

An Online Dictionary and Thesaurus

Dikshit Kumar, Agam Kumar, Man Singh, Archana Patel*, Sarika Jain

Department of Computer Applications, National Institute of Technology, Kurukshetra, Haryana, India

Abstract

Semantic similarity and definition of words are very crucial part of the data; therefore various online dictionary and thesaurus had been studied in the literature and it was found that wordnet is the most famous and better one. WordNet design is inspired by current psycholinguistic theories of human lexical memory. It is an ontology-based system which organizes English nouns, verbs, and adjectives into synonym sets, each representing one underlying lexical concept. Although wordnet is a more efficient dictionary and thesaurus then also there are lots of problems in it which make it not so good. Therefore in this paper we will discuss all the challenges of existing online dictionaries and thesauruses and then we will conclude that why WordNet is good but not the best system. This state of the art paper clarifies the difference between all existing online dictionaries and thesauruses by using an effective parameter.

Keywords: WordNet, ontology, dictionary, thesaurus

***Author for Correspondence** E-mail: archanamca92@gmail.com

INTRODUCTION

A dictionary, sometimes known as a wordbook, is a collection of words in one or more specific language often arranged alphabetically which may include information on definition usage, pronunciations, translation, etc.; and a thesaurus is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms) [6]. Various dictionaries and thesauruses are available in the world but WordNet is most widely used lexical system. WordNet is the project of the Princeton University which started in 1985 [1]. There are various versions of wordnet been released like WordNet 1.5, WordNet 1.6, WordNet 1.7, WordNet 1.7.1, WordNet 2.0, WordNet 2.1, WordNet 3.0, and WordNet 3.1; whereas the latest version of wordnet is 3.1 which has been released in November 2013. WordNet 3.0 is available for Unix-like systems (including: Linux, Mac OS X, Solaris) and the most recent windows version of WordNet is 2.1, released in March 2005 [2]. WordNet database contains 155,287 words which are organized in 117,659 synsets for a total of 206,941 words sense pairs. It is a lexical database for the English language which groups the words into sets of synonyms

called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members. Examples of semantic relations used by WordNet are synonymy, antonyms, hyponymy [3]. A word more than one sense is called the polysemous and that word share at least one sense in common is called synonymous. Since a semantic relation is a relation between meanings, and meaning can be symbolized by synsets, therefore it is natural to think of semantic relation as pointers between synsets. A synset is a collection of synonymous words which links the other synset according to a number of different possible relationships between the synsets (e.g., is-a, has-a, is-part-of, and others).

Although other dictionaries and thesauruses have many features but WordNet provides a more effective combination of traditional lexicographic information and modern computing. Wordnet stores all concepts in the knowledge base called ontology [4]. The ontology is “an explicit specification of the conceptualization of the domain” [5]. It typically contains representations and descriptions of object, attribute, relationship, values, axioms etc. There are several key

points for ontology as follows:

- The ontology makes things explicit;
- The ontology is supposed to be formal;
- The ontology concerns some specific domain; and
- The ontology represents a conceptualization.

The key purpose of the ontology is to provide a vocabulary with which to model the information and knowledge to be put in the system, and to frame the input to and output from the system. The ontology determines what can and what cannot be said in an application. The role it plays for a knowledge base is akin to that played by the conceptual schema of a database [6]. Ontologies are broadly utilized in the field of knowledge acquisition, knowledge sharing, knowledge re-use, domain theory development, verification and validation. There are basically three types of ontologies namely, lightweight ontologies, upper or top ontologies and core or domain ontologies, whereas Wordnet is best known and most widely used lightweight ontology [7]. With respect to the database system, ontology is similar to the hierarchical and relational models, but intended for modeling knowledge about class elements, their attributes, and relationships among them [8].

The remainder of the paper is prepared as follows. Next part of the paper introduces review of related works. Then it presents the comparative analysis of all existing lexical systems. After that we discuss the scope and gaps of wordnet and describe the applications of lexical system. Last part of the paper concludes the discussion and future works.

RELATED WORKS

There are various dictionaries and thesauruses available in the literature which provide a concrete description of the word. Each online dictionary and thesaurus has some unique and important features but there is not a single online dictionary and thesaurus which contains the entire important feature. Therefore in this section we will highlight all systems and talk about its importance.

OxfordDictionaries.com

It was published in 1992 (1996 in website form). This dictionary content focuses on

underway English and admits modern meanings and importantly tells about the uses of words. In this dictionary, if we search any word and searching word has more than one meaning in modernistic English, then most significant and common meaning in latest English are given first and less common and inexhaustible words are listed below. The database of this dictionary offers world english as well as American English as different lexicons [9].

Dictionary.Com

It is an online dictionary published on 14th May 1995 which provides the Hindi meaning and synonym of desired word [10]. It also gives the definition (sense) of the desired word and shows how many times a word can be used in the form of noun and verb. Dictionary.com presents the synonym according to alphabetical order, complexity and lengthwise [11]. It provides synonym without any semantic relationship between them and does not display the frequency count and sense key of word.

Cambridge Advanced Learner's Dictionary

In 1995, Cambridge University published its dictionary for English users or learners. Cambridge dictionary provided free access to an online dictionary and thesaurus in 1999. The dictionary has over 140,000 words, phrases, and meanings. This dictionary contains the Cambridge English Corpus of more than 1.5 billion words of real English [12].

Merriam-Webster

Merriam-Webster online began offering these dictionaries completely free of cost in 1996. This contains dictionary of synonyms, English usage, slang, medical terms, geography biography, proper names, Spanish/English, sports terms, and numerous others [13].

MacmillanDictionary.com

This dictionary for progressive learners was initially published in 2002 by Macmillan Education [14]. This electronic dictionary is available at zero cost on the Web. The main advantage of Macmillan Dictionary is that it always has updated content regularly with some latest words. Latest version of Macmillan Dictionary has facility to find

synonyms for any word, meaning or phrase [15, 16].

Wiktionary

Wiktionary is online dictionary published on 12 December 2002 by Daniel Alston and this idea was given by Larry Sanger, co-founder of Wikipedia. It is a combined project to make a free-content multilingual dictionary and it illustrates words of all the languages by using definitions and descriptions in English. Its main purpose is not to include the definition of the word, but rather to provide more information to understand it. Therefore synonyms, antonyms, etymologies, sample quotations and translations are included [17].

Shabdkosh

Shabdkosh was first brought online in 2003 by Maneesh Soni. It is an online dictionary which provides the Hindi meaning, synonyms and the definition (sense) of the desired word [18]. As similar to dictionary.com, it also does not hold any relationship between the synonyms, sense and does not give frequency count as well as sense key.

TheFreeDictionary.com

It was published on 5th June 2003. TheFreeDictionary.com is an American online dictionary that collects information from a various sources and provides the graphical representation of the synonymous [19]. TheFreeDictionary.com along with TheFreeLibrary.com is the world's most broad online database of reference information. It is a multi-dictionary search engine with a customizable portal feature.

CollinsDictionary.com

The Collins Dictionary was first published on the web on 31 December 2011. This dictionary contains the words of the British English. It uses the full capacity of computer databases and typesetting in its preparation. There are total 12 editions available of this dictionary and 12th is the latest edition, which was published in October 2014 [20]. The website also describes the example sentences which show the effective word usage from the Collins Bank of English Corpus [21].

CollinsDictionary.com increased the strength of words by adding 6500 new scrabble words to their database. These words are based on terms related to and influenced by food, technology, slang, social media and more [22].

Word Web Online

It was brought out on 5th December 2016. It is an international English dictionary and thesaurus partly based on the WordNet database. The thesaurus is incorporated into the vocabulary as well as categorization of the relationship between words, are shown under the definition. The word web dictionary database contains antonyms, synonyms, type of ('papaya' is listed as a type of 'tree'), types ('play' lists several sub-types of play, including 'passion play'), parts (under 'Department', listed parts include 'faculty' and 'student'), part of ('wheel' is a part of 'vehicle' like car, truck etc) and similar (words that are not synonyms, but nearby; 'good' is listed as similar to 'nice'). Users can also actively improve the efficiency of dictionary and thesaurus by submitting errors (such as missing words/phrases/senses) and enhancement requests [23].

Synonym.Com

It only gives synonyms and antonyms of a particular word and does not show any type of definitions. This database provides the synonym without any semantic relationship between words, sense and does not give the frequency count and sense key of word [24].

FrameNet

FrameNet is a project to make a database of words of English that is useful for both human and machine. From the user's point of view, FrameNet is a dictionary of more than 13,000 word senses, most of them with annotated examples that shows the meaning and usage. For example: "He teaches English" essentially describes the same basic situation as "English is taught by him" just from different perspective. It changes the sentence without changing its meaning [25].

Power Thesaurus

Power Thesaurus is a fast, user-friendly and broad online thesaurus. The synonymic and

other associations are suggested, rated (according to the length, alphabets and syllable) and used by its visitors [26]. It provides an interface by which user can select the words with the help of special characters for example search tab contains following information:

Special characters '?' and '*':

- ????: shows 3-letter terms.
- a??e: 4-letter terms starting with 'a' and ending with 'e'.
- a*: terms starting with 'a'.
- *ment: terms ending with 'ment'.

If nothing is found, then alternative search will try to find the terms that:

- Start with searched query;
- Sound like it; or
- Similar to it.

Hinkhoj

HinKhoj is widely used Hindi English Dictionary in India. It also provides synonyms, antonyms and translation. Hinkhoj also helps to learn English by improving vocabulary. It has over 15 million application installs on Android platform and also has over 500000 words and meanings in database [27].

WordNet

WordNet is a database of words having a semantic lexicon for the various language, that organizes words into groups called synsets (i.e., synonym sets). When building a category model for a document, words related to a feature already in the model (and satisfying some desired WordNet relationship) are extracted from the WordNet database and integrated into the model [28]. WordNet is used in two ways, either as a standalone knowledge base or as a mean to augment existing RDF/OWL ontologies. It is more than only an electronic dictionary; though as the “net” in the name suggests, WordNet not only contains words and their definitions, but also a complete set of relations exists among the word senses. In particular, the hyponymy relation between noun synsets induces a taxonomical structure of concepts [29]. WordNet is more powerful in comparison of other available dictionaries and thesauruses because it contains more semantic relations

among the words and the fundamental architecture is based on ontology.

COMPARATIVE STUDY

There are various features which must be included into lexical system. After studying all existing online lexical system we found that none of the systems have all features. In this section we will clearly define which features are missing in which lexical system and compare all these online dictionaries and thesauruses along with effective parameters. Table 1 describes a comparative analysis of online lexical system along with effective parameter. The most important features of lexical systems are as follows:

- Synonyms: Synonym are those words which have at least one sense in common.
- Definition: Definition is a concise explanation of the meaning of a word or phrase or symbol.
- Multilingual: It supports more than one language.
- Complexity of the word: It refers those word which have easy pronunciation.
- Visualization: It specifies the graphical representation of the words with their synonym.
- Ontology Based: It is knowledge repository which extracts implicit information.
- Alphabet Ordering: It provides synonym of a word according to the alphabetical order.
- Sound and Part of Speech.

The comparison Table 1 shows that there is none lexical system which is perfect. Our main concentration is on wordnet because it is based on an ontology which has the capacity to infer hidden knowledge from the knowledge base.

GAPS AND SCOPE

The wordnet is the most famous lexical system in comparison of other existing systems. Therefore after studying the WordNet, we experienced that there is some touchstone in which it needs to work. In this section, we will explain all those features which are required for all lexical systems but it is not available in wordnet.

Table 1: Comparison of Different Dictionary and Thesaurus.

Sr. No.	Title	Synonym	Definition (Sense)	Multilingual	Complexity of the Word	Visualization	Ontology Based	Alphabet Ordering	Sound	Part of Speech
A.	Oxforddictionaries.Com	YES	YES	NO	NO	NO	NO	YES	YES	NO
B.	Dictionary.Com	YES	YES	NO	YES	NO	NO	YES	YES	NO
C.	Cambridge Dictionary	YES	YES	YES	NO	NO	NO	YES	YES	NO
D.	Merriam-Webster	YES	YES	NO	NO	NO	NO	YES	YES	NO
E.	Macmillandictionary.Com	YES	YES	NO	NO	NO	NO	YES	YES	NO
F.	Wiktionary	YES	YES	YES	NO	NO	NO	YES	YES	NO
G.	Shabdkosh	YES	YES	NO	NO	NO	YES	NO	YES	NO
H.	Thefreedictionary.Com	YES	YES	YES	NO	YES	NO	NO	YES	NO
I.	Collinsdictionary.Com	YES	YES	YES	NO	NO	NO	YES	YES	NO
J.	Wordweb Online	YES	YES	YES	NO	NO	YES	YES	NO	NO
K.	Synonym.Com	YES	YES	NO	NO	NO	-	NO	NO	NO
L.	Power Thesaurus	YES	YES	NO	YES	NO	NO	YES	YES	NO
M.	Hinkhoj	YES	YES	YES	NO	NO	NO	NO	NO	NO
N.	Wordnet 3.1	YES	YES	YES	NO	NO	YES	YES	NO	NO

a). Sentence Categorization: It is an essential feature for lexical systems which means to classify the sentence into the parts of speech. Suppose if user gives the input 'He is a Boy'. Then, output should be presented as follows:

{“He (pronoun)
Is (verb)
A (article)
Boy (noun)}.

b). WordNet includes the lexical categories noun, verb, adjective and adverb but ignores preposition.

c). Wordnet does not categorize the given synonym of a word according to complexity and length. Suppose if user wants to find out the synonyms of “Target” then there should be an option on interface by which user can select those words which have maximum length in terms of alphabet, maximum complexity of word and relevance which rank synonyms and suggest the best matches based on how closely a synonym’s sense matches the sense you selected.

d). Wordnet has less effective customization and user interface.

e). Wordnet does not provide the visualization of the relationship between the words. Visualization provides a clear picture of all the words related to each other in some way.

f). Semantic Sentence Facility: Semantic similarity between the sentences is an important feature of the lexical system but unfortunately it does not hold in the wordnet. The semantic sentences are those sentences

which have different ways of saying but have same meaning for example “John sold a car to Mary” OR “Mary bought a car from john”. Both sentences have same sense or meaning but have different way of saying.

These above features play very crucial role in the lexical system. There are different projects where WordNet has great importance and used as a tool. Some major applications of WordNet are as follows:

- Automatic crossword puzzle generators.
- Natural Language Processing: Information retrieval, Automatic text classification and Text summarization, Text mining etc.
- Artificial Intelligence: Ontology mapping, merging, alignment.
- Semantic search with quick response time.
- To understand the difference between different parts of speech.
- To calculate the similarity between the words.

This paper provides guidelines for creating a modern WordNet which will be fully based on semantic technologies like ontology, SPARQL etc. and with the help of these technologies, we can easily resolve all problems which are mentioned above.

CONCLUSION

The main purpose of this paper is to describe that why WordNet is better than other existing

lexical systems. WordNet has a broad future work and large scope. The visualization of the relationship among the words, categorization of the sentence and semantic sentence will make WordNet more effective and attractive. The visualization will aid to know the many lexical relationships that can be investigated through WordNet easily. This paper provides us with new opportunities to estimate and examine scientific works, discover the network of authors, reviewers, and editors, and attempt to present potential research topic trends.

REFERENCES

1. Miller G. *WordNet: A Lexical Database for English.Com*. ACM; Nov 1995.
2. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/WordNet>
3. Miller GA, Charles WG. Contextual Correlates of Semantic Similarity. *Lang Cogn Process*. 1991; 6(1): 1–28p.
4. Patel A, Jain S. Formalisms of Representing Knowledge. *Procedia Comput Sci*. 2018; 125: 542–549p.
5. Gruber TR. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In: *Formal Ontology in Conceptual Analysis and Knowledge Representation*. Kluwer Academic Publishers, in Press. Substantial Revision of Paper Presented at the *International Workshop on Formal Ontology*. 1993.
6. Miller George A, Richard Beckwith, Christiane Fellbaum, *et al*. Introduction to WordNet: An On-Line Lexical Database. *Int J Lexicograph*. 1990; 3(4): 235–244p.
7. Trevor Bench-Capon. *Ontologies in AI and Law*. In *Inteligencia Artificial Aplicada Al Derecho, Memoria Del Congreso Internacional De Culturas Y Sistemas Jurídicos Comparados*, Ciudad De México, Instituto De Investigaciones Jurídicas; 2005; 65–85p.
8. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/OntologyWordNet>
9. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/OxfordDictionaries.com>
10. Dictionary.com WHOIS, DNS, & Domain Info: Domain Tools Page at DomainTools.com. Retrieved Jun 16, 2016.
11. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/Dictionary.com>
12. Cambridgeorg. (2019). Cambridgeorg. Retrieved 8 April, 2019, from <https://dictionary.cambridge.org/about.html>
13. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/Merriam-Webster>
14. Bejoint H. *The Lexicography of English*. Oxford University Press; 2010; 186–189p.
15. Edemariam Aida. Online Dictionaries: Which is Best? *The Guardian*. 30 Aug 2010.
16. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from https://en.wikipedia.org/wiki/Macmillan_English_Dictionary_for_Advanced_Learners
17. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/Wiktionary>
18. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://hi.wikipedia.org/wiki/Shabdkosh>
19. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/TheFreeDictionary.com>
20. *Collins English Dictionary*. 12th Edn. Collins.co.uk. HarperCollins Publishers Ltd. Retrieved 6 Dec 2014.
21. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from https://en.wikipedia.org/wiki/Collins_English_Dictionary
22. Collinsdictionarycom. (2019). Collinsdictionarycom. Retrieved 8 April, 2019, from <https://www.collinsdictionary.com/word-lovers-blog/new/discover-the-new-scrabble-words,210,HCb.html>
23. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/WordWeb>

24. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://simple.wikipedia.org/wiki/Synonym>
25. Wikipediaorg. (2019). Wikipediaorg. Retrieved 8 April, 2019, from <https://en.wikipedia.org/wiki/FrameNet>
26. Powerthesaurusorg. (2019). Powerthesaurusorg. Retrieved 8 April, 2019, from https://www.powerthesaurus.org/_about
27. Powerthesaurusorg. (2019). Powerthesaurusorg. Retrieved 8 April, 2019, from https://www.powerthesaurus.org/_about
28. Miller GA, Beckwith R, Fellbaum CD, et al. WordNet: An Online Lexical Database. *Int J Lexicograph*. 1990; 3: 235–244p. <https://en.wikipedia.org/wiki/WordNet>.
29. Valerio Basile. WordNet as an Ontology for Generation. In *WebNLG 2015 1st International Workshop on Natural Language Generation from the Semantic Web*. 2015.

Cite this Article

Dikshit Kumar, Agam Kumar, Man Singh, Archana Patel, Sarika Jain. An Online Dictionary and Thesaurus. *Journal of Artificial Intelligence Research & Advances*. 2019; 6(1): 32–38p.