndexWords(String lemma)	Return a set of IndexWords, with each element in the set corresponding to all <i>POS</i> of the <i>lemma</i> in which synsets are present.
IndexWordSet		lookupMorphedIndexWord(POS pos, String lemma)	Returns a set of IndexWord for all root forms of the <i>lemma</i> for the specified <i>POS</i> .
IndexWordSet		lookupAllMorphedIndexWords(String lemma)	Returns a set of IndexWord for all root forms of the <i>lemma</i> for all <i>POS</i> in which synsets are present.
<i>Other Classes</i>			
<i>Class</i>	<i>Return Type</i>	<i>Name</i>	<i>Description</i>
Synset	Pointer[]	getPointers()	Return an array of pointers for this <i>Synset</i> , 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 <i>PointerTarget</i> Class.
Pointer	Pointer Target	getTarget()	Gets the target of this pointer, generally a <i>Synset</i> for semantic relations, and a <i>Word</i> for lexical relations