PROCEEDINGS
OF THE V INTERNATIONAL SCIENTIFIC
CONFERENCE
ADVANCED INFORMATION
SYSTEMS AND TECHNOLOGIES

AIST-2017
(Sumy, May 17–19, 2017)

Site structure analysis and optimization

Alexander Volsky

NTUU «Igor Sikorsky Kyiv Polytechnic Institute», volsalex@yandex.ua

Abstract – **The article contains information about current situation of site development, reasons why sites are so essential for activity of organizations and why researches of their structure improvement are so important. Short analysis of previous researches is provided, with their negative features and ways for further researches being exposed. Several ways to formulate recommendations and requirements for site structure are suggested, as well as methods for its optimization. A method for implementation of research results as software product is proposed. The information about an example of such software which performs analysis of sites of educational institution is provided too.**

Keywords – **site structure; optimization; site type**

I. INTRODUCTION

Currently, information technologies quickly spread around the world and are used in various fields of activity. The Internet becomes an important and integral part of the life of a modern man. Now websites are the main source of information for a huge number of Internet users. Therefore, most organizations and companies create their own websites, which represent them in the Web, as well as provide information about the capabilities, features and services of the organization. Since the vast majority of sites are aimed at providing their visitors with information placed on them, and in some cases offer additional functionality, such as purchases, the questions about attracting visitors (target audience) to the site and their retention (the transfer of visitors to the site permanent users) are very important nowadays, especially if there is competition with other companies or similar information resources. This is the reason why organizations create their own website templates, define design features, content and try to organize the site in such a way that it is highly rated by search engine robots and placed among the first positions of search results in various search engines.

The user's choice of a site is influenced by several factors, among which are: uniqueness of t content, time of access to the site and search for necessary information on it, the intuitive notion of structure and functional actions, etc.

One of the most important factors is the convenience of the site structure for the user. It should be simple, logical and understandable for the visitor. If a user is forced to spend a lot of time searching for the information he needs or trying to understand organization of the page hierarchy, he is likely to simply leave the site or will not want to visit it in the future.

Therefore, research on the organization of a good and user-friendly structure of the site is essential for the effective activity of organizations at the current time.

Further in the article, the results of previous studies of the structure of the site are reviewed, their negative features and possibilities for further research and development are discussed, and ideas and methods are proposed to determine the set of requirements and recommendations for site structure that will help make it more user-friendly. An example of using research results in a software product is given as well.

II. CURRENT SITUATION AND FORMULATION OF THE PROBLEM

Based on the results of previous studies, four main types of site structure were identified:

Linear - the site does not allocate sections, and pages are linked together in sequence. A page usually contains links to previous, next and main pages. Sometimes these sites can have several page sequences, with their pages available only for certain categories of users. Linear structure is most often used in websites where online books, presentations and step-by-step descriptions of certain processes are posted.

Hierarchical - pages in the site are organized in the form of a multi-level hierarchy, which can be described as a tree structure. In the structure there is a main page that contains a menu with links to sections of the site that represent pages of next level of hierarchy. Sections may in turn contain links to subsections or relevant information. This is a universal structure that is suitable for almost any resource: home web page, organization or corporate website, information site and others.

Network - has all properties of a hierarchical structure, but also gives the ability to quickly jump from a page of one section to a page of another that is logically linked to it, or from a page belonging to one section to page of the same level of hierarchy in other section (in contrast to the hierarchical structure in which transition is possible only up and down between pages on the hierarchical levels belonging to one section). Most often this structure is used in Internet shops and information portals. However, it is also the most difficult to implement and has big amount of internal links, so it is not suitable for large sites.

Combined - combines several other structures. Often a site can have a hierarchical structure, which in certain parts of the site transforms into a linear one.

An example of such a structure is a catalogue of articles - its content can begin with hierarchically arranged pages, but the content of large sections and pages is divided into several consecutive pages.

Also, there is a defined set of recommendations for site structure organization:

Three-click rule is an unofficial rule of site structure organization. It assumes that the user should be able to find any information he needs (the web page of the site) passing through three or less links, starting from the homepage of the site.

The rule of seven - the navigation menu of the site, as well as lists of subsections of the section, should not contain more than seven items representing sections of the site.

This rule is based on the property of short-term human memory to remember up to seven entities at a time and assumes that because of this it will be more difficult for the user to search for information he needs on a site where the menu breaks this rule.

Navigation rule - for the convenience of users it is recommended to post a link to the main page of site on all its pages. It is also common to use site logo as a link to the main page.

In addition, it is recommended to create and display on each page a navigation bar that includes links to the main sections of the site (and, possibly, to the subsections of the current section).

Map rule – it is highly recommended to create a special page on a website that contains a hierarchical list of links to all its pages. This way, if a user was confused by diversity of sections and subsections, he can visit this page and search for a page which contains the information he is looking for.

However, it is a weak solution if the site structure is bad as many users are unaware that such a page exists and will simply leave the site.

This classification of types of site structures and set of recommendations are commonly used, but they are often vague and contain few specific requirements. They also don't take into account site features and goals.

For example, number of sections on certain sites is determined by number of product categories sold on it, or by number of subject areas, information about which it provides. Some portals and sites are also very big and have a large number of pages, so they have to create pages of fourth or bigger level to make navigation for their users easier.

This is why specific requirements for different kinds of websites should be made as well. The suggested ways to solve these problems are proposed in the next section.

III. RESEARCH MATERIALS AND RESULTS

At the moment, there is a set of requirements and recommendations for site structure optimization, for example, Google's Webmaster recommendations [1] or services that perform site analysis and find existing errors (for example, Netpeak Spider [2]). However, recommendations are usually given in general form and do not take into account the features and purpose of the site. Using a detailed classification of sites makes it possible to make recommendations for improving the structure more accurate and take into account the specifics of the site. However, currently, there is a problem - no single classification of sites [3,4], so introduction of a unified site classification is needed. An example of such classification is shown in Table 1.

TABLE III.   MAIN SITE CLASSES

№	Site class
1	Business card website
2	Information site
3	Corporate image site
4	Online contact site
5	Online store
6	Personal project
7	Promotional site
8	Forum site
9	Online Service site
10	Web portal
11	MFA,MFS-site

Based on this classification, it is possible for each class to determine:
- recommended site structure
- the allowed number of pages
- the presence of certain sections in the main menu of the site
- special set of recommendations

To determine these parameters, a special statistical research for each class of sites should be made. The point of research is that each participant will have to find certain information on the site. To do this, he will work with two or more versions of the same site, with differences in structure, navigation, and content organization. For the study, a set of parameters is defined that serve as criteria for the convenience of the structure of each version of the site for the user. Such criteria can be:

1) Information search time – it is one of the main indicators of the quality of the site structure. Any flaws in it, such as a big nesting level of pages, confusing navigation, illogical organization of hierarchy, automatically increase this indicator. Accordingly, the lower its value, the better the structure of the site is.

2) The user's personal rating - a subjective user's estimation of the convenience of the site structure, which takes the value from 1 to 10 and is determined by the user himself after working with the site version.

Other possible criteria are given in [5].

Also, it is possible to determine the type of site structure and requirements for it based on the number of pages in it. To do this, at the initial stage, recommendations of web studios can be used, for example "Antula" [6] (Table 2) with further analysis of the type of site and user behavior when the structure is changed.

TABLE IV.   RECOMMENDED SITE STRUCTURE

№	Number of pages	Recommended structure
1	<40	One-level structure
2	40-700	Two-level structure, with third nesting level possible for 10-15% of pages
3	700+	Three-level structure, with fourth nesting level possible for 10-15% of pages

Based on the results of such researches, the specific requirements for the optimal structure of the site of each class can be determined. In the future, software product implementation of research results can be made as a service that analyzes site structure of the chosen site, compares its parameters with parameters determined by the requirements for the structure of this class of sites, and generates a set of recommendations for analyzed site structure optimization.

For personal research, a class of information sites and its subclass - sites of educational institutions was selected. The research was made on the websites of the faculties and departments of the Igor Sikorsky Kyiv Polytechnic Institute. The reasons for this choice were:

1) a large number of faculties and departments in the university, so there is a big number of relevant sites;

2) significant differences in the structure and design of these sites;

3) the necessity to update and optimize these sites

The importance of the research is also increased because of the fact that the sites quality directly affects the opportunities for attracting new enrollees, the sites position in the search engines, and the position of the educational institution in the ratings that take into account its electronic resources, such as Webometrics [7].

Target audience of such sites includes enrollees and their parents, students, as well as employees and managers of the educational institution. The site must provide them with the information necessary for each category. Therefore, the main criteria for analyzing these sites are:

1) number of pages of the site and accordance of site structure to recommended;

2) nesting level of pages;

3) number and length of items in the main menu of the site;

4) link quality - the presence in the link text of words associated with the content of the page

For the research, special software "Edu Spider" was created. The structural scheme of the program is shown in Fig. 1.
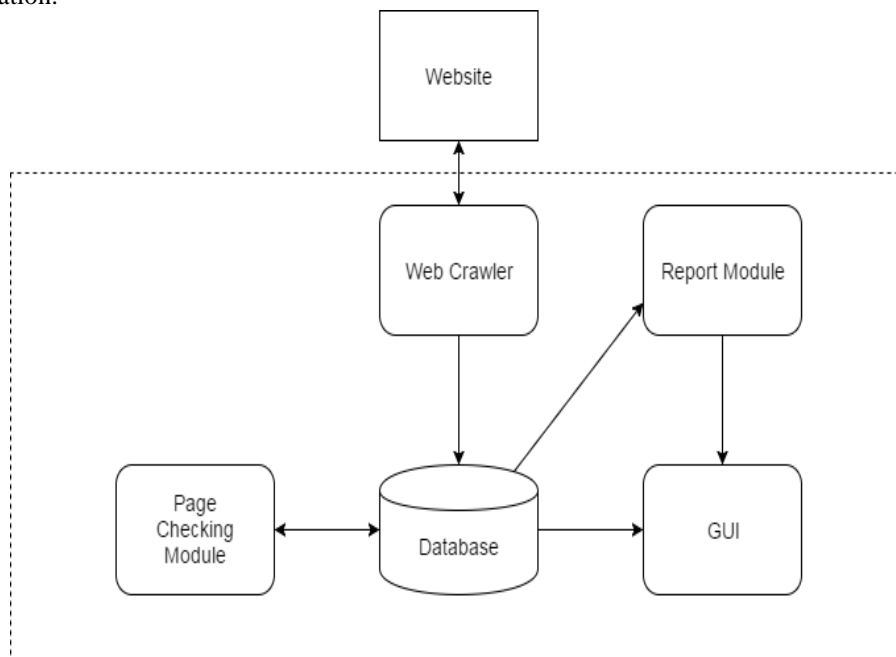

Figure 1.   Structural scheme

The web crawler performs crawling of site internal pages and gathers the necessary information that is saved in the database. After that, the page checking module checks whether the site structure and certain page parameters match requirements and recommended values. If there are disparities and errors, the module adds information about them to the database. Information from the database, if necessary, is processed in the report module and displayed in the user interface, giving him information about the results of the site analysis.

Also, the program builds a tree graph of site pages, with problem pages being highlighted (Figure 2).
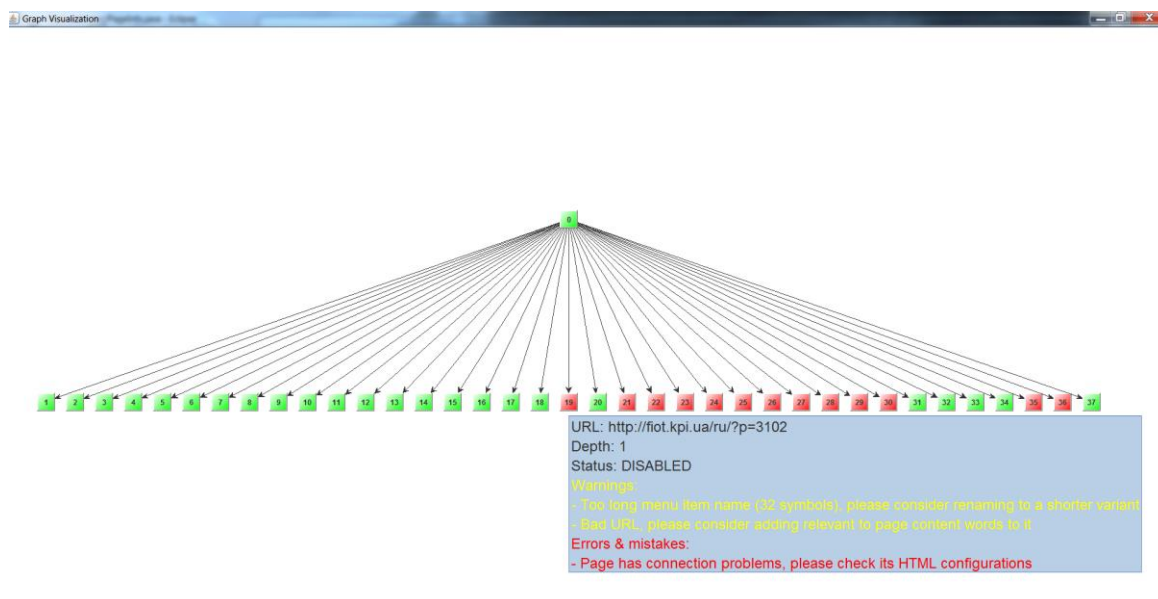

Figure 2.   Tree graph of site pages

Each page of the site is represented by a rectangular-shaped node of a tree with a unique page identifier. Pages where no problems or inconsistencies were found are highlighted in green. Problem pages are highlighted in red. When hovering over a page, the user can see the general information about the page: its link, nesting level and status. Nesting level shows the minimum number of transitions from the site main page needed to get to this page. The status determines ability to connect to the page. The problem pages additionally show information about detected inconsistencies and recommendations how to remove them. Possible errors and inconsistencies of the recommended page parameters are indicated in yellow. Explicit errors and violations of requirements are indicated in red.

CONCLUSIONS

In this article, the main features and problems of site structure analysis and optimization were shown. Several methods have been proposed to solve the problems discussed. They include the introduction of a general classification of sites, the definition of specific requirements for each class of sites, as well as methods that can be used to determine the appropriate structure for the site based on the number of its pages. These solutions can make the requirements more specific and take into account the specific features and purpose of the site, which will complement the set and correct the main shortcomings of existing requirements.

Also, the directions for further research and some features of their conduct were indicated. Based on the results of these studies, it is possible to create special software that will be able to automatically analyze the site structure and offer opportunities for its optimization.

Based on the research, special software was developed that performs the analysis of the websites of the higher educational institution Igor Sikorsky Kyiv Polytechnic Institute and offers recommendations how to remove discovered shortcomings.

REFERENCES

[1] *Google Webmaster Guidelines.* (n.d). Retrieved from https://support.google.com/webmasters/answer/35769?hl=en

[2] *Netpeak Spider: Your Personal SEO crawler*. (n.d). Retrieved from https://netpeaksoftware.com/spider

[3] Koksharov, S. (2010, June 24). *Internet Marketing site classification*. Retrieved from https://www.seonews.ru/analytics/klassifikatsiya-saytov-v-internet-marketinge/ (in Russian)

[4] *Yandex Catalogue site classification*. (n.d). Retrieved from https://yandex.ru/support/catalogue/basic-info/classification.html (in Russian)

[5] Kosarekov, I. (2013, July 2). *Site quality criteria*. Retrieved from https://blog.mcsite.ua/quality-site.html (in Russian)

[6] *About Antula Web Studio.* (n.d). Retrieved from http://www.antula.ru/about-firm.htm (in Russian)

[7] *About Webometrics.* (n.d). Retrieved from http://www.webometrics.info/en/About_Us