

## QUANTIFICATION OF COMPETITIVE VALUE OF DOCUMENTS

P. Šimek, J. Vaněk, J. Jarolímek

**Received: May 7, 2009**

### Abstract

ŠIMEK, P., VANĚK, J., JAROLÍMEK, J.: *Quantification of competitive value of documents*. Acta univ. agric. et silvic. Mendel. Brun., 2009, LVII, No. 5, pp. 285–290

The majority of Internet users use the global network to search for different information using full-text search engines such as Google, Yahoo!, or Seznam. The web presentation operators are trying, with the help of different optimization techniques, to get to the top places in the results of fulltext search engines. Right there is a great importance of Search Engine Optimization and Search Engine Marketing, because normal users usually try links only on the first few pages of the fulltext search engines results on certain keywords and in catalogs they use primarily hierarchically higher placed links in each category. Key to success is the application of optimization methods which deal with the issue of keywords, structure and quality of content, domain names, individual sites and quantity and reliability of backward links. The process is demanding, long-lasting and without a guaranteed outcome. A website operator without advanced analytical tools do not identify the contribution of individual documents from which the entire web site consists. If the web presentation operators want to have an overview of their documents and web site in global, it is appropriate to quantify these positions in a specific way, depending on specific key words. For this purpose serves the quantification of competitive value of documents, which consequently sets global competitive value of a web site. Quantification of competitive values is performed on a specific full-text search engine. For each full-text search engine can be and often are, different results. According to published reports of ClickZ agency or Market Share is according to the number of searches by English-speaking users most widely used Google search engine, which has a market share of more than 80%. The whole procedure of quantification of competitive values is common, however, the initial step which is the analysis of keywords depends on a choice of the fulltext search engine.

competition, web site, fulltext search engine, document, Google, keyword

Meaning of „competitive strength“ has developed through the years. In the beginning was the competitive strenght in obtaining disposable sources, nowadays it can be seen as a broader set of competitive advantages(1). An electronic form of presentation of a company towards its surroundings is becoming one of the basic sources of competitiveness. The website presentations themselves with the ongoing development of the Internet come increasingly to the forefront of interest, primarily in terms of obtaining information. Whether created to present personal views, knowledge and skills, more or less artistic creations of individuals and interest groups, or as a source of information, news and

scientific knowledge, they comprise a lot of data, in which it is very difficult to search.

To create a high-quality Internet project, whether it's a company website, an electronic magazine or Web services, is undoubtedly a difficult issue. Unfortunately, many high quality projects remain hidden to end-users deep in the internals of the Internet and thus loses both parties involved – the web site operator loses profit and the user (potential customer) loses the quality content of the product. Enormous number of existing web sites makes orientation difficult and reduces the possibility of finding relevant information. On any web page – any information – can occur in two ways: either by entering the exact

address in a web browser, or by following the link from another site (1).

Interfaces, which intermediate Internet visitors a link guiding to the desired information, are the search servers. In most of them a visitor is asked to enter a sought word (keyword) or phrase and the search engine offers list of links (completed by a short description) that relate to the sought word or phrase. If the query satisfies more links, these are ranked according to a perceived relevance. This relevance is fully directed by the search engine.

Right there is a great importance of optimizing web pages for fulltext search engines (SEO<sup>1</sup>) and marketing based search engines (SEM<sup>2</sup>), because normal users usually try links only on the first few pages of the fulltext search engines results on certain keywords and in catalogs they use primarily hierarchically higher placed links in each category. As more and more websites accrued, the search engines became more frequent and popular (2).

Key to success is the application of optimization methods which deal with the issue of keywords, structure and quality of content, domain names, individual sites and quantity and reliability of backward links. The process is demanding, long-lasting and without a guaranteed outcome. Many web sites operators are therefore resorting to an alternative, which represents a number of variants of paid display of links.

### TARGET

A website operator without advanced analytical tools do not identify the contribution of individual documents from which the entire web site consists. The aim is to quantify the contribution of various documents of a web site by specific keywords, thus determine their competitive value, and subsequently quantify the global competitive value for the whole web site for the purposes of comparison with competitive web presentation.

Quantification of competitive values is performed on a specific full-text search engine. For each full-text search engine can be and often are, different results. According to published reports of ClickZ agency or Market Share is according to the number of searches by English-speaking users most widely used Google search engine, which has a market share of more than 80% (2). The share of search engines is of course difficult to measure and there are different approaches. But there is no dispute that Google leads globally. That is why the methodology of quantification of competitive values of documents is focused on this search engine, even if it is applicable to other search engines.

### METHODICAL PROCEDURE OF COMPETITIVE VALUES QUANTIFICATION

The whole procedure of quantification of competitive values is common, however, the initial step which is the analysis of keywords depends on a choice of the fulltext search engine. Analysis of keywords is essential, as properly accomplished analysis influences resulting competitive values. The first step is to get as many keywords as possible, on that the document, eventually the whole web site can be optimized. To obtain the width of keywords, it is possible to use a wide range of software tools such as Google AdWords, eTarget, Wordtracker or Overture Search Term Suggestion Tool. Selection of appropriate software tools is not important, very important is the work of SW itself, the distribution of keywords and keywords selection.

In any software tools for the acquisition of synonyms of keywords, it is necessary to gradually enter the primary keyword, which has been carefully determined. SW tool Google AdWords after entering a keyword or phrase, language selection and country or territory, returns a number of synonyms with valuable data on search volume, advertisers competition, average search volume, search trends, eventually the highest volume of searches. All offered synonyms of the relevant keyword are ranked by relevance and in some cases, the so-called other keywords are offered as well.

The first step of keywords analysis is a simple and quick calculation of relevance of each submitted keyword, which is calculated from the total number of offered key words and phrases and the order itself of the specific keyword and phrase. For example, if the total number of offered keywords is 150, the first, most relevant keyword or phrase, has a value of 1 ( $150/150$  – in most cases it is the primary keyword), and the last, least relevant keyword or phrase, has a value of  $1/150$ .

In the second step it is necessary to allocate keywords by a certain scoring method for “more specific keywords” and “other keywords” (if offered by the used application), for example, “more specific keywords” are evaluated by two points, and “additional keywords” only one point. If the used software application offers several groups of keywords, it is possible to use an appropriate scoring scale.

Determining of the relevance of individual key words and phrases and their distribution according to the specificity, is followed by permanent withdrawal of totally inappropriate and meaningless keywords and phrases through reduction in their number. That is to say that the deleted keywords or phrases absolutely depart from the original specified keyword (eg the phrase does include the pri-

1 Search Engine Optimization

2 Search Engine Marketing

mary keyword, but by itself refers to another topic), or contain domain names of the first order, the so-called Top Level Domain, or words that are in implicit use of the fulltext search engine ignored in the search, for example, conjunction, numbers, prepositions, etc.

It is necessary to assess and assign the purget, relevant keywords with a priority, ie. that all the key words and phrases are compiled respectively. The primary task is to summarize the key words and phrases in a table and draw up a matrix for the application of methods of multi-criteria analysis. In the first column of the table are the individual keywords and in the first line the individual criteria. From each application software for the analysis of keywords emerge somewhat different keywords (synonyms) and other criteria. The procedure by the analysis and selection of key words and phrases (the determination of their order) is the same or very similar.

To determine the weights of cardinal information about the preferences of individual criteria it is possible to use several methods, assuming that the web site operator is able and willing to determine not only the order of importance of criteria, but also the ratio of importance between pairs of criteria. The most widely used methods of this area are a scoring method, which transforms the scoring evaluation of criteria importance to a form of a weight vector, and Saaty's method of quantitative pair comparison which derives the weight vector from the information on estimation of relative weights ratio. After determining of criteria weight and accomplishing of checks, whether all criteria are maximizational, it is appropriate to use for the evaluation of individual keywords specific methods that require the cardinal information. There is a number of methods, which require to enter the cardinal information on criteria in form of weights and on variants in form of a criteria matrix with cardinal values, such as the weighted sum method and TOPSIS method (2).

$$CVW = c_1 \left( \frac{1}{o_1} + \frac{1}{o_2} + \dots + \frac{1}{o_k} \right) + c_2 \left( \frac{1}{o_1} + \frac{1}{o_2} + \dots + \frac{1}{o_k} \right) + \dots + c_i \left( \frac{1}{o_1} + \frac{1}{o_2} + \dots + \frac{1}{o_k} \right)$$

The resulting keywords and phrases from the keywords analysis are the basis for determining competitive values of documents. The procedure is such that each key word or phrase entered into the search engines such as Google and the search results themselves are subsequently analyzed. For the sake of objectiveness of searching, it is necessary that before entering a keyword in the search engine temporary cookies are erased that relate to a specific web browser and used search engine.

The overall ranking itself of competitive documents is provided by arrangement of resulting values of a simple multiplication of a reciprocal value of the order of the document in search results and keyword value derived from the keywords analysis. To obtain the most competitive web sites, the results of the order of competitive documents are classified according to domain names (domain), in the process the values of the same domains are summed.

Indicator of a competitive value of a document (CVD) <sup>3</sup> is calculated for each document from the search result according to the formula

$$CVD = c_i \cdot \frac{1}{o}$$

CVD.. indicator of competitive value of a document within keyword or key phrase.

$c_i$  ..... value of a multi-criteria variants analysis of keyword analysis (eg the value of the TOPSIS method or methods of the weighted sum).

$o$  ..... position of a document in the search results.

Another important indicator within the keyword value is a competitive value of a web site (CVWi) <sup>4</sup>, which is the sum of competing values of documents coming from one domain name. The formula for the calculation of CVWi is

$$CVWi = c_i \left( \frac{1}{o_1} + \frac{1}{o_2} + \dots + \frac{1}{o_k} \right) = c_i \sum_{j=1}^k \frac{1}{o_j}$$

CVWi .... indicator of competitive value of a website within keyword or phrase.

$o_1 - o_k$  ..... individual positions of documents in the search results coming under one domain name.

The third useful indicator is a competitive value of a web site (CVW) for all keywords, thus the sum of competitive values of a website from the individual keywords (CVWi), described by formula

or

$$CVW = \sum_{i=1}^n \left( c_i \sum_{j=1}^k \frac{1}{o_j} \right) = \sum_{i=1}^n CVWi$$

CVW.. indicator of competitive value of a website for all keywords or phrases.

$c_1 - c_i$  ... values from multi-criteria variants analysis of keywords analysis (eg, the value of the TOPSIS method or methods of the weighted sum).

3 Competitive Value of Document

4 Competitive Value of Web Site

## DISCUSSION

For the full-text Google search engine through the use of web application AdWords it is possible to use for the keywords analysis criteria in the form of advertiser competition, average search volume, and the relevance of specific words. Saaty's method is used to determine the weights of criteria, it is appropriate, if there is only one evaluator who evaluates. It is a method of quantitative pair comparison of criteria. For evaluation of pair comparison of criteria is used a 9-point scale. In the process it is possible to use the intermediate stages (values 2, 4, 6, 8):

- 1 – equivalent criteria  $i$  and  $j$
- 2 – slightly preferred criterion  $i$  before  $j$
- 5 – strongly preferred criterion  $i$  before  $j$
- 7 – very much preferred criterion  $i$  before  $j$
- 9 – absolutely preferred criterion  $i$  before  $j$

The evaluator, therefore, compares each pair of criteria and records size of preferences of the  $i$ -th criteria due to the  $j$ -th criterion in Saaty's matrix  $S = (s_{ij})$  (3).

Through Saaty's method of quantitative pair comparison the individual weights of criteria for con-

structed table of keywords as follows can be easily and quickly determined:

k1 – Advertiser Competition: 0.2650

k2 – Average Search Volume: 0.5924

k3 – Relevance: 0.1113

k4 – The specific words: 0.0313

Further steps in the calculation of competitive values are dependent on concrete search results, thus which documents are displayed in the search results and how many documents are included in the calculation of competitive values (eg the first 30 documents). The ideal case is if the top of search results on the given keyword, which are included in the calculation, displays only the documents from the operator's web site. On the contrary, the worst option is, if the search results do not display even one link to the operator's web site. The whole methodology of quantification of competitive values only discovers which position take the given documents and the whole websites in comparison with competitive documents and the whole web site. If the values are the highest in comparison with the competition, the documents and the operator's whole web site are the most competitive on the given keywords.

I: Exemplary table for calculation of keywords and phrases order

Keyword	Competitiveness of advertisers k1	Average search volume k2	Relevance k3	Specific words k4
word 1	value 1,1	value 1,2	value 1,3	value 1,4
word 2	value 2,1	value 2,2	value 2,3	value 2,4
word 3	value 3,1	value 3,2	value 3,3	value 3,4
word 4	value 4,1	value 4,2	value 4,3	value 4,4
word 5	value 5,1	value 5,2	value 5,3	value 5,4

## SUMMARY

An electronic form of presentation of a company towards its surroundings is becoming one of the basic sources of competitiveness. The website presentations themselves with the ongoing development of the Internet come increasingly to the forefront of interest, primarily in terms of obtaining information. Interfaces, which intermediate Internet visitors a link guiding to the desired information, are the search servers. In most of them a visitor is asked to enter a sought word (keyword) or phrase and the search engine offers list of links (completed by a short description) that relate to the sought word or phrase. If the query satisfies more links, these are ranked according to a perceived relevance. This relevance is fully directed by the search engine.

A website operator without advanced analytical tools do not identify the contribution of individual documents from which the entire web site consists. The aim is to quantify the contribution of various documents of a web site by specific keywords, thus determine their competitive value, and subsequently quantify the global competitive value for the whole web site for the purposes of comparison with competitive web presentation.

The whole procedure of quantification of competitive values is common, however, the initial step which is the analysis of keywords depends on a choice of the fulltext search engine. Analysis of keywords is essential, as properly accomplished analysis influences resulting competitive values. For the full-text Google search engine through the use of web application AdWords it is possible to use for the keywords analysis criteria in the form of advertiser competition, average search volume, and the relevance of specific words. To determine the weights of cardinal information about the preferences of individual criteria it is used methods of multi-critical analysis variants.

Indicator of a competitive value of a document (CVD) is calculated for each document from the search result. Another important indicator within the keyword value is a competitive value of a web



site (CVWi), which is the sum of competing values of documents coming from one domain name. The third useful indicator is a competitive value of a web site (CVW) for all keywords, thus the sum of competitive values of a website from the individual keywords (CVWi).

## SOUHRN

### Kvantifikace konkurenčních hodnot dokumentů

Elektronická forma prezentace firmy vůči svému okolí se stává jedním ze základních zdrojů konkurenceschopnosti. Samotné webové prezentace s nezpomalujícím rozvojem internetu se dostávají stále více do popředí zájmu především z hlediska získávání informací. Rozhraním, zprostředkujícím návštěvníkům internetu odkaz vedoucí k požadovaným informacím, jsou vyhledávací servery. Ve většině z nich je návštěvník vyzván k zadání hledaného slova (klíčového slova) nebo fráze a vyhledávač mu nabídne seznam odkazů (doplňených stručným popisem), které s hledaným slovem nebo frází souvisejí. Pokud dotazu vyhovuje více odkazů, jsou seřazeny podle domnělé relevance. Tato relevance je plně v režii vyhledávacího stroje.

Provozovatel webové prezentace bez pokročilých analytických nástrojů neurčí přínos jednotlivých dokumentů, ze kterých se celý web site skládá. Cílem je tedy kvantifikovat přínos jednotlivých dokumentů web site na konkrétní klíčová slova, určit tedy jejich konkurenční hodnotu, a následně kvantifikovat globální konkurenční hodnotu za celý web site pro potřeby porovnání s konkurenčními webovými prezentacemi.

Celý postup kvantifikace konkurenčních hodnot je obecný, nicméně prvotní krok, kterým je analýza klíčových slov, je závislý na výběru fulltextového vyhledávače. Analýza klíčových slov je zásadní, neboť správně provedená analýza ovlivňuje výsledné konkurenční hodnoty. Pro fulltextový vyhledávač Google za použití webové aplikace AdWords lze pro analýzu klíčových slov použít kritéria v podobě konkurence inzerentů, průměrného objemu vyhledávání, relevance a konkrétních slov. Pro stanovení vah z kardinální informace o preferencích jednotlivých kritérií jsou využity metody vícekritériální analýzy variant.

Ukazatel konkurenční hodnoty dokumentu (CVD) je vypočítán pro každý dokument z výsledku vyhledávání. Druhým důležitým ukazatelem v rámci klíčového slova je konkurenční hodnota web site (CVW<sub>i</sub>), která je součtem konkurenčních hodnot dokumentů pocházejících z jednoho doménového jména. Třetím přínosným ukazatelem je konkurenční hodnota website (CVW) za všechna klíčová slova, tedy součet konkurenčních hodnot website z jednotlivých klíčových slov (CVW<sub>j</sub>).

konkurence, web site, fulltextový vyhledávač, dokument, Google, klíčové slovo

The implementation of the project was supported within the development programmes of the Ministry of Education, Youth and Sports of the Czech Republic and at the CULS was maintained as a solution of an institutional research intention MSM 6046070906 „Economics of resources of Czech agriculture and their efficient use in frame of multifunctional agri-food systems“.

## REFERENCES

- CZADEK, M., 2006: *Projektové řízení jako zdroj konkurenceschopnosti v ICT*. Systems integration 2006. Prague. 362–369.
- VARGA, A., 2009: Materials for Implementing Interactive Media. P5 Vuorovaikutteisen median toteutus 3 os. [online] [citation: 24<sup>th</sup> of April, 2009]. <http://www.uta.fi/hyper/opetus/05kevat/p5>.
- SMIČKA, R., 2004: *Optimalizace pro vyhledávače – SEO: Jak zvýšit návštěvnost webu*. Dubany: Jaroslava Smičková, 2004. ISBN 80-239-2961-5.
- MARKET SHARE, 2009: *Search Engine Market Share*. Market Share. [online] [citation: 24<sup>th</sup> of April, 2009]. <http://marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4>.
- BROŽOVÁ, H., HOUŠKA, M., ŠUBRT, T., 2003: *Modely pro vícekritériální rozhodování*. Praha: ČZU v Praze, 2003. ISBN 80-213-1019-7.

## Address

Ing. Pavel Šimek, Ph.D., Ing. Jiří Vaněk, Ph.D., Ing. Jan Jarolímek, Ph.D., Katedra informačních technologií, Česká zemědělská univerzita v Praze, Kamýčká 129, 165 21 Praha 6 – Suchbát, Česká republika, e-mail: simmek@pef.czu.cz, vanek@pef.czu.cz, jarolimek@pef.czu.cz

