

Enhancing Access to Open Corpus Educational Content: Learning in the Wild

Séamus Lawless

University of Dublin, Trinity College
ORI GF 31, Dublin 2, Ireland
+35318961335

slawless@cs.tcd.ie

Lucy Hederman

University of Dublin, Trinity College
ORI GF 13, Dublin 2, Ireland
+35318962245

lucy.hederman@cs.tcd.ie

Vincent Wade

University of Dublin, Trinity College
ORI GF 33, Dublin 2, Ireland
+35318962091

vincent.wade@cs.tcd.ie

ABSTRACT

The World Wide Web (WWW) provides access to a vast array of interconnected educational content on almost every subject imaginable. A great deal of this content is ideal for incorporation into personalised eLearning experiences. However the discovery, harvesting and incorporation of appropriate educational material have proven to be complex and arduous tasks.

Traditional educational hypertext systems are based upon the generation of links and anchors between content objects [1]. However the dynamic incorporation of open corpus educational content in eLearning requires the generation of a relationship between educational concepts and the hypertext documents. One approach to create this overlay between concept and content is to use a Mindmap interface to allow learners to explore and associate hypertext content with knowledge maps of their own creation.

This paper presents the Open Corpus Content Service (OCCS), a framework that uses the hypertext structure of the WWW to provide methods of educational content discovery and harvesting. The OCCS semantically examines linked content on both the WWW and in digital content repositories, and creates concept specific caches of content. The paper also introduces U-CREATe, a novel user-driven Mindmap interface for supporting the exploration and assembly of content cached by the OCCS, in a pedagogically meaningful manner. The combination of these systems benefits both the educator and learner, empowering the learner through ownership of the educational experience and allowing the educator to focus on the pedagogical design of educational offerings rather than content authoring.

Categories and Subject Descriptors

H.3.3 [Information Systems]: Information Search and Retrieval;

H.5.2 [Information Interfaces and Presentation]: User Interfaces;

H.5.4 [Information Interfaces and Presentation]: Hypertext / Hypermedia;

H.1.2 [Information Systems]: User Systems (Human Factors);

General Terms

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT'08, June 19–21, 2008, Pittsburgh, Pennsylvania, USA.

Copyright 2008 ACM 978-1-59593-985-2/08/06...\$5.00.

Design, Experimentation, Human Factors, Theory

Keywords

Open Corpus Content, Information Retrieval, eLearning.

1. INTRODUCTION

eLearning environments are attempting to respond to the demand for personalised interactive learning experiences by providing increased support for functionality such as personalisation, adaptivity [2] and dynamic learning object generation [3]. As the current generation of eLearning systems attempt to support such features, one of the most significant problems being addressed is the traditional closed corpus nature of these systems. They rely solely upon bespoke, proprietary content [4]. In the second generation of adaptive eLearning systems there has been a significant shift towards the separation of personalisation and adaptivity information from the physical learning content [5][6]. This provides the opportunity to utilise content from external open corpus sources, such as the World Wide Web (WWW) and digital content repositories, in the generation of eLearning offerings. Content can be selected and inserted into a learning experience in a sequence that suits each individual learner.

Educational hypertext residing on the WWW is a resource that has yet to be thoroughly exploited in the field of eLearning. Such content, referred to as open corpus content, may be defined as any content that is freely available for non-commercial use by the general public or educational institutions. Content can be sourced from web pages, scholarly research papers, digital content repositories, commercial training repositories, forums, blogs, etc. Many countries have invested in the creation of national digital content repositories to encourage the reuse of learning resources. Merlot [7] in the United States of America, Jorum [8] in the UK and the NDLR [9] in Ireland are just three examples of such repositories.

Some educational hypertext systems, such as KBS Hyperbook [10], allow the incorporation of content sourced from the WWW. However, leveraging such content for use within these systems requires significant manual effort, on the part of a domain expert or course designer, in advance of incorporation. The content must be sourced, not a trivial task, and then tagged according to a specific metadata schema to enable integration [11]. This is an arduous task which teachers and course designers must undertake in the course of developing learning offerings, however, this leads to unnecessary cognitive load; the educator's efforts are better exerted on the pedagogical aspects of learning experience design. In learner-driven eLearning experiences it will be difficult to

empower learners to tackle such labour intensive tasks unless suitably motivated through incentives.

Content on the WWW is inherently unstructured or lacks consistency in its structuring. If the vast amount of knowledge and information available on the WWW is to be leveraged for use in interactive learner-centric eLearning systems methods of surmounting the heterogeneity of this content must be implemented. Apriori means of content discovery and classification are necessary to make it possible to better group and describe open corpus content. Dynamic user interfaces and eLearning systems which enable individual learners, teachers and course designers to explore, assemble and re-use content will be of the utmost importance.

It is of vital importance that such tools and eLearning systems address educational concerns. As learning moves from static tutor-centric delivery mechanisms to interactive learner-centric experiences, eLearning is seen as a method of supporting such dynamic interaction with the knowledge domain as a support to the traditional teacher-driven methods of education [12][13]. These systems can provide ways in which students can explore subject domains and construct their own knowledge maps and conceptual models of the topics they are exploring.

This paper introduces two systems which, when combined, enable the delivery of interactive, pedagogically meaningful learning experiences using hypertext content sourced from the WWW. The paper begins by presenting the Open Corpus Content Service (OCCS). The architecture of the system is discussed and system components are detailed. The design of U-CREAtE, a user interface for the learner-driven creation of knowledge maps and exploration of open corpus content is then discussed. The learning theories that governed the design of this user interface are explained. The paper closes with a discussion of the conclusions to date and the future work planned for this research.

2. OCCS - OPEN CORPUS CONTENT SERVICE

OCCS is a service which can leverage open corpus educational content for use in eLearning environments. The OCCS is responsible for the discovery, harvesting and indexing of such content. A focussed web crawler traverses hyperlinks, categorises content by topic and harvests the content deemed relevant to the subject of the crawl. Cached content is then indexed and made searchable via a web interface. The OCCS is provided as an autonomous service that can replace the current method of content sourcing in elearning environments, so as to minimize the impact on the current architecture of such systems.

Discovery: The OCCS implements a focused web crawler which conducts crawls across the WWW classifying the content it encounters and generating caches of candidate learning content in specific subject domains. This crawler can also identify candidate content in selected digital repositories, and add this content to the cache. The crawler is based upon the Internet Archive's Heritrix crawler [14]. Heritrix is an open source, extensible, web-scale, archival quality web crawler. The crawler is written in Java and can perform *broad crawling*, *focused crawling*, *continuous crawling* and *experimental crawling*. Heritrix executes crawls in a similar fashion to the majority of web crawlers; A hyperlink is chosen from among those scheduled, it is fetched and analysed,

the results are archived, any discovered hyperlinks of interest are added to the schedule and the process is then repeated incrementally.

Harvesting: In the OCCS web crawler, Heritrix is combined with a language guesser called JTCL [15] and a statistical text classifier called Rainbow [16] to create a focused, rather than general purpose, crawler. These tools govern what content is harvested and added to the content cache and what content is disregarded.

The language guesser JTCL is used to limit the content to be passed for classification to specific languages; for the purposes of this research we only examine English language content. JTCL is a Java implementation of TextCat, a written language identification program based upon the text categorisation algorithm presented in [17]. Currently the guesser can identify 69 natural languages.

Rainbow is used to classify the content discovered by Heritrix and perform the "focussing" aspect of the crawl. Rainbow is a statistical text classifier based upon BOW [18], a collection of C libraries for text mining and retrieval produced at Carnegie Mellon. Rainbow must be trained in advance for the subject of each focused crawl performed. Rainbow is provided with a combination of key word files and positive and negative training sets created in conjunction with the Google API. A statistical model of the subject domain is generated based on this training set. Rainbow compares all content to this model and rates it for relevancy to the subject area. A relevancy boundary is manually set in advance of each crawl and this dictates how on-topic the content must be to be included in the cache. If the content is deemed by Rainbow to be accurate enough to the subject area to achieve a rating above this boundary then the content is included in the subject specific cache and its links are added to the crawler hyperlink queue. If the content achieves a rating below the relevancy boundary it, and its subsequent links, are discarded.

Indexing: Once each focused crawl is complete and a cache of subject specific content has been generated, it is necessary to index the cache to make it searchable. All content archived by Heritrix is stored in an ARC file. NutchWAX [19] has been developed by the Internet Archive and open source partners as a method of indexing ARC files. NutchWAX stands for Nutch and Web Archive Extensions. Nutch [20] is open source web-search and indexing software built upon Lucene Java. The Web Archive Extensions include adaptation of the Nutch fetcher step to process ARC files rather than crawl the WWW and plug-ins to add extra fields to the index. NutchWAX uses the Hadoop [21] framework to run indexing jobs. Hadoop is an open source implementation of Google's mapreduce and GFS utilities. A list of the ARC files contained in the content cache is provided to NutchWAX. Each ARC file is imported, all the content objects or hyperlinks are taken in turn, parsed and indexed. Upon completion of this incremental process, an index is created for the entire collection of ARCs.

Content Visualisation: NutchWAX can be deployed under a servlet such as Tomcat to provide a text-based search interface for the content cache. This allows the user to input a search string and search across the entire content cache for candidate content objects. The results are ranked according to Nutch's relevancy rating. However there is a problem with the visualisation of the search results if this method of visualisation is used in the OCCS.

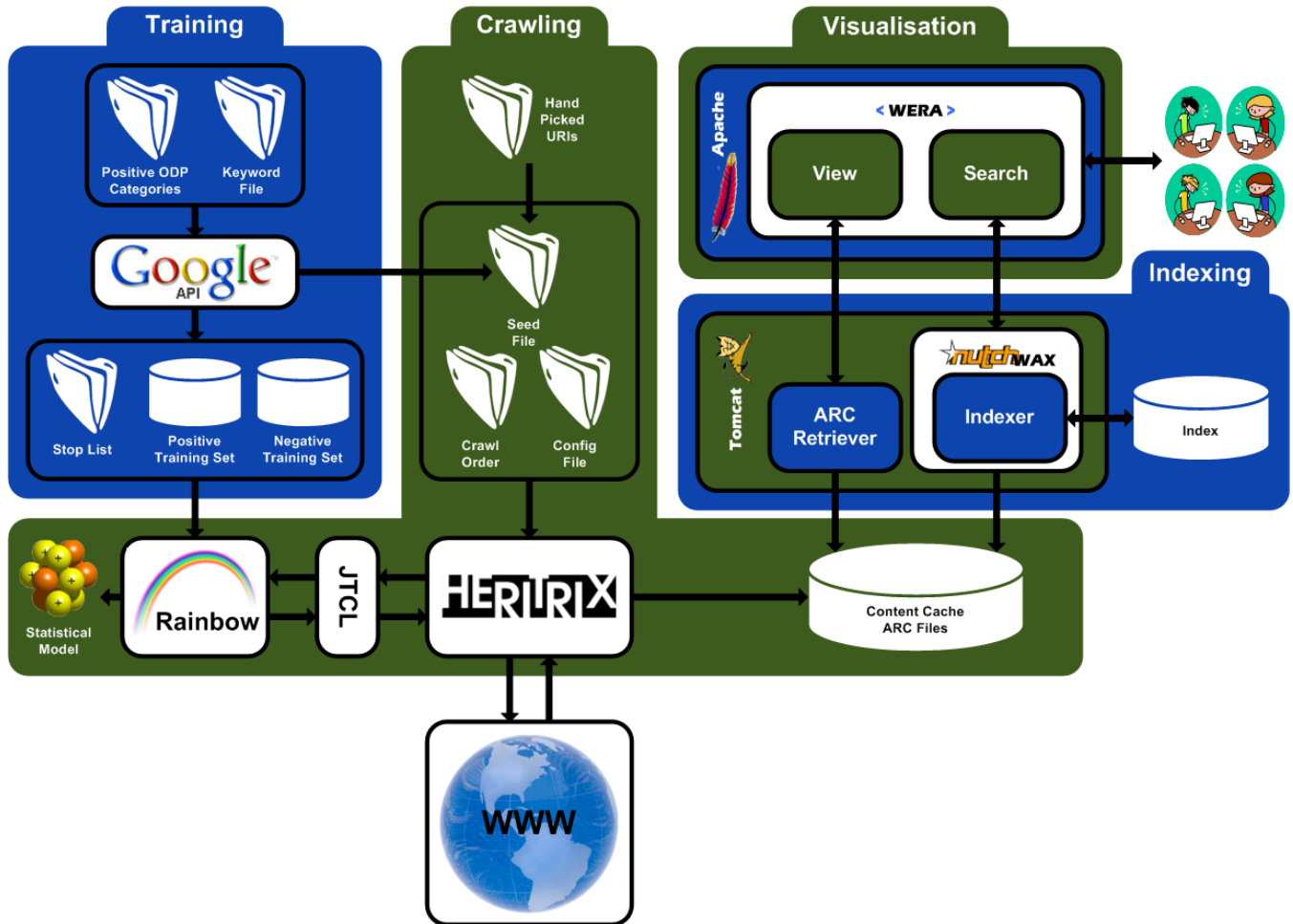


Figure 1. OCCS Architecture

When a search is performed on a regular search engine and a result clicked on, the user is redirected back to the original hyperlink where the content was discovered. In the case of this research, it is necessary to redirect the user to the actual archived piece of content. This is required as further manipulation of the content cache will take place after the content has been harvested. Research is ongoing into metadata enhancement, semantic annotation and vocabulary mappings which can be performed on the cached content to improve content classification and identification. The content residing at the original hyperlink could subsequently be altered which could render these detailed descriptions inaccurate, and could result in undesired content being delivered to the elearning environment.

Two methods of linking a content index with a web archive and visualising the archived material were researched. The Internet Archive's Wayback Machine [22] and the open source project

Wera [23]. The Wayback machine is not currently compatible with NutchWAX so Wera was selected for OCCS query resolution and visualisation of search results. WERA stands for "Web Archive Access". It is an archive viewer application that gives access to ARC file collections as well as the ability to perform free-text search. WERA is based upon, and replaces the NwaToolset. It is built using PHP and Java and utilizes open standards like the http protocol and XML extensively for communication between different parts of the system.

Research is continuing into methods of addressing various interoperability issues that the OCCS encounters when attempting to facilitate the use of content sourced on the WWW. Content discovered on the WWW may have insufficient, or in some cases no associated metadata descriptions. Semi /Automatic semantic annotation of content could be implemented to generate detailed metadata descriptions. Metadata descriptions can also be

implemented in varying and sometimes, incompatible metadata standards. Semantic mappings to a canonical metadata model using a fixed taxonomy of terms could be employed to overcome these issues.

2.1 OCCS Evaluation

To evaluate the OCCS it was decided to carry out a focused crawl for a particular subject area and examine the quality of the content returned from the indexed cache that was generated.

Crawl Preparation: There are a number of factors that affect the content returned from a focused web crawl. The first is a keywords file manually prepared and provided to Rainbow. For every keyword in the file a query is sent to the Google API. The top results for each of these queries are added to Rainbow's training set as positive domain examples. The first result for each query is also added to the seeds.txt file as a seed hyperlink for the crawl.

The second factor that can affect the content cache generation is the Rainbow relevancy boundary, described above. Only crawled content that Rainbow rates as being above this boundary is included in the content cache. Upon completion, the training step generated a seed file. The seed file contained 131 hyperlinks. These hyperlinks are the start points for the web crawler and can dictate the range of sites that the crawler visits.

Crawl Execution: The focused crawl was initiated and ran for a total of 46 hours 41 minutes. At that point 36,196 pages had been added to the content cache and this was deemed sufficient for the purposes of the evaluation. During the course of the crawl 473,259 hyperlinks had been discovered with 370,064 scheduled for download. Of these 67,144 had been downloaded and passed to JTCL. 61,527 hyperlinks were labelled as English and passed to Rainbow, which adjudged 36,196 results to be above the 90% relevancy boundary and were included in the content cache. Once the focused web crawl was complete the NutchWAX indexing was executed.

Evaluation: To evaluate the quality and validity of the results it was necessary to create a means of assessment whereby domain experts could semantically analyse the content and document its accuracy and applicability. It was decided that this evaluation should be conducted via the WERA interface, where selected users could perform specific search queries on the index for a pre-defined intended audience. These users were then required to rate each of the top ten content results for:

- Relevancy,
- Accuracy,
- Quality,
- The rank at which the results drop below an acceptable standard,
- The rank of the best overall result.

These ratings were performed using a Likert scale [24]. This is a type of psychometric response scale used in questionnaires, and is the most widely used scale in survey research. When responding to a Likert questionnaire item, respondents specify their level of agreement to a statement. Having discussed this option with Mary Sharp, a data analysis expert from the Department of Computer Science in Trinity College, it was decided to use a ten point Likert scale, as this provides no midpoint on the scale. Users will tend to

err on the side of caution and choose a neutral response if one is provided.

Three domain experts (DBMS lecturers) were selected to analyse and critique the results returned by the search engine. The analysis of the search results delivered for each query performed on the content index produced some interesting and valuable results. The overall relevancy of the content returned by the search engine was quite high. There were at least a couple of highly relevant, high quality results for each query performed. This was an essentially important outcome in the context of this research work. It was fundamental that the focused crawling performed was caching content of both high relevancy for the subject area and also of high quality.

The general pattern was for the relevancy of results to deteriorate a little lower down the results list; this is normal and in fact is even desired. There was general correlation between assessors on all the characteristics examined. This gives confidence in the results returned. There was a reduction in the number of commercial results appearing in the content listing when compared to a search performed in any of the most popular web search engines. This was an intentional consequence of the educational nature of the seed set of web sites provided to the focused crawler. These seed sites were entirely educational in nature, as educational sites would tend to link to other educational material rather than commercial sites this resulted in the crawler largely remaining within the educational domain. This seed set was chosen as the content required for the purposes of this research was educational in nature.

This evaluation proved that the OCCS was creating searchable subject specific caches of high quality, highly relevant educational content.

3. USER INTERFACE

Current eLearning interfaces are involved with the construction of online learning experiences, i.e. class and course management, while educational hypertext systems are concerned with the generation of links for content navigation. These systems are useful but their effectiveness in supporting learner-centric educational experiences is limited. The ability to engage learners and allow more active involvement in the construction of eLearning experiences is required. However, it is critically important that the teacher / tutor be empowered and not feel disenfranchised from the learning experience. In a blended teaching approach, adaptive, learner-centric eLearning does not replace the teacher; in fact it transforms and enhances their role within the learning experience. The teacher can use such interfaces and systems to increase the potential educational effectiveness of learning offerings. Through a mixture of pedagogy, subject matter expertise and adaptivity, better informed eLearning systems and user interfaces can be developed to actively engage and support the learner.

Extensive work has been carried out in the area of mapping pedagogical theory to educational practice to support better pedagogically-informed eLearning design [25][26][27]. These new interfaces, frameworks and systems lead to the modification of educational roles. Learners become responsible for their learning and are more actively engaged. The role of the teacher transforms from disseminator of information to facilitator of

knowledge acquisition, providing a support framework for the educational process [12].

U-CREATe is a learner-centric user interface being developed to combine the benefits of learning theories, such as constructionism and enquiry-based learning, with the techniques used in mindmapping and the ability of the OCCS to leverage large volumes of open corpus educational content. The learning theories involved in the design of U-CREATe are described in detail in the following section. The U-CREATe interface, its integration with the OCCS, and the knowledge representation technique known as mindmapping are then presented.

3.1 Learning Theories

For a number of years educational policy makers have been suggesting a move from static education to interactive personalised education [13][28]. More activity focused pedagogies are evolving, establishing the foundations for learner-centric educational environments. Learning theories, such as andragogy [29] and constructivism [30][31], advocate learning experiences, which address individual learners' motivations and promote learners' responsibility and engagement. The sharing of initiative, control and responsibility transports educational theories into the 21st century, facilitating contextualised, "anytime, anywhere" learning [32][33].

3.1.1 Constructivism and Constructionism

Constructivism is a cognitive theory developed by Jean Piaget. It regards learning as a dynamic process which consists of the formation of mental models or "constructs of understanding" by the learner. In this view of learning, the students actively construct new ideas or concepts based on their existing knowledge, their current educational context and by dynamically interacting with the learning media. It implies that learners do not passively absorb information but actively generate their own knowledge structures.

Constructionism is an educational method, pioneered by Seymour Papert, based on the constructivist learning theory. It shares constructivism's connotation of learning as "building knowledge structures". Constructionism builds upon the notions of constructivism by adding the idea that this formulation of knowledge structures occurs most prolifically in a context where the learner is "consciously engaged in constructing a public entity, whether it's a sand castle on the beach or a theory of the universe" [34]. This implies that individuals learn best when they are in the active roles of designer and constructor and are increasingly motivated if the item they are constructing will be seen, critiqued or used by their peers.

Papert uses the diffusion of cybernetic construction kits, such as Lego Mindstorms™, into the lives of children as an example of how constructionist methods could change the context of the learning of mathematics. The construction and programming of such models will not only improve children's understanding of mathematical concepts, the children will also have more motivation to further learn maths if it enables them to build better, more complex Mindstorm models.

One practical outcome of this theory of human learning is that the learning medium must create a situation where the learner has the freedom to exercise judgment about what is to be learned and at what pace. Papert states that the real benefit of computers in

education will be seen when educational systems begin to shift the balance between the instructional transfer of knowledge *to* students and the production/construction of knowledge *by* students.

In her PhD thesis[35], Idel Harel documented experiments which showed that children's attention could be held for an hour a day for periods of several months by making, as opposed to using, educational software, even if the children consider the subject of the software to be boring in its classroom form.

3.1.2 Enquiry-Based Learning

An old adage states: "Tell me and I forget, show me and I remember, involve me and I understand." The final element of this statement is the basis of the educational theory of enquiry-based learning. Enquiry-based learning, as the term suggests, is based upon a process of enquiry which actively involves participants in learning by encouraging discussion, questioning and investigation [36]. Involvement and engagement in the educational process leads to understanding. To be involved in this process the learner must possess or acquire the skills to seek resolutions to questions and problems while constructing new knowledge models of the subject area.

Enquiry-based learning as an educational method was first pioneered by an American educationalist and psychologist named John Dewey [37]. Dewey adopted the approach as a challenge to the traditional information-dissemination approach to education as this provided no opportunity for creative thought or for students to actively participate in a learning experience. Dewey's methods were later developed upon by psychologists such as Jerome S. Bruner who believed that education should stimulate and build expertise in the process of learning as opposed to the absorption of facts and memorising of concepts [38].

The enquiry-based learning approach is focused on using and consuming content as a means to develop information-processing and problem-solving skills. This approach is learner-centric, empowering the teacher as a facilitator of learning rather than the disseminator of information. Students are involved in the construction of knowledge through active involvement, and are thus dynamically engaged in the learning experience. It is essentially important that teachers remain in a close guiding role throughout such inquiry-based learning approaches. The idea is not to remove the teacher from the learning experience but to allow them to act as a facilitator and guide rather than merely a transmitter of curricula.

Daniel Edelson identifies the five most significant challenges to successful implementation of this approach [39], they are; motivation, the ability of learners to perform the required tasks, the ability of students to formulate questions based on their previous knowledge and experience, the ability to manage extended activities, the need to work within the practical constraints of the learning environment. This final challenge is where the OCCS can be most effective in supporting enquiry-based learning user interfaces. OCCS can provide high quality and highly relevant educational resources to the student.

3.2 U-CRETE – User-driven Content Retrieval, Exploration and Assembly Toolkit for eLearning

U-CRETE is a learner-centric user interface which employs mindmapping functionality to allow students and teachers to construct knowledge maps of certain concepts or subject areas. The interface is linked to the OCCS so that learners and educators can explore topic-focussed open corpus educational content as they build and develop their Mindmaps. U-CRETE combines the benefits of the learning theories described above, such as constructionism and enquiry-based learning with the ability of the OCCS to leverage large volumes of open corpus educational content using a mindmapping approach.

Mindmapping is a technique used to visualise, structure and classify knowledge. A diagram is generated, representing an individual's knowledge map of a subject or idea. The diagram forms a map of concepts (or nodes) and the relationships between these concepts. Concepts are usually connected in a downward-branching hierarchical structure. Mindmapping systems are considered "Knowledge Representation" tools and can be used to visually represent an individual's knowledge of a subject area.

Mindmapping in its raw form dates back many hundreds of years, however the modern form of mindmapping has been more recently formalised. Tony Buzan, an English psychologist, researched the use of mindmapping [40] and developed his own mindmapping software. Mindmaps have since been used across academia and industry to represent the knowledge of individuals and groups and to stimulate idea generation.

U-CRETE is developed upon FreeMind, an open source mindmapping system written in Java. FreeMind provides mindmapping functionality combined with an easy-to-operate hierarchical editor. The FreeMind interface combines a main toolbar across the top of the workspace, containing buttons for common operations, such as node addition, text formatting and map viewing options. A vertical toolbar along the left side of the workspace contains icons which can be added to nodes in the map.

FreeMind is quite feature rich. There are many options for navigation within a Mindmap, clicking on the map background allows you to drag the map around the workspace, it is possible to horizontally and vertically scroll a map using the mouse wheel. Clicking on a node in a map causes all child nodes to fold, so they are hidden from view. You can drag and drop single or multiple selected nodes in a map from one location to another. One of the reasons FreeMind was chosen as the basis for U-CRETE was that it offers a variety of options for linking other content to Mindmaps. It is possible to create links to web pages, local folders, executables (to launch programs), and any file on a local computer or network. In addition, you can drag and drop text or lists of files from outside applications into FreeMind.

U-CRETE contains additional innovative functionality. There are navigation points within the interface which link to the OCCS. This allows learners to search for high quality, topic specific educational content as they construct their knowledge maps. This can help the learner to formalise or further their knowledge of a concept. Multiple links to relevant content can now be added to each node which allows learners to collate content by concept.

This generates an overlay map between subject areas and educational hypertext. It is also possible in U-CRETE to browse the hyperlinks associated with a node in a separate window and select links to view in a browser.

U-CRETE and OCCS combine to create a learner-driven educational environment which supports educational theory through an innovative practical solution. U-CRETE is an educational tool that strongly complements the learning theory of constructionism as it allows students to actively construct visual models of subject areas and concept spaces and represent their knowledge models in a meaningful way. Learners are consciously engaged in constructing a Mindmap of a subject area which can subsequently be used during further exploration of the topic.

The interface is integrated with the OCCS, which allows a learner or educator to explore open corpus content which has been sourced from the WWW as they create a Mindmap for a particular subject area. Relevant content can be attached to each concept in the Mindmap. This learner-driven exploration and assembly of content constitutes an enquiry-based learning experience, as the student is encouraged to seek out content on a specific subject and select the most relevant and applicable pieces of content to be included in their Mindmap. Such experiences are based upon the active participation and investigation of learners within a subject domain. The ability to deliver educational experiences that support these two learning theories in a single system is a valuable and innovative educational development. Extensive usability testing and functionality evaluation of U-CRETE and OCCS are ongoing, the results of which will be published in detail in the near future.

4. CONCLUSION

A number of key issues that enhance the access to open corpus educational content on the WWW were presented in this paper. The OCCS was introduced and methods of content discovery, identification and harvesting were discussed in detail. Issues which can affect both the sourcing and categorisation of content were explored. Once content is located and harvested there are indexing and visualisation challenges that need to be met, the methods implemented to meet these challenges in the OCCS are presented.

U-CRETE, a user interface for knowledge representation and the exploration and assembly of open corpus content is then presented. The learning theories that guided the design and implementation of U-CRETE are explained. Finally mindmapping and the integration of U-CRETE and the OCCS are discussed.

Through the provision of a dynamic user interface, U-CRETE, it is possible to combine the benefits of learning theories such as constructionism and enquiry-based learning, the advantages of knowledge representation techniques such as mindmapping and the ability of the OCCS to leverage the vast amounts of educational hypertext available on the WWW.

U-CRETE provides the opportunity for learners and educators alike to generate visual representations and maps of the concepts and subject areas that they are exploring. It enables individual learners, teachers and course designers to explore and assemble highly relevant educational content sourced from the WWW and digital content repositories. This content is incorporated into their

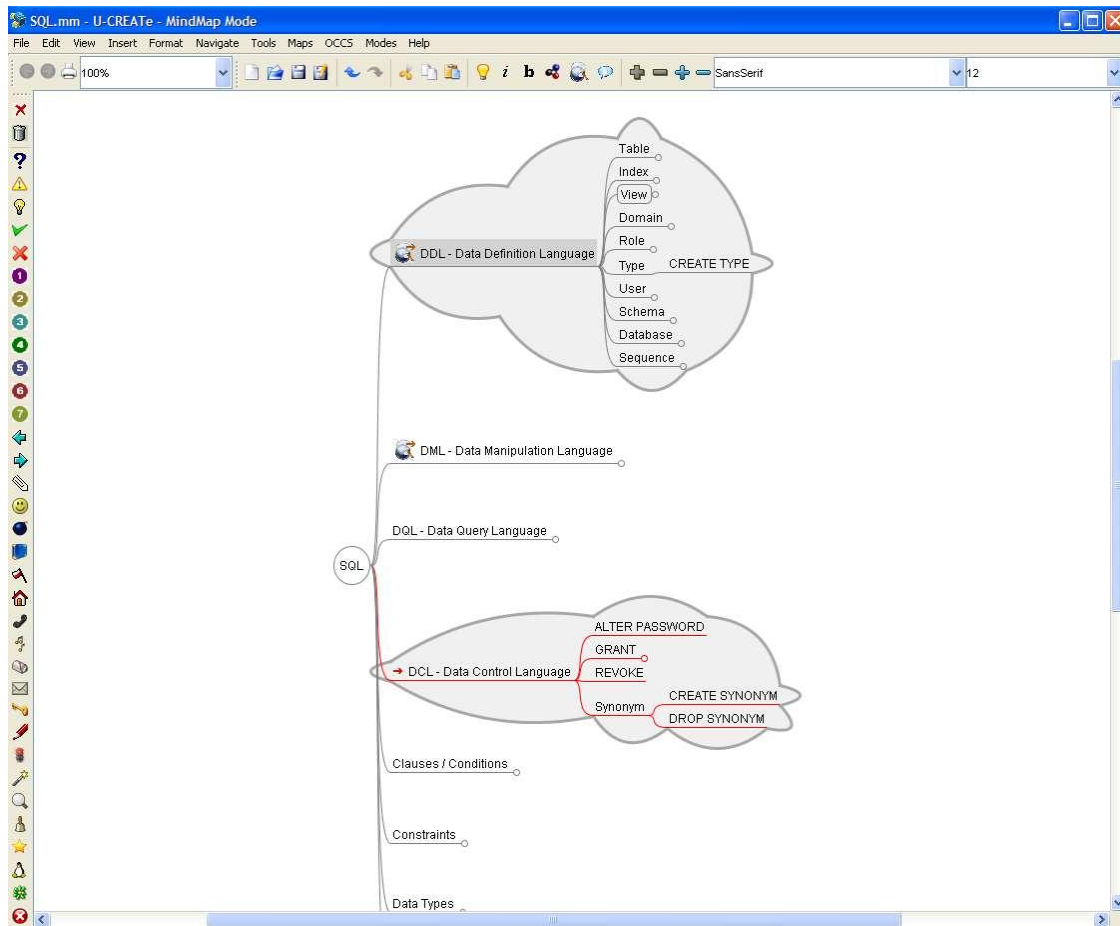


Figure 2. U-CREATE User Interface

knowledge maps, thus creating a binding layer between the educational concepts and the hypertext.

U-CREATE and the OCCS combine to provide a novel and powerful tool that delivers huge educational benefits for the learner through tailored pedagogical experiences, enables educators to concentrate on the pedagogical aspects of learning experience design through the reduction of content authoring and integration, and provides a first attempt to tap into the massive potential of open corpus educational content for eLearning.

5. ACKNOWLEDGMENTS

This research is funded by the Embark Initiative of the Irish Research Council for Science Engineering and Technology; funded by the National Development Plan.

6. REFERENCES

- [1] Henze, N. "Open Adaptive Hypermedia: An approach to adaptive information presentation on the Web". First International Conference on Universal Access in Human-Computer Interaction (UAHCI 2001), held jointly with HCI International 2001, 5-10 August, New Orleans, Louisiana, USA, 2001.
- [2] Brusilovsky, P. & Peylo, C. "Adaptive and intelligent Web-based educational systems". International Journal of Artificial Intelligence in Education, 13(2-4), 159-172, 2003.
- [3] Brady, A., Conlan, O., Wade, V. "Towards the Dynamic Personalized Selection and Creation of Learning Objects", E-Learn 05, World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education, Vancouver, B.C. (eds) Richards, G., AACE, 1903-1909, 2005.
- [4] Brusilovsky, P. and Henze, N. "Open corpus adaptive educational hypermedia". In: P. Brusilovsky, A. Kobsa & W. Neidl (eds.): The Adaptive Web: Methods and Strategies of Web Personalization. LNCS 4321, Berlin Heidelberg New York: Springer-Verlag, 671-696, 2007.
- [5] Brusilovsky, P. "Adaptive Educational Hypermedia: From Generation to Generation". In M. Grigoriadou et al. (eds): Proceedings of the 4th Hellenic Conference with International Participation in "Information and Communication Technologies in Education", New Technologies Publications, 19-33, 2004.

- [6] Conlan, O. *"The Multi-Model, Metadata Driven Approach to Personalised eLearning Services"*. PhD Thesis, Trinity College Dublin, 2005.
- [7] Merlot, Multimedia Educational Resource for Learning and Online Teaching. <http://www.merlot.org>
- [8] Jorum, UK Higher Education Institutions Digital Repository. <http://www.jorum.co.uk>
- [9] NDLR, National Digital Learning Repository. <http://www.learningcontent.edu.ie>
- [10] Henze, N. and Nejdil, W. *"Extendible adaptive hypermedia courseware: Integrating different courses and Web material"*. In: P. Brusilovsky, O. Stock and C. Strapparava (eds.) *Adaptive Hypermedia and Adaptive Web-Based Systems. Lecture Notes in Computer Science, Proceedings of Adaptive Hypermedia and Adaptive Web-based Systems, AH2000, Trento, Italy, August 28-30, 2000*. Berlin: Springer-Verlag, 109-120, 2000.
- [11] Henze, N., Nejdil, W. *"Adaptation in Open Corpus Hypermedia"*. *International Journal of Artificial Intelligence in Education: Special Issue on Adaptive and Intelligent Web-Based Systems*, 12, 325-350, 2001.
- [12] Carey, D. *"Teacher roles and technology integration: moving from teacher as director to teacher as facilitator"*, *Computers in the Schools*, v.9 n.2-3, p.105-118, 1993.
- [13] Laurillard, D. *"Rethinking University Teaching: A Framework for the Effective Use of Educational Technology"*, Routledge & Kegan, P., 1993.
- [14] Heritrix - The Internet Archive's open-source, extensible, web-scale, archival-quality web crawler project. <http://crawler.archive.org>
- [15] Java Text Categorization Library (JTCL). Java implementation of the TextCat language guesser. <http://textcat.sourceforge.net/>
- [16] Rainbow – Statistical Text Classification. <http://www.cs.cmu.edu/~mccallum/bow/rainbow/>
- [17] Cavnar, W. B.; Trenkle, J. M. "N-Gram-Based Text Categorization". In *Proceedings of Third Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, NV, UNLV Publications/Reprographics*, pp. 161-175, 11-13 April 1994.
- [18] Bow: A toolkit for statistical language modelling, text retrieval, classification and clustering. <http://www.cs.cmu.edu/~mccallum/bow/>
- [19] NutchWAX – Nutch and Web Archive Extensions for searching web archive collections. <http://archive-access.sourceforge.net/projects/nutch/>
- [20] Nutch open source search engine. <http://lucene.apache.org/nutch/>
- [21] Hadoop framework for running applications on large clusters of commodity hardware. <http://lucene.apache.org/hadoop/>
- [22] Wayback Machine – Web Archive Search Engine. <http://www.archive.org/web/web.php>
- [23] Wera – Web Archive Access. <http://archive-access.sourceforge.net/projects/wera/>
- [24] Likert, R. "A Technique for the Measurement of Attitudes" *Archives of Psychology* 140, 55. 1932
- [25] T. Mayes and S. de Freitas. "Review of e-learning theories, frameworks and models", *JISC e-Learning Models Desk Study*, 2004.
- [26] J. Scott. "Assessing the relevance of the review of e-learning theories, frameworks and models and the mapping table to designers", *JISC e-Learning Models Desk Study*, 2004.
- [27] J. Scott. "Learning Specifications and Standards", *JISC e-Learning Models Desk Study*, 2004.
- [28] Secretary of State for Education and Skills. "Department for Education and Skills: Five Year Strategy for Children and Learners", 2005. <http://www.dfes.gov.uk/publications/5yearstrategy/>
- [29] M.S. Knowles, "Andragogy in Action: Applying Modern Principles of Adult Learning" (The Jossey-Bass Higher Education Series). San Francisco : Jossey-Bass, 1984.
- [30] C.T. Fosnot, "Constructivism: Theory, Perspectives and Practice." New York : Teachers College Press, 1996.
- [31] T.M Duffy and D.J. Cunningham. "Constructivism: Implications for the Design and Delivery of Instruction" in *Handbook of Research for Educational Communication and Technology* [ed.] D.H. Jonassen. New York : Simon & Schuster; MacMillan, 1996. pp. 170-198.
- [32] Dagger, D., Wade, V., Conlan, O., "Towards "anytime, anywhere" learning: The role and realization of dynamic terminal personalization in adaptive eLearning". In *Ed-Media 2003, World Conference on Educational Multimedia, Hypermedia and Telecommunications, Hawaii, 2003*.
- [33] Horton, W., "Designing Web-Based Training: How to Teach Anyone, Anything, Anywhere, Anytime", New York: John Wiley & Sons. 2000.
- [34] Papert, S. & Harel, I., "Constructionism - Chapter 1: Situating Constructionism", Ablex Publishing Corporation. 1991
- [35] Harel, I., "Software design for Learning: Children's Constructions of Meanings for Fractions and Logo Programming". Doctoral Dissertation. MIT laboratory, Cambridge MA, USA. 1988.
- [36] Crabtree, H., "Improving Student Learning Using an Enquiry Based Approach". In *Learning based on the Process of Enquiry. Conference Proceedings, Curriculum Innovation, University of Manchester*, 2003. pp 77-84.
- [37] Dewey, J., "Democracy and Education: An Introduction to the Philosophy of Education" 1916.
- [38] Bruner, J. S., "Towards a Theory for Instruction" The Belknap Press of Harvard University Press, Cambridge Mass, 1966.
- [39] Edelson, D. C. et al., "Addressing the Challenges of Inquiry Based Learning through Technology and Curriculum

Design” In Journal of the Learning Sciences, 8, 1999. pp
391-450.

[40] Buzan, T. & Buzan, B., “The Mind Map Book” New York:
Penguin Group, 1993.