International Journal of Integrated Engineering:
Special Issue 2018: Data Information Engineering, Vol. 10 No. 6 (2018) p. 19-25.
 DOI: https://doi.org/10.30880/ijie.2018.10.06.003

Search Engine Optimization Algorithms for Page Ranking: Comparative Study

Arif Ullah[1], Nazri Mohd Nawi[2*], Edi Sutoyo[3], Asim Shazad[4], Sundas Naqeeb Khan[5], Muhammad Aamir[6]

[1,2,4,5,6]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Johor, Malaysia

[3]School of Industrial Engineering, Telkom University, 40257 Bandung, West Java, Indonesia

Received 28 June 2018; accepted 5 August 2018, available online 24 August 2018

Abstract: No doubt that every second in our daily routine, the number of visitors that connect to the internet increase day by day due to the fast growing of World Wide Web. Until this day there are more than 11.3 billion web pages in the World Wide Web. In the modern era of technology and advance computation, world page ranking is becoming a common feature of modern retrieval system. However, many of us did not realise that any query in search engine will display both relevant and irrelevant data that can cause overhead to the search engine and will affect the page ranking process. Therefore, a new optimization technique to improve the existing search engine optimization in increasing the page ranking is needed. This paper presents a review and comparative study of different existing page ranking algorithms for search engine optimization. This paper also explores some improvements that are needed to overcome the current problem in this field by testing on some real case data. The simulation result's analysis clearly shows that there is a need for new optimization technique which can reduce the complexity and user overhead by displaying only related data which will reduce over heading in search engine.

Keywords: Search, Engine, Offline Page Ranking, On Line Page Ranking, Optimization Technique

1. Introduction

Nowadays, majority of the web traffic is driven by commercial search engines. Search engine is define when a program is used for searching for any type of documents or files using specific keyword, then the search results return a list of documents based on the key words. Furthermore, there is also a definition that defines search engine as a computer program that based on word in finding document from the data base or World Wide Web. Nowadays there are various kinds of search engines that are viable on internet with their own techniques and specialties such as Google, Bing and yahoo. The first search engine known as veronica is text base search engine. Whereas, FTP (File Transfer Protocol) files also been used as search engine [1]. World Wide Web (WWW) is scattered different information pages resource and it consists of data and hyperlink. Recently, with the growth of WWW, it becomes difficult to find the required information that matches the user demand and interest.

Even though, search engines are smart, but they still need help in term of getting the correct materials as requested by users. Therefore, major engines are always working to improve their technology to crawl the web more deeply in order to return better results to users. However, there are still some limitations that bee identified particularly on how search engines can operate because, the right SEO can help user to thousands of visitors and increased attention.Whereas, selecting wrong strategies can hide or bury the user site deep in the search results where visibility is minimal.

In addition, the process of making content available to search engines, SEO also helps boost rankings so that content will be placed where searchers will more readily find it. The Internet is becoming increasingly competitive, and those companies who perform SEO will have a very good advantage in visitors and customers..Figure 1 shows the different type of search engine that are available nowadays.

Moreover, there are hundred and millions of linked page that form the result which have been retrieved, organized and presented by search engine. This result pages may have some of the information that may not be useful and irrelevant for the user. Therefore, for this scenario web page ranking algorithms play an important role to optimize this kind of result [2].

Web search becoming more and more popular search site in search engine. In this scenario it is necessary to provide proper and well reverent document and quality information to the use when it request. And it is main challenge to overcome the problems research take help of algorithms and there are different kinds of algorithms are design and modified to achieved the goal [3].

*Corresponding author: nazri@uthm.edu.my

Fig.1: Type of search engines

Furthermore, there are different kinds of algorithm that can be used for this process. Some of these algorithms are: (a) HITS, (b) PageRank, and (c) weighted PageRank generic algorithm, (d) Particle Swarm Optimization, (e) distance rank algorithm and (f) Ontology Page Ranking algorithms. Moreover, the web is the most important resource of information with different level. Therefore, getting its information from different level of web required a different background. Normally, the used of information by users is recoded in web page logs and process of analysis web log files [4]. Figure 2 shows the different steps of search document from web.

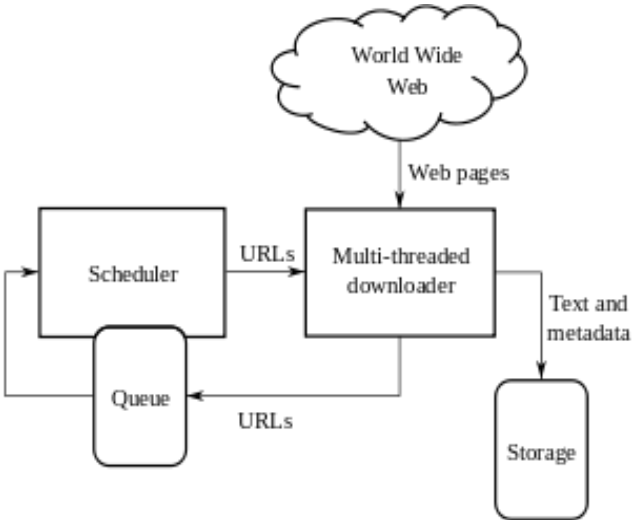

Fig. 2: Searching process.

There are several components that are needed for web search engine[5]. However, the main components are given as illustrate in Figure 3 which shows the three main component of web. According to this step the process must include:

- User Interface
- Parser
- Web Crawler
- Database
- Ranking Engine

User interface: is the part of web search engine which interacting with the user and allowing them to do query and view the result of query.

Parser: These components of web search engine provide term or key word extraction for both sides. Parser determines the key word which used in the queries in web document and they have been scanning by the crawler. Moreover, parser also used other term such as Tokenization, Normalization, Stemming Stop word handling.

Web Crawler: This component is simple and automated programs or script that scans or crawls through the internet pages for creating an index of the data. Alternatively, there are also other names for a web crawler including web spider, web robot, bot, crawler, and automatic indexer. When a spider visits a web page it then read text, hyperlink and content of various tag such as key word and Meta tags [6].

Database: Lastly, the website is included in the search engine's database and its page ranking process. Role of data base in web raking is important because it used for different purpose like page store text storage and different kind of query are their which help to ranking. Different kind of data base is used in ranking like user and system data base [7].

Fig. 3: There main components of web search

In 2006, Kennedy [8] had made a comparison between two algorithms which used Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) algorithm in improving the quality of service in web. According to the simulation results, Particle Swarm Optimization (PSO) gave very best result in both condition single and multi-use service.

However, due to dynamic change in text and structure of web page in WWW some additional challenges had been made for classification of web text and its structure. Furthermore, due to interconnected nature of hypertext it also provided future opportunities for researchers in which can assist the process. Therefore, Sk-Nearest Neighbor and Support Vector Machine (SVM) are used for the classification of text and web page [9]. According to Kritzinger's findings, the distance rank algorithm is intelligent algorithms which used for measurement of distance between different pages in web and it measure on the basis of distance between two pages and rank them. The limitation of this algorithm is that the

crawler tends to visit more and the calculation takes more time.

Mekky and Atwan [10] in 2015had proposed a ranking algorithm and suggested an algorithm that consists of user feedback. In addition, the use of feedback was measured by using suggested items. They found out that the algorithms that used query such as Ontology Page Ranking Algorithms had become famous algorithms since it provide an efficient way to reduce some irrelevant information from www. The algorithm defined a specific way of information in specific domain which help user to get their requested information easily using this algorithms.

Later, the similarities base algorithm is used to measure the similarities between queries using model. The algorithms known as similarity measure algorithm where in each single model for ranking each page they are compared with corresponding document and been rank using that model. Ranking model in this algorithm is the combination of various models of similar training queries [11].

Berkhin, P. [12] in 2005 had designed and modified further those page rank algorithm by giving weight. According to his work, the weights give more value for important pages rather than diving value among its outgoing linked pages. Furthermore, each outgoing link will get value and been proposed as an important. The results showed that the proposed method work well than page rank algorithm. In e–ranking algorithm, the result was arranged from searching by considering each rank link. In this method vector space model is used for document representation. Moreover, Co-sign were used for similarity function to get matched term related particular query from web log and a few level of semi structured data. Then the sign used for two vectors query and link terms were calculated using equation [13].

Recently, there are new trends of research to integrate or hybrid some optimization techniques with meta-heuristic techniques to solve some complex problems. This is because the hybridization will introduced more efficient algorithms which overcome each other limitations. In 2105, Chiroma et. al. [14] had improved the effectiveness of the ANN by hybridized the cuckoo search algorithm with accelerated particle swarm optimization for training the ANN to build a model for the prediction of OPEC CO_2 emissions. The proposed model predicts OPEC CO_2 emissions for 3, 6, 9, 12 and 16 years with an improved accuracy and speed over the state-of-the-art methods.

This follow by Nawi et. al. [15] in 2014 where a hybrid technique of Accelerated Particle Swarm Optimization was proposed by using Levenberg Marquardt known as (APSO_LM) had achieved faster convergence rate and to avoid local minima problem. In addition, the proposed techniques provided faster training for solving pattern recognition problems using the numerical optimization technique The results clearly showed that the hybridization had significantly improve the current optimization technique and had inspired many researchers to hybrid other optimization techniques in improving search engine optimization.

2. Numerical Model

Search Engine Optimization (SEO) is a technique that optimizes web pages or whole website in making the search engine friendly and helping user to get higher rank in result. In other words, SEO will make web pages more visible for user when they search with the help of search engine. Search engine is one of the main important methods for improving the visibility of the website or web page and also a very important method that can improve the search result [16].

According to the SEO, all the web pages will appear on top when users search them because due to more visitors. Furthermore, SEO may search different kind of data such as image, video, academic news, and other thing. Search Engine Optimization (SEO) is an internet marketing policy that used widely for improving the volume and quality of customer traffic to a company through search engine. At present there are many popular search engines in use and include Google, Bing, Yahoo [17].

3.1 Search Engine Optimization Flavours

Recently, there three popular flavour of search engine optimization such as: (a) white hat SEO, (b) Gray Hat SEO and (c) Black Hat SEO. In White Hat SEO, the technique is known for good guy where SEO is the best way of search optimization and its support all type of search engine and it used nature way to update the result. This is because it used nature way of technique to update. If the website is regularly updated with quality and unique content, it gets better links from relevant niche websites and blogs. This means that the webmaster does not take a single attempt to mislead search engine and does not try to cheat [18].

Meanwhile, for Gray Hat SEO, this type of technique needs the webmaster to use some irrelevant technique for their web site. Simply the web master may buy or exchange links with other website in order to get better result. However, this type of flavor is not acceptable by search engine and that is why this type of technique is not used for long process in ranking.

Last but not least, the Black Hat SEO where in this type of search engine optimization the web master is doing spamming in the link or making irrelevant link with some irrelevant niche website. This type of technique is dangerous for long term period because it regularly update in search engine algorithm crawler and will easily know and identify the spammed technique that been used to optimize the websites [19].

The size of the WWW is rapidly growing day by day and at the same time the number of request or queries handled by SEO has also grown incredibly. Due to the increasing in number of users on the web, the request on web search engine also growing exponentially. In addition, the search engine need to process according to that result. Therefore the search engines have applied different techniques in order to extract only relevant documents from the database which can provide intended information to the users. Moreover, to present the documents in an ordered manner, web page ranking

methods are applied which can arrange the documents in order based on their relevance, importance and content score and use web mining techniques to order them [20]. Figure 4 shows the different types of page ranking algorithms.

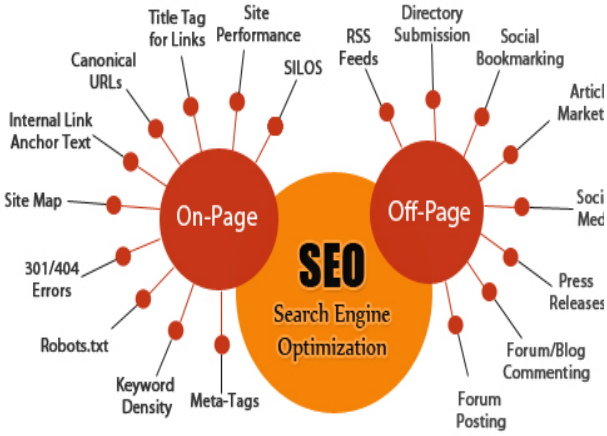

Fig. 4: Type of Page Ranking

Search engine optimizations are divided into two parts which are: (a) Internal SEO or On-Page Search Engine Optimization, and (b) External SEO or Off-Page Search Engine Optimization.

On Page SEO: it refers to the settings that apply on the website which optimized for search engines. On page SEO is one of the popular optimization which implement on web site and the most important elements are given below:

- Page Titles
- Meta Descriptions
- Meta Tags
- URL Structure
- Body Tags (H1, H2, H3, H4, etc.)
- Keyword Density.
- Image SEO.
- Internal Linking
- Google Authorship verification for all pages
- Top quality fresh content

Off page SEO search engine optimization is known as link building. In off page search engine optimization there are two type of linking first one is called incoming link and these are trust link because it depends from where the links are coming [21].

The second are negative link because they are paid. These off site Off-page SEO used type of Off-page SEO such as Social Networking Sites also known as "Online Reputation Management" which contain of:

- Blogging
- Blog Marketing
- Forum Marketing
- Search Engine Submission
- Directory Submission
- Social Bookmarking
- Link Baiting
- Photo Sharing
- Video Marketing
- Business Reviews
- Local Listings
- Article Submission
- Social Shopping Network
- Answer Questions

4 Results and Discussion

All the algorithms that described above are effective for page ranking by using search engine optimization. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) algorithm used for improving the quality of service in web. Distance rank algorithm is an intelligent algorithm which was used for measurement of distance between different pages in web and it measure on the basis of distance between two pages and rank them. Suggestive algorithm it consist of user feedback where they used feedback measure by using suggested items. Some algorithms also used query technique which used social tagging system according to the suggestion that make the page raking. Algorithms such as Ontology Page Ranking provided more efficient way in reducing some irrelevant information from WWW.

It defines a specific way of information in specific domain which helps the user to get their information easily using this algorithm. Whereas, e–ranking algorithm known for future arrangement the result from searching by considering each rank link. In this method vector space model is used for document representation. Last but not least, Co-sign is used for similarity function to get matched term related particularly query from web log and some level semi structured data. Thus in order to broaden down the research on those identified approaches of ranking algorithms, Table 1 provides the comparative analysis between the different page ranking algorithms on various attributes and parameters.

Table 1: A comparison of algorithms

Algorithm name	Author	Main technique	Result	Year
Genetic algorithm	Palanikkumar & Kouslary	QOS single user	Good	2012
Particle swarm optimization	Palanikkumar & Kouslary	QOS Multi user	Normal	2012
Distance rank algorithm	Kumar	Measurement of distance	Good	2010
Ranking and suggestive algorithm	M.vojnovic	Feedback	Good	2014
Ontology page ranking algorithms	Weng and shie	Similarity	Good	2009
Weighted PageRank according	Xing	Value for important page	Normal	2013
Re-ranking	Ratheash	Vector	Normal	2013

algorithm	machine

Table 2: Page ranking techniques

Algorithm name	Offline paging	Online paging
Genetic algorithm	Page Titles Meta Descriptions	Blogging Blog Marketing
Particle swarm optimization	Meta Tags URL Structure	Directory Submission
Distance rank algorithm	URL Structure	URL Structure
Ranking and suggestive algorithm	Image SEO. Internal Linking	Link Baiting
Ontology page ranking algorithms	Link Baiting	URL Structure
Weighted PageRank according	Business Reviews	Keyword Density
Re-ranking algorithm	Business Reviews	Keyword Density

Table 2 shows the different techniques that had been used along with algorithms by researcher in their research. Furthermore, the parameter performances were checked by using different ranking techniques as can be seen in Figure 5 until Figure 8.

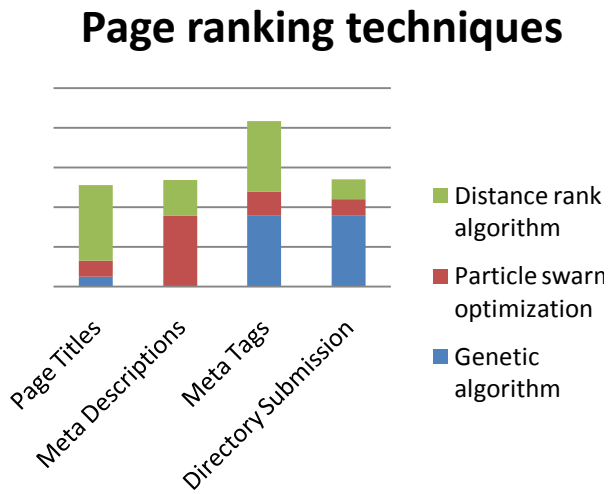

Fig 5: Page ranking element

In Figure 5 shows the result of different algorithms which used different parameter of page ranking such as page title, Metadescriptions, and directory. These elements are applied different algorithms by different scholar. As we can see in the figure that all algorithms work properly in Meta tags as well as directory submission.

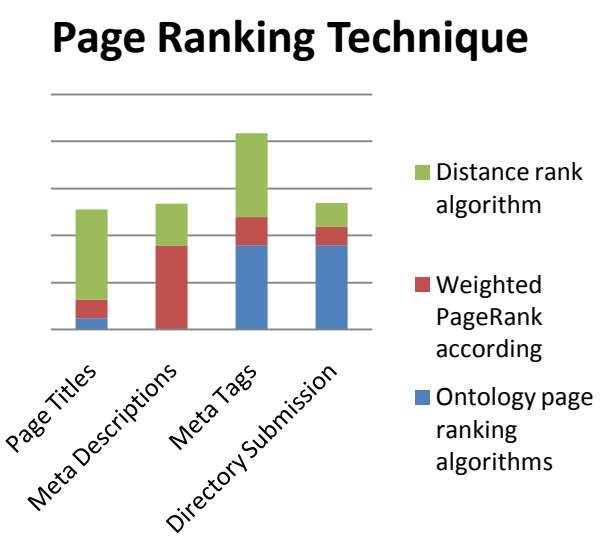

Fig 6: Page ranking element.

Figure 6 shows the result of different alorgithms in page ranking used different elememt. Based on four criterias, the result demonstrates that all algorithms are working properly in meta tags. As in directory submission it only work on ontology ranking algorithms but not on others.

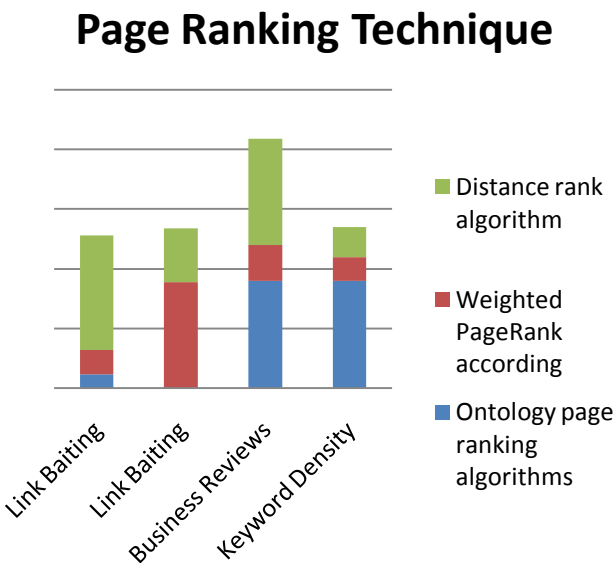

Fig 7: Page ranking element

According to the Figure 7 the distance rank and weighted page rank alorithms used different page ranking element. The result business element of page working properly .As compaire to othere element so business review is more compatable with all algorihms.

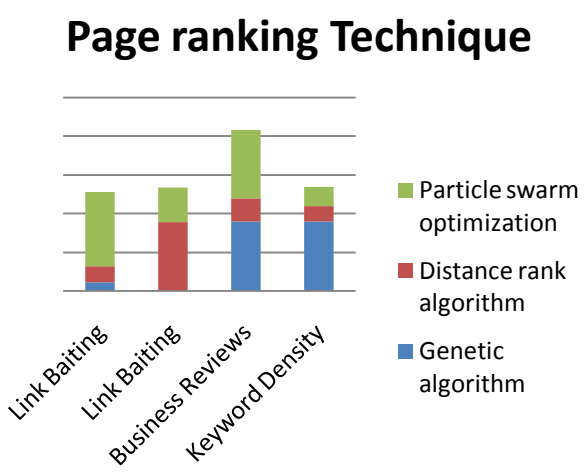

Fig 8: Page ranking element.

According to Figure 8 there are different element of off-page ranking are checked with different algorithm and according to their working business review is working more properly with the three algorithms as compaire to othier element .

5 Summary

Page ranking is a one of the crucial area in information retrieval from domain. Developing an effective searching engine in any domain needs a proper selection of good type of algorithms, technique, an effective parameter and characteristics. All the systems can retrieve relevant results from document. However, one cannot rank the different ranking algorithms in terms of performance as different approaches of page ranking algorithms are suitable for different applications. Therefore this paper is just a comparative analysis that reviews various researches about page ranking algorithm.The analysis results will help organisation in deciding the proper algorithm that can be used in search engine optimization. Over all we selected six algorithms which are used for page ranking along with different parameter .According to result algorithm's working properly with some element some element missing it means that we need different algorithm's for offline page ranking and online page ranking. In future we are planning to discuss the short coming of these algorithms for different type of element in both online and offline page ranking.

Acknowledgement

The authors would like to thank Universiti Tun Hussein Onn Malaysia (UTHM) and Ministry of Higher Education (MOHE) Malaysia for financially supporting this Research under IGSP grants note U420 and under Trans-displinary Research Grant Scheme (TRGS) vote no. T003.

References

[1] Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. Computer networks, 56(18), 3825-3833.

[2] Baeza-Yates, R., Castillo, C., Junqueira, F., Plachouras, V., & Silvestri,

[3] F. (2007, April). Challenges on distributed web retrieval. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 6-20). IEEE.

[4] Sherman, C., & Price, G. (2001). The invisible Web: Uncovering information sources search engines can't see. Information Today, Inc..

[5] Alguliev, R. M., Aliguliyev, R. M., & Mehdiyev, C. A. (2013). An optimization approach to automatic generic document summarization. Computational Intelligence, 29(1), 129-155.

[6] Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177). ACM.

[7] Guy, I. (2016, July). Searching by talking: Analysis of voice queries on mobile web search. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 35-44). ACM.

[8] Kennedy, J. (2006). Swarm intelligence. In Handbook of nature-inspired and innovative computing (pp. 187-219). Springer US.

[9] Kritzinger, W. T. (2005). The effect webpage body keywords location has on ranking in search engines results: an empirical study (Doctoral dissertation, Cape Peninsula University of Technology).

[10] El-gayar, M. M., Mekky, N., & Atwan, A. (2015). Efficient proposed framework for semantic search engine using new semantic ranking algorithm. International Journal of Advanced Computer Science and Applications, 6(8).

[11] Dennis, J. (2016). Search Engine Optimization and the Long Tail of Web Search.

[12] Berkhin, P. (2005). A survey on pagerank computing. Internet Mathematics, 2(1), 73-120.

[13] Jerkovic, J. I. (2009). SEO warrior: essential techniques for increasing web visibility. " O'Reilly Media, Inc.".

[14] Chiroma, H., Abdul Kareem, S., Khan, A., Nawi, N.M., YaU Gital, A., Shuib, L., Abu Bakar, A. I., Rahman, M. Z., Herawan, T.(2015). Global warming: Predicting OPEC carbon dioxide emissions from petroleum consumption using neural network and hybrid cuckoo search algorithm. PloS One, 10(8).

[15] Nawi, N.M., Khan, A., Rehman, M. Z., Aziz, M. A., Abawajy, J. H., Herawat, T. (2014). An accelerated particle swarm optimization based Levenberg marquardt back propagation algorithm. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)Volume 8835, 2014, Pages 245-253.

[16] Fishkin, R., & Høgenhaven, T. (2013). Inbound marketing and SEO: Insights from the Moz Blog. John Wiley & Sons.

[17] Bailyn, E. (2013). SEO Made Easy: Everything You Need to Know about SEO and Nothing More. Que Publishing.

[18] Gudivada, V. N., Raghavan, V. V., Grosky, W. I., & Kasanagottu, R. (1997). Information retrieval on the world wide web. IEEE Internet Computing, 1(5), 58-68.

[19] Jerkovic, J. I. (2009). SEO warrior: essential techniques for increasing web visibility. " O'Reilly Media, Inc.".

[20] Clay, Bruce, and Susan Esparza. Search engine optimization all-in-one for dummies. John Wiley & Sons, 2011.

[21] Sharma, Pratibha, et al. "Information Retrieval in Web Crawling Using Population Based, and Local Search Based Meta-heuristics: A Review." Proceedings of Sixth International Conference on Soft Computing for Problem Solving. Springer, Singapore, 2012