Hindi WordNet Data and Associated Software License Agreement

The Indian Institute of Technology Bombay (hereinafter referred to as "IIT Bombay"), represented for the purpose of the signature of this agreement by **the Dean of Research and Development, IIT Bombay or by his authorized representative Dr. Pushpak Bhattacharyya, Professor of the Department of Computer Science and Engineering, IIT Bombay**

and

(hereinafter referred to as "the Licensee")
Who is adhering to the terms and conditions of this agreement in order to use the Hindi WordNet.

Have agreed as follows:

Article I - Definition of the Hindi WordNet

> The Hindi WordNet is a repository of Hindi words connected by lexical and semantic relations along with the browsing interface and associated software (see schedule A too)

Article II – Ownership and License

1. IIT Bombay is and at all times shall remain the sole and exclusive owner of a set of linguistic and non-linguistic data and support documentation along with the browsing interface and associated software, which together constitute the Hindi WordNet.

2. IIT Bombay grants the Licensee the non-exclusive, non transferable right to use, re-work and build upon the Hindi WordNet according to the terms of this agreement and only to the extent required for the Licensee to carry out works specified in Article III below of the present agreement.

Article III - Purpose of the License

> The Licensee shall endeavor to use the Hindi WordNet only for the purposes of research and development.

Article IV - Delivery of the Hindi WordNet and site of use

1. *IIT Bombay shall provide the Licensee with the data files of the Hindi WordNet and the associated software along with the associated software as specified in "Schedule A" of this Agreement, attached hereto and incorporated herein.*

2. The website where the Hindi WordNet is installed has the URL as
   www.cfilt.iitb.ac.in

3. The Hindi WordNet shall not be launched at or transferred to, or accessed by, any disc-based computer system at any other site without prior permission in writing from IIT Bombay.

Article V - Charges
1. No charges shall be imposed by IIT Bombay to the Licensee for the use of the Hindi WordNet for the purposes described in Article III of this Agreement
2. Any use of the Hindi WordNet beyond the scope of those specified in Article III shall be the subject of a separate Agreement
3. Any use of the Hindi WordNet, or parts thereof, in any documentation, application or service, where that documentation, application or service is charged for by the Licensee, shall be the subject of a separate Agreement between IIT Bombay and the Licensee.

Article VI - Ownership, unauthorized copy and use
1. The Licensee acknowledges the copyrights and any other proprietary rights protecting the Hindi WordNet in all its publications and communications.
2. The Licensee hereby agrees not to assign, pledge, lend, create security interest in, sell, trade, copy, or surrender possession of the Hindi WordNet, nor shall the Licensee permit any claims, liens, or encumbrances to be incurred thereon. The restrictions of use specified in Article III shall also apply to this article.
3. The source files of the Hindi WordNet API are licensed under GNU General Public License (GPL), as published by the Free Software Foundation; either version 2 or (at your option) any later version.

Article VII - Exclusion of warranties
Other than as stated in this Agreement, the Hindi WordNet provides no warranties, explicit or implicit, including, but not limited to, any implied warranty of fitness for a particular purpose.

Article VIII - Limitation of liability
In no event shall IIT Bombay be liable for the loss of profits or indirect, special, or consequential damages of any nature by the Licensee or any third party, even if they have been advised of the possibility of such damages. It is agreed that the remedies of the Licensee set forth in this Agreement shall be the exclusive remedies of the Licensee.

Article IX - Indemnification
The Licensee agrees to indemnify, defend, and save IIT Bombay from and against any and all claims, liabilities, costs, and expenses which IIT Bombay may incur, suffer, or is required to pay by reason of claims or suits of a third party arising out of or related to information processed by the Hindi WordNet and/or the Licensee's negligence or willful misconduct with respect to the Hindi WordNet.

Article X - Arbitration and entire agreement
Any dispute or controversy arising out of or relating to this Agreement, its construction or its actual or alleged breach will be decided by ARBITRARTION. If the mediation does not result in a resolution of such dispute or controversy, it will be finally decided by an appropriate method of alternate dispute resolution, including without limitation, arbitration, conducted in the city of Mumbai, India in accordance with the Laws of India.

In cases of disputes arising out of or pertaining to this agreement .1 or more persons shall be appointed as arbitrator/s mutually by consent of both the parties, and (that ARBITRATION shall take place at a place mutually agreed upon (preferably Mumbai) and also the decision given by such ARBITRATOR/S WILL BE FINAL AND BINDING ON BOTH THE PARTIES. and no appeal shall lie in any court of law thereafter.

SCHEDULE A

The proprietary package known as 'Hindi WordNet' including data, associated software, browsing interface and related manuals (please see below).

AGREED:

For	For
IIT Bombay	The Licensee
Signature	signature

	Name:
Title: Dean, Research and Development	Title:
Address: IIT Bombay	Address
Mumbai- 400 076	
India	
E-mail: deanrnd@admin.iitb.ac.in	Email:

Date:	Date:

Schedule A

Hindi WordNet Documentation

The *CD* contains the Hindi WordNet – a lexical database. To use the Hindi WordNet CD, please read the copyright notice and the documentation below. References to the relevant papers are also included.

The Hindi WordNet is an ongoing effort. To browse the latest version of the Hindi WordNet, Please visit http://www.cfilt.iitb.ac.in/WordNet/webhwn/

COPYRIGHT NOTICE

© 2006, CFILT, IIT Bombay, Mumbai, All Rights Reserved

We would like to thank the Hindi WordNet team members for their valuable contribution.
Hindi WordNet Team: Pushpak Bhattacharyya, Prabhakar Pandey, Laxmi Kashyap, Debasri Chakrabarti, Manish Sinha, Mahesh Kumar Reddy, Nitin Verma, Arati Sharma, Madhu Sharma, Dipak Kumar Narayan and Sanjay Kumar Jha.

References
- Manish Sinha, Mahesh Kumar Reddy and Pushpak Bhattacharyya, *An Approach towards Construction and Application of Multilingual Indo-WordNet*, 3rd Global WordNet Conference (GWC 06), Jeju Island, Korea, January, 2006.
- Manish Sinha, Mahesh Kumar Reddy, Prabhakar Pande, Laxmi Kashyap and Pushpak Bhattacharyya, *Hindi Word Sense Disambiguation,* International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems, Delhi, India, November 2004.
- N. Verma and P. Bhattacharyya, *Automatic Generation of Multilingual Lexicon by using WordNet,* International Conference on Convergence of Knowledge, Culture, Language and Information Technology. Library of Alexandria, Egypt, December 2003.
- N. Verma and P. Bhattacharyya, *Automatic Lexicon Generation through WordNet,* Global WordNet Conference (GWC-2004), Czech Republic, January 2004.

- D. Chakrabarti and P. Bhattacharyya, *Creation of English and Hindi Verb Hierarchies and their Application to Hindi WordNet Building and English-Hindi MT,* Global WordNet Conference (GWC-2004), Czech Republic, January 2004.
- G. Ramakrishnan, Prithviraj, A. Deepa, P. Bhattacharyya, and S. Chakrabarti, *Soft Word Sense Disambiguation,* Global WordNet Conference (GWC-2004), Czech Republic, January 2004.
- D. Narayan and P. Bhattacharyya, *Using Verb Noun Association for Word Sense Disambiguation*, International Conference on Natural Language Processing (ICON 2002), Mumbai, India, December 2002.
- Dipak Narayan, Debasri Chakrabarti, Prabhakar Pande and P. Bhattacharyya, *An Experience in Building the Indo WordNet - a WordNet for Hindi*, First International Conference on Global WordNet, Mysore, India, January 2002.
- S. Jha, D. Narayan, P. Pande, P. Bhattacharyya, *A WordNet for Hindi,* International Workshop on Lexical Resources in Natural Language Processing, Hyderabad, India, January 2001.

Description of the 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 jahaaM praathamik yaa maadhyamik star kii aupacaarik 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 पाँचवीं तक की शिक्षा दी जाती है" (is vidyaalay me

pahalii se pancvii 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.
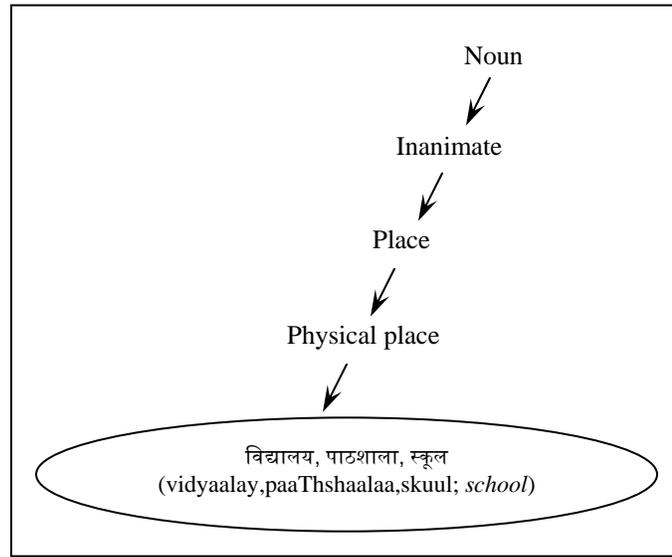
Figure1. Ontology for the synset of *school*

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 *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, bilvapatra; *a leaf of a tree named bela*)
    ==> पत्ता, पात, पर्ण, पत्र, दल (pattaa, paat, parNa, patra, 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:

> जड, मूल, सोर (jaR, muul, sor; *root*)
>> ==> पेड, वृक्ष, पादप, द्रुम, तरु, विटप, रूख, रूँख, अग्निप, अग (peR, vriksh, paadap, drum, taruu, viTap, ruukh, ruuMkh, aghrip, ag; *tree*)

Here, जड (jaR; *root*) is the part of पेड (peR ; *tree*), meaning that जड (jaR; *root*) is the meronym of पेड (peR ; *tree*) and पेड (peR ; *tree*) is the holonym of जड (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:

> मुस्कुराना, 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.
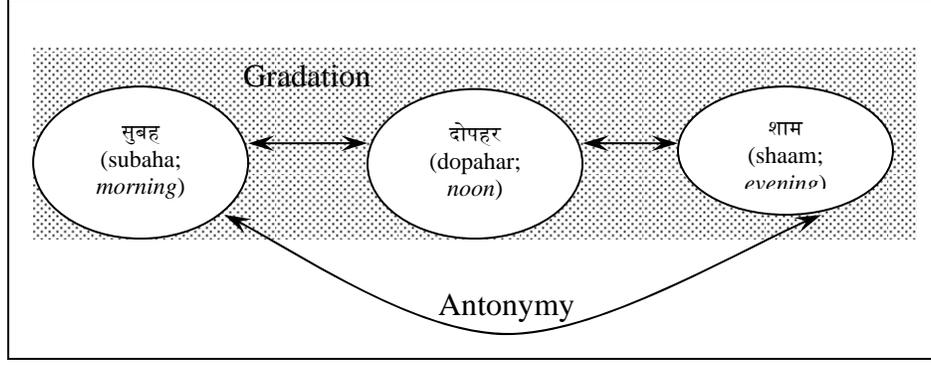
Figure 2. Gradation relation

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, पौंरना, 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,शख़्स,शख्स,जन (vyakti, maanas, sakhs, jan; *person*)
==> तैरना,पैरना,पौंरना,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*)
==> 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,विहग (pakshii, ciRiyaa, panchi, khag, parindaa, vihanga, 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,शख़्स,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 (kabhii, 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, kramaanusaar, yathaakram, silsilevaar, baarii-baarii se, kramvaar; *step by step*)

==> क्रम,सिलसिला,शृंखला,अनुक्रम,अनुक्रमणिका (kram, silsilaa, shrinkhalaa, anukram, anukramaNikaa; *series*)

Current Status of 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: **26,208**
Total number of unique words: **56,928**

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

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

KEYBOARD/की-बोर्ड

Space | Backspace | Reset | Collapse

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

हिन्दी शब्दतंत्र **Hindi Wordnet**

आम...
Noun(2)

1. (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 इंद्रासनी फल की संज्ञा दी है"*
2. (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)

1. (R) सामान्य, **आम**, साधारण, कामचलाऊ, मामूली, अविशिष्ट, अविशेष, अदिव्य - जिसमें कोई विशेषता न हो या अच्छे से कुछ हल्के दरज़े का   *"यह सामान्य साड़ी है"*
2. (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 खोजें(Search)

KEYBOARD/की-बोर्ड

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

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

सी.एफ़.आई.एल.टी. मुख्यपृष्ठ	हिन्दी शब्दतंत्र
CFILT home	Hindi Wordnet

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

हिन्दी शब्दतंत्र Hindi Wordnet

आम...

Noun(2)

1. (R) **आम**, रसाल, आब, अंब, अम्ब - एक फल जो खाया या चूसा जाता है   "तोता पेड़ पर बैठकर आम खा रहा है / शास्त्रों ने आम को इंद्रासती फल की संज्ञा दी है"
   - A. Ontology Nodes
   - B. Hypernymy (is a kind of ...)
     - (R) खाद्य फल, खाद्य-फल, खाद्यफल - वह फल जो खाया जाता हो "आम एक खाद्य फल है"
       - (R) फल, फर, प्रसून - वनस्पति में होने वाला गूदे या बीज से भरपूर बीजकोष जो किसी विशिष्ट ऋतु 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 "जाइलम और फ्लोयम वनस्पति अंग है"
           - (R) वनस्पति भाग, पेड़-पौधे का भाग, वनस्पति का भाग - किसी पेड़-पौधे का कोई भाग "तना, जड़ आदि वनस्पति भाग है"
             - (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, चीज़, चीज - वास्तविक या कल्पित सत्ता "हवा 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"
   - C. Hyponymy (... is a kind of)
   - D. Meronymy - Component Object (part of ...)
   - E. Holonymy - Component Object (is a part of)
   - F. Holonymy - Stuff Object (is a part of)

2. (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)

1. (R) सामान्य, **आम**, साधारण, कामचलाऊ, मामूली, अविशिष्ट, अविशेष, अदिव्य - जिसमें कोई विशेषता न हो या अच्छे से कुछ हल्के दरज़े का "यह सामान्य साड़ी है"

2. (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)