# Approaches To Indexing In The UK

Brian Kelly, UK Web Focus, UKOLN, University of Bath, Bath, BA2 7AY
Email: <B.Kelly@ukoln.ac.uk> URL: <http://www.ukoln.ac.uk/web-focus/>

## Abstract

*This paper discusses the approaches made in indexing institutional web sites in two public sector communities in the UK: Higher Educational Institutions (HEIs) and Public Libraries. The paper reports on surveys made of these communities and discusses the findings. A summary of other development work is given. A summary of other development work is given, including volunteer initiatives and projects funded by the eLib Programme and JISC.*

## Background

When choosing software to index an organisational web service you may choose to read reviews in Internet magazines, attend trade shows and read documentation provided by the software vendors. However it can also be useful in seeing the approaches taken by similar organisations. It can be helpful to see if there is a clear leader within your community – for example you will be able to see if you organisation is being left behind.

In July /August 1999 a survey of indexing software used in UK University web sites was carried out. A similar survey of UK Public Library web sites was carried out in January 2000. The results of these surveys are freely available and are intended to provide a useful resource for these communities.

## Survey of UK Universities

A survey of UK University and University College web sites was carried out in July / August 1999. The survey made use of the HESA list of University and University College web sites [1]. The results of the survey [2] and a report [3] have been published.

A total of about 160 University and University College web sites were surveyed. Since the initial report was published information concerning a number of changes has been received. An updated summary has been published [4]. A brief summary of the latest findings is given in Table 1.

| Indexing Software | Total |
|---|---|
| ht://Dig | 32 |
| eXcite | 17 |
| Microsoft | 15 |
| Ultraseek | 9 |
| Harvest | 7 |
| SWISH / SWISH-E | 4 |
| Webinator | 4 |
| Netscape | 3 |
| wwwwais | 3 |
| Other | 13 |
| None | 50 |

**Table 1: Indexing Software Used in UK University and College Web Sites**

### Discussion of Findings

It is perhaps surprising, as the UK Higher Education community was an early adopter of the Web, that about 30% of web sites appear not to provide a search facility. Although the total may not be quite this high, since a search facility may be available which was not found in the survey, it is unlikely that the numbers differ significantly from those given.

The most popular product is **ht://Dig** [5] which is used by 32 institutions (up from 25 in the original survey in August 1999). This software is freely available, and a new version was released in December 1999. It uses a robot which enables multiple servers to be indexed.

The **eXcite** [6] software is used by 17 institutions (down from 19 in the original survey). This software is also freely available. However the eXcite web pages have not been updated since January 1998, when a security warning was given.

**Microsoft** [7] software is used by 15 institutions (up from 12 in the original survey). Several products are available, which are freely available (e.g. Index Server) or bundled with a server product (e.g. SiteServer).

**Ultraseek** [8] is also used by 9 institutions (up from 7). Ultraseek is a licensed product, which is expensive, but is also very powerful.

**Harvest** [9] is used by 8 institutions (down from 8 in the original survey). Harvest is freely available.

Three institutions made use of third party services to index their web site. Two institutions made use of **FreeFind** [10] and one used the public **AltaVista** search engine [11].

## Survey of UK Public Libraries

The survey of UK Public Library web sites was carried out in January 2000. The survey made use of the Harden list of Public Library web sites [12]. The results of the survey have been published [13].

A total of 137 Public Library web sites were surveyed. A brief summary of the findings is given in Table 2.

| Indexing Software | Total |
|---|---|
| Microsoft | 18 |
| Domino | 3 |
| Muscat | 3 |
| eXcite | 3 |
| ht://Dig | 2 |
| Other | 16 |
| None | 98 |

**Table 2: Indexing Software Used in UK Public Libraries**

### Discussion of Findings

Perhaps the most surprising finding from the survey was the large number of web sites (49%) which did not appear to provide a search facility.

Of the web sites which provided a search facility, 45% made use of **Microsoft** indexing software. **Lotus Domino** [14] is used by 3 public libraries. This is a licensed product, which is part of the Domino server. **Muscat** [15] is used by 3 public libraries. This is also a licensed product.

Public Library web sites differ from University web sites in that a Public Library web site is often part of a Council web site. A Public Library web site will often use the search facility provided by the Council web site. In many cases it was not possible to restrict a search to the Public Library area of the Council web site.

## Comparisons

The UK Higher Education community has been involved in web developments since the early days of the web. This community is often able to make use of good technical resources, such as postgraduate students. The community is keen on use of open source software.

Public libraries in the UK, in contrast, have embraced web technologies more recently. Although it has technical expertise to implement OPAC systems, it does not have the range of technical expertise available in the HE community. The Public Library community appears to prefer shrink-wrapped solutions, often running on an NT platform.

# Other Developments

## Volunteer Initiatives

**ACDC** [16] provides an interesting example of an unfunded project to provide an index of the UK Higher Education community. ACDC relied on volunteer effort to use Harvest to provide a distributed index of resources. Unfortunately it appear that ACDC is no longer being developed.

A number of interesting developments have started within institutions. **Maestro** [17] makes use of a robot developed for the OS/2 platform to provide an index of Scottish resources.

The North East Universities [18] provide what appears to be a cross-searching service across Universities in the north east, although this is, in fact, an interface to the AltaVista and HotBot public search engines.

## eLib Developments

Within the UK Higher Education community the eLib Programme [19] has been instrumental in much of the development work in the area of Digital Libraries. Phase 3 of eLib is concentrating in the development of "Hybrid Libraries" which will enable users to find resources not only on web sites, but also other electronic resources (e.g. OPACs) and "real-world" resources (e.g. books, items in museums and special collections, etc.). The Hybrid Library projects do not limit themselves to resources held within institutions, but may have a regional or subject-based perspective. **MusicOnline** [20], for example, enables users to search for music resources throughout the country and **BUILDER** [21] provides a search across other Phase 3 projects.

## Commercial Developments

There is an argument that, rather than developing an infrastructure for searching across UK University web sites, we should simply make use of commercial services which provide national searching facilities, such as. UKmax [22] or SearchUK [23]. However it is not certain that such services would be interested in engaging in discussions with the community over the communities' specialist requirements.

## JISC Initiatives

JISC are developing the DNER (Distributed National Electronic Resource) [24] which aims to provide seamless access to electronic resources available on a variety of national services, such as MIMAS, NISS and BIDS. The DNER approach focuses on the importance of standards, including standards such as Dublin Core, Z39.50, LDAP, etc.

An example of a JISC service which will be a part of the DNER is the RDN (Resource Discovery Network). The RDN provides an example of seamless access to disparate resources though its cross-searching demonstrator [25].

# Conclusions

This paper has given an overview of the approaches taken within the UK Higher Education community to enable members of the community to find resources provided by the community or of direct relevance to the community. We have seen the approaches taken within institutions to the provision of search facilities across institutional web sites. We then discussed a number of volunteer initiatives aimed at providing search facilities across regions or across the country. We then described eLib Phase 3 projects which are addressing the needs of end users to find resources, which may be located on a web site, within a backend database or OPAC, or may be a physical resource, such as a book. We concluded by mentioning the DNER which aims to provide seamless access to distributed national electronic resources.

# References

1   *Higher Education Universities and Colleges*, HESA
    <URL: http://www.hesa.ac.uk/links/He_inst.htm>
2   *Survey Of UK HE Institutional Search Engines - Summer1999*
    <URL: http://www.ariadne.ac.uk/issue21/webwatch/survey.html>
3   *WebWatch: UK University Search Engines*, Ariadne issue 21
    <URL: http://www.ariadne.ac.uk/issue21/webwatch/>
4   *Survey Of UK HE Institutional Search Engines – April 2000*
    <URL: http://www.ukoln.ac.uk/web-focus/surveys/
    uk-he-search-engines/survey-apr2000.html>
5   *ht://Dig,* <URL: http://www.htdig.org/>
6   *eXcite,* <URL: http://www.excite.com/navigate/>
7   *Microsoft,* <URL: http://www.microsoft.com/>
8   *Ultraseek,* <URL: http://software.infoseek.com/>
9   *Harvest,* <URL: http://www.tardis.ed.ac.uk/harvest/>
10  *FreeFind,* <URL: http://www.freefind.com/>
11  *AltaVista*, <URL: http://www.altavista.com/>
12  *The UK Public Libraries Page*, <URL: http://dspace.dial.pipex.com/
    town/square/ac940/ukpublib.html>
13  *Survey of UK Public Library Search Engines*, <URL: http://www.ukoln.ac.uk/
    web-focus/surveys/pub-lib-search-jan-2000/survey.html>
14  *Lotus Domino*, <URL: http://www.lotus.com/home.nsf/welcome/domino/>
15  *Muscat*, <URL: http://www.muscat.com/>
16  *ACDC*, University of Kent at Canterbury, <URL: http://acdc.hensa.ac.uk/>
17  *Maestro,* University of Dundee, <URL: http://somis.ais.dundee.ac.uk/search/
    www-gtw?server=Search+the+University>
18  *Unis4ne*, <URL: http://www.unis4ne.ac.uk/>
19  *eLib*, <URL: http://www.ukoln.ac.uk/services/elib/>
20  *MusicOnline*, <URL: http://www.musiconline.ac.uk/>
21  *BUILDER*, <URL: http://www.builder.bham.ac.uk/>
22  *UK Max*, <URL:  http://www.ukmax.co.uk/>
23  *SearchUK*, <URL: http://www.searchuk.co.uk/>
24  *DNER*, JISC, <URL: http://www.jisc.ac.uk/pub99/dner_desc.html>
25  *RDN*, <URL: http://www.rdn.ac.uk/>