Methodical procedure for optimization of a document for fulltext search engines

Pavel Simek¹, Jan Jarolimek¹, Jiri Vanek¹

¹ Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Information Technologies, simek@pef.czu.cz

Abstract

An electronic form of company presentation towards its environment is becoming one of the basic sources of competitive advantage. To develop a high-quality internet project is, without question, a demanding matter; in particular, in case of a company web presentation, an electronic magazine or a web service. Nowadays, fulltext search engine servers become an interface which mediates a link leading to the required information to a visitor of the Internet. In the majority of these servers, the visitor is asked to enter the sought word or phrase. The search engine offers a list of links (supplemented with a brief description) which relates to the found word or phrase. If the request meets more links, these are listed according to their relevancy. This relevancy is fully on the expenses of the search engine.

The proposed methodical procedure for a successful optimization of documents (MPOD) goes partly out of the already known procedures of SEO activities; partly out of the significancy of particular factors that influence a position of documents in the outcomes of searching; as well as out of the longtime practical experiments. The whole methodical procedure for optimization of a document can be summarised to the following activities; whereas for the first three a web application is compiled:

1. Key words analysis
2. Analysis of competition with calculation of competitive value of a document
3. Current stage analysis (if it exists)
4. MPOD Copywriting
5. Optimization of a document and server
6. Registration to catalogues and production of retrace links
7. Effectivity analysis

The effectivity of MPOD differs to a great extent according to the thematic area and the aim of a website. Competition and number of service receivers (readers, potential customers) are the main factors which has an influence on effectivity. An indispensable aspect is time

Keywords: Internet, searching, fulltext, search engine, optimization, document, methodical procedure for a optimization of documents

Introduction

At present, information and communication technologies are developing with an unusual speed. Commercial subjects, subjects of state administration, municipal governments, as well as
individuals are obliged to use ICT more and more; they have still bigger requirements that lead to improvement, discounts and mass spreading of these technologies. Constant shortening of developmental, realization and production cycles is, at the same time, a consequence of explosive increase of IT applications and theirs intergrowth to all spheres of people’s activities [3]. At the same time, ICT are one of the crucial growth and productivity factors of advanced economies [4]. In the period of beginning of companies „computerisation“, ICT were perceived as one of the solutions how to acquire advantage in competition [1].

An electronic form of a company presentation towards its environment is becoming one of the basic sources of competitive advantage; and web presentations with continual development of the Internet are coming more and more in centre of interest; especially, with the view to obtaining information. Whether they were created to present personal opinions, knowledges and skills, more or less artistic work of individuals or interest groups; or as sources of information, news and scientific knowledges; they represent many data, in which it is very difficult to search.

To develop a high-quality internet project is, without question, a demanding matter; in particular, in case of a company web presentation, an electronic magazine or a web service. Unfortunately, many quality projects remain hidden for the end user deep in the internals of the Internet; and so, it is losing both interested sides – webside operators are losing their profit; users (potential customers) are losing a quality content or product.

Nowadays, fulltext search engine servers become an interface which mediates a link leading to the required information to a visitor of the Internet. In the majority of these servers, the visitor is asked to enter a sought word or phrase. The search engine offers a list of links (supplemented with a brief description) which relates to the sought word or phrase. If the request meets more links, these are listed according to their relevancy. This relevancy is fully on the expenses of the search engine.

**Optimization of a WWW page**

Right here has a significant importance an optimization of WWW pages for fulltext search engines (SEO3) and marketing (SEM4), established on the search engines. And this mainly because of the fact that common users usually use links of the first few pages in results of the fulltext search engines on definite key words and in catalogues they use primarily hierarchically higher set links in each categories.

Key to success is an application of optimization methods which deals with the issue of key words; quality and structure of content; domain names of individual pages; and number and credibility of retrace links. The process is demanding, longlasting and without guaranteed result.

Many web providers turn to an alternative which is represented by a whole range of variants of a paid display of links.

Todays biggest players on the market of fulltext search engines are Google, Yahoo!, Microsoft and ASK. It is speculated that AOL, which uses the Google technology, will be bought by some of the „big players“. The participation of search engines is, of course, difficult to measure; and various approaches exist. Although, there are no doubts that Google leads globally.

---

1 Information and Communication Technologies
2 Information Technologies
3 Search Engine Optimization
4 Search Engine Marketing
While the traditional Search Engine Optimization is oriented more on a partial arrangement of web pages with the aim to place the links in the fulltext search engines in the best way; the Search Engine Marketing on the bases of attentive analysis at first forms an effective strategy and it is further applied not only to the field of typical fulltext search engines, but also to catalogues of pages and pay-per-click search engines.

Nowadays, there is no publicly known and open MPOD for the widely spread fulltext search engines. At the same time, according to the statistics of renowned agencies, e.g. DoubleClick, the usage of the fulltext search engines for searching of information in worldwide web Internet is still growing; and website providers invest considerable financial amounts to creation and arrangements of their own documents and whole websites; or they buy text promotional campaigns on the exactly defined words, and this just because of the fact that their links emerge on the first places in the results of searching on a specific key word.

The biggest problem of diverse procedures and methods of WWW pages optimization for fulltext search engines are connected partly with inaccurately elaborated individual steps of optimization (or lack of specific steps); further stagnacy of optimization at specific point (step) and dynamical development of algorithm of the search engines.

**Methodical procedure for optimization of a document for WWW search engines**

The proposed methodical procedure for a succesful optimization of documents (MPOD) goes partly out of the already published procedures of SEO activities; and partly out of the significance of particular factors that influence a position of documents in the outcomes of searching.

After a longtime practical experiments and studying of accessible sources on the given topic, it is possible to summarise the whole methodical procedure for optimization of a document to the following activities; whereas for the first three a web application is compiled:

- Key words analysis
- Analysis of competition
- Current stage of a website analysis (if it exists)
- MPOD Copywriting
- Optimization of a document and server
- Registration to catalogues and creation of retrace links
- Effectivity analysis

At first, it is needed to define primary key words from the relevant branch and region – the so called „springboards“ of the whole process of a document optimization. The website provider should at first consider a target group of users (and thus potential customers), and from this group choose 1–10 primary key words.

**Key words analysis**

A key words analysis is for the methodical procedure of a document optimization essential; because of the fact that the rightly chosen key words significantly influence success of all MPOD steps. At first, it is needed to obtain as many key words as possible, on which it is possible to optimize the document (or the whole website). For acquisition of breadth of the key words, it is possible to use a whole range of SW tools, e.g. Google AdWords, eTarget, Wordtracker or Overture Search Term Suggestion Tool. Selection of a relevant SW tool is not important; what is important is work with the SW itself, division of key words and selection of key words.
The first step of the key word analysis in MPOD is a simple and fast calculation of relevancy of each offered key word which is calculated from the whole number of the offered key words and phrases, and order of a specific key word and phrase itself. In the second step it is needed to divide the key words by a certain score method to "specific key words" and "other key words" (if they were offered by the Google AdWords application), for example "specific key words" are scored by two points and "other key words" only by one point. The purified, relevant key words must be evaluated and assigned with priority; that means that a website provider or MPOD expert compiles an order from all key words and phrases; from this order each key word and phrase will be used by the process of MPOD. Primary task is to summarise the key words and phrases to a table, and to compile a matrix for application of methods of poly-criterial analysis of variants.

**Analysis of competition**

The second step of the MPOD is, after the key word analysis, a detailed analysis of competitive documents (websites), or noncommercial documents (websites), aimed at the similar topic. The basis for an analysis of competition are the resulting key words and phrases from the key words analysis. In principle, each key word or phrase is put to the search engines, e.g. Google and subsequently; results of searching themselves are being analysed. The result of the analysis of competition is a file of tables (for each key word or phrase there are few tables), which have an important declaring value, e.g.:

- Position of a document (o) – takes the value of 1-30.
- Google Toolbar PageRank (GTPR) pages – takes the value of 0-10.
- Title (title) – takes the value of 0 or 1 and informs about representation of a key word in a title of a document.
- h1 (h1) – h6 (h7) – takes the value of 0 or 1 and informs about representation of a key word in a first level heading in a document. In sporadic cases this indicator can also takes the value of more than 1. In this case it informs about an incorrect use of higher number of the highest level heading in one document.
- Description (des) – takes the values of 0 or 1 and informes about representation of a key word in the legend (caption) of a document.
- KeyWords (kw) – take the value of 0 or 1 and inform about representation of a key word in a list of key words of a document.
- Index Age (i.a.) – takes the value of counting numbers and informs about an indexation age of a document in hours.
- Number of words (w) – takes the value of counting numbers and informs about the whole number of words in the body of a document (in the body of a document, if it concerns XHTML document).
- Density of a key word (d.kw.) – takes the value of real numbers and informs about the proportion of representation of a key word or phrase on the total number of words in a document (number of key words/total number of words).
- Etc.

The main contribution of the analysis of competition is, besides an overview of whole range of important factors, a calculation of a competitive value of a document under a specific CVD key word, calculation of a competitive value of website under a specific CVW_i key word and calculation of a total competitive value CVW website.
Current stage of a website analysis (if it exists)
If a website provider already owns a particular WWW presentation, it is needed to accomplish an analysis of current stage, and thus a certain audit which partly copies the procedures from analysis of competition. An MPOD expert or website provider calculates a competitive value of each document for each key word or phrase from the key word analysis. If in the results of searching a link to any document of a website provider does not appear, competitive value of a document is zero.
Further very important indicators of each document are identical with the analysis of competition; namely, Google Toolbar PageRank of each document; number of retrace links; average value of Google Toolbar PageRank documents referring to the analysed document of a website provider and relevancy of contents of documents.

MPOD Copywriting
Copywriting for web differs from a classical copywriting by the so called „flightiness of a user“. Users are able to read a long solid text on a paper; however, on the web they read in a different way. On the pages a user mostly roams with eyes. He/she looks for clues, jumps from one place to another. A text for web must be short, with short paragraphs. Otherwise the reader gets tired and leaves. Writing for web is distinguished by a frequent use of bold print and italics – by highlighted parts of the text. And thus, sharing of information is a discipline which needs an appropriate training, clear objectives and relevant means [2].
Basis for creation of documents content is again the key word analysis. Although the modern fulltext search engines do not define relevancy of page according to the absolute frequency of the key words in a text, they assign a considerable significance to number of word occurrence in the text. They make this in connection to a total length of the text and this either linearly (the so called density of key word) or nonlinearly. The density between 2 and 10 per cent is considered as optimal for the important key words. From a certain limit, bigger number of key word occurrence already does not influence the position of a document in the results of searching. Frequency of the key words should correspond with results of the key word analysis; that means that the text should involve those key words, which are placed on the first places. Density of key words should not be overestimated. In part, it is not by far the only criterion, according to which search engines evaluate the pages; and in part, the stylistic quality, information value and language diversity of the text are usually more important.

Optimization of documents and server
Each document from the whole complex of providers‘ website must go through the MPOD. The basis for the MPOD is primary the key word analysis, the analysis of competition and the copywriting (possibly also an analysis of a current stage of existing website of pages). By the optimization MPOD expert must take into account:
• the importance of a key word which is a result of the key word analysis
• influence of XHTML marks and other factors on the search robots
• possible number of occurrence of XHTML marks in a document and influence of other factors on indexation of search robots
possible number of occurrence of a key word in a document, which is a result of copywriting of documents’ competitive value and other factors from the analysis of competition.

Effectivity analysis
Each provider of the Internet presentation should care if, and how it fulfils its purpose. The best way to gain the needed information is through an analysis of data about number, aims and
activities of website visitors. There are three possibilities, how to reach it. And these are: through compiling of server logs; through setting some of the measuring systems, e.g. Google Analytics; or through a usage of own measuring system.

For evaluation of the effectivity of expended sources for a document optimization (website) it is important to monitor the following key data:

- number of visits and number of viewed documents
- proportion of new visitors towards those who are coming back to the website
- geographical partition of visitors
- linking websites (sources), from which the visitors come
- number of conversions
- key words, through which the visitors came from the search engines
- which documents are mostly visited within the website
- through which document are the visitors coming
- through which document are the visitors leaving
- behaviour of the visitors on the whole website (ways in which they go through the website)

The number of visitors itself informs more in detail about unique entries on the website per specific period; average entries to the website per particular shorter period, e.g. days; and how many visitors came on the website for the first time.

**Discussion**

The above described issue is very wide and considerably exceeds the extend of the submitted text. As sources for proposing the methodical procedure for optimization of a document for fulltext WWW search engines served, in particular, an evaluation of influence of factors on evaluation of document relevancy by the search engine. Resulting rate of influence of individual factors was set partly from cca four-year survey of diverse continuous placing of 10 web presentations from different spheres and „Google Search Engine Ranking Factors V2“ document, in which participated 37 world leading specialists in the field of search engines.

In the last period, more and more emphasis is put on increasing of relevancy and specification of provided searching results. It emerged that it is possible to reach good results, namely through the identification of suitable key words and their connection to the content of pages.

The effectivity of the MPOD differs to a great extent according to the thematic area and the aim of a website. Competition and number of service receivers (readers, potential customers) are the main factors which has an influence on effectivity. An indispensable aspect is time. Although the SEM works more quickly, it works just at the time when it is continuously financed. Evaluation of the MPOD prove that the effect of optimization appears after few months, but it remains apparent for couple of years.

*Pieces of knowledge introduced in this paper resulted from 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“.*

---

5 Conversion sets proportion of a total number of the visitors and those who fulfilled a specific, in advance defined target.
References


