

A Visual Exploratory Meta Data Search Engine Solution based on Cloud Computing

Mohammed Najah Mahdi

M.Sc.

College of Graduate Studies

College of Information Technology

University Tenaga National Malaysia

Abdul Rahim Ahmad

Associate Professor

College of Graduate Studies

College of Information Technology

University Tenaga National Malaysia

Roslan Ismail

Associate Professor

College of Graduate Studies

College of Information Technology

University Tenaga National Malaysia

ABSTRACT

With the rapid upsurge of web use and internet connectivity (Access), the traditional search engines face significant challenges in the contemporary searches, response speed and the storage of bulk pages. The search engine contained in the cloud can help alleviate the identified shortcomings. This is simply because cloud computing is advantageous in mass data processing and storage. Through an analysis of a visual exploratory Meta data search engine on a cloud platform, there will be strategies adopted to optimize the search engine in order to improve the overall performance and effectiveness of search.

Keywords

Visual Exploratory, META DATA Search Engine, Cloud Computing, Information Retrieval.

1. INTRODUCTION

Cloud based storage enables massive online storage of data. In the cloud storage, a full text search engine is critical for finding documents. A visual exploratory Meta data search engine suits the search of information in a cloud. In cloud computing, elasticity and reliability are of great importance. A visual exploratory search engine solution based on cloud computing environment has a major focus on the improvement of the traditional search engines through the provision of a fresh strategy for users to achieve reliable, personalized and graphical representation of the search items [3]. This strategy is achieved through; collection of raw information by a new meta-search engine in order to rapidly access large proportions of raw information, indexing and analysis of raw information through automatic similarity calculation and creation of semantic links on the basis of cloud computing environments and finally user interfaces through creation of an advancement of traditional result use list in the representation and this is replaced by graphic display of the search results [8].

The traditional search engines which created a human-computer interaction has been very successful in the presentation of search results to the users in ordered list in the order of relevance to the request as it is the case with the commercial search engines like Google, Yahoo and Bing. However in the recent times visual exploratory Meta data search has been accorded significant research interest. This is for the simple reason that the strategy enables the users to conduct exploratory, adventurous and discovery in information -vested world [21].

The growing trend characterized by more active user engagement in the search is mainly steered by the increase in web data and knowledge, the technical urge for selection, navigation and the trial and error strategies in internet use. The bottom-line of this is that there is intrinsic need by the users for information and online learning. The visual exploratory Meta data search has been empowered by the methodological advances in the manner in which information is retrieved, the interaction between the computers and the humans, information visualization and the information engineering and storage [9].

Visual exploratory meta data search engines on cloud computing platforms have developed a new paradigm creating the necessity for the development of a wide range of mobile applications. However in order to use this strategy with maximum effectively, there has to be a broad array of publicly accessible data to be established and maintained. Cloud-based data storage is an optimum solution in deploying the visual exploratory Meta data search engine [16].

2. PROBLEM STATEMENT

In the recent times, research has mainly focused on the vertical search engines and has delivered the expected results. However, much of the research has only concentrated on the application segment of the search engines meaning that the areas of storage and accessibility have been broadly overlooked. With the advancement of the internet technologies together with the growth of 3G and 4G wireless networks, the online users of the web and the bulkiness of the pages has tremendously grown with time. The ordinary traditional model of the search engine cannot cope with the advancements in the network technologies and consequently the search engine continues to face the challenge of how bulk data in the networks can be stored and accessed or processed faster [15]. The cloud computing platform provides new avenues to develop amicable and feasible solutions to the problems due to its capability to handle bulk data storage and processing. Embedding the visual exploratory search engine on the basis of cloud computing is indeed an ideal solution [13].

3. RESEARCH OBJECTIVE

The objective of this research is to find out the manner in which a visual exploratory Meta data search engine solution based on cloud computing is effective compared to the classic open-source browsers. It seeks to investigate how the aspects of speed processing and mass data storage are achieved through this novel strategy. At the end of the research, the researcher should be in a position to recommend or dismiss

the visual exploratory Meta data search engine solution based on cloud computing to contemporary businesses.

4. LITERATURE REVIEW

A search engine is the most effectual instrument for use to browse the World Wide Web. It has gained massive popularity in application in internet exploration. Without a search engine, it would be difficult to access the information on the internet since it is impossible to search for single websites for information. It is based on particular strategies which apply particular computer programs in collection of information on the internet, organization of the information and retrieval of the information by users [4]. Search engines comprise of five parts namely: fetcher, parser, indexer, retriever and the user interface. The classical search engines operate in a more focused manner such that they cannot perform parallel operations efficiently and give the users timely searches [17].

Research on the distributed computing framework of the visual exploratory Meta data search engine architecture indicates that compatibility with cloud computing is essential for handling mass data. This is facilitated by its efficient distribution of the computing framework; MapReduce and uses a HDFS to distribute system files hence has a mass data capability. A visual exploratory Meta data search engine on a cloud platform will definitely alleviate the problem of mass data storage and processing. The engine will also be improved in real-time searches and the speed of responses [18].

A visual exploratory search engine solution based on cloud computing environment has a major focus on the improvement of the traditional search engines through the provision of a fresh strategy for users to achieve reliable, personalized and graphical representation of the search items. This strategy is achieved through; collection of raw information by a new meta-search engine in order to rapidly access large proportions of raw information, indexing and analysis of raw information through automatic similarity calculation and creation of semantic links on the basis of cloud computing environments and finally user interfaces through creation of an advancement of traditional result use list in the representation and this is replaced by graphic display of the search results [12].

Search engines play a major role in ecommerce. They enable customers to search the large amount of information on the internet so as to make informed purchasing decisions. It is therefore important for a firm with a virtual store to have an effective search engine solution that can enable its customers get the right information efficiently [11]. There are four main types of search engines in the market currently. The first type is the crawler-based search engines. This type of search engines is made up of a crawler, an index and search engine software. The crawler is responsible for visiting a given web page, reading the information on the web page then following the links to other pages within a given website [11]. In addition, the crawler is responsible for revisiting a web page regularly to look for changes. The index is responsible for storing information regarding the web pages found by the crawler. The search engine software is responsible for accepting user-entered query, interpreting the query and going through the records in the index to find matches before ranking the pages in its result page [7]. Examples of crawler based search engines include Google and Ask. Figure one below shows how a crawler based search engines work.

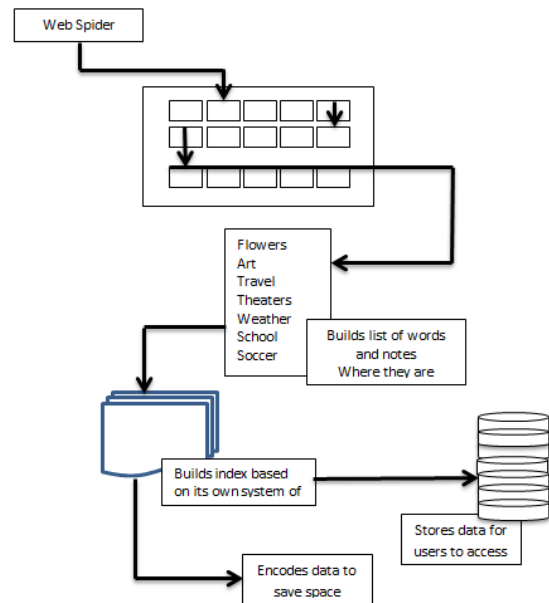


Fig 1: The workings of a crawler based search engine [7]

The first step in the working of a crawler based search engine is searching the internet and collecting web pages. The task is done by crawlers which are programs that automate repetitive tasks at very high speeds such as building lists of words found on selected web pages. A crawler in crawler based search engines begin its work by visiting heavily used servers and popular databases [7]. It then visits popular web sites, indexes the words it finds before following links in the visited web pages. This enables the crawler to travel faster across widely used parts of the internet. At each step, the crawler carefully chooses the words and web pages to index using preset policies [7].

After collecting relevant information regarding the crawled web pages, the crawler stores the information in an index [7]. The data is then encoded to save storage space. When a user enters a query in given crawler based search engine, the search engine software sifts through the results and ranks them according to their relevancy in its result page [7].

The second type of search engines is the human powered directories. These are search engines that rely on humans for their web page listings. The search engines therefore get their listings of web pages from the submissions of various web page masters [7]. Examples of human powered directories include Yahoo, Galaxy and dmoz. The main advantage of human powered directories is that they have high content that are more appropriate compared to the crawler based search engines [5]. This is because the pages are submitted manually and appropriate measures are taken by editors to ensure that the pages are of high quality. It is however important to point out that submitting web pages manually is time consuming given the huge amount of information on the internet [7]. Moreover, changes made to an already submitted web page will not be reflected unless the page is resubmitted. Finally, it is impossible to change the rankings of pages in a given directory once the ranking is done [19].

The third type of search engines is the hybrid search engines. These are search engines that incorporate the features of crawler based search engines and human powered directories [7]. Some of the examples of hybrid search engines include MSN and Google. The fourth type of search engines is the Meta search engines. These are search engines that fetch

results from other search engines and display the results on their result pages [14]. These types of search engines therefore fetches results from other types of search engines, combine the results then rank the result before displaying the result on their result pages [7]. This type of search engines were useful at a time when search engines were less savvy and most search engines had significantly unique indices. An example of a Meta search engine is the MetaCrawler.

Compared to the classic search engines the visual exploratory Meta data search engine on a cloud platform is advantageous in that:

4.1 It deduces vast semantic relationships between the request and related concepts from broad meta-search results obtained from search engines like Google and making representations in the form of graphs.

4.2 The exploratory search enables the users freely and effectively explores adventure and acquaintance in the vast information world.

4.3 The new search engine is formulated on the basis of a personalized model thus has the ability to provide different users with different search experiences [1].

5. FINDINGS

Visual exploratory search involves multiple iteration and return sets that call for the user's cognition and interpretation. The items can be represented in various formats like graphs and texts which may require the web searchers to navigate, compare and reach qualitative and quantitative conclusions [10]. The human memory easily recognizes a keyword more easily than making out the same. Browsable information formats in the form of relational graphs and other visual exploratory structures give an outline of the content being searched thus enabling the searcher to navigate through the pool of information from a broad range to a finer scale. This therefore implies that the visual exploratory meta-data can greatly increase the experience of the user in learning and discovering [2].

The classic search engines do not permit the representation of a wide range of web information and knowledge in the form of relationship graphs and large meta-searches hence limit the learning and discovering experiences. The visual exploratory search engines allow the users a wide range of liberty to naturally and flexibly navigate through the query relationships in multi-scales. This allows for abstraction and extortion of information by users in their personalized levels and fineness [6].

The visual exploratory search is also easily compatible with smartphones with cloud connectivity such as Android and iOS devices. The devices offer a user-friendly interface hence have become more popular as well as their multiple functionality hence facilities easy navigation through vast data hence suiting it to cloud connectivity. The smart devices also have unique internet connectivity to Wi-Fi and 3G wireless networks thus providing quick data access whenever needed. This leads to the conclusion that the integration of the novel search engine with the smartphones could achieve great efficiency with regards to data access [20].

6. CONCLUSION AND RECOMMENDATIONS

From the analysis between cloud computing and Meta data search engine, there is a good indication pertaining to the storage of bulk data and speedy processing of the mass data. This implies that deploying a visual exploratory Meta data search engine to cloud computing platform can help alleviate the current problems with the classic search engines. However one shortcoming with this novel technology is that there is only one web site with bounded data sets. This implies that the mass data processing capability of cloud computing could be hampered.

It is therefore recommendable that future research should focus on alleviating this shortcoming just as the introduction of this search engine alleviates the problem of data processing and storage in the classic search engines.

7. REFERENCES

- [1] Castano, S., Ferrara, A. & Montanelli, S. (2012). Structured data clouding across multiple webs. *Information Systems*, 37(Semantic Web Data Management), 352-371
- [2] Chuang, J. Manning, C.D. & Heer. J. (2012). Termite: visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 74–77
- [3] Dhamdhere, S. N. (2014). *Cloud Computing and Virtualization Technologies in Libraries*. Hershey, PA: Information Science Reference.
- [4] Javadi, B., Tomko, M. & Sinnott, R. O. (2013). Decentralized orchestration of data-centric workflows in Cloud environments. *Future Generation Computer Systems*, 29 (1)
- [5] Kassim, J. M. & Rahmany, M. 2009, *Introduction to Semantic Search Engine*. Available from: <http://ieeexplore.ieee.org> [22April 2013].
- [6] Kitsos, I., Magoutis, K. & Tzitzikas, Y. (2014). Scalable entity-based summarization of web search results using MapReduce. *Distributed & Parallel Databases*, 32(3), 405
- [7] Levene, M. 2010. *An Introduction to Search Engines and Web Navigation*. John Wiley & Sons, New York.
- [8] Li, K., Shih, T. K. & Li, Q. (2014). *Cloud Computing and Digital Media: Fundamentals, Techniques, and Applications*. Hoboken: Taylor and Francis
- [9] Lloret Mauri, J., Lin, K., Rodrigues, J. & IGI, G. (2014). *Mobile Networks and Cloud Computing Convergence for Progressive Services and Applications*. Hershey, Pennsylvania: IGI Global.
- [10] Menard, E. & Smithglass, M. (2014). Digital image access: an exploration of the best practices of online resources. *Library Hi Tech*, 32(1), 98
- [11] Morris, M. R. & Teevan, J. 2010. *Collaborative Web Search: Who, What, Where, When, and Why*. Available from: <http://www.morganclaypool.com> [22 April 2013].
- [12] Ouyang, J., Zheng H., Kong F. & Liu T (2013). ' Visual Exploratory Search of Relationship Graphs on Smartphones '. *PLoS ONE* 8(11)

- [13] Ramachandran, M. (2014). *Advances in Cloud Computing Research*. Hauppauge, New York: Nova Science Publishers, Inc.
- [14] Romsaiyud, W. & Premchaiswadi, W. 2012. Exploring Web Search Behavior Patterns to Personalize the Search Results. Available from: <http://ieeexplore.ieee.org> [22 April 2013].
- [15] Schutt, R. & O'Neil, C. (2013). *Doing data science: Straight talk from the frontline*. Sebastopol, California: O'Reilly
- [16] Srinivasan, S. S. (2014). *Cloud Computing Basics*. Berlin: Springer
- [17] Tang, J., Tang, X. & Chen, D. (2012). A Visual Exploratory Search Engine Solution Based on Cloud Computing. 2012 Second International Conference On Cloud & Green Computing, 368
- [18] Vidal, J. C., Lama, M., Otero-García, E. & Bugarín, A. (2014). Graph-based semantic annotation for enriching educational content with linked data. *Knowledge-Based Systems*, 5529-42
- [19] Xiang, X, Shi, Y & Guo L. 2004. Rich Metadata Searches Using the JXTA Content Manager Service. Available from: <http://ieeexplore.ieee.org> [22 April 2013]
- [20] Yang, B & Ding J. (2012). 'A New Model of Search Engine based on Cloud Computing'. *International Journal of Digital Content Technology and its Application*.5 (6) 236-243
- [21] Yang, C. & Huang, Q. (2014). *Spatial Cloud Computing: A Practical Approach*. Boca Raton: Taylor & Francis.