INTERACT
INTEGRATE
IMPACT

Proceedings of the 20th Annual Conference
of the Australasian Society for Computers in
Learning in Tertiary Education (ASCILITE)

Adelaide, Australia
7–10 December 2003

Editors
Geoffrey Crisp, Di Thiele, Ingrid Scholten, Sandra Barker, Judi Baron

ascilite
**Australasian Society For Computers
In Learning In Tertiary Education**

THE TECHNICAL DEVELOPMENT AND BENEFITS OF A METADATA AGGREGATION AND INSERTION TOOL

Kieren Pitts and Kate Sharp
Institute for Learning and Research Technology
University of Bristol, Bristol, UK
Kieren.Pitts@bristol.ac.uk

Abstract

This paper describes the technical development of a metadata aggregation, management and insertion tool known as M.A.I.M. (MAIM Aggregates and Inserts Metadata). The paper will examine the methodology and technologies used and the benefits of such a tool to a large online educational resource. The paper will also present a case study discussing the use of MAIM on the Biz/ed service - a unique service for students, teachers and lecturers of business, economics, accounting, leisure and recreation and travel and tourism. This paper will also discuss some of the navigational and resource integration tools that the MAIM software makes possible and how some of these have been integrated within the Biz/ed service.

Keywords

MAIM, metadata, aggregation, insertion, e-learning, distributed authors, online learning, resource integration, resource discovery, Perl, PostgreSQL

Introduction

One of the key components to any online resource or learning object ("any entity, digital or non-digital, that may be used for learning, education or training." (IEEE LTSC, 2002)) is metadata - "data that describes other data" (CETIS, 2001). Metadata is structured data used to describe a resource and can provide information including a textual description, author details, creation date and rights management information for a document. Most popular Web search engines do not rely on metadata to describe or rank search results as this method is open to abuse through "key word spamming" (Stanley, 1997). However, metadata has a vital role in describing learning objects (Currier & Barton, 2003), site searching (searches limited within a Web site) and for content management. Metadata is almost entirely hidden from the average user viewing a Web page using a traditional browser but it can be used to enrich their browsing experience. For example, users may be able to view dynamic lists of associated resources or merely to navigate up to the parent page for a particular section through URIs (Uniform Resource Identifiers) given in the metadata - this can be particularly important if a site is entered from a link in search engine results. In terms of content management out of date content can be quickly identified and author information can be listed within a Web page using metadata. In terms of e-learning, metadata can be used to identify resources relevant for individual topics and aid subsequent packaging of related content into discrete units.

There are a large number of metadata standards in use but, as Dekkers and Weibel (2003) state, the Dublin Core Metadata standard is the primary online metadata standard. The Dublin Core Metadata Element Set (DCMI, 2003) contains fifteen elements allowing the broad description of any online resource. The Dublin Core Element Set is intended to be easy to create and CETIS (2002) found that over 91% of respondents involved in delivering online learning materials or systems within the UK education sector were aware of and/or using Dublin Core metadata. However a similar study in Finland found that only 21% of Web sites for educational institutions contained metadata of any kind (Hirvasniemi & Öörni, 2001).

Despite the obvious benefits of using metadata there are a number of problems restricting its use:

- Generation - it is time consuming and not always easy to produce good metadata. The author of a resource may not have sufficient skills to produce an HTML document or, as discussed by Currier & Barton (2003), may not be the most appropriate and consistent metadata creator. Expert metadata authors may be required to produce accurate metadata and validate data entered by others.

- Ease of updating - in most cases metadata will be held within the code of the HTML page. If these data are to be edited directly then it requires someone with adequate technical training and access to the Web server.

- Distributed information - metadata is usually held only within an individual Web page. Under these circumstances it is difficult to cross search for related resources in a large Web site.

- Consistency - large educational resources frequently have teams of distributed authors working on different areas of content. Consequently it is difficult to maintain consistency within the metadata and wide variation in metadata quality can lead to a decrease in its effectiveness.

- Error checking and reporting - when metadata is held solely within the pages of a Web site it is difficult to check for errors.

The aim of this paper is to discuss the design and implementation of software tools enabling the management of metadata for large educational resources. The tools represent an ongoing project with the ultimate aim being a suite of tools that can be made freely available to facilitate the centralisation of metadata and an increase in interoperability between e-learning resources.

Initial Development

Variation in the level of awareness and use of metadata in educational institutions and the potential difficulties involved with adding and maintaining metadata across a large Web site led to the development of a suite of tools designed to address these problems. The primary aim was to produce a set of open-source tools that could be used to maintain metadata for any online resource. These tools were developed at the Institute for Learning and Research Technology (ILRT, 2003) by the principal author. During March 2002 work began on the suite of tools (and associated database) collectively known as M.A.I.M. (MAIM Aggregates and Inserts Metadata). The four key areas addressed by these tools are:

- Harvesting - as outlined above the level of metadata use within educational Web sites is very varied. However, whatever the level of metadata available it is important that it is harvested in to the database automatically and no additional effort is required to re-enter existing data. In addition, when new pages are added to a site metadata should be harvested into the database rather than the data produced using the editing tools and inserted back into the pages (although both techniques should be possible). The metadata harvester is a command line program written in Perl (Perl, 2002). It requests Web pages in the same manner as a traditional Web browser and extracts data from the HTML meta tags. Harvested metadata is stored in a PostgreSQL database (PostgreSQL, 2003). The harvester also assigns a unique ID number to each page and stores the location of the file on the file system of the Web server. The harvesting tool can be scheduled to run at regular intervals and, on encountering a new Web page (i.e. one that has not already been entered into the database), the metadata is harvested into the database. The MAIM software therefore maintains an incrementally updated copy of all the metadata (plus filename and directory paths) for the static HTML pages on a Web site.

- Editing - metadata editing might be undertaken by numerous distributed authors, therefore a Web-based front-end allowing data to be viewed, entered and/or altered using simple forms was required. The metadata editor is written in Perl and changes to the metadata made through a series of Web forms. Changes are written back to the database and that resource given a 'modified' flag.

- Insertion - when metadata has been changed it must be inserted back into the relevant page, this is achieved through a command line Perl program. Using a list of updated records from the database (identified by the flagging process described above) the inserter removes existing metadata from a Web page and replaces it with updated data. The updated document is then saved back to the file system replacing the original. The insertion tool can be run on demand or set to run at regular intervals.

- Validation - the MAIM tool can identify Web pages missing important metadata (e.g. descriptions or key words). The validation tool is written in Perl and through a series of Web pages users can identify and fix errors. After correcting the mistake the changes are written back to the database and the resource flagged as modified. The metadata will then be corrected during the next run of the insertion tool.

Dublin Core is the most widely used metadata standard and the tools were designed to use Dublin Core natively. However, Dublin Core is not the only metadata standard and in order to make the MAIM software suite as useful as possible it is designed to work with a different standard if necessary.

Biz/ed Case Study

Biz/ed (Biz/ed, 2003) is a unique service for students, teachers and lecturers of business, economics, accounting, leisure and recreation and travel and tourism. Prior to September 2002 the service contained some 14,000 unique URIs comprising roughly 4,700 static XHTML documents (XHTML, 2003) each containing seven pieces of Dublin Core metadata. The validity and accuracy of this data could not be checked without examining each document individually. Furthermore the Biz/ed Web site was so huge and covered such a diverse range of resources that, despite the inclusion of a site search; site map and numerous help pages, it was often described as being difficult to navigate and find specific resources quickly.

MAIMing Biz/ed

During the summer of 2002 the Biz/ed Web site underwent a major redesign and during this time the existing metadata was harvested in to the MAIM software. Three content developers for the Biz/ed service then updated the existing metadata and added new information. After accessing the MAIM metadata editor the content developers were presented with a list of directories each containing a list of HTML files. Through this file listing the metadata for each page could be edited (Fig. 1).
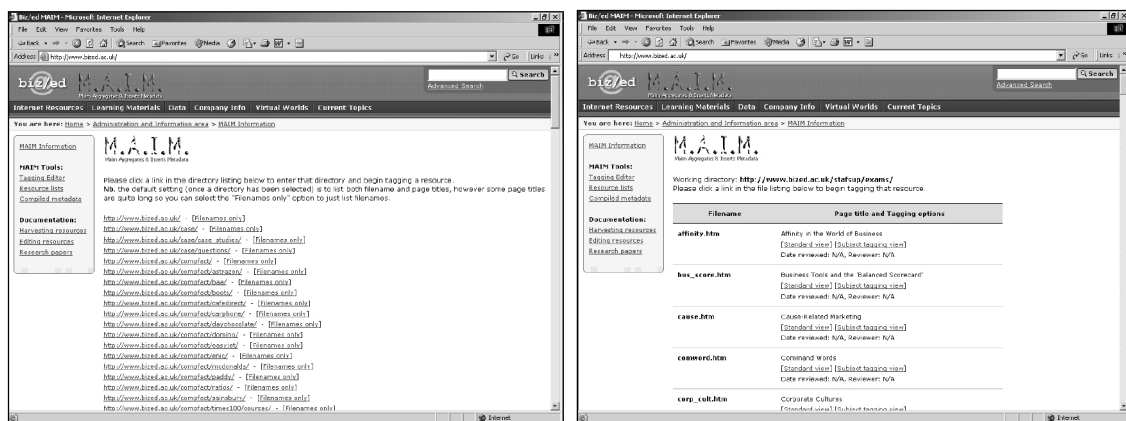

Figure 1 - Directory listings (left) and file listings (right) selection pages.

Once a file had been selected the existing metadata for that page was displayed in a Web form. The data could then be reviewed, edited and resubmitted (Fig. 2).
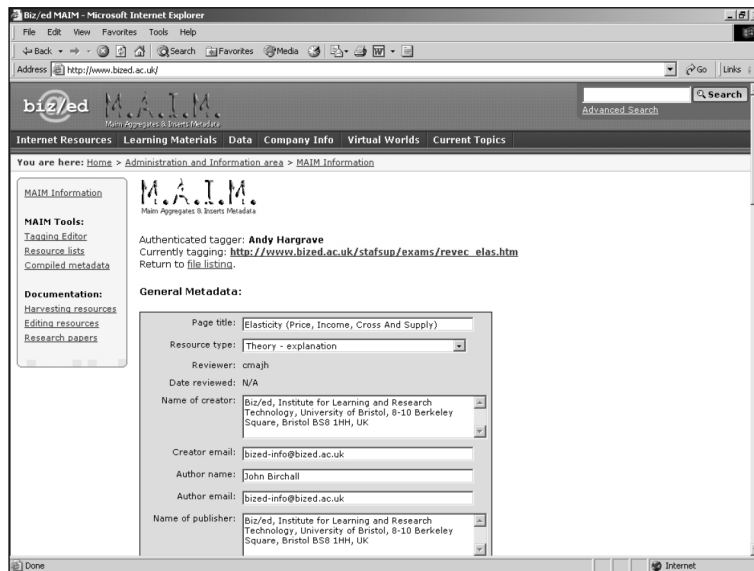
Figure 2 - the current MAIM metadata editor.

Using data entered and maintained through the MAIM software it was possible to develop and employ novel navigation methods on the Biz/ed Web site. One such development involved the tagging of individual Web pages according to their relevance to a particular subject, academic level and, where appropriate, examination body (Fig. 3). The content developers could also provide a list of relevant terms associated with each page that could be searched upon using a glossary tool already available on the site. Using a similar technique the content developers were able to add related terms that could be used to search within an archive of business news items ("In the News" or Chronology terms), a catalogue of related Internet resources (Catalogue terms) and a list of theory explanations (Theory terms) (Fig. 4).
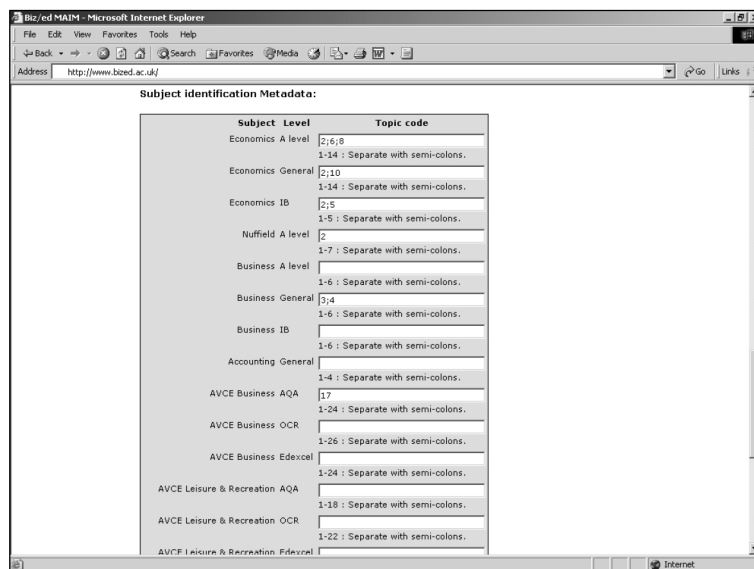

Figure 3 - The MAIM metadata editor illustrating the tagging of a resource for its relevance to individual subjects and academic levels.
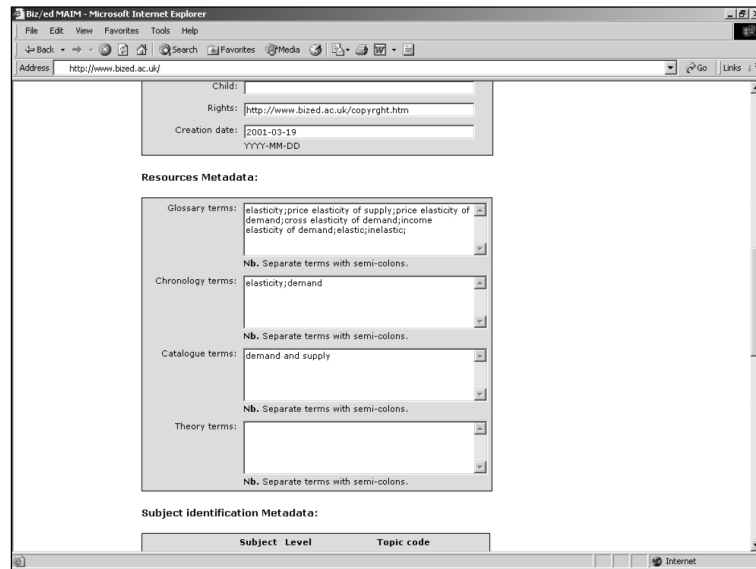
Figure 4 - MAIM metadata editor showing entry of Glossary, Chronology, Catalogue and Theory terms.

At this time the error checking part of the MAIM software also generated a list of Web pages that had not been reviewed. This listing provided an indication of how much of the Web site had been reviewed and made sure all of the resources had been checked prior to the relaunch of the site.

Once metadata for all the pages had been reviewed the insertion tool was used to insert the new data, this data comprised some 18 tags as opposed to the original seven. The increase was due to changes in the Dublin Core standard and the additional amount of information needed to accurately describe each page. The metadata insertion tool was also used to generate lists of Web pages applicable to different subjects, academic levels and examination bodies. A "Quickjump" box (Fig. 5) was added to the Biz/ed home page allowing access to these resource lists. Students used three drop down boxes to identify the course they were studying, on submission a Perl script (developed separately from the MAIM software) redirected students to resources directly relevant to their course. The list of resources was broken down to individual topics and by learning resource type, for example interactive worksheets or theory explanations (Fig. 6).
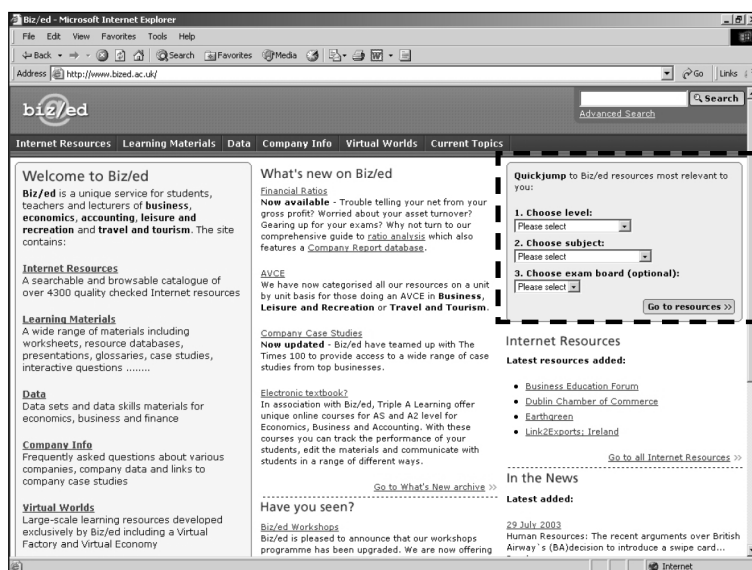

Figure 5 - Biz/ed homepage with the Quickjump box highlighted (dashed box). This allows students to directly access resources relevant to their course.
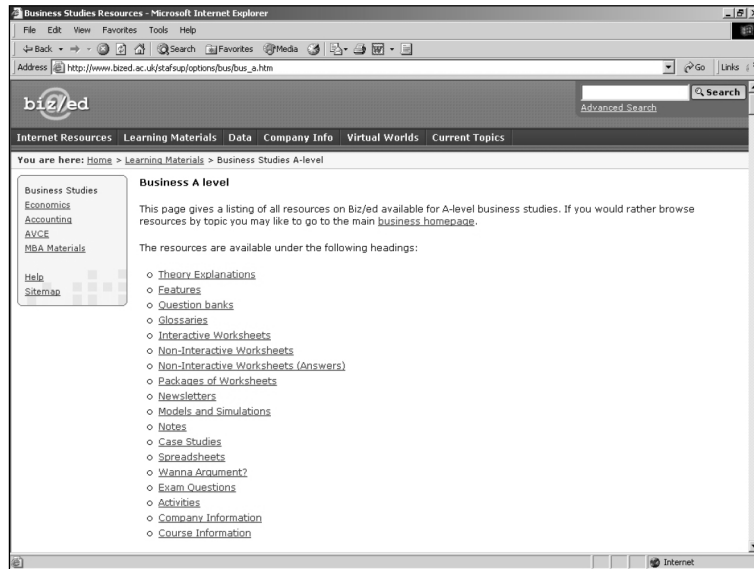
*Figure 6 - The list of Biz/ed resources displayed to a student entering "A-level" and
"Business Studies" in to the Quickjump box on the Biz/ed homepage.*

The insertion tool also produced sub-navigation drop down menus that allowed access to the Biz/ed glossary, "In the News" archive, "Internet Resources" catalogue and a set of theory files (Fig. 7). After selecting a resource from the drop down list a Perl script (developed separately from the MAIM software) redirected the student to the relevant resource. For example a student selecting term "time series" from the Definitions (Glossary) drop down menu would be redirected to a Web page within the Biz/ed glossary containing the definition of "time series". By maintaining this information within the MAIM software it could be edited and terms added, validated or removed as necessary.
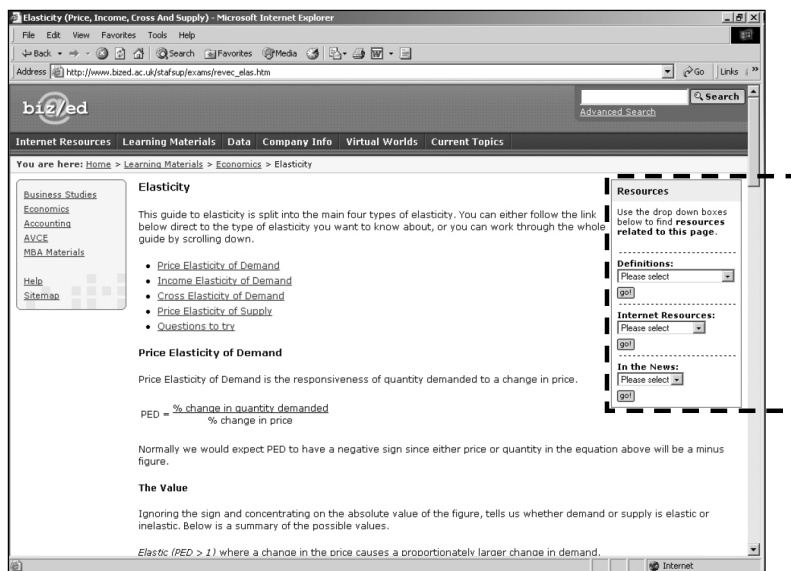

Figure 7 - The additional navigation tools present on many of Biz/ed's pages made possible by MAIM.

The results of the MAIMing process

After the relaunch of the Biz/ed Web site a marked increase in the number of monthly page accesses was observed reaching over twice the pre-launch level within only three months of the relaunch (Fig. 8).
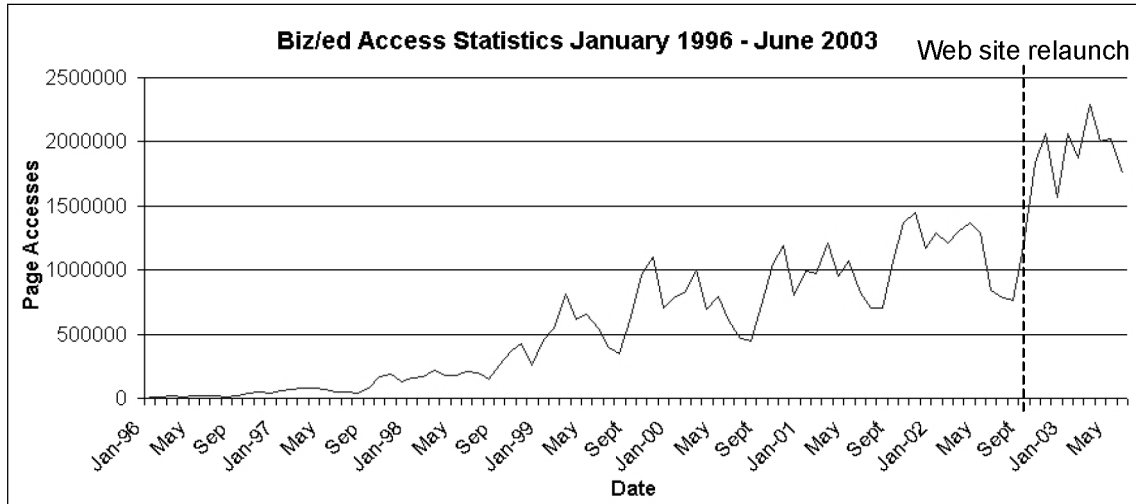
Figure 8 - Monthly page access statistics for the Biz/ed Web site before and after the relaunch of the Web site during September 2002. The dashed line represents the date of the Biz/ed relaunch.

Obviously some of the increase in traffic was due to the redesign of the Web site and could not be directly attributed to the use of data stored by the MAIM software. However examination of the log files for the Quickjump box and the drop down navigation tools provided an insight in to the usage of these navigation tools.

Quickjump box
Basic analysis of the log file for the Quickjump log file showed that it is used an average of 1004 times every day.
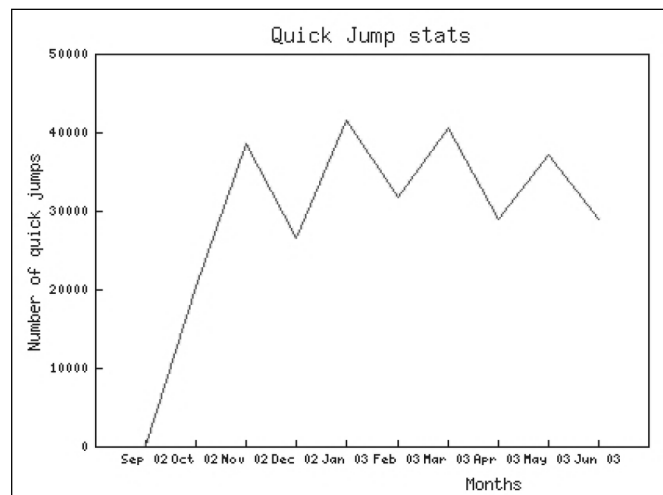

Figure 9 - The pattern of usage of the Quickjump box located on the home page of the Biz/ed site since its implementation in September 2002.

In addition, analysis of the Quickjump box usage log provided an indication of the number of students studying particular subjects and using the Biz/ed service. This type of information would not usually be available without asking users to register before accessing the site or complete a user survey. It was found that the three most popular groups of students were A-level Business Studies (22.8%), A-level Economics (17.7%) and AVCE Business (16.9%).

Resource menu

Examination of usage logs for the four types of drop down menu available on the Biz/ed site showed that they are used daily and considering the fact that these resources are not available on all pages within the Biz/ed site and contain differing numbers of terms they do represent a useful and popular navigational tool (Table. 1).

Drop down menu	Click throughs per day/number of terms
Glossary terms (Definitions)	83/3704 = 0.0224
In the News (archive of news items)	17/1827 = 0.0095
Internet Resources (catalogue of reviewed Web sites)	76/2340 = 0.0326
Theory (links to Theory explanations)	4/301 = 0.0124

Table 1 - shows the pattern of usage for the four drop down navigation tools.

Added value for users

The drop down menus and Quickjump box have improved resource discovery and offer new ways to access information but despite the increase in traffic on the Web site it is difficult to measure the effect of the additional navigational tools themselves. These tools were specifically designed to seamlessly integrate with the existing Web site and although we can examine usage figures (as above) no comparisons can be drawn as these tools are novel and no equivalent was available prior to the use of the MAIM software. However, positive email feedback and a decrease in the number of requests for help indicate that these tools aid students in finding relevant resources and understanding the resources they do find (Biz/ed, pers. comm.). For example:

"What a fantastic site! I just stumbled across it by accident, but I'll be certain to spend some more time here for some time to come." Dan Kiser, US (student)

"In the development of this course I have found Biz/ed to be supremely useful, and I can not thank you enough for providing this service. The materials are of a high quality, whilst being succinct and simple to access." Robin Tankard, UK (IT Tutor)

The Quickjump box in particular offers new ways to access information by providing a global view of relevant resources for a topic without replacing the traditional browse route whereby students move between pages on a topic in a sequential fashion. In addition to raising the profile of resources within the Biz/ed service the global view may benefit students not suited to a more linear, sequential learning style. A further improvement is the inclusion of drop down boxes of terms on the pages on which they occur. Students no longer need to stop and refer to offline dictionaries or glossaries but can instantly look up terms they are unfamiliar with as they progress through the learning materials - since the implementation of this feature there has been a distinct decrease in the number of emails received by the team asking for clarification of terms and an increase in positive feedback (Biz/ed, pers. comm.).

"I find your website really helpful, especially the definitions and all in the glossary"
Vijay Patel, UK (student)

"I am finding that your economics glossary definitions are straight and to the point, and the graphs are fantastic." Jessica Tirado, USA (student)

The addition of the "In the News" drop down menu is of particular benefit to students more suited to an intuitive learning style as these students can use the menus to explore relationships between theories and actual news stories.

A further benefit of the Quickjump feature on the home page is that resources relevant to a topic are categorised by the type of resource. Consequently students preferring more active learning methods can try out interactive worksheets and simulations whilst those preferring a reflective approach can select

static resources and work trough them at their own pace. A further benefit of the Quickjump box is that it can be used to collect demographic information on users without requiring them to register or supply any personal data. This is extremely useful in the development of new resources as under-represented subject areas and/or popular subjects can be identified and authors commissioned to produce resources.

In addition to the added value for end-users the content developers and curriculum specialists have also been very complimentary of the MAIM software, for example:

"I have found the MAIM software very straightforward and user friendly."
Gemma McShane, Curriculum Advisor.

Further developments

Biz/ed employs a team of subject authors on a temporary basis, these authors produce resources within a topic area and once available on the site they are harvested into the MAIM database. It is important that these authors can only edit the metadata for resources they have developed. To meet this requirement the MAIM software was altered during March 2003 to incorporate user roles ranging from developer (access to all data for all resources) to a subject author (access to a sub-set of metadata for resources they have written). These changes made it possible for a team of Scottish curriculum experts to add metadata relating Biz/ed resources to subjects covered in Scotland - prior to this the subject specific tagging of Biz/ed only applied to the English and Welsh curricula.

The Biz/ed Web site features a wide range of resources to suit all learning styles and these can contain highly interactive content. Since accessibility is of paramount importance to all educational resources it is important to be able to identify resources that might present accessibility issues. These issues can then be addressed as applicable. Using the MAIM software on Biz/ed all of the resources on the site were identified as belonging to one of 36 categories, using this system lists of resources containing interactive content can quickly be produced and potential problems identified.

This paper has been concerned with the use of metadata within (X)HTML documents but other file formats can also contain metadata and this is likely to become increasingly widespread. Using the MAIM software any resource linked from a Web page on a Web site can have metadata associated with it. Depending on the file format this data may be able to be incorporated directly into the file itself, for example both PNG (Portable Network Graphics) and SVG (Scalable Vector Graphics) can have metadata incorporated within them.

Conclusions

We hope that through this paper we have not only demonstrated the importance and usefulness of metadata within educational resources but also given a practical description of how a metadata aggregation and insertion tool has made an important difference to the way that a large online resource is used. The MAIM tool has not only provided multiple routes throughout a resource tailored to a students' preferred learning style but also provided demographic information on courses being taken by students using the Web site - this allows the content team to develop resources for particular curriculum areas. Despite the time investment required in producing accurate metadata we hope to have demonstrated that the benefits of this approach far outstrip the time taken to produce and maintain the data.

References

Biz/ed (2003), *Biz/ed* [Online] Available: http://www.bized.ac.uk [22 July 2003].
CETIS (2001) *CETIS Reference: Metadata. CETIS.* [Online] Available:
    http://www.cetis.ac.uk/encyclopedia/entries/20010905154527 [22 July 2003].
CETIS (2002) *Survey of Educational Metadata Use. CETIS.* Online] Available:
    http://metadata.cetis.ac.uk/usage_survey/ed_md_survey_2002-07-29.pdf [22 July 2003].

Currier, S. & Barton, J. (2003). Quality Assurance for Digital Learning Object Repositories: How Should Metadata Be Created? In Cook, J. and McConnell, D. (Eds.) (2003). *Communities of Practice*. ALT-C 2003 Research Proceedings. Held 8-10 September, 2003, University of Sheffield & Sheffield Hallam University.

DCMI (2003) *Dublin Core Metadata Element Set, Version 1.1: Reference Description* [Online] Available: http://dublincore.org/documents/2003/06/02/dces/ [24 July 2003].

Dekkers, M.& Weibel, S. (2003) State of the Dublin Core Metadata Initiative, April 2003. *D-Lib Magazine*, Vol. 9, Number 4. [Online]. Available: http://www.dlib.org/dlib/april03/weibel/04weibel.html [22 July 2003]

Hirvasniemi, S. & Öörni, K. (2001) *Educational Information in the Web: Discussing the Metadata Requirements for a Web Service Guiding Citizens' Education*. Proc. Int'l Conf. On Dublin Core and Metadata Applications.

IEEE LTSC (2002), *Draft standard for Learning Object Metadata*. [Online] Available: http://ltsc.ieee.org/doc/wg12/LOM_WD6_4.pdf [01 September 2003].

ILRT (2003), *Institute for Learning and Research Technology, University of Bristol, Bristol, UK.* [Online]. Available: http://www.ilrt.bris.ac.uk [24 July 2003]

Perl (2003), *The source for Perl*. [Online] Available: http://www.perl.com [24 July 2003].

PostgreSQL (2003), *PostgreSQL* [Online] Available: http://www.postgresql.org [24 July 2003].

Stanley, T. (1997). *Keyword Spamming: Cheat Your Way To The Top* [Online] Available: http://www.ariadne.ac.uk/issue10/search-engines/intro.html [22 July 2003].

XHTML (2003), *W3C HTML Home Page* [Online] Available: http://www.w3.org/MarkUp/ [24 July 2003].