How to Help Users Search Web Directory

Igor Ljubi, Goran Doboš

Department of Telecommunications

Faculty of Electrical Engineering and Computing

University of Zagreb, Croatia

Summary

- About WWW.HR project
  - Web directory
- How Users Search
- How to Help Users Search
- Analysis
- Results
- Conclusion

WWW.HR

- Web based information service supported by CARNet
  - established in 1994.
  - thematic portal, providing regional information concerning Croatia
  - two services:
    - Facts about Croatia
    - Web directory

WWW.HR Search Index

- site name in Croatian and English
- site description in Croatian and English
- site URL
- category names in Croatian and English
- META keywords extracted from the submitted page

Supported queries

- **all keywords**
  - default query, eg. "zagreb live"
- **logical expressions**
  - eg. "+zagreb -live", "zagreb or live"
- **wildcard queries**
  - eg. "europe*"
- **phrases**
  - eg. "history of shipbuilding"

How to Help Users Search

- Corecting description of site
  - Filling missed information prior to accepting site
  - Checking spelling and/or grammar
  - Putting site into appropriate category

- METADATA
  - Keyword spamming

Analysis

- Monitoring period: April - June, 2005
  - 1534 new sites added to Directory
  - 558 sites were denied
    - Non Croatian sites
- What was analysed:
  - Complete site description
  - Bilingual description
  - spelling
  - grammar
  - METADATA information

Results

- ## 30% of sites doesn't have English description

  - Time consuming due to large number of sites
  - Involves copying a few sentences form "About us" or similar page

- ## METADATA

  - Low usage of DC metadata
  - Abuse detected mostly on non-commercial pages
    - Keyword spamming is detected and eliminated

Solutions

- METADATA generator
- Users can input keywords relevant to their sites
  - Keywords; DC.Subject; DC:Title; Description are genereated with appropriate HTML code
  - Users are instructed how and where to put this code
  - Refference to CDA0027
- Upgrades:
  - Checking inputed keywords

Solutions (2)

- MWP has detected that 45% of Croatian Web pages have metadata information
- 53.5% of sites submitted to WWW.HR has KEYWORD meta tag
- 48.4% has DESCRIPTION meta tag

- Only 4% of sites have Dublin Core metadata

Conclusion

- More than 1500 submitted sites analysed
- Poor description – spelling/grammar
- Minor usage of metadata
  - users fail to recognize the advantages offered
- high percentage commonly generated metadata
- Tool for generating metadata information based on user's input

Thank you for your attention.

... and don't forget to visit Croatian

Homepage at www.hr ☺