This is the published version:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30046156

Every reasonable effort has been made to ensure that permission has been obtained for items included in Deakin Research Online. If you believe that your rights have been infringed by this repository, please contact drosupport@deakin.edu.au

Copyright : 2001, Southern Cross University
Resource Synergy

Improving the discovery of quality resources to support and information-dense learning environment

Albert Ip, Technical Consultant, Education Network Australia (EdNA) [HREF 1], University of Melbourne, Parkville, 3010. albert@DLS.au.com

Michael Currie, EdNA Higher Education Project Manager, Dept of Information Systems, University of Melbourne [HREF 1a], Parkville, 3010. m.currie@dis.unimelb.edu.au

Prof. Iain Morrison, Dept of Information Systems [HREF 1b], The University of Melbourne, Parkville, 3010. j.morrison@dis.unimelb.edu.au

Abstract

This paper analyses the resources currently used in information-dense learning environments and identifies a growing use of resources beyond those created specifically for education. This highlights a limitation of current metadata standardisation efforts which concentrate on educational resources. After describing the search process in terms of a specific data model, it describes subject gateways and their role in the collection and discovery of quality resources.

The paper promotes collaboration between subject gateways to achieve a distributed collection of high quality resources in order to support an information-dense learning environment and addresses issues of interoperability and autonomy. Based on the recognition of the value of subject gateways, the authors set out four levels of collaboration and propose a path to achieve such goals.

Introduction

The dramatic change in our capacity to encode, transmit and locate information in a digital form has triggered a shift of paradigm in teaching and learning towards enquiry-based and resource-based learning techniques leading towards information-dense learning environments. The ability to provide rich sources of information in the daily teaching and learning at both school and university levels may have significant impact on the design of these learning environments.

Learning Environments

Technology-enhanced learning environments refer to those educational settings where some element of instructional technology is used to support the interaction between the teacher, the learner and the educational content (Ip and Naidu, in press). Students in these educational settings are often required to pursue self-directed and open-ended inquiries. These environments are designed not so much to instruct as to provide contexts wherein understanding and insight can be cultivated.

Suitably designed environments comprise a complex arrangement of several components. Perkins (1991) has labeled these components as “information banks, symbol pads, phenomenaria, construction kits, and task managers”. This paper focuses on information banks, which are sources or repositories of information such as textbooks, teachers, encyclopedia, and other such media.

Irrespective of whether the design is based on an information delivery or student-centered model, an information bank is always an important element. In pre-Internet times when libraries as text repositories were the main reference source, the challenge was always to find sufficient information to meet user needs. Today the challenge is to find useful and relevant information from the flood of often dubious data available online. It has often been compared to finding a needle in a haystack. Hence the role of the information bank in a learning environment needs to change to meet these technological changes.

Information Discovery Data Model

Ip, Currie and Morrison (2000a) have previously investigated the adequacy of commercial search engines in locating and describing appropriate resources on the Web for educational use. They argued that using the criteria of size, coverage and precision, general search services were inadequate for the task of accessing the most appropriate resources on the Web. One solution to this problem has been the application of standardized metadata tags describing key features of the resource.

Ip, Currie, Morrison and Mason (1999) proposed a data model to explain the relationship between resources, metadata and the search process. This model uses three data types and two processes to describe resource discovery. Type 1 data refers to the resources, Type 2 to the database supporting the search service. Type 3 includes the mechanisms that control the selection of metadata elements to be included in the Type 2
Search sites build their value by Type 2 data collection and search fulfillment which is the process of matching the search query with the Type 2 data, providing a set of search results and optionally applying a "ranking" algorithm to the search results before presentation to the searcher. It is important to note that most search services do not store the Type 1 data. Via a hyperlinking mechanism, the searchers are directed to the original location where the resource resides.

**Subject Gateways**

The effective use of metadata is best demonstrated with specialized resource aggregators called Subject Gateways (SG). Subject Gateways are subject-based resource discovery guides that provide links to information resources (documents, collections, sites or services), predominantly accessible via the Internet. Resource description and subject classification are the most important characteristics of such guides. [HREF19].

Established Gateways in Australia include Agrigate, AVEL (Australasian Virtual Engineering Library), ADT(Australian Digital Theses Program), Australian Science at Work, Bright Sparcs, Education Network Australia (EdNA), MetaChem, PADI (Preserving Access to Digital Information). Other gateways under development in Australia are AUSTLIT: the Australian Literature Gateway, Australian Trade Union Heritage Resources Gateway, Lawzone and Weblaw.

Koch (2000) refers to a spectrum of services, from "... link lists with minimal description and shallow subject structure (called subject gateways) to subject services with high standards for quality-control and with rich description and structure, (called quality-controlled subject gateways)".

Subject gateways exist to meet the resource needs of their communities of interest. They tend to place a higher value on returning quality resources than general search engines - however we may interpret 'quality'. The value of any SG is in finding more efficient and effective ways to build, index and facilitate retrieval from special interest collections sifted or culled from the massive underlying search space. This sifting and culling depends on the domain knowledge of the SG owners. Such domain expertise is hard to find and replicate in general purpose search services.

Most SGs utilise sophisticated metadata schemas as opposed to a general dearth of metadata on the Web at large, as noted by Lawrence and Giles (1999). These sites are normally characterized by relatively small collections and associated resource descriptors, the use of sophisticated metadata and the adoption of metadata standards. What they do offer is a high degree of selectivity towards resources that offer special value to their identified community.

**Non-educationally Focused Resources**

One of the primary metadata standards in the world is the Dublin Core Metadata Initiative. The DCMI Education Working Group [HREF2] states as its role, "The objectives of the Working Group in 2000-2001 are to continue discussion and development of proposals for the use of Dublin Core metadata in the description of educational resources" (our italics). The authors contend that there is an assumption that educational metadata should be applied to resources that were created 'for educational use'. In fact, there are many resources that were not created with any educational purpose in mind but which have been used or adapted by course designers and educators. This is often the case with information from the Web. The authors created the term 'non-educationally-focused' (NEF) resources to describe this type of information (Ip, Morrison, Currie and Mason, 2000). The extensive use of NEF resources is a feature of student-centred learning.

This subtle but significant difference between 'educational resources' and 'educational description of resources' needs to be understood by current educational metadata efforts in order to create an environment for SGs to operate in a collaborative manner. It seems to us that education metadata standards setting efforts have failed to address the specific issues related to the use of NEF resources.

Without a standard to describe NEF resources, SGs are forced to implement ad hoc mechanisms to enable their users to create and support the use of information-dense learning environments.

**Quality of Search**

A simple interpretation of 'quality of search' may be fitness for purpose. Search sites aim to meet the needs of their user community. A site that purports to focus on, say, Curriculum and Standards Framework Level 4 resources, can be assessed on how well it matches that aim independently of the quality of the resources provided. That is, we need to differentiate between the quality of the resource and the quality of the site in meeting its stated aim.

Based on the model suggested by Ip, et al (1999), the quality and coverage of any search is only as good as the quality and breadth of the Type 2 data used to support the search. The size of the collection (i.e., breadth) of the Type 2 data determines the coverage or the comprehensiveness of the search results. Because they are not owned by the search site, the quality of the linked resources is an outcome of the selection processes employed based on the search site owners' domain subject expertise. The value of any subject gateway is the 'trust' it has from its users in providing this authoritative assessment of quality.

However, the overall quality of the search result (i.e., appropriateness and precision) depends on the quality of the Type 2 data supported by the ranking algorithm in use by the search engine. To achieve this quality, specialized search sites would depend on the sophistication of the metadata (Type 3 data), the coverage of the resources within.
the confines of their charter (size of Type 2 data), the quality of the Type 1 resources being included (selectively applied when including resources into the Type 2 data set) and the ability to return the appropriate results (Type 3 data) based on their unique understanding of their clientele.

The major issue faced by general purpose search engines in their Type 2 data representation of the collection is the lack of a semantic context as to how a keyword or phrase is used in the original Type 1 resource. This will have significant impact on the ability of search engines to rank the search results. Metadata potentially can solve this issue by developing a sophisticated ontology of the classification scheme and mapping keywords into concepts utilising a thesaurus or “open” or “controlled” vocabulary. Fig. 2 shows the use of a thesaurus by a searcher to assist in providing more contextual information about the keyword being used in a search process on EdNA Online. Such technique would require the availability of a thesaurus at the client side that matches the underlying Type 2 data classification used by EdNA Online.

The effectiveness of a search engine, irrespective of the underlying mechanism, is to be judged by the final user experience. Among the six criteria proposed by Lancaster and Fayen (1973) ‘Coverage’ and ‘Precision’ are directly related to the underlying Type 2 data quality.

Collaboration v Autonomy

While subject gateways enable their target group to access a range of quality resources, the relatively small size of their collections can limit wider use and promotion. Scalability is becoming a problem as the growth of the Internet increases. (Huxley, 2000). It is increasingly recognized that raising the level of collaboration between gateways can be of benefit, particularly to those users whose search query is not precisely fulfilled by the targeted gateway. This is particularly applicable to those working in cross-disciplinary areas, e.g. someone investigating the development of farming techniques in Australia could be interested in agriculture (Agrigate), farm machinery (AVEL) and agricultural chemicals (MetaChem).

Cooperation to varying levels already exists. The Australian Subject Gateway Forum [HREF 3] is an example as is the RDN in the UK. Areas of cooperation include software and metadata standardization, research and policy sharing, intellectual property rights and common terminology.

A number of international projects are also addressing collaboration issues. The Isaac Project [HREF 4] and IMesh Toolkit [HREF 5] are concerned with technical issues such as the development of compatible software for managing sites, while the CORC project [HREF 6] aims to assist resource sharing between participating libraries. The UK Metadata for Education Group (MEG) [HREF 7] is a new group aimed at improving collaboration between UK libraries. While it does not specifically address resource sharing, its Concord sets principles for interoperable metadata standards and resource descriptions. The Renardus project’s [HREF 8] aim is to build an academic subject gateway service for Europe by developing a system that will ’broker’ data from a range of existing distributed gateways and other Internet-accessible collections across Europe. (Huxley, 2000)

However cooperation has always been a thorny issue, at odds with the valued autonomy of subject gateways. Each of these subject gateways has arisen out of an identified need of a specific user group at a particular point in time. Funding for gateways has traditionally been tightly focused (e.g. the Australian ARC grants) and usually managed by a particular university. It is this independence that enables the unique focus and organization that adds value to the SG. The value created by SG is tightly coupled to the expertise of the SG owner or the key participants in the specific subject domain in which SG operates. This is a feature of the database that holds the collected Type 2 data and the specific adaptation of metadata standards (Type 3 data in Ip’s model). No doubt, many owners query the usefulness of collaboration or have legitimate concerns that collaboration has the potential to diminish their particular focus.

Domain-Specific Issues

All subject domains have their own specific requirements and it is unlikely that any semantic standardisation can meet all the needs of every domain. In other words, domain specific efforts in standardising metadata must remain within the domain specific community. This is the rationale behind the decision to make Dublin Core elements extensible by additional qualifiers.

At noted above, a prime value of subject gateways is the ‘trust’ its community users have and the unique understanding of the clientele demonstrated in the customisation of the underlying Type 2 data used to provide the search service. Forcing all subject gateways into the same metadata standard (Type 3 data) would not be welcomed by the subject gateway owners. For example, the common elements of Teaching and Learning Databases are about ‘education’ and ‘pedagogy’, and standards for these categories will most probably be based on the IMS schema. However, there is a need to capture the domain specific issues in the Type 2 data as well. If a SG is devoted to only one subject domain, it can use a subject specific extension to the metadata standard, such as the MetaChem standard for chemistry. The knowledge and experience of the domain-specific SG owner in creating metadata is adding value to the database in the view of its users. However, the domain specificity of the metadata standard may compromise the ability of the SG to be searched by other sites. There is a conflict between the specificity of the data, and the need for interoperability between SGs.

Any collaborative framework must recognise this apparent contradiction and work creatively to enable a solution. One suggestion may be that the collaborative framework specifies the standard of expressing domain specific semantics, so that cross-domain searches may be done without the search engine really understanding the
semantics. In other words, when a domain-specific query is passed to a SG that knows of other SGs which may fulfill the search, it would only need to be able to pass the query in standard format without full understanding of the domain specific details.

**Policy Implications**

The Web is aptly named, given that it is an ever-expanding web of inter-connectedness, of networks and nodes, and rich in information resources and communications opportunities. The 'information society' has arrived and is setting the scene for social, economic and cultural re-configuration for the new millennium. The proposed strategy of establishing a collaboration framework needs to leverage upon both the creative use of information resources and the creative use of the communication opportunities.

This strategy is in line with a Commonwealth's "Strategic Framework for the Information Economy" (December 1998) [HREF 9] which states that, "The private sector is driving, and will continue to drive, the transition to the information economy."

This sentence is taken in the context of "... the role of governments to provide an environment conducive to investment in new technology, to the formation and growth of new enterprises, and to the acquisition of information technology skills and knowledge ...", but we believe the underlying principle is valid for a strategy for enabling collaboration and competition between SGs.

**Interoperability between Subject Gateways**

Technical 'interoperability' refers to the technical communication protocols for data exchange between Information Resource services. While some efforts are focusing on this, they do not recognize the different value-added contributions each SG adds to the information they provide within their specialist area. A longer-term solution should address broader issues than just technical interoperability.

A collaboration framework that allows these gateways to be bridged in an easy and transparent way will:

- produce greater value out of the considerable resources that have been committed to supporting various SG initiatives.
- recognise and support the value of the diversity
- facilitate cooperation and collaboration on a level playing field between SGs
- provide a common tool set
- improve end user experiences (via different levels of engagement within the framework)
- develop mechanisms to take advantage of the value added to the data by content experts.

Ip’s data model articulates the existence of three data types which could be used to understand the different level of engagement or commitment to collaboration. We propose a staged collaboration framework with different levels of engagement between SG providers based on this understanding:

- Minimum
- Intermediate
- High
- Maximum

**Minimum**

This is the beginning of engagement among SGs. It may be as simple as exchanging of banners and engaging in discussion. When a searcher is with a SG, banners of other SG are available to encourage them to try other SGs. Exchange of banners is currently being tried in several gateways.

A forum or dialog between SGs may have been established either formally (as in the case of Australian Subject Gateways) or informally through personal contacts between SG owners in conferences. For example, at the first meeting of Subject Gateway owners held on 22nd February, 2000, the forum discussed issues such as use of a common terminology, selection of software, IP issues pertaining to the ownership of the metadata and the issue of sharing, choice of metadata standards, branding and promotion, research into international subject gateways collaboration initiatives and issue of sustainability. Clearly, there is value to the SGs even at this level of collaboration.

**Intermediate**

This is the extension of the minimum level of engagement. Each SG is able to identify the focus of the other SGs whose banners are hosted and to make recommendations based on key words to target specific hosted gateways. In other words, SG owners are starting to understand the other SGs domain of expertise and can make sensible recommendation to their own community. At this level of engagement, the 'trust' is passed on to the other SG. However, the main asset (Type 2 and Type 3 data) of individual SGs is not compromised or shared.

**High**

This would involve some forms of transparent searching across a range of SGs based on search APIs provided by other SGs. This cross-searching would necessitate an agreement to use compatible protocols. This can be unidirectional such as the provision of a search API by one gateway to other gateways or to client organisations. In some versions, participating SGs will share a common minimum set of metadata.

In assigning information keywords, categories, metatags and associations to other information, a SG is enriching...
The information considerably. This enrichment does not reside with the Type 1 data but with the associations inherent in the process of ‘locating’ it within a body of information (i.e. Type 2 data). For the processes involved with cross-searching SGs, common metadata (Type 2 data in our data model) is the main asset used by persons looking for information. However, different SGs (due to the nature of the domain or otherwise) will add additional elements - extensions which may not be defined in standards such as DC or IMS. There are different underlying data models of such ‘extensions’ and such underlying data models will contain the main ‘value-adding’ component which was inserted by the SG expert.

Such rules, coding mechanisms, classifications and so on have been identified as Type 3 data. However, without an understanding of the Type 3 data of other SGs, a hosting SG generally cannot meaningfully interpret the search results and can only return the result as provided from the underlying supported SG. Some form of integration, such as reduction of duplication may be performed, but ranking is unlikely to be effective. This form of collaboration involves both the propagation of “trust” and the sharing of assets (in the form of Type 2 data) by providing services to meet other SG needs. However, the expertise of the subject gateway owners remains with the individual SG in the sense that other participating SG is not able to contribute to building up the size of other SG because the underlying selection criteria, special additional metadata elements and the enhanced schemas are not shared or understood.

**Maximum**

Maximum interoperability involves full interoperability at both syntactic and semantic levels (utilising Type 2 & Type 3 data).

Type 3 data is an important part of the valuable asset of any SG, assisting it to remain competitive. No doubt, the value is expressed as Type 2 data of the SG, however, those rules which applied to their Type 2 data define the structure of the Type 2 data. Any collaborative framework which attempts to shrink-fit SGs into the same semantic model and stay at that level of engagement will fail because it will lack the ability to integrate this value-added attribute. Any successful approach to bridging SGs will need to ensure that gateway developers share their Type 3 data in a way which is transparent, easily navigated and yet without compromising the value-adding ability of SGs.

The importance of recognising Type 3 data is gaining international attention. Most metadata organisations including the ISO [HREF 10] and the DCMI [HREF 11] use the term registry to refer to the database of Type 3 data and repository to the database of Type 2 data. The clear articulation of the concept is a positive step towards a solution, however an immense amount of work is still needed.

Maximum level of engagement describes this ideal situation at which there is transparent machine sharing of Type 3 data or meaningful access to registry. The work of W3C consortium in defining the “semantic web” is moving in a direction to meet this demand.

**Collaboration Framework**

This collaboration framework is built on an understanding of the basic value of SGs. At the lowest level of collaboration, SGs are sharing their “trust” by recommending services from other gateways. However, such recommendations are not based on a full understanding of the other SGs. As the collaboration level increases, this understanding of the other SGs becomes more apparent. SGs are starting to engage in sharing not only the intangible notion of ‘trust’, but the value that has been built, i.e. the Type 2 data in the form of search APIs. However, before reaching a maximum level of collaboration, the notion of competition still limits the full exchange of the expertise as expressed in the Type 3 data.

Even if the political will exists, the transition from High to Maximum level of engagement of collaboration involves serious technical hurdles. At an early stage of the High level of engagement, a subset of elements which form the common core of the collaborating SGs may be shared. This is what Renardus is attempting to achieve [HREF 12]. Mega-searching among the collaborating SGs operates on these common elements.

Our own experience in performing the cross-walking between IEEE LOM and IMS metadata specification reveals that some elements in one specification may be mapped to the other specification. However, there are a number of elements which cannot be satisfactorily mapped. While the semantics of these elements are in the same general area, there are sufficient differences to force us to keep both. For those elements which have the same semantic meaning, the list of best practice keywords and controlled vocabularies presents yet another challenge. If our cross-walking of two specifications within the education domain is any indication of the issue, any cross-walking exercise between metadata specification from wider domain will definitely be a very difficult, if not impossible task.

A possible solution is to make use of the knowledge we can gain from cross-walking exercise and allow for ‘multi-stage’ searching. The initial search may be based on the limited number of elements which are either common or have been through a cross-walking exercise and hence are mapped as having the same general semantic meaning. Further refinement of the search will have to depend on the underlying SG which provided the service.

The ability to express the semantic meaning of elements depends on the expressive ability of the coding mechanism for Type 3 data. Technical mechanisms for expressing Type 3 data are in the horizon. One such possibility is using W3C RDF to describe knowledge domains and express Type 3 data in RDF.

Collaboration framework is more than just the technical possibility of cross machine searching (meaningful mega-searching). The recognition of the value adding and creation of a business model to support this value chain is equally important.
Conclusion

The increasing size and commercialization of the Internet continues to present challenges to those wishing to effectively access online resources in an information-dense learning environment. While Subject Gateways are able to provide quality resources that meet the needs of their particular audiences, there exists a need to create a synergy of effort through the development and implementation of increasingly transparent and powerful linkages between gateways. Ip's data model sheds some light on a possible path of increasing level of collaboration. Further research and discussion in this area will provide an effective argument for sustainability and ongoing investment in subject gateways.

References

Broder, A. et al. (2000). Graph Structure in the Web [HREF 13]

National Library of Australia (2000). 'Summary of outcomes of an Australian subject gateway owners' meeting on 22nd February 2000' [HREF 18]

Hypertext References

HREF1 http://www.edna.edu.au/
HREF 1a http://www.unimelb.edu.au
HREF 1b http://www.dis.unimelb.edu.au
HREF2 http://dublincore.org/groups/education/
HREF5 http://www.imesh.org/toolkit/
HREF6 http://www.oclc.org/oclc/corc/
HREF7 http://www.ukoln.ac.uk/metadata/education/documents/concord.html
HREF 8 http://www.renardus.org
HREF 10 http://hmrha.hirs.osd.mil/mrc/
HREF 11
Copyright

Albert Ip, Michael Currie, Iain Morrison, © 2001. The authors assign to Southern Cross University and other educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Southern Cross University to publish this document in full on the World Wide Web and on CD-ROM and in printed form with the conference papers and for the document to be published on mirrors on the World Wide Web.