

# Dynamic Contextual eLearning - Dynamic Content Discovery, Capture and Learning Object Generation from Open Corpus Sources

Séamus Lawless, Vincent Wade, Owen Conlan

Knowledge and Data Engineering Group  
Trinity College, Dublin, Ireland  
{slawless, Vincent.Wade, Owen.Conlan}@cs.tcd.ie  
<http://kdeg.cs.tcd.ie>

**Abstract.** Adaptive Web Systems are seen as one of the main ways in which tailored learning experiences may be provided to individuals. However, such systems are restricted by the fact that they are reliant on content that has been specifically developed for each system. There are many sources of open corpus content available, the primary examples of which are the Worldwide Web and Open Corpus digital content repositories. However, the nature of this material is either inherently unstructured or lacks consistency in its structuring. These structural inconsistencies are both semantic, relating to the vocabulary/ontology used, and syntactic, relating to the metadata standards implemented. It is possible to facilitate personalized eLearning across a broad range of resources by leveraging such resources in an adaptive way. This paper seeks to explore the issues and challenges encountered in the sourcing, harvesting and provision of such open corpus content for Adaptive Hypermedia Systems.

## 1 Introduction

eLearning across all environments, be they commercial, academic or individual, can be greatly enhanced if the learning experience can be satisfied both immediately (i.e. at the time required) and contextually (i.e. tailored to the experience, goals and personal preferences of the individual learner). Such learning requirements may range from very specific learning needs e.g. how to carry out a particular activity, to more broad learning goals, e.g. develop knowledge of a particular discipline. Learning theories, such as Andragogy [Knowles, 84] and Constructivism, advocate learning experiences, which address individual learners' motivations and promote learners' responsibility and engagement.

Next generation learning environments are responding to these demands by attempting to support such aspects as personalisation, adaptivity and on-demand learning object generation [Brusilovsky, 01]. Adaptive Hypermedia is seen as one of the key areas of delivering personalised "just-for-you" eLearning. The benefit of such learning is that it can be dynamically suited to the learner's goals, competence, prior experience, learning style, learning preferences, learning strategy. This empowers the learner as the learning experience and the learning activities are more suited to the particular student.

One of the most significant problems with personalised eLearning is that they usually access content which has been specifically developed for the personalised learning system. This is particularly the case with Intelligent Tutoring Systems where, in fact, the personalisation is embedded in the content itself [Conlan et al., 02]. However in the next generation of adaptive systems the sequencing for the adaptivity is separated from the actual content. This leaves the opportunity whereby content can be selected and inserted into a sequence that suits the learner, however, in such systems (Knowledge Tree, Aha!, APeLS) the content is, in most cases, still sourced from a private repository of learning resources. This paper examines some of the challenges and issues which are encountered when attempting to provide such dynamic adaptivity using open corpus content rather than content tailored to the system involved.

The paper begins with an introduction into the current state of the art in Adaptive Hypermedia Systems. It briefly discusses how Adaptive Hypermedia Systems function and explains how they currently source and utilize learning content. The paper then proceeds by identifying the relevant standards and technologies which currently support the distributed access to learning objects and learning content on the web. The key challenges in the dynamic searching, identification, harvesting, customisation and learning object generation from open corpus sources are then discussed and information supplied regarding methods of implementation. The paper concludes by summing up proposals for future developments in this field and how it is intended to proceed with such development.

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

## ***2 State of the Art – Adaptive Hypermedia***

Adaptive Hypermedia Systems attempt to alleviate some of the difficulties encountered in Hypermedia systems by adapting the system individually for each user [Conlan, 04]. The system uses information on each user to create a user-model and this model then dictates how the system is altered to benefit that particular user.

### **2.1 Attributes for Adaptation**

Many attributes of a user are taken into account in the user model. These properties then control what adaptation is performed within an Adaptive Hypermedia System [Brusilovsky et al., 98]. The attributes can include,

- Cultural Background – This can dictate some of the basic methods of information presentation such as native language or dialect and the scale of measures and weights (i.e. metric versus imperial) to be used.
- User Preferences – The systems user interface can be adapted to the user's preferences, generally through the use of option or preference menus [Helander et al., 97]. This approach gives the user an improved sense of familiarity and comfort when using the interface.
- Communication Style and Needs – A user may have specific requirements, for instance, disabled users may require specialised input and output devices to be able to communicate adequately with the system. An example of this type of adaptivity is the AVANTI system [Brusilovsky & De Bra, 98]. Communication style can also vary from user to user, for example, some users prefer strict guidelines to complete a process, whereas other users may prefer a more open spectrum of options to allow them to dictate how they progress through a process.
- Cognitive and Learning Style – Users differ in the manner in which they prefer learning material to be presented. An example of this is the VARK learning preference model (Visual, Aural, Read / Write, Kinesthetic) for the presentation of information. The CAMELEON system [Ottmann et al., 98] is an example of a tutoring system that adapts on learning style.
- Prior Knowledge and Expertise – There is a connection between learning objects and the knowledge that is required to comprehend them. The previous knowledge and expertise of a user is used to dictate what learning objects are made accessible to each user. Systems that provide adaptivity in this manner are ALEKS [Doignon & Falmange, 99] and RATH [Hockemeyer et al., 98].
- Aims and Goals – Users may differ in their opinions of the aims and goals of a particular learning exercise. A system could adapt learning content to ensure that the user's aims and goals are achieved.

### **2.2 Methods of Adaptation**

All the aspects of a user that may contribute to the adaptivity of a system and the aspects of context in that system are collectively known as, the axes of adaptivity. The methods that may be used to realise this adaptation towards the axes are known as the methods of adaptivity.

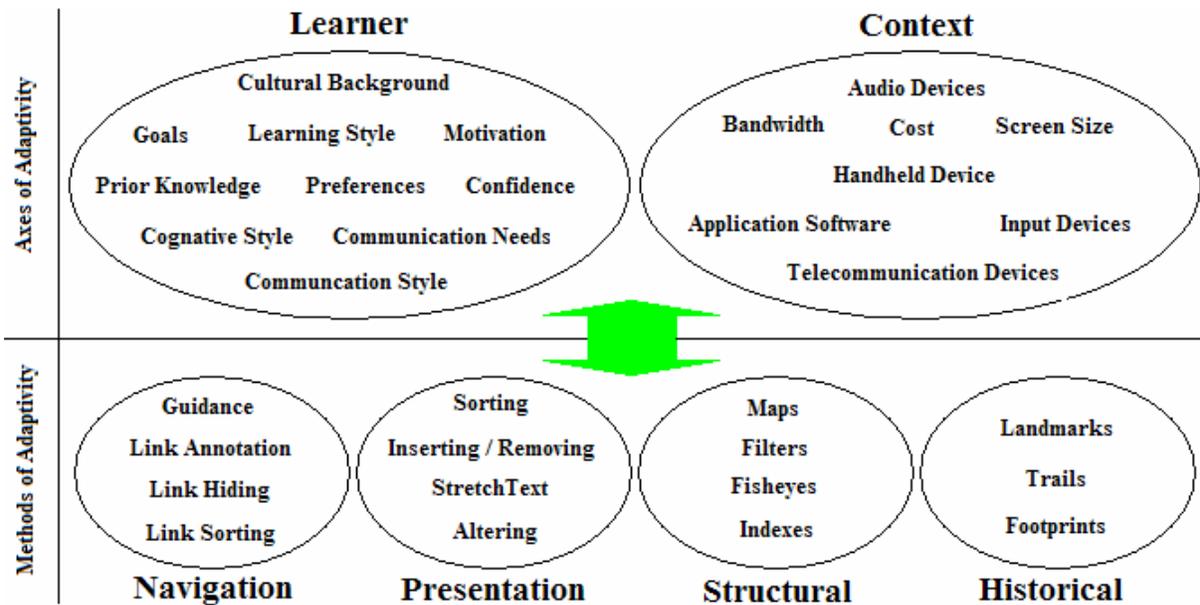


Figure 1, Axes and Methods of Adaptivity

The methods of adaptivity are grouped into the four categories described below. Each of these methods employs various means of realising adaptation towards the axes [Brusilovsky, 96] [Brusilovsky et al., 98] [De Bra et al., 99].

- Adaptive Navigation – This guides the user through the system by customising the link structure and format as a direct result of information retrieved from the user model. This dictates how a user navigates through the system and the level of freedom afforded them to navigate non-linearly [Eklund et al., 98]. Adaptive navigation is applied using many methods, including link sorting, hiding, annotation, guidance and relevance.
- Adaptive Presentation – This is the customisation of content to match attributes specified by the user model. The granularity of this customisation can vary from single words, to pages or to the substitution of media formats. Content can be altered in many ways to meet with a user’s needs. It can be altered to include additional information that is required by the user on a particular subject or in a particular section. It can also be altered to include prerequisite information that a user requires before they can successfully comprehend later sections.
- Structural Adaptation – This form of adaptation is used to give the user an accurate mental representation of the hypertext space involved. It is intended to give the user a sense of where they are positioned within the environment and an image of the environment as a whole.
- Historical Adaptation – This attempts to give a time context to the user as they progress through the system. History Trails and Footprints are made by the system periodically to track a user’s navigation and progress through the system. Landmarks can be made by the user to signify important stages of progress. Progression cues may also be customised by the system for the user.

### 2.3 Content and Metadata

One of the major downfalls in Adaptive Hypermedia Systems and personalised eLearning is that the systems generally access content which has been developed and tailored for that specific system. In a lot of cases, references to particular content fragments or sets of content objects are hard coded in the adaptive mechanisms. In fact, in Intelligent Tutoring Systems the personalisation elements are generally embedded in the content itself [Conlan et al., 02]. This restricts the reusability of content and the ability of the system to use any external content.

Metadata standards and specifications provide the mechanisms to describe all the elements of an Adaptive Hypermedia system, including the content itself. They are designed to make digital content more independent of services and the hardware used to deliver it. They facilitate content discovery and retrieval as well as providing information about how that content may be used. Some of the standards currently in wide-scale use on the worldwide web include,

- ADL SCORM [SCORM] – The Sharable Content Object Reference Model is divided into three main categories. The Content Aggregation Model (CAM), Run-Time Environment (RTE) and Sequencing and Navigation (SN).
- IEEE LOM [LOM] – The Institute of Electrical and Electronics Engineers (IEEE) Learning Object Metadata (LOM) aims to provide an extensive metadata description for learning objects.
- IMS Learning Resource Metadata [IMS] – The IMS Learning Resource Metadata Model is an implementation of IEEE LOM version 3.5 with some modifications.
- DC / DCMI [DC] – The Dublin Core (DC) Metadata Initiative (DCMI) specifies a common list of semantic annotations for the description of internet resources. Many non-DC metadata data models, such as IMS, have included mappings of their elements to DC.

## **2.4 Analysis**

The reusability of the content in Adaptive Hypermedia Systems is severely restricted by the fact that not all systems use the same metadata standard or the same vocabulary when tagging content. This creates structural and semantic interoperability problems. Not all metadata standards are interoperable and thus some systems may be unable to process content that uses a metadata standard it is unfamiliar with. Without the appropriate metadata or knowledge of the vocabulary used, each system may have no knowledge of what information a piece of content actually contains.

Some current adaptive web systems and intelligent tutoring systems have adaptivity embedded in the content that it uses. It is almost impossible to predict all axes of adaptivity that a learning object could be required to adapt on for each individual learner. This means that if a piece of content has adaptivity rules embedded that do not suit all usage scenarios, it greatly reduces the possible re-use of this content.

By implementing a method whereby open corpus content may be used by adaptive web systems, a situation is created whereby adaptivity and content may remain discrete. This would improve the ability of a system to re-use its learning objects and improve the quantity, variety and quality of content available to the user.

## **3 Open Corpus Content**

Open Corpus Content is content that is freely available for use by any educational institution or system. Such content is available in public repositories, such as ARIADNE [ARIADNE], ARROW [ARROW] and EducaNext [EducaNext], commercial repositories, such as ContentDM [ContentDM] and Lydia [Lydia], and via the worldwide web.

Countries such as the UK, USA, Canada and Australia, are increasingly investing in national repositories to encourage the re-use of digital content. Examples of such initiatives include, ARIADNE Foundation (European Union) [ARIADNE], Education Network Australia Online (Australia) [EdNA], eduSource (Canada) [eduSource] and BELLE (Canada) [BELLE], Multimedia Educational Resources for Learning and Online Teaching and Gateway to Educational Materials (USA) [MERLOT] and the National Institute of Multimedia Education (JAPAN) [NIME].

Open corpus learning content can exist in numerous formats. It can be wrapped and tagged as a learning object or as an entire course; it can take the form of video, audio or graphical streaming; it can be in the form of documents, pdf files or PowerPoint presentations. Any text regardless of its granularity from single words up to entire pages also constitutes open corpus content.

If the vast amounts of open corpus content that is available can be leveraged for Adaptive Hypermedia Systems, it will be possible to facilitate Dynamic Contextual eLearning. The best-case scenario is to enable adaptive course generators to use “any content, anywhere”.

## **4 Challenges and Implementation**

To implement the usage of open corpus content in Adaptive Hypermedia Systems, methods for sourcing, harvesting and delivery of this content need to be identified. Content and locations where content is available will need to be sourced. A method of harvesting and delivery of the content will need to be implemented, once the content has been identified. It will also be necessary to perform a dynamic translation of metadata between the search information

model and the open-corpus metadata information model, as no assumptions can be made regarding the metadata standard applied to the content, or indeed, if any is applied at all.

#### **4.1 Search Functionality**

It will be possible to source content from various locations when searching. The content could be available via a “system-aware” repository, this implies a repository where the content is structured and commonly stored as Learning Objects (LO) and the metadata standard applied will be known to the system. These repositories could be local, national, corporate or belonging to a partner organisation. Searching such repositories is vastly simplified as the system is aware of the structure of the content and the metadata standard required.

The content may also be available via a foreign or “non-system-aware” repository. These could be international, public or commercial. In this situation the content will in the majority of cases, still be stored as LOs, however the search function cannot assume the metadata standard that is involved and it may need to investigate the structure of the metadata used.

Finally the content may be available via the worldwide web. In this case no assumptions can be made regarding the structure of the content. The content may simply be text, however it could also be an image or an audio stream. It may not be stored in any structured fashion and there may not be metadata available to describe the content.

In order to cope with these various situations, any open-corpus content search function will need to be a series of search strategies. These strategies will dictate the search arena (i.e. the locations to be searched) and in what order they should be searched. These strategies will consist of,

1. Searching through “system-aware” repositories where all the content is stored as Learning Objects or courses and all the content is described using a system-aware metadata standard. This will provide the fastest results as the system can assume that the metadata is structured correctly.
2. Running an indexing facility to source content in open corpus repositories and on the worldwide web, and store references to the sourced content in a system search cache.
3. Searching through this cache of external content references that has been created from foreign repositories and the worldwide web.

#### **4.2 Metadata and Metadata Conflicts**

The cache that the indexing function creates will need to be able to handle multiple ontologies as references will be added for content that uses various metadata standards. The metadata may be SCORM, Dublin-Core or one of a variety of other metadata standards, some may be proprietary, and some sourced content may not have metadata applied to it at all. There are numerous options for handling such metadata interoperability conflicts, for instance,

- A translation could take place at run-time to match the search query to the content metadata. However this would be very inefficient. The translation would have to take place for every LO sourced as the search progresses. Also, if an LO is referenced multiple times in searches, the translation may have to be performed on the object many times.
- All the sourced content could have a reference added to the cache that is compliant with one metadata standard. A metadata listing would need to be created for content that is retrieved without metadata.
- Mappings could be made to a canonical model using a fixed vocabulary ontology. These mappings would be required to ensure that semantic differences do not cause content searches to miss suitable content.

The ability to resolve differences and conflicts between metadata standards and ensure interoperability between metadata standards would be a major step in the development of the semantic web.

#### **4.3 Integrating an Open Corpus Content Search Service into Adaptive Web Systems**

To implement the sourcing, harvesting and delivery of open corpus content, it is proposed to provide a Learning Object search and construct service to the Adaptive Engine that has been developed by the Knowledge and Data Engineering Group (KDEG) in Trinity College Dublin [Conlan et al., 02]. The adaptive engine is an Adaptive Hypermedia System that provides the facilities for reconciling multiple models, primarily those representing Content, User and Narrative, to produce individualized learning content. This Adaptive Engine, also known as APeLS [Conlan and Wade, 04], does not embed adaptive logic in the content or the engine itself. Therefore, the

adaptive logic, embodied in the Narrative model, remains separate to the content, thus facilitating the use of content chosen at runtime.

The aim for each search performed is to deliver a structured learning object to the adaptive engine. The adaptive engine does not need to be aware of where this content has been sourced. This service could be seen as a ready to use learning object constructor for adaptive hypermedia.

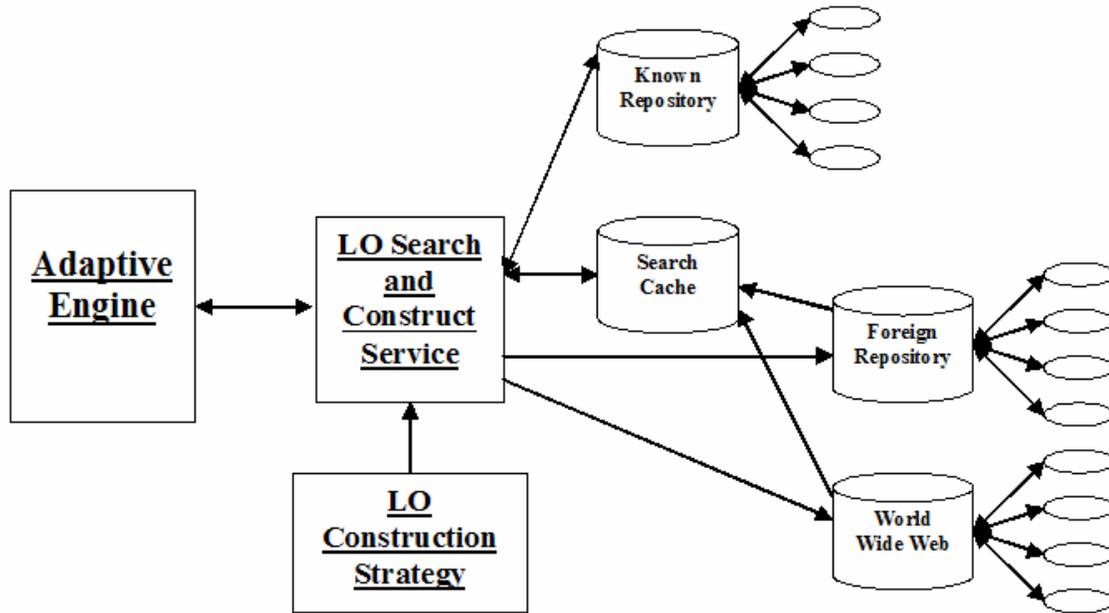


Figure 2, Search and Construct Service Architecture

#### 4.4 Proposed Architecture

Currently the adaptive engine functions by searching on a pre-defined repository or repositories. The search makes use of Candidate Content Groups (CCG). A CCG is a list of LOs manually grouped together that each teach a similar concept in a variety of ways. This creates a group of candidate objects for the course generator to choose from. Sourcing and defining the selected repository and manually creating the CCG by associating LOs create a large overhead in course generation. Narratives refer to the concept identifier of the CCG, thus maintaining abstraction between the adaptive logic and the content.

It is proposed to provide the new LO search and construct function as a stand-alone service that can replace the concept of a CCG in the search architecture of the adaptive engine. This will minimize the impact on the current architecture of the system. The author of the narrative may still provide information on content location if he/she wishes to use their own content.

The service can generate connections to repositories and the worldwide web on the fly. As a result, only one general candidate selector would be required. This candidate selector will consist of information on the learner that is available in the narrative. Information and rules on how to structure the LOs that are generated will also be provided to the service.

#### 4.5 Content Delivery

Once the required content has been retrieved it needs to be delivered to the adaptive engine in a format that can be used in adaptive course generation. If the content is sourced from a system-known repository then any metadata conflicts should be minor and the LO could be delivered with minimal modification.

If the content is sourced from a foreign repository or from the worldwide web then the structure of the content and the metadata applied to it could be uncertain. In this situation an analysis of the content will have to take place to ascertain its granularity and exact nature.

Information regarding the learners preferences and the structure of the learning object that is required will be provided to the LO construct service. The service will use this information to make decisions on what pieces of retrieved content to include in the LO and the sequencing of the individual pieces of content. Once constructed, the LO will be tagged using a fixed metadata standard and a single common vocabulary ontology to ensure ease of re-use in the system.

## 5 Conclusion

There are many challenges in implementing open corpus content usage in Adaptive Hypermedia Systems. Semantic and Syntactic interoperability is one of the main issues. The variety of metadata standards applied to content on the worldwide web and in repositories currently limits the interoperability of content and systems. The lack of a standard vocabulary in describing content also makes semantic matching of searches to relevant content a difficult task. These interoperability issues will not only effect the sourcing of content but also the re-use of any learning objects generated by the service.

Once content has been sourced, its granularity can dictate how useful it may be to an adaptive course generator. Small pieces of content such as paragraphs of text or single images are more useful than full pages of content where only sections may be applicable to the course being created. The sourcing of suitably fine grained content and the sequencing of this content in the generation of learning objects will be a further challenge.

An issue that will arise in the harvesting of content will be intellectual property rights and digital rights management. The ownership of digital content and permissions around its harvesting and use will be a major issue in all areas of society in the next number of years. Some public repositories, such as the Maricopa Learning Exchange [MLX] are trying to address this by applying licensing to all its stored content. However content residing openly on the worldwide web currently has no such standards applied.

The flexibility that would be created by addressing these issues and enabling Adaptive Hypermedia Systems to use open corpus content would provide a major step forward for adaptive eLearning and represent much greater learning possibilities for the user.

Huge quantities of information are available via the worldwide web on almost every subject imaginable. By allowing the user to access this volume of data, a situation is created where the user has a much greater choice of information and the style in which this information is presented. This in turn will encourage students to use eLearning technologies to support their studies and improve their knowledge. Improving the interoperability of metadata standards would also be a major step forward in all aspects of the semantic web.

## 6 Bibliography

- [ARIADNE] Ariadne foundation for the European Knowledge Pool. Available online at <http://www.ariadne-eu.org/>
- [ARROW] Australian Research Repositories Online to the World (Arrow). Available online at <http://arrow.edu.au/>
- [BELLE] Broadband Enabled Lifelong Learning Environment. Available online at <http://belle.netera.ca/>
- [Brusilovsky, 96] Brusilovsky, Peter. Methods and Techniques of Adaptive Hypermedia. In: User Modelling and User-Adapted Interaction 6: 87-129, Kluwer academic publishers, 1996. Available online at <http://www.contrib.andrew.cmu.edu/~plb/UMUAI.ps>
- [Brusilovsky et al., 98] Brusilovsky, Eklund, Schwarz. Web-Based Education for All: A tool for Development Adaptive Courseware. In Computer Networks and ISDN Systems (Proceedings of Seventh International Worldwide Conference, 14-18 April 1998) 30(1-7), 291-300. Available online at [http://www2.sis.pitt.edu/academics/publications/attachment/y1998\\_brusilovsky3.pdf](http://www2.sis.pitt.edu/academics/publications/attachment/y1998_brusilovsky3.pdf)
- [Brusilovsky & De Bra, 98] Brusilovsky, Peter & De Bra, Paul (1998). *Second Workshop on Adaptive Hypertext and Hypermedia*. Eindhoven University of Technology, NL. Online available at <http://www.wis.win.tue.nl/ah98/Proceedings.html>
- [Brusilovsky, 01] Brusilovsky, Peter. Adaptive Hypermedia. In User Modelling and User-Adapted Interaction, 87-110, Springer, March 2001. Available online at <http://www.springerlink.com/media/6CCWPGQWQTAG6XXTWC2L/Contributions/L/N/3/J/LN3J16077V24K774.pdf>

[Conlan, 04] Owen Conlan, The Multi-Model, Metadata Driven Approach to Personalized eLearning, Ph.D. Thesis submitted to the Department of Computer Science, Trinity College, Dublin (2004).

[Conlan et al., 02] Conlan, O.; Wade, V.; Bruen, C.; Gargan, M. Multi-Model, Metadata Driven Approach to Adaptive Hypermedia Services for Personalized eLearning. Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Malaga, Spain, May 2002  
[https://www.cs.tcd.ie/Owen.Conlan/publications/AH2002v0.99e11\\_Conlan.pdf](https://www.cs.tcd.ie/Owen.Conlan/publications/AH2002v0.99e11_Conlan.pdf)

[Conlan and Wade, 04] Conlan, O., Wade, V. Evaluation of APeLS - An Adaptive eLearning Service based on the Multi-model, Metadata-driven Approach. Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Eindhoven, The Netherlands, August 23-26 2004

[ContentDM] Digital collection management software by DiMeMa Inc. Available online at <http://contentdm.com/>

[DC] Dublin Core Metadata Initiative (DC or DCMI). Available online at <http://dublincore.org/>

[De Bra & Calvi, 98] De Bra, P., Calvi, L. (1998). *AHA: a Generic Adaptive Hypermedia System*. In Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia, pp. 5-12, Pittsburgh, 1998. (Proceedings available as CSN 98-12, TUE, or online at <http://www.wis.win.tue.nl/ah98/>)

[De Bra et al., 99] De Bra, Brusilovsky, Houben. Adaptive Hypermedia: From Systems to Framework. ACM Computing Surveys. Available online at <http://www.cs.brown.edu/memex/ACMCSHT/25/25.html>

[Doignon & Falmagne, 99] Doignon, J. & Falmagne, J. (1999). *Knowledge Spaces*. Berlin: Springer Verlag.

[EdNA] Education Network Australia Online. Available at <http://www.edna.edu.au>

[EducaNext] Portal for Learning Resources. Available online at <http://www.educanext.org/ubp>

[eduSource] Canadian network of learning object repositories. Available online at <http://www.edusource.ca>

[Eklund et al., 98] Eklund, J., Brusilovsky, P. (1998). *The Value of Adaptivity in Hypermedia Learning Environments: A Short Review of Empirical Evidence*. <http://www.wis.win.tue.nl/ah98/Eklund.html>, 1998

[Helander et al., 97] Helander, M., Landauer, T. & Prabhu, P. (1997). *Handbook of Human-Computer Interaction*. 2nd edition. Amsterdam, NL: Elsevier.

[Hockemeyer et al., 98] Hockemeyer, C., Held T., and Albert, D. (1998). *RATH A relational adaptive tutoring hypertext WWW-environment based on knowledge space theory*. Proceedings of CALISCE'98, 4th International conference on Computer Aided Learning and Instruction in Science and Engineering, Goteborg, Sweden, pp. 417-423.

[IMS] IMS Global Learning Consortium. Available online at <http://www.imsglobal.org/>

[Knowles, 84] Knowles, M.S. (1984). *Andragogy in Action*. San Francisco: Jossey-Bass.

[LOM] The Institute of Electrical and Electronics Engineers (IEEE) Learning Object Metadata (LOM). Available online at <http://standards.ieee.org/>

[Lydia] LydiaLearn, provides global access to eLearning resources. Available online at <http://www.lydialearn.com/>

[MERLOT] Multimedia Educational Resource for Learning and Online Teaching. Available online at <http://www.merlot.org>

[NIME] National Institute of Multimedia Education. Japan. Available online at <http://www.nime.ac.jp/index-e.html>

[Ottmann, et al., 98] Ottmann, T. & Tomek, I. (1998). *Proceedings of ED-MEDIA/ED-TELECOM 98 - World conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications*. Charlottesville, VA: Association for the Advancement of Computers in Education (AAACE).

[SCORM] Advanced Distributed Learning - Sharable Content Object Reference Model (SCORM)  
<http://www.adlnet.org>