

# Experience of Librarians in Design and Development of the National Web Directory of Iranian Sites, *Iranhoo*

Kayvan Kousha<sup>1</sup>

Mahsid Abdoli<sup>2</sup>

As the volume of Internet resources grows, there is a need to provide better and effective access to organized and high quality information on the World Wide Web. This paper describes how librarians at an Iranian-German joint venture company compiled a directory of Iranian sites, *Iranhoo*, with Persian language interface. The whole processes, from searching and evaluating for relevant sites to content description and arranging them in hierarchical subject categories are described. The paper also gives attention to traditional skills and knowledge of librarians to build and maintain national web directories to serve special information needs of users with different cultural and social values.

## 1. Introduction

Traditionally librarian is known as a person responsible for dissemination and organization of printed materials. With the advent of Internet and ever-increasing number of web sites this picture has rapidly changed to "Cybrarian", a person who use digital information on Internet for information services. Librarians have been information providers for centuries. They now have the opportunity to use web environment to provide information services. Web-based databases, reference sources, electronic journals, library OPACs etc. can be accessed through World Wide Web. Much attention

has been given in the LIS literature to search engines and their effectiveness for locating information (for example 1-7). However, little mention has been made of subject directories that exist to guide users to information on the web (8-10). The directories can be compiled and maintained by commercial companies, research institutes or libraries. Whereas most maintained by commercial companies.

As web directories are human-compiled of web sites in subject tree or hierarchical structure, traditional knowledge and experience of librarians can be used for designing and developing them. For the librarian traditional tasks of cataloging, indexing and searching for information are equivalent to

---

<sup>1</sup> PhD Student, LIS Department, University of Tehran, Iran  
Content Manager, *Iranhoo* Web Directory Project  
kkoosha@ut.ac.ir

<sup>2</sup> Reference Librarian, National Library of Iran  
Mahshid\_abdoli@hotmail.com

design and development of web directories. This has given the librarian a new dynamic role to play and serve in the internet environment. While search engines must often

web directories can be considered as more appropriate tool for finding higher quality information on the world wide web. Web directory are completely based on individuals who find, evaluate, describe and organize web sites. So they should have skills for: Searching and finding information on the web, evaluating the content of web sites, writing a brief informative description about web sites and organizing the web sites based on subject hierarchical structure. We can see close relation between above skills and librarians' knowledge and experience.

This paper has an overview on the role and experience of librarians who compiled and organized the content of a Persian national web directory, *Iranhoo* ([www.iranhoo.com](http://www.iranhoo.com)), and social and cultural impact of their job to attract ten thousands of Iranian end-users daily to access to about 4,000 Iranian web sites through Persian search interface as well as 500 browseable subject categories.

## 2. What is a Web Directory?

A Web directory is a subject tree or hierarchical list of selected web sites. Under each main category (topic) there's a list of sub-categories each of which may be linked to another set of sub-categories, and so on. The user can browse through the directory by going from the more general to the more specific topics and find the list of related web sites or hyperlinks. Construction of subject tree in web directories is a very important issue. Some Web directories have their own proprietary subject tree such as Yahoo ([www.yahoo.com](http://www.yahoo.com)) or Open Directory ([www.dmoz.org](http://www.dmoz.org)). There are several web

wade through piles of irrelevant sites to find what users are looking for based on a database of web sites created by a software program called a "robot" or "spider",

directories using classification schemes like Library of Congress Subject Headings (LCSH) and Dewey Decimal Classification (DDC) for organizing web sites. Although, the subject tree is important tool for navigation and finding desired information, many web directories use search interface in basic or advanced mode either to retrieve categories or indexed web sites in their database.

Directories are human-compiled and they rely to some extent on human intuition. Unlike search engines, which use robots to find and index web pages with any content, web directories are usually assembled by teams of editors who are subject specialists. Open directory is one of the largest web directory managed by more than 56,000 volunteer editors. Although, Web directories can often point users to relevant information with just a few mouse clicks, they often index only web sites not web pages in the depth of the sites. So, they cover smaller portion of information on the web. Summarizing the content of selected Internet resources, they let the end-user more easily and quickly find out whether the sites are related to his/her information needs or not.

Web directories have some advantages (such as, organized by team of editors, database of quality sites, easy and quick access to information through subject tree, retrieve related information, useful when searching on a broad topic) and disadvantages (fewer sites in the database, more dead links in database, much more expensive to construct and maintain, often not useful when searching for very specific topic in web pages).

### 3. Librarians as the Content Specialist of Web Directories

The environment in which librarians work is changing in terms of greater access to a range of information, increased speed in acquiring information, greater complexity in locating related information and analyzing and evaluation information. Librarians can efficiently accomplish many of the tasks related to design and development of web directories. We can see many web directory projects which librarians have built and maintained them. The traditional skills of librarians in searching, evaluating, indexing and abstracting and organizing the information can be used for creation and development of web directories. There are several important projects in which librarians have important role as the content specialist of web sites:

*BUBL LINK* is a catalogue of over 11,000 selected Internet resources covering all academic subject areas and cataloged according to DDC. All items are selected, evaluated, catalogued and described by academic librarians in UK. Links are checked and fixed each month. (11). *INFOMINE* is another librarian built directory of over 100,000 Internet resources relevant to faculty, students, and research staff at the university level. It is the efforts of over 20 librarians from the several universities and colleges in North America. All selected Internet resources are catalogued according to Library of Congress Subject Headings (12). *The Scout Report Archives* is a searchable and browseable database of more than 14,500 critical annotations of carefully selected Internet sites and mailing lists by professional librarians and content specialists. The selected sites are summarized and organized according to Library of Congress Subject

Heading (13). *The Librarians' Index to the Internet* is searchable, annotated subject directory of more than 10,000 reliable, trustworthy, librarian-selected Internet resources organized according to Library of Congress Subject Headings (14). Moreover, there are many annotated directories available in the libraries' web sites which are compiled by librarians to help users to locate and access useful Internet resources in various subjects.

All the above projects are known as the most important subject indexes on the web containing highly evaluated and useful sites in academic level. In creation of these databases, librarians had the most important role. The question is that why traditional skills and knowledge of librarians can well be used for creation and development of web directories and Internet subject indexes? We can seek the answer in the four main content management skills of information that librarians traditionally are familiar with them:

#### 3.1. Skill one. Effective Searching on Web

Traditionally librarian is familiar with formation of search strategy such as selection of databases, search terms, operators and fields. In the web environment, they can use past knowledge and experience to find related information and guide end users to them through a hyperlink structure. However, search engines usually retrieve piles of irrelevant information and this makes locating related information more difficult than traditional databases. Basically, many search concepts in traditional databases are the same in Internet search tools, such as using Boolean operators, field searching, truncation, nesting and search limits (language, date and etc). Librarians experience with end-users' information needs as well as selection of

appropriate search terms can help them perform more effective search for compiling high quality information for web directories. In other words, librarians are three, or four, or five times more efficient and more effective than end users are at performing their own information searching activities (15). This, probably, is one reason for why librarians can play a role in designing and development of web directories and subject indexes.

### *3.2. Skill Two. Evaluation of the Quality of Information on Web*

In a broad sense, librarians or information professionals evaluate information in order to judge quality and/or appropriateness of information content. Unlike printed world, on the web anyone can publish any information with any content. In other words, Internet is a media lacking the authority control structure of traditional print sources. Robot-based search engines retrieves web pages with any content and users should have the skill for evaluating Internet resources. But web directories suppose to be a database of evaluated web sites by experts. Although, many researchers have the skills for evaluating Internet resources on their related subject areas, librarians generally can apply some criteria of selection of print material in the libraries for evaluation of Internet sources. However, there are some criteria that differ. Some of these criteria include authority, reliability, credibility, structure, content, currency, usefulness, design, and special considerations. Some web directories such as Open Directory is maintained by community editors who evaluate sites for inclusion in the directory based on certain criteria and instruction.

### *3.3. Skill Three. Content Description of Web Sites*

Librarians are thoroughly familiar with abstracting and indexing print materials. Although, research is being conducted in the area of machine generated abstracts, there will continue to be a role for human abstractors. Summary notes about web sites or web pages would be highly useful for Internet users to assess content and value of retrieved documents. Some web directories use brief or full description of web sites. Unlike search engines which automatically generate description from texts or Meta Tags of web pages, directories use humans to describe the subject and content of the web sites. It is clear librarians with skill of abstracting can likely do this task more efficient and effectively.

### *Skill Four. Organization of Selected Web Sites*

Resources on the Internet have now grown to such an extent that there is an urgent need for their collection management and organization. Web directories or subject indexes are basically browsable through subject hierarchical structure. Under each main topic there's a list of sub-topics each of which may be linked to another set of sub-topics. The question is that who will construct this subject network and its referral links to related topics. Librarians and cataloguers can use cataloguing and organization knowledge and practices in the web environment to organize web sites. Thus, several web directories use DDC or LC classifications for organization of web sites and many others use their own subject scheme. Librarians subject approach to web sites would be more helpful to build and develop the subject tree as a main retrieval tool in web directories.

#### 4. Librarians' Experience in Building and Developing a National Web Directory

In October 2000, joint venture Iranian-German Company began its work in Iran to develop a national search tools for locating Iranian sites through Persian Interface. Primary study showed that there is no comprehensive web directory with Persian interface and the existing tools have many limitations in terms of quality of indexed web sites, subject hierarchical structure, search interface and options. In average more than 10,000 Persian language users search in the directory to locate related sites daily. Although, Persian interface of Google and Open Directory can be used by Iranian users, Google is a search engine and Open Directory have indexed small portion of Iranian sites without any advanced search options. On the whole the following limitations in existing web directories were the motivation for designing and developing a national web directory of Iranian web sites:

- Lack of Persian search interface.
- Lack of subject hierarchical structure for locating a list of related web sites.
- Lack of comprehensive and logical subject tree among existing directories.
- Lack of advanced search options for locating information.
- Lack of informative description about web sites.
- Lack of add or submit function for introducing web sites for inclusion in database.
- Lack of criteria for selection and evaluation of web sites in accordance with

cultural, social, political and national principals of Iran among existing directories.

##### 4.1 Team Selection

Two main sections were developed for designing and developing of Iranhoo directory: Content Section and Technical Section.

##### 4.1.1. Content section

Librarians as content specialist were responsible for doing following tasks:

- *Searching and locating Iranian web sites through search engines, web directories or other web-based resources.* Librarians responsible for locating Iranian web sites were well-familiar with search tools and techniques on the web.
- *Content evaluation of located web sites based on approved criteria.* Librarians in this section were responsible for content evaluation of web sites in terms of usefulness for Iranian users, overlapping and repetitive content with other sites not useful for users. The sites with illegal contents in Iran including pornography, illegal political activities were not included. Under construction sites, sites with many dead links and heavy graphical designs that make them very slow to retrieve were ignored too.
- *Determination of main-categories and related sub-categories based on hierarchal subject tree.* Librarians in this section use Persian Subject Headings of the National Library of Iran (NLI) and other Persian thesauruses to determine categories for inclusion of located and evaluated Iranian web sites.
- *Description of content of web sites.* As mentioned earlier, librarians are familiar with abstracting and indexing techniques, so they can describe the content and subject of

web sites more effectively and make a appropriate decision on which category the web sites should included in. In the Content Section, librarians have critical role for writing informative description about selected web sites.

- *Determination of search interfaces and options.* Comments and recommendations of librarians who are in fact familiar with searching in traditional databases on CD-ROM and Internet were important for designing search capabilities and interface of the web directory.

#### 4.1.2. Technical Section

Computer experts were responsible for doing following tasks:

- Programming for database and other parts of the site.
- Computer graphical design such as icons, buttons, search interface, images and etc
- Network management and security.
- Data transfer and control.

### 4.2. An Overview on Iranhoo Web Directory

**5.1.1. Size of the Database.** Until April 1, 2003 about 4,000 Iranian web sites were located, evaluated and added to the Iranhoo database. Most of the selected sites are in Persian. Comparing Iranhoo's database size with other international web directories such as Open Directory with 1700 sites and Yahoo with 1200 sites indicated that it has a larger database among popular web directories. Every week, more Iranian sites are located or submitted and added to the database.

**5.1.2 Currency of Information.** While it is very difficult to review all added sites in a certain period of time, librarians check the sites in the database every two month in terms of possible content change or accessibilities of the URL. If the content of the added site change, they write an updated description about. To check the possible change of URLs or disappearance of the added sites, a robot is used to identify and remove dead links. If the site was not found by robot within 10 days, the system automatically would remove the site permanently from the database. The system shows last updated time of each retrieved entry (web site) in front of it. So, the user can easily find out when the site is located and described by editors.

**5.1.3. Number of Categories.** There are 18 main categories in the first page of the directory including Education, Business, Health and Medicine, Entertainment, Society and Culture, Transportation, Home, News and Media, Government, Social Science, Science, Arts and Humanities, Computer, Kids and Teens, Reference, Islam, Sports, and Tourism. Totally, there are more than 510 sub-categories that lead the users to the list of related web sites. To access web sites related to some narrow topics, users should browse through 3 or more subject layers. For example to locate online directories on Iranian Oil and Gas companies a user should browse through following links: *Business> Industries> Energy> Oil and Gas> directories.* Like other major web directories on the web, librarians may add, remove or edit sub-categories everyday.

**5.1.4. Methods for Locating Web Sites.** There are two methods to locate new web sites: Searching or browsing in other related search engines and web directories or studying print media to identify new Iranian

web sites. To accelerate the process of locating and adding new web sites, users or web sites owners can themselves register sites through *Submit a URL* function. Thus, anyone can enter title, URL and description of the desired site and submit it to the web directory in order to be evaluated by librarians and added to the database. Users can also select and suggest a sub-category related to the content of the submitting web sites.

**5.1.5. Content Description of Sites.** As mentioned earlier, librarians write a short description (15-30 words) about the content of each site. It seemed very difficult to introduce the whole parts of the selected sites in one or two sentences. So, librarians with practical experience of abstracting and indexing were selected to do this job. They tried to write an informative description about each sites to help end-users make a better decision about subject and content of the retrieved sites.

**5.1.6. Search Capabilities.** Although, browsing subject categories is the main information seeking method in web directories, many major directories have designed search interfaces in basic or advanced mode to help users find related sites or categories. Iranhoo has two basic and advanced search interfaces in Persian. Basic search supports all Boolean operators (AND, OR and NOT). The search default is AND operator. While it is impossible to use explicit Boolean operators AND, OR and NOT in Persian search interface, the signs +, || and - were used to perform the operators accordingly.

In the advanced search, keywords can be searched in the title, description and URL field separately. Through pull-down menu, users can use operators among fields. Moreover, the signs +, || and - can be used

to perform related operators just like basic search mode. There is an option to perform Exact Word Search in advanced search mode. As a default, the system use stemming function to retrieve related words. Users can also search within retrieved results. Moreover, as a default, keywords are searched in the title and description of sites as well as keywords in categories. So, users can either select related categories or sites. In preferences section users can set the results per page from 5 to 50.

**5.1.7. Displaying of Information.** After initial search, the total number of hits is displayed. The Number of indexed sites under each main or sub-categories is shown too, so users can find out how many sites are listed under each categories. Each entry (indexed site) consists of title, description and URL of sites. Last updated time of each indexed sites and visited times of them by users are shown.

The ranking algorithm of sites in each category is based on visited times by Iranhoo's users. In other words, a site with higher hits in the directory, displays higher in the related category which is indexed. The system not only automatically generates a ranking list in each category, but also shows the most popular (visited) Iranian sites in the entire database. So, users can easily see the top 10, 20, 30 or more visited sites which are searched and visited by directory's users. In the first page of the site the total number of categories and indexed sites as well as new added sites within a month is shown.

**5.1.8. Other Sections of the Site.** Although, Iranhoo Web Directory is the most important and visited section of the site, there are several other services which basically are developed by efforts and opinions of librarians, including online searchable database of Iranian companies,

searchable database of Iranians E-mail addresses, Kids and Teens as well as free online magazines on Health, Computer and Internet, Science and Technology, Business and Economics and Sports in Persian. The content of all above sections are provided by librarians.

## 6. Conclusion

Librarians may be developing a tendency to use their traditional skills and knowledge to embrace the ever-growing internet resources, such as searching, evaluating, and knowledge management. Building and maintaining web directories is one way in which librarians' traditional knowledge and skills are being applied in a non traditional context. But these skills and knowledge are still needed to be modified to develop new methods for searching; evaluating and selecting Internet resources. The challenge of cyberspace offers librarians the opportunity of partnership to make a database of selected and evaluated web sites. Since they use subject cataloguing, information retrieval and selection of print materials in their traditional activities, they can offer similar services in the Internet environment. In other words, one way to *strengthen information content* on the web is building librarians compiled web directories.

*Iranhoo Web Directory* is one of the successful example of using skills and knowledge of librarians to build and develop a national largest database of selected Iranian web sites with Persian language interface. Lack of a national search tool in many developing countries has caused users search their desired keywords in other languages. Although, there are several international popular search engines (such as

Google, AltaVista and All the Web) with multi-language search interface, the lack of national search tools in many developing countries has caused users search their desired keywords in other languages. Moreover, any country has its own cultural and social values which want to be preserved. The question is that if searching and retrieving in popular search tools can *guarantee culture and values* of nations especially with non-English languages and specific information needs. For instance Open Directory Project where more than 53,000 volunteer editors help it to build the largest human-edited web directory with more than 3.8 million selected web sites. It is not far reaching that librarians in any countries working in different fields of science begin to develop national directories or subject indexes and provide appropriate grounds for cultural exchange. Planning *strategic alliances with various user communities* like researchers and content specialist to locate and evaluate web sites can definitely be helpful to build a database of high quality web sites. Through this network of editors with the same cultural and social backgrounds, social and cultural responsibilities of libraries in bridging the gaps, searching and locating related web sites in non-English language, can be filled. If cultural expression is to be a reality on the Internet, it must be made available to the majority. Cultural values needs to be organized so that it can be made available through the Internet. Building multilingual interfaces in the national web directories can expand cultural and social exchanges among nations. This is one of the ongoing projects of Iranhoo web directory.



## References

1. M. Landoni and S. Bell. *Information retrieval techniques for evaluating search engines: a critical overview*. *Aslib Proceedings*, 52 (3) Mar 2000, p.124-129.
2. C. Hume. Internet search engines and robots: what they are and how to use them. *Journal of Internet Cataloging*, 2 (3/4) 2000, p.29-45.
3. K. I. Munson. *Internet search engines: understanding their design to improve information retrieval*. *Journal of Internet Cataloging*, 2 (3/4) 2000, p.47-60.
4. S. J. Clarke. *Search engines for the World Wide Web: an evaluation of recent developments*. *Journal of Internet Cataloging*, 2 (3/4) 2000, p.81-93.
5. C. Oppenheim and et al. *The evaluation of WWW search engines*. *Journal of Documentation*, 56 (2) Mar 2000, p.190-211.
6. H. Chu and M. Rosenthal. *Search engines for the World Wide Web: a comparative study and evaluation methodology*. Proceedings of the 5Ninth Annual Meeting of the American Society for Information Science, 1996, p.127-135.
7. C. Schwartz. *Web search engines*. *Journal of the American Society for Information Science*, 49 (11) Sep 1998, p.973-982.
8. A. Griffiths. *Setting up a subject directory of Web sites: a case study of Management Link*. *Library Management*, 20 (1 and 2) 1999, p.12-18.
9. S. Gandhi. *Proliferation and categories of Internet directories: a database of Internet subject directories*. *Reference and User Services Quarterly*, 37 (4) 1998, p.319-31.
10. G. R. Notess. Comparing net directories. *Database*, 20 (1) Feb/Mar 97, p.61-64.
11. BUBL Links. [Online]. Available: <<http://bubl.ac.uk/link/>>. [2 April 2003].
12. INFOMINE. [Online]. Available: <<http://infomine.ucr.edu/>> [2 April 2003].
13. The Scout Report Archives. [Online]. Available: <<http://scout.wisc.edu/>> [1 April 2003].
14. Librarians' Index to the Internet. . [Online]. Available: <<http://www.lii.org/>> [1 April 2003].
15. J. Griffiths. *The changing role of librarians: managing new technologies in libraries*. In *Library and information services in astronomy*, Garching, Germany, 10-12 may 1995. Available: <<http://www.eso.org/gen-fac/meetings/lisai/papers/griffiths/josemarie-griffiths.html>> . [10 Jan 2003].