Formalized Analysis of the Web-Site Structure

Natalia Pasichnyk, Viktor Spilchuk, Ruslan Shevchuk, Iryna Spivak

Abstract - **In this paper the information technology of support of the website structure analysis is given.**

Keywords – **Website, Structure Analysis, Information Technology.**

I. INTRODUCTION

The structure of a Website and its segments is the basis for its information content, which defines its popularity and usefulness to a large extent [1]. The design and modification of a website begin with the manipulations with its structure. The choices that are made at this stage define the major volumes of the further developers' work. In case of successful analysis such choices provide for significant increase of the website popularity and usefulness among its target audience. Intuitive analysis of this process is not always efficient and the cost of the mistakes is too high on this stage. Methods of the formal analysis of this problem are scarce, thus development of these methods is a timely task, which is the subject of the current study.

II. INFORMATION TECHNOLOGY OF SUPPORT OF THE WEBSITE STRUCTURE ANALYSIS

The segments of a website should contain timely for the target audience information in order to provide for its appeal. At the same time the segment should contain the basic information to support website's integrity. The researcher reviews the structure of a few competitive websites for intuitive specification of the developed segment's structure The flaw of such analysis lays in its labour-intensiveness and the possibility to skip unique structure elements among large masses of repeated information. The use of information technologies ensures that the analysis of the large scale of the website structures is complete and that the excess review of the similar pages is avoided.

Generally, the titles of the pages that make up the website structure are presented in the main-page menu. However its formalized parsing is hard to provide, since such menus are often realized via the JavaScript technology. Thus it's better to conduct the analysis of the website structure through the links in the HTML code of its sitemap.

The first problem of the automation of website's structure analysis lays in the synonymic display of a term. It is resolved using the following term-recognition predicate Cn by its linguistic interpretation $LC(CN) = \{LCm_i(CN)\}_{i=1}^{n}$, which includes a certain number of components LCm_i. The components are identified according to following relation

$$Bases = < BsP, Base, Wrd, Lg\_ID >$$

$$BaseSym = < Base\_ID, BsP, Id\_SA, BsPCn >,$$

$$IdB(LCm_i(CN)) = \pi_{Base\_ID}(\sigma_{Wrd=LCm_i(CN) \cap Bas(BsP)=BasSyn(BsP)}(Bas, BasSyn)) \quad (1),$$

where BsP is identifier of the base representative, Base is base of word, Wrd is word, Lg_ID is language identifier, Base_ID is base identifier, Id_SA is subject domain identifier, BsPCn is number of the contacts with the base representative.

A term is considered as recognized if a set of the bases that represent it are found. The recognized term is displayed by its most often mentioned linguistic interpretation. The unrecognized term is displayed as is. The user can add it to some synonymic class of the current subject area after the analysis of the given text.

The next problem is that the HTML code of the sitemap contains many repetitions, excess detalization and unnecessary for our analysis information. So the information has to be filtered.

This filtration is conducted by developer by defining specific options during the review of the first website's structure with minimal refining of the filtration on the next websites' sitemaps.

Such technology gives the possibility to quickly analyze the structures of a great number of sites, yet bring the attention to its specialties and unique elements. The choice of the websites for analysis is conducted on the automated processing of the search-engine return list for the requests, which specify the subject segment. The developer modifies the structure of the under-development site basing on the model of the segment's structure which is automatically generated upon the analysis of the set of segments in the information system.

The offered technology also includes the generation of the HTML code of a website-development sitemap, which is used to range the elements of a future changes set, that are done through the web-questionnaire for such elements' popularity.

III. CONCLUSION

The information technology of support of the website structure analysis which minimizes its realization efforts on a set of websites of the subject segment is offered. The support of questioning the popularity of the website's structure development options is presumed. This approach allows formalizing the prognosis of competing websites' popularity for the target audience.

REFERENCES

[1] N.R. Pasichnyk and M.P. Dyvak, " Formalism in the quality site creating problem," *Scientific papers of the DonNTU, Informatics, Cybernetics and Computing Devices series*", vol. 14(188) 2011, pp. 325-329 (in Ukrainian) http://www.nbuv.gov.ua/portal/natural/Npdntu_ikot/2011_14/4_13.pdf .