WWW.HR – How to Help Users Search the Directory?

Authors: Igor Ljubi, Goran Doboš
Faculty of Electrical Engineering and Computing, University of Zagreb,
Croatia

Extended Abstract

1. INTRODUCTION
Searching the Internet and finding relevant information quickly can often become
"mission impossible".
Usual ways of finding information on the Web are through the use of Web search
engines and Web
directories. It is important to understand how the users execute their search
queries. In this paper we
analyze how the administrators of sites listed in our Web directory can help the
users to get more
relevant results while searching the Directory.

2. CASE STUDY

WWW.HR is a Web-based information service supported by the Croatian Academic and
Research
Network – CARNet. WWW.HR consists of two services: general facts about Croatia
and a Web
directory. The directory is a hierarchically organized, fully searchable
catalogue of Croatian or Croatia-
related Web sites. Its top level contains 14 categories and an ever growing
number of subcategories [3].
Sites that are listed in the Directory are submitted by their authors. Upon the
submission, the site search
index is created based on the information provided through submission forms:

- Site URL,
- Site name in Croatian and English,
- Site description in Croatian and English,
- Category names in Croatian and English,
- META keywords extracted from the page.

Clearly, if those 5 parameters don't correctly describe the site, users trying
to find the particular
information could miss this site, although it is listed in Directory. Therefore,
the site administrator must fill
the missing information in before accepting the submission into the Directory.
When the user executes
the query, the directory search index database is examined and matching results
are returned to the
user.

3. ANALYSIS
To assess how the administrators can help users in getting more relevant results
while searching the
Directory, we have monitored new site submissions from April till June
2005,analysing the information

that is important to our search engine. During that period, 1544 new sites were added to the Directory,
while 558 sites were discarded (mostly because they were not Croatian-related sites).

The monitoring process revealed that 30% of site submissions don't provide site name and/or site
description in English language. This is the first point where administrator has to intervene. It is usually
done by checking if the site has English version, and copying a few sentences from "About us" or similar
section. Due to large number of submitted sites (daily average around 60), this method requires a great
deal of time.

META keywords can also be an obstacle in providing relevant results to user's queries, since authors of
Web sites are known to put "popular" words in META keyword to attract visitors (keyword spamming).
Some efforts in this area have already been done [2], but there is still a need for a tool, that would
generate metadata from the Web page that is being submitted. Of course, this also calls for cooperation
with site creators, since only they can put metadata on their pages.

A step in this direction is the tool that is available to the users of our Directory through on-line site
submission form. Submitters are encouraged to add relevant keywords to their submission form. Upon
submitting it, the tool generates Dublin Core and HTML metadata, which are delivered to the submitter
with the instructions on how to put this metadata information to their Web site.

Metadata information available on sites listed in Directory enables more thorough searching of our
content. This information is extracted from the page and used to generate index of Web sites available
in the Directory. Recent measurements of Croatian Web space [5] has shown that almost 45% of
Croatian Web pages have some type of metadata information. 53.5% of sites submitted to WWW.HR
Directory had KEYWORD meta tag, while 48.4% of submitted sites had DESCRIPTION meta tag. It is
interesting that only about 4% of metadata information was Dublin Core metadata.

4. CONCLUSION
We have conducted the analysis of submission forms for WWW.HR's Directory of Croatian Web sites in
order to discover what administrators can do to help users in more successful searching of the
Directory. Our research concluded that beside correcting and/or adding the missing parts of submission
forms, we can help by generating metadata information on those sites as well as educate our users on
more sophisticated methods of searching the Directory. A metadata generator has been introduced to
our submitting form. Since the pages listed in the Directory represent entry points to the respective Web
sites, they should contain accurate metadata information. With helping their administrators to add this
information to their sites, we are pursuing our goal to have 100% of listed pages described with accurate
metadata information.

5. REFERENCES
[1] Gordan Gledec, Igor Ljubi "How users search WWW.HR directory", CARNet Users Conference 2004.
[2] Igor Ljubi, Gordana Stojšiæ: "A tool for adding Dublin Core metadata to Web pages", CARNet Users Conference 2002.
[3] Maja Matijaševiæ, Gordan Gledec: "Implications of Web resource composite for Web server

workload characteristics", Proceedings of NEW2AN 2004, Sankt Petersburg, Russia, February
2004, pp. 189-193.
[4] Damir Juriæ, Maja Matijaševiæ, Gordan Gledec: "Using Ontologies To Improve Search On
WWW.HR Web Directory", to be presented at CUC 2004.
[5] http://www.srce.hr/mwp