

Comprehensive Analysis of Word Sensing Tool and Techniques for Enhancing Classification Accuracy of Query String

Sunita Mahajan^{1*} and Vijay Rana²

¹Department of Computer Science, Arni University, India

²Department of Computer Science, SBBS University, India

Abstract

Word sense in the field of natural language processing (NLP) is a corner stone for appropriate word selection. A word can contain more than one sense, but machine can't extract the actual sense of the given particular content. Implication of this situation is mismatch between the user requirements and result generates through the machine. e.g., User wants to search a query "What is word sensing?" The machine can't find the relation between these two words "word", "sensing". Relationship between words cannot be extracted by the machine and more results corresponding to sensing is displayed and user requirements corresponding to "Word Sensing" as a whole are rejected. primary reason for this mismatch is due to static dictionary possessed by web servers. Techniques we are analysis different types of techniques and algorithms for the word sense. The major techniques are which used to word sense are knowledge based approaches are based on different knowledge sources as machine readable dictionaries to extract the sense like thesauri, Word net are machine readable dictionaries to find the word sense, Supervised learning technique is a manually extract the sense from the data. In this process trained the target words through the labelling, unsupervised learning technique in this process words are no needs to be trained target data are based on the clustering, Semi-Supervised learning technique is a hybrid approach of the supervised and unsupervised. In this process target words are based on the particular content. Tools for building word database to be accessed by the web applications including Word Net, Image Net and Babel Net are discussed in this literature. Our Contribution we conduct comprehensive review of knowledge based, supervised, unsupervised and semi-supervised learning techniques used in the field of word sensing and detect the best word sensing mechanism for fetching only relevant material from the web while decreasing the execution time for content retrieval.

Keywords: Execution time; Knowledge base; Word sense

Introduction

Search engine is a famous term for an Information retrieval system [1,2]. The function of a search engine is to mine requested meaningful information from the massive database of resources presented on the internet. Search engines are critical public tool for searching the necessary information from global web base. There are many different types of search engines to acquire the information you are searching for, from the internet. Internet usage has been extremely increased in modern days with the easy to use search engines like Google, Bing and Yahoo [3]. A search engine search user query is a requested for information from the web. The requested information is mixed type like text, images, audio, video and other types of data. Every time a user puts a string in a search engine [4]. The strings act as the keywords. Search engine compares the user query in the search request with the indexed pages from the database. Since it is likely that more than one page contains the search string, search engine starts calculating the relevancy of each of the pages in its index with the search string [5]. There are various algorithms like Panda, Penguin, Hummingbird, PageRank to calculate relevancy. Each of these algorithms has different relative weights for common factors like keyword density, links, or meta tags. That is why different search engines give different search results pages for the same search string. It is a known fact that all major search engines periodically change their algorithms [6]. Search engine can be described as crawling, data mining, indexing and query processing. The popular search engines are crawler based search engines. The crawler based search engines are Google, Bing, Yahoo (Figure 1). Crawling is also known as spider. The spider runs on the internet and visits websites and read the information on presented Web Pages try to fetch the meaningful Web Pages. The extracted meaningful webpages and their links are load on their database. Google's web crawler is known as Google bot. In case some webpages are not be searched that are

called deep web [7,8]. The indexer analysis the contents, links of the web pages that are searched by the spider. The spider maintains the index so that similar contents should be matched again and again also unmatched contents is discarded with aggregate execution speed. The index contains content along with address of the URL where it is located. Query processing user can find some information on the internet the user query contains some words. Something to keep in mind, you are not searching the entire internet when you search. most search engines have a list of stop words, words that will not be searched. most search engines will not search for the, and, it, be, will, etc. users are only searching an index of the internet, after the pre-processing extract the Keywords from the user query. The keywords are matched with database and extract the information to be provided by the user [9-11]. Natural language processing (NLP) is becoming more essential area of Human Computer Interaction. It is the branch of AI with includes information extraction, machine translation, and speech recognition and also interacts with search engine to extract appropriate results. The general aim of such search interface or NLP system is to make the machine able to understand user query. NLP techniques which can make the computer understand the languages naturally used by humans. While natural language may be the easiest symbol system for people to learn and use, it has proved to be the

***Corresponding author:** Sunita Mahajan, Department of Computer Science, Arni University, India, Tel: 18001200049; E-mail: sunitamahajan2603@gmail.com

Received July 13, 2018; **Accepted** July 26, 2018; **Published** August 10, 2018

Citation: Mahajan S, Rana V (2018) Comprehensive Analysis of Word Sensing Tool and Techniques for Enhancing Classification Accuracy of Query String. J Comput Sci Syst Biol 11: 265-275. doi:10.4172/jcsb.1000284

Copyright: © 2018 Mahajan S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

hardest for a computer to master. Neuro linguistic programming is refers to neurology language functions. The overall speed execution can be enhanced by the use of NLP [12,13]. The pattern analysis is used in this phase, where pattern analysis indicates behaviour or expectation of user in performing desire operation. Operation can be enhanced further by including decision using fuzzy membership function. Every membership function has certain weight associated. The query string will matched with the membership function, if match occurs than the sentence prediction take place. With the help of neurology language functions, all the query parts not required to be searched again and again, the membership function with higher weighted values are selected for display through this approach. Word sense is the main feature to improve web search. Web search becomes need of the user for being improve the search efficiency User can input their query on a search engine. But the search engine can't try to identify the user opinion behind the searched query, because some search engines are not searched for stop words the, and, to, it, be, will etc. They can try to extract the actual keywords from the user query [14-17]. Those keywords are matched with database and extract the webpages check their indexing. Their search indexing is high that pages to be provided by the user, without understand what the demand of the user behind this search string. Rest of the paper is organized under section 2 gives the detail of techniques used for word sensing.

Machine Learning Techniques

Machine Learning is also part of the AI in area of computer science that techniques to provide the capacity to improve the performance on a specific task. We describe the techniques of machine learning that are knowledge based, supervised learning, unsupervised learning, and semi-supervised learning.

Knowledge based

The knowledge based is also known as dictionary based technique to exploit knowledge resources. There resources are low performance rather than supervised and unsupervised learning techniques. The more powerful knowledge resource is present as WORDNET and other knowledge resource are corpus based [18,19].

WORDNET: The Word Net is online lexical dictionary sense representation. The latest version is presented WORDNET 3.1 created

and maintained at Princeton University. The concept of Word Net in the term of synonyms. The resource is based on the basic six relations to extract the sense of the particular word. Their six semantic relations [20-24] (Figure 2).

Context representation: Human can be used natural language in the natural language polysemy words are presented in the input. The alternative senses of a polysemous word is a matter of distinguishing between different sets of linguistic contexts in which the word form can be used to express the word sense, like Apple.

- Apple is presented Fruit.
- Apple is presented Electronics.

In this given example human can understand the meaning of the word “apple”, but machine can't be understand this situation. The WordNet is help to solve this of problems. WordNet can find the appropriate meaning of the word in the given context. The semantic relations can be defined between words and between word senses than are incorporated into WordNet.

Corpus based: The corpus is also known as structure as set of text is based knowledge resources is another method of sense representation. In this knowledge resource data is presented in structured form.

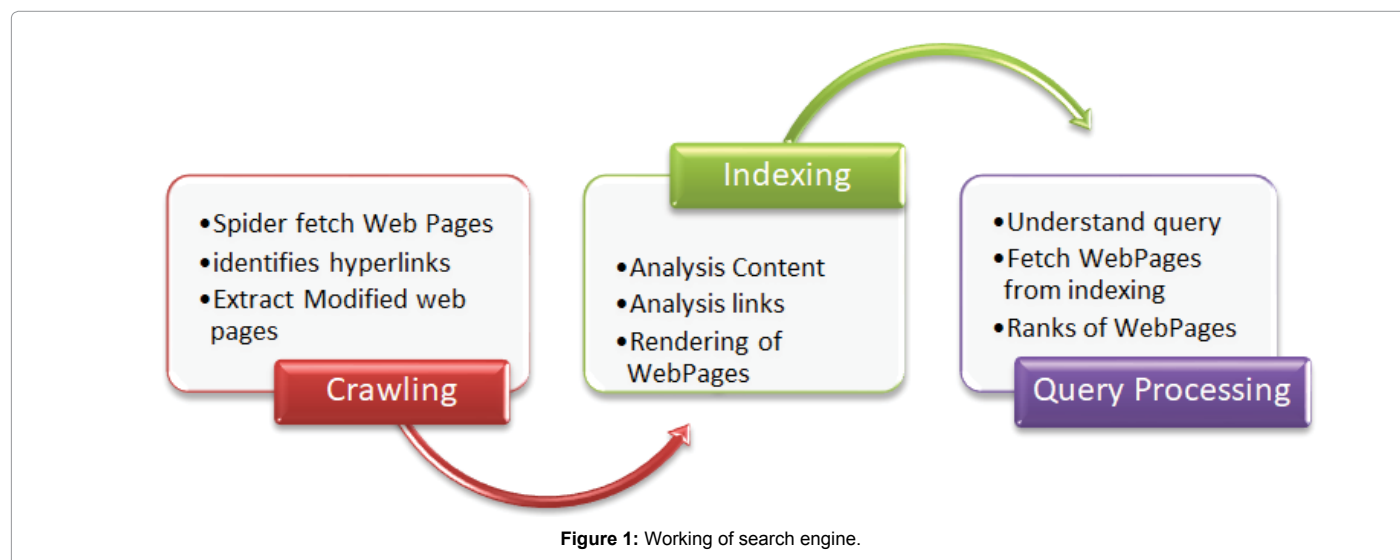
Supervised learning technique

Supervised learning technique is a manually extract the meaningful information from the data. In this process all data trained the target words through the labeling. The function of labeling each input data can contain pair of information that is the process of labelling [25].

Algorithms: The following algorithms are used in the supervised learning techniques.

Support vector machine: SVM understands the data used for the classification and regression. SVM is dividing the input values into positive and negative groups. Find the distance to the closest positive and negative called support vectors. In other words, support vector machines (SVMs) tend at the same time to minimize the empirical classification error and maximize the geometric margin between positive and negative examples.

Decision tree: A decision tree is an analytical model used to stand



for classification rules with a tree structure that recursively partitions the training data set. Each inner node of a decision tree represents a test on a feature value, and each branch represents an outcome of the test. A prediction is made when a terminal node is reached. The tree is traversed and, after the results shows in the form no-yes-no path.

Naive bayes: A naive bayes algorithm is a probabilistic classifier based on the purpose of Bayes' theorem. It relies on the computation of the conditional probability of each sense of a word given the features is depend upon the given context.

Neural networks: A neural network is an interconnected cluster of artificial neurons that uses a computational model for processing data based on a connectionist approach. Pairs of input extract the response are input to the learning program. The utilization of input features to partition the training contexts into without overlapping sets corresponding to the preferred responses. The new groups are provided; link weights are progressively adjusted so that the output unit representing the desired response has a larger activation than any other output unit.

KNN: k-Nearest Neighbor is stands for the k-Nearest Neighbor

algorithm, is depend upon one of the highest-performing methods [26].

Decision list: Decision list provide the appropriate sense to the given input. It can be checked the results as a list of "IF THEN ELSE" rule based. The training set provides the better result is based on the number of matched results.

Unsupervised learning techniques

In the Unsupervised learning techniques is the different aim to the supervised, knowledge based approaches. In this technique there is no labeled data in the input data. The techniques provide the better result because it will check the number occurrence in given word, number of occurrence divide into other clusters. After then it will check the higher propriety.

Algorithms: There are number of algorithms are presented in the unsupervised learning techniques associated with the terms of clustering [27].

Context clustering: Is the first approach of the unsupervised

Semantic Relation	Syntactic Category	Examples
Synonymy (similar)	N, V, Aj, Av	pipe, tube rise, ascend sad, unhappy rapidly, speedily
Antonymy (opposite)	Aj, Av, (N, V)	wet, dry powerful, powerless friendly, unfriendly rapidly, slowly
Hyponymy (subordinate)	N	sugar maple, maple maple, tree tree, plant
Meronymy (part)	N	brim, hat gin, martini ship, fleet
Troponymy (manner)	V	march, walk whisper, speak
Entailment	V	drive, ride divorce, marry

Figure 2: Word Net semantic relations. N-Nouns, AJ-Adjectives, V-Verbs, AV-Adverbs.

learning techniques is based on context clustering. The occurrence of the target word in a corpus represent as a vector. the vectors are compared with number of corpus, each identify a sense of the target word.

Word clustering: Word clustering is the task of the divide the number of words into a certain number of clusters. Each cluster is required to consist of words that are similar to one another in syntactic or semantic construct and dissimilar to words in distinctive groups. Word clustering generalizes specific features by considering the common characteristics and ignoring the specific characteristics among the individual features [28].

Co-occurrence: Co-occurrence refers to an above-chance frequency of occurrence of two terms from a text corpus alongside each other in a certain order. Co-occurrence in this linguistic sense can be interpreted as an indicator of semantic proximity or an idiomatic expression. Corpus linguistics and its statistical analyses reveal patterns of co-occurrences within a language and enable to work out typical collocations for its lexical items [29].

Semi-supervised learning: Semi-supervised learning is a hybrid approach of supervised and unsupervised learning techniques. In this technique that also make used labeled or unlabeled data training under these process large amounts of amount of data is trained for machine. some machines want to be some trained data as well as untrained data may be used, in case theses technique is helpful for the machine training [30].

Tools

Tools for building word database to be accessed by the web applications including corpus, ontology, Word Net, Image Net, Babel Net, IndoWord Net and Sem Eval are discussed in this literature [31,32].

WORDNET

WORDNET is online English lexical dictionary for using sense representation linked together with semantic relationship. It is group of English words into set synonyms called synsets provides with short definition with usage example. Synonyms are words that have similar meanings. A synonym set, or synset, is a group of synonyms. A synset, therefore, corresponds to an abstract concept. the latest version is presented WORDNET 3.1 created and maintained at Princeton University. The resource is based on the basic six relations to extract the sense of the particular word. Their six relations are N-nouns, AJ-adjectives, V-verbs, AV-adverbs (Figure 3).

Synonymy (Silimar): The term synonymy means those words their sense are similar like truck, lorry, sad unhappy [33].

Antonym (opposite): It express the opposite concepts like bad, happy, day, night, poor, rich.

Hyponymy (subordinate and it inverse): subordinate and it inverse are transitive relations between synsets.

Metonymy-(part and it inverse): part and it inverse are complex semantic relations.

Tropology (manner): tropology is for verbs what hyponymy is for nouns, although the resulting hierarchies are much shallower.

Working of WORDNET3.1

Word net is online dictionary available on the internet. User can

easily access this dictionary. The latest is version of WORDNET3.1. User can add a word in Wordnet3.1 interface. Example- Apple. The results are shown as number of sense is presented in the word Apple and show the sentence (Figure 4).

Context representation

Human can be used natural language in the natural language polysemy words are presented in the input. The alternative senses of a polysemous word is a matter of distinguishing between different sets of linguistic contexts in which the word form can be used to express the word sense, like Apple,

- Apple is presented Fruit,
- Apple is presented Electronics.

In this given example human can understand the meaning of the word “apple”, but machine can’t be understand this situation. The WordNet is help to solve this of problems. WordNet can find the appropriate meaning of the word in the given context. The semantic relations can be defined between words and between word senses than are incorporated into WordNet.

Applications: There are number of applications of the Wordnet it also helpful in speech recognition, allows us to search and query the text on the basis of the topic as identify by the lexical trees that are formed rather than the basis of keywords, which is most common from the search mechanism on text.

There is major problem of the WordNet, we can’t easily create a gateway with asp.net. There jar file is also available in Java, php but not present in .NET.

Corpus based: The corpus is collection of the text. The structure as set of text is based knowledge resources is another method of sense representation. In this knowledge resource data is presented in structured form.

IMAGENET: ImageNet is hierarchy of the Wordnet. ImageNet is an image database organized as Wordnet key. In a project is a large visual database (Figures 5 and 6).

BABELNET: BabelNet is both a multilingual encyclopedic dictionary, with lexicographic and encyclopedic coverage of terms in 271 languages. Multilingual lexicalized semantic network and ontology developed at the Sapienza University of Rome, at the Department of Computer Science [34] (Figures 7 and 8; Table 1).

Conclusion for the Comparison Analysis for Tools

Firstly WORDNET a text based similarity finder lexical dictionary based on the noun, verb, and adverb adjective. The main problem with Wordnet is the linking part with is not user friendly only jar file is provided for the corresponding to Wordnet which can’t be linked with all the web specific applications (Table 2). Image Net on the other hands is similar to Wordnet with the difference that larger image set can be treacle using Image Net but categorization of information not possible. The classification of the images is the biggest problem with Image Net. In other words different images with different sizes have to be adjusted within the ImageNet performing the operation. Babel Net is the advancement of Word Net and derived information from Wikipedia, dbpedia. Interface phase general problem using Babel Net. Which means it can be linked with the specific application and it can’t be generalized. The comparative analysis indicated that it is clear that knowledge base can be customized to perform efficiency parsing of

Semantic Relation	Syntactic Category	Examples
Synonymy (similar)	N, V, Aj, Av	pipe, tube rise, ascend sad, unhappy rapidly, speedily
Antonymy (opposite)	Aj, Av, (N, V)	wet, dry powerful, powerless friendly, unfriendly rapidly, slowly
Hyponymy (subordinate)	N	sugar maple, maple maple, tree tree, plant
Meronymy (part)	N	brim, hat gin, martini ship, fleet
Troponymy (manner)	V	march, walk whisper, speak
Entailment	V	drive, ride divorce, marry

Figure 3: WordNet semantic relations.

WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) **apple** (fruit with red or yellow or green skin and sweet to tart crisp whitish flesh)
- [S:](#) (n) **apple**, [orchard apple tree](#), [Malus pumila](#) (native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits)

Figure 4: Word Net 3.1.

IMAGENET

14,197,122 images, 21841 synsets indexed

[Explore](#) [Download](#) [Challenges](#) [Publications](#) [CoolStuff](#) [About](#)

Not logged in. [Login](#) | [Signup](#)

ImageNet is an image database organized according to the **WordNet** hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

[Click here](#) to learn more about ImageNet, [Click here](#) to join the ImageNet mailing list.



What do these images have in common? *Find out!*

Check out the ImageNet Challenge on Kaggle!

© 2016 Stanford Vision Lab, Stanford University, Princeton University support@image-net.org Copyright infringement

Figure 5: Image Net.

14,197,122 images, 21841 synsets indexed

Not logged in. [Login](#) | [Signup](#)

Your query "apple" matches 49 synsets.

Rank by [Popularity](#) [Depth](#)

Synset: apple
Definition: fruit with red or yellow or green skin and sweet to tart crisp whitish flesh.
Popularity percentile: 94%
Depth in WordNet: 6

Synset: May apple
Definition: edible but insipid fruit of the May apple plant.
Popularity percentile: 94%
Depth in WordNet: 8

Synset: apple, orchard, apple tree, Malus pumila
Definition: native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits.
Popularity percentile: 91%
Depth in WordNet: 12

Synset: mayapple, May apple, wild mandrake, Podophyllum peltatum
Definition: North American herb with poisonous root stock and edible though insipid fruit.
Popularity percentile: 85%
Depth in WordNet: 8

Synset: cooking apple
Definition: an apple used primarily in cooking for pies and applesauce etc.
Popularity percentile: 80%
Depth in WordNet: 7

Synset: apple butter
Definition: thick dark spicy puree of apples.
Popularity percentile: 80%
Depth in WordNet: 9

Synset: rose apple
Definition: fragrant oval yellowish tropical fruit used in jellies and confections.
Popularity percentile: 79%
Depth in WordNet: 6

Synset: apple juice
Definition: the juice of apples.

Figure 6: Results based on Image Net.

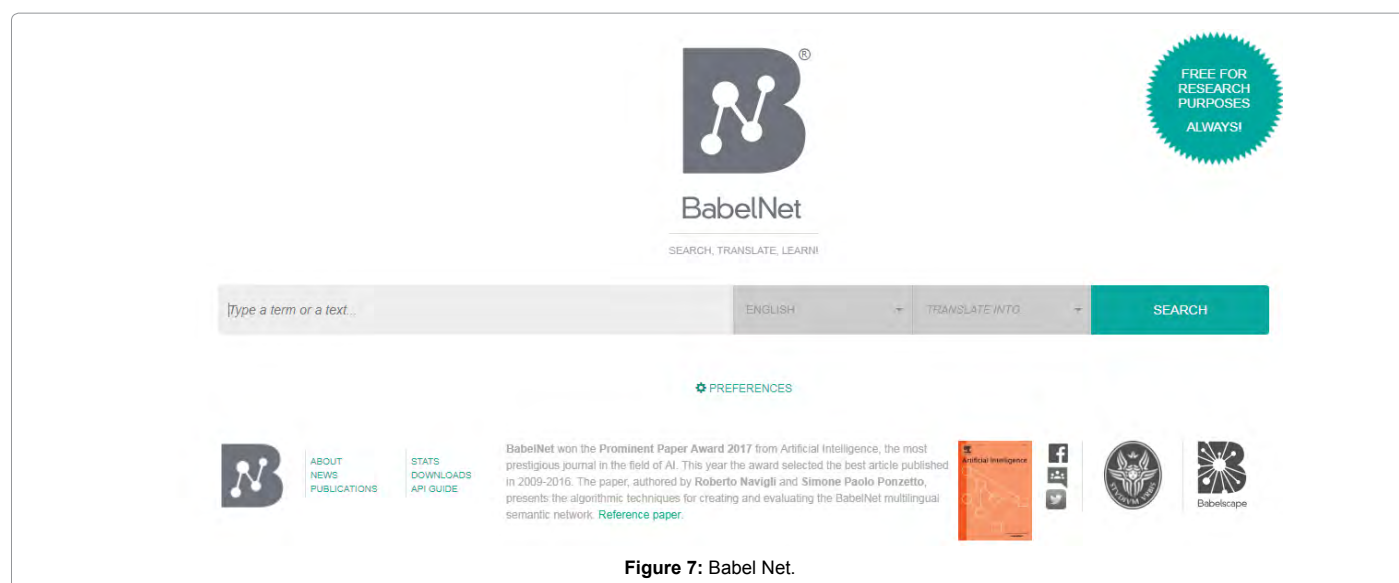


Figure 7: Babel Net.

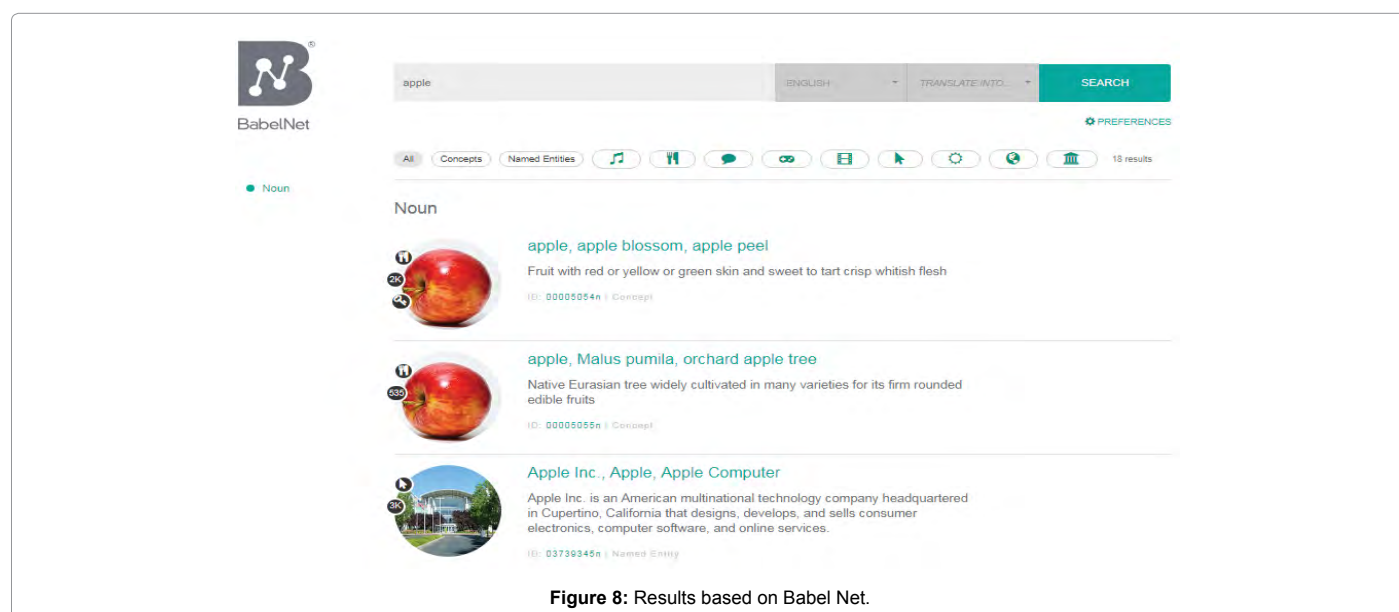


Figure 8: Results based on Babel Net.

the text. The layered within the proposed system can be customized in a manner to accumulate can be parsing quickly. In other words the hybridization is required in order to achieve optimization of the result. The multi-objective environment to be created by using the application of existing supervised, unsupervised, semi-supervised and knowledge bases. By using the customization the sense originated to the string can be better also execution speed also increases. The personalized result is required to be generated which may not be possible using individual approaches. The layers of multilayer perceptron merge along with back propagation can be used for optimization. This optimization can have multiple objectives associated with it. These objectives can be expressed as fitness function. Fitness function can either specify profit or cost. Profit in terms of fitness function must be maximized and cost in case of fitness function must be minimized. The changes to existing techniques including supervised, unsupervised and knowledge base is required for optimization.

Our Contribution

The existing systems can be modified by the use of the phases of the query processing.

Pre-processing

User Query processing improves the search technology. In the pre-processing user query contains number of words. Some search engines are eliminating the stop-words in the given string and pick the remaining keywords. Provide the result is based on the keywords. Sometimes search engines are not providing the accurate results. Our proposed system will be based on the Multi-objective, because some search engine provides the less suggested words for spell correction. Our proposed model, after the pre-processing find the sense of the remaining keywords for provides more accurate result to the users and generates the effective URL [35-39].

Sr. No.	WORDNET	IMAGENET	BABELNET
1.	Lexical database for the English language, a combination of dictionary and thesaurus [32].	Used Image Database Organized according to the WORDNET [31].	Multilingual lexicalized semantic network and ontology based developed [24].
2.	The database contains about 160.000 words, organized in about 120.000 synsets, for a total of about 200.000 word-sense pairs	Features, of object bounding boxes for about 1 million pictures and of object attributes, both annotated and verified through Amazon Mechanical Turk.	A multilingual ontology, a large machine-readable encyclopedic dictionary and a multilingual semantic network [35]
3.	WordNet was created in the Cognitive Science Laboratory of Princeton University under the direction of psychology professor George Armitage Miller, starting in 1985, and has been directed in recent years by Christiane Fellbaum	ImageNet is developed by research team from the universities of Stanford, Princeton, Carolina. The project is sponsored by the Stanford Vision Lab, Stanford University, Princeton University, Google Research and A9, a subsidiary of Amazon.com based in Palo Alto, California, that develops search and advertising technology.	BabelNet is a multilingual lexicalized semantic network and ontology developed at the Sapienza University of Rome, at the Department of Computer Science Linguistic Computing Laboratory. BabelNet was automatically created by linking Wikipedia to the most popular computational lexicon of the English language, WordNet
4.	Wordnet Contains Six relations are Synonymy, Antonymy, Hyponymy, Meronymy, Troponymy	semantic network which connects concepts and named entities in a very large network of semantic relations, made up of more than 13 million entries	Relations come either from Wikipedia hyperlinks (in any of the covered languages) or WordNet.
5.	It is accessible to human users via a web browser, but its primary use is in automatic natural language processing and artificial intelligence applications.	Online accessible for the users	Online accessible for the users
6.	Programmed in Python language.	ImageNet using similar methodology but by accessing large image database	Python with Fuzzy based system [25]used to develop this software
7.	Commonly used find the semantic similarity between words[36]	Our goal is to show that ImageNet can serve as a useful resource for visual recognition applications such as object recognition, image classification and object localization [4].	The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network.
8.	It is only supported English Language and Text based.	It is based on Multilanguage and Image based	It is Multilanguage based and provides description show the image.
9.	Find sense noun, verb, adverb based	Find the sense is based on WORDNET and retrieve more information.	Find the sense between two words show more information.
10.	Use for find semantic similarity between two words.	Use for find semantic similarity between two words [37].	Use for find semantic relatedness between two words.

Table 1: Comparative analysis for the tools WORD NET, IMAGE NET, BABEL NE.

Sr. No.	Knowledge based	Unsupervised	Supervised	Semi-Supervised
1.	Knowledge based method used knowledge resources.	Unsupervised methods do not use any data to adjust internal parameters.	Supervised methods, on the other hand, exploit all given data to optimize parameters such as weights or thresholds.	Semi-supervised methods [38] use only part of the data for parameter optimization, for instance a subset of known network interactions.
2.	In knowledge base aim is to building knowledge and understanding, regardless of its potential applications	In unsupervised learning the aim is to discover groups of similar instances within the data. In this approach, we have no information about the class label of data or how many classes there are.	We assessed and compared the performance of these algorithms to determine if supervised classification outperformed unsupervised clustering and if so which algorithms were most effective	We assessed and compared the performance of these algorithms to determine if supervised classification outperformed unsupervised clustering and if so which algorithms were most effective.
3.	Knowledge base train data from corpora of texts, either Unlabelled or annotated with word senses, to machine-readable dictionaries, thesauri, glossaries, ontologies	Unsupervised learning has no labelled data in training process. A combination of both labelled and unlabelled data	Supervised learning assumes that a set of training data has been provided	Semi-supervised learning is a class of machine learning techniques [38]that utilizes both labelled and unlabelled data.
4.	Knowledge base applications are to improve the interaction with human and machine.	Unsupervised learning encompasses many other problems involving summarizing and explaining various key features of data [38].	Supervised learning applications are widely used such as decision trees, naive Bayes, and support vector machines for their easy evaluation and acceptance to human [38].	semi-supervised learning approaches are explored for practical applications in different areas, such as information retrieval, text classification, and bioinformatics
5.	Knowledge base is summarized Thesauri, ontologies, and Machine-readable dictionaries, WORDNET.	Unsupervised learning are summarized into clustering algorithms like K-Means, KNN, Decision tree, Decision list, Naïve Bayes .	Supervised learning is summarized into two categories: generative algorithms and discriminative algorithms [38].	Semi-supervised learning methods are include self-training, co-training, EM with generative mixture models, graph-based algorithms.

6.	Knowledge base sources are unsalable according to situation.	In the unsupervised learning the sources are scalable.	In the supervised learning sources are unsalable.	In the semi-supervised learning sources are Maybe or may not scalable.
7.	Knowledge base is less expensive to supervised learning	In the unsupervised learning, it is more less expensive rather than supervised knowledge base.	In the supervised learning is more expensive, because it may need to trained data with the help of a trainer.	In semi-supervised it's Depend upon situation to be trained or not trained data.
8.	In the knowledge base compared larger data sources.	In the unsupervised compared larger data sources labelled or unlabelled.	In the supervised smaller and historical data sources are compared that is labelled.	In the semi-supervised intermediate data compared that labelled and unlabelled.
9.	Knowledge base is hybrid of back and feed forward network	In the unsupervised back propagation network is used.	In the supervised feed forward network is used.	In the hybrid approach multilayer perceptron is required.
10.	In the knowledge base more time consumed but easily to optimize. Large data set used.	In the unsupervised data optimization more difficult. Because large data set be used.	In the supervised data optimization is easy. Because undersized data set be used.	In the semi-supervised intermediate data used so more time consumed for the data optimization

Table 2: Comparison between Knowledge, Unsupervised, Supervised, Semi-Supervised Techniques.

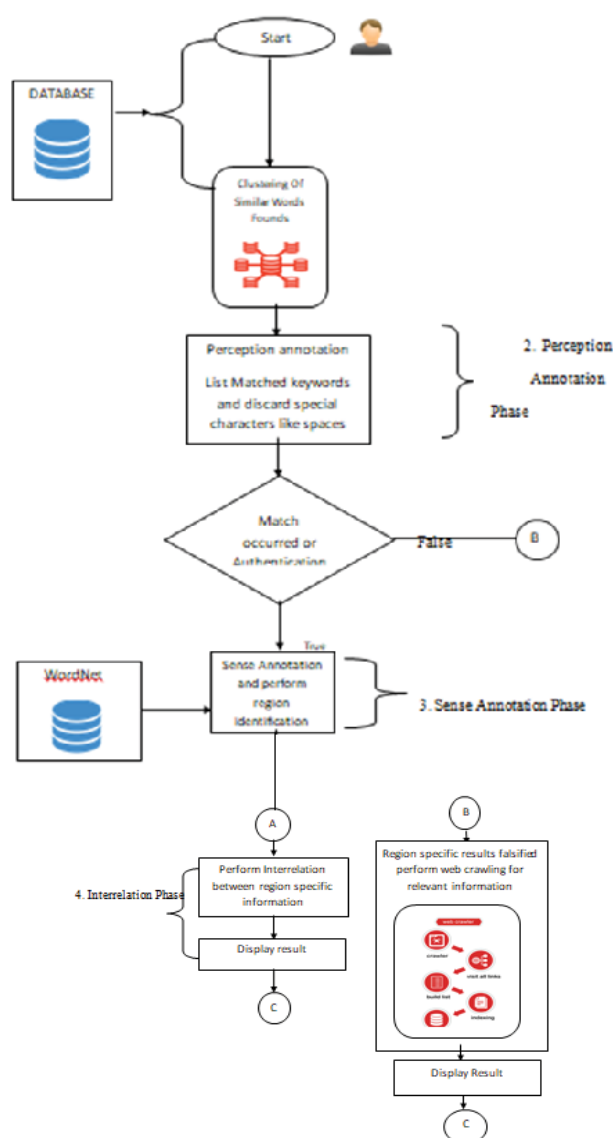


Figure 9: Perception based Semantics model for knowledge Extraction.

Clustering

The clustering is the main mechanism of the proposed model. the words or URL which is most often searched are grouped together in the buffer. If same URL is fetched again and again then web server is not invoked again it will save large amount of time which will be consumed while searching for the URL. clustering done will be more probable values.

Semantic intermediates

This analysis phase will be performed after clustering. In this case the user performances are noted and strings are processed by looking at the earlier search query by the use of semantic intermediates.

Proposed Model

Excepted parametric outcomes

The Parametric outcomes are based on perception based semantics model is to improve the results efficiency; SPBS is the change in user preference from manipulations of seek outcome by search engine providers. Our model re-ranking search results manipulate according to user perception. information extraction enables the generic extraction and minimizes re-processing of data. due to, this provides automated query generation mechanism so that, casual users no need to learn the query language in order to perform mining [40]. The basic function of this model is to extract the information based on user input (Figure 9).

Conclusion

We have conducted comprehensive survey of various techniques associated with word sensing used by optimized search engine. The search waste large amount of time in identify content being searched and extracted meaningful information. In other terms to rectified techniques such as supervised, unsupervised, semi-supervised and knowledge based approaches. The techniques which are analysis lack of clustering mechanism which hencing the execution speed in sensing the desire results. Also this paper compared with existing techniques metrics and demerits and find out that the perception based semantics model for knowledge extraction based system is best among the available techniques. The modification of the existing system in term of improve the pre-processing phase, find the sense of the user query with based on the clustering and reduced exaction time significant.

References

- Gupta V (2007) A keyword searching algorithm for search engines. In: Innovations in Information Technology. 4th International Conference on Dubai.
- Miller GA, Beckwith R, Fellbaum C, Katherine DG, Miller J (1990) Introduction to Word Net. International Journal of Lexicography 3: 235-244.
- Singhal A (2001) Modern information retrieval: A brief overview. IEEE Data Eng Bull 24: 35-43.
- Lee K, Lee J, Kwan MP (2017) Location-based service using ontology-based semantic queries: A study with a focus on indoor activities in a university context. Computers Environment and Urban Systems 62: 41-52.
- Fathalla S, Hassan M, El-Sayed M (2012) A hybrid method for user query reformation and classification. In Computer Theory and Applications (ICCTA). 22nd International Conference Alexandria, Egypt.
- Chandra S, Sultana N, Paira S, Alam SKS (2017) A brief study and analysis of different Searching Algorithms. 2nd Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India.
- Kumar A, Majumdar R, Kakkar A (2015) Search Engine Optimization: A Game of Page ranking. 2nd International Conference on Computing for Sustainable Global Development (INDIA.Com). New Delhi, India.
- Aravindhan R, Shanmugalakshmi R (2014) Comparative Analysis of Web 3.0 Search Engines: A Survey Report. International Conference on Advanced Computing and Communication Systems, Coimbatore, India.
- Cui M, Hu S (2011) Search engine optimization research for website promotion. International Conference of Information Technology, Computer Engineering and Management Sciences. Jiangsu, China.
- Tsai SE, Chen YS, Tsai CY, Tu SW (2010) Improving query suggestion by utilizing user intent. In: Information Reuse and Integration. IEEE International Conference on Las Vegas, USA.
- Makvana K, Shah P, Shah P (2014) A novel approach to personalize web search through user profiling and query reformulation. In Data Mining and Intelligent Computing (ICDMIC), International Conference, New Delhi, India.
- Sturt J, Ali S, Robertson W, Metcalfe D, Grove A, et al. (2012) Neurolinguistic programming: A systematic review of the effects on health outcomes. Br J Gen Pract 62: 757-764.
- Javadi M, Saeid Y, Mokhtari N, Sirati JM (2014) The effect of neuro linguistic programing strategies training on nurses emotional intelligence: A randomized clinical trial study. J Appl Environ Biol Sci 4: 238-244.
- Singh H, Gupta V (2015) Performance analysis of recent Word Sense Disambiguation techniques. 2nd Int. Conf. Recent Adv. Eng. Comput. Sci., Chandigarh, India.
- Sachdeva P, Verma S, Singh SK (2014) An Improved Approach to Word Sense Disambiguation. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Noida, India.
- Mivule K (2016) Web search query privacy, an end-user perspective. Journal of Information Security 8: 56-74.
- Shekarpour S, Marx E, Auer S, Sheth AP (2017) RQUERY: Rewriting Natural Language Queries on Knowledge Graphs to Alleviate the Vocabulary Mismatch Problem. 31st AAAI Conference on Artificial Intelligence, San Francisco, California, USA.
- Chaplot DS, Salakhutdinov R (2018) Knowledge-based Word Sense Disambiguation using Topic Models.
- Pfisterer D, Romer K, Bimschas D, Kleine O, Mietz R, et al. (2011) SPITFIRE: Towards a Semantic Web of Things. IEEE Communications Magazine 49: 1-8.
- Kolte SG, Bhirud SG (2008) Word Sense Disambiguation Using WordNet Domains. First Int. Conf. Emerg. Trends Eng. Technol., Nagpur, India.
- Kolte SG, Bhirud SG (2012) Word Sense Disambiguation using WordNet. 1st International Conference on Emerging Trends in Engineering and Technology 6: 183-199.
- Naskar SK, Bandyopadhyay B (2007) Word Sense Disambiguation Using Extended WordNet. Int. Conf. Comput. Theory Appl., Kolkata, India.
- Dhungana UR, Shakya S, Baral K, Sharma B (2015) Word Sense Disambiguation using WSD specific WordNet of polysemy words. 9th Int. Conf. Semant. Comput. IEEE Anaheim, USA.
- Ustalov D, Panchenko A (2017) A Tool for Effective Extraction of Synsets and Semantic Relations from BabelNet. Siberian Symposium on Data Science and Engineering (SSDSE), Novosibirsk, Russia.
- Chen Y (2010) A knowledge-based approach for Textual Information Extraction from Mixed Text/Graphics Complex Document Images. IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey.
- Sun S, Wang Y (2010) K-Nearest Neighbour Clustering Algorithm Based on Kernel Methods. In Intelligent Systems (GCIS). 2nd WRI Global Congress, Wuhan, China.
- Navigli R, Crisafulli G (2010) Inducing word senses to improve web search result clustering. In Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics.
- Gan G, Ng MKP (2017) K-Means Clustering with Outlier Removal. Pattern Recognit Lett 90: 8-14.
- Trillo R, Gracia J, Espinoza M, Mena E (2007) Discovering the Semantics of User Keywords. J Univer Comput Sci 13: 1908-1935.
- Huang G, Song S, Gupta JND, Wu C (2014) Semi-supervised and unsupervised extreme learning machines. IEEE Trans Cybern 44: 2405-2417.
- Deng J, Dong W, Socher R, Li LJ, Li K, et al. (2009) ImageNet: A large-scale

- hierarchical image database. Cvpr, Miami, FL, USA.
32. Miller GA (1995) WordNet: a lexical database for English. Commun ACM 38: 39-41.
33. Espinoza M, Trillo R, Gracia J, Mena E (2006) Discovering and merging keyword senses using ontology matching. CEUR Workshop Proc, 225: 1-252.
34. Rashidghalam H, Taherkhani M, Mahmoudi F (2006) Text summarization using concept graph and BabelNet knowledge base. Artificial Intelligence and Robotics (IRAN OPEN), Qazvin, Iran.
35. Vandenbussche P, Atemezang YGA, Poveda-Villalón M, Vatan B (2017) Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. Semantic Web 8: 437-452.
36. Shoeleh F, Azimzadeh M, Mirzaei A, Farhoodi M (2016) Similarity based Automatic Web Search Engine Evaluation. Tehran, Iran.
37. Mala V, Lobiyal DK (2017) Semantic and keyword based web techniques in information retrieval. Proceeding - IEEE Int. Conf. Comput. Commun. Autom. (ICCCA), Noida, India.
38. Nora BM, Lemnar C, Potolea R (2010) Semi-supervised learning with lexical knowledge for opinion mining. IEEE 6th International Conference on Intelligent Computer Communication and Processing.
39. Koudas N, Srivastava D (2005) Data Stream Query Processing. 21st Int. Conf. Data Eng. (ICDE).
40. Baeza-Yates R (2004) Web usage mining in search engines. Web Min Appl Tech.