Analysis of Website Ranking in Search Engine using Multi-criteria Decision PROMETHEE Method: Implementation, Verification and Validation

Nita Solehati[1], Joonsoo Bae[1]* and Hyerim Bae[2]
[1]Department of Industrial and Information Systems Engineering, Chonbuk National University, Republic of Korea
nitasolehati@jbnu.ac.kr, jsbae@jbnu.ac.kr
[2]Department of Industrial Engineering, Pusan National University, South Korea
hrbae@pusan.ac.kr

Abstract. Over past decade, internet has been growing very fast and has become an integral part in modern business activity. As such, search engine - a tool to find online data, e.g., Google, Bing, Yahoo etc - has become the most popular website application among others. It is widely known that the highest rank website is the most popular one for which it is beneficial not only in terms of advertisement rate, but also can promote the website for business activity. Therefore, every company expects their website to appear in the top 10 or in the first page of search engine website; this fact makes web developers highly competitive to get into the highest rank on the search engine. In this paper, we explore and evaluate the ability of PROMETHEE method and its implementation on the customized Visual PROMETHEE software to index and rank some websites based on multi-criteria analysis. The Visual PROMETHEE's results are verified with analytical solution derived manually and validated against Google result. This method allows company to easily compare their website popularity among competitors' website in the search engine which can be used as a powerful tool to boost their marketing strategy.

Keywords: google; multi-criteria analysis; PROMETHEE; rank; search engines.

1. Introduction

Search engine machine, e.g. Google, Yahoo, Bing etc has become an indispensable tool in modern activity as nowadays people will usually find almost anything through internet. The fast growing of hardware, software and internet connection makes people to easily get access to the internet and, hence, extensively use search engine from their laptop, tablet and smart phone etc. Among many of the search engine available in the internet, Google has become the most popular search engine in the world followed by Bing, Yahoo, Ask and AOL [1]. Google market share accounts for around 65% in the search engine application, while Bing-powered (Bing and Yahoo!) accounts for 28.62%, and the remaining 66 search engines for 6% [2].

For instance, the highest rank website which comes out from the search engine is usually the most popular one. This is beneficial to the owner of the website as it can increase their income from advertisement and, most importantly, for the company or business related activity. Becoming the highest rank website in the search engine means that they can promote their

* Corresponding author

product better than their competitor. Therefore, every company expects their website to appear in the highest rank or at least at the top 10 or in the first page of search engine website. This fact makes web developers highly competitive to get into the highest rank on the search engine. In order to boost their popularity rank, many search engine optimization (SEO) are currently available for this purpose [3]. While some SEO methods are stay within the guidelines laid out by the major search engines; some others, however, are found to violate the guidelines and may risk to be penalized or to be banned from the search engines [4]. Thus web developer need to be extra careful when implementing SEO to their website to avoid penalized from the search engine. Furthermore, one of the main drawbacks of the SEO is that they cannot be used to predict the result of the search engines which is important on designing a website.

In our previous work [5], we analyzed the popularity of a website in search engine using PROMETHEE method, we have also shown that the result from our analytical solution of PROMETHEE method are in good agreement as compared to that of the results from search engines, i.e., Google and Yahoo, counterpart. To extend our work on the implementation of PROMETHEE for search engine application, the aim of the work presented herein is twofold: *(i)* Implementation of Visual PROMETHEE software for prediction of search engine websites popularity; and *(ii)* Verification and validation of Visual PROMETHEE software results. The former concerns on the accuracy of the solution from Visual PROMETHEE software, for which we compare the results from the Visual PROMETHEE software with one from our analytical solution of PROMETHEE method; while the latter focuses on the validity of the results from Visual PROMETHEE software as compared to the experimental evidence from search engine, i.e., Google, results. The verification and validation are rigorously analyzed using statistical approach by evaluating 30 keywords and 120 websites.

2. Methodology

PROMETHEE is a multicriteria decision making methodology which has been widely used to solve for several multicriteria decision applications ranging from environmental to computer science to stock market to medicine etc.; details of the various PROMETHEE application can be found in review paper by Behzadian et al [7]. Basically, PROMETHEE is an outranking multicriteria decision aid method [8,9]. The main principles upon which the PROMETHEE method based-on are extension of the notation of criteria, development of a valued outranking relation, and exploitation of the outranking relation. In the PROMETHEE, the method is implemented in six steps:

- *First step:* the first step involves selection of alternative determination. Here, we need to determine some of the alternatives that are already exist and later will be selected as the solutions. For verification and validation purposes, we take top four websites appearing from Google result as alternatives. In our study, we verify and validate 30 keywords and four alternative websites from each keyword which, overall, it ends up with 120 websites; the details of the all keywords and websites that we analyze can be found in Table Al in the Appendix. Note that the keywords used in this study are chosen randomly.
- *Second step:* we determine some criteria that are used in decision-making processes. In order to predict the search engine result - in this paper we limit to Google search - we determine the search criteria by Google which consist of five factors: keywords, link, content, design and address [5, 10]; for each factor, we give the value scale ranging from 1 to 5 depending on the analysis of the website, and then analyze five factors for all 120 sites one by one; all of the factors are listed in Table 1. Further, in factor number 3 and 4, i.e.,

content and design, it comprises of five and four subfactor, respectively. Each sub-factor is analyzed separately and the total scale is calculated by summing all the subfactors.

Table 1. Factors and subfactors used for evaluation of PROMETHEE website popularity ranking.

Factor	How the value is determined	Scale
Keyword $(f_1)_{max}$	Meta tag keyword in HTML code	1-5
Link $(f_2)_{max}$	*Toolbar check pagerank*	*1-5*
Content $(f_3)_{min}$	• How many duplication in the sites page	1-5
	• How relevant the keyword with the content	1-5
	• How many page that is attempted for phishing, Trojan, virus and other badware	1-5
	• How many pages that has affiliation program or non-original software	1-5
	• How many pages that is used as a cheating page	1-5
Design $(f_4)_{min}$	• How many *link scheme* it has	1 -5
	• How many hidden text or hidden link in the page	1 -5
	• How many wrong/dead HTML code	1 -5
	• How many link in the page	1 -5
Address$(f_5)_{max}$	How same website address chosen with the keyword	1 -5

- *Third step:* we assign the weight value (degrees-of-importance) for each criterion. This step is found to be an important step as all of the assigned weight need to be judged objectively. Thus, a clear and consistent standard need to be established in order to correctly assign the degree-of-importance weight of each factor and subfactors to avoid misjudgment. In Google, however, the weight value for the search algorithm becomes their top secret against competitors. Thus, in this study, we assume the weight value in the following order: keywords, link, address, content, design.

Table 2: Initial matrix formation

alternative f	- 1 (max)	f2(max)	f3(min)	f4(min)	f5(max)
al	f1(ai)	f2(a1)	f3(a1)	fa(al)	$f_5(a_i)$
a2	f1(a2)	f2(a2)	f3(a2)	f4(a2)	f5(a2)
a_3	f1(a3)	f2(a3)	$f_3(a_3)$	f4(a3)	$f_5(a_3)$
a4	f1(a4)	f2(a4)	f3(a4)	f4(a4)	$f_5(a_4)$

- *Fourth step:* we calculate and format the matrix according to the minimum and maximum values for each criterion according the given evaluation scales as illustrated in Table 2. The factor with maximum value means that the higher the scale, the better the performance of the site; while for the minimum factor, the site performance is higher as the scale is decreased.
- *Fifth step:* we search basic matrix and matrix transpose. For the maximum factor condition, the matrix is formed by dividing row over column; whereas for the minimum factor condition, it is in opposite way, column divided by row. The basic matrix is defined by

$$\text{basic matrix} = \text{initial matrix} / \sum_{n=1}^{/5=4} f, (a_n)$$

The basic matrix is then transposed. The *leaving flow, 0+,* is calculated from the average value of the basic matrix with the weight value; while the *entering flow,* **0",** is the average value of the transpose matrix with the weight value. The *net flow* is defined from the difference between the first two values.

- *Sixth step:* all the websites are ranked based on the net flow, **0+_ 0";** that is to calculate positive and negative preference flows for each alternative. *Leaving flow, 0+,* the positive flow expresses how much an alternative is *dominating* (power) the other ones. While *Entering flow,* **0",** the negative flow relates to how much it is *dominated* (weakness) by the other ones, after which, we calculate the *net flow.*

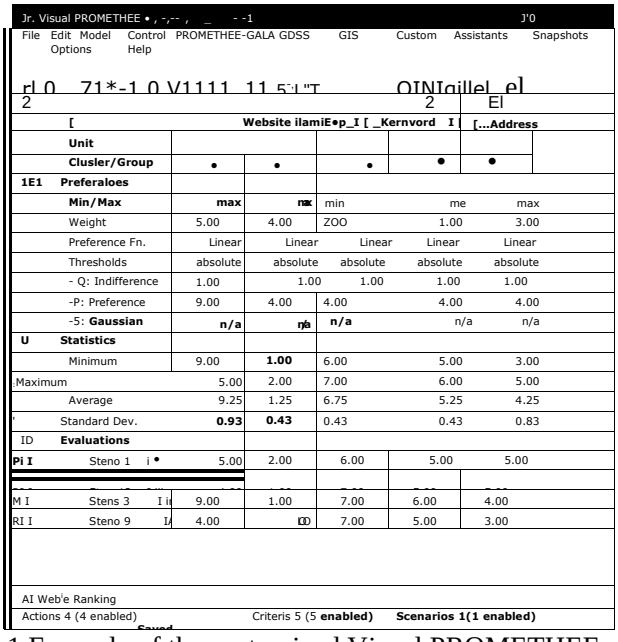

Figure 1 Example of the customized Visual PROMETHEE software.

To assist and expedite on the calculation and post-processing data, we utilize the Visual PROMETHEE software. Note that for this purpose, we customized the Visual PROMETHEE software into one scenario, four actions and five criteria for each keyword; the name for each actions and criterion are adjusted according to our problem in second step; the minimum and maximum values, weight and evaluation scale are set accordingly as in third step, as can be inferred from Figure 1. The matrix formation (fourth step), matrix search and transpose (fifth step) and ranking decision based on the *net flow* are done by the Visual PROMETHEE software. Further, to ensure the accuracy of the results produced by Visual PROMETHEE software, we first verify and compare the accuracy of the Visual PROMETHEE software solution with our analytical solution by calculate and solve for the matrix manually for the matrix formation (fourth step), search and transpose (fifth step) and the *net flow, leaving flow* and *entering flow* (sixth step).

3. Results and discussion

The experiments are carried out by analyzing 30 keywords with total 120 websites; the keywords and website can be found in Table Al in the Appendix, the details of the basic PROMETHEE values for all five factors in the 120 websites are listed in Table A2 in the Appendix. Note that due to strict page limitation, the data in the appendices are available upon request to the corresponding author. In the following, we analyze the Visual PROMETHEE software results and verify with solution from analytical calculation and validate with the evidence obtained from Google search engine.

First, we verify the accuracy of the results produced by Visual PROMETHEE software and compare them with our manually calculated solution. We note that the results from software in terms of ranking, *leaving flow*, *entering flow* and *net flow* are exactly match the same with that of our manually calculated solution. This confirms that the Visual PROMETHEE software solves for exactly the same equations, matrix and procedures as in our analytical PROMETHEE method.

Table 3. Validation of PROMETHEE rank results against Google results (short version).

No	Google rank	0^+- 0"	0^+	0"	PROMETHEE rank
1	1	0,3009	0,3009	0,0000	1
2	2	0,0694	0,1157	0,0463	2
...					
41	1	0.0231	0.0231	0.0000	1=
42	2	0.0231	0.0231	0.0000	1=
43	3	0.0000	0.0000	0.0000	2
44	4	-0.0463	0.0000	0.0463	3
120	4	-0.1157	0.0000	0.1157	4

Besides verification of the software results with analytical solution derived from the first principle concept, validation of the proposed method with experimental evidence from Google search engine results is of paramount importance for the judgment of successful method. The results of *leaving flow*, *entering flow* and *net flow* together with the rank results from both Visual PROMETHEE software as well as Google results are summarized in Table 3. Here several features are apparent; foremost among them is the good agreement between PROMETHEE results and Google results. On closer inspection, however, we note that the results from PROMETHEE are somewhat different than that from Google results, especially for the case where the website has very close similarity. For example, in the websites number 41 and 42 with the keyword of "buy ginseng online", both website, i.e., http://www.onlineginsengstore.com/ and http://www.buyginseng.org/ have very much similarities; thus, albeit the degrees-of-importance for scales of factors and subfactors are different, it, surprisingly, turns out to have similar *net flow* output which, in turn, leads to the same *net flow* and, thus, same ranking. Similar cases also occur for some other websites. This deviation may be attributed to the fact that our assumption on the weight value for the degree-of-importance of the factors (third step) is, perhaps, slightly different with that is used by Google (weight value is secret in Google) and, thus, leads to the deviation in some results. A thorough systematical adaption to determine the weight value for the degree-of-importance on Google search engine, which is beyond the scope of present work, will be considered in future. However, we highlight that the predicted results from our PROMETHEE method are found to

be sufficient enough for our purpose as the total relative error of prediction based on statistical analysis are less than **4%.** Thus, statistically speaking, the confidence level of the PROMETHEE method is found to be — 96%. Clearly PROMETHEE method and Visual PROMETHEE [6] is a reliable method and tool to predict website popularity rank, which shows potential to be used for practical applications.

4. Concluding remarks

A study has been conducted to demonstrate the ability of PROMETHEE method to predict the search engine rank of the websites by customizing Visual PROMETHEE software. The software results are verified with analytical solution derived manually from the first principle with 100% accuracy. The results is further validated against results from Google with large amount of data comprising 120 websites and 30 keywords; based on statistical analysis, the accuracy of PROMETHEE's prediction are — 96% which is good enough for practical purpose. This method can be very useful for company which has aggressive competitor in order to outperform their website visibility in the search engine result. Future work will focus on the rigorous adaption of weighting value to search for best combination of the degree-of-importance similar to one used by Google. The *feed-back loop concept* will also be introduced in the future work with the aim to improve the website rank and its implementation to the real website rank.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0025650).

References

[1]Top 15 search engines, http://www.ebizmba.com/articles/search-engines

[2]World Net Marketshare, http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4.

[3]N. Yalcm, U. K6se, What is search engine optimization: SEO, Procedia Social and Behavioral Sciences 9, pp. 487-493, 2010.

[4]R.A. Malaga, Search Engine Optimization—Black and White Hat Approaches, Advances in Computers 78, pp. 1-39, 2010

[5]N. Solehati, Y. Kang, J. Bae, Analysis of Popularity Site in Search Engine using PROMETHEE Method, Korean Institute of Industrial Engineers Conference, Seoul, Korea, 2011.

[6]Visual Promethee, 2012, http://www.promethee-gaia.net/software.html

[7]M. Behzadiana, R.B. Kazemzadehb, A. Albadvib, M. Aghdasib, PROMETHEE: A comprehensive literature review on methodologies and applications, European Journal of Operational Research 200, pp. 198-215, 2010.

[8]Brans, J.P., and Vincke, Ph. (1985), "A preference ranking organization method. The PROMETHEE method for MCDM" Management Science 31 / 6, 647-656.

[9]Brans, J-P, Vincke, Ph, Mareschal, B (1986) How to select and how to rank projects: The PROMETHEE method, European Journal of Operational Research 24, pp. 228-238.

[10] A. Esuli, F. Sebastiani, Page ranking WordNet synsets: An application to opinion related properties, In Proceeding of the 35th Meeting of the Association for Computational Linguistics, Prague, CZ. 2007, pp. 424-431.