Appendix to: Personalized Access to Distributed Learning Repositories – PADLR –

S. Staab & R. Studer

May 25, 2001

Abstract

Topics in education are changing with an ever faster pace. Especially in the field of life-long learning the aspects that need to be taught by information providers must keep up to date with the process in its field. The courseware watchdog is a comprehensive module which allows users to focus on existing subfields of a discipline, but thereby be aware of important drifts and tendencies in the field.

1 Module: Courseware Watchdog

The courseware watchdog consists of four major components:

1. A focused crawler that gathers data from relevant educational media sources
2. A subjective clustering algorithm that allows to group educational media with similar contents together following different types of ontology similarity criteria
3. An intelligent browsing capability, which allows to understand similarities and differences of educational media sources
4. A mechanism for updating the ontology in order to reflect a drift of courseware topics

These four components are elaborated in the following submodules.

The significant difference to modules like the automatic extraction of metadata and the personal search engine is the attempt at using unsupervised learning methodology and intelligent browsing facilities in order to detect trends and tendencies rather than in classifying or retrieving according to existing structures. Thus, the courseware watchdog nicely complements these other work packages.

1.1 Focused Crawler

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer,Staab)

Working Title. Focused Crawler.

Problem Description. In order to detect related material and topics it is necessary to observe related sites and check them for congruency with one's own interests (cf. [3] on focused crawling for a Human Ressource Management problem). For this purpose, it is necessary to have a sophisticated crawler that takes advantage of one's focus such as expressed in the ontology and in one's documents.

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Crawler: Re-use existing crawlers to collect metadata and relevant documents from "seed" sites. Because existing crawlers typically "only" deal with gathering but rarely with selection of links according to content, the existing crawler has to be adapted in a twofold way. Adapt crawler to focus on ontology-congruent documents. Adapt crawler to focus on documents similar to the ones a user has given as preferable.

Dissemination, Testbeds and Evaluation. Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia. The prototype will be integrated with the Edutella infrasturcture at the end of the first year.

Collaboration and Scholarly Exchange. Interactions with other modules:

1. Edutella module (Exchange Facilities / Basic Infrastructure)
2. Automatic extraction of metadata and ontological information
3. Personal Search Engine

Use research visits (2 weeks up to 3 months) (Braunschweig, Hannover, Stanford, Stockholm, Uppsala) in order to integrate design and development with other modules. In particular, we expect fruitful interaction with the workpackage on metadata extraction, because based on additional metadata accuracy of crawling may be increased.

Budget Overview (including overhead costs):

AIFB: 70K first year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

1.2 Subjective Clustering

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer, Staab)

Working Title. Clustering with multiple, ontology-based views.

Problem Description. In order to detect trends and tendencies it is necessary to find outliers or groups of outliers. Because users typically do not want to exactly specify their complete profile and because users' profiles tend to change rather often (which is a major problem for recommender systems), we want to give the user views onto existing educational media. For this purpose, there have been developed clustering algorithms. Typically, however, clustering was used to give a single, "optimal" view on (learning) components. This is not suitable to account for the plurality of views that exist when looking at educational media. E.g. the same or similar slides on the internet might be used for introducing social effects of the internet or its technical foundations. We have recently developed preliminary clustering mechanisms that allow to provide *subjective views* onto documents [2], which are based on an underlying ontology. For instance, one view may concentrate on differences and similarities along the ontology parts that deal with social effects and another view may concentrate on ontology parts that focus on technology. The objective of this workpackage lies in the elaboration of these techniques and their realization within a framework that makes them applicable within Edutella.

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Adapt ontology for clustering: The ontology determines the view onto the resources. If there are too few concepts and relations in the ontology there may be the necessity to remodel parts of the ontology or — preferably — to provide methods to derive a slightly changed ontology from the given one such that it is suitable for clustering.

Implement Clustering: The clustering mechanism must be implemented and integrated in the overall framework. In particular, it must build on the crawled information and the users' own data.

Dissemination, Testbeds and Evaluation Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia.

Collaboration and Scholarly Exchange Interactions with other modules:

1. Edutella module (Exchange Facilities / Basic Infrastructure)
2. Focused Crawler

Budget Overview (including overhead costs):

AIFB: 35K first year, 35K second year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

1.3 Browsing of Watchdog Data.

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer, Staab)

Working Title. Browsing of Watchdog Data

Problem Description. Results from crawling and clustering need to be visualized. We plan to blend techniques from formal concept analysis and open conceptual hypermedia [1] in order to visualize the similarities and differences between documents and clustering results through lattices [6].

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Analysis of Underlying Data. This is necessary in order to understand underlying properties of data and metadata, such as "which data lies on a scale", etc. This analysis serves as input to the FCA-based browsing.

FCA-based Browsing. FCA-based browsing allows the user to explore similarities and differences between educational media.

Dissemination, Testbeds and Evaluation Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia. At the end of this workpackage it will be integrated into the overall environment.

Collaboration and Scholarly Exchange Interactions with other modules:

1. Edutella module (Exchange Facilities / Basic Infrastructure)
2. Clustering with multiple, ontology-based views

Use research visits (2 weeks up to 3 months) (Braunschweig, Hannover, Stanford, Stockholm, Uppsala) in order to integrate design and development within this module and with other modules. Workshops will be suitable to promote collaboration as well.

Budget Overview (including overhead costs):

AIFB: 35K second year, 35K third year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

1.4 Ontology Evolution

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer)

Working Title. Ontology evolution.

Problem Description. The watchdog depends on having a reasonable ontology of current topics in order to distinguish upcoming topics and in order to cluster and present educational material. AIFB will extend its experiences from ontology learning and detection of new conceptual relationships [5, 4] to ontology evolution. Of particular interest in the latter case, are the tendencies of drift in usage of terms, while "conventional" ontology learning rather considers text corpora to be static. Thus, we may extend the given ontology and adapt the overall results of the courseware watchdog.

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Adapt existing learning environment: AIFB already has a set of tools which may be used for this step. However, it is necessary to augment and adapt them in order to consider the different nature of ontology evolution (in constrast to ontology extraction), taking into account temporal and spatial heterogenety of educational media.

Dissemination, Testbeds and Evaluation. Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia.

Collaboration and Scholarly Exchange Interactions with other modules:

1. Automatic extraction of metadata and ontological information

Use research visits (2 weeks up to 3 months) (Stanford) in order to integrate design and development with other modules. AIFB has been organizing two workshops on ontology learning and is planning to continue this "tradition".

Budget Overview (including overhead costs):

AIFB: 70K second year, 35K third year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

References

[1] Carole Goble, Sean Bechhofer, Leslie Carr, David De Roure, and Wendy Hall. Conceptual open hypermedia = the semantic web? In *Proceedings 2nd Intl. Workshop on the Semantic Web (SemWeb'2001), Hong Kong, China, May 1, 2001*, pages 44–50, 2001.

[2] Andreas Hotho, Alexander Maedche, and Steffen Staab. Ontology-based text clustering. In *Proc. of the Workshop "Text Learning: Beyond Supervision" at IJCAI 2001. Seattle, WA, USA, August 6, 2001*, 2001.

[3] Andreas Hotho, Alexander Maedche, Steffen Staab, and Rudi Studer. A topic broker for human resource management. *Journal of Universal Computer Science*. to appear.

[4] Alexander Maedche and Steffen Staab. Discovering conceptual relations from text. In W. Horn, editor, *ECAI 2000 — Proceedings of the 14th European Conference on Artificial Intelligence, Berlin, August 21-25, 2000*, Amsterdam, 2000. IOS Press.

[5] Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2), March/April 2001.

[6] Gerd Stumme and Rudolf Wille, editors. *Begriffliche Wissensverarbeitung — Methoden und Anwendungen*. Springer, Heidelberg, 2000.