ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research, volume 205
The 2nd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2018)

Analysis of Semantic Retrieval System Based on Information Resource Management

A Comparative Analysis of Wolfram Alpha and Google

Bin Ye
School of Economics
Zhejiang University
Hangzhou, China

***Abstract*—Wolfram alpha is a computational knowledge search engine. This paper tries to evaluate its search results from the point of view of users and compare it with Google to evaluate the advantages and disadvantages of the search engine.**

Keywords—Wolfram alpha; Google; semantic retrieval

I. INTRODUCTION

In the era of knowledge economy, the retrieval and utilization of vast amounts of information is an important research and application field currently. At present, the practical information management system is mainly based on classified catalogue or keyword matching. The former is not efficient for massive information processing, and the depth is limited. The latter has certain limitations in the semantics of information and semantic disclosure. Solving the problems such as information organization, knowledge representation, machine understanding and human-machine interaction is very important and urgent for improving the efficiency of information utilization. In recent years, intelligent semantic retrieval provides an opportunity to solve the above problems, and provides an effective way to information management. Previously, portals and search engines have been known as the two core technologies of the Internet. With the development of technology, is it possible that the third core technology emerge in the Internet? Tim Berners Lee, the father of the internet, believes it is possible. He once said publicly that the "future web" will make it easy for anyone to link to any even bits of information, such as a photograph or a bank report, the so-called semantic web. He formally proposed the semantic web architecture at the International XML Conference in 2000. The architecture embodies the basic steps of Semantic Web development, each of which will generate added value, thereby implementing the Semantic Web with incremental method. The steps of research and development should be as follows:

- Providing common syntax for machine-understandable declarations;
- Establishing a common glossary;
- Reaching agreement with logical language;
- Exchanging evidence in this language. [1]

The aim of Semantic Web is to make the information on the Web have machine-understandable semantics. That is to say, it can make machine be accessed and explain the content of the World Wide Web, in order to satisfy the effective access and search of distributed heterogeneous information on the World Wide Web by intelligent software agents. At present, semantic search has become a hot development direction of search engines. For example, hakia, powerset, Freebase, Wolfram alpha are all well-known semantic search engines, the author hopes to evaluate the Wolfram alpha and compare the evaluation results of Google, trying to analyze the advantages and disadvantages of this semantic search engine.

II. INTRODUCTION TO WOLFRAM ALPHA

Wolfram Alpha is a computational knowledge search engine. It can not only produce the calculation results, but also provide the users with a variety of chart analysis of data information.

Wolfram, the developer, says the working principle of Wolfram Alpha is different from Google Web search. In the Wolfram Alpha service, the search engine returns answers directly to the user rather than a bunch of web links after the user types the question to be queried in the search box. [2]

Nova Spivack, a search engine researcher in the US, believes Wolfram Alpha uses patented algorithms to build a huge database of information. As a result, when a user types a query question in Wolfram Alpha, the search engine returns the corresponding detailed answer directly. Wolfram Alpha's innovation is that it delivers what you're looking for right away, rather than just showing you what might be relevant, like Google. As for the search processing mechanism Wolfram Alpha, in addition to its ability of information processing and extraction, it is also outstanding in seeking items that meet the requirements in the knowledge base through the inference mechanism, as well as its unique powerful computing power. Through this new type of search engine, people will no longer get the existing online ready-made information, but the information generated through server computing. By introducing powerful "computing"

capabilities, Wolfram Alpha can help users realize the statistical work based on keywords or problems they input. Wolfram Alpha will also open application programming interfaces (API) to external developers. In other words, external developers will be able to build more targeted search services based on Wolfram Alpha.

III. EVALUATION METHOD

Summarizing the evaluation index proposed by domestic scholars, the author evaluate the main search engine from the following aspects: (1) Index database performance, including the index quantity of specified site, update frequency, indexing range; (2) retrieval function, including primary retrieval, advanced retrieval and characteristic vertical searching products; (3) retrieval effect, including recall ratio, precision ratio and time lag; (4) result display, referring to the statistics of the retrieval results and the display format of the retrieval results; (5) User burden, including personalized settings, help function, information reorganization, language translation, learning function and memory function. [3]

In the process of evaluation, the indexes that need to be calculated are the index quantity of specified sites, recall ratio, precision ratio and the irrelevant link rate (heavy chain rate, dead chain rate and false chain rate) involved in the calculation of precision rate. [4] The retrieval results of Wolfram alpha are the knowledge results through possibilistic computation, and there is no relevant data such as index quantity, and the performance of database and the retrieval effect are basically to meet the information needs of users. Therefore, a number of test topics can be selected to conduct qualitative analysis and evaluation from the satisfaction degree of users' information needs. [5]

IV. SEARCH THEME

In order to reflect the users' search needs objectively, and get close to the actual search content of most users as much as possible, the author selected the top five English keywords in the popular trend of American region in the Google hot list at 16:00 on June 29 as the test subject, which are billy mays dead, don Cornelius, tevin Campbell, ojays, willie mays respectively. [6] In addition, the key words with strong knowledge of each subject are selected as the test subject, which are gdp Chinese/india, x ^ 6-x ^ 5 + 20 = 0, stopping distance 70 mph respectively.

V. TEST RESULTS

The search result of billy mays dead: the top five results of Google are three news articles, one wiki encyclopedia, and one YouTube video, as shown in "Fig. 1"; Wolfram alpha gets the identity and birth time attributes of the keyword by separating billy mays and dead, as shown in "Fig. 2".

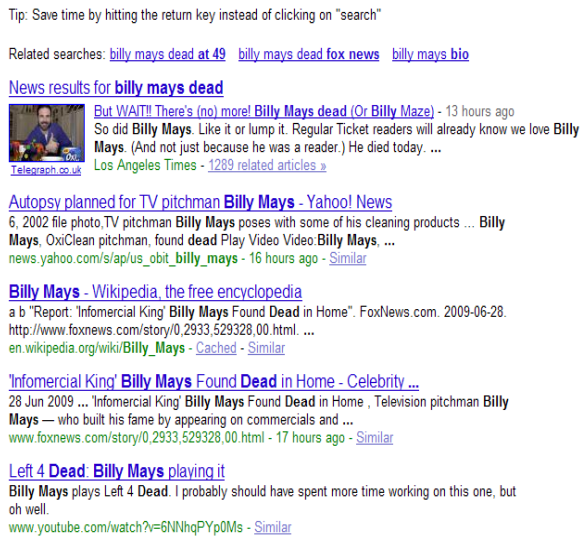

Fig. 1. Search result of billy mays dead in Google.

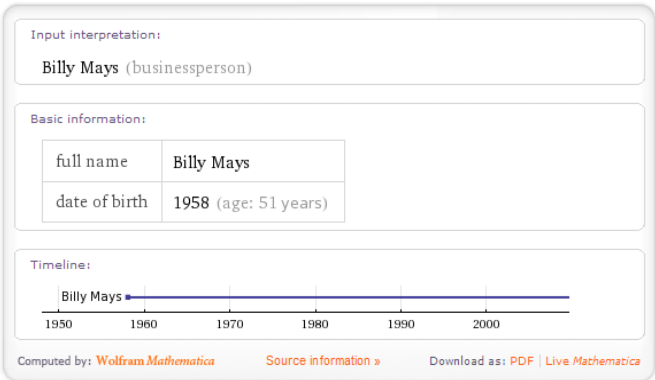

Fig. 2. Search result of billy mays dead in Wolfram alpha.

The search results of Don Cornelius: the top five results of Google include two YouTube videos, one wiki encyclopedia, one picture and one news article, as shown in "Fig. 3"; Wolfram alpha has only statistics and graphs of the name of don and cornelius, as shown in "Fig. 4". The results of this test can reflect the shortage of index quantity of the Wolfram alpha.

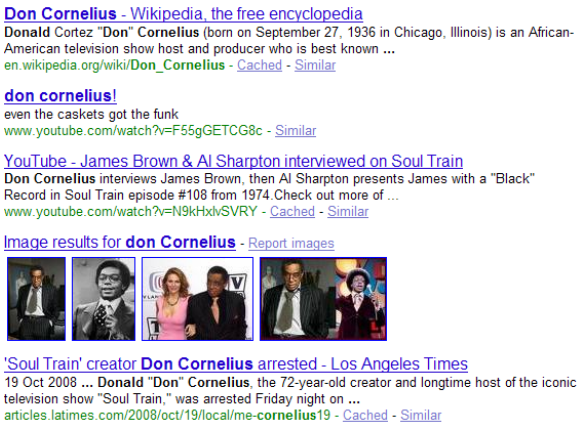

Fig. 3. The search results of Don Cornelius in Google.

Fig. 4. The search results of Don Cornelius in Wolfram alpha.

The search results of tevin Campbell: the top five results of Google include two YouTube videos, one wiki encyclopedia, one picture, and one news article, as shown in "Fig. 5"; Wolfram alpha briefly shows the key attributes of the keyword, such as the identity, date of birth, and place of birth and so on.

Fig. 5. The search results of tevin Campbell in Google.

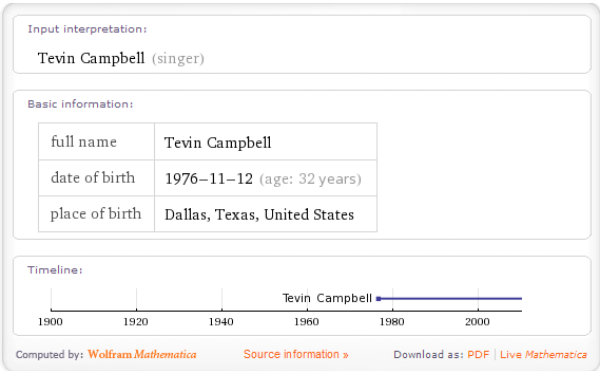

Fig. 6. The search results of tevin Campbell in Wolfram alpha.

The search results of Ojays: the top five results of Google include one wiki encyclopedia, one home page, one Amazon sale and two YouTube videos, as shown in "Fig. 7"; Wolfram alpha returns no matching results, and it recommends jays search, as shown in "Fig. 8", which is clearly not according to the original search keyword. This result also shows the lack of index quantity of Wolfram alpha.

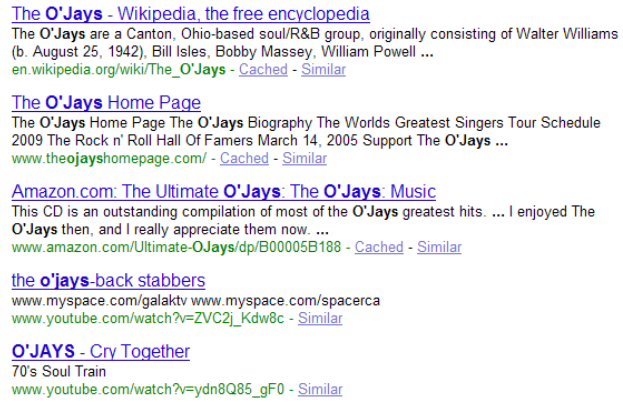

Fig. 7. The search results of Ojays in Google.

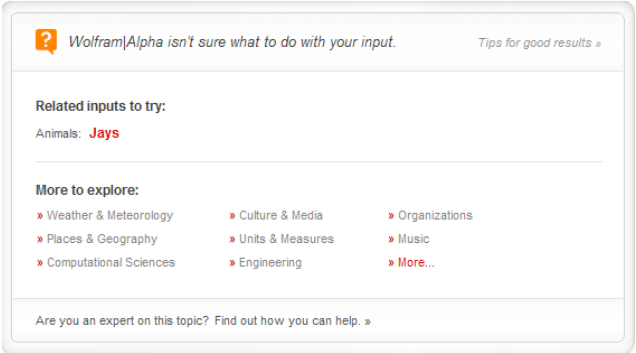

Fig. 8. The search results of Ojays in Wolfram alpha.

The search results of willie mays: The top five results in Google include a wiki encyclopedia, a picture, a YouTube video and two presentations, as shown in "Fig. 9"; Wolfram alpha shows the main attributes of the keyword, such as the identity, date of birth, and place of birth, as shown in "Fig. 10".

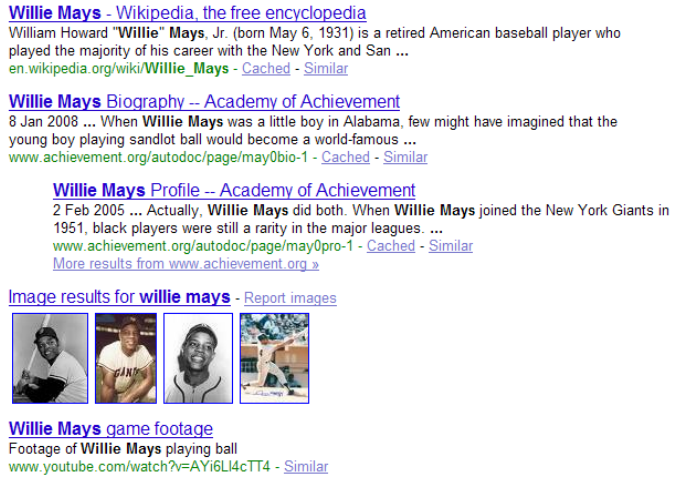

Fig. 9. The search results of willie mays in Google.

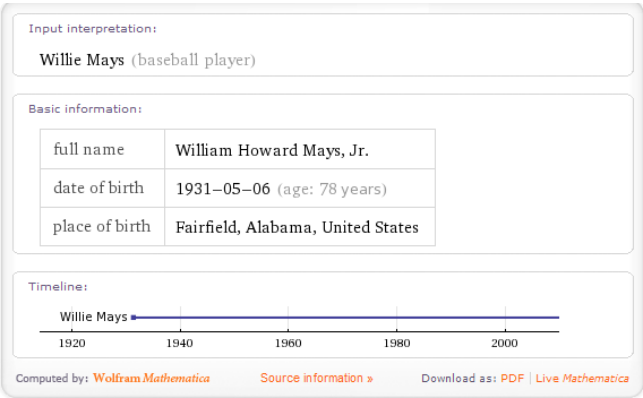

Fig. 10. The search results of willie mays in Wolfram alpha.

According to the results of top five keyword search test in Google, the results of Google are obviously more in line with users' needs. Of course, the test keywords themselves are more beneficial to Google. The top five search results in Google are fairly regular, generally including wiki encyclopedia, YouTube video, real-time authoritative news and related images of important websites. It should be said to consider the needs of users from all angles. In contrast, Wolfram alpha also presents its own features. That is displaying the important attributes of keywords and presenting them in a clear and structured form, generally including simple images. Of course, the information is obviously not as good as the results in Google, and the important attributes of its results can be easily obtained in Google. In addition, Wolfram alpha also expose the problem of insufficient indexes. However, the author believes that although the results of Wolfram are not satisfactory, the idea of organizing scattered information into structured knowledge is in line with the development trend of search engines, and the search of keywords with less knowledge cannot represent the real value of Wolfram alpha.

As for the search results of gdp china/india, the top five results in Google show the websites about GDP in China and India and users can get answers through reading and understanding by themselves, while Wolfram alpha gives the results succinctly and directly, as shown in "Fig. 11" and "Fig. 12".

Fig. 11. The search results of gdp china/india in Google.

Fig. 12. The search results of gdp china/india in Wolfram alpha.

For search results of x ^ 6-x ^ 5 + 20 = 0, Google gives a number of sites that provide mathematical calculations, but none of these sites offer the ideal answer; Wolfram gets the answer directly as shown in "Fig. 13" and "Fig. 14".

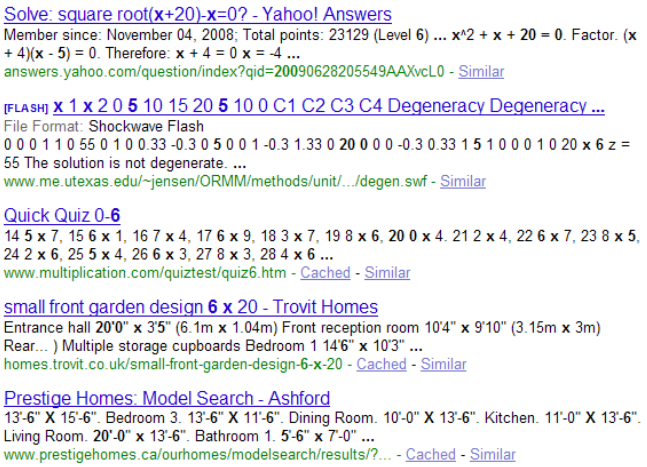

Fig. 13. The search results of x ^ 6-x ^ 5 + 20 = 0 in Google.

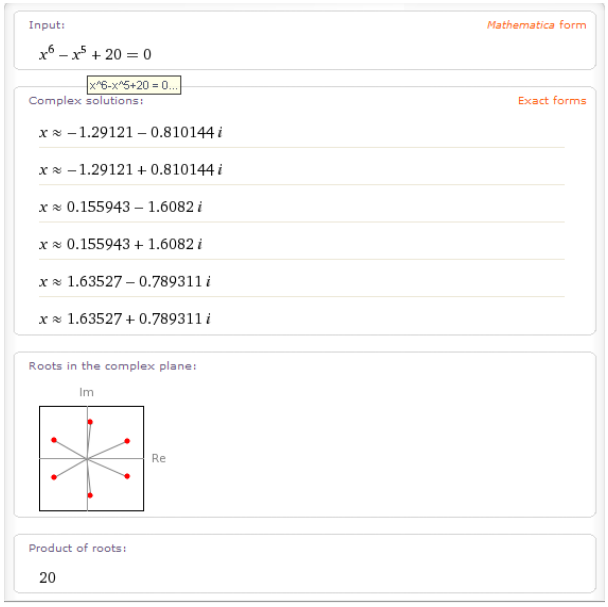

Fig. 14. The search results of x ^ 6-x ^ 5 + 20 = 0 in Wolfram alpha.

About search results of stopping distance 70 mph, Google searches out two texts on parking distance, two xenus, a Yahoo question and answer. Users can get the answer by reading and understanding the text, while Wolfram alpha directly get the answer, as shown in "Fig. 15" and "Fig. 16".

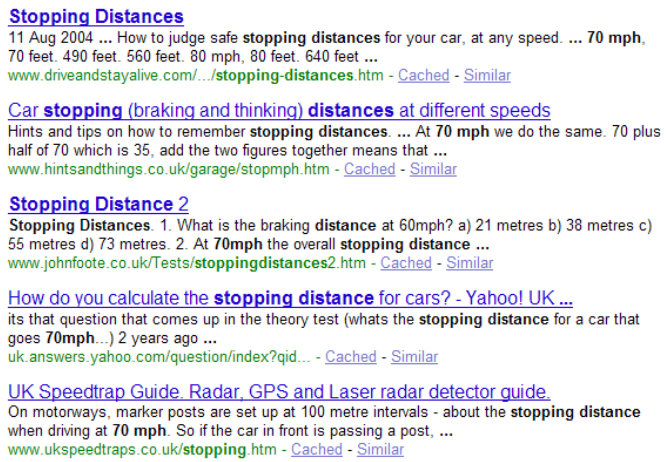

Fig. 15. The search results of stopping distance 70 mph in Google.

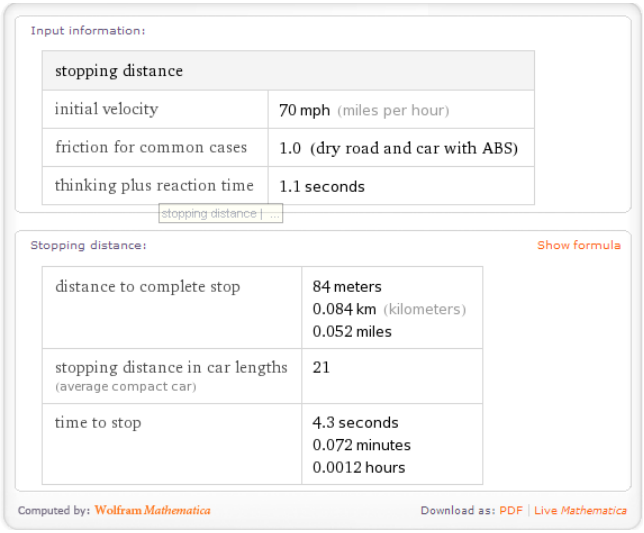

Fig. 16. The search results of stopping distance 70 mph in Wolfram alpha.

According to the results of the knowledge keywords, Wolfram alpha is more suitable for the information needs of users, which is mainly owing to its powerful semantic computing background and knowledge understanding system, including 29 major categories of disciplines to provide the majority of commonly used knowledge reasoning, such as: "Fig. 17". Google, by contrast, can only use wiki, Yahoo Q & A sites to meet the user's need for in-depth knowledge search, which is far from the complete system of Wolfram alpha.

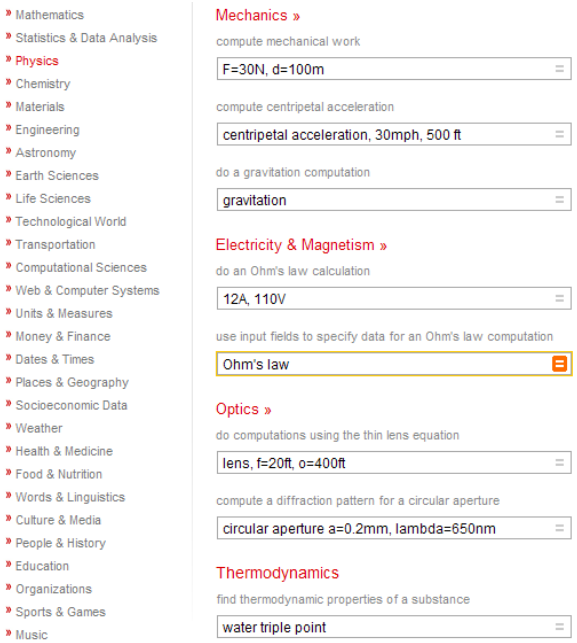

Fig. 17. The powerful semantic computing background and knowledge understanding system of Wolfram alpha.

VI. CONCLUSION

In this way, Wolfram Alpha has taken a big step towards the ultimate goal of the Semantic Web — making information on the Web machine-understandable. Google

can only provide web links based on page rank, while Wolfram Alpha gives you effective answers to questions. In short, it elevates the information search to the level of knowledge search. Of course, the results of the test show that Wolfram alpha still has a number of fatal defects that make it unable to compete with Google. The first is the lack of index since it did not grasp the time-sensitive pages and index in time; Secondly, for the popular keywords that are not knowledgeable, it needs to return the results in more detail. If Wolfram alpha is to become the next generation of search engine, it can not only be based on the search of key words which require high threshold of professional knowledge, but also organize the popular information, excavate the user's preference demand for popular information, and use its powerful computing background to provide the content after arranging the information of the web page. Thirdly, it is necessary to consolidate the complete subject system, further develop the inference model of new knowledge with the experts of each subject, count the user's preference demand, deal well with the relationship between subjects, and provide the inference knowledge consistent with the demand. Imagine the next generation of search engines: the barrier between the natural language that humans use on a daily basis and the language of machines is almost invisible, and search engines can understand and return the natural language.

REFERENCES

[1] Feng Zhiyong, Li Wenjie, Li Xiaohong, Ontology Engineering and Its Application, p. 52, Beijing: Tsinghua University Press, 2007.

[2] Wolfram alpha：http://www27.wolframalpha.com/.

[3] Zhang Junhua, Han Quanhui, main performance evaluation of five Chinese integrated search engines. Information Science, vol. 26, no. 9: 1353-1356, 2008, 9.

[4] Yue Zhen, Evaluation of Retrieval Performance of Four Chinese Search Engines, Intelligence Science, vol. 23, No. 6: 921-927, 2005, 6.

[5] Powerset vs. Google: The Completely Premature Head-to-Head：http://www.readwriteweb.com/archives/powerset_vs_google.php.

[6] Hot trends of the United States: http://www.google.cn/rebang/detail?bid=12100006 & utm_source = zh_CN-cn-et-abt & utm_medium=et.