

PARTITIONING DATA TO INCREASE WEBSITE VISIBILITY ON SEARCH ENGINE

Kirubahar. J¹, Mannar Mannan. J²

*¹PG Scholar, ²Teaching Assistant, Department of IT, Anna University Regional Centre,
Coimbatore, Tamilnadu (India)*

ABSTRACT

Currently in search engine, users unable to get exactly appropriate result. It takes long time to generate result to the user. In terms of cost its overhead by getting data. Getting a higher number of data results in a higher overhead. In our search result loss of unwanted information. So we partition the data and retrieving the result based on user level. Users have to specify their user level. Search engine will returned the result based on the user level and key word. The main scope of the project is get accurate search result to the user. So we can get the expected result and it will be reduced time to our search document. It can reduce the unwanted data. In this paper we first formally introduce Term Partitioning as a new optimization problem. Then we get the process of relevant search web links. It may easier to get our search result easier. And each web link shown by a ranking process.

Key Terms: Search Engine, Indexed Pages, Crawlers, SEO, SERP, Horizontal Partitioning

I. INTRODUCTION

Search engine optimization is a process of affecting the visibility to the user on a webpage in a search engine that gives a organic search results. In a search engine that can include image search, local search, video search, academic search, news search and industry specific vertical search engines. In search engine, we get higher ranking on the search result page and more frequently link appears in the search result list, most of the users can get these kind of methodology in a search engines. . WWW users and content providers give optimizing sites for search engines in the mid-1990s, as the first search engines were on the early Web. All the web users needed to do submit the address of a page, or URL, to the various engines which would post a "spider" to "crawl" that page, extract links to other pages from it, and return information found on the page to be indexed. The operation calls for a search engine spider downloading a page and storing it on the search engine's own server, where a second plan, recognized as an indexer, extracts various information about the page, such as the words it checks and where these are settled, as well as any weight for specific words, and all links the page contains, which are then put into a scheduler for crawling at a later appointment. The network creates new challenges for data recovery. The amount of data on the web is maturing rapidly, as considerably as the number of new users inexperienced in the artistic production of web research. People are likely to surf the network using its link graph, often beginning with high quality human maintained indices such as Yahoo! Or with search engines. Human maintained lists cover popular topics effectively but are subjective, expensive to build and maintain, slow to amend, and cannot cover all issues. Automated search engines that rely on keyword matching usually return too many low quality matches. To build things worse, some advertisers attempt to earn people's attention

by calling for measures meant to mislead automated search engines. We have established a large-scale search engine which addresses many of the problems of systems. It makes especially heavy use of the additional structure present in hypertext to provide a good deal higher quality search solutions.

1.1. Relationship and methods in search engine

The existing leading search engines like Google, Bing, Yahoo!. From that most of the search engines use crawlers to find pages for their algorithmic search results then they are applying the indexing operation. Pages that are linked from other search engine indexed pages do not necessitate to be taken because they are set up automatically. And to bring the network links they use the red worms. Search engine crawlers may look at a number of different factors when crawling a website. Not every page is indexed by the search engines.

Search engine optimization is ever shifting and evolving, and then it's important to save up with optimization trends and techniques. In the previous articles in our SEO series we dealt with the bedrock of search engine optimization as well as some tips & tricks of the craft. We will explore more advanced techniques and reveal why they are significant. Be warned though, this is the stuff that can get very time consuming and the consequences are likely to be less spectacular. Even so, the effort may bump your ranking just enough to propel you off the second page to the always coveted top ten and first page.

In this paper we are using a partitioning technique in that we can be used of horizontal partitioning is a database design principle whereby rows of a database table are taken separately, instead than being split into columns (which is what normalization and vertical partitioning do, to differing extents). Each partition forms part of a shard, which may in turn be located on a separate database server or physical position.

On that point are numerous advantages of the horizontal partitioning approach. Since the tables are split up and dispersed into multiple servers, the total number of courses in each table in each database is scaled down. This cuts down the index size, which generally improves search performance. A database shard can be located on separate hardware, and multiple shades can be located on multiple cars. This enables a distribution of the database over a heavy number of machines, which entails that the database performance can be broadcast out over multiple machines, greatly improving operation. In accession, if the database shard is based on some real-world, segmentation of the data (e.g., European customers v.v. American customers) then it may be possible to infer the appropriate shard membership easily and automatically, and query only the relevant shard.

1.2 Need for Study

This report analyses the search engine algorithms and ranking principle, we hashed out the situation structure, keywords, single-page optimization, and search engine penalties. In add-on, search engine optimization techniques developed in late years, which have a really significant function in the corporate internet site ranking. We use some search engine optimization technology applications to discuss this increasing technology. Due to Webmasters and content providers began optimizing sites for search engines in the mid-1990s, as the first search engines were cataloguing the early Web. Initially, all webmasters needed to serve was to submit the address of a page, or URL, to the various engines which would post a "spider" to "crawl" that page, extract links to other pages from it, and return information found on the page to be indexed. The operation calls for a search engine spider downloading a page and storing it on the search engine's own server, where a second plan, recognized as an indexer, extracts various information about the page, such as the language. It checks and where these are

settled, as well as any weight for specific words, and all links the page contains, which are then put into a scheduler for crawling at a later appointment.

II. RELATED WORK

"How to Use Search Engine Optimization Techniques to Increase Website Visibility". [1] What contributes to search engine rankings? And what can web content creators and webmasters do to create their content and sites easier to find by audiences using search engines? Key concepts: Search engine rankings are shaped by three classes of participants: search engine companies, programmers and search engine optimization practitioners, and search engine users. Key lessons: By using three key lessons, professional communicators can create it more comfortable for audiences to get their web content through search engines: (1) consider their web content's audiences and website's competitors when analyzing keywords; (2) insert keywords into web text that will appear on search engine results pages, and (3) involve their web content with other websites and web content creators. Implications: Because successful search engine optimization requires considerable time, professional communicators should progressively apply these lessons in the sequence presented in this tutorial and should keep up to date with frequently changing ranking algorithms and with the associated changing practices of search optimization professionals.

"Worst practices in search engine optimization"[5]. Many online companies have become aware of the importance of ranking well in the search engines. A recent article reveals that 62% of search engine users click only on answers that appear on the first search engine results page (SERP) and less than 10% of users click on results that appear. In parliamentary law to rank comfortably in the SERPs companies have commenced to utilize search engine optimization techniques (SEO). That is, they manipulate the site's content and meta tags, as well as endeavour to attract incoming links from other websites. Nevertheless, certain SEO techniques directly violate the guidelines published by the search engines. While the specific guidelines vary a bit, they can all be summarized up as: show the same content to search engines as you present to users.

"Comparing rankings of search results on the web"[2] In this report we introduce a number of bills that compare rankings of search engine results. We use these amounts to five questions that were monitored daily for two periods of 14 or 21 days apiece. Rankings of the different search engines (Google, Yahoo! and Teoma for text searches and Google, Yahoo! And Pic search for image searches) are compared on a everyday basis, in addition to longitudinal comparisons of the same engine for the same query over time. The answers and rankings of the two points are compared as well.

"Standard parameters for searching behaviour in search engines and their empirical evaluation" [6] Search engines have become the most significant medium for Internet users to find pages on the WWW. They assist clients to decrease their information overload, and heighten the sales of commercial sites in different ways. For these reasons, the exploration of and changes in (human) online searching behaviour has become a topic of especial importance. This theme will help search engine and website administrators and developers to monitor online searching behaviour properly and to derive strategies from the information gathered. We define standard parameters against which search engines can be quantified and compared. These parameters as well reflect the online searching behaviour of search engine users.

"Analyzing Google rankings through search engine optimization, data"[3] The report gives the outcomes of a study into 50 highly optimized web pages that were produced as part of a Search Engine Optimization

competition. The work concentrates on the most popular techniques that were employed to rank highest in this rivalry, and includes an analysis on the use of Page Rank, number of pages, number of in-links, domain age and the role of third party sites such as directories and social bookmarking websites. A separate study was formed into 50 non-optimized web pages for comparing. The report offers insight into the techniques that successful Search Engine Optimizers use to secure a page ranks highly in Google.

"The issue of user intent on the stability of search engine results"[4] Previous work has demonstrated that search engine queries can be sorted according to the purpose of the searcher (i.e., why is the user searching, what specifically do they signify to execute). In this clause, we report an experiment in which four sets of queries, each set representing different user intent, are repeatedly submitted to three search engines over a period of 60 days. Employing a variety of measurements, we identify the overall stability of the search engine results recorded for each group.

2.1 Search Engine Optimization

Webmasters and content providers began optimizing sites for search engines in the mid-1990s, as the first search engines were cataloging the early Web. Initially, all webmasters needed to serve was to submit the address of a page, or URL, to the various engines which would post a "spider" to "crawl" that page, extract links to other pages from it, and return information found on the page to be indexed. Keyword research is the cognitive operation of finding out which keyword phrases people are using to lookup for the products or services you offered.

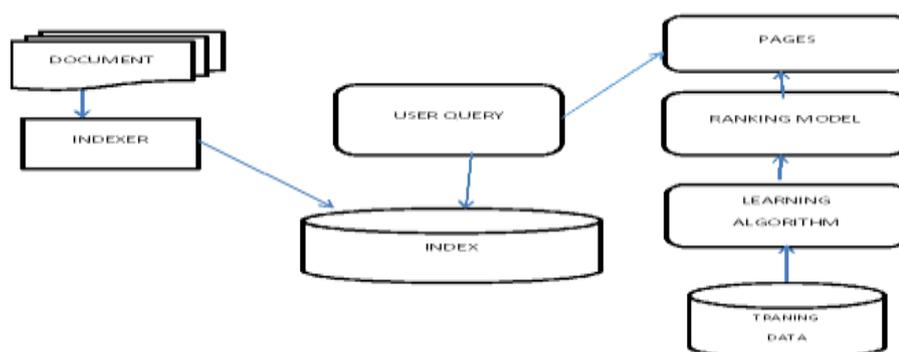


Fig 1 Basic architecture of search engine optimization

Keyword research also involves determining which keywords are most relevant to your website: there is a big difference between optimizing a website for keywords that are currently bringing traffic to your website and optimizing for keywords that are better suited for your website.

Keyword research is a relatively complex procedure that calls for multiple tools to provide quality keyword phrases and data on their relative popularity and expected effectiveness. Taking keyword research should be the beginning point for anyone interested in SEO for their internet site. Because SEO is a process, there are no guarantees of immediate success. But we will guarantee that without conducting effective and thorough

keyword research, subsequent attempts at optimizing a website or page for search engines will be far less efficient.

Process includes following steps:

- Determine the composition and scope of the client’s business or establishment. We have interviews with the client and review their existing web site to better interpret the client organization’s products and services.
- Discover new keywords, We generate new keywords based on synonyms, semantically related terms, and tangential phrases.
- Evaluate the list of keywords. Using SEO tools and supporting data, we evaluate the intensity of the generated keywords.
- Pare down and select a final list of keywords. We narrow down the list of keywords based on their relevancy, the client’s specifications, and important metric data relating to those terms. At the final stage of the keyword research process, we’ll supply you with a detailed report that includes an executive summary of our determinations, our recommendations for keyword phrases to optimize your situation for, supporting data, analysis of the data, and suggestions for next steps.

2.2 Disadvantages in Existing System

It requires long time to generate answers to the user. In conditions of cost its overhead by getting data. Getting a highest number of data results in a higher overhead. In our search results loss of undesirable information. Temporary disconnections are possible, hence additional mechanisms are needed.

Every search engine uses different complex mathematical patterns to generate search results. The consequences for a specific query are then exhibited on the SERP. Search engine algorithms consider the key elements of a web page, including the page title, content and keyword density, and come up with a ranking for where to put the results along the pages.

III. SYSTEM AECHITECTURE DESIGN

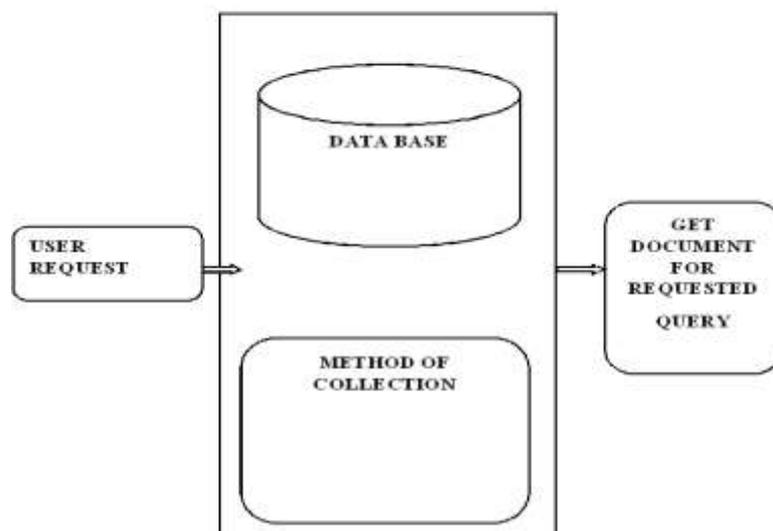


Fig 2 System Architecture Design

3.1 Detailed Description

Currently in search engine, users unable to get exactly appropriate result. So we partition the data based on user level. Hence we can reduce users search accuracy & time and also search engine running time. The advantages of the proposed system are its help's to retrieve the related data easier. Then retrieve the data short span of time. They can apply to make it easier for audiences to find their web content and websites in our search engine.

In this paper search engine partitioning the data to increase the website visibility.

Partitioning levels

- Beginner level
- Expert level

After process of user level selection then user enters the query then its enters into the process of

1. Data Base
2. Method of document collection

Before get content out of your MySQL database, you must know how to establish a connection to MySQL from inside a PHP script. To perform basic queries from within MySQL is very easy. This article will show you how to get up and running.

Let's get started. The first thing to do is connect to the database. The function to connect to MySQL is called my sql connect. This function returns a resource which is a pointer to the database connection. It's also called a database handle, and we'll use it in later function

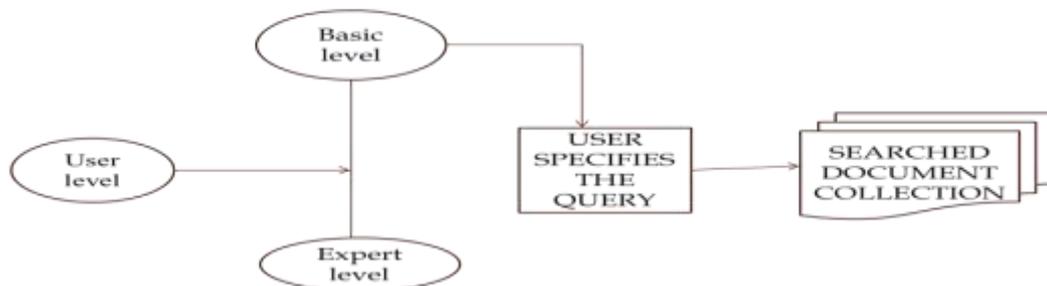


Fig 3 Selecting user level

Then we get the process of relevant search web links. It may easier to get our search result easier. And each web link shown by a ranking process.

IV. ALGORITHM FOR RANKING PROCESS

In the HITS algorithm, the first step is to retrieve the most relevant pages to the search query. This set is called the root set and can be obtained by taking the top n pages returned by a text-based search algorithm. A base set is generated by augmenting the root set with all the web pages that are linked from it and some of the pages that link to it. The web pages in the base set and all hyperlinks among those pages form a focused sub graph. The HITS computation is performed only on this focused sub graph. According to Kleinberg the reason for constructing a base set is to ensure that most (or many) of the strongest authorities are included.

Authority Update: Update each node's Authority score to be equal to the sum of the Hub Scores of each node that points to it. That is, a node is given a high authority score by being linked from pages that are recognized as Hubs for information.

Hub Update: Update each node's Hub Score to be equal to the sum of the Authority Scores of each node that it points to. That is, a node is given a high hub score by linking to nodes that are considered to be authorities on the subject.

V. CONCLUSION AND FUTURE WORK

An information provider wishes to dispatch specific information to the user. Through a user level, we designed the SEO process which selects “levels” and lets them act as data broadcasters within a region of interest. So, the user can get the appropriate result easier. We also showed that a significant further increase in performance can be achieved by efficiently managing by a user of the content.

Now I did the process of study based process only in my search engine. My future work is to built SEO partitioning process for all knowledge based process and interested to put on this process in android. In the future more granular partitioning of data can be implemented. By making use of learning algorithms search engine can designed to learn users level using the users searching pattern.

REFERENCES

- 1.Bar-Ilan. j , M. Mat-Hassan, and M. Levene, “Methods for comparing rankings of search engine results,” *Comput. Netw.*, vol. 50, no. 10, pp. 1448–1463, 2006.
- 2.Bar-Ilan. j "Comparing rankings of search results on the web", *Inf. Process. Manage.*, vol. 41, no. 6, pp.1511 -1519 2005.
- 3.Brin. s and L. Page. (1998). "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, pp.107 -117 1998.
- 4.Grappone. J and G. Couzin *Search Engine Optimization: An Hour a Day*,, 2011 :Wiley Search engine ranking factors. 2011.
- 5.Hochstotter. N and D. Lewandowski "What users see—Structures in search engine results pages", *Inf. Sci.*, vol. 179, no. 12, pp.1796 -1812 2009.
- 6.Hindman. M , K. Tsioutsoulis and J. A. Johnson. (2003). *Googlearchy : How a few heavily-linked sites dominate politics on the web.*

7. Inrona. L and H. Nissenbaum "Shaping the web: Why the politics of search engines matters", Inf. Soc., vol. 16, no. 3, pp.169 -185 2000.
8. Jhon B. Killoran "How to Use Search Engine Optimization Techniques to Increase Website Visibility" IEEE transactions on professional communication, vol. 56, no. 1, march 2013.
9. Jansen. B and P. R. Molina "The effectiveness of web search engines for retrieving relevant ecommerce links", Inf. Process. Manage., vol. 42, pp.1075 -1098 2006 .
10. Killoran. J. B "Writing for robots: Search engine optimization of technical communication business web sites", Tech. Commun., vol. 57, no. 2, pp.161 -181 2010.