

Proficient Data Mining Approach for Search Engine Optimization

Dr.P. Sujatha and K. Kavitha

Abstract--- In today's digital era, technological advancement such as internet of things, cloud computing, sensor networks, and video surveillance has drastically change the type and nature of data. Such data no longer implies simple text, but includes audio, video, images and animations. Consequently the rapid growth of information on the web makes search engines as a replacement of traditional media. However retrieving the relevant information from billions of web pages is the toughest task for the search engine. Search Engine Optimization (SEO) provides big remedy for displaying the relevant web pages based on ranking. This paper presents proficient data mining approach for Search Engine Optimization for improving the ranking of the website much easier. In addition this paper discusses about three essential SEO techniques such as keyword generation, directory submission and link exchanges.

Keywords--- Data Mining, Search Engines, Search Engine Optimization, SEO Technique

I. INTRODUCTION

Due to Computerization in all fields such as business, banking, science and healthcare, huge amount of data is available nowadays. Data mining is an emerging research field that helps in retrieving information from large databases. Data mining is the process of transforming data into useful information. It is also known as mining of knowledge from data. The main goal of data mining is to extract information from a data set and transform it into an

easily understandable structure for further use [2]. The main tasks of data mining are,

Association rule learning: It identifies the relationship (association) between various attributes in large data sets. The main goal of association rule is to develop large number of rules from data repository. Association rules are employed in many application areas such as intrusion detection, bioinformatics and web usage mining.

Clustering: It is the process of analyzing the set of objects and grouping them that have similar patterns. However the patterns are not known in advance. For example, in an employee database, clusters may be formed for employees who earned more than eighty thousand.

Classification: It groups the information into different classes based on previously known structures. Classification comes under supervised learning because it uses known trained data set to form patterns. For example, an employee data may be classified according to designation.

Outlier Analysis: This task identifies the records that do not match the usual patterns. It is useful for fraud detection in credit card transaction.

a. Applications of Data Mining

Researches use Data mining for the following applications.

- Search engine optimization (SEO)
- Medical and health care.
- Fraud detection
- Computer Security
- Education
- Banking
- Marketing

Dr.P. Sujatha, Associate Professor, Department of Computer Science, Vels University, Chennai, India. E-mail:sujinagi@gmail.com

K. Kavitha, Research Scholar, School of Computing Sciences, Vels University, Chennai, India. E-mail:kavimct16@gmail.com

This paper focuses with one of the emerging research areas known as Search Engine Optimization and how data mining is helpful for this.

b. Role of Data Mining in SEO

Internet is an enormous and alive data collection which includes countless hyperlinks and huge volumes of information. All web sites include dynamic data and they may not follow any uniform structure. Information available in all web pages is raw data and also it is not indexed. In general, user concentrates only with small portion of information in the web by leaving irrelevant data. Therefore searching such information has become more complex and difficult. Hence SEO requires effective data mining technique for search engine optimization.

II. LITERATURE REVIEW

SurajAnjaana [7] suggested manual submission as the best technique since two things are different such as directory and target users. Bo Xing et al. [4] discussed many features of information retrieval vehicles as well as search engine marketing. To boost websites paid placement or SEO methods are suggested. However SEO is not costly compared with paid placement. In [6] David discussed directory submission as a best method to create links as well as to increase web rankings. They provide links to other databases as well as to other search engines.

Abhishek et.al [8] described the importance of keyword that acts as a short phrase in the search engine for extracting relevant information. Many of the search engines use keyword as a tool for ranking websites. Different types of keywords such as Query log, proximity searches, advertiser log mining and Meta tag crawlers are discussed.

Najork et.al [9] discussed about link exchanges and how this method allow websites to exchange links with other related websites with high page rank.

III. SEARCH ENGINES

Search engines are programs that are used to search documents using specific keywords and search information

on the World Wide Web. List of documents will be returned by the search engine in which the keywords were found. Google, Yahoo, AOL Search, Bing, Web Crawler, Dog Pile, Alta Vista are some of popular search engines.

a. Goals of Search Engine

The main goal of search engine is,

- Quality
- Speed

Quality – Effectively retrieving the most relevant documents for the given keyword.

Speed – Efficiently processing the user queries as fast as possible and return the set of documents.

b. Characteristics of Search Engine

- Performance: How fast the search engine works and how fast it gives the results?
- User Friendly: It indicates how for search is easy to work with the interface.
- Consistency: It specifies the reliability of the search engine.
- Results Display: Results should be ranked by relevancy and most relevant documents to be displayed on the top.

c. Working of Search Engine

Search engine is mainly used for retrieving information from data bases of HTML documents in the web. Generally Web based search engine work as follows. First the search engine creates an index that contains word in each document and also pointer to its location inside the document. Next, it matches the given query against an index. When the match occurs, it saves the retrieved information of many web pages. Then the web crawlers (spider) retrieved these pages and submit to the users [3].

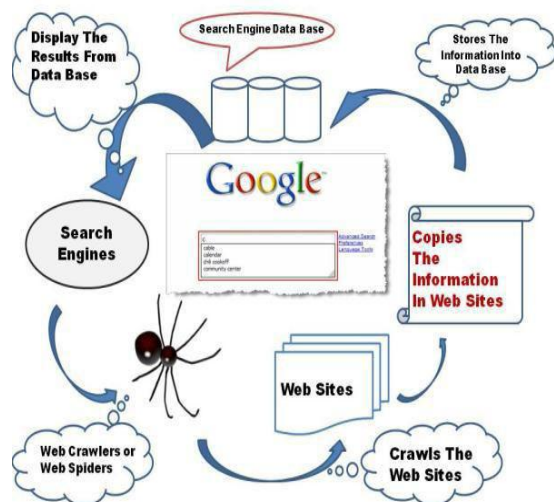


Fig. 1: Working of Search Engine [3]

Types of Search Engine: Based on the functioning, search engines are classified as,

- **Human Power Directories**

It depends on the human for submitting information as well as for indexing. This type of search engine never contains full text or webpage to which it is linked.

- **Crawler Based Search Engine**

It uses automated software (crawlers) for searching a website. Crawlers read the content of each website and also read the website's meta tags. Then the crawler follows the links of the website to create index. At last the crawler returns all information to a central depository in which the data is stored and indexed. Crawlers retrieve huge amount of information.

- **Hybrid Search Engine**

This is the combination of crawler based search engine and human power directories. It is different from text oriented search engine (Google) or directly based searched engine (Yahoo). In these search engines each program operates by comparing a set of metadata.

IV. SEARCH ENGINE OPTIMIZATION

It is a method for affecting the visibility of website in search engine [5]. In other words, SEO is the procedure for

improving the visibility of a web page in search engine through the unpaid searched results [10]. The target of optimization includes different searches such as video search, image search, academic search, local search.

In general, searching information in the web uses two methods: paid placement and search engine optimization. The first method displays results but amount should be paid for improving the search ranking of the website. The amount is calculated by placing keywords. Hence SEO is adopted to optimize web sites with free of cost. As well as web spammers use SEO techniques for improving low level search ranking web sites into highest search ranking websites.

V. SEARCH ENGINES OPTIMIZATION TOOLS

To optimize the functionality of the search engine, some operators are used that are known as search engine optimization tools.

- **Keyword Tool:** This tool is used for website classification as well as for regulating keywords deployment columns. It includes the tools for keyword density analysis, competitor analysis and keyword research. Keyword selector tool, keyword external tool are few example for keyword tool.
- **Link Tool:** To increase the ranking of the website link tool uses link popularity spider simulator.
- **Usability Tool:** This tool is used to test display effects of page in different resolution, different operating system, and different browser. It includes HTML and CSS validation, Firefox extension, and Page speed test.
- **Keyword Strategy:** It includes strategies for selecting appropriate keywords that should be related with products, service and area.
- **High duality incoming Link:** This tool is used to submit the website to search engine directories and find websites to exchange links. Import link, outbound links, internal link are some of the examples.

VI. SEO TECHNIQUES

The following are the three techniques used by search engine optimization.

- **Keyword Generation**

Every search engine needs keyword to elaborate information based on the given keyword. Hence keywords places vital role in SEO. The selection of keyword is based on the subject what the user is going to search. In a short period of time, with the help of the selected keyword searching rank of website will be high [13]. The process is progressed using different online tools such as yahoo keyword selector tool, Google Ad words, word tracker and Thesaurus etc. Using such tools, it is enough to type one word related with website. However supplying such keywords alone will not assure the popularity of website.

- **Directory Submission**

This is one of the most important SEO techniques for creating incoming links to a website through related page and category [12]. It creates incoming links to a website through indexed page. Any website needs to be ranked to get excellent business results. The best approach for ranking the website is the manual submission to directories. In the web, Internet directory acts as a platform for linking with other websites. Free services will be provided by many directories for any website in the directory. Some of the following information related with website is required for directory submission. URL, description, title and keyword, email and category are such information.

- **Link Exchanges**

To start business in the internet, reciprocal link or link exchange is needed for any web site to connect with any other websites. Link exchange is the SEO method to create link with other websites and vice versa. Some of the link exchanges are like: Exemplify hobbies on web pages directly or send email regarding hobbies for link exchange. Related websites with good page rank are only needed to build reciprocal links. Website title, keywords, description,

URL and email address are essential to generate reciprocal links. Link exchange is essential for boosting website traffic from user who clicks on the website link. It also increases ranking in web search engine. However sharing of all information is the main drawback of link exchange.

VII. DIFFERENT SEARCHING APPROACHES

Based on the way of searching and retrieving the information, the search engine is classified as follows.

- **Keyword-based search:** Google & Yahoo are some of the search engines that use keyword based search. Such searching uses specific keyword to find the documents that are matched with the given keyword.
- **Random Surfing:** It follows web linkage pointers for searching the relevant document.
- **Querying deep Web sources:** This method hides data behind searchable data base query forms that could not be accessed through static URL links.

VIII. CONCLUSION

This paper discussed an efficient data mining approach for Search Engine Optimization for improving the ranking of the website much easier. It also described different Search Engine Optimization tools. In addition this paper discussed about three essential SEO techniques such as keyword generation, directory submission and link exchanges and how they are helpful for displaying the relevant web pages based on ranking.

REFERENCE

- [1] Jiawei Han Kevin and Chen-Chuan Chang, 2002, "Data mining for web intelligence", IEEE Computers, Vol. 35, PP. 64-70.
- [2] "Data Mining Curriculum", ACM SIGKDD 2006-04-30.
- [3] Palvi Arora and Tarun Bhalla, 2014, "A Synonym Based Approach of Data Mining in Search Engine Optimization", International Journal of Computer Trends and Technology (IJCTT), Vol. 12, No. 4.
- [4] Bo Xing, Zhangxi Lin, 2006, "The Impact of Search Engine Optimization on Online Advertising Market", The Eight International Conferences on

- Electronic Commerce (ICEC'2006), pp. 519-529, ACM Electronic Commerce.
- [5] Rajesh Singh and S.k Gupta, 2013 , "Search engine optimization using Data mining approach", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Vol. 2, No. 9, PP. 28-32.
- [6] David Eaves. The ultimate guide to directory submissions, <http://www.webcredible.co.uk/user-friendly-resources/searchengine-optimisation/directory-submission.shtml>
- [7] SurajAnjaana, "The Importance of SEO Manual Directory Submission", http://www.ewebmarketing.com.au/seo/696_search_engine_optimisation_article.html.
- [8] Abhishek, V. and Hosanagar, K, 2007, "Keyword generation for search engine advertising using semantic similarity between terms", Proceedings of the Ninth international Conference on Electronic Commerce (ICEC'2007), PP. 89-94, ACM Electronic Commerce.
- [9] Allan Gardyne. How To Get Reciprocal Links, <http://www.associateprograms.com/articles/48/1/How-to-getreciprocal-links/>.
- [10] Solihinand Niko, 2013, "Search Engine Optimization: A Survey of Current Best Practices", Technical Library,PP. 151.
- [11] Mike Barus. Link Exchange and One Way Links Using Web Directories, <http://www.article99.com/internet-marketing/directorymarketing/article.php?art=21628>.
- [12] Ricky Mondal. Directory Submission – A Way Of Getting One Links, <http://www.article99.com/internet-marketing/directory-marketing/article.Php?art=63811>.
- [13] Terry Fung, "How to Generate Keywords for PPC Advertising", <http://www.howtodothings.com/business/a1693-how-to-generatekeywords-for-ppc-advertising.html>.