



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: VII Month of publication: July 2017

DOI:

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Resolving Word Sense Ambiguity Using WordNet and SensEval

Subha Mahajan¹, Rakesh Kumar², Vibhakar Mansotra³

^{1,2,3}Department of Computer Science and IT, University of Jammu, Jammu

Abstract: *Sense ambiguity has always been syndicated with human language and its evaluation. It is a significant aspect of any kind of language whether English or Hindi or any other. The words that are persistently used in language are often being assigned to more than one distinct meaning in the real world, thus resulting in ambiguity. One of the biggest challenges faced by Language Translators, Search engines and Machine learning systems is to involve in what sense a specific word is being used. Word Sense Disambiguation (WSD) is a process of computationally identifying the accurate sense of a word in a context. Lot of studies have been carried out over the past few years for WSD and its approaches and algorithm. The paper advocates the use of lexical resources i.e. WordNet for determining the lexical category of target word from SensEval dataset and extract the word class form from WordNet and proposed an algorithm for determining the sense that in which context the word is being used in order to improve the accuracy of information retrieval.*

Keywords: Word Sense Disambiguation, POS, POS Tags, WordNet, SensEval

I. INTRODUCTION

With the advancement in the field of technology and explosive growth of data will boost the systems of data availability and collection to be accurate so that users should be provided with relevant resources because in present scenario people are mainly depended on the web search engines for searching any kind of data. When a person is searching any kind of information he/she expects the relevant information especially when there is a case of information overload. It gets difficult for the user to get relevant information in any language when the word or phrases have more than sense. The goal of retrieving all and only the most relevant information is still a far cry. For example, the phrase “she her child in an accident”, the word ‘accident’ means a crash, while in the “”, in this sentence the word “second” refers to the position in series. The problem can be reduced upto an extent by the concept of disambiguation of a word. When a word has more than one meaning then it is probably considered an ambiguity. Hence, Word sense disambiguation (WSD) is termed as an open problem of an ontology and natural language processing with a process of identifying a correct sense of a word in a given context. WSD is the empowering automation for applications such as Information Retrieval, Machine Translation, Information Extraction and Speech Recognition.

WSD plays important role in improving the quality of the computer so as to comprehend in what sense a specific word is used. WSD was first formulated as a distinct ciphering task during early days of machine translation in late 1940s, making one of the oldest problem of computational semantics. The problem was continued as a challenging task until there was a availability of resources. In 1980 there was prodigious development in the area of WSD research when a large scale lexical resources and corpora came into existence. In 1990s, NLP provided three major developments for WSD: online dictionary WordNet which is used as an online sense inventory, statistical methodologies which are used as sense classification problems and SensEval which includes training and testing data and dictionary is mapped with WordNet. Different others SensEval evaluation exercises have come up for machine translation in different languages like French, Italian, German etc. so that researchers can share and upgrade their views in this research area.

A. WordNet

WordNet is an electronic lexical resource based on lexical category which contains English noun, verb, adverb, adjective which are organised into synonym sets, each representing one underlying lexical concept [1]. WordNet plays important role in the work described in this paper. Synonymy and Polysemy does not make it possible to have one to one mapping of word forms to the meanings. This was the creativity behind the consortium of the WordNet into synset. A synset is a set of synonymous words whose main purpose is to apprehend a unique meaning. For example the word ‘accident’ can have two different meanings, viz. a mishap and a chance event. The synonym sets, {accident, mishap} and {accident, chance event} can serve as unambiguous designators of these two meanings of ‘accident’. The most important relation for WordNet is synonymy which marks the similarity of meaning,

since the ability to judge the relation between word forms is a prior condition for the representation of meanings. The next familiar relation is antonymy which is for a given word is a word with opposite meaning. Antonymy provides the central organizing principle for the adjectives and adverbs in WordNet. Unlike synonymy and antonymy which are lexical relations between word forms, hypernymy / hyponymy is a semantic relation between word concepts i.e, synsets.

B. Parts-of-Speech (POS)

Parts-of-speech also known as POS, word classes, or syntactic categories POS are useful as they large amount of information of a word and its neighbours. They determine the English word classes which are based on syntactic and morphological function, grouping words that have similar neighbouring words. syntactic structure around the word make part-of-speech tagging an important component of syntactic parsing and morphological affixes which can influence stemming to enhance the performance of informational retrieval, and can help in summarization for improving the selection of nouns or other important words from a document. Parts of speech are useful features for finding named entities like people or organizations in text and other information extraction task. Parts-of-speech can be divided into two broad categories: Closed class types which includes relatively fixed membership, such as prepositions—new prepositions are rarely coined. Closed class words are generally stop words like of, it, and, or you, which tend to be very short, occur frequently, and these stop word often have structuring uses in grammar. Secondly is the Open class types which includes nouns and verbs in which new nouns and verbs like iPhone or to fax are continually being created or borrowed. Any given corpus may have different open class words, but sufficiently large corpora, likely share the set of closed class words. Four major open classes occur in the languages of the world: nouns, verbs, adjectives, and adverbs. English has all four, although not every language does. The syntactic class noun includes the words based on people, places, name or things, and others as well. The verb class includes most of the words referring to actions and processes thus forming the main part of the predicate of a sentence, such as hear, become, happen, draw, go etc. Adjectives class includes many terms for properties or qualities. Most languages have adjectives for the concepts of color (blue, red), age (old, adult), and quality (nice, cruel), but there are languages without adjectives. Adverbs class form , that which amends the meaning of an adjective, verb, or other adverb, expressing manner, place, time, or degree

C. Part-of-Speech Tagging

Part-of-speech tagging (tagging for short) is the process of assigning a part-of- speech marker to each word in an input text, after the tokenization process takes place which includes partitioning of the query , removal of unwanted symbols like commas, quotation marks, punctuations etc., from words and removal of stop words. Output of tokenized words are input in to a tagging and words are assigned with appropriate tagset, and the output is a sequence of tags, a single suitable tag for each word is done. Thus tagging is a disambiguation task, words are ambiguous i.e. they have more than one possible part-of-speech and the idea is to find the correct tag in a particular context. While there are different lists of parts-of-speech in English ,but the main tags which are generally used all over the world are adjective noun, adverb, and verb

II. LITERATURE REVIEW

Udaya Raj Dhungana et al. [2]developed a new model of WordNet that organizes the different sense of polysemy words and also single sense words based on the clue words is used to disambiguate the correct sense of polysemy words from the clue words which are noun, verb, adjective or adverb. The idea is applied with the help of Knowledge based Word Sense Disambiguation (WSD) algorithms in order to find the correct meaning of the word in a given context.

Khaled Abdalgader [3]proposed a computationally efficient word sense disambiguation method that improves the text similarity measures through correct identification of word sense .Also compare the WordNet glosses of the target word with a context vector which comprise the remaining words in the text fragment surrounding the target word. Paper also describes the measure of short text similarity instead of word overlap and evaluates the resulting method on the benchmark Microsoft Research Paraphrase Corpus in order to show the significant improvement in performance of word sense disambiguation .

Neha Kumari [4]surveyed the different research papers which suggests different methods to calculate to calculate the similarity of sentences and they mainly focus on the one or more feature for example word, structure and semantic information etc. The objective was to find various techniques used for calculating similarity between two sentences , even if the sentences do not share similar surface form in order to retrieve useful information out of large amounts of data and also used for question answer portals.

Prity Bala [5]developed a WSD tool using knowledge based approach with Hindi WordNet as external lexical resources, which is built from co-occurrence and collocation and it includes synset or synonyms which belong to either noun, verb, adjective, or

adverb, and are called as part-of- speech (POS) tagger. It also involves incorporation of word knowledge to disambiguate words and specify exactly, which sense of several word is being used.

F.B. Dian Paskalis [6] implemented WSD in Lucene using query expansion with thesaurus and relevance feedback and again re-implemented the Extended Lesk algorithm to disambiguate the query using WordNet. The approach uses 20 words for the iteration process and provides better understanding of WSD in information retrieval system performance. Author have also analyzed the implementation of query expansion, word sense disambiguation (WSD), iterated relevance feedback, and some retrieval variations.

Roshan Karwa [7] proposed a hybrid approach of lexical knowledge and world knowledge of English WordNet which developed at Princeton University. Also SemCor corpus and the JAWS library (Java API for WordNet searching) has been used for this purpose for resolving ambiguity in a sentence and provide information to the users.

Francisco João Pinto [8] describes the experimentation performed to test the effectiveness of automatic query expansion and word sense disambiguation (WSD) using short and long query on a topic TREC under vector model using linguistic information extracted from WordNet. And observed that short queries and long queries is not able to improve retrieval performance without the selection of the correct meaning of the words but the results are better using short queries.

Samhith.K et al. [9] presented a methodology so as to disambiguate the word senses of polysemous words with the help of Lexical Categories present in WordNet a English Lexical database. They have also applied an algorithm to the data scraped from Wikipedia pages of words belonging to the lexical category of given words which is determined by using neighbouring words of its context so as to find the correct sense of word with the help of modified Lesks algorithm.

Ravi Sinha [10] presented a comparative evaluations using several measures of word semantic similarity and several algorithms for graph centrality based on unsupervised graph-based method for word sense disambiguation so as to enhance the performance competing with the state-of-the-art in unsupervised word sense disambiguation, when the right combination of similarity metrics and graph centrality algorithms is performed on standard data sets.

K.Neeraja [11] observed the current state of the art of WSD and surveyed on the different approaches of WSD which reduces the ambiguity of words in telugu language and used dictionary as WordNet. Also studied approaches for word sense disambiguity for telugu nouns and concluded the problem of word sense disambiguation by a combination of different machine learning algorithms.

Ioannis P. Klapaftis [12] presented a description of SenseEval -2010 dataset used for WSI task and a new evaluation setting for sense induction methods. Also contributed their research firstly on detailed analysis of SenseEval 2010 WSI evaluation results and identify the shortcomings of current evaluation measures. Secondly, defining the methods able to perform well above the Most Frequent Sense (MFS) baseline in highly skewed distributions.

Dr. Bushra Kh. AlSaidi [13] proposed a genetic algorithm to solve word sense disambiguation problem which can automatically identify the intended meaning of word in context without any additional resource and can be effectively applied for many optimization problems. The algorithm is evaluated so as to produce lot of sense to the ambiguities of word and thus system creates dynamic, and up-to-date word sense in a highly automatic method.

III. PROPOSED WORK AND EXPERIMENTAL SETUP

The proposed method involves POS tagging of the sentence so as to extract the nouns from the sentences and also get the noun forms of the word(target word). Then that target word is compared with the training corpus of SenseEval 1 an online dataset which is provided by sussex is used for semantic evaluations. The SenseEval dataset contains 35 words. We have taken 10 different sentences containing different ambiguity and correspondingly taken training and testing corpus of SenseEval 1 dataset along with a dictionary key as sense inventory which are mapped with lexical database WordNet 1.5 and WordNet 1.6 that provides distribution according to the part of speech. The target word has different senses and each sense had a numerical unique identifier, which is formatted in the form of <tag sense number>, which is further used in our analysis. The process is explained below in the figure 1.

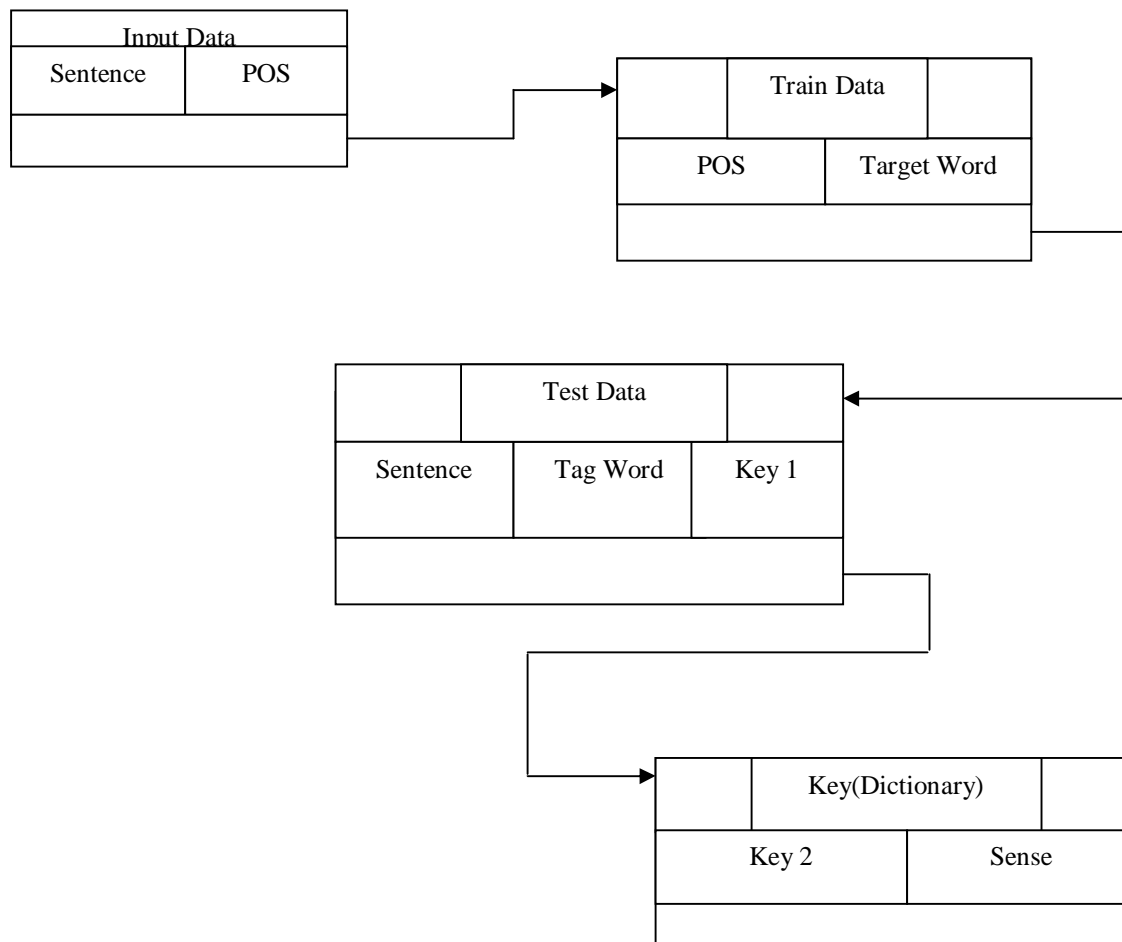


Figure1: Showing interface relationship between the datasets used during the experiments

IV. RESULTS AND DISCUSSION

The analysis of the text was carried out on 10 sentences of 10 different target words which involves three phases, in the first phase the noun forms of the targeted sentences are extracted from Table 1. The detailed description of the sentences with respect to Part-of-speech are described in Table 2 which is done with the help of POS tags which are provided by Penn Treebank tagset as shown in Table 3. These extracted nouns forms are matched against train data and score is generated based on the number of matches found as shown in Table 4. In the phase two, the generated table is again matched with test data to generate a unique tag code which can be seen in Table 5. we have taken 5 target words each in Train and Test data in Table 4 and Table 5 respectively as sample in order to show that how the process takes place. In the third phase these tag codes are matched against the codes present in the dictionary database to find the exact sense pertaining to each sentence provided as an input to the model and finally the results are tabulated in Table 6.

TABLE I

Table 1: Showing input text with target word

Sno.	Target Sentence	Target word
1	There have been ten major accidents at nuclear power plants since the first at Windscale in 1957	accident
2	The band of boys gave a great performance	band
3	It is a fair bet that most Independent readers did more than their share of the nation's borrowing last year	bet
4	we have an excess of electricity demand at the moment	excess
5	I never saw my float as the fish cruised up and down	float

6	Onions are good for digestive system	onion
7	she kept her promise of not telling our secret	promise
8	When I offered to carry her sack she waved me aside	sack
9	Give the scrap of meat to the dogs and watch them fight	scrap
10	There is then no effective sanction against illegal tapping	sanction

TABLE II

Table 2: Showing POS Tagging of input text

Sno.	POS Tagging	Extract Nouns
1	"There/EX" "have/VBP" "been/VBN" "ten/CD" "major/JJ" "accidents/NNS" "at/IN" "nuclear/JJ" "power/NN" "plants/NNS" "since/IN" "the/DT" "first/JJ" "at/IN" "Windscale/NNP" "in/IN" "1957/CD"	"accidents" "power" "plants" "Windscale"
2	"The/DT" "band/NN" "of/IN" "boys/NNS" "gave/VBD" "a/DT" "great/JJ" "performance/NN"	"band" "boys" "performance"
3	"It/PRP" "is/VBZ" "a/DT" "fair/JJ" "bet/NN" "that/IN" "most/JJS" "Independent/JJ" "readers/NNS" "did/VBD" "more/JJR" "than/IN" "their/PRP\$" "share/NN" "of/IN" "the/DT" "nation/NN" "'s/POS" "borrowing/NN" "last/JJ" "year/NN"	"bet" "readers" "share" "nation" "borrowing" "year"
4	"we/PRP" "have/VBP" "an/DT" "excess/NN" "of/IN" "electricity/NN" "demand/NN" "at/IN" "the/DT" "moment/NN"	"excess" "electricity" "demand" "moment"
5	"I/PRP" "never/RB" "saw/VBD" "my/PRP\$" "float/NN" "as/IN" "the/DT" "fish/NN"	"float" "fish"
6	"onions/NNS" "are/VBP" "good/JJ" "for/IN" "digestive/JJ" "system/NN"	"onions" "system"
7	"she/PRP" "kept/VBD" "her/PRP\$" "promise/NN" "of/IN" "not/RB" "telling/VBG" "our/PRP\$" "secret/NN"	"promise" "secret"
8	"Give/VB" "the/DT" "scrap/NN" "of/IN" "meat/NN" "to/TO" "the/DT" "dogs/NNS" "and/CC" "watch/VB" "them/PRP" "fight/VB"	"scrap" "meat" "dogs"
9	"When/WRB" "I/PRP" "offered/VBD" "to/TO" "carry/VB" "her/PRP" "sack/NN" "she/PRP" "waved/VBD" "me/PRP" "aside/RB"	"sack"
10	"There/EX" "is/VBZ" "then/RB" "no/DT" "effective/JJ" "sanction/NN" "against/IN" "illegal/JJ" "tapping/NN"	"sanction" "tapping"

TABLE III

Table 3: part-of-speech tags

Tag	Description
JJ	Adjective
JJR	adjective , comparative
JJS	adjective ,superlative
NN	noun ,singular or mass
NNS	noun, plural
NNP	proper noun ,singular
NNPS	proper noun , plural
RB	Adverb
RBR	adverb , comparative
RBS	adverb superlative
VB	verb base form
VBD	verb past tense

VBG	verb gerund
VBN	verb past participle
VBP	verb non-3sg pres
VBZ	verb 3sg pres

TABLE IV

Table 4: Showing Train Data

Sno.	Target Word	Train Data
1	accident	"as there have already been two accidents there, barriers were suggested."
2	accident	"this results from lower than expected north sea oil production, which is taking longer than expected to recover from the serious accidents of the last two years."
3	accident	"finally the men gave up, grudgingly admitted that the room was safe, and left muttering about `freak accidents and `acts of god"
4	accident	"if you have two drivers like this in formula one you would have accidents every week.
5	accident	"and what we have today are entirely modern creations, syntheses, the result of countless unrecorded accidents rather than the national spirit seeking to realise itself."
6	accident	"two roundabouts were likely to give rise to more accidents he said. "
7	accident	"most holiday injuries we see are caused by scooter accidents."
8	accident	"and on how often you have expensive accidents
9	accident	"the accident survey showed june was the worst month for accidents, with 59; november (55) came second. "
10	accident	"the least number of accidents was in december &dash. only 36. "
11	band	"bowllly was singing there with the roy fox band"
12	band	"indeed a tremendous compliment to the band and manager scott piering welcomed the exercise. "
13	band	"he planned to schedule it as the band began the tour which was to net them #100m each."
14	band	"the single is already big in the clubs and has been remixed by the band"
15	band	"so when his band primal scream released `loaded\", one of the best winter dance records, everything seemed sweet."
16	band	"the band approached the event with some trepidation."
17	band	"they are the last rock'n'roll band before a new and better rock'n'roll race creed begins."
18	band	"you've got to hear the band sounding crap sometimes. "
19	band	"the recording quality is much superior to that given to the hanover band or even norrington. "
20	band	"the hothouse flowers are the band most easily tagged with the title nouveau hippies "
21	bet	"a friend of mine followed us round and he was starting to panic because he'd had a bet on the boss.
22	bet	"(a fair bet, though, is that the remarkable anita roddick will adjust the orientation of the body shop chain to her maturing customers.) "
23	bet	"minimum bet is #2, maximum #3,000. "
24	bet	"it is probably more than coincidence that protection was the best backed horse for the race yesterday, being cut to 9-2 from 11-2 by the sponsors, after they laid a bet of #2,000 each-way. "
25	bet	"a vague pledge to boost direct-grant schools looks a good bet. "
26	bet	"by the end of the season it looked a daft bet, but it almost came off when kerrera finished second to her stable companion, musical bliss. "
27	bet	"that sort of performance makes him look a sound bet in a hennessy field that is certainly not a vintage one. "

28	bet	"fanatical (next best 2.35), who finished third to firelight fiesta over the course in a race that has worked out exceptionally well, looks another sound bet. "
29	bet	"before the invention of the sliding fader your best bet was either to encourage the artist to leave the studio in the song's later stages or, in the case of an orchestra, to walk out with the microphone. "
30	bet	"it's a fair bet you would end up with something very close to the 348.
31	excess	"we drink and eat in the daytime but not at night and so there is an excess of fluid and salts in the body in the daytime, and a deficit during the night. "
32	excess	"whereas middle and upper classes have negotiated and established salary levels far in excess of need over a period of decades! "
33	excess	"the r and a's projected attendance figure for the week is around 225,000 &dash. in excess of the record 205,000 at royal lytham in 1988 and a massive increase on the 193,000 who swarmed over the old course in 1984."
34	excess	"sometimes there is an excess clause whereby you have to meet the first #25 or #50 of any damage for which you are claiming."
35	excess	"but now the anticipated figure is in excess of #600m . "
36	excess	"it still cannot arise where the claimant has capital in excess of #6,000 &dash. and where capital is between #3,000 and #6,000, the notional income taken into account from it may, when added to other income, rule out this particular type of help. "
37	excess	"the department of transport's invention about the cost of fatal road accidents becomes transmuted, through several magnification processes, into the anti-alcohol lobby's bald statement that `at 1987 costs, the deaths and injuries from alcohol-related road accidents represent a cost in excess of #800 million". "
38	excess	"the company says all the motorist has to pay for is the initial test and the re-test and a small excess. "
39	excess	"the cost of a successful concierge system in a typical london borough is in excess of #200 million. "
40	excess	"it's a very extreme film; a story about fame and excess, about people who acquire a very specific audience, people who really watch them and reflect them."
41	float	"i have set up a mormark microlite 13 foot float rod, a shimano 2500s reel loaded with 2 lb maxima coupled with a 1.5 lb double strength hook length. "
42	float	"to save time i make up my float rigs beforehand and put them on pole winders."
43	float	"on tackle, when fishing a waggler, arthur attaches a swivel to the float so `it falls over nicely".
44	float	"he explains why you don't use light lines in a wind; how to shot a stick float and how to keep squats and maggots fresh."
45	float	"the kidnapping of plastic ducks belonging to those involved in the water privatisation is not confined to schroders, the merchant bankers to the float "
46	float	"groaning under the weight of their wallets, the government's advisers on water privatisation made a predictable splash at monday night's auction of irreverent artefacts from the float"
47	float	"bream and roach showing off the ferry boards and boatyard stretches to stick float or feeder."
48	float	"water float "
49	float	"but instead of trying to create a new currency which can be pegged, the new economy minister, erman gonzalez, was expected last night to announce a free float of the austral against the dollar.
50	float	"a hit with stick float anglers but could be won on the pole or waggler "

TABLE V
Table 5: Showing Test Data

Sno.	Target Word	Test Data	Tag Code	Sense
1	accident	"The accident appeared to have little effect on the Christmas party, except to lengthen it considerably. "	532675	crash
2	accident	"The accident occurred on the Saturday of the annual Popular Flying Association (PFA) rally at Cranfield."	532675	crash
3	accident	"More than 250,000 people have to go to hospital each year because of accidents in the kitchen."	532675	crash
4	accident	"We all read accident reports."	538889	crashmod
5	accident	"It had been in an accident and it was worthless."	532675	crash
6	accident	"The doctors suggest a strategy to combine the promotion of helmets with moves to reduce alcoholism, and therefore drink-related accidents."	532675	crash
7	accident	"However, it will be met with anger by Eurotunnel's critics who argue that the risk to the public will be increased in the event of accidents or fires inside the tunnel."	532675	crash
8	accident	"unhurt from a heavy accident while testing tyres at Estoril last week."	532675	crash
9	accident	"We know that all of us will eventually die from disease, natural disaster, accidents or whatever."	532675	crash
10	accident	"All the legislation in the world couldn't have coped with an accident of such violence."	532675	crash
11	band	"How can a band of vociferous and woolly-minded objectors deny the hope of life to thousands of young sufferers by stopping this vital research for the sake of a few unformed embryonic cells?"	532735	group
12	band	"In the process they've lost a certain edge, for at the end The Cateran had the rage and the fury of a band at its peak falling apart."	532736	mus
13	band	"Meanwhile, the show is nearly over and the band strike up `I Love You Love`."	532736	mus
14	band	"The band have recorded `Just Like tom Thumb's Blues for Imaginary's Bob Dylan compilation."	532735	group
15	band	"Mr McCormick submitted that if any one of the mother's reasons was possibly valid then it could not be said that the mother's refusal to consent was outside the reasonable band"	532734	range
16	band	"The band who have just finished a tour supporting Soul II Soul are currently playing PAs around the county."	532735	group
17	band	"The malts vary so much that a growing band of devotees collects them, comparing the Lowlanders and Highlanders, or the product of one island with that of the next."	532735	group
18	band	"These countries, too, have their unreconstructed fascists and nazis, their residual lunatic right-wing fringe, and their broader bands of sympathizers with various aspects of rightist thinking. "	532734	range
19	band	"Nowadays they are merely noisy in the manner of a rock band"	532736	mus
20	band	"Paul Ince joins the growing band of midfielder-turned defenders with a fine display in Manchester United's televised 1-0 win over Nottingham Forest, but United show few signs of having shaken off their current malaise."	532735	group
21	bet	"Market research shows that more people wish to bet on evening racing than on Sunday fixtures."	519007	gamlbr

22	bet	"You're just betting that if you call me a coward I'll be angry enough to do it."	520051	speculate
23	bet	"But it is a fair bet that all conductors of Otello on record have learnt from his example &dash. each of the current versions has a fine dramatic sweep and none is seriously out of touch with the idiom."	519912	probability
24	bet	"In fact, one Edinburgh newspaper has been insisting since the selection was unveiled that the latest capital XV are unworthy of even an each-way bet "	519914	wager
25	bet	"The horse closest to Desert Orchid in the Gold Cup betting is the Irish crack, Carvill's Hill, who is still suffering from the recurring back problem that has dogged his career. "	520056	odds
26	bet	"That gave them their one-way bet" "	520057	speculation
27	bet	"it is a fair bet that a good many punters were not amused. "	519912	probability
28	bet	"the best bet now looks to be the alliance &and. leicester." 519912	519912	probability
29	bet	"it's a fair bet that if a certain white horse could talk, he would put his hoof on his heart and say `nobody does it better"519912	519912	probability
30	bet	at all events, they should catch the film while they can, for it is a fair bet that he will no longer be saving christmas come 25 december."	519912	probability
31	excess	"But if you let your baby overheat, excess sweat can irritate the skin folds, and there is the (rare) danger of heat stroke.	512405	toomuch
32	excess	"If, however, your Account is overdrawn we may require any amount outstanding in excess of one month's transfer to be repaid."	512402	morethan
33	excess	"The increase partly reflects a relaxation in the issue of Monetary Stabilisation Bonds, an instrument used by the monetary authorities to absorb excess liquidity."	512405	toomuch
34	excess	"Fanaticism or fundamentalism are precisely the kind of excess which dies out because of the evolutionary process."	512403	aglut
35	excess	"The overweight person, on the other hand, is usually not so mindful of internal sensations as a cue for eating or abstaining, and is less likely to take exercise as an antidote to excess indulgence."	512405	toomuch
36	excess	"If any individual issue is subscribed less than 1.75 times then the excess shares clawed back will be held by the Government."	512405	toomuch
37	excess	"English officials have reacted to press excesses with an amateurish petulance."	512404	ott
38	excess	"But we didn't adopt the Italian approach to styling, which has too much excess detail."	512405	toomuch
39	excess	"This demonstrates how property values have risen, and if this rate of increase continued over the coming 10 years then by the year 2000 the average house price would be in excess of #160,000."	512402	morethan
40	excess	"Another ruse was to offer houses at low rent or for sale, or to offer signing-on fees far in excess of the #10 theoretically permitted by the FA. "	512402	morethan
41	float	"The leading barge was tilted at right angles and the rest floated gently apart."	523379	swimadv
42	float	"It is shrugging off current market conditions to float the business via an introduction sponsored by Schroders, the merchant bank."	523224	shares
43	float	"Nicholas Ridley, the Secretary of State for Trade and Industry, refused to rule out the possibility of the Government dropping its veto on takeover bids for other state-owned businesses which have been floated on the stock market."	523224	shares

44	float	"The orange river picked up the rejection letter and open cigarette packet in its path and floated them over onto X's lap."	523379	swimadv
45	float	"Fishing at Sandford with a stick float last week, Richard took a cracking bag of fish weighing 41.0.0, half of which were dace!"	523248	rod
46	float	"On tackle, when fishing a waggler, Arthur attaches a swivel to the float so `it falls over nicely" 523248	523248	rod
47	float	"Navratilova next misjudged a floated backhand which nicked the line, and a forehand winner followed by four first serves in the next game put Graf back on course at 5-5."	523280	sport
48	float	"Many believe this could be the time to try the package of reforms floated recently by Dr Nikolai Shmelyov in Moscow. "	523236	idea
49	float	"On set, Shrader devotes most of his efforts to working out the movements of what he calls his floating rectangle" &dash."	523310	app
50	float	"To many devout Hindus, the sanctity of the Ganges is inviolable, and no amount of raw sewage, floating garbage or charred corpses can detract from its essential holiness."	523373	swimming

TABLE VI

Table 6:Showing Tag code with sense

Sno.	Word	Tag code	Sense
1	accidents	532675	crash
2	accidents	538889	crashmod
3	band	532734	range
4	band	532735	group
5	band	532736	mus
6	bet	519907	gamblr
7	bet	519912	probability
8	bet	519914	wager
9	bet	520051	speculate
10	bet	520056	odds
11	bet	520057	speculation
12	excess	512402	morethan
13	excess	512403	aglut
14	excess	512404	ott
15	excess	512405	toomuch
16	excess	512830	overind
17	float	523224	shares
18	float	523236	idea
19	float	523248	rod
20	float	523280	sport
21	float	523310	app
22	float	523373	swimming
23	float	523379	swimadv
24	onion	528344	plant
25	onion	528347	veg
26	promise	537527	vow
27	promise	537614	ingadj
28	promise	537626	success

29	promise	538409	keep
30	scrap	534532	bit
31	scrap	534533	cutup
32	scrap	534534	cancel
33	scrap	534544	woodhams
34	scrap	534560	scbook
35	scrap	534577	morsel
36	sack	504751	dismiss
37	sack	504753	firing
38	sack	504755	material
39	sack	504756	bag
40	sanction	506284	penalty
41	sanction	506385	econaction
42	sanction	506429	approve

V. CONCLUSION

The model was able to evaluate the target sentences with an accuracy of 60% successfully for the target sentences belonging to 10 different categories as shown in table 1. The target word is giving appropriate sense but also giving the sense close to the context in which the word is being used. It is thereby concluded that lexical based analysis although is time consuming but can efficiently evaluate the sentences with the said accuracy. In the future plan large dataset for test and train can be used along with more entries in the dictionary to enhance accuracy in information retrieval.

REFERENCES

- [1] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller, "Introduction to WordNet: An On-line Lexical Database"(Revised August 1993)
- [2] Udaya Raj Dhungana, Subarna Shakya, Kabita Baral and Bharat Sharma, "WORD SENSE DISAMBIGUATION USING WSD SPECIFIC WORDNET OF POLYSEMY WORDS", International Journal on Natural Language Computing (IJNLC) Vol. 3, No.4, August 2014
- [3] Khaled Abdalgader, "Text-Fragment Similarity Measurement using Word Sense Identification", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 24 (2016)
- [4] Neha Kumari, Sukhbir Kaur, "Online Assessment of Similarity between Sentences in Question Analogous System: A Review Paper", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 5, May 201
- [5] Prity Bala, "Knowledge Based Approach for Word Sense Disambiguation using Hindi Wordnet", The International Journal Of Engineering And Science (IJES), Volume 2, Issue 4
- [6] F.B. Dian Paskalis, M.L. Khodra, "Word Sense Disambiguation In Information Retrieval Using Query Expansion", 2011 International Conference on Electrical Engineering and Informatics 17-19 July 2011,
- [7] Roshan Karwa and Manoj Chandak, "AN HYBRID APPROACH TO WORD SENSE DISAMBIGUATION WITH AND WITHOUT LEARNED KNOWLEDGE", International Journal on Natural Language Computing (IJNLC) Vol. 4, No.2, April 2015
- [8] Francisco João Pinto, Carme Fernández Pérez-San Julián, "Automatic query expansion and word sense disambiguation with long and short queries using WordNet under vector model" Actas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos, Vol. 2, No. 2, 2008
- [9] Samhith.K, Arun Tilak.S, G.Panda, "Word Sense Disambiguation using WordNet Lexical Categories", International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016
- [10] Ravi Sinha and Rada Mihalcea, "Unsupervised Graph-based Word Sense Disambiguation Using Measures of Word Semantic Similarity", Department of Computer Science and Engineering University of North Texas ravisinha
- [11] K.Neeraja, Dr. B. Padmaja Rani, "Approaches for Word Sense Disambiguation: Current State of The Art", International Journal of Electronics Communication and Computer Engineering, Volume 6, Issue (5) Sept, 3rd National Conference on Research Trends in Computer Science & Technology NCRCTCST-2015,
- [12] Ioannis P. Klapaftis, Suresh Manandhar, "Evaluating Word Sense Induction and Disambiguation Methods", Springer Science+Business Media Dordrecht 2013 Lang Resources & Evaluation
- [13] Dr. Bushra Kh. AlSaidi, "Automatic Approach for Word Sense Disambiguation Using Genetic Algorithms", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 1, 2016



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)