

**The Use of Search Engines, Web Directories and Indices on the World
Wide Web for Architectural Research in Nigeria:
Resources and Methods**

By

**Dr. Olu Ola Ogunsote and Dr. (Mrs.) Bogda Prucnal-Ogunsote
Department of Architecture, School of Environmental Technology
Federal University of Technology, Akure**

Paper presented at the

Year 2003 Annual General Meeting and Conference

Of the

Association of Architectural Educators in Nigeria (AARCHES)

Theme: Methodological Issues in Architectural Research

Held at the Ahmadu Bello University, Zaria

September 24 – September 27, 2003

Abstract

The Internet has become an increasingly popular tool for architectural research in Nigeria, despite the fact that the information obtained from the World Wide Web often has little local content. This is partly because there are few web sites maintained by Nigerian architects about Nigerian architecture, and partly because finding the little that is relevant to Nigeria is made difficult by the sheer expanse of the Web. While the number of Nigerian sites on the Web grows daily, finding information on the Web is not getting easier. This paper discusses the three primary tools used to search for information on the Web: search engines, indices and directories. The choice of a search strategy and the characteristics of selected search engines and directories are discussed. Several resources for e-books, e-journals, digital collections as well as online sources of printed books and periodicals are provided. It is concluded that architects need to put more information on the web while the logic of search engines undergoes improvement.

What are search engines?

A search engine is a software tool that crawls the Web searching for information. The search engine identifies which web sites (pages) on the Internet contain the relevant information. This makes looking for information on the Web easy for a researcher, but the quality of information retrieved depends on the choice of, and proper use of the search engine.

How Search Engines work

A search engine builds a database of the contents of the Web by retrieving documents and indexing them. Search engines search the Web using software tools called robots, spiders or crawlers. They automatically follow hyperlinks from one document to the next around the Web. When a new site is discovered, information is sent back to the main site for indexing. Search engines regularly search the Web to update their indices because of the fast rate at which information posted on web sites change. When a query is entered into a search engine, the query is checked against the search engine's indices and the best matches are returned as hits.

Keyword and Concept Searching

The two primary methods of text searching are keyword and concept searching. Keyword searching is the most common, and the search engines determine which words on a web site are most significant. Usually, words that are found at the top of a document, titles, headings, subheadings and words that are repeated several times are indexed as keywords. Some search engines however index all words. Keyword searching can however produce inaccurate results since it does not consider the order of words, singular and plural forms, verb tenses, and similar words.

Concept-based search engines such as Excite try to determine the meaning of the query. They are therefore capable of returning hits on documents that are about the subject of interest, even if the words in the query do not exactly match the words in the document. These engines use clustering, a technique of examining words in relation to other words found nearby. The frequency of occurrence of several words or phrases that are tagged to signal a particular concept can then be used to conclude that the document is about a

particular subject. While smarter than keyword searching, concept based searching is not artificial intelligence and is far from perfect, relying mainly on statistical analysis. The best results are obtained when a lot of words are entered in the search.

Getting more accurate results

The accuracy of results can be improved by refining the search. A basic search uses keywords without additional options. A refined search allows the use of Boolean operators such as AND, OR, and NOT as well as proximal locators such as NEAR and FOLLOWED BY. Phrases can be searched for by enclosing the phrase in quotation marks, in which case an exact match with the phrase is found.

Relevancy Ranking

A confidence or relevancy ranking is an indication of how closely the search engine thinks the results match the query. The most useful results should therefore have the highest ranking and should be placed at the top of the list. This is critical considering that many queries return hundreds, if not thousands or millions of web pages. Unfortunately, relevancy rankings are often unsatisfactory, though there are attempts to improve the logic. Some engines use both the frequency and positioning of keywords to determine relevancy, while AltaVista allows advanced queries that assign relevance weights to query terms. See Tables 1, 2, 3 and 4

Table 1: Characteristics of selected popular search engines. Source: www.monash.com.

Search Engine	Type of search / options	Domains searched	Relevance ranking	Rating
Alta Vista	Keyword search, with simple, advanced and refined options	Web, Usenet	Ranks according to how many of your search terms a page contains, where in the document, and how close to one another the search terms are	A fast, powerful search engine with enough bells and whistles to do an extremely complex search. Fast searches, capitalization and proper nouns recognized, largest database; finds things others don't. It will search on both words and on phrases, including names and titles. However multiple pages from the same site show up too frequently; some curious relevancy rankings, especially on Simple search.
Excite	Concept and keyword searches with simple and refined options	Web, Usenet and classified ads	Confidence percentile provided on all searches	An intelligent search engine that uses concept-based searching. Has a large index and excellent summaries. Hits can be viewed grouped by confidence or grouped by web site. However it does not specify the format, size or total number of hits.
Google	Keyword search	Web	Frequency ranking	Google is thorough and fast. It's technology considers pages that are linked to from other sites to be more important than pages that only have a few links from other sites. In other words, if many webmasters consider a website valuable enough to create a link to it, Google considers that a good reason to justify a high ranking for that site
Hotbot	Keyword search with simple, modified and expert options	Web	Search terms in the title are ranked higher than search terms	Claims to be fast because of the use of parallel processing, which distributes the load of queries as well as the database over several work stations. Some limitations still on Boolean operators, and the

Search Engine	Type of search / options	Domains searched	Relevance ranking	Rating
			in the text.	help files still aren't very good.
Infoseek	Keyword search, with simple but powerful options	Web, Usenet, Usenet FAQs, Reviews, Topics	Web, Usenet, Usenet FAQs, Reviews, Topics	Infoseek has added an extensive catalogue section for subject-oriented searching. You can also cross-reference your search terms with similar catalogue subject items and searches come back with subjects automatically appended. You can also search images. Fast, flexible, reliable searching. Good output, which gives the URL, the size of the document and the relevancy score. Allows you to see similar pages. Full-text indexing, allows capital letters and phrases.
Lycos	Keyword search, with basic and advanced options. Can search on image and sound files.	Web, Usenet, News, Stocks, Weather, Multi-media	None	Lycos is gradually becoming less of a search engine, it seems, and more of a Yahoo-like subject index. Large database. Comprehensive results given--i.e., the date of the document, its size, etc. Lycos indexes the frequency with which documents are linked to by other documents to make sure the most popular web sites are found and indexed before the less popular ones.
Webcrawler	Keyword search, with simple and refined options	Web, Usenet	Frequency calculated -- computes the total number of times your keywords appear in the document and divides it by the total number of words in the document. Webcrawler returns surprisingly relevant results.	Easy to use. Popular on the Web because it belongs to AOL and there are a lot of websurfers who sign on from AOL. Publishes usage statistics on their site. Also provides a service by which you can check to see whether a particular URL is in their index, and, if so, when it was last visited by their "spider." There is also some fascinating information about how Webcrawler's search strategy works. However, Speed seems to be slowing down a little recently. Its previous weakness--no way to refine search--has been eliminated with the addition of Boolean operators.
Yahoo!	Keyword search with simple and advanced options	Yahoo's index, Usenet, E-mail addresses	Since Yahoo returns relatively few hits (it will never return more than 100), it's not clear how results are ranked.	Although not precisely a search engine site, Yahoo is an important Web resource. It works as an hierarchical subject index, allowing you to drill down from the general to the specific. Yahoo is an attempt to organize and catalogue the Web. Yahoo also has search capabilities. You can search the Yahoo index (note: when you do this you are not searching the entire Web). If your query gets no hits in this manner, Yahoo offers you the option of searching the Alta Vista, which does search the entire Web. Yahoo will also automatically feed your query into the other major search engine sites if you so desire. Thus, Yahoo has the capacity to act as a kind of meta-search engine.

What are indices?

An index is a listing of the contents of the web, usually in a particular category. The index is maintained by an individual or organisation, and updated regularly. The Subject Index of the Web is an alphabetical listing of subjects with further sub-listings. The KWSnet Web Resources Architecture, Decorative Arts and Interior Design Index

(www.kwsnet.com/artsarch.html) for example has links to numerous architecture-related sites including periodicals, organisations, projects and buildings, et cetera. Architecture Web Sites (<http://ejw.i8.com/archiweb.htm>) contains links to architecture related sites with grouping by topic, location, building, designer, organization, et cetera.

Table 2: Selected search engines and directories.

Search Engine	Address	Comment
37.com	www.37.com	A directory with links to 37 search engines
Addresses.com	www.addresses.com	Free search for email addresses, white pages, yellow pages, phone numbers, and zip codes. Also offers reverse lookup.
All the Web	www.Alltheweb.com	Has a simple interface and a huge database. Also includes news, pictures, videos and music.
Ask Jeeves!	www.askjeeves.com	Ask Jeeves Incorporated. The principal activity of the Group is the provision of online personal service infrastructure for companies seeking to target, acquire, convert and retain customers online.
Direct Hit	www.directhit.com	Web search engine that utilizes popularity information to determine more relevant results.
Dogpile	www.dogpile.com	A meta-search engine including the white and yellow pages.
eWebsite	www.eWebsite.com	A search engine and directory
FindWhat.com	www.Findwhat.com	FindWhat.Com. The Company develops and markets performance-based advertising services for the internet.
Librarian's Index	www.lii.org	A web directory of the Library of California
MSN	www.msn.com	A web directory similar to Yahoo!
Questfinder	www.questfinder.com	Allows capitalization, categorization and phrasing
Search King	www.searchking.com	A portal network directory
SearchPort	www.searchport.org	A meta-search engine and directory.
Teoma	www.teoma.com	
WiseNut	www.wisenut.com	A new engine based on similar technology to Google.

Table 3: Arts and Humanities search engines

Search Engine	Address	Description
ADAM	www.adam.ac.uk	UK art, design, architecture and media Information
Art-Bridge	www.art-bridge.com	Directory and searchable database of art resources
Getty Information Institute	www.ahip.getty.edu	Bridging art and technology
Librarians Index	www.lii.org	Arts, History, Philosophy, Science, etc..
Voice of the Shuttle	www.vos.ucsb.edu	Humanities research directory.

Table 4: Science and Technology search engines

Search Engine	Address	Description
COS	www.cos.gdb.org	Searching in expertise, facilities and inventions.
EESE	www.ee vl.ac.uk/ee se	Engineering E-journal SE, over 100 engineering e-journals.
IBM	www.patent.womplex.ibm.com	IBM's Patent Server, search technology showcase.
Knowledge Express	www.knowledgeexpress.com	An information service designed specifically for technology transfer, licensing and business development departments, providing searching, alerting services, mailing lists, and reporting capabilities.
Leeds	www.leeds.gov.uk/library/services/patents.html	22 million documents, patents.
SciSeek	www.sciseek.com	A search engine with categories

Web directories

A web directory (also called a subject directory) is a catalogue very similar to the familiar library card catalogue. The catalogue organises the web into major topics, with a list of topics under each major topic. Each topic has a list of subtopics, and each subtopic has a list of sub-subtopics, and so on. Yahoo for example is a subject-tree style catalogue that organizes the Web into 14 major topics namely Arts, Business and Economy, Computers and Internet, Education, Entertainment, Government, Health, News, Recreation, Reference, Regional, Science, Social Science, Society and Culture. See Table 5.

Table 5: Selected web directories and portals.

Internet Directory	Address	Comment
4anything	www.4anything.com	A network of web guides that help users find information for thousands of topics of interest and cities.
American Access to Information Guide	www.ameri-globe.com	A directory and guide to American businesses, professional firms, government offices, associations and other public and private sector institutions.
AOL Search	www.search.aol.com	AOLs search tool, uses the Open Directory Project as its backbone.
ArteNetwork	www.artenetwork.com	International free art portal.
Arts, Sciences, and Humanities News & Directory	www.toppest.com	News headlines and a directory of websites related to the visual arts, performing arts, creative arts, sciences, humanities, and luxury lifestyles.
Galaxy	www.galaxy.com	Searchable Internet directory whose mission is to provide contextually relevant, trustworthy information
Internet Public Library	www.ipl.org	Includes directories of online texts, newspapers, magazines, reference materials.
Lerra.com	www.lerra.com	Search engine and directory that organizes Internet content into thousands of targeted categories.
MagPortal.com	www.magportal.com	Search engine and directory allowing users to locate free magazine articles on the Web. Site supplies free article feeds and users can browse by category.
Main Portals	www.mainportals.com	Guide to the main Internet portals and search sites organized by subject country.
Microsoft Network	www.msn.com	Microsoft's Internet access service and directory.
Search Engine Guide	www.searchengineguide.com	Guide to search engines, portals and directories.
The Associations Directory	www.marketingsource.com/associations	The Associations Database is a searchable directory of business related associations around the world. You may search by title, area code, city, state, zip code or category.
Topsitelinks.com	www.topsiteslinks.com	Web directory that provides links to reviewed sites in various categories for different countries.
Yahoo!	www.yahoo.com	Yahoo! Inc. The Company is a global Internet communications, commerce and media company. The company offers a wide range of branded network of services for a broad range of audiences, based on demographic, key-subject and geographic interests.

Web portals and Web hubs

These are places where people go to get information about a multitude of subjects, and even to chat, send email and form online communities. These are actually search engines but now incorporating web directories into their sites. Examples are Excite, Infoseek, Lycos and Alta Vista.

The Search Strategy

The choice of a search strategy is determined by the type of information we are looking for. For broad general information, start with a web directory. Use a search engine for narrow, specific information.

Online Libraries

There are several public, federal and university libraries on the web that feature online catalogues and digital collections. The digital collections often contain photographs, maps, archives and text. The libraries however usually require registration, often with an annual fee to access the collections online. See Tables 6 and 7.

Table 6: Selected Public and Federal Libraries

Library	Web address
Cal State Library	www.sunsite.berkeley.edu
Chicago Public Library	www.chipublib.org
Library of Congress	www.lcweb.loc.gov
New York Public Library	www.nypl.org
Philadelphia Free Library	www.library.phila.gov
Queens Borough Public Library	www.queens.lib.ny.us

Table 7: Selected University Libraries

Library	Web address
Columbia University Library	www.columbia.edu/cu/lweb
Cornell University Library	www.campusgw.library.cornell.edu
Harvard University Library	www-hcl.harvard.edu
The University of the Witwatersrand Library	www.wits.ac.za/library
MIT Library	www.libraries.mit.edu
Princeton University Library	www.infoshare1.princeton.edu:2003
University of Chicago Library	www.lib.uchicago.edu
Yale University Library	www.library.yale.edu

RIBA Library

The Royal Institute of British Architects (RIBA) library (riba-library.com) is a loans library with books collection, periodicals collection, special collections, drawings and photographs. There are also exhibitions, publications, manuscripts and archives. The services include online catalogue, information and reprography. It is available to the public for a fee, and to members at a discount.. The RIBA Library catalogue online is a

free index to over 30,000 books, 150,000 articles from 300 periodicals in 20 languages, and 23,000 records of drawings, photos, and manuscripts. It includes records for material which has been catalogued or indexed since the 1980's.

Some 2,000 periodical titles (of which some 700 are current) are held. Journal holdings include full sets of Architectural Review, The Builder, the RIBA Journal and Country Life. Photocopies of periodical articles can be supplied. The notice board covers building and construction regulations, some special projects and architectural journals published in the UK.

Architecture Magazines

Majority of international architecture magazines have web sites giving the profile of the magazine and containing subscription information.

Online (Electronic) Architecture Journals and Magazines

There are several architecture magazines that are available online. The University of Berkeley digital library (www.lib.berkeley.edu/ENVI), for example, has an environmental design library with a collection of electronic journals. While many of the journals are restricted to registered faculty members, most are free. The specific articles can be downloaded as PDF files. See Table 8.

Online books

There are several sources of digital books on the web. Archnet.org is an online community for architects, planners, urban designers, interior designers, landscape architects, and scholars, with a special focus on the Islamic world. It has a digital library with publications, images and a gallery. Publications can be downloaded from the site. You can search for publications by author, title, building type, country, language and specific keywords.

The Online Books Page (onlinebooks.library.upenn.edu) lists more than 20,000 online books in English. All the books are free for personal and non-commercial use. You can search by author and title, and browse by subject. The site also lists freely accessible archives of magazines, journals, newspapers and other periodicals.

Architecture books

There are thousands of books that can be ordered online, usually at a discount. Popular web sites with architecture books include Amazon.com, eBay.com, Architectstore.com, half.com, HalfPriceBooks.com, ecampus.com, alibris.com, builderspace.com and overstock.com.

Encyclopaedia

Many encyclopaedia are available on the web on CDs or in print, and these can be ordered online, for example through amazon.com. Many of these are internet-linked.

There are also several encyclopaedia available online. Bartleby.com for example provides several reference materials free of charge.

Britannica.com contains all the 32 volumes of Encyclopaedia Britannica complete with pop-up dictionary and research tools such as table of contents, complete index entries and citation information. There is an annual subscription of about \$60, but the first 72 hours is free.

Table 8: Selected electronic journals

Periodical	Web address	Comment
Architect Designed Houses	www.archmedia.com	Architect Designed Houses showcases domestic architecture by selected Australian architects. Architect Designed Houses features houses and renovations to suit a variety of styles and budgets.
Architecture Australia	www.archmedia.com.au	The official national magazine of the Royal Australian Institute of Architects. Published bi-monthly, it reviews the latest in Australian architecture and covers important issues relating to the profession.
Architecture SOUTH	www.specsite.com/w_grfx/007.a.html	journal of South-eastern United States architecture.
Design Intelligence	www.di.net	Design Intelligence is a newsletter dealing with business and strategic issues for design professionals.
Architecture Resources	www.lib.berkeley.edu/ENVI/Architecture.html	The environmental design library of the University of Berkeley
Architecture Electronic Journals	www.ejw.i8.com/archie-j.html	A listing of many free electronic journals
African Studies Ejournals	www.columbia.edu/cu/lweb/indiv/africa/ejournals.html	Electronic journals and newspapers on Africa
Architecture ejournals	www.lib.strath.ac.uk/engweb/arcej.html	A University of Strathclyde list of electronic journals of relevance to Architecture available via the Library catalogue.
Electronic Journals and magazines	www.usg.edu/galileo/internet/electronic/elecjour.html	An index of electronic journals and texts, including architecture journals
Architecture Journals	www.architectstore.com/magazine.html	A web site with links to over 100 architecture magazines

Schools of Architecture

Many schools of architecture around the world list their programmes, curriculum and faculty on the Web. Every school of architecture will eventually register a presence on

the Web, making exchange of ideas and information between lecturers, students and architects easier.

Conclusion

Architecture is a global phenomenon, and architectural research in Nigeria requires knowledge of sources from other countries. The findings should however be relevant to the Nigerian experience, hence the need to compare foreign sources with the local sources. While the information about Nigerian architecture is growing on the web, it is important to speed up the process through grants to academic and research institutions. While the logic of search engines is being improved daily to find more relevant information on the web, the engines cannot find what has not been put on the web.

Acknowledgements

The information about the web sites was obtained from the web sites. Details and links were obtained by search using various search engines, especially Google, Lycos, Alltheweb.com and 37.com.

References

1. <http://archnet.org/library> (2003).
2. <http://archnet.org/library/documents> (2003).
3. <http://digital.nypl.org> (2003).
4. <http://lcweb.loc.gov> (2003).
5. <http://onlinebooks.library.upenn.edu> (2003).
6. <http://webdev.uwaterloo.ca/ejournals> (2003).
7. <http://www.allsearchengines.com/scitech.html> (2003).
8. <http://www.architecture.com/go/Architecture/Reference> (2003).
9. <http://www.architecturecafe.com> (2003).
10. <http://www.columbia.edu/cu/lweb/eresources/ejournals> (2003).
11. <http://www.di.net> (2003).
12. <http://www.lib.berkeley.edu/ENVI> (2003).
13. <http://www.library.phila.gov/register> (2003).
14. <http://www.monash.com> (2003).
15. <http://www.msn.com> (2003)
16. <http://www.pbs.org/saf> (2003)
17. <http://www.science-frontiers.com> 2003)
18. <http://www.searchalot.com> (2003).
19. <http://www.wits.ac.za/library> (2003)
20. <http://www.yahoo.com> 2003).