IndoWordNet Dictionary:
An Online Multilingual Dictionary using IndoWordNet

Hanumant Redkar
Center for Indian Language Technology,
Indian Institute of Technology Bombay, India
hanumantredkar@gmail.com

Sandhya Singh
Center for Indian Language Technology,
Indian Institute of Technology Bombay, India
sandhya.singh@gmail.com

Nilesh Joshi
Center for Indian Language Technology,
Indian Institute of Technology Bombay, India
joshinilesh60@gmail.com

Anupam Ghosh
Center for Indian Language Technology,
Indian Institute of Technology Bombay, India
anupam.ghsh@gmail.com

Pushpak Bhattacharyya
Center for Indian Language Technology,
Indian Institute of Technology Bombay, India
pushpakbh@gmail.com

Abstract

India is a country with diverse culture, language and varied heritage. Due to this, it is very rich in languages and their dialects. Being a multilingual society, a multilingual dictionary becomes its need and one of the major resources to support a language. There are dictionaries for many Indian languages, but very few are available in multiple languages. WordNet is one of the most prominent lexical resources in the field of Natural Language Processing and Information Retrieval, while IndoWordNet is an integrated multilingual WordNet for Indian languages. These resources are used by researchers to experiment and resolve the issues in multilinguality through computation. However, there are few cases where WordNet is used by the non-researchers or general public. This paper focuses on providing an online interface – *IndoWordNet Dictionary* to non-researchers as well as researchers. It is developed to render multilingual WordNet information of 19 Indian languages in a dictionary format. The WordNet information is rendered in multiple views such as: sense based, thesaurus based, word usage based and language based. English WordNet information is also rendered using this interface. The IndoWordNet dictionary will help language researchers as well as general public to know meanings of a word in multiple Indian languages.

1 Introduction

Language is a constituent element of civilization. In a country like India, diversity is its primary aspect. This leads to varied languages and their dialects. There are numerous languages in India which belong to different language families. These language families are Indo-Aryan, Dravidian, Sino-Tibetan, Tibeto-Burman and Austro-Asiatic. The major ones are the Indo-Aryan, spoken by most of Indians while Dravidian spoken by southern Indians. The Eighth Schedule of the Indian Constitution lists 22 languages, which have been referred to as scheduled languages and given recognition, status and official encouragement.

A Dictionary can be called as a resource dealing with the individual words of a language along with its orthography, pronunciation, usage, synonyms, derivation, history, etymology, *etc*. arranged in an order for convenience of referencing the words. Various criterions used for classifying this resource are - density of entries, number of languages involved, nature of entries, degree of concentration on strictly lexical data, axis of time, arrangement of entries, purpose, prospective user, *etc*. Some of the common types of dictionaries are[1]-

- **Encyclopedia:** Single or multi-volume publication that contains accumulated and authoritative knowledge on a subject arranged alphabetically. E.g. Britannica encyclopedia.
- **Thesaurus:** Thesaurus is a dictionary that lists words in groups of synonyms and related concepts.
- **Etymological Dictionary:** An etymological dictionary discusses the etymology/origin of the words listed. It is the product of research in historical linguistics.
- **Dialect Dictionary:** These dictionaries deal with the words of a particular geographical region or social group which are non standard.
- **Specialized Dictionary:** These dictionaries covers relatively restricted set of phenomena.
- **Bilingual or Multilingual Dictionary:** These are linguistic dictionaries in two or more languages.
- **Reverse Dictionary**: These dictionaries are based on the concept/idea/definition to words.
- **Learner's Dictionary:** These dictionaries are meant for foreign students/tourists to learn the usage of the word in language.
- **Phonetic Dictionary**: These dictionaries help in searching the words by the way they sound.
- **Visual Dictionary**: These dictionaries use pictures to illustrate the meaning of words.

WordNet is a lexical resource composed of synsets and semantic relations. Synsets are sets of synonyms. They are linked by semantic relations like hypernymy, meronymy, *etc.* and lexical relations like antonymy, gradation, *etc*. (Miller et al., 1990; Fellbaum, 1998). IndoWordNet[2] is a linked structure of WordNets of 19 different Indian languages from Indo-Aryan, Dravidian and Sino-Tibetan families (Bhattacharyya, 2010). Other popular multilingual WordNets are: EuroWordNet, which is a linked WordNet for European languages (Vossen, 1999) and BalkaNet, which is a linked WordNet for Balkan Languages (Christodoulakis, 2002). The most innovative aspect of WordNets is that lexical information is organized in terms of meaning; that is, a synset contains words of the same part-of-speech which have approximately the same meaning. Thus, it is synonymy that functions as the essential principle in the construction of WordNets (Vincze et al., 2008). This feature of WordNet is most important for the dictionary construction.

IndoWordNet is used in the field of Natural Language Processing tasks like Machine Translation, Information Retrieval, Information Extraction, *etc*. But, not much has been explored to use this resource beyond research labs. In this paper, we present an interface – *IndoWordNet Dictionary* (*IWN Dictionary*) in the form of multilingual online dictionary which uses IndoWordNet as a resource. The primary focus of this interface is to provide synset information in a systematic and classified manner which is rendered in multiple views.

The rest of the paper is organized as follows: Section 2 justifies the need of IndoWordNet Dictionary. Section 3 details the IndoWordNet Dictionary, its components, followed by its design and layout. Section 4 gives the features of the dictionary. Section 5 lists its limitations. Finally, the conclusion, scope and enhancements to the interface are presented.

2 Need for IndoWordNet Dictionary

Our work on developing IWN Dictionary interface is motivated from various available online resources. To name some: `langtolang.com`[3] which includes cross-lingual references across 47 non-Indian languages, `wordreference.com`[4] which includes 17 non-Indian languages, and others being `logosdictionary.org`[5] and `xobodo.org`[6] which has multiple languages including some Indian languages. But, all these resources render not more than two languages at a given instance.

Further survey is done, which reveals that Mohanty et al. (2008) had developed a tool for

[1] http://www.ciil-ebooks.net/html/lexico/link5.htm

[2] http://www.cfilt.iitb.ac.in/indowordnet/

[3] http://www.langtolang.com/

[4] http://www.wordreference.com

[5] http://www.logosdictionary.org/

[6] http://www.xobdo.org/

multilingual dictionary development process to create and link the synset based lexical resource for machine translation purpose. The aim was to simplify the process of synset creation and to link it with different Indian language WordNets. The tool was mainly used by lexicographers involved in the process of creating various Indian language WordNets. Also, Sinha et al. (2006) who have designed a browsable bilingual interface for viewing WordNet information in two languages Hindi and Marathi. The input to this browser is a search string in any of the two languages and the output is the search result for both the languages. The primary usage of this interface is to help users get the semantic information of the search string in both Hindi and Marathi. However, Sarma et al. (2012) built a multilingual dictionary considering three languages, *viz.*, Assamese, Bodo and Hindi. The dictionary interface allows searching between Hindi-Assamese and Hindi-Bodo language pairs at a time.

All these interfaces mentioned above could display the meanings in at most two languages with all the linguistic information available in the WordNet at a time. Hence, we have developed a web based interface to render multilingual WordNet information in a dictionary format. This interface is developed keeping in mind the general public, apart from researchers.

Following points justify the need of online multilingual dictionary in today’s time:

- To know or understand other languages.
- To find the meaning of a word in multiple languages on a single platform.
- To understand other languages with the help of a pivot language (referring one of the known languages).
- To understand synonymous words in input language as well as in another languages.
- To understand additional information like part-of-speech, gender, etc. in input as well as other different languages.
- To understand the semantic variations in different languages.
- To fulfil the social need of bridging the communication gap.
- To understand the script of various languages.
- To understand a local language when relocated to that area.
- To improve on one’s own language vocabulary.
- To address social and educational needs.

3 IndoWordNet Multilingual Dictionary

3.1 What is IndoWordNet Dictionary?

IndoWordNet Dictionary[7] or *IWN Dictionary* is an online interface to render multilingual IndoWordNet information in the dictionary format. It allows user to view the results in multiple formats as per the need. Also, user can view the result in multiple languages simultaneously. The look and feel of the IWN Dictionary is kept same as a traditional dictionary keeping in mind the user adaptability. So far, it renders WordNet information of 19 Indian languages. These languages are: Assamese, Bodo, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Odia, Punjabi, Sanskrit, Tamil, Telugu and Urdu. The WordNet information is also rendered in English. The English WordNet information is taken from Princeton University[8] website and imported into IndoWordNet database structure (Prabhu, et al., 2012). The data is imported to this database using DB Import tool developed under the Indradhanush WordNet Project[9]. The work is still going on to include other non-Indian languages and storing them in the World WordNet Database Structure proposed by Redkar et al. (2015).

3.2 Modules of IWN Dictionary

The IWN Dictionary has two major modules: IWN Dictionary Search Module and IWN Dictionary Display Module. They are explained in details in the next section. Figure 1 shows the block diagram of the IWN Dictionary.

3.2.1 IWN Dictionary Search Module

The main function of the IWN Dictionary Search module is to process the input word in the form which can be used to search the database and re-

[7] http://www.cfilt.iitb.ac.in/wordnet/iwndictionary

[8] https://wordnet.princeton.edu/

[9] http://indradhanush.unigoa.ac.in/

trieve relevant information. Its function can be split in three sub modules which involve analyzing a word, database lookup and multilingual data extraction. Each of these sub modules are described in details below -

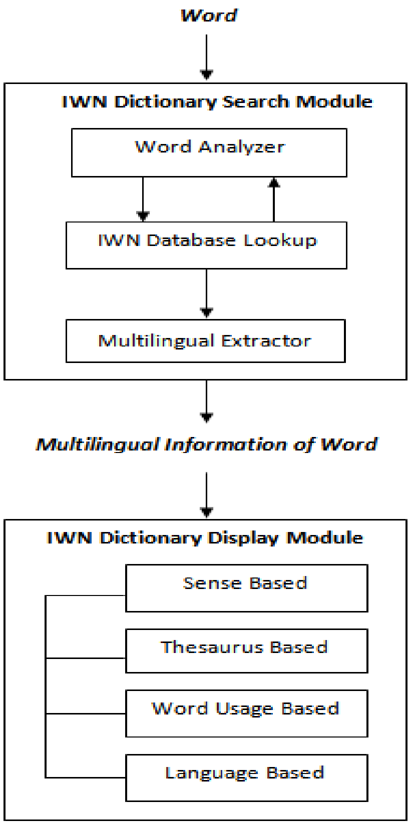


Figure 1 Block diagram of IWN Dictionary.

Word Analyzer

Word Analyzer analyses and processes the input word. It checks whether it is in its root form. If it is in its root form then it is directly passed on to the IWN Database lookup module, else it will be processed to find the closest possible word in the database. The concept of human mediated lemmatizer is adopted from the work of Bhattacharyya et al., (2014) to find the closest possible words of an input word. Here, the trie data structure is created out of the vocabulary in the input language and trie structure is navigated to find the match between the input word and an entry in trie. Accordingly, the closest possible word(s) is populated and given for the next module for the database lookup.

IWN Database Lookup

In IWN Database Lookup module, the word received from Word Analyzer is searched in the IWN database and the corresponding synset ids of all the senses of word are sent to the Multilingual Extractor engine for further processing. If the word is not found in the database then the control is sent back to the Word Analyzer module to look for other similar or closest words.

Multilingual Extractor

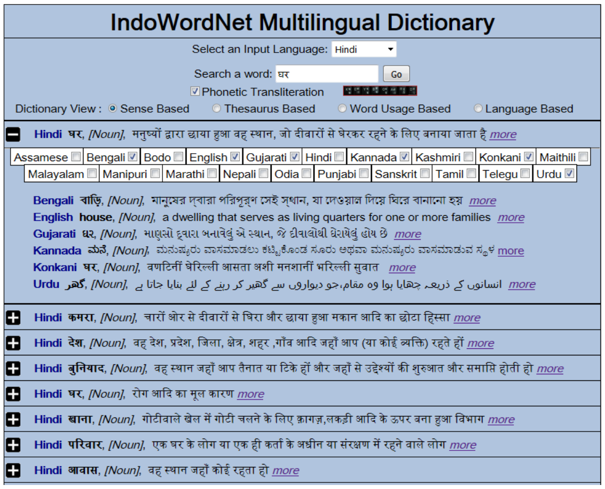
The basic task of Multilingual Extractor module is to take the input synset ids and extract synset information of all the languages available in the IWN database. This extracted synset information, i.e. the multilingual senses of an input word are then sent to the IWN Dictionary Display module for rending information in the dictionary format.

3.2.2 IWN Dictionary Display Module

The extracted multilingual senses of a given word are given to the IWN Dictionary Display module for rendering purpose. This is an important module of an IWN Dictionary which renders and ranks information based on the user response and statistics. Here, *IWN Click-Based-Rendering Algorithm* is used to rank most frequently used senses using this interface. The basic task of this algorithm is to record and maintain the number of user clicks on a sense being viewed and accordingly rank the sense in the output interface. Similarly, this algorithm is applied to all the other views mentioned below. The multilingual IndoWordNet data can be rendered using different views:

- **Sense Based view:** All the meanings of an input word are displayed with respect to the senses available in the IWN database. Figure 2 shows the sense based view of an IWN Dictionary.
- **Thesaurus Based view:** In thesaurus based view, synonymous words in each language are rendered. Here, user can click on any of the words in the list to go to see other senses of that word. This is also displayed in the thesaurus based view. Figure 3 shows the thesaurus based view of an IWN Dictionary.

- **Word Usage Based view:** In word usage based view, usage of an input word with respect to the languages is rendered. Here, the examples of a synset are rendered. Figure 4 shows the word usage based view of an IWN Dictionary.
- **Language Based view:** In language based view, meanings of a word are rendered with respect to the languages. Here, for each sense of a word, the meaning in all the languages is rendered in horizontal tabbed format. Figure 5 shows the language based view of an IWN Dictionary.


IndoWordNet Multilingual Dictionary

Select an Input Language: Hindi

Search a word: घर Go

☑ Phonetic Transliteration

Dictionary View : ● Sense Based ○ Thesaurus Based ○ Word Usage Based ○ Language Based

Hindi **घर**, *[Noun]*, मनुष्यों द्वारा छाया हुआ वह स्थान, जो दीवारों से घेरकर रहने के लिए बनाया जाता है *more*

Assamese ☐ | Bengali ☑ | Bodo ☐ | English ☑ | Gujarati ☑ | Hindi ☐ | Kannada ☑ | Kashmiri ☐ | Konkani ☑ | Maithili ☐
Malayalam ☐ | Manipuri ☐ | Marathi ☐ | Nepali ☐ | Odia ☐ | Punjabi ☐ | Sanskrit ☐ | Tamil ☐ | Telegu ☐ | Urdu ☑

Bengali **বাড়ি**, *[Noun]*, মানুষের দ্বারা পরিপূর্ণ সেই স্থান, যা দেওয়াল দিয়ে ঘিরে বানানো হয় *more*
English **house**, *[Noun]*, a dwelling that serves as living quarters for one or more families *more*
Gujarati **ઘર**, *[Noun]*, માણસો દ્વારા બનાવેલું એ સ્થાન, જે દીવાલોથી ઘેરાયેલું હોય છે *more*
Kannada **ಮನೆ**, *[Noun]*, ಮನುಷ್ಯರು ವಾಸಮಾಡಲು ಕಟ್ಟಿಕೊಂಡ ಸೂರು ಅಥವಾ ಮನುಷ್ಯರು ವಾಸಮಾಡುವ ಸ್ಥಳ more
Konkani **घर**, *[Noun]*, वणटिनीं घेरिल्ली आसता अशी मनशानीं भरिल्ली सुवात *more*
Urdu **گھر**, *[Noun]*, انسانوں کے ذریعہ چھایا ہوا وہ مقام،جو دیواروں سے گھیر کر رہنے کے لئے بنایا جاتا ہے *more*

Hindi **कमरा**, *[Noun]*, चारों ओर से दीवारों से घिरा और छाया हुआ मकान आदि का छोटा हिस्सा *more*

Hindi **देश**, *[Noun]*, वह देश, प्रदेश, जिला, क्षेत्र, शहर ,गाँव आदि जहाँ आप (या कोई व्यक्ति) रहते हों *more*

Hindi **बुनियाद**, *[Noun]*, वह स्थान जहाँ आप तैनात या टिके हों और जहाँ से उद्देश्यों की शुरुआत और समाप्ति होती हो *more*

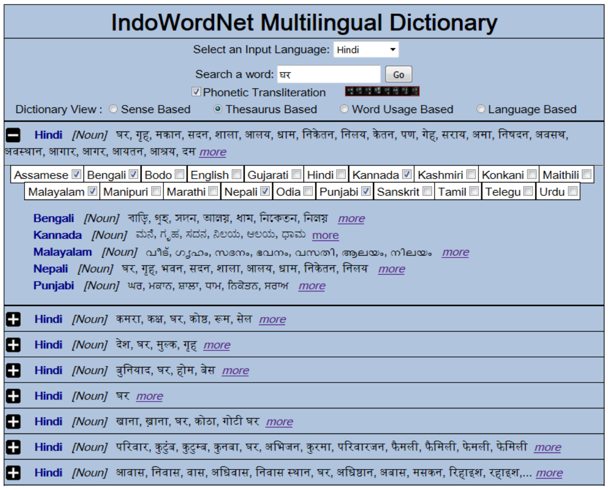
Hindi **घर**, *[Noun]*, रोग आदि का मूल कारण *more*

Hindi **खाना**, *[Noun]*, गोटीवाले खेल में गोटी चलने के लिए क़ागज़,लकड़ी आदि के ऊपर बना हुआ विभाग *more*

Hindi **परिवार**, *[Noun]*, एक घर के लोग या एक ही कर्ता के अधीन या संरक्षण में रहने वाले लोग *more*

Hindi **आवास**, *[Noun]*, वह स्थान जहाँ कोई रहता हो *more*

Figure 2 IndoWordNet Dictionary Interface showing multiple languages with sense based view.


IndoWordNet Multilingual Dictionary

Select an Input Language: Hindi

Search a word: घर Go

☑ Phonetic Transliteration

Dictionary View : ○ Sense Based ● Thesaurus Based ○ Word Usage Based ○ Language Based

Hindi *[Noun]* घर, गृह, मकान, सदन, शाला, आलय, धाम, निकेतन, निलय, केतन, पण, गेह, सराय, अमा, निषदन, अवसथ, अवस्थान, आगार, आगर, आयतन, आश्रय, दम *more*

Assamese ☑ | Bengali ☑ | Bodo ☐ | English ☐ | Gujarati ☐ | Hindi ☐ | Kannada ☑ | Kashmiri ☐ | Konkani ☐ | Maithili ☐
Malayalam ☑ | Manipuri ☐ | Marathi ☐ | Nepali ☑ | Odia ☐ | Punjabi ☑ | Sanskrit ☐ | Tamil ☐ | Telegu ☐ | Urdu ☐

Bengali *[Noun]* বাড়ি, গৃহ, সদন, আলয়, ধাম, নিকেতন, নিলয় *more*
Kannada *[Noun]* ಮನೆ, ಗೃಹ, ಸದನ, ನಿಲಯ, ಆಲಯ, ಧಾಮ more
Malayalam *[Noun]* വീട്, ഗൃഹം, സദനം, ഭവനം, വസതി, ആലയം, നിലയം *more*
Nepali *[Noun]* घर, गृह, भवन, सदन, शाला, आलय, धाम, निकेतन, निलय *more*
Punjabi *[Noun]* ਘਰ, ਮਕਾਨ, ਸ਼ਾਲਾ, ਧਾਮ, ਨਿਕੇਤਨ, ਸਰਾਅ *more*

Hindi *[Noun]* कमरा, कक्ष, घर, कोष्ठ, रूम, सेल *more*

Hindi *[Noun]* देश, घर, मुल्क, गृह *more*

Hindi *[Noun]* बुनियाद, घर, होम, बेस *more*

Hindi *[Noun]* घर *more*

Hindi *[Noun]* खाना, ख़ाना, घर, कोठा, गोटी घर *more*

Hindi *[Noun]* परिवार, कुटुंब, कुटुम्ब, कुनबा, घर, अभिजन, कुरमा, परिवारजन, फैमली, फैमिली, फेमली, फेमिली *more*

Hindi *[Noun]* आवास, निवास, वास, अधिवास, निवास स्थान, घर, अधिष्ठान, अबास, मसकन, रिहाइश, रहाइश,... *more*

Figure 3 IndoWordNet Dictionary Interface showing multiple languages with thesaurus based view.

Apart from these views, transliteration information is displayed for each of these views. The transliteration information is seen once a user hovers on any of the content in the dictionary view. For now, we have used transliteration into a roman script using the Indic NLP Library[10] developed by Kunchukuttan et al. (2015). However, we are in the process to implement transliteration in all the Indian languages, so that anyone can read other language in their chosen *transliteration language*.

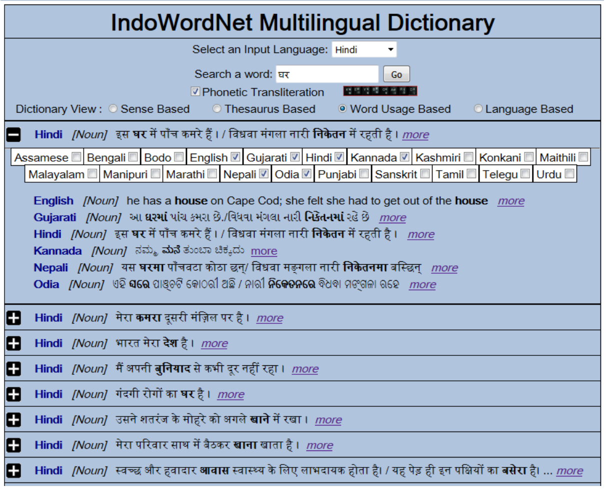
3.3 Design and Layout of IWN Dictionary

The IWN Dictionary is designed keeping in mind, its simplicity and usability, more importantly, the user friendliness of the system. The frontend is designed and developed with PHP, CSS, JavaScript, etc. and at the backend MySQL database is used. The IWN dictionary data is retrieved from IndoWordNet database using IndoWordNet APIs (Prabhugaonkar et al., 2012). Figure 2, figure 3, figure 4 and figure 5 shows sample screen shots of the IWN Multilingual Dictionary Interface with different views. In all these views, initial information is always shown in the source language. This initial information is rendered on the horizontal tab in all the views. User can click on plus or minus button to expand or collapse to see the details in each tab respectively. Also, user can check/uncheck the checkbox next to each Language in *sense based view*, *thesaurus based view* and *word usage based view* to see information in the checked language.

In figure 3, we can see that the source language is Hindi and the target languages are Bengali, Kannada, Malayalam, Nepali and Punjabi. Here, *thesaurus based view* is enabled; Hence, the IWN Dictionary renders synonymous words in form of thesaurus. On clicking hyperlink *more*, a user can see the meaning of these set of words. Also, user can click on any of the words in the list which takes him/her to multilingual senses of that particular word.

In figure 4, we can see that the source language is Hindi and the target languages are English, Gujarati, Hindi, Kannada, Nepali and Odia. Here, *word usage based view* is enabled; hence, the IWN Dictionary renders usage of a word by showing the example sentences in a synset.

[10] http://anoopkunchukuttan.github.io/indic_nlp_library/



IndoWordNet Multilingual Dictionary

Select an Input Language: Hindi

Search a word: घर

☑ Phonetic Transliteration

Dictionary View : ○ Sense Based ○ Thesaurus Based ● Word Usage Based ○ Language Based

Hindi [Noun] इस **घर** में पाँच कमरे हैं। / विधवा मंगला नारी **निकेतन** में रहती है। *more*

Assamese □ Bengali □ Bodo □ English ☑ Gujarati ☑ Hindi ☑ Kannada ☑ Kashmiri □ Konkani □ Maithili □
Malayalam □ Manipuri □ Marathi □ Nepali ☑ Odia ☑ Punjabi □ Sanskrit □ Tamil □ Telegu □ Urdu □

English [Noun] he has a **house** on Cape Cod; she felt she had to get out of the **house** *more*
Gujarati [Noun] આ **ઘરમાં** પાંચ કમરા છે./વિધવા મંગલા નારી **નિકેતનમાં** રહે છે *more*
Hindi [Noun] इस **घर** में पाँच कमरे हैं। / विधवा मंगला नारी **निकेतन** में रहती है। *more*
Kannada [Noun] ನಮ್ಮ **ಮನೆ** ತುಂಬಾ ಚಿಕ್ಕದು more
Nepali [Noun] यस **घरमा** पाँचवटा कोठा छन्/ विधवा मङ्गला नारी **निकेतनमा** बस्छिन् *more*
Odia [Noun] ଏହି **ଘରେ** ପାଞ୍ଚଟି କୋଠରୀ ଅଛି / ନାରୀ **ନିକେତନରେ** ବିଧବା ମଙ୍ଗଳା ରହେ *more*

Hindi [Noun] मेरा **कमरा** दूसरी मंज़िल पर है। *more*

Hindi [Noun] भारत मेरा **देश** है। *more*

Hindi [Noun] मैं अपनी **बुनियाद** से कभी दूर नहीं रहा। *more*

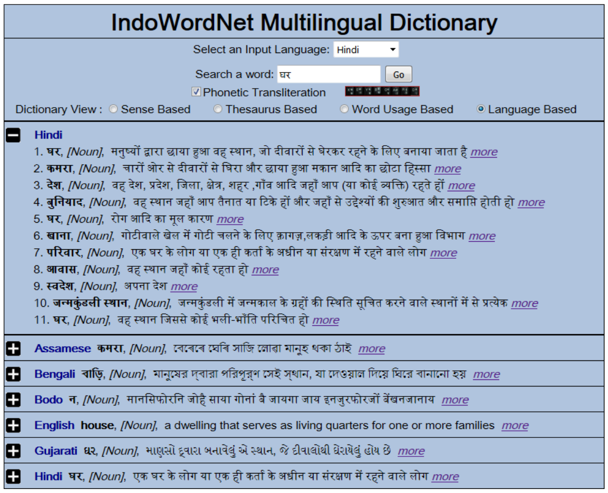
Hindi [Noun] गंदगी रोगों का **घर** है। *more*

Hindi [Noun] उसने शतरंज के मोहरे को अगले **खाने** में रखा। *more*

Hindi [Noun] मेरा परिवार साथ में बैठकर **खाना** खाता है। *more*

Hindi [Noun] स्वच्छ और हवादार **आवास** स्वास्थ्य के लिए लाभदायक होता है। / यह पेड़ ही इन पक्षियों का **बसेरा** है। ... *more*

Figure 4 IndoWordNet Dictionary Interface showing multiple languages with word usage based view.



IndoWordNet Multilingual Dictionary

Select an Input Language: Hindi

Search a word: घर

☑ Phonetic Transliteration

Dictionary View : ○ Sense Based ○ Thesaurus Based ○ Word Usage Based ● Language Based

Hindi
1. **घर**, *[Noun]*, मनुष्यों द्वारा छाया हुआ वह स्थान, जो दीवारों से घेरकर रहने के लिए बनाया जाता है *more*
2. **कमरा**, *[Noun]*, चारों ओर से दीवारों से घिरा और छाया हुआ मकान आदि का छोटा हिस्सा *more*
3. **देश**, *[Noun]*, वह देश, प्रदेश, जिला, क्षेत्र, शहर ,गाँव आदि जहाँ आप (या कोई व्यक्ति) रहते हों *more*
4. **बुनियाद**, *[Noun]*, वह स्थान जहाँ आप तैनात या टिके हों और जहाँ से उद्देश्यों की शुरुआत और समाप्ति होती हो *more*
5. **घर**, *[Noun]*, रोग आदि का मूल कारण *more*
6. **खाना**, *[Noun]*, गोटीवाले खेल में गोटी चलने के लिए क़ागज़,लकड़ी आदि के ऊपर बना हुआ विभाग *more*
7. **परिवार**, *[Noun]*, एक घर के लोग या एक ही कर्ता के अधीन या संरक्षण में रहने वाले लोग *more*
8. **आवास**, *[Noun]*, वह स्थान जहाँ कोई रहता हो *more*
9. **स्वदेश**, *[Noun]*, अपना देश *more*
10. **जन्मकुंडली स्थान**, *[Noun]*, जन्मकुंडली में जन्मकाल के ग्रहों की स्थिति सूचित करने वाले स्थानों में से प्रत्येक *more*
11. **घर**, *[Noun]*, वह स्थान जिससे कोई भली-भाँति परिचित हो *more*

Assamese **कमरा**, *[Noun]*, বেৰেৰে ঘেৰি সাজি লোৱা মানুহ থকা ঠাই *more*

Bengali **बाड़ि**, *[Noun]*, মানুষের দ্বারা পরিপূর্ণ সেই স্থান, যা দেওয়াল দিয়ে ঘিরে বানানো হয় *more*

Bodo **न**, *[Noun]*, मानसिफोरनि जोहै साया गोनां बै जायगा जाय इनजुरफोरजों बेंखनजानाय *more*

English **house**, *[Noun]*, a dwelling that serves as living quarters for one or more families *more*

Gujarati **ઘર**, *[Noun]*, માણસો દ્વારા બનાવેલું એ સ્થાન, જે દીવાલોથી ઘેરાયેલું હોય છે *more*

Hindi **घर**, *[Noun]*, एक घर के लोग या एक ही कर्ता के अधीन या संरक्षण में रहने वाले लोग *more*

Figure 5 IndoWordNet Dictionary Interface showing multiple languages with language based view.

In figure 5, we can see that the source language is Hindi and in the target languages initially the information is rendered in Hindi, However, a user can click on any of the horizontal bars to see all senses in each language. Here, *language based view* is enabled where all senses of an input word are rendered in each of the horizontal language tabs.

The procedure to use this IWN multilingual dictionary is as follows:

- User selects an input language.
- User types in a word. To input a word, he can enable phonetic transliteration or use onscreen keyboard.
- User selects the Dictionary View; by default, *sense based view* is enabled and information is rendered sense wise.
- User clicks on button GO. Once he/she clicks on button GO, the meanings of an input word are displayed in horizontal tabbed format.
- User can toggle between different views by clicking on view sense based, thesaurus based, language based, word usage based radio buttons.
- User can click on *more* to get additional information under each of these views.
- The transliterated information is provided under each of these views. User has to click on checkbox ‘view transliteration’ for this purpose; this is available under the hyper link *more*. However, user sees transliteration of each language on mouse *hover* in English by default.

4 Features of IWN Dictionary

The salient features of IndoWordNet Dictionary are as follows:

- Renders information in multiple languages at a time.
- Different views to display the information: *sense based, thesaurus based, word usage based, language based.*
- Transliteration feature available in different languages which helps in reading effectively.
- Can assist language translation task where multilinguality is primary concern.
- Can be referred for educational and social needs.
- Automatic word suggestion is assisted.
- Similar / closest word suggestion is assisted.
- Usage analysis and statistics available.
- Provides statistics such as most frequently searched word with respect to the selected language, input language searched, etc.

5 Limitations of IWN Dictionary

- It is a concept based dictionary, i.e., for all the synonymous words it will show only one gloss or a concept definition.

- Based entirely on WordNet as a backend resource which is an ongoing development resource so issues/errors in WordNet data will also be propagated here.
- Issues of WordNet can be its limitations like word spellings as found in corpus used for creating WordNet.
- Since the development of WordNet in all mentioned languages is not at same pace, at times we may not get the meaning in a language if it does not exist.
- The availability of words and their meaning depends upon the how much the WordNet data is available.
- Features such as pronunciation, sound, pictures, domain, origin of a word, etc. are not incorporated due to unavailability of the same in the IndoWordNet.

6 Conclusion

In this paper, we have presented an online interface *viz*. *IWN Dictionary*. The IWN Dictionary interface renders IndoWordNet information in a systematic and classified manner. There are various views; such as *sense based, thesaurus based, word usage based, language based* to view the IndoWordNet information. Transliteration feature is also provided in this interface for user readability. Rendering is done based on the user response and usage.

7 Future Scope and Enhancements

In future, we plan to expand this interface by including morph analyzer to process inflected input word. The ontological details along with semantic and lexical relations can be incorporated in the dictionary interface. Further, there can be a feature to link IndoWordNet dictionary to other available dictionaries along with the foreign language WordNets using World WordNet Database Structure. The features like picture, pronunciation, domain, source of a word, glossary, *etc*. can be implemented in the future versions. The transliteration feature to effectively read in one's native language can be implemented in this interface.

Acknowledgments

We graciously thank all the members of CFILT[11] lab, IIT Bombay for providing necessary help and guidance needed for the development of *IndoWordNet Dictionary*. Further, we sincerely thank the members IndoWordNet community and the users of IndoWordNet resource.

References

Anoop Kunchukuttan, Ratish Puduppully, Pushpak Bhattacharyya, *Brahmi-Net: A transliteration and script conversion system for languages of the Indian subcontinent*, Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2105) - Human Language Technologies: System Demonstrations. 2015

Christiane Fellbaum, (ed.) 1998, *WordNet: An Electronic Lexical Database*. The MIT Press.

Christodoulakis, Dimitris N. 2002. *BalkaNet: A Multilingual Semantic Network for Balkan Languages*. EUROPRIX Summer School, Salzburg Austria, September, 2002.

George Miller, R., Fellbaum, C., Gross, D., Miller, K. J. 1990. *Introduction to wordnet: An on-line lexical database*. International journal of lexicography, OUP. (pp. 3.4: 235-244).

Hanumant Redkar, Sudha Bhingardive, Diptesh Kanojia, and Pushpak Bhattacharyya. 2015. *World WordNet Database Structure: An Efficient Schema for Storing Information of WordNets of the World*. In Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, Texas.

Manish Sinha, Mahesh Reddy and Pushpak Bhattacharyya. 2006. *An Approach towards Construction and Application of Multilingual Indo-WordNet*. 3rd Global Wordnet Conference (GWC 06), Jeju Island, Korea, January, 2006.

Neha R Prabhugaonkar, Apurva S Nagvenkar, Ramdas N Karmali. 2012. *IndoWordNet Application Programming Interfaces*. COLING2012, Mumbai, India.

Pushpak Bhattacharyya. 2010. *IndoWordNet*. In the Proceedings of Lexical Resources Engineering Conference (LREC), Malta.

[11] http://www.cfilt.iitb.ac.in/

Pushpak Bhattacharyya, Ankit Bahuguna, Lavita Talukdar, and Bornali Phukan. . 2014 *Facilitating Multi-Lingual Sense Annotation: Human Mediated Lemmatizer*. In Global WordNet Conference, 2014.

Piek Vossen, (ed.) 1999. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks for European languages*. Kluwer Academic Publishers, Dordrecht.

Rajat Mohanty, Pushpak Bhattacharyya, Prabhakar Pande, Shraddha Kalele, Mitesh Khapra and Aditya Sharma. 2008. *Synset based multilingual dictionary: Insights, applications and challenges.* Global WordNet Conference, Hungary, January 22-25.

Shikhar Sarma, Kr Sarma Dibyajyoti, Biswajit Brahma, Mayashree Mahanta, Himadri Bharali, and Utpal Saikia. 2012. *Building Multilingual Lexical Resources Using Wordnets: Structure, Design and Implementation.* 24th International Conference on Computational Linguistics. 2012.

Venkatesh Prabhu, Shilpa Desai, Hanumant Redkar, Neha Prabhugaonkar, Apurva Nagvenkar, Ramdas, Karmali. 2012. *An Efficient Database Design for IndoWordNet Development Using Hybrid Approach*. COLING 2012, Mumbai, India. p 229.

Veronika Vincze, György Szarvas, and János Csirik. 2008. *Why are wordnets important*. Proc. of 2nd ECC, Malta, WSEAS Press (2008): 316-322.