Natural Language Processing-A vital shift

T.PrasannaPoornima[1] and Dr.K.Selvam[2]

1 Research Scholar, Department of Computer Applications, Dr. MGR Educational and Research Institute University, Chennai – 95, Tamil Nadu, India, 9840252299,
2 Professor, Department of Computer Applications, Dr. MGR Educational and Research Institute University, Chennai – 95, Tamil Nadu, India, 9840724613,

ABSTRACT

The aim of this paper is to hold forth to the usage of advanced search options available in search engines related to relevance, discussing the pros and cons and give grounds for shifting to NLP - IT's primary area of pursuit. Advanced search options aid the following: rules out unwanted results, extracts specific results and apparently helps in time conservation. Concurrently users are compelled to meet with exact solution to keywords, rearing and development of knowledge library, information matching and refining, and booming the intelligentization for searcher. As a deduction, it is undeniably evident to transit for NLP.
Keywords: Web search, NLP, advanced search options.

Corresponding Author: PrasannaPoornima.T

INTRODUCTION

Information professionals use advanced search options in a meticulous manner. But for a user who has less experience, its absolute necessity to have at least one of the following ideas: if It's a need to understand a topic, to find descriptive results, to do research in a specific area, for the timeliness of information, to know facts, for opinions and views, specific type of media or potentially new to the internet environment. With an effect on every part of scope, the multilingualism of web content continues to burst. For quick searchers, the content from distant countries and multiple languages appear suddenly and create new space for excavating once buried information resources. To excavate such resources in languages that the user is not familiar with, the search engines come out with a variety of language aids. Searchers have a variety of progressively skeptical tools to retrieve information content from many languages.

To consider the classification, search engines fall under one of the following categories:
Language of the interface
Translated search
Language limit

The interface language is the first and most frequently talked about. Importance of it depends on the individual search. A listing of more than hundred languages is seen on clicking Google's language tools. These are the interface language choices listed by Google as "Use the Google Interface in Your Language." The language options shown just changes Google's homepage text and the text on results pages, help files, button text, and other words used by Google within the interface itself. The effect on searching and results is not up to the mark. Searching the same query with the English gets the same results. Language limits of a particular search engine can be used for searching pages published primarily in a particular

language or languages. The language limits are normally masked on the advanced search pages. The machine translation tools are often buried even further on completely separate pages which serve to fetch unfinished and inaccurate translations between languages.

Boolean search strategies do not locate materials. Boolean vs. relevance ranked queries retrieve only 50 percent of the documents in a collection. Ranking based on relevance uses algorithms in statistics to compare all documents in a database.

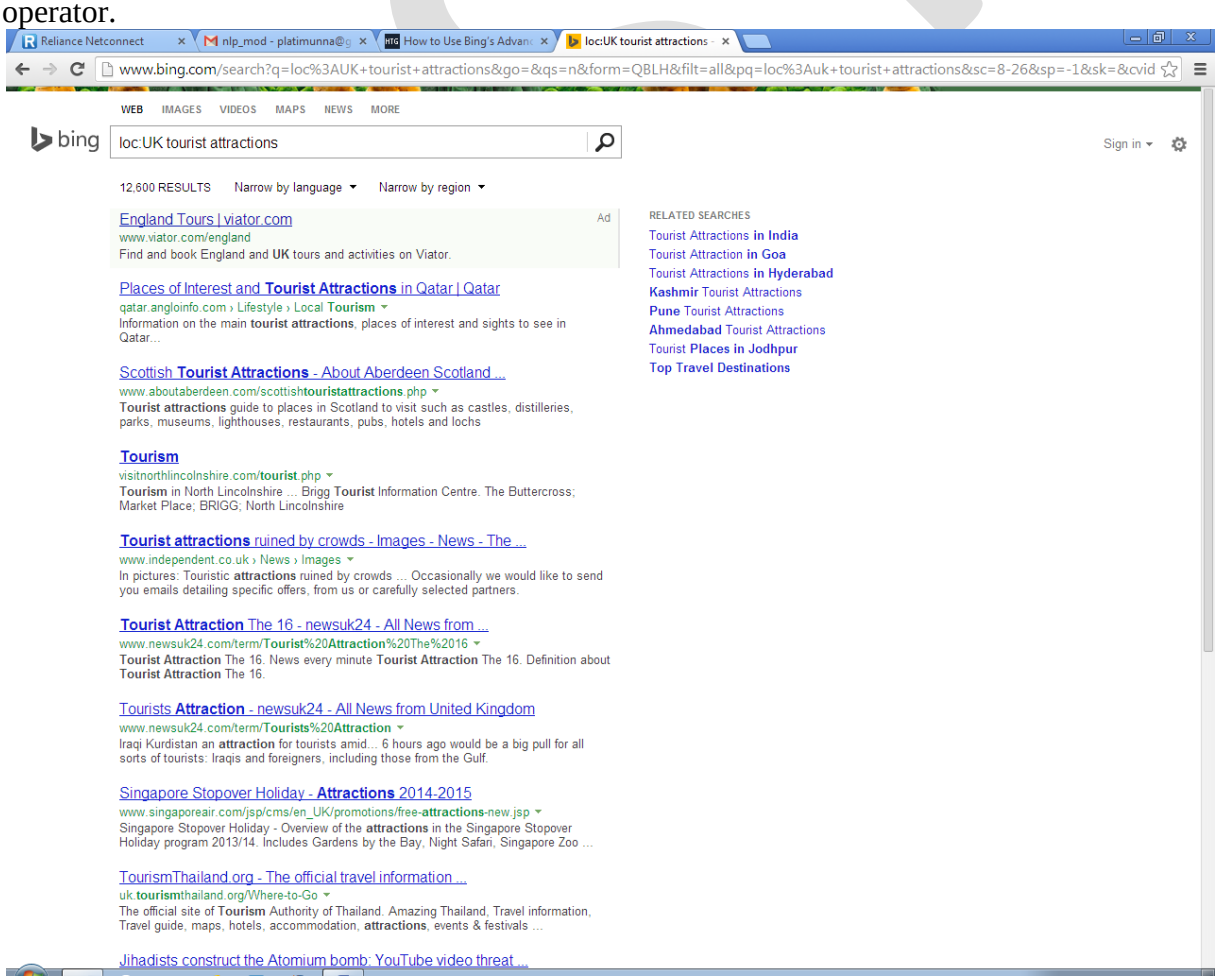
Basically, the number of times a word in a query occurs is counted in each document. And the documents with the highest number of query words, and also the greatest frequency of occurrence of each word, are rated as the most relevant. Such statistically based systems do not recognize the context or meaning and use statistical means for prediction of the relevance of a particular document. Boolean software is based on the principle of "either/or" basis. They search for exact matches. Either the word, one search for is available or not. Upon forgetting to truncate or adding a likely synonym, it's likely that we miss the documents that might have been retrieved along with them.

Recently the search engine Bing has also arrived with many search facilities equipped to cater the needs of the searcher, but that needs highly trained professional searchers to receive the desired output.

Some of the operators available at Bing are discussed that seems to be promising but depends on user's technique adaptability, learning capacity and individual's time tolerance when requisite results are difficult to fetch.

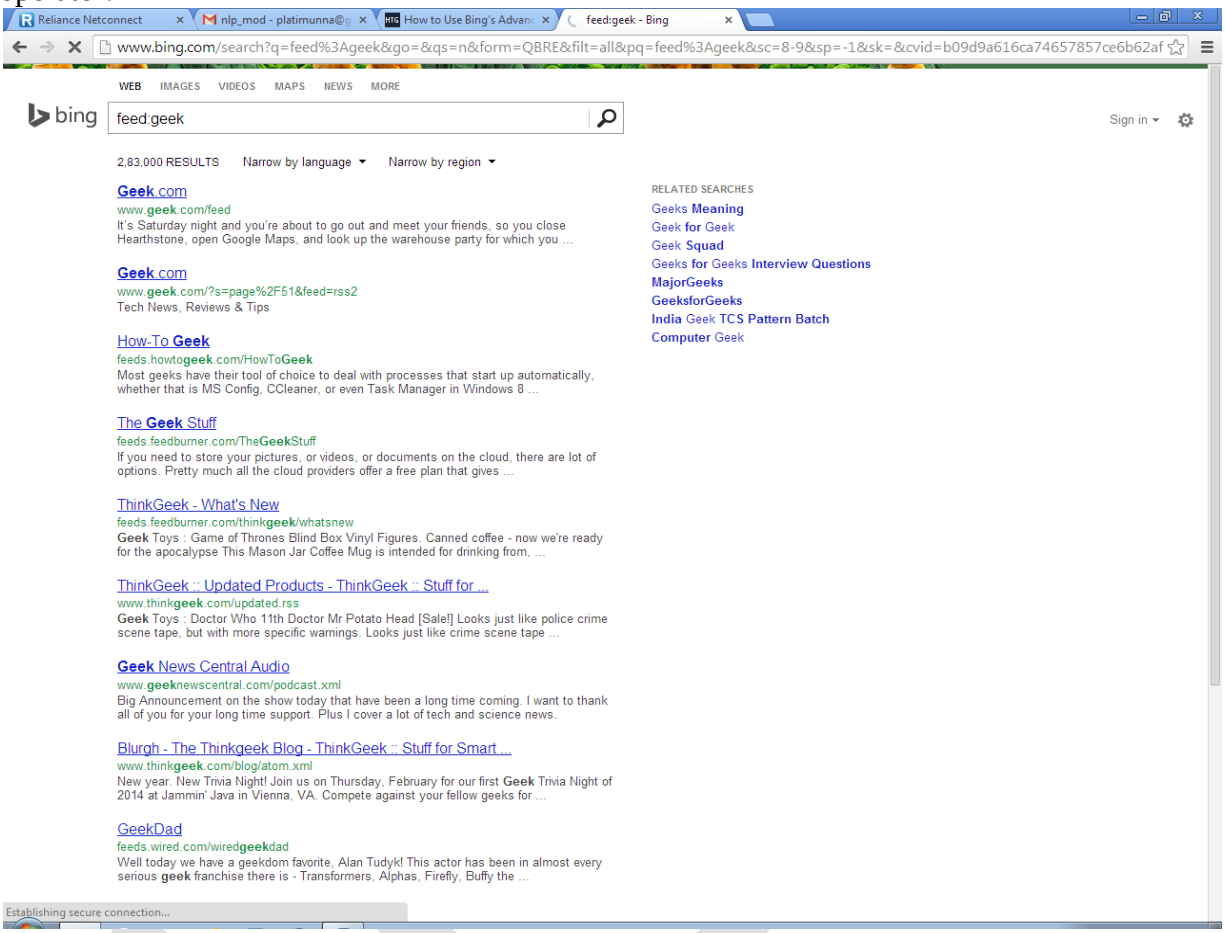
To substantiate the same, two screenshots are presented from the search engine Bing viz. loc and feed

The following screenshot is taken from the search engine "Bing" and illustrates the 'loc' operator.



Use the loc: operator to specify a particular location. For example, the above query returns tourist attractions in the United Kingdom.

The following screenshot is taken from the search engine "Bing" and illustrates the 'feed' operator.



Use the feed: operator to find web feeds that contain a word. Its purpose is to find blogs about a topic. For example, find feeds that contain the word "geek" with the following query:

feed: geek

Connecting search operators, joining together complex queries out of several different operators can be yielding the undoubted power for a tech savvy but for a naïve user these illustrations become a mystery. So the best way to leap all these hurdles is to switch over to NLP in search engines. Though there exists many such NLPs performing specific tasks, still it has to be utilized to the fullest such that users from different levels of knowledge could use information retrieval without any hassle. Let's have a quick brief of what NLP does in this aspect of catering the needs of the user.

Natural Language Processing (NLP)

Natural language (NL) processing has procured noticeable progress in searching and has become stereotyped in search tools where human interaction is required.

A great demand has been shown for natural queries not only in the field of database searching but also in Internet search engines. There must be less restrictions on the syntax with which users could articulate their natural queries. NL systems are yet to achieve the accuracy and reliability expected of them and is an active area of study. Several natural language interfaces have been created by specialists in computational linguistics to achieve the expected accuracy and reliability. The search to be "truly natural" refers to different types of linguistic articulation. A native speaker must be able to understand and use sensibly. Natural

articulation cannot be achieved through fixed design of templates and powerful-words dictionary. There are numerous NLP s tailored to the user's needs. Some of them are Thinknowlogy, Bitextor, TagAligner, RelEx, AlchemyAPI Android SDK, AlchemyAPI Java SDK, and Wintermute.

A neat NLP must be able to either comprehend the natural articulation or to enlist and assess all feasible interpretations. As experience aids to improve interpretation (either understanding or enumeration-evaluation), an NLP is expected to learn from its performance, including both successes and failures. Learning involves numerous tasks all through the query process. The user's articulation might be responded intelligently by the NLP and preferably solicit more information about the query. By confirming the meanings and results with the user, NLP could assist evaluation and bag correctness. One more possible case is that NLP could retain usage patterns so that performance would be improved.
The recent invention, Google glass though failed in the adoption test still indicates the power of search in the future. It's always a wondering aspect of questioning a computer in conversational language and answered relevantly when it's been years struggling for a query formulation in machine format to query a database. It also becomes evident for data owners to organize data logically and conform to relevant data conventions such that it's accessed by all possible spiders.
Normal search engines take up a content-first approach: When a collection of documents is provided, how would the content be understood stands a big mystery. The language of the content expressed does not match with the expressed user's need.
An algorithm to understand concepts and establish relationships between concepts and answer complex questions is given focus. Thus to quickly parse full questions instead of word-by word and rank answers from the content available that is indexed is given significance. Other than NLP, Google released Caffeine that tried to improve indexing speed and users were catered with novel results.

APPROACH FOR THE PROBLEM

The query "who was the inventor of television" helps the machine to have an explicit fact: the inventor of the television. There exists no uncertainty with derivation of place, time, person concerned with this event.
But think of another query like "I want to write on the whiteboard", states that the requirement is a marker, easily inferred by humans. As opposed to a machine, it is difficult to be resolved as there is no evidence for the need of a marker.
A possible solution would be to put user-first and not content-first approach. Assess the inputs of the user and codify them, then represent an understanding of intents and needs. Then a subject-matter expert (SME) or a content analyst would provide the appropriate result. Though it seems painstaking still, only one user has to go via the traditional search, thereby recording the steps. Then codify the same and it is repeatable. Whenever another user faces the same situation, returning the already codified result becomes consistently repeated and in no time.
The termination of popular services such as Google Reader, iGoogle, and the + symbol for exact searching substantiate that emerging NLP techniques are replacing the former services and are an active area of study in the near future.

Recently face book has implemented its new technique called graph search that has leveraged natural language interface. Though graph search seems impractical for normal search engines, a brief note on the underlying logic would be appropriate.

A keyword-based system would not be the best choice because of the fact that keywords, which usually consist of nouns or proper nouns, can be vague in their intent. For example, "friends Face book" can mean "friends on Face book," "friends who work at Face book Inc.," or "friends who like Face book the Page." Keywords, in general, are good for matching objects in the graph but not for matching connections between the objects.

Face book has built a system tolerant of grammatically incorrect user queries and picking up traditional key word searches. Moreover, the query suggestions are always constructed in natural language, thus allowing user to have a prior knowledge whether the system has correctly understood user's intent before selecting any suggestion and executing a search. It also suggests options options for completing your search as you type in to the type ahead, demonstrating what kinds of queries it can understand. Also for queries completely out of the scope of grammar, a mechanism to detect them is built and falls back to entity suggestions or web search suggestions.

Thus a robust NLP to recognize traditional keyword search, identify grammatically incorrect queries, providing auto-complete options as a user enters search terms or phrases must be designed to perform exact searches across the abundance of data. And such natural language interface identifying common synonyms and related forms by culling extensive lists of words that could be used interchangeably needs to be built.

What a search engine refers to:

Search engines like Google, Ask, and Bing serve as an online tool for searching web sites. These search for any site, any subject, and any type of file. Search engines for specific purpose will search one particular subject or one particular type of file. Web sites are grouped by search engines with the help of software programs called "spiders" that crawl the web and search for web sites. Web site creators can also register their web site with a specific search engine to confirm that their web site will be enclosed by that search engine.

What a subject directory refers to:

While subject directories like Yahoo directory refers to a search engine that clusters its web sites by subject areas. Only a limited number of web sites are searched. Web sites in a subject directory are enlisted by people and those web sites are selected based on quality, relevance, or some other criteria. This proves that subject directories do not turn out many hits for a search as search engines do.

What a Meta search engine refers to:

On the other hand Meta search engines like Dog Pile, Search.com searches multiple search engines and subject directories at the same time. Type in one search and it will give results for a number of search engines and subject directories.

CONCLUSION

It is clearly proven that for an unschooled user, these different categories of search paths tend to distract the search process. Henceforth a logical shift to a better approach for easy access of exact result set becomes mandatory. Thus NLP comes along with a wide variety of customized solutions tailored to the needs of the searcher.

REFERENCES

[1] B. Jansen and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs". Info. Processing & Management, Vol. 42, Issue 1, 2006

[2] Jurafsky, Daniel, and James H. Martin. 2009. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Upper Saddle River, NJ: Pearson Education, 2nd edition

[3] Bilal, D. (2012). Ranking, Relevance Judgment, and Precision of Information Retrieval on Children's Queries: Evaluation of Google, Yahoo!, Bing, Yahoo! Kids, and Ask Kids.

Journal of the American Society for Information Science and Technology, 63(9), 1879–1896.
[4] Google Search Quality Rating Guidelines. (2012). http://static.googleusercontent.com/media/www.google.com/de//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf
[5] List-Handley. C. J. (2008) Information literacy and technology. 4th ed. Dubuque, Iowa: Kendall/Hunt
[6] Barker, D. I., Barker, M. S., & Pinard, K. T. (2010). Internet research: Illustrated. Boston, MA: Course Technology.
[7] D. Hall and D. Klein. Finding Cognate Groups Using Phylogenies. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 1030–1039, Uppsala, Sweden, July 2010. Association for Computational Linguistics.