Impact of web query morphology and ambiguity on search engine's performance

Anand P. Dwivedi
M.I.P.S, Kanpur, India
dwivedi_anand@hotmail.com,

M. Tech., Banasthali University
Rajasthan, India
mishra.arti199@gmail.com

Sanjay K. Dwivedi
B.B.A. University, Lucknow, India
skd200@yahoo.com

Prof. V.K. Pathak
H.B.T.I., Kanpur, India
vinaypathak.hbti@gmail.com

Abstract-The effectiveness of Internet search engines are often hampered by the deficiencies in user queries and the reluctance or inability of users to build less ambiguous multi-word queries. This is mainly because of the language morphologies and word ambiguities. Neither simple query expansion techniques nor enhanced indexing mechanisms have been satisfactory in addressing these problems, because these methods do not consider the user context or knowledge of the problem domain. This paper covers the comprehensive analysis of web queries in English to know the impact of morphology (especially root word) and ambiguity on the Google search engine. We have used different query sets to test each of these aspects. Our results show that the search engines normally find themselves incompetent to understand and resolve these issues on behalf of user. Therefore, the mean average precision of Google search engine has reduced in case of the query set having any type of ambiguities as compared to the other queries.

Keywords: query ambiguity, morphology, search engine, precision

I. Introduction

The term '**morphology**' refers to the study of the internal structure of words, and of the systematic form-meaning correspondences between words.

Morphology is the study of the structure of words. The structure of words can also be studied to show how the meaning of a given morpheme, or its relation to the rest of the word, varies from one complex word to another. Consider how sun works in the following words: sunbeam, sunburn, sundial, sunflower, sunglasses, sunlight, sunrise, and sun-spot (scientific sense), and sun-spot (tourist sense), and suntan. Inflection does not really yield "new" words, but alters the form of existing ones for specific reasons of grammar. Derivation, on the other hand, does lead to the creation of new words.

[2]Morphology is the field of linguistic which studies word structure and formation. It is composed of [3,4]*inflectional morphology* and *derivational*

Arti Mishra

Morphology. Inflection is defined as the use of morphological methods to form inflectional word form from a lexeme. Inflectional word forms indicate grammatical relations between words. Derivational morphology is concerned with the derivation of new words from other words using derivational affixes.

Compounding is another method to form new words. A compound word (or a compound) is defined as a word formed from two or more words written together. The component words are themselves independent words (free morphemes).

A morpheme is a smallest unit of a language which has a meaning. [1,4]Morphemes are classified into free morphemes and bound morphemes. Free morphemes appear as independent words. For e.g. In English, {red}, {house} and {when} are free morphemes. Bound morphemes do not constitute independent words, but are attached to other morphemes or words. Bound morphemes are also called affixes.

Morphological structure of English language has a great impact on the performance of the search engines[1]. In this study we have focused on two factors of language morphology that can change or modify a web query i.e. query with root word and query with various senses.

II. METHODS OF EVALUATION OF SEARCH ENGINES

Following methods are used for the evaluation of search engines:

(i) Precision (P): is the fraction of retrieval documents that are relevant. A high precision means that everything returned was a relevant result, but one might not have found all the relevant items (which would imply low recall).

There are variations in the ways of the precision is calculated. [7]TREC almost always uses binary relevance judgments-"either a document is relevant to a query or it is not". [6]Chu & Rosenthal (1996) used a three-level relevance score (relevant, somewhat relevant, and irrelevant) while Gordon and Pathak

(1999) used a four-level relevance judgment (highly relevant, somewhat relevant, somewhat irrelevant, and highly irrelevant).

(ii) Recall (R): is the fraction of relevant documents that are retrieved. A high recall means we haven't missed anything but we may have a lot of useless results to sift through (which would imply low precision). But Recall is a difficult measure to calculate because it requires the knowledge of the total number of relevant items in the collection. Chu & Rosenthal's Web search engine study omitted recall as an evaluation measure because they consider it "impossible to assume how many relevant items are there for a particular query in the huge and ever changing Web systems[6]"

Based on the documents retrieved by a search engine (relevant, non relevant), Table 1 below shows the method of computations of precision and recall

TABLE 1. : PRECISION AND RECALL COMPUTATION TABLE

	Relevant	Nonrelevant
Retrieved	True positives (tp) - Correct result	False positives (fp)- Unexpected result
Not retrieved	False negatives (fn) - Missing result	True negatives (tn) - Correct absence of result

The precision and recall can be calculated by the formula shown below:

Precision = tp/ (tp+fp)

Recall = tp/ (tp+fn)

(iii) *Mean Average Precision* (MAP): Most standard among the TREC community is *Mean Average Precision* (MAP), which provides a single-figure measure of quality across recall levels. Among evaluation measures, MAP has been shown to have especially good discrimination and stability. For a single information need, Average Precision is the average of the precision value obtained for the set of top k documents existing after each relevant document is retrieved, and this value is then averaged over information needs.

MAP = Average Precision/ No. of queries

When a relevant document is not retrieved at all, the precision value in the above equation is taken to be 0.

These methods are used as the users always want to desired documents, and can be assumed to have a certain tolerance for seeing some false positives providing that they get some useful information. The measure of precision and recall concentrate the evaluation on the return of true positive, asking what percentage of the relevant documents have been found and how many false positive have also been returned.

III. EVALUATION METHODOLOGY

1. Selection of Test Queries: The U.S. National Institute of Standards and Technology (NIST) have run a large IR test based evaluation series since 1992. Within this framework, there have been many tracks over a range of different test collections, but the best known test collections are the ones used for the TREC Ad Hoc track during the first eight TREC evaluations between 1992 and 1999. TRECs 6 through 8 provide 150 information needs over about 528,000 newswire and Foreign Broadcast Information Service articles. In this work, we have framed the queries based on the TREC pattern and also from the web search engine's log. So our set of test queries used for the evaluation of search engines in this study have a good mix of standard TREC queries and actual user queries from the search engine's log.

2. Human Relevance Judgments: one of the important issues in performance evaluation of search engines is that whenever human relevance judgment is used, there is a variation in who makes the judgments. TREC leaves relevance judgments to experts or to a panel of experts (Voorchees & Harman, 2001)[7.] However some other researchers (e.g. Chu and Rosenthal, 1996)[6] used human relevance judgment made by researchers themselves. [12]Gordon and Pathak (1999) emphasized that relevance judgments can only be made by individual with the original information need. In this study, the human relevance judgments have been done using a mix of the approaches followed by Voorchees et.al (2001) and Chu et.al. (1996).

3. Precision: There are variations in the ways how precision is calculated. In this study, the precision is calculated on the binary relevance judgment approach followed by TREC[5] - "either a document is relevant to a query or it is not".

4. Recall: [6]Chu & Rosenthal's Web search engine study omitted recall as an evaluation measure because they consider it "impossible to assume how many relevant items there are for a particular query in the huge and ever changing Web systems". in this study too we have omitted the recall as an evaluation measure for the similar reasons.

The computation of precision has been done as follows: Suppose an IR system returns 8 relevant documents and 10 non-relevant documents. There are a total of 20 relevant documents in the collection.

tp (true positive) = 8

fp (false positive) = 10

fn (false negative) = 20-8=12

Precision = tp/ (tp+fp)

= 8/ (8+10)

= 8/18

=0.44

Average Precision = sum of all precision/ No. of queries

Mean Average Precision = av. precision/ No. of queries

IV. EVALUATION

To search for the desired documents, a web query can be formed in different ways. The keywords of the user query may or may not be in root word format. Further, a query terms may have more than one sense. However, the searchers may not be aware of the impacts of such intricacies of queries on the retrieval relevancy. In this section we will evaluate each of the above two issues of queries.

1) Root word of the keywords: In English prefixes and suffixes (collectively called affixes) are normally used (e.g. s, es, dis, ness, ing etc.) with morpheme (root word) and new words are constructed. These new words are called morphological variants of the stem. For ex.: increase +ing = increasing, or dis + able = disable. Or happy + ness = happiness.

While searching on the web the query terms given by the users may not be in root form. As there is no restriction/help about how to choose or select the query term, same query may be formed with different morphological variations of its terms. This may lead to variation of results and the relevancy of results by search engines. To analyze this, we took a real time test of Google search engine using a set of 20 web queries (as per the discussion in the previous section). These queries are listed in table 2, and to properly analyze the result each query has been formed twice – with root words and without root words

TABLE 2: TEST QUERY SET FOR ROOT WORD ANALYSIS

Query with root word	Query without root word
1.1 Civil Service exam	1.2 Civil service examination
2.1 Mercury level in bird	2.2 Mercury levels in birds
3.1 water waste in India	3.2 water wastage in India
4.1 Fund and grants institution	4.3 Funding and grants institution
5.1 beds sharing with children	5.2 beds sharing with children’s
6.1 mercury levels is increase	6.3 mercury levels is increasing
7.1 The temperature is decrease	7.3 The temperature is decreasing
8.1 Native language of India	8.2 Native languages of india
9.1 merit of democracy	9.2 merits of democracy
10.1 Use of computer	10.2 Uses of computer
11.1demerit of democracy	11.2 demerits of democracy
12.1 advantage of mobile phones	12.2 advantages of mobile phones
13.1 disadvantage of mobile phones	13.2 disadvantages of mobile phones
14.1 Imagine power	14.2 Imagination power
15.1 power of battery	15.2 power of batteries
16.1 liberty of information act forms	16.2 liberties of information act forms
17.1 Game is begin	17.2 Game is beginning
18.1 Choose right path	18.2 Choosing the right path
19.1 Problem is examine	19.2 Problem is examined
20.1 English query	20.2 English queries

We then performed Google test for each pair of query set (table 2) and precision values are computed as shown below in the Tables 3 & 4.

Table 3.Precision computation for queries with root words on google (using table 2)

Query	Doc. Retrieved	Precision @10
1.1	5,350,000	0.55
2.1	35,100,000	0.57
3.1	71,800,000	0.5
4.1	114,000,000	0.66
5.1	25,500,000	0.66
6.1	68,900,000	0.77
7.1	125,000,000	0.77
8.1	5,990,000	0.37
9.1	17,800,000	0.88
10.1	2,900,000,000	0.66
11.1	369,000	0.66
12.1	112,000,000	0.62
13.1	1,270,000	0.88
14.1	126,000,000	0.66
15.1	572,000,000	0.5
16.1	18,700,000	0.6
17.1	17,456,000	0.62
18.1	18,187,000	0.7
19.1	26,432,000	0.57
20.1	9,876,000	0.66
Mean Average Precision = 0.643		

TABLE 4. Precision computation for queries without root words on google (using table 2)

Query	Doc. Retrieved	Precision @10
1.2	7,920,000	0.55
2.2	26,000,000	0.55
3.2	162,000	0.44
4.2	94,300,000	0.44
5.2	27,200,000	0.57
6.3	68,400,000	0.62
7.3	26,300,000	0.44
8.2	2,780,000	0.77
9.2	7,870,000	0.55
10.2	572,000,000	0.37
11.2	194,000	0.77
12.2	10,200,000	0.44
13.2	1,980,000	0.62
14.2	112,000,000	0.55
15.2	556,000,000	0.55
16.2	15,600,000	0.5
17.2	12,768,000	0.55
18.2	13,145,000	0.6
19.2	23,564,000	0.44
20.2	7,956,000	0.57
Mean Average Precision = 0.5445		

From the Tables 3 & 4, it is clear that when queries are in root form, search engine generally indexes more documents (comparing columns II of tables 3 & 4) i.e. the documents Retrieved are higher. The mean average precision for the root word queries is also higher. It shows that the root word queries are better understood by the search engines. .

***2) Sense Ambiguity (Ambiguous keywords)**:* Many words are polysemous in nature that is they have multiple possible meaning and senses. Finding the correct sense of the words in the given context is an intricate task. Various researchers (specially Eric Brill [8], [9]Argaw, [10]Navigili and Christopher Stokoe and John Tait[11]) have justified the role of Word Sense Disambiguation in the improvement of performance of web searching for English and other languages. Ambiguous keywords deflate the relevancy of the results.

We considered 20 queries (based on our discussion in para III) which are normally ambiguous in nature (a query has been considered ambiguous if one of the term of query is ambiguous). Further, in order to analyze the impact of ambiguity over search engine's performance we have tried to manually disambiguate each query with the help of WordNet Database and the search engine in consideration and have shown the effect of ambiguity on the performance of the search engines. This is shown in table 5 where the left side column has query with ambiguity and right side column has manually redesignated query without ambiguity same query

Table :5 : TEST QUERY SET FOR AMBIGUITY ANALYSIS

Query with ambiguous word (in bold)	Query with unambiguous words
1.1 Wall paint is blue	1.2 Wall color is blue
2.1 The train is standing on the platform	2.2 The train is standing on the railway platform
3.1 There are four seasons in a year	3.2 There are four cycle in a year
4.1 critical case	4.2 critical situation
5.1 A bug terminates a program	5.2 A error terminates a program
6.1 Python are found mostly in rainy season	6.2 Python snakes are found mostly in rainy season
7.1 Draw the figure of a flower	7.2 Draw the diagram of a flower
8.1 Close the door	8.2 Shut the door
9.1 There should be a break between two lectures	9.2 There should be a gap between two lectures
10.1 The river is dry	10.2 The river is empty
11.1 Score of team India in World cup	11.2 Run of team India in World cup
12.1 balance in my phone	12.2 money in my phone
13.1 live in present	13.2 live in today
14.1 aim of a doctor	14.2 duty of a doctor
15.1 the pitch of sound is high	15.2 the level of sound is high
16.1 Use of cosine function	16.2 Use of cosine expression
17.1 The chair of conference	17.2 The head of conference
18.1 Exercise is necessary to keep our body fit	18.2 Physical Exercise is necessary to keep our body fit
19.1 interest in science	19.2 favorite is science
20.1 major accident	20.2 big accident

The above queries are tested on the Google search engine and the results are shown below in the Tables 6 & 7

Table 6. Precision computation for ambiguity using google(using table 5)

query	Doc. retrieved	Precision @10
1.1	140,000,000	0.44
2.1	31,600,000	0.66
3.1	2,860,000	0.37
4.1	175,000,000	0.55
5.1	2,550,000	0.5
6.1	1,020,000,000	0.55
7.1	18,400,000	0.66
8.1	435,000,000	0.33
9.1	2,210,000	0.75
10.1	662,000,000	0.37

11.1	4,420,000	0.22
12.1	325,000	0.44
13.1	12,600,000	0.62
14.1	9,260,000,000	0.44
15.1	16,200,000	0.5
16.1	338,000,000	0.55
17.1	174,000,000	0.66
18.1	335,000,000	0.55
19.1	45,100,000	0.44
20.1	683,000,000	0.75
Mean average precision = 0.5175		

Table 7. Precision computation for ambiguity using google(using table 5)

query	Doc. retrieved	Precision @10
1.2	374,000,000	0.33
2.2	187,000,000	0.77
3.2	3,150,000	0.44
4.2	374,000,000	0.33
5.2	95,000,000	0.44
6.2	363,000,000	0.55
7.2	66,000,000	0.37
8.2	78,998,000	0.75
9.2	123,000,000	0.44
10.2	112,342,000	0.87
11.2	145,000,000	0.75
12.2	786,000,000	0.66
13.2	111,498,000	0.77
14.2	15,700,000	0.66
15.2	27,200,000	0.57
16.2	68,400,000	0.62
17.2	26,300,000	0.44
18.2	2,780,000	0.77
19.2	572,000,000	0.5
20.2	18,700,000	0.6
Mean average precision = 0.5815		

After examining and comparing the precision values of each queries (Tables 6 & 7), we found that after manual disambiguation of the queries, the precision of 13 out of the 20 queries has improved. The mean average precision has also improved. This shows that the ambiguity in web query can result in poor relevancy of results. Sometimes ambiguity in queries produces adverse results.

IV. DISCUSSION

We have done an extensive analysis of the impact of morphology and ambiguity issues of web queries. The google, though has been capable of searching very efficiently still not very capable of understanding user's intension and the context of queries. A minor change in the query term (at least from the searchers point of view) may result in considerable change in precision. The comparison of mean average precision for each test shows that the relevancy of search engines has been improved.

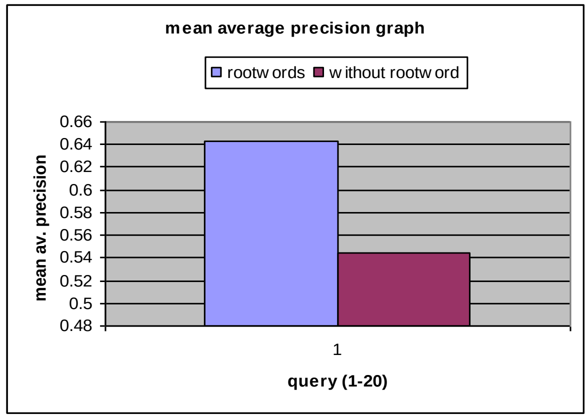

Fig 1. MAP for Root words vs. without root words queries

The fig.1 above shows the comparison between the mean average precision of the queries with root words and without root words. The significant increase in mean average precision for queries with root words shows that the search engines better understand root word queries over all other forms.

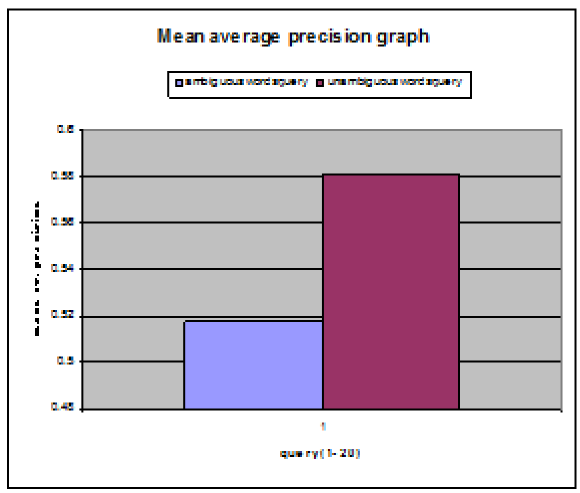

Fig 2. MAP for ambiguous vs. non ambiguous queries

In fig.2, the mean average precision of the queries with ambiguous keywords are found significantly low as compared with the ones without ambiguity.

V. CONCLUSION

The issues discussed in this paper towards query formation at the end user level are generally ignored by common web searchers. In fact, these factors can be very important in improving the performance of search engines. In this paper, we made an effort to highlight these factors for English queries. Our results show that the performances of the search engines are affected by these factors. The query term ambiguity may sometimes drastically reduce the relevancy of a search engine. The ambiguity detection and disambiguation of web queries are therefore essential which may require some short of human intervention.

REFERENCES

[1] Dwivedi SK, Rastogi P, Goutam R. Impact of language morphologies on Search Engines Performance for Hindi and English language, IJCSIS, Vol. 8, No.3, 2010J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Karlsson, F. Yleinen Kielitiede. [General linguistics]. Helsinki: Helsinki University Press, 1998. [In Finnish].

[3] Bybee, J.L. Morphology: a study of the relation between meaning and form. Amsterdam- Philadelphia: Jhon Benjamins Publishing Company, 1985.

[4] Matthews, P.H. Morphology. Cambridge University Press, 1991.

[5] Akmajian, A., Demers, R., Farmer, A. and Harnish R. Linguistics: an introduction to language and communication. Cambridge, MA: The MIT Press, 1990.

[6] Chu, H., & Rosenthal, M. (1996). Search engines for the World Wide Web: a comparative study and evaluation methodology. In Proceedings of the 59th annual meeting of the American Society for Information Science (pp. 127-135). Medford, NJ: Information Today.

[7] Voorhees, E.M., & Harman, D. (2001). Overview of TREC 2001. NIST Special Publication 500-250: The 10th text retrieval conference (TREC 2001) (pp. 1-15). Retrieved 17 December 2002 from http://trec.nist.gov/pubs/trec10/papers/overview_10.pdf.

[8] E. Bril and S. Vassilvitskii, "Using WebGraph Distance for Relevance Feedback in Web Search" in Proceedings of SIGIR'06, Seattle, Washington, USA, pp. -147-153, 2006.

[9] A. A. Argaw, "Amharic-English Information Retrieval with Pseudo Relevance Feedback", in Proceedings of 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, pp. 119-126, 2007.

[10] R. Navigili and P. Velardi, "Structural Semantic Interconnection: a knowledge-based approach to Word Sense Disambiguation", in Journal of Pattern Analysis and Machine Intelligence, Volume 27, Issue 7, pp. 1075 – 1086, July 2005.

[11] C. Stokoe and J. Tait, "Towards a Sense Based Document Representation for Internet Information Retrieval", in Proceedings of SIGIR'03, July 28- August 1, Toronto, Canada, pp. 791-795,2003.

[12] Gordon M, Pathak P, Finding information on World Wide Web: the retrieval effectiveness of search engines, Information Processing and Management 141-180, 35(1999).