Knowledge-based Content Navigation in *e*-Learning Applications

M. E. S. Mendes, E. Martinez, L. Sacks
Dept. of E&EE, University College London
mmendes@ee.ucl.ac.uk

Abstract: In this paper, a framework for organizing and navigating online learning material with respect to the semantic context of documents is described. We propose the use of a fuzzy clustering algorithm and TopicMaps to discover and represent knowledge, respectively. We present a prototype implementation of the framework for semantic browsing of a test collection of RFC documents.

1. Introduction

Since the Internet became easily accessible to millions of people around world, a vast range of Web services have emerged for the most varied purposes. In recent years, network-based teaching and learning – *e*-Learning – has been widespread, especially since standardizing initiatives for learning technologies [1] have begun. The work presented in this paper is in relation with the CANDLE project (http://www.candle.eu.org/), a European Commission funded project that is developing an *e*-Learning service mainly focused on the delivery of sharable and reusable online learning material.

Although *e*-Learning applications lack face-to-face interaction between teachers and students, they present the major advantage of enabling people to access learning facilities regardless of their location and at the time that is most convenient to them. Large networked repositories of learning material may be accessed by students, but tools are needed to narrow down the available resources to a particular individual based on the learning context, i.e., take into account learning objectives, pedagogical approaches, user profile, etc. Hence, tools are required to determine which documents are the most relevant for a given student, who wants to learn a particular subject.

Document relevance should be computed based on two components. Firstly, the set of concepts associated with each document should be identified. Thus, there needs to be a way to classify and organize materials in terms of knowledge domains. Secondly, the learning context should be considered, since the learning goals and pedagogical models may impact on the way materials are structured and consequently, on the definition of relevant links.

Our approach is to use fuzzy clustering to identify relationships between learning materials and to dynamically organize them into knowledge domains. The discovered relationships are then used to adaptively link documents. In the remaining sections of this paper we detail on this approach. In section 2, we present some background on knowledge representation and we propose use of fuzzy clustering for dynamic knowledge discovery and automated creation of ontologies. We also discuss possible ways of representing ontologies using XML (eXtensible Markup Language) [2] and specifically the TopicMap standard (XTM) [3]. In section 3, we look into the integration of this standard in adaptive hypermedia systems to build customized learning sessions and in section 4, we describe a prototype implementation of the knowledge representation framework. Finally, in section 5 we present some conclusions.

2. Dynamic Knowledge Discovery and Representation

Knowledge representation is becoming evermore important in applications that deal with large amounts data. The interpretation and sharing of data between systems in meaningful ways using automated tools requires knowledge about the data itself. It is therefore necessary to have semantic descriptions of the information so that knowledge can be generated and exchanged. Metadata (i.e. data about data) serves that purpose since it can provide rich descriptions about resources.

In *e*-Learning applications metadata is also fundamental for describing online learning materials and among other information, it can capture the subjects or knowledge domains associated with each document. This information is needed for customizing learning sessions, but it is not sufficient. For locating relevant content, the system needs to know how the subjects are interrelated in an abstract knowledge space and follow the associations to suggest links to the student.

There are several approaches to build the knowledge space one of which is to manually create an ontology of the problem domain. But this approach as some limitations namely its maintenance effort and subjectivity [4]. The approach we have followed is to have a process of dynamic knowledge discovery and representation using fuzzy clustering.

2.1 Fuzzy Clustering

Clustering algorithms aim at grouping data elements according to some similarity measure so that related elements are attributed to the same cluster. These algorithms can be used to find document relationships based on their metadata or full-text content. We have found that for our purposes, fuzzy clustering techniques would be more suitable than hard clustering methods [4], like agglomerative hierarchical clustering and the partitional K-Means that are the most common for document clustering. Instead of placing documents in a single cluster fuzzy clustering generates a fuzzy partition of the document space and hence, documents may have membership in several clusters simultaneously.

So far in our trials we have been using a new modified version of the Fuzzy C-Means algorithm [4], which is one of the most popular fuzzy clustering methods. The algorithm generates relationships among documents in the context of clusters, possibly uncovering unobvious associations through the analysis of the fuzzy memberships. The output of the algorithm can be seen as clusters representing knowledge domains, with relationships that form the abstract knowledge space.

The formal representation of the clustering results can use similar tools to those used to represent ontologies in networked systems. The RDF (Resource Description Framework) standard has been suggested for the Semantic Web [5], since it is very flexible to represent data relationships. Since this standard does not provide mechanisms to transport data between systems, XML has been selected for encoding and transporting RDF. Recently, a new standard – TopicMaps [3] – has also been proposed for knowledge representation. It has been specified under two different syntaxes: the ISO/IEC 13250 Document Type Definition (DTD), which is based on SGML/HyTime and the XTM 1.0 DTD, which has been designed to be optimised for the Web, and is expressed in XML [6]. We have decided to use TopicMaps, which are described next.

2.2 TopicMaps

TopicMaps consist of a new standardised tool developed to model and manage knowledge structures and information resources. They are based on the concept of the traditional book index (list of information), but with the expanded capabilities of a thesaurus (interrelation of terms) and glossary (list of terms and definitions), all comprised into a simple and scalable model [3]. Figure 1 contains this model, which is very simple but effective:

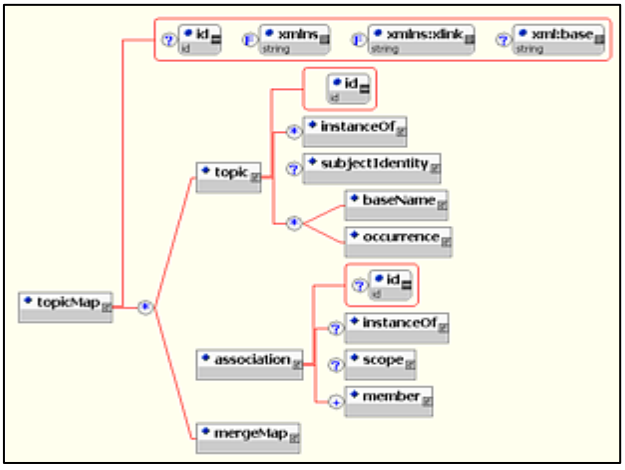

Figure 1. TopicMap Document Type Definition

- knowledge "elements" (any set of subjects, ideas, concepts, themes) are classified into objects called *topics*,
- any kind of relationship among them can be made explicit in the form of *associations*,
- most of the topics can be addressed by one or more specific resources (*occurrences*),
- topics, associations and occurrences can be *scoped* for different contexts.

The main advantages of this technology are:

1. Its standardisation provides flexibility and adaptability on its implementation.
2. The provision of an intuitive and meaningful (semantic) structure is translated into more efficient navigation and browsing processes among resources.
3. The ability to manage information resources and knowledge in the same tool.
4. And great scalability, as its architecture is designed to facilitate the merging of different topic maps without copying or modifying them.

One of the most obvious direct uses of TopicMaps is the navigation and browsing of information resources, although the combination of all its advantages makes them a very powerful tool for a wide range of information, knowledge and content management applications.

3. Adaptive Knowledge-based Content Navigation

In general, adaptive navigation systems enable personalised access to hyper-linked information. Adaptation in hypermedia systems can be provided at two different levels: at the presentation-level and at the link-level [7]. Although both kinds of adaptation should be present in *e*-Learning systems that aim at providing flexible learning environments, link-level adaptation is particularly important to guide students through their learning path.

Link-level adaptation deals with the discovery and display of relevant links to a given user. There are mainly two approaches for dynamically defining the links. One is to log the user's actions so that the system can suggest links based on past information. The other approach consists on keeping a record of the user's current knowledge and interests in a profile and then search for pages that match the individual needs.

There are several adaptive educational hypermedia systems that implement link-level adaptation [8,9,10,11]. The basic mechanism for adapting content to each student is based on the representation of both domain knowledge (domain model) and student knowledge level (student model) and uses the notion of pre-requisites and outcomes. Pre-requisites are basically the set of concepts a user needs to know to access a document and outcomes the set of concepts he/she is expected to acquire after reading it. The system then analyses the student's knowledge – stored in the profile as weighted concepts – to determine which links should be made available.

The knowledge domains are usually modelled manually by experts in the area. Instead of using such static models, fuzzy clusters and TopicMaps can be used for link-adaptation and for browsing the knowledge space. In [12], the authors also suggest the use of document clustering for adaptively linking resources. Their approach is to use hard clustering and adapt links in context of the user's interests and of the documents' contents. The advantage of having fuzzy clusters is that links can be ordered by degree of relevance, computed from the fuzzy memberships, and unobvious links may be revealed.

4. Prototype

A prototype of the knowledge representation framework has been implemented and the diagram shown in Figure 2 illustrates the complete process of knowledge discovery and representation:

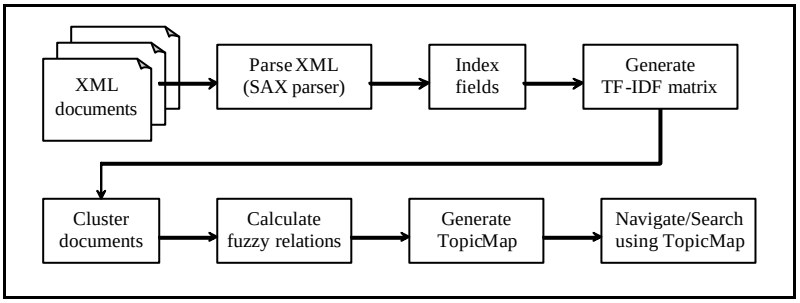

Figure 2. Knowledge Representation Framework

In our experiments [4,14] we have worked with a test collection of RFC documents (IETF standards) that was parsed to extract a set of *k* indexing terms for each document. Documents were then represented as a *k*-dimensional term vector using the well-known Vector Space Model (VSM) of information retrieval [13]. Term weights were generated as a function of their frequency within the document and of their inverse document frequency (TF-IDF), i.e. the number of documents where the term occurred. Then, the modified Fuzzy C-Means algorithm [4] was applied to the TF-IDF data. The algorithm uses the cosine measure to assess the similarity between documents and iteratively updates the clusters centres until an objective function is optimized. We ran the algorithm for a fixed number of clusters and subsequently a TopicMap was automatically generated. In Figure 3, a UML class diagram representing the clustering outcome is shown.

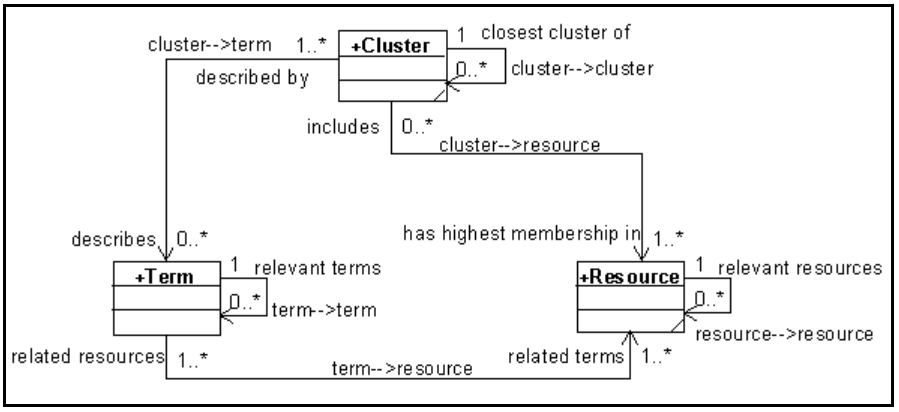

Figure 3. UML class diagram for the clustering TopicMap template

Cluster-cluster relationships were computed as the cosine similarity between each pair of cluster prototypes, cluster-term relationships were obtained from the weight of the terms in the prototype vector and cluster-resource relationships consisted on the fuzzy memberships of each document in the cluster. Resource-resource relationships were computed as fuzzy relations and finally, term-resource relationships were derived in context of the clusters' prototypes.

The *Omnigator* browser (www.ontopia.com) was used to visualize and navigate the TopicMap.

5. Conclusions

We have described a framework for organizing and navigating online learning material with respect to the semantic context of documents. We have proposed the use of a modified version of the Fuzzy C-Means clustering algorithm and of the TopicMap standard to discover and represent knowledge, respectively. Furthermore, we have explored ways of using this framework to adaptively link documents. The prototype implementation has shown that the algorithm was indeed able to find useful relationships among a test collection of RFC documents.

6. Acknowledgments

This work has been supported by *Fundação para a Ciência e a Tecnologia* through the PRAXIS XXI scholarship programme.

7. References

[1] IEEE Learning Technology Standards Committee (LTSC): http://ltsc.ieee.org/

[2] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler. "Extensible Markup Language (XML) 1.0", W3C Consortium, Oct. 2000. Available at: http://www.w3.org/TR/ 2000/REC-xml-20001006

[3] S. Pepper, G. Moore. "XML Topic Maps (XTM) 1.0", TopicMaps.Org Authoring Group, Aug. 2001. Available at: http://www.topicmaps.org/xtm/index.html

[4] M. E. S. Mendes, L. Sacks. "Dynamic Knowledge Representation for *e*-Learning Applications." Proc. of the 2001 BISC International Workshop on Fuzzy Logic and the Internet, FLINT 2001, U. C. Berkeley, Aug. 2001.

[5] T. Berners-Lee, J. Hendler, O. Lassila. "The Semantic Web". Scientific American, May 2001.

[6] S. R. Newcomb, M. Biezunski. "The Structure of Topic Maps Foundations". June 2001: http://www.topicmaps.net/

[7] P. De Bra, P. Brusilovsky, G. Houben. "Adaptive Hypermedia: From Systems to Framework". ACM Computing Surveys, Vol. 31, No. 4es, Dec. 1999.

[8] P. Brusilovsky, J. Eklund, E. Schwarz. "Web-based education for all: A tool for developing adaptive courseware". Computer Networks and ISDN Systems, Vol. 30, No. 1-7, pp. 291-300, Apr. 1998.

[9] L. Calvi, P. De Bra. "Improving the Usability of Hypertext Courseware through Adaptive Linking". Proceedings of the Eighth ACM conference on Hypertext and Hypermedia 1997.

[10] D. Pilar da Silva, R. Van Durm, E. Duval and H. Olivié. "A Simple Model for Adaptive Courseware Navigation". Proceedings of INFWET '97, Nov. 1997

[11] G. Weber, M. Specht. "User Modelling and Adaptive Navigation Support in WWW-Based Tutoring Systems". Proceedings of the 6th International Conference on User Modelling, UM'97, pp. 289-300, 1997.

[12] S. R. El-Beltagy, W. Hall, D. De Roure, L. Carr. "Linking in Context". Proceedings of the 12th ACM conference on Hypertext and Hypermedia 2001, HT 2001, pp. 151-160.

[13] R. Baeza-Yates, B. Ribeiro-Neto. "Modern Information Retrieval".ACM Press, New York, 1999.

[14] M. E. S. Mendes, L. Sacks. "Assessment of the Performance of Fuzzy Cluster Analysis in the Classification of RFC Documents", Proc. of the London Communications Symposium, Sep. 2000, London.