



A Survey on Development of Search Engine

Prof. Mulik A.T¹, Heena Palkar²

Professor, Computer Science & Engg, DACOE, Karad, India¹

Student, Computer Science & Engg, DACOE, Karad, India²

Abstract: Search engine is the procedure used to improve the visibility of the search result for a free search engine or website and webpage. This paper aims to improve the search time of search engine, search function is an interface between a search engine and users. This function receives user's queries and returns relevant pages to the queries. There are some notable open source search engine projects but due to their complex design and implementation this project are still not a good for web search engine technology. There are lots of open source frameworks available to build an inverted index of web documents used for information retrieval.

Keywords: Search Engine, Open source frameworks, Information retrieval.

I. INTRODUCTION

A web search engine is an information retrieval system designed to help information find which is stored on World Wide Web (or simply the web). Due to fast growth in the size of the web, web search engines face a big performance challenges such as storage requirements, data retrieval rate, query processing time. Large search engines have to be able to processes tens of thousands of queries per second on tens of billions of documents, making query throughput a critical issue.

To satisfy this heavy workload web search engines use variety of performance optimization including query optimization, high-speed processing. Web based search engine are related to their effectiveness in retrieving the closely relevant information with efficiency over the internet. For the instance a Google has done the great job in improving both the effectiveness of information retrieval and the efficiency of the query performance.

II. LITERATURE SURVEY

An Kongres Pengajaran dan Pembelajaran UKM, 2010.[1]

Documents are in various formats such as email, websites, memos, letters and articles. The text transformation transforms documents into index terms or a feature which involves lexical analysis (parsing-tokenizing-stop word removal-stemming). Text acquisition identifies and stores documents for indexing. Index creation Takes index terms and creates data structures (indexes) to support fast Searching.

User interaction Supports creation and refinement of query, display of results. Ranking Uses query and indexes to generate ranked list of documents. Evaluation Monitors and measures effectiveness and efficiency.

Application of Full Text Search Engine Based on Lucene Rujia Gao¹, Danying Li², Wanlong Li¹, Yaze Dong³ [2]

With the rapid development of Internet and web information, Internet users how to remove the impurities and quickly and easily to gain the information they need in the vast ocean of information to become a hot research topic in this field. The core of information search is retrieval technology. Search technology provided us with the information retrieval tool according to the content of data rather than the external features based on a variety of computer data such text. Create all the possible terms in the index which are searched by network users as well as help people to oversee and arrange broad data and empower organize clients to rapidly and effortlessly recover any information they need. Lucene has become one of the most highly praise and most popular information retrieval library. Firstly, we should build a text database which is used to store all information retrieved by the user, then determine text model of retrieval system. Create index with the model according to the text of database. Search requests are submitted by the users and information retrieval systems to preprocess and search the information eventually return user the information. Indexing can greatly improve the speed of information retrieval. Which way do you use depends on the scale of information retrieval system. After indexing the documents, you can start to search information you need.

A Novel Compressed Index-Query Web Search Engine Model [3]

At the point when the record is prepared the looking procedure can be performed through question interface, a client enters an inquiry into a web index (ordinarily by utilizing watchwords), the motor looks at its file and gives a posting of best coordinated Web pages as per its criteria, typically with a short synopsis containing the record's title and in some cases parts of the content. At this stage, the outcomes positioned, where positioning is a relationship between an arrangement of things with the end goal that,



for any two things, the first is either "positioned higher than", "positioned lower than" or "positioned equivalent" to the second. IR frameworks perform association report accumulations and make it conceivable to get to reports fulfilling the requirements of a client. All through this theory the meaning of a report is utilized as takes after "An archive is a unit of content that is returned because of questions

Design and Implementation of Search Engine Using Vector Space Model for Personalized Search [4].

The Internet can enable people to get the information more efficiently. This leads to information in enormous forms, and thus network information grows exponentially. A search engine is a web-based tool that enables users to locate information on the World Wide Web. As the Search Engines (SE) gives us lots of information so that SE is more essential in our life.

III. BACKGROUND AND MOTIVATION

Existing search engine are very slow because they work on only indexed data and it require large time to response the query. To avoid this we design the search engine system in which data is stored on the database and then apply the paging technique in which each record is stored on the separate page and this pages are listed sequentially by indexing technique .To develop this system for fast semantic business retrieval search engine.

IV. CONCLUSION

Search engines are programs that search document for specified keyword on search for information on the World Wide Web and returns with a list of documents where the keywords are found. Today's internet users are very positive what search engine already do and they feel good about their experiences. When searching the internet search engine offer users vast and impressive amount of information available with speed. Search engine is the process of internet marketing and 90% of the web user turning to search engines to search information online.

REFERENCES

- [1] Kongres Pengajaran dan Pembelajaran UKM, 2010 Development of Search Engines using Lucene: An Experience Masnizah Mohd Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia
- [2] Application of Full Text Search Engine Based on Lucene Rujia Gao1, Danying Li2, Wanlong Li1, Yaze Dong3
- [3] Design and Implementation of Simple Web Search Engine
- [4] Search Engine Optimization Using Data Mining Approach Khattab O. Khorsheed, Magda M. Madbouly, Shawkat K. Guirguis
- [5] On Indexing Mechanism in Geographical Information Retrieval System Xing LIN, Bo YU, Yifang Ban Geoinformatics, Royal Institute of Technology (KTH), Stockholm, Sweden Information Sciences and Technology, the Pennsylvania State University, USA
- [6] Application of Full Text Search Engine Based on Lucene Rujia Gao1, Danying Li2, Wanlong Li1, Yaze Dong3 1Computer Science

- and Technology, Changchun University of Technology, Changchun, China
- [7] High Performance Full Text Search for Database Content By Peter Zaitsev, Vadim Tkachenko Percona Ltd
- [8] A Novel Compressed Index-Query Web Search Engine Model