

An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval

Lloyd Sokvitne
Manager (Information Systems Development)
State Library of Tasmania
Lloyd.Sokvitne@central.tased.edu.au

Abstract:

Presents the results of a study into the effectiveness of Dublin Core metadata for retrieval. The metadata from twenty Australian government and educational organisations was analysed. The metadata was assessed for its capacity to facilitate title, creator/publisher, and subject access. From the results, it emerged that title metadata added little value, that creator/publisher access was flawed by inconsistent name formats, and that the subject terms used were too broad to produce increased precision and decreased recall. Specific suggestions are made as to how metadata creation in these three areas can be improved, and the relationship of Dublin Core to information foraging theory is explored.

Background

The past five years have seen the development of a number of methodologies to improve the ability to locate resources available over the World Wide Web. One of the most significant recent outcomes has been the development of the Dublin Core (DC) metadata schema. This set of 15 elements attempts to provide a basic and agreed set of descriptive categories that enable effective information retrieval over the World Wide Web. Although initially focussed on Document Like Objects (DLOs), Dublin Core quickly expanded its semantic scope to include the full range of information formats, from text to images and multimedia, and included physical as well as electronic resources.

The importance and persuasiveness of Dublin Core (DC) can be seen in that it has become the basis for a range of specialised retrieval schemes. The Australian Government Locator Service (AGLS) has built onto Dublin Core to meet the extra needs of the Australian government sector. Australia's Education Network (EdNA) has built on key semantics within Dublin Core, and adds additional fields for educational requirements.

The success of Dublin Core and its extensions (eg. AGLS) is illustrated by the rapid adoption of metadata as a base standard by large information producing sectors such as government, education, museums, etc. Metadata is particularly seen as an important and correct strategic decision by government because of its ability to enhance the delivery of information to the general community. Formal recommendations abound within the government environment across Australia recommending the use of AGLS for web publishing. Systems such as the Business Entry Point (BEP) utilise metadata as a base for description and retrieval. Emerging cross-jurisdictional systems such as GOVERNET assume a rich AGLS metadata environment.

This in-principle adoption of the AGLS standard has not yet led to the significant production of actual metadata. The absence of fully developed and user-friendly metadata creation tools means that the utilisation and production of metadata has lagged behind the theory. Search engines that utilise DC and extended metadata schema are only just becoming available. This has also led to a dilemma: without these search engines there is little incentive to create the metadata content, without a significant body of metadata there is little incentive to develop an appropriate search engine.

One of the underlying assumptions of DC metadata and the Web discovery model is that Web information searchers are seeking information in a logical and stable way, and in a way that can be anticipated at the time that individual resources are created and catalogued. Recent work in 'information foraging theory' raises a number of serious questions as to the validity of this approach.

Information foraging theory suggests that information seekers 'sniff' out information, and that they follow processes analogous to the concepts of 'hunting and gathering' as taken from the biological world. The hunting process occurs when the required information is

identified by following a wide-ranging trail of references or possible sources until the searcher discovers the resource that actually meets their needs.

The path that the searcher takes to identify the needed resource is not necessarily a direct one, with decisions being taken on an assessment of likely returns as well as hard results. After this hunting process has identified an area of high quality information, these information dense areas are revisited whenever needed. The user is likely to modify their search techniques based on their needs at any given time and as a result of the feedback received during the foraging process.

It is not clear how the DC metadata discovery model will apply in an information foraging environment. Given that the nature of information gathering could be highly subjective as it is based on changing user perceptions, it may be unproductive to rely on a system where the individual creator of an information resource provides the searching data at the time of creation.

Objectives of this paper

The development of an agreed semantic base for Dublin Core has taken great effort and engendered significant debate. The current acceptance of Dublin Core should not obscure the issue that there is still a ‘discovery’ problem to solve. Although the development of the metadata elements and their meaning is now reaching conclusion, the effectiveness or suitability of Dublin Core must still be proven in practice.

This paper is based on the conviction that the value of any retrieval methodology should first and foremost be measured against its actual ability to deliver effective retrieval. It is the ability of a populated Dublin Core metadata set to facilitate retrieval that needs to be tested and investigated. This should be done while there is still time to adapt or modify its semantics or application standards and processes if needed.

This paper attempts to provide an early indicative and objective analysis of the actual data that populates the elements within a sample Dublin Core set from a number of large Australian organisations publishing in the Web environment. This analysis will be based on the ability of key DC fields to meet the primary and original metadata objective: to facilitate retrieval to that resource over the World Wide Web.

Research hypothesis

The development of Dublin Core has focused on general semantics (and to some extent syntax), with little attention to the actual population of the elements with content and the real relationship of that content to the retrieval process. The use of the schema qualifier and the ability of metadata elements to reference controlled vocabularies has built a theoretical base to improve the quality and consistency of the content. This flexible

theoretical base has sidelined a real analysis of how consistent content will actually be delivered in practice.

Information and indexing specialists have been grappling with the problem of effective retrieval from large datasets for decades. From this experience, it follows that the real measure of metadata indexing in the Web environment will be its ability to provide high precision and low recall. The Web searcher using metadata should be delivered a contained set of results with a high proportion of relevant returns. This research project will measure the ability of actual metadata content to improve precision and reduce recall when searching for that resource.

It is not necessary to await the development of search engines to assess the indexing quality of metadata content. The current content in metadata elements can be analysed simply as raw indexing data, and tested against basic indexing and retrieval criteria. These indexing criteria are independent of the end tools: sophisticated tools will be unable to add quality and precision if that does not first exist in the indexing data itself.

As well as measuring the content of DC elements against basic indexing criteria, it is important to assess the effectiveness of metadata against the freetext capabilities of search engines. The concern over the high recall and low precision of this approach was one of the driving forces of DC development, but it remains to be seen whether DC implementations will improve the precision and recall ratio of standard freetext searching.

Because the Internet has inconsistent and at times slow network speeds, the Web searcher is likely to rely heavily on the results that are presented in the first few screens delivered. Against this environment, precision can be identified closely with ranking. The ability of metadata to help rank results is likely to be dependent on the functionality of the software used, and has therefore been excluded as a research component in this study. The DC metadata elements themselves do not address the question of ranking as a semantic or functional issue.

Research principles and methodology

This research begins with the assumption that Dublin Core metadata exists to facilitate retrieval by searchers from outside the originating organisation and without specialised knowledge about the location or types of resources they are likely to return. From this assumption, there are three basic retrieval needs that are likely to form the focus of searching activity. These are

- retrieval by title - accords to DC.Title field
- retrieval by creator/publisher - accords to DC.Creator, DC.Contributor, and DC.Publisher fields
- retrieval by subject - accords to DC.Subject field.

The semantic differences between the creator, contributor, publisher metadata fields are not likely to be widely understood in the user community. These three fields were combined and treated as a single retrieval strategy for the purposes of this research.

Retrieval by title has been measured against these criteria:

- 1) Does the Dublin Core title reflect what the average viewer of the Web page would call the title?

The title used for retrieval must reflect what the searcher is likely to search for. Although the HTML title tag is meant to reflect the page title, most users would expect the dominant title that physically presents on the displayed Web page to be the real title, and this title is what they would cite or search for. The absence of a DC title element that reflects the 'apparent' title reflects a failure in indexing policy rather than in DC itself, as the DC.Title element can be repeated to provide alternate title entries.

- 2) Does the Dublin Core title element duplicate the HTML title tag?

The DC.Title field will be checked to see if it simply duplicates the HTML element tag <title>. There may be a potential loss of retrieval capacity if DC.Title merely replicates a retrieval point that already exists.

- 3) Is the Dublin Core title duplicated in the text?

The DC.Title entry will be checked against the content of the web page to see if the title phrase also occurs in the normal text displayed in of the page. This check is to ascertain how the use of freetext searching across web pages to search for a title phrase compares with searches that rely on the DC.Title element.

Retrieval by DC.Creator, DC.Contributor, and DC.Publisher has been assessed using these criteria:

- 1) Is the DC.Creator/contributor/publisher duplicated in the text?

This criterion serves to assess how the extra effort taken to populate these metadata elements compares to the capacity of freetext searching to retrieve the same information.

- 2) Is the DC.Creator correctly attributed?

This criterion is used to ascertain whether the correct entity is given attribution as creator, and uses the philosophical stance that corporate pages and organisational data should reflect corporate authorship rather than indicate the person creating the HTML page.

3) Is the DC.Publisher in Library of Congress Name Authorities (LCNA) equivalent format?

This criteria is applied in order to assess the likely value of the publisher name as a consistent retrieval point across organisations, jurisdictions, and the World Wide Web itself. Publisher names that assume the main organisation name and just cite a division or branch, for example, are not likely to facilitate retrieval from outside that organisation.

The need for subject retrieval leads to an analysis of the DC.Subject element. Although there is an obvious expectation that retrieval software will add post-coordinate indexing capacity to the DC.Subject element, the quality of the index terms/phrases assigned to a resource is of basic importance. Within this context, the types of indexing error identified by F. W. Lancaster are a relevant basis for assessing the terms and phrases found in the DC.Subject element. These five indexing errors have been applied as follows:

1) the indexer fails to follow indexing policy, particularly related to the exhaustivity of indexing

Given the open and diverse nature of Web resources to be analysed and the lack of a single indexing policy, I have defined this to include only very basic errors in indexing:

- a) an obvious over-supply of terms - over 25 terms for a single page
- b) cases where the DC.Subject content has clearly just been replicated across different resources, irrespective of the individual content of the pages.

2) the indexer fails to correctly apply the indexing vocabulary

Lancaster uses this to describe the incorrect use of terms against the rules of a thesaurus or subject heading scheme. I have used this simply to identify misspellings or ungrammatical/non-sensible fragments located in the DC.Subject element.

3) the indexer applies terms at the wrong level of specificity

The use of terms that are either too broad or too narrow for the subject matter does not enable the retrieval process to work efficiently. It is this concept that places the greatest demands on the skills and expertise of an indexer. For the purpose of this analysis, I have split this area into two specific types of error: the use of terms that are too general for the resource, and the use of terms that are too specific. Terms are considered too general if they describe a broad concept that extends well beyond the content contained in the resource. Terms are considered too narrow if that term describes a concept that is a minor or unimportant component of the resource.

It is not possible to provide an absolute objective measure of specificity and the evaluation of keywords is potentially affected by the subjective views of the assessor. To counter this factor, only obvious errors will be counted and the results should be treated as indicative rather than absolute.

4) the indexer uses an obviously incorrect term

The use of incorrect terms has a negative effect on the ability to retrieve that resource.

5) the indexer omits an important term

The omission of terms that should be used also has a negative impact on the ability of the searcher to locate a needed resource.

The occurrence of the last three types of indexing error have significant relevance to information foraging theory, where effective retrieval needs the careful choice of terms that target the needs of the end user.

The type of metadata records to be assessed was defined as metadata embedded in Web documents, and provided by government and educational organisations, rather than from commercial or community sectors. This scope was chosen because of the significant policy decisions to adopt metadata that are occurring at most levels of government. It appears the government sector, more than any other, will see the widespread adoption of metadata in the next few years. This also would enable the outcomes from this project to have some potential real benefit to this sector.

Metadata that is included within databases and metadata repositories was excluded because such metadata itself is usually difficult to access directly. The characteristics and functionality of such databases tend to obscure the value of the metadata content itself, and differences in software functionality make it difficult to compare metadata performance from different databases and different organisations.

Libraries were specially excluded from the research sample. It was assumed that the experience and expertise of libraries should enable them to provide high quality metadata (although this is an untested assumption). More importantly, it is the non-library sectors that will be creating the bulk of metadata in the future, and the philosophical base of DC metadata to date assumes application by people without librarianship training.

The current metadata environment

It proved difficult to find and utilise a large number and wide range of government and educational Web pages with valid DC metadata records. The most obvious problem was the scarcity of embedded metadata. Individual organisations applied metadata inconsistently to their web sites, most commonly at higher levels but not lower down where real content could be found. The metadata used was variable in quality, showing different states of DC development, but also different understandings of what metadata should be used. Certain organisations seemed to apply a mechanistic approach, even choosing to simply reapply an identical metadata set across different publications and resources.

Because of these problems, my earlier intention to sample a large number (250+) of metadata records became impractical. To reach such a number I would have had to include a disproportionate number of records from a small number of organisations. The results would have been distorted by the metadata cataloguing policies of just a few large organisations.

On the basis of this preliminary research, it also became evident that clear patterns about metadata usage emerged from as few as five records at a given site, and that it was not necessary to exhaustively analyse large datasets at each location. As stated above, the larger a set chosen from a given metadata producer, the more likely it was to unbalance the overall results.

Sampling processes

Against this background, major Australian government entry points (federal and state) were visited and agencies chosen at random. Home pages and content pages were checked for metadata, and where possible search engines were used to identify metadata producing agencies, divisions, or sections. Twenty organisations producing sufficient embedded Dublin Core metadata of minimum standard were identified using this method: 7 from the Commonwealth government, 10 from state and local government jurisdictions, and 3 from the education (University) sector. Five records from each organisation were selected randomly on the basis of following every second link on the home pages. The embedded metadata was then printed out from the source HTML and compared visually to the Web page as it presented in the Web browser.

Results

Title as retrieval point

Criteria	Occurrence
a) title used in DC.Title replicates the on-screen Web page title	48%
b) title duplicated in HTML tag	59%
c) title duplicated in body of Web page	47%

These results clearly indicated that the DC.Title element failed to adequately provide the title that most searchers would expect to search for, that is, the obvious title that appears on the Web page. From an indexing point of view, a 48% pass rate for DC.Title is extremely low. To some extent this may be explained by the high degree of duplication between the DC.Title and HTML title tag, which also did not reflect the apparent Web page title. This high degree of duplication may be an indicator of metadata generating software that assumes the DC.Title element can simply be populated with the contents of the HTML title tag.

It was not always easy to identify an obvious title on a Web page due to site stylistic conventions, frames, and template applications. However this potential confusion over the possible title of a Web page for a subsequent searcher could be redressed by the careful and occasionally repeated use of the DC.Title field. The use of the DC.Title element to merely replicate the existing HTML title tag has no value when that title tag itself has no retrieval value.

A significant number of Web pages used a graphical image to present the main title (44%). Of these, only slightly more than half of the pages (53%) used DC.Title as an alternative method to describe the apparent title. Freetext search engines would not provide alternative access to the title of the Web page if it is only presented as a graphical image. The careful use of DC.Title in these cases would be of great value to the searcher.

Creator/Contributor/Publisher as retrieval point

Criteria	Occurrence
a) creator/contributor/publisher duplicated in Web page	58%
b) creator and publisher elements assigned correctly	65%
c) publisher uses a recognised name authority standard	22%

Although the creator and publisher had a lower error rate than title, there still seemed to be some confusion amongst metadata creators, as well as Web page authors and designers, as to what the roles of creator and publisher were. To some extent the Web publishing environment does not expect or require conventions on its pages as to who published the page and who was responsible for the content. Given the focus on government publishing it is perhaps not surprising that simple copyright statements were commonly presented without any further indication of intellectual ownership or publishing processes on the page.

The most significant result was the inconsistency in the formats used for the organisation responsible for the resource. Many publisher elements contained names where the specific section was provided without reference to the jurisdiction or parent organisation. In other cases the order of parent organisation and unit name varied, even within the metadata records of one organisation. This measurement against the LCNA format is not to indicate a preference for Library of Congress Name Authorities in the Web environment, but to indicate the likely inconsistencies that will occur in publisher data if there is no authority control.

Subject as retrieval point

a) subject terms too numerous (over 25 terms) or simply repeated across resources - 19%

The worst case encountered was a Web page that contained 47 words and phrases in the DC.Subject element, and virtually all of these were too general for the content. Although this can easily be dismissed as an aberration, more worrying were the three organisations that simply repeated the DC.Subject

element content set across all the pages, irrespective of the real content. This clearly showed no understanding of the purpose or use of DC.Subject.

b) subject terms misspelled or grammatically incorrect - 7%

Although small in number, these errors reflected poor quality control over the metadata - eg in one case a university actually misspelled the word 'university' in the DC.Subject field.

c) subject terms too general: 79%

This emerged as the single biggest problem encountered and indicates the poor understanding that most indexers have of the need for specificity in keyword indexing. Common errors included providing the name of the jurisdiction, state or even country in the subject metadata, or including broad based topics that concerned the organisation generally but that were not reflected on the page that the metadata was meant to describe.

d) subject terms too narrow: 12%

The use of terms too narrow for the actual content of the page occurred relatively infrequently, although it was a concern that a high number of these pages also contained terms that were too general. This is the scattergun approach at its worst.

e) subject terms incorrect: 17%

The relatively high number of incorrect subject terms indicates that quality assurance processes are rarely applied to metadata addition.

f) subject terms missed: 23%

There appeared to be a high number of occurrences where apparently important topics on a web page were not reflected in the subject terms chosen for the metadata. However most of these errors came from just three organisations where it is obvious that there has been a failure to communicate the purpose of the DC.Subject field to the personnel actually creating the metadata. This suggests that the processes to allow Web authors within these organisation to create Dublin Core metadata is not being accompanied by information and training about the purposes of metadata.

There were 960 subject terms/phrases used within the sample metadata set, giving an average number of 9.6 terms per metadata record. Three records were found to have nonsense content in the DC.Subject element. An assessment of the occurrence of the DC.Subject terms in the text of the document was made, but with great variability in the results across pages and organisations. On average, 33% of DC.Subject terms also appeared in the text of the documents.

It is not possible to say how much of the subject metadata was created by the authors of the Web resources themselves. The poor results do seem to verify that trained metadata cataloguers or professional indexers did not create the metadata.

The metadata creators in this sample did not appear to understand the needs of possible searchers and the data did not exhibit the type of flexibility needed to provide alternatives that tune in to particular searching strategies. This suggests that the creation of retrieval metadata by untrained authors is not likely to promote subsequent discovery. Users in the information foraging model would be poorly served by such retrieval data.

Methodology issues

The problems that forced a relatively small sample set (100 records) were counter balanced by the high degree of consistency of the results across most organisations. The small number of organisations that were identified as providing embedded metadata in Australia also made the sample more significant. Occasionally an organisation would exhibit a consistent and contrary trend to the average, but this itself reflects the diversity of metadata application. This variability would itself be a worthwhile addition to a future research project but does not undermine the conclusions as to metadata usage that were identified. Overall I am confident that the trends and results identified in this study cannot be dismissed because of statistical errors due to a small total population sample.

It is possible that an evaluation of the specificity of the metadata keyword sample done by another researcher could produce some variation in results. This was mitigated because only obvious errors were included and the 'real' state of keyword metadata is therefore likely to be understated. Authoritative research into the appropriateness of subject keywords will require a team of assessors for objectivity and a highly developed instrument for measuring index term specificity. There is undoubtedly scope for such an approach in the future.

Conclusions

- 1) That DC.Title as currently used does not add value for retrieval purposes. This is because it commonly replicates the HTML title, and fails to reflect the apparent title presented on the actual Web page.
- 2) That the DC.Creator/contributor/publisher fields have the potential to add value to the retrieval process, but that inconsistent name formats render this potential lost.
- 3) That the DC.Subject element is not currently being used effectively to provide effective subject retrieval.

There seemed to be a fundamental and consistent misunderstanding or lack of awareness and training in what the DC.Subject field is for and how it should be used effectively. On current indications, an increase in metadata records would not

improve the recall/precision ratio exhibited by freetext search engines, but merely duplicate it. Strategic decisions to adopt DC or AGLS will not provide any return to the organisations and sectors involved unless there is an accompanying development of the policies and skills to populate the DC.Subject element.

- 4) That the Dublin Core standard will have questionable value as a discovery tool unless the elements are able to be populated and used correctly. This population of Dublin Core elements with data will need agreed and precise prescriptions and rules concerning their use.

Specific DC implementation recommendations

- 1) that DC.Title should be used to reflect the apparent title of Web pages

Such an implementation decision would not require the addition of skilled professionals to the metadata creation process, but it does require the human interpretation and evaluation of Web pages to assess what the apparent title actually is. The DC.Title element adds no value if it is populated automatically by a software program reading the HTML title tag or file name.

- 2) that work should begin to establish a methodology for the consistent application of creator/publisher names in DC metadata

The ability to search effectively for creator and publisher metadata across institutions and organisations implies adherence to a standard that encompasses a variety of needs and that maintains a wide perspective. The development of such a standard suitable for the Web is outside of the internal metadata cataloguing abilities of a individual organisation, and must be tackled either sector by sector (eg government, education, etc), or best of all across sectors by a national organisation. It would seem logical for the library community to lead in this process.

- 3) that the nature of subject term assignment within DC.Subject be seriously and professionally addressed before creating a large body of metadata within an organisation.

The creation of subject metadata must be built on a developed conceptual base, with appropriate tools and support processes, and with adequate training. This means that the organisation about to embark on the process of creating metadata must first address the indexing needs of its clientele and the characteristics of its content. From this it should develop an indexing policy that suits those needs, develop the appropriate software tools to facilitate consistent and accurate indexing, and implement training programmes to ensure that the metadata content creators are skilled in their tasks. Finally quality assurance processes must be developed to ensure that all of this effort is not wasted.

The retrieval flexibility that seems necessary within an information foraging model of Web behaviour indicates that subject metadata should not, or not only, be added at the time a resource is created. The subsequent addition of index terms would allow the use of third-party indexing expertise and would allow a flexible and varied approach to the provision of terms. This would allow terms to be added or removed to meet current and changing needs.

- 4) that an analysis similar to that employed by this paper be regularly conducted by organisations on the metadata that is produced within the organisation.

The purpose of such activity would be to ensure quality control and that appropriate outcomes were emerging from the process of metadata creation.

Future directions for metadata development

An important and successful assumption of DC is that information discovery on the World Wide Web should be built on flexible and varied approaches. However the semantics of many of the DC elements have only provided a broad interpretation of semantic intent, whereas the actual usefulness of these elements requires a detailed set of application instructions. Given the collegiate nature of DC development, it is unlikely that there would be agreement at this level of detail except within specific sectors, where either the nature of the material or the nature of its use allow such agreement. The results of this research indicate that this should most certainly now happen within the Australian government sector.

The subsequent assessment of Web content by third parties, who add enough indexing information to selected resources to allow these resources to be 'sniffed out' by information gatherers is another significant direction that should be pursued. This assessment by third parties would not focus on cataloguing or identifying every record that exists, but on identifying key information nodes or retrieval points. These are the resources that once identified by a user lead them onto the rich detail that they require, and that are revisited again whenever similar information is required.

Work is already beginning in this area through the development of subject gateways, and it is appropriate and important that libraries play a leading role in this process. But libraries may also need to develop a conceptual framework that helps to identify the nature and characteristics of key information nodes or intersections on the Web, and then develop a schema that allows the management of such nodes.

Bibliography

Metadata

Australian Government Locator Service (AGLS). Available: <http://www.naa.gov.au>. 24 September 1999

Business Entry Point (BEP). Available: <http://www.bep.gov.au>. 24 September 1999

Digital Libraries: Metadata Resources, IFLA. Available: <http://www.ifla.org/II/metadata.htm>. 24 September 1999

Dublin Core (DC). Available: <http://purl.org/DC/>. 24 September 1999

Education Network Australia (EdNA). Available: <http://www.edna.edu.au/>. 24 September 1999

Meta Matters, National Library of Australia. Available: <http://www.nla.gov.au/meta/>. 24 September 1999.

World Wide Web Consortium (W3C). Available: <http://www.w3c.org/>. 24 September 1999

Information Foraging Theory

Card, Stuart K..The WebBook and the Web Forager: An Information Workspace for the World-Wide Web. by Stuart K. Card, George G. Robertson, and William York. Available: <http://siglink.acm.org/sigchi/chi96/proceedings/papers/Card/skc1txt.html>. 24 September 1999.

Gerhard, Susan L.. Browsing In Context. Available: <http://www.twurl.com/chi-browsing-in-context.htm>; and, Browsing In Context -- an Information Foraging Model. Available: http://www.twurl.com/onr_report/foraging.html. 24 September 1999

Nielsen, Jakob. Why Web Users Scan Instead of Read, (Sidebar to Jakob Nielsen's column on how users read on the Web). Available: <http://www.useit.com/alertbox/whyscanning.html>. 24 September 1999

Pirolli, Peter. Information Foraging in Information Access Environments. by Peter Pirolli and Stuart Card' CHI '95 Proceedings, Papers. Available: http://www.acm.org/sigchi/chi95/proceedings/papers/ppp_bdy.htm. 24 September 1999

Robertson, George. Exploring Information Spaces. Available:
<http://research.microsoft.com/~ggr/navigation/exploring.html>. 24 September 1999

Sandstrom, Pamela Effrein. Scholars as Subsistence Foragers. Bulletin of the American Society for Information Science. Volume 25. No 3, February/March 1999. Available:
<http://www.asis.org/Bulletin/Feb-99/sandstrom.html>. 24 September 1999

Books - Indexing

F. W. Lancaster, F. W.. Indexing and Abstracting in Theory and Practice. Champaign, Illinois: University of Illinois, 1991.

Wellisch, Hans. H.. Indexing from A to Z. 2nd ed. New York: H. W. Wilson, 1995.

Foskett, A. C.. The Subject Approach to Information. Clive Bingley Ltd: London, 1982.