Building a global legal index: a work in progress

Madeleine Davis

The article describes the process of developing, expanding and refining the directory of thousands of selected legal websites available on the World Law Index of the Australasian Legal Information Institute (AustLII).

The Australasian Legal Information Institute (AustLII) (http://www.austlii.edu.au) has a commitment to maximize the availability of free legal information via the World Wide Web (the Web). It does this in four ways:

- through its databases of primary legal materials. AustLII hosts over 80 full-text databases of Australian, United Kingdom (BAII), South Pacific (Fiji, Vanuatu, East Timor) and New Zealand legislation and case law;
- through its special collections databases (e.g. the Australian Treaties Library; Indigenous Law Resources, Workplace Relations Law Project);
- through its extensive secondary materials (e.g. law reform reports, online journals);
- through its World Law index (a Yahoo-type directory).

World Law is a catalogue or directory of over 10,000 links to the world's best law sites on the Web. There are links to legislation from over 100 jurisdictions worldwide, hundreds of law journals and much other material. World Law’s subject index catalogues materials by over 80 subjects. As well as providing links to law websites, AustLII also allows the full text of most of these sites to be searched. The search facility already covers over 12Gb of legal texts and this is growing rapidly. The search facility is built by a ‘web robot’ sent to capture information from sites in the index. AustLII’s ultimate objective is to find and index all types of legal resources located anywhere on the Web. It is a global legal research facility with up to 300,000 ‘hits’ (pages accessed) per day.

AustLII’s achievement has been to provide an accessible interface for the user by integrating its research tools – combining the ability to browse the information (via a hierarchical index) with AustLII’s own sophisticated search mechanism.

Organizing and structuring the information

Printed indexes are well understood and have been used for several centuries. They are fundamental to any research activity. The Web has taken indexing into an entirely different realm. Indexers were once able to index with a general understanding of the audience and the material. Now they must understand much wider user expectations and how this new audience interacts with information on the Web.

Distributing information via a website is still relatively new and navigation of the information in this new format is still in its infancy. However, the same indexing standards apply – consistency, clarity and organization (Moys, 1993: 16). The user must know what to expect by a logical progression through the information; the user must understand the term or concept; and the entries need to be properly organized, not just in alphabetical order but with multiple entry points into the information. The web indexer must still anticipate how the user will dive into the information.

AustLII has borrowed extensively from traditional indexing methods. This is AustLII’s value-adding, based on subject indexing, cross-referencing of material, inserting multiple entry points into the information and formal structuring of the categories of information.

Adapting indexing techniques to the Web has entailed adjusting and modifying the normal indexing rules. AustLII uses a highly structured hierarchical system that emphasizes context. On AustLII’s World Law pages, each menu screen presents users with choices and then leads them to specified categories. Only one heading level appears on the early screens; this is in contrast to the flatter structures of back-of-the-book indexes, where up to three or four levels of information are laid out. However, a back-of-the-book index provides only a one-way path between the index and the content. A web index using hypertext provides a fundamentally different model for user interaction with the information. AustLII’s World Law provides a generative or associative browsing environment, allowing the user to make connections between terms as well as providing multiple levels of pathways and routes through the information. Linking sites by country or region, by subject matter and by source or type provides a rich and multifaceted perspective for the user to connect and navigate the content at a number of different levels in the index.

Although the number of levels of subcategories are limited as much as possible, users may need to go through several layers to reach the contents or directory pages. This decision is traded off against the user’s ability to target exactly the information they want. User frustration with ‘drilling down’ through a number of levels before getting to information is offset by the specificity of the information that can finally be retrieved.

The user can click through several levels by expanding a term in the index categories. In the actual directories or
contents pages the links are annotated to allow the user to pick the most relevant link under a particular category.

The advantage of this type of index is its flexibility: the category entries can be expanded so that subject areas can be kept up to date and relevant; the directories or contents pages have no limit to the number of entries. The challenge for the indexer is to select the right subset of subjects or categories to include on the screen and still work within the various structures and vocabularies of the law. The organization of current information must also take into account future developments. As the content of the hierarchical scheme grows, the hierarchy is modified and further subdivided to fit new material.

And of course, unlike the limitation imposed by printed indexes, links can be made to the full text of statutes, cases, commentary, treaties and so on, which may be hosted on the AustLII site or elsewhere on the Web. Online indexing brings together a multiplicity of resources. Primary as well as secondary legal material can be indexed and links are made to individual pages or subsites within a website, not just to the main homepage.

The information retrieved from linked websites depends on the structures of the websites visited. Many websites do not have good directory structures to hold their information or use frames to display the information. And some websites simply do not allow the web robot to retrieve any information.

Categories and subcategories

AustLII’s collection of over 10,000 links to websites starts at the World page (http://www.austlii.edu.au/links/World). The links in the catalogue are divided into categories, with many categories divided into further subcategories. Each category or subcategory is on a separate web page of the catalogue.

The top subcategories of the World page (Fig. 1) are:

- **Countries**: Legal websites from every country in the world are included in the catalogue. Countries are also grouped in some cases by language, religion, culture or political affiliation.
- **Other indexes**: Links to other valuable legal indexes, directories and catalogues on the Web.
- **Regions**: Pages for each region of the world, including links to each country in the region and to regional organizations.
- **Search engines**: Links to Web search engines, which are not restricted to legal information.
- **Subject index**: As well as being indexed under their source or type (the middle subcategories of the World page below), almost all links are also indexed under their subject matter.

The middle subcategories of the World page (Fig. 1) are:

- **Courts & case law**: Courts and case-law collections for each country and for international courts.
- **Education**: Law schools and other legal education facilities for each country.
- **International**: This is the starting point for all international legal information. Subcategories include treaties and international agreements; international courts; inter-governmental organizations; international non-governmental organizations; public international law and private international law.
- **Law journals**: Law journals, some with full text, others with subscription details only.
- **Law libraries**: International law library associations, and law libraries for each country.
- **Law reform**: Law reform organizations for each country where available.
- **Lawyers**: Lawyers’ organizations by country (and international organizations).
- **Legislation**: Sources of legislation for each country.
- **Parliaments**: Parliament websites for each country.
- **Publishers & bookshops**: Legal publishers and bookshops, both international and by country.
- **Research**: Research tools, dictionaries, encyclopaedias, legal history, etc.

AustLII’s most successful adaptation of indexing technique is in cross-indexing the material though double-posting, triple-posting, and so on. Most websites are listed in at least three ways, allowing the user more than one way to browse for the information:

- by the country or region to which they relate (or under one of the subcategories of ‘International’ if they do not relate to a particular country);
- by the subject matter they concern (on one of the Subject Index pages);
- by the source or type of information they concern (under one of the subcategories in the middle level of the World page).

The structure of other pages in World Law

Individual subject/country pages have a three-part structure (Fig. 2):

1. The top division lists subcategories of the catalogue organized by type or source of information.
2. At the second level, subcategories are organized by subject matter. See references and see also references to related categories are also placed here.

3. The links to individual websites (not shown) – the content of the catalogue – are located here, in alphabetical order by the name of the site or topic heading within the website. The addition of annotations (usually a comprehensive description of the site content) to the actual directories of links assists users in judging whether the link is relevant for their needs. This information might take the form of a précis of the contents of a site, an outline of the type and extent of the information on a site or, if the link is pointing to a general site or government or agency site, then the link indicates it is to the home page. AustLII does not indulge in star ratings; only links that have been assessed for their relevance are listed.

- Stored searches, links that run predefined full-text searches using the search engine, are also located here in their own distinctive box (Fig. 3). This is really an embedded search in the subject/country pages – AustLII does the work for you! Stored searches are predefined searches of World Law, each with a different search scope, which will find hundreds of web pages about the subject or country concerned. The main advantages of these predefined searches over normal links are that:
  (a) the results of the search are updated every time the link is accessed;
  (b) the expertise of World Law’s indexing staff is used to create the stored search, which may be a higher level of expertise in searching than that of most users (i.e. full boolean searches are embedded);
  (c) The information retrieved has been gleaned across the whole of the World Law catalogue and includes captured information from websites linked under another category in the index;
  (d) Stored searches for country pages can be constructed as multilingual searches (e.g. the stored search for Cambodia will give results for cambod* or kamboji* or camboya* or camboja* or kambodsha* or cambogia* or kampuchea*). Search terms with asterisks match all words that begin with the specified text. A search on ‘cambod*’, for example, will retrieve terms including ‘cambodian’ and ‘cambodia’.

The first and second divisions are essentially for navigation around the catalogue and the third level is where the content of the catalogue is found and information can be stored from throughout the catalogue.

Navigating the structure (the browsing hierarchy)

The issue of accessibility on the AustLII site is most important. Pages are deliberately minimalist, designed to be legible and short, to download quickly and to be viewed on a variety of user platforms. In addition, the necessity for scrolling down pages has also been kept to a minimum by offering two columns of category information wherever possible.

Today users expect information to be accessible in a more-or-less seamless way, whatever the physical source. Cross-indexing of websites, double-posting and even triple-posting means that there is usually more than one way to enter and browse the information in World Law. In addition, navigation support is continuous (Fig. 4): there is a clear indication of what part of the catalogue any one page belongs to, and it is linked back to the main page for that section so that users can track where they have been in the hierarchy. At the top of every page in the catalogue is the page’s hierarchical location. In other words, a list shows the hierarchy of categories and subcategories (starting with World) that give it its position in the catalogue. The user can click on the name of any category in the hierarchy to go back to the page in the catalogue.
On many pages, links to some or all subcategories of the catalogue are followed by the (@ symbol, which shows that this is a cross-reference to another part of the category hierarchy. This occurs in two situations, which in a library catalogue or back-of-the-book index would often be called a see reference. In the Subject Index, see references direct users from terms that are not used to terms that are used (e.g. Bankruptcy @ Insolvency & Bankruptcy, Trade Practices @ Competition Law). It also indicates throughout the directory where links from double or triple entries direct the user to another pathway in the catalogue (e.g. Information about privatization in Canada is situated under World>>Country>>Canada but an @ link is made from the World>>Subject Index>>Privatisation>>By Country page).

See also references, which suggest related concepts, are clearly differentiated on the screen and they are usually reciprocal. In World Law, the use of see and see also references also affects the scope of searches.

**Surfing and classifying**

This is where the dual aspects of the job enter; indexing can only be undertaken after the material has been found and vetted. AustLII cannot develop the directory categories until there is something to put behind them. Unlike book and periodical indexing, or even website content indexing, the AustLII process is twofold:

1. first find the information; and
2. then classify it and add it to the World Law pages.

There has been a huge expansion of the amount of legal information on the Web during the late 1990s. In particular, governments, universities, legal associations and international organizations have developed websites with extensive information. The exponential growth in the availability of legal decisions by the inclusion of the deliberations of courts and tribunals, and the free access provided to much of this material, has also added to the amount of legal information that is now both freely available and accessible via the Web. Retrieval and indexing of this material from conventional print media is not a practical option in many cases.

**Keeping up to date**

The explosion of information on the Web has given rise to a new problem – finding and keeping up to date with these new resources. Given the technical sophistication of the Web, the discovery of new sites is surprisingly unsophisticated and must be undertaken through a combination of several methods:

- AustLII participates in Internet lists and discussion groups. INT-LAW, the international law librarians list, is one of the most proactive legal groups on the Web and a mine of information about new sites;
- AustLII belongs to a number of email alert lists;
- individual law firm sites often contain information, news and publications online;
- search engines on the Web, including law-specific search engines;
- review of other legal directory, law portal and catalogue sites on the Web, particularly those developed by major international law libraries;
- ‘Add a link’ to AustLII: other sites may submit links to be added to the directories after checking and approval. These sites can choose the AustLII subject or source category most relevant for their site and request a link to their own site;
- ‘mining’ of government and university sites world wide.

Once found, there are two basic criteria for choosing to link to sites:

- **Technical:** some websites make it technically difficult to link to anything other than their home page. It is then not appropriate to index it under any specific subject heading or enable the information on the site to be retrieved by AustLII for search purposes.
- **Intellectual:** users are directed to quality sites and sources that contain related information. Sites are listed in World Law if they appear to be a serious attempt to provide useful legal information. Details of the origin, creator or publisher of the website are listed so the user can see before clicking through where the information is coming from – a government, academic or international organization site, for example.

Additional, not specifically legal, material is also added – governments and parliaments, departments and agencies, international organizations, search engines, reference material and special topics such as Poverty Reduction, Corruption, Money-Laundering and Good Governance, have been developed.

User trust is important to AustLII. Websites are frequently added to and restructured. There may be outdated content or lost links and it is a sure way to lose credibility, particularly with new or inexperienced users, when this occurs. Links are only valuable to the user if they are maintained. AustLII’s links pages are checked automatically but the indexer has the task of finding and fixing broken links.

The Web provides the opportunity to facilitate international collaborative efforts. A group of volunteers is always able to accomplish more than one person. The success of the Librarians’ Index to the Internet (LII) (http://www.lii.org) in the early 1990s, the DMOZ Open Directory Project (http://www.dmoz.org) and the Social Science Information Gateway (SOSIG) (http://www.sosig.ac.uk) all show this. AustLII is also developing a contributing editors concept. So far, collaboration with the Department of Foreign Affairs (Australia), Elections Canada, Seoul University (Korea) and the University of Otago (New Zealand) have been established to utilize individual expertise and develop specific topics. Attracting further interest and the cooperation of others around the world with a commitment to the provision of free access to legal information via the Web is an ongoing goal.

From an indexer’s perspective the experience of working at AustLII and the challenges involved has led me to appreciate that the skills I already have can be transferred to indexing the new medium – the Book of the Web!
The bliss of excessive fondness

Dr Samuel Johnson suggested to his friend Samuel Richardson the compilation of an index to Richardson’s novel, Clarissa, so that ‘when the reader recollects any incident, he may easily find it’. Richardson accordingly compiled an Index Rerum to Clarissa in 1751, and in 1754 provided for Sir Charles Grandison an ‘Index Historical and Characteristical of the Seven Volumes of this Work’. In 1755 he published as a separate work of 450 pages A Collection of Such of the Moral and Instructive Sentiments, Maxims, Cautions and Reflexions, Contained in the Histories of Pamela, Clarissa, and Sir Charles Grandison, described in the Preface as a ‘General Index both of Maxims and Reflections’.

The collection of sentiments from Clarissa alone runs to 135 pages in what we would now call A5 size. Its entries are listed in alphabetically arranged sections with group (synonymous) headings, the first being:

- Adversity
- Affliction
- Calamity
- Misfortune

This is followed by a page and a half of aphorisms from the book. Next comes:

- Advice and Cautions to Women

which begins

EVERY one’s eyes are upon the conduct, the visits, the visitors of a young Lady made early independent, i. 120. [125].

There are nearly seven pages of such cautions, followed by

- Air and Manner
- Address

A specimen section in full is:

- Learning

A LETTER’D education too generally sets the children of the poor above those servile offices, by which the business of the world is carried on’, iii. 363. [iv. 148]

‘Take the world thro’ there are twenty happy people among the unletter’d, to one among those who have had a school-education’, iii. 364 [iv. 148]

‘Yet who would not wish to lift to some little distinction and gentle usefulness, the person he desires to reward!’ [ibid.]

Madeleine Davis is a back-of-the-book and loose-leaf indexer who has ‘morphed’ into a web surfer, indexer and classifier and is currently employed as a legal indexer on the AustLII website. This article is based on a paper given at AustLII’s Law via the Internet ‘99 conference (Davis et al. 1999) and a speech given by Madeleine at the Annual Maria Gemenis Award Dinner, NSW Special Libraries Group, 28 July 1999. Email: mdavis@austlii.edu.au

The Indexer Vol. 22 No. 3 April 2001