Scientific Research

Website Search Engine Optimization: Geographical and Cultural Point of View

Osama Rababah[1], Muhannad Al-Shboul[2], Fawaz Al-Zaghoul[3], Rawan Ghnemat[4]

[1]Department of Business Information Technology, The University of Jordan, Amman, Jordan
[2]Department of Curriculum and Instruction, The University of Jordan, Amman, Jordan
[3]Department of Computer Information Systems, The University of Jordan, Amman, Jordan
[4]Department of Computer Science, Princess Sumaya University for Technology, Amman, Jordan
Email: o.Rababah@ju.edu.jo, malshboul@ju.edu.jo, fawaz@ju.edu.jo, r.ghnemat@psut.edu.jo

Abstract

The concept of Webpage visibility is usually linked to search engine optimization (SEO), and it is based on global in-link metric [1]. SEO is the process of designing Webpages to optimize its potential to rank high on search engines, preferably on the first page of the results page. The purpose of this research study is to analyze the influence of local geographical area, in terms of cultural values, and the effect of local society keywords in increasing Website visibility. Websites were analyzed by accessing the source code of their homepages through Google Chrome browser. Statistical analysis methods were selected to assess and analyze the results of the SEO and search engine visibility (SEV). The results obtained suggest that the development of Web indicators to be included should consider a local idea of visibility, and consider a certain geographical context. The geographical region that the researchers are considering in this research is the Hashemite kingdom of Jordan (HKJ). The results obtained also suggest that the use of social culture keywords leads to increase the Website visibility in search engines as well as localizes the search area such as google.jo, which localizes the search for HKJ.

1. Introduction

In Web search, given a query, a search engine returns the matched documents in a ranked list to meet the user's

information need. Ranking models play a central role in search engines [2]-[4]. Currently, almost all the existing Website SEO models consider only the current query and the Websites, but do not take into account any geographical or cultural information such as the previous queries in that area and local society keywords. In other words, almost all current ranking models are insensitive to the culture of the country.

Information retrieval research has well recognized that user information is very helpful in achieving good search results. People and cultural information may provide hints about the user's intent of Web search and help to make better matching to increase the visibility of Websites [5]-[7]. For example, if a user writes "google.com" in HKJ, Google redirects the search to the domain "google.jo" to provide local search. Any absence of people's preference and cultural information in Website may affect Website visibility.

When a user queries certain keywords or phrases, a search engine will typically return results in two forms: organic search results and paid listings.

Organic search results are the natural listings that are suggested by search algorithms. Search engines use complex algorithms to determine relevant pages and suggest them to the users as per their request. However, Websites have to be crawled by search engine spiders, before they can be displayed as organic search results. Crawling services are mostly offered free of cost by the search engines [8].

Whereas paid listings are short contextual text advertisements within search results screens, with links to the advertiser's Website; advertisers compete with one another and bid for keyword sponsorship. Text advertisement (consisting of uniform resource locator (URL) and short description) of the highest bidder will appear on the top of the paid listing results when sponsored keywords are queried. The text advertisement of other bidders (second highest, third highest, etc.) are displayed in the descending order on the same column under the advertisement of the highest bidder. Paid listings are normally displayed above or on the right hand column of the organic search results [9].

There are various SEO guidelines available on the Web. The top three players in the industry (Google, Microsoft Live Search, and Yahoo) have developed guidelines focusing on their own search engines. All three guidelines are designed to improve Websites SEO. The main themes of SEO are similar in various search engine providers. For example, they all emphasize on uniqueness of content, meta-data (title and description), navigation, and human centric Web design. Among the three search engines, Google is the dominant player of the search engine providers [10]. Moreover, it is fast and reliable, works on massive scale, and produces useful results much of the time; it searches the full texts of documents, and indexes multiple documents types including HTML, PDF, Word Documents. Therefore, this study focuses on increasing Websites SEO in local Google search engine [10].

Most existing studies on SEO have focused on usability. Usability is a broad scientific discipline based on rigorous application of the observation, measurement, and principles of design that are useful for the creation and maintenance of Websites [11]. People access Websites through search engines or directly by typing URL in the browser address bar. However, the majority use search engines; they do not usually browse beyond the first ten or twenty items displayed on the first page of any search results. This shows that the Websites displayed on the first page of search results have the best chance of being visited. Therefore, it is important for any Website developer to put emphasis on SEV while developing the Website. SEV is especially important to promote any Website or product [12]. This research study provides a basic strategy for SEO and SEV in a local geographical area and it demonstrates the effect of people culture and behavior on that area as classified by search engine providers.

The rest of this paper is structured as follows: Section 2 introduces a literature review; Section 3 provides description and motivation of methods used; Section 4 presents discussion and analysis of results; Section 5 summarizes the conclusion of the study; finally, Section 6 describes the future work.

2. Literature Review

According to Zhang and Dimitroff [13], search engine optimization (SEO) is the process of identifying factors in a webpage which would impact search engine accessibility to it and fine-tuning the many elements of a website so it can achieve the highest possible visibility when a search engine responds to a relevant query. Thus, SEO aims at achieving good search engine accessibility for Webpages, high visibility in a search engine results, and improvement of the chances the Webpages are retrieved.

The basic understanding of search engine optimization starts with understanding how a search engine works.

There are three basic types of Web search engines; crawler-based search engines, human powered directories, and hybrid search engines or mixed results. SEO only applies to the crawler-based search engine, which is what the larger search engines (Google and Yahoo) are using [14].

Several optimizing SEO techniques have been researched and identified by [15]-[17]. Some of these include providing keyword rich Website content throughout all Webpages [10] [12] [18] [19], using keywords in the title and description meta tags in the Website code [9], and keeping the graphics, forms and frames to a minimum [8] [10] [11] [18] [19]. The researchers [15] [17] recommended a link development strategy whereby other good quality and relevant Websites are used to develop or implement a link back to the intended Website [12]. According to the researchers' knowledge, there is no clear strategy for analyzing effect of the geographical area and people culture on Website ranking. In this paper, the researchers analyzed the influence of geographical area (Jordan) and a local culture on Website search engine ranking, and identify the effect and the relationship of the local society keywords in increasing Website ranking.

3. Description and Motivation of Methods

Research study, human analysis, and statistical analysis methods were primarily used to conduct this study. The aim of this research is to determine the suitable keywords for Website visibility used on search engines by Jordanian people. Although, there are many tools available to conduct SEO analysis, only few of them are available for free of charge [20]. The ones that are free did not deliver adequate data required to conduct SEO analysis. Therefore, the researchers of this paper analyzed the Websites by accessing the source code of its homepage through Google Chrome browser. The statistical analysis methods were selected to assess and analyses the results of the SEO and SEV in Jordan.

4. Discussion and Analysis of Results

The proposed strategy for analyzing effect of the geographical area and people culture on Website ranking is based on search engine history of keyword weight analysis which is found on the list of the same geographical area. The main steps for the proposed strategy are summarized in **Figure 1**.

4.1. Keywords Identification and Gathering

Keywords are represented by a list of terms and weights associated with those terms; keywords were collected by taking the first 10 terms that appear in Google suggested history list for each letter of the alphabet in both Arabic and English in Jordan, for a period of three months per different sector. People in HKJ usually type keyword phrases while gathering information suitable to the way they think and their interests through search engines. The keywords were collected with the help of many keyword search analyzing tools [21].
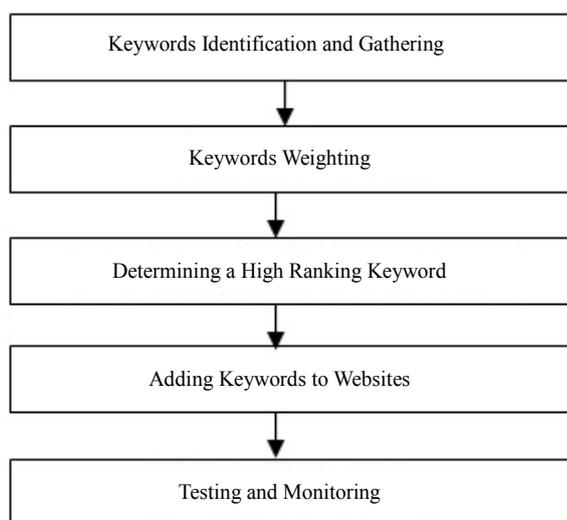

Figure 1. Proposed strategy steps.

4.2. Keyword Weighting

After collecting the keywords from the suggested search query history lists (google.jo); the weights of each term were computed by using Term Frequency-Inverse Document Frequency (TF-IDF). After that, the keywords with the highest occurrence for all the English and Arabic letters were saved. This gave us twenty six keywords in English language and twenty eight keywords for Arabic language.

4.3. TF-IDF Weighting

The Term Frequency-Inverse Document Frequency (TF-IDF) is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus [21]. It is often used as a weighting factor in information retrieval and text mining. The TF-IDF value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus, which gives us an indicator about the words that are generally more common than others [22].

Variations of the TF-IDF weighting scheme are often used by search engines as a central tool in scoring and ranking Websites that are relevance to a user query. TF-IDF can be successfully used for stop-words filtering in various subject fields including text summarization and classification. Term weights may be assigned for different reasons. We considered a weighting in a search engine history list which obtained by the user in his/her local area [23].

To obtain TF-IDF estimates for each letter, the researchers of this study use the inverse frequency of the terms that appear in the search history list for all alphabetic (Arabic and English) using Google search engine. However, in the case of google.jo, we weighted the history list keywords by TF-IDF as described in Equation (1) below [22] [24].

$$TF - IDF : (t, d, D) = tf(t, d) * idf(t, D) \tag{1}$$

where:
- $tf(t, d) = 0.5 + \left(0.5 * f(t, d)\right) / \left(\max(fw, d) : w \in d\right)$.
- t is the number of occurrence of term t in d.
- d is the number of documents that saved from search history lists.
- $idf(t, D) = \log\left(N / \left(1 + \left|\{d \in D : t \in d\}\right|\right)\right)$.
- N is the total number of documents gathered.
- $\left|\{d \in D : t \in d\}\right|$ represents the number of documents where the term t appear.

A keyword or term would have the same weight in different requests for the same letter by different users and different browsers. When relevant search are taken into account, the same term may have a different value for different requests for different letters by different or same users. An obvious weighting function, derived from different starting points, has been proposed by Barkla [25] and by Miller [26] [27], The researchers of this paper have enhanced the function proposed by Barkla [25] and by Miller [26] [27] to calculate the keyword/term weight in search engine history list as described in Equation (2).

$$Weight = \log \frac{\left(\dfrac{R}{r}\right)}{N} \tag{2}$$

where:
- R is the number of keywords for every letter.
- r is the number of occurrence of keywords for different letters.
- N is the number of chosen letters.

4.4. Ranking Methodology

Many researchers used the user profile to rank the top results returned by a search engine to bring up results that are more relevant to the user [28]. In this study, the researchers used the search engine (Google) profile for each term rather than user profiles because Google history list profile gave better results for ranking than user profile. Google history list is based on Google algorithms search history. This allows us to take advantage of the search

engines data to obtain the ranking results, by taking all the terms appear for all the alphabetical letters according to the geographical location.

After calculating the keywords weighting over the three months and recording the results, the researchers of this study have found that some keywords got higher values than the others. So for each of the Arabic and English alphabets, the researchers of this study have chosen the keyword according to its highest rank as shown below in Equation (3) and Equation (4) respectively.

$$\text{Highest Ranked keyword} = \max\left(avgKw\right) \quad\quad (3)$$

$$avgKw = \sum\nolimits_{0}^{90} kw/90 \qu\quad (4)$$

where:
- kw is the weight for each keyword.
- $avgKw$ is the average kw weight for 90 days.

However, **Table 1** shows the English letters average word weight over three months, whereas **Table 2** shows the Arabic letters average word weight over three months.

Table 1. English letters average word weight over three months.

Letter	Keyword with Highest Ranking	Keyword Percent of Ranking over Three Months
A	Amman	92.325%
B	BBC Arabic	90.218%
C	Candy Crush	94.948%
D	Dubizzle	91.829%
E	Eduwave	97.837%
F	Facebook	98.452%
G	Google Translate	96.934%
H	Hashemite University	92.627%
I	Ibis Hotel Amman	89.348%
J	Jordan Weather	99.478%
K	Khaberni	95.358%
L	Linkedin	96.214%
M	Myegy	91.478%
N	nasa	97.215%
O	Orange Jordan	92.146%
P	Panet	93.875%
Q	Queen Alia Airport	97.957%
R	Royal Jordanian	91.986%
S	Saraya	94.298%
T	Translate	97.452%
U	University of Jordan	98.369%
V	Viber	92.258%
W	Weather in Jordan	96.147%
X	Xe Exchange	97.214%
Y	Youtube	93.962%
Z	Zain Jordan	98.943%

Table 2. Arabic letters average word weight over three months.

Letter	Keyword with Highest Ranking	Keyword Percent of Ranking over Three Months
أ	الجامعة الاردنية	95.325%
ب	برنامج البرنامج	94.368%
ت	ترجمة گوگل	97.214%
ث	ثقف نفسك فيس بوك	98.236%
ج	جيران	96.214%
ح	حالة الطقس في الاردن	97.369%
خ	خبرني	96.214%
د	دائرة الارصاد الجوية	95.258%
ذ	ذهب	92.354%
ر	رؤيا	93.214%
ز	زمان الوصل	94.279%
س	سعر الذهب في الاردن	92.354%
ش	شركة الكهرباء الاردنية	97.235%
ص	صباح الخير	95.246%
ض	ضرب المثل	93.214%
ط	طقس الاردن	92.358%
ظ	ظريف	95.214%
ع	عيد الحب	91.387%
غ	غنيلي تغنيلك	96.325%
ف	فيس بوك	97.224%
ق	قناة رؤيا	96.325%
ك	كورة	94.458%
ل	لعبة من سيربح المليون	95.349%
م	مترجم	98.235%
ن	نزار قباني	94.258%
ه	هوتميل	96.347%
و	وزارة التربية والتعليم	97.325%
ي	يوتيوب	98.268%

As shown in **Table 1** and **Table 2** above, the researchers selected keywords in English and Arabic for each letter alphabetically, respectively. Then, these selected keywords were tested by Google Chrome search engine over three months duration. After that, the researcher reported the ranking of these selected keywords along with their visibility using percentages representation as shown in these two indicated tables. So, the percentage representation for each letter either in English or in Arabic reflects the influence of the proposed SEO strategy on the SEV.

4.4. Monitoring and Testing

To test the proposed strategy of the study, the researchers have chosen 45 Websites from different sectors:

Health, Business, Real Estate, Education, Construction, Beauty, Restaurants, Tourism, and Agriculture. For each of the above mentioned sectors, the researchers follow the following process:

1) A Web search, for a given query was conducted.

2) From the returned matched Websites, the researchers have chosen Websites appeared in the 5th and 6th pages and store their ranks in alexa.com.

3) The keywords got highest rank has been added to the Meta description and Meta keywords for the respective Websites.

4) Keep monitoring by conducting further Web searches and store their new update in alexa.com ranking values.

The researchers have chosen five different Websites from different sectors, and applied the proposed strategy on them to test its effectiveness and functionality. The outcome of applying the proposed strategy shows significant increase in the visibility of those selected local Websites. Furthermore, after comparing the new ranks with those Websites picked earlier, a superior improvement occurs in the Websites visibility in search engines.

The proposed strategy is essential to the search engine optimization, because SEO is often about making small modifications to parts of Website. When viewed individually, these changes might seem like incremental improvements, but when combined with other optimizations, they could have a noticeable impact on Website's user experience and performance in organic search results.

5. Conclusions

In this paper, the researchers analyzed the influence of local geographical area and cultural values as well as the effect of local society keywords in increasing Website visibility. The researchers of the study conducted an empirical study on real search logs and developed a strategy to determine the suitable keywords for Website visibility based on search engine history keyword weight analysis that was found on the list of the same geographical area. The researchers further enhanced the function proposed by Barkla [25] and by Miller [26] to calculate the weight in the search engine history list. The experimental results verified the effectiveness of the proposed strategy to increase the Website visibility in search engines that localized the search area such as google.jo.

The results of the study revealed that the use of local society keywords enhanced local Websites visibility in search engines, which indicated the efficiency and effectiveness of our proposed strategy. Consequently, the whole study is giving a whole new dimension in understanding the influence of SEO proposed strategy on various Websites by using any search engine. In short, the researchers find that the proposed strategy regarding search engine optimization does affect the ranking of the Websites visibility. Hence, it can be concluded that the proposed strategy leads to the ranking or popularity of a Website in an improved manner. This is important because basic SEO is fundamental and essential since it can help individuals position their Websites properly to be found at the most critical points in the buying process or when people need their sites.

Last but not least, in this paper, the researchers describe the design and initial implementation of a geographic search engine prototype. Geographic search engines provide a flexible interface to the Web that allows users to constrain and order search results in an intuitive manner, by focusing a query on a particular geographic region or a certain language. Geographic search technology has recently received significant commercial interest, but there has been only a limited amount of academic work. Our prototype performs massive extraction of geographic and culture features. In addition, this study helps fill in the gap in the current Information Technology knowledge base regarding search engine optimization and search engine visibility.

6. Future Work

As a future work, it is recommended to conduct more studies on Website search engine optimization in the same geographical area and cultural values to validate the findings of this study. However, validation data from several sites would be required to give a more precise correction figure on SEO.

References

[1] Martinez-Torres, M. and Diaz-Fernandez, M. (2013) A Study of Global and Local Visibility as Web Indicators of Research Production. *Research Evaluation*, **22**, 157-168. http://dx.doi.org/10.1093/reseval/rvt003

[2] Kim, J.Y., Collins-Thompson, K., Bennett, P.N. and Dumais, S.T. (2012) Characterizing Web Content, User Interests,

and Search Behavior by Reading Level and Topic. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, Seattle, 8-12 February 2012, 213-222. http://dx.doi.org/10.1145/2124295.2124323

[3]   Dou, Z., Song, R. and Wen, J.R. (2007) A Large-Scale Evaluation and Analysis of Personalized Search Strategies. *Proceedings of the* 16*th International ACM Conference on World Wide Web*, Edmonton, 8-12 May 2007, 581-590. http://dx.doi.org/10.1145/1242572.1242651

[4]   Liu, S., Zou, Q. and Chu, W.W. (2004) Configurable Indexing and Ranking for XML Information Retrieval. *Proceedings of the* 27*th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sheffield, 25-29 July 2004, 88-95.

[5]   Radlinski, F., Kurup, M. and Joachims, T. (2008) How Does Click-through Data Reflect Retrieval Quality? *Proceedings of the* 17*th ACM Conference on Information and Knowledge Management*, Napa Valley, 26-30 October 2008, 43-52.

[6]   Sanderson, M. (2008) Ambiguous Queries: Test Collections Need More Sense. *Proceedings of the* 31*st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Singapore, 20-24 July 2008, 499-506.

[7]   Barry, C. and Charleton, D. (2009) In Search of Search Engine Marketing Strategy amongst SME's in Ireland. *Communications in Computer and Information Science*, **48**, 113-124. http://dx.doi.org/10.1007/978-3-642-05197-5_8

[8]   Digital River (2014) What is Organic Search? http://www.domsreport.com/www/developer-resource.com

[9]   Green, D.C. (2003) Search Engine Marketing: Why It Benefits Us All. *Business Information Review*, **20**, 195-202. http://dx.doi.org/10.1177/0266382103204005

[10]   Shah, B.P. (2009) Search Engine Visibility of National Tourism Promotion Websites: A Case of Nepal. *Proceedings of the* 3*rd International ACM Conference on Theory and Practice of Electronic Governance*, Bogotá, 10-13 November 2009, 287-292.

[11]   Pearrow, M. (2007) Web Usability Handbook. 2nd Edition, Charles River Media Inc., Boston.

[12]   Xiang, Z., Wöber, K. and Fesenmaier, D.R. (2008) Representation of the Online Tourism Domain in Search Engines. *Journal of Travel Research*, **47**, 137-150. http://dx.doi.org/10.1177/0047287508321193

[13]   Zhang, J. and Dimitroff, A. (2005) The Impact of Metadata Implementation on Webpage Visibility in Search Engine Results (Part II). *Information Processing & Management*, **41**, 691-715. http://dx.doi.org/10.1016/j.ipm.2003.12.002

[14]   Bifet, A., Castillo, C., Chirita, P.A. and Weber, I. (2005) An Analysis of Factors Used in Search Engine Ranking. *Proceedings of the* 14*th International World Wide Web Conference*, Chiba, 10-14 May 2005, 1-10. http://airweb.cse.lehigh.edu/2005/bifet.pdf

[15]   Joachims, T. (2002) Optimizing Search Engines Using Click-Through Data. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Alberta, 23-25 July 2002, 133-142. http://dx.doi.org/10.1145/775047.775067

[16]   Thomas, P. and Hawking, D. (2006) Evaluation by Comparing Result Sets in Context. *Proceedings of the* 15*th ACM International Conference on Information and Knowledge Management*, Arlington, 5-11 November 2006, 94-101.

[17]   Radlinski, F. and Craswell, N. (2010) Comparing the Sensitivity of Information Retrieval Metrics. *Proceedings of the* 33*rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Geneva, 19-23 July 2010, 667-674.

[18]   Brin, S. and Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, **30**, 107-117. http://dx.doi.org/10.1016/S0169-7552(98)00110-X

[19]   Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E. and Li, H. (2010) Context-Aware Ranking in Web Search. *Proceedings of the* 33*rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Geneva, 19-23 July 2010, 451-458.

[20]   Seidl, T. and Kriegel, H.P. (1998) Optimal Multi-Step K-Nearest Neighbor Search. *ACM SIGMOD Record*, **27**, 154-165. http://dx.doi.org/10.1145/276305.276319

[21]   Stockwell, J. (2011) Keyword Research Tools. http://www.keyworddiscovery.com/static/keyword-tool-reviews.pdf

[22]   Rajaraman, A. and Ullman, J.D. (2011) Mining of Massive Datasets. Cambridge University Press, Cambridge.

[23]   Schultz, J. and Fristedt, J. (2005) Calling All Search Engines. *Association Management International*, **25**, 8-13. http://www.associationmanagement.co.uk/AMI/Home/AMI/Home.aspx

[24]   Manning, C.D., Raghavan, P. and Schütze, H. (2008) Introduction to Information Retrieval. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511809071

[25]   Barkla, J.K. (1969) Construction of Weighted Term Profiles by Measuring Frequency and Specificity in Relevant

Items. *Proceedings the* 2*nd International Cranfield Conference on Mechanized Information Storage and Retrieval Systems*, Cranfield, 2-5 September 1969, 22-36.

[26] Miller, W.L. (1970) The Evaluation of Large Information Retrieval Systems with Application to Medlars. Ph.D. Thesis, Newcastle University, Newcastle.

[27] Miller, W.L. (1971) A Probabilistic Search Strategy for Medlars. *Journal of Documentation*, **27**, 254-266. http://dx.doi.org/10.1108/eb026520

[28] Teevan, J., Dumais, S.T. and Horvitz, E. (2005) Personalizing Search via Automated Analysis of Interests and Activities. *Proceedings of the* 28*th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, 15-19 August 2005, 449-456.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.