

# Advanced Search Query for Identifying Web-Forum Threads Relevant to Given Subject Area

Peleshchyshyn Andriy, Korzh Roman, Oksana Tymovchak-Maksymets

**Abstract** – The paper presents a formalized search query to global search engines, such as Google, Yahoo etc. for detecting Web forum threads in which discussion subject is relevant to a given subject area.

**Keywords** – Web forum, thread, search engine, formalized search query.

## I. INTRODUCTION

Search engines are an extremely efficient means of identifying web-communities in particular Web forums in which discussion subject belongs to the given subject area. However, basic queries to search engines using keywords are effective only for keyword search on a Web page. They are less effective in filtering web pages of certain types. In addition, keyword search results give a lot of information noise (irrelevant threads, non-community web pages etc.) Thus, Web forum page search requires, apart from basic, some additional features which enable advanced search.

## II. ADVANCED QUERY DESCRIPTION

For detecting relevant Web communities the authors have developed a formalized advanced search query to global search engines.

The query consists of the following operations:

1. operation of Web community detection;
2. operation of relevance determination.

The **operation of Web community detection** enables identification and selection of those web pages which belong to Web communities. Functionally, this operation consists of:

- operator for Web address analysis
- Web community indicator.

**Operator for Web address analysis** is a search filter which removes those web sites and web pages which Web address does not match the defined restriction on the Web address structure. Formal markers in URL address point to community web pages. Therefore, operator for Web address analysis selects pages with special URL address structure, which is set by a Web community indicator.

**Web community indicator** is a formal sign, which shows that the website or web page belongs to Web community. Web community indicator is part of the URL address of the website or web page by which web communities can be identified.

Special indicator is determined by its software platform. Each platform has some special indicators of web pages that

Andriy Peleshchyshyn - Lviv Polytechnic National University, S. Bandery str., 12, Lviv, 79013, UKRAINE, e-mail: apele@ridne.net

Oksana Tymovchak-Maksymets - Lviv Polytechnic National University, S. Bandery str., 12, Lviv, 79013, UKRAINE, e-mail: tymovchak@ridne.net

are part of a community. Indicators for popular web platforms are presented in Table 1.

**Operation of relevance determination** reveals if the identified community Web page is relevant to the given subject area. This operation consists of:

- operator for subject analysis;
- relevance indicator.

TABLE I. POPULAR WEB COMMUNITY PLATFORM INDICATORS

Web community platform	Web community indicator		
	Indicator base	Refining Yes/No	Indicator refining
Platform independent	forum	No	-
phpBB	viewtopic	No	-
vBulletin	showthread	No	-
Simple Forum Machines	index.php?to pic=	Yes	Powered by SMF
myBB	showthread	Yes	-
	thread.php	Yes	Powered by myBB
Invision Power Board	showtopic	Yes	Powered by Invision Power Board
XMB, PHP-Fusion, Discuz etc.	viewthread	No	-
Word Press	“topic.php”	Yes	Powered by ExBB, Powered by BBpress

**Operator for subject analysis** is search filter that removes from the list of web community pages those which are not connected to the given subject. Web page relevance to the subject are established by its heading - word or word combination in the page title, in `<title>` tag.

**Relevance indicator** is a word or phrase in the title of a Web page which proves that Web community page is relevant to the given subject. As the title of a web page in most cases coincides with the subject of Web forum thread, the presence of a specific word or phrase in the title indicates that the discussion may relate to the subject.

## III. CONCLUSIONS

The developed universal search query can be employed in any search engine to identify relevant Web forum pages in the given subject area. This task is part of the complex task of identifying customer experience in Web forum discussions.