A Cloud Computing with Search Engines

R.Hari Priyanka[1] R.Malathi Ravindran[2]
[1]M. Phil. Student [2]Assistant Professor
[1,2]Department of Computer Science
[1,2]NGM College Pollachi India

Abstract— Cloud computing can be defined as a large-scale and distributed computing paradigm which is driven by economies of scale, in which a pool of dynamically-scalable, abstracted, virtualized, managed computing power, platforms, storage, and services are delivered on demand to the customers through the Internet. search engines are programs that search documents for specific keywords and returns a list of the documents where the keywords were found. The term is often used to specifically describe systems like Google, Bing, and Yahoo! search that enable user to search for documents on the World Wide Web. In this chapter we discussed about the search engine optimization, search systems, search engine methods.

Key words: Search Engine, Search Optimization, Search Systems, Web Search Engine

I. INTRODUCTION

Search Engines has become a vital need and daily tool for internet users and online advertising media. Today, internet market top Search engines are turning profit from advertising, entertainment, social media networks, daily use applications (e.g. maps) and online product sales and services. Internet advertising/online advertising revenue. Internet traffic and business is growing day by day which indicates a massive growth in internet marketing and web development field. It is the art of designing, developing, modifying and coding Web pages, so that they can achieve high rankings in the search results and converting high user traffic. It rapidly changes their ranking algorithms with the passage of time but SEO basics remain same throughout. Search engine such as Google and Bing are the keys to finding the answers you seek in this huge mass of data.

II. LITERATURE REVIEW

Cloud search systems have been implemented for both commercial use and research studies. User behavior analysis is important for both Web information retrieval technologies and commercial search engine algorithms. With the expansion of information data, the current search engine is facing some serious problems, such as limited storage space and computing power[1]. search engines on cloud computing platforms have developed a new paradigm creating the necessity for the development of a wide range of mobile applications.[4],[5] 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 [8]. 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 [10].

III. SEARCH ENGINE OPTIMIZATION

Search engine optimization can be performed by two kind of practices/methods which are White-hat SEO and Black-hat SEO. White-hat SEO called ethical or legal SEO in which activities carried out according to the guidelines, rules and policies of search engines and normally most of SEO practitioners follow it. Whereas Black-hat SEO refers to the illegal SEO practices against search engine rules and regulations which derived to get quick search ranking results. There are billions of websites live on the internet. Most of them are based on managing, adding and modifying content. A content management system allows users to share, access control, contribute data and enhance communication between users in a collaborative environment. Joomla, Drupal, WordPress are the examples of today's most popular CMS applications. Most of CMS systems were not designed and developed according to all SEO standards. There are some specific problems with the content generated by every content management system.

IV. SEARCH SYSTEMS

The search system contains two types namely:

- Keyword based web search engine
- Faceted search system.

The users can have an intuition and basic idea on how roughly the keyword-based web search engines (such as Google and Bing) and the hierarchical faceted search engine work.

A. Keyword-Based Web Search Engine

On the user interface of most keyword-based web search engines, the key component is a text box where the user can enter any combinations of keywords as the query. As the "search" button is clicked, a HTTP request containing the query is sent to the server. Then, the server processes the query, searches the indexed documents and returns a list of most relevant results back to the user. The most challenging problem for the search engine is how to retrieve the relevant ones from billions of indexed documents in a millisecond timescale based on the keywords.

The first model is called Boolean retrieval model. In this model, keywords in the query are in combinations with logic operators (and, or, not) and each document is treated as a set of words. In order to quickly locate the documents matching the query, an inverted index list is built for each of the words appearing in all the indexed documents. The second model is called vector space model, which overcomes the problems of the Boolean retrieval model. This model has been widely used in the modern web search engines as the basic function for information retrieval. The basic idea of vector space model is that both the query and the document are converted to word vectors

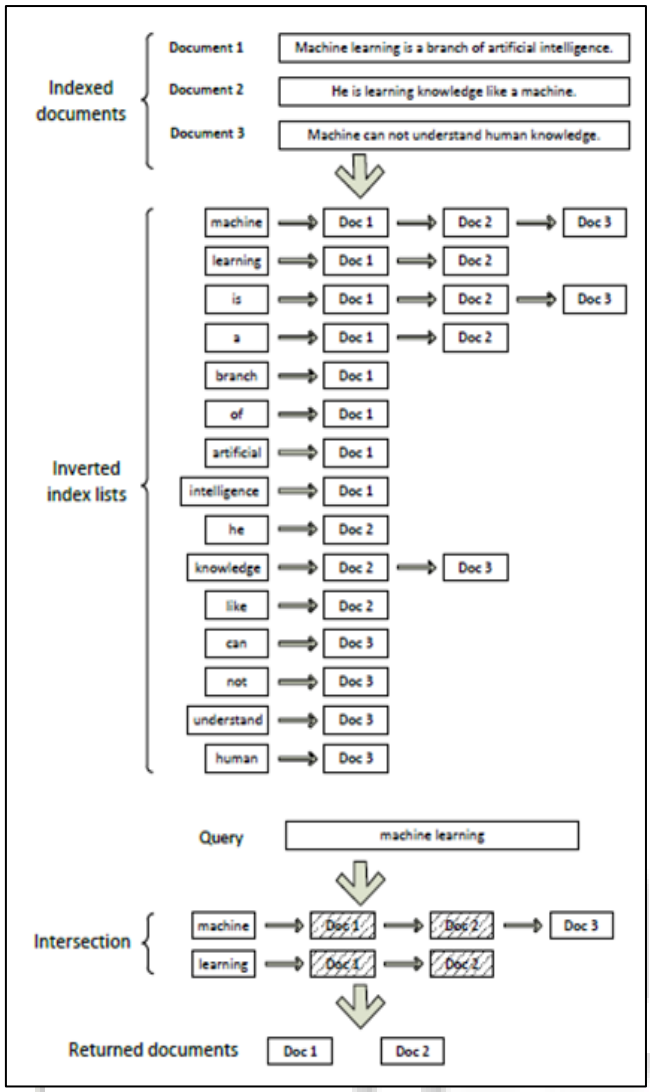

Fig. 1: Model

B. Faceted Search System

Faceted search systems, such as in Amazon.com and Kijiji.ca, allow users to explore a collection of information by applying multiple filters in the facets. The search system can usually have more than one facet, and each facet can be either a flat list or a structured hierarchy of various categories. For example, in Amazon.com, the top facet is the department facet, where all the products are categorized into a hierarchy. If we search "iPad" in the category "tablets", besides the department facet, additional facets (e.g., display size, hard drive size, brand, price, etc.) will be displayed to the user (shown in Figure 2)

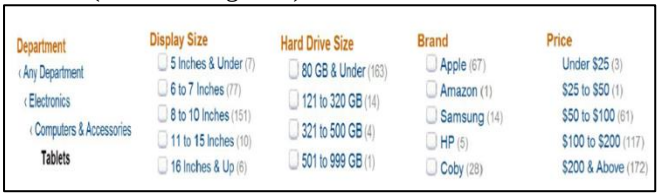

Fig. 2: Description

Usually users search in two ways in the faceted search system. The first way is to directly search the keywords in the search box, and then the user can not only obtain a flat list of results matching the keywords but also a group of facets with the results classified into their categories (sometimes displaying the number of the results). Afterwards, the user can iteratively refine the search results by clicking the categories of those facets.

V. Search Engine Methods

- Data Preprocessing
- Web Crawling and Caching
- Indexing and Searching
- Web Information Retrieval Using Particle Swarm Optimization

A. Data Preprocessing

The data preprocessing is to optimize the search process and to maximize storage capacity, recent crawled web pages are pre-processed before indexes can be built and the pages safely stored. Stemming is a process that reduces words by removing stem, thereby mapping them to the same root. The important difference between Stemming and Lemmatization is that the later reduces words to their roots, using vocabulary and morphological analysis of words, therefore removing ending and returning back to the dictionary form of a word, also known as lemma. Stemming is a much simpler heuristic process, which chops the end of words in the hope to perform these operations correctly in most of the times.

B. Web Crawling and Caching

A web crawler forms an integral part of any search engine. The basic task of a crawler is to fetch pages, parse them to get some more URLs, and then fetch these URLs to get even more URLs. In this process crawler can also log these pages or perform several other operations on pages fetched according to the requirements of the search engines. The Google search engine is a distributed system that uses multiple machines for crawling. The crawler consists of five functional components running in different processes. A URL server process reads URLs out of a file and forward them to multiple crawler processes

C. Indexing and Searching

The indexing and searching process the internet users perform billions of queries on web search engine daily. Most of common search engines like Yahoo, Google and Bing fetch results in milliseconds for the particular query, however it has been seen that results from the search queries vary (different) in different search engines respectively. Each web search engine depends upon one or more crawlers to present the content for its operation. Crawlers use a starting set of uniform resource locators (URL). The crawler retrieves (i-e copies and stores) the content on the sites specified by the URLs.

D. Web Information Retrieval Using Particle Swarm Optimization

PSO is an evolutionary computation method, which is clearly different from other evolutionary-type methods that does not use the filtering operation (such as crossover and/or mutation) and the members of the whole population are maintained through the search procedure. In order to find an optimal or near-optimal solution to the problem, PSO updates the current generation of particles (each particle is a candidate solution to the problem) using the information

about the best solution obtained by each particle and the entire population.

VI. CONCLUSION

Cloud computing is a new paradigm of computing utilities that promises to provide more flexibility, less expense, and more efficiency in IT services to end users. It is a combination of computation, software, data access and also provides storage services. search engines are programs that search documents for specific keywords and returns a list of the documents where the keywords were found. A search engine is a program or a software which fetches/retrieves data, files and information from a collective database or from the computer network. A Web search engine has three main parts. These are Web crawler, Indexer and searching. Firstly this paper presents search engine and search engine optimization. Secondly focused on search systems and its types. Thirdly this paper presents search methods and its types.

REFERENCES

[1] S. Grzonkowski, P. M. Corcoran, and T. Coughlin, "Security analysis of authentication protocols for next-generation mobile and CE cloud services," in Proc. IEEE Int. Conf. Consumer Electron., 2011, Berlin, Germany, 2011, pp. 83–87.
[2] D. X. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp. 44–55.
[3] C. Wang, N. Cao, K. Ren, and W. J. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, Aug. 2012.
[4] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst., Genova, ITALY, 2010, pp. 253–262.
[5] Bhushan Lal Sahu Rajesh Tiwari, " A Comprehensive Study on Cloud Computing", International Journal ofAdvanced Research in Computer Science and Software Engineering, Volume 2, Issue 9, September 2012 ISSN:2277 128X.
[6] Armbrust M et al (2009) Above the clouds: a Berkeley view ofcloud computing. UC Berkeley Technical Report. [10] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," in Proc. Conf. Comput. Commun. Secur., 2012, pp. 965–976.
[7] D. Cash, J. Jaeger, S. Jarecki, C. Jutla, H. Krawczyk, M. C. Rosu, and M. Steiner, "Dynamic searchable encryption in very large databases: Data structures and implementation," in Proc. Netw. Distrib. Syst. Security Symp., vol. 14, 2014
[8] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," in Proc. Conf. Comput. Commun. Secur., 2012, pp. 965–976.
[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYPT, 2004.
[10] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficiently searchable encryption," in Proc. of CRYPTO, 2007.

