Workshop on
Multimedia and Internet Technologies
26th-28th February, 2001
DRTC, Bangalore

Paper: CB

Web-based company directories: A case study with IT industry

Umakant Gupta and K.T. Anuradha

NCSI, IISc, Bangalore 560 012

Abstract

The World Wide Web (WWW) is a popular platform for delivering digitized information. All types of information sources are available on WWW such as primary sources (e-journals), secondary sources (bibliographic databases) and tertiary sources (directories). Web-based information sources have some specific features compared to other forms (print, CD-ROM). In this paper we have discussed the steps involved in setting up a web-based company directory of Information Technology (IT) industry.

As the first step, a few important web-based company directory services are analyzed. Based on the analysis, meta tags, indexing and search features required for effective web-based company directory service are defined. About 50 company web pages related to IT industries are retrieved by surfing the Net. The collected information is compiled and mounted as a MYSQL database. PERL is used to access the database. The sample directory has features such as hierarchical browsing, and field based searching.

1   Introduction

The World Wide Web (WWW) is a popular platform for delivering digitized information. It allows users to access the information services easily, but the users face the challenge of finding and using information that is accurate and reliable. Browsers such as Netscape and Internet Explorer have demystified the Internet, and made its contents accessible to users who have minimum technical expertise. Consequently, the notion of the Internet as a virtual library available at the click of a mouse is becoming increasingly attractive, particularly to libraries with limited resources and small collections. On the web all type of sources are available like primary sources (E-Journals), secondary sources (Bibliographic databases) and tertiary sources (Directories). However, using the web to access accurate and reliable information is a complicated proposition.

2   What are Company Directories

Company directories are sources that are defined as a list of companies arranged systematically, usually in alphabetical or classified order giving details such as address, phone number(s), key persons, financial and product information.

Company directories are usually available in the form of multi-volume printed sets. Since they are both cumbersome and expensive, the printed directories are typically available only in libraries. Other forms of company directories, such as on CD-ROM, are available, but they are also very expensive. Since past few years, company directories have become available on the Web. As a result, more people from outside the libraries can access the directories, and there have been substantive changes in the directories themselves. For example, some directories have added fields such as URLs, key persons and financial information of the companies. Another example is the added capacity to search by different approach like company name or region or products/services or categories or ticker or full text of the directories. Contrary to this, print version typically offers only two or three avenues of access i.e. by company name or region. So one result of the new

web-based environment is that company directories are now more easily accessible to the general public in terms of cost and availability. Another result is the new features that have been added which respond to the possibilities provided by the new technological environment.

3   Locating the Company Directories on the Web

There are many company directories available on the web, some are freely available and some are on subscription basis. Following is a sample list of web-based company directories.

    a)   Hoover's Online

        http://www.hoovers.com/company/dir/0,2116,4798,00.html

    b) Corporate Information *www.corporateinformation.com*

    c) Thomas Register *www.thomasregister.com*

    d) Kompass *www.kompass.com*

    e) Lycos Companies Online *www.companiesonline.com*

    f) AllBusiness.com *www.comfind.com*

4   Analysis of Web-Based Company Directories

To judge the impact and reliability of web-based company directories, first three sites are chosen for comparative analysis. These web-based company directories are evaluated both as web sites and informational databases. Here are the some categories chosen for this analysis with examples of types of questions asked during the assessment.

4.1   Authority: Who is responsible for this information? What is the authority/expertise of person or group who created the site?

4.2   Accuracy: How reliable and accurate is the information and data source? Is it clear who has the ultimate responsibility for ensuring accuracy?

4.3   Objectivity: Is the information biased? Are the goals of the site clearly stated? Who is the intended audience?

4.4 Currency: When was the information created and when is it updated? Are the links up to date?

4.5 Coverage: What is the extent and depth of the information? If there are print equivalents, does the coverage differ?

4.6 Accessibility: How easy/difficult is it to access the site? How easy/difficult is it to find information? How flexible are the searching options?

4.7 Search Features: Which types of search features are provided by the directories?

4.8 Browsing Features: Is the browsing features provided by directories? If yes, then through which fields browsing facility is available.

4.9 Help: Is the help facility is available? Is it user friendly?

The following table shows the comparative analysis of web-based company directories.

Company Directories	Authority	Currency	Coverage	Contents	Accessibility
Hoover's Online	Provides information about sponsor	No information	15,000 (companies all over the world)	Exhaustive information	Free & subscription
Corporate Information	,,	Weekly	3,00,000 (companies all over the world)	All basic information (name, Address, Financial information)	Free
Thomas Register	,,	No information	1,56,194 (companies only American and Canadian)	Brief information (name, products, address)	Free

Company Directories	Boolean search	Feedback search	Fields search	Browsing	Truncation
Hoover's Online	And	Yes	Name, Co.keywords people, news -----	Yes(five approaches)	No
Corporate Information	And, or	No	Name, ticker	Region, categories	Automatic truncation
Thomas Register	And,or, not	Yes	Name, brand, product or service	No	No

Company Directories	Exact match	Advanced search	Help
Hoover's Online	Yes	Yes (only for members)	Yes(not examples)
Corporate Information	No	No	Yes(with examples)
Thomas Register	Yes	No	No

5   Features of Web-Based Company Directories

On the basis of analysis, following features of web -based company directories are defined:

5.1 Meta Tags (Content Information)

➢ Contact Information: Company name, Contact person, Full address of the company, E-Mail address, Phone number, Fax number, Toll free number.

➢ General Information: Number of employees, Year established, Symbol.

➢ Description (Activities of Company)

➢ Web site: Home page address, Products/services, Downloads, Site search (with link)

➢ Key persons: Chairman, President, Chief Financial Officer, Secretary.

➢ Financial Information: Market Capitalization, Sales and Rate of Growth, Annual Report, Stock Chart, Earnings / Dividends, Currency, Fiscal Year Ends.

➢ Top Competitors.

➢ Products and Services defined using Primary Industrial Code

5.2 Indexing Features

According to ANSI, "An index is a systematic guide to items contained in, or concepts derived from, a collection. These items or derived concepts are represented by entries arranged in a known or stated searchable order, such as alphabetical, chronological, or numerical."

To provide proper guidance to find out the required company information, approaches should be provided by company direc tories through company name, ticker, company category, products or services, region. The following features are also considered to fulfill these approaches.

➢ Use of product and service catalogues to standardize the product and service headings.

➢ Provide hierarchical relations among the products and services headings.

➢ Index the company name in as many ways as possible.

5.3 Search Features

Following search features should be available in the company directory:

➢ Boolean searches.

➢ Feedback searches.

➢ Exact matching.

Directories should also support browsing features through categories of company, region as well as the first alphabet of company name so that users can easily find the required company.

6   System Design

Setting up a web-based company directory demands minimum hardware and software. In this study, MYSQL is used for creating the company database and PERL scripts have been written in RH Linux 6.0 to search and browse the company database. For setting up web-based company directory of IT industry following steps have been taken.

6.1 Collect the Required Metadata

To create any database, metadata is required. In this study, the web is used as an information source (company web site and other sources) to collect the required company data such as company name, key persons, financial information, products and services.

6.2 Classify all IT Companies

Classification of IT companies has been made to provide the searching and browsing facility through particular class of the companies. In this study nine broad categories of the IT companies are identified and each category is again classified into subcategories. The following is the list of broad categories:

1) IT Service Providers.
2) Software Products.
3) Hardware Products.
4) Associations & Agencies.
5) Education & Training.
6) Manufactures & Distributors.
7) Internet.
8) Publications.
9) Telecommunication.

6.3 Database Design

MYSQL, a SQL (Structure Query Language) database server that is in public domain, is used to create company database. MYSQL is a client server implementation that consists of a server daemon mysqld and many different clients, programs/libraries.

For this study, one database "directory" is created. Since MYSQL does not support foreign key concept, for providing hierarchical browsing to mysql database, twenty-three tables (for each broad categories and subcategories which has

subdivisions) including one table for company information are created under the database "directory". The structure of all the category tables (twenty-two tables) is the same. An example of category table is given below. The following example shows the record fields for the IT Services category and company information.

| ID	Number for identification of a particular record
CODE	Number for identification of group of subcategory
SERVICE	Name of subcategory
LINK SERVICE	URL address for providing the link to subcategories
HIERARCHY	Name of broad category

Table 1: Record fields for category table

ID	Number for identification of records
NAME	Name of company
ADDRESS	Address of company
PHONE	Phone numbers of company
EMAIL	E-Mail address of company
DESCRIPTION	General information about company
KEY PERSONS	Name of the key persons of the company
LINK MAP	URL address of the map of the company
LINK ANNUAL REPORT	URL address of the annul report of the company
LINK STOCKCHART	URL address of the stock chart of the company
LINK COMPANY	URL addresses of the company web sites
COMPETITORS	Name of the competitors of the company
LINK COMPETITORS	URL addresses of the competitors web site
SYMBOL	Unique symbol of the company
ESTABLIS HMENT YEAR	Establishment year of the company
FISCAL YEAR END	Financial end month of the company
PRODUCTS/SERVICES	Products and Services of the company
CODES	Codes for the identification of the categories of the company

Table 2: Record fields for company information table

6.4 Search Interface

Using PERL and CGI the user interface is created for searching and browsing of the database. The search can be done by company name, products and services or category and this search can be restricted by country and browsing can be done by category of the company and first alphabet of company name. The database supports truncation and search is case insensitive in nature.

The following are the screenshots of the search interface and the corresponding search result:
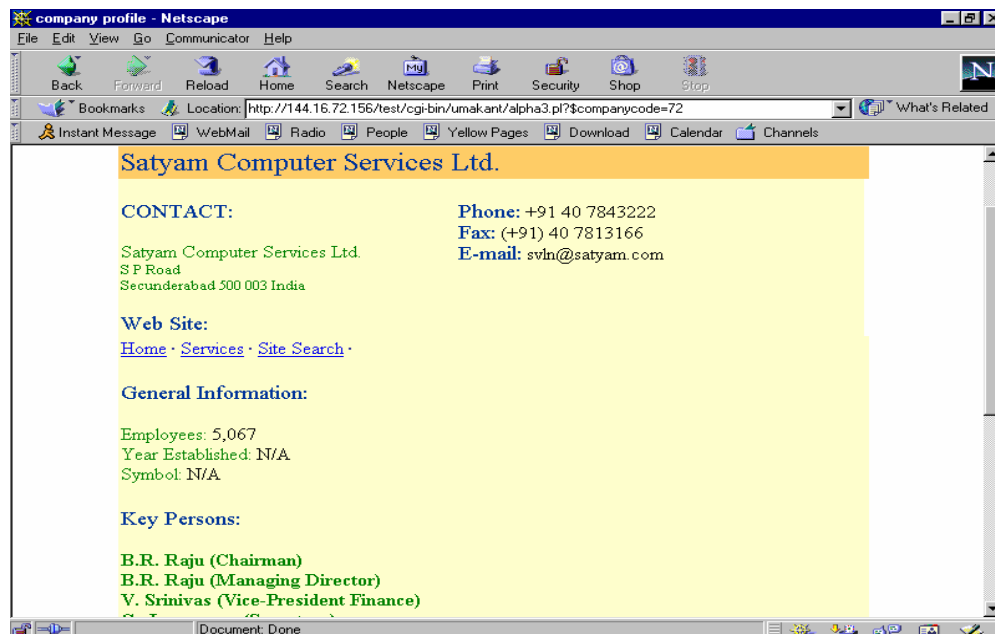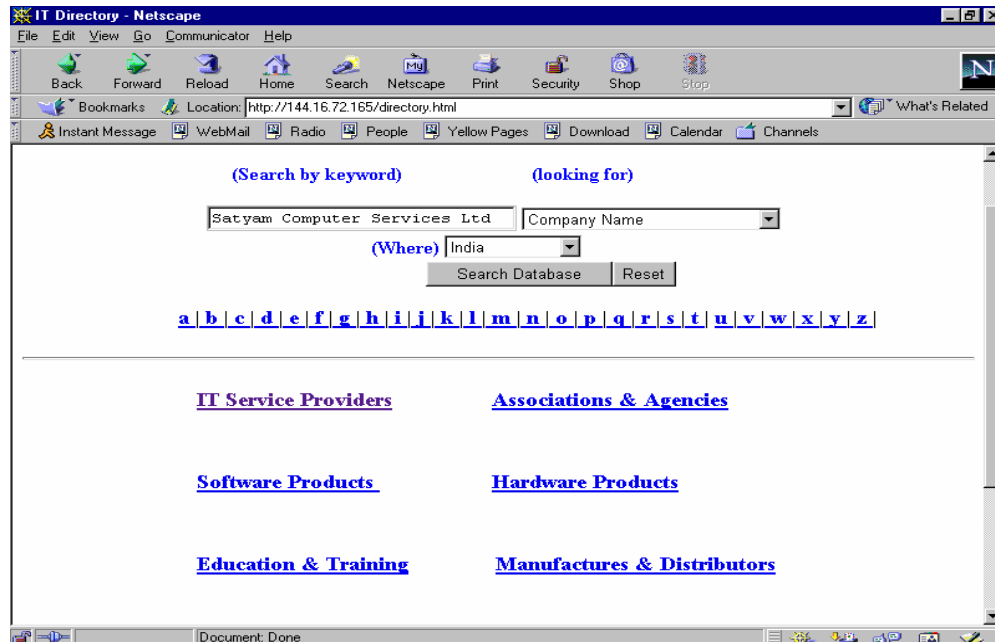
7   References

1. ARMSTRONG (Chris J.). Database Quality Criteria.

    Gopher://ukoln.bath.ac.uk:7070/00/BUBL_Main_Menu/H/H2/H2C28/H2C280 02_Database_Quality_Criteria

2. GRASSIAN (Esther). Thinking Critically About World Wide Web Resources.*www.library.ucla.edu/libraries/college/instruct/critical.htm*

3. Devshed-Mysql. *www.devshed.com/Server_Side/MySQL*

4. Learning Perl Made Easy!! *http://ucat.ncsi.iisc.ernet.in/ tutorial/perl /perlmain.html*