

LIS566

Lecture 4

Internet Directories I

Thomas Krichel

<http://openlib.org/home/krichel>

2001-10-21

Structure

- introduction
- the dmoz project
- what makes for a quality resource gateway?

Reading

Project Web Site <http://www.dmoz.org>

metadata

• meta: Greek for about

• metadata: "data about data"

• metametadata: "data about metadata"

usually used in on the context of description of digital resources,
esp. those available on the internet.

quote

"There is one thing that I know about metadata, it is expensive
to produce."

William Y. Arms

Krichev's strategy to create a free metadata

1. Have people—possibly volunteers—maintain the metadata
catalog for free.

2. Have a lot of catalogers, so each can do it in their spare
time.

3. Reward cataloger by widely distributing their work.

According to them

The Open Directory Project is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors.

Claim that there is a historic precedence in the Oxford English Dictionary.

Formerly known as "GnuHoo", then "NewHoo", then acquired by NetScape, and called "dmoz".

implements 1 and 2 and 3

1. dmoz is maintained by volunteers "net-citizen". No special qualifications required, but claimed to be experts.

2. There are about 30,000 volunteers (they claim).

3. Powers the core directory services for the Web's largest and most popular search engines and portals, including NetScape Search AOL Search Google Lycos HotBot DirectHit

and hundreds of others, so they say. Catalogs 2,3M URLs

Who runs the ODP?

• hosted and administered by Netscape Communication Corporation, as a non-commercial entity

• a very small staff responsible for

– editorial policies and direction

– community management and development

– systems engineering

• but "first and foremost a self-regulating community".

Web directory

Purpose of ODP is to list and categorize web sites.

- no ranking

- no promotion

- no explicit search engine optimization

Social Contract

Inspired by and derived from, the Debian Social Contract

- "The Open Directory Will Remain 100% Proprietary content is added

- "We Give Back to the Web Community" aim for
 - * comprehensive * "high quality"
 - * user-friendly * widely available taxonomy
 - * selective * fair treatment of all sites

- "We Don't Hide Our Official Editorial Policies"

- "We Provide an Open Invitation to Join"

- accept editors from all walks of life

- each application to become an editor is being reviewed

- applicants have to show

- * fairness

- * impartiality

- * objectivity

- "We Encourage a Self-Regulating Community"

- "Our Priorities are Our Data Users and the Community"

- "Users Not Meeting The Free Use License"

- unattributed use of the data

basic editing

- add a new site

- add a submitted site

- move a submitted site

criteria for adding sites

- Does the site contain useful content? Is it content-rich?

- Are the site's contents not currently covered in the category?

- Is the site current? Does it appear to be well-maintained?

- Does the site appeal to a broad range of users? Does it provide diversity to the category?

probably inappropriate

- affiliate links

- mirror sites

- multi-level marketing

- illegal sites

- sites that do not resolve to a domain

- product listings

- deep-linking

non-English sites

Non-English sites

- all listed under "world"

- translated sites may appear in several categories

Adult Sites only in the "Adult" category

describing sites

titles should be

- informative concise

- no words like "homepage", "online"

Descriptions should be

- short, factual, non-promotional

- no acronyms

- no repetition of title information

- use third person

Categories

No editor "owns" her categories.

The guidelines for creating sub-categories are relatively vague.

There can be "@links", like "see" in a conventional catalog.

There can be related categories, like "see also" in a conventional catalog.

Types of Editors

- "foot soldier"
- Editorial Editors
- Content Editors
- Category Moderators
- Meta Editors

- Removal of editors
- Failure to comply with the guidelines
- Inability to function well within the Open Directory community
- Poor editing
- Uncivil behavior toward submitters, other editors, or DMOZ Staff
- Violation of ODP forum and email privacy
- Self-promotion (such as site cooling, and title or description manipulation)
- Discriminating against or tampering with competitors' listings
- Spamming the directory

critique

from <http://www.donotgo.com/godoz.htm>

... These menus are not a logically categorized version of the search database. Typically, they contain only one out of every hundred sites in the search engine data base. This appalling lack of representation is made worse by the fact that "coverage" is dictated by the interests of the volunteers who are, as a group, not representative of the general population.

I have described these "armies" as a horrible mix of corrupt generals and untrained privates, since there are only two kinds of "guide" volunteer: The passionate, often self-interested, "subject spammer" and the virtuously motivated, but web-ignorant, "want-to-belonger".

critics discussion

on <http://groups.yahoo.com/group/xodp>, we have xodp: "A discussion group for ODP expatriates, ODP malcontents who are thinking of leaving ODP, and various critics of ODP."

from <http://www.traffick.com/story.asp?StoryID=59>

- Lack of representativeness and lack of transparency.

- Incentive for corruption and excessive categorization of low-quality sites.

- The "open" directory is owned by a \$300 billion company.