Quick Business Information Retrieval using Dot Lucene

Prof. Mulik A.T[1], Komal Londhe[2]

Professor, Computer Science & Engg, DACOE, Karad, India [1]

Student, Computer Science & Engg, DACOE, Karad, India [2]

Abstract: This system works on indexed data so it is easy and fast retrieval system over large data with millions and trillions of columns. This system will provide the highly accurate and semantic result with minimal time. Two major components of search engines are the indexing and query process. The indexing process aims to create data structures or the indexes that allows the searching. The querying process will use the structures and user queries to generate a ranked list of documents. Search engines should include the concept, models, techniques and the processes of IR. Search function is an interface between a search engine and users. This function receives user's queries and returns relevant pages to the queries. The pages usually are sorted according to some criteria. 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. Ordering can enormously enhance the speed of data recovery which way do you utilize and relies upon the size of data recovery framework. After indexing the documents search requests are submitted by the users and information recovery frameworks and seek the data in the long run return to client.

Keywords: Search Engine, Indexing, Paging, Lucene.

I. INTRODUCTION

Lucene is a Java library which is able to perform the indexing and searching process. Lucene full text search is mainly composed of searching, indexing and analysis three modules. Searching module is mainly used for interacting to users. Analysis module is responsible for preprocessing document information. The principal role of indexing module is to enhance the speed of retrieval. It allows the development of a content-based information retrieval systems or applications such as search engines. Information retrieval (IR) is a field concerned with the structure, searching, organization, analysis, and retrieval of information. It emphasizes on the process of match user queries to the index in finding relevant documents. Lucene is an open source Java library which supports the process and techniques of information retrieval. Lucene is able to index text from a various formats such as HTML, PDF and Microsoft Word, and also in various languages. Applications for example, Amazon are among the business application that utilizations Lucene for indexing and permitting effective searching.

II. RELATED WORK

An Kongres Pengajaran dan Pembelajaran UKM, 2010[1]

Text acquisition are identifies and stores documents for indexing. Documents are in various formats such as websites, email, memos, letters and articles. The text transformations are transforms documents into index terms or features which involves lexical analysis. Index creation is takes index terms and creates data structures (indexes)

to support fast Searching. User interaction Supports to the creation and refinement of query and display of results. Ranking uses query and indexes to generate ranked list of various documents. Evaluation monitors and measures effectiveness and efficiency.

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

With the fast advancement of web information and internet, internet clients how to remove the polluting influences rapidly and effortlessly to pick up the data they require in the unlimited data. The centre of data inquiry is recovery innovation search technology gave us with the information retrieval tool according to the content of information rather than the external features based on a variety of computer data such as text. Create all the possible terms in the index network users to quickly and easily retrieve any information they need. Lucene has become one of the most highly praise and most popular information retrieval library. Firstly, we should assemble a text database which is used to store all data recovered by the client, then determine text model of recovery system. Create index with the model related to the text of database. Indexing can improve the speed of information retrieval. Which way do you use depends on the scale of information retrieval system. After indexing the information, you can start to search information you need. Search requests are given by the users and information retrieval systems to process and search the information and return the result to the user.

III. PROPOSED SYSTEM

Information recovery is for the most part considered as a segment of software engineering that arrangement with the representation, storage, and access of data. The principle point of data recovery show (IR) is to "finding significant learning base data or a record that satisfy client needs". There are three essential procedures an IR show needs are the representation of the data of the archives, the interpretation of the client's data require, and the correlation of these two representations.

IV. IMPLEMENTATION

4.1 LUCENE :
Lucene .Net is a line by line port of apache Lucene which is a high performance, full featured text search engine library written entirely in java .It is a technology suitable for nearly any application that requires full text search. Lucene is faster than Sql query and it is absolutely fast when searching for some text.

4.2 INDEXING:
Search engine indexing gathers, parses and stores information to facilitate quick and accurate information retrieval. The purpose of storing and indexing is to improve speed and performance in finding relevant documents for a search query. The reason for putting away a list is to enhance speed and execution in finding significant records for a pursuit question. Without a file, the web crawler would examine each archive in the corpus, which would require extensive time and processing power. For instance, while a list of 10,000 archives can be questioned inside milliseconds, a consecutive sweep of each word in 10,000 expansive records could take hours. The extra PC stockpiling required to store the file, and the significant increment in the time required for a redesign to happen, are exchanged off for the time spared amid data recovery.

4.3. SEARCH:
In the wake of ordering the records, you can begin to inquiry data you require. Look solicitations are put together by the clients and data recovery frameworks to preprocess and seek the data in the end return client the data.
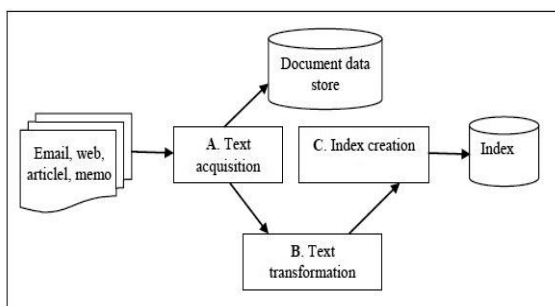

Figure 4.2 Indexing process in search engines

4.3 PAGING:
Paging is the process of dividing a document into discrete pages either electronic pages or printed pages. In paging stores each record is stored on separate page and display this information to the user as per the all details.
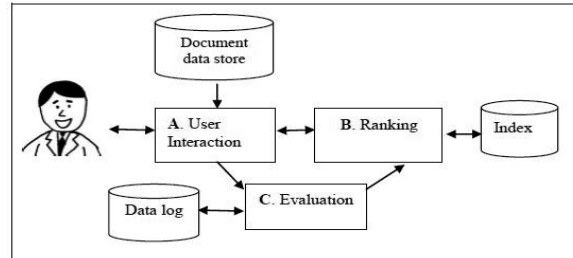

Figure 4.3 paging process in search engines

4.4 SORT RESULT
After the data recovery framework look the data that the clients need and it will channel or sort the data by making a specific govern after that arrival the client related data

V. CONCLUSION

We have presented web search engine software that is suitable for researches and learning purposes because of its simplicity, portability, and modifiability. We provided many scores functions to sort relevant pages to user queries. This offer users vast and impressive amounts of information, available with a speed. This system is useful for business owner for give their product advertisement and for knowing the feedback of customer in the form of ratings. This is market place for the various kind of information which can define the company's profile. This is work in the two bride categories one is they use to provide information about business and they serve the customer towards the right choice for them. It gives more freedom to the business owner to interact with the people. For the customers, it is more beneficial of using directory sites to make a proper decision and make a benefit of them mostly in terms of buying a product or any services. This is not only the medium of making base for business it can also use for huge change in business.

REFERENCES

[1] Kongres Pengajaran dan Pembelajaran UKM, 2010 Development of Search Engines using Lucene: An Experience Masnizah Mohd Faculty of Information Science and Technology, Universiti 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] Application of Full Text Search Engine Based on Lucene Rujia Gao1, Danying Li2, Wanlong Li1, Yaze Dong3 Computer Science and Technology,Changchun University of Technology, Changchun, China
[6] On Indexing Mechanism in Geographical Information Retrieval System Xing LIN*, Bo YU**, Yifang Ban
[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