# Accessing Distributed Learning Repositories through a Courseware Watchdog

Christoph Schmitz, Steffen Staab, Rudi Studer, Gerd Stumme, Julien Tane

Learning Lab Lower Saxony (L3S), Expo Plaza 1, D--30539
Hannover, Germany; www.learninglab.de;
<lastname>@learninglab.de

Institute for Applied Informatics and Formal
Description Methods (AIFB),
University of Karlsruhe, D--76128
Karlsruhe, Germany;
www.aifb.uni-karlsruhe.de/WBS/

**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 emerging topics. 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.

## Introduction

In many domains, knowledge is becoming obsolete with an ever faster pace. Therefore, life long learning is gaining high importance for employees in order to keep up-to-date with recent developments in their respective fields (Fischer 2000). For supporting life long learning, E-Learning plays an important role in enabling on-demand learning of the individual employee (Maurer & Sappe 2001). However, only having online access to predefined course material that is stored in a single centralized repository does not meet the needs of such a learning scenario. Rather, the E-Learning approach should support the tailoring of the learning material to the specific needs of the employee being determined by the task at hand and his or her previous experience (Nejdl 2001). Furthermore, the individual learner should be able to keep track of recent developments and new trends in the relevant domains.

The same holds for lecturers, as the life cycle of courseware becomes shorter, and new courses on upcoming topics have to be offered more and more frequently. As there do not yet exist 'classic' text books on these topics, other ways to obtain new courseware material have to be gone. One relevant information source is of course the WWW; but as in old times one will also contact peers, i. e., lecturers and researchers working on the same domain, and ask them if they can help out with some material. As an individual learner, the lecturer benefits from technical support in accessing and tailoring these sources, and in discovering new trends.

The Courseware Watchdog that is described in this paper addresses these challenges by a comprehensive approach that exploits concepts from the Semantic Web, such as ontologies, in an E-Learning scenario (Stojanovic et al 2001). It is part of the PADLR framework (Personalized Access to Distributed Learning Repositories) that builds upon a peer-to-peer approach for supporting personalized access to learning .[1]

[1] http: //www.learninglab.de/pdf/L3S_padlr_17.pdf

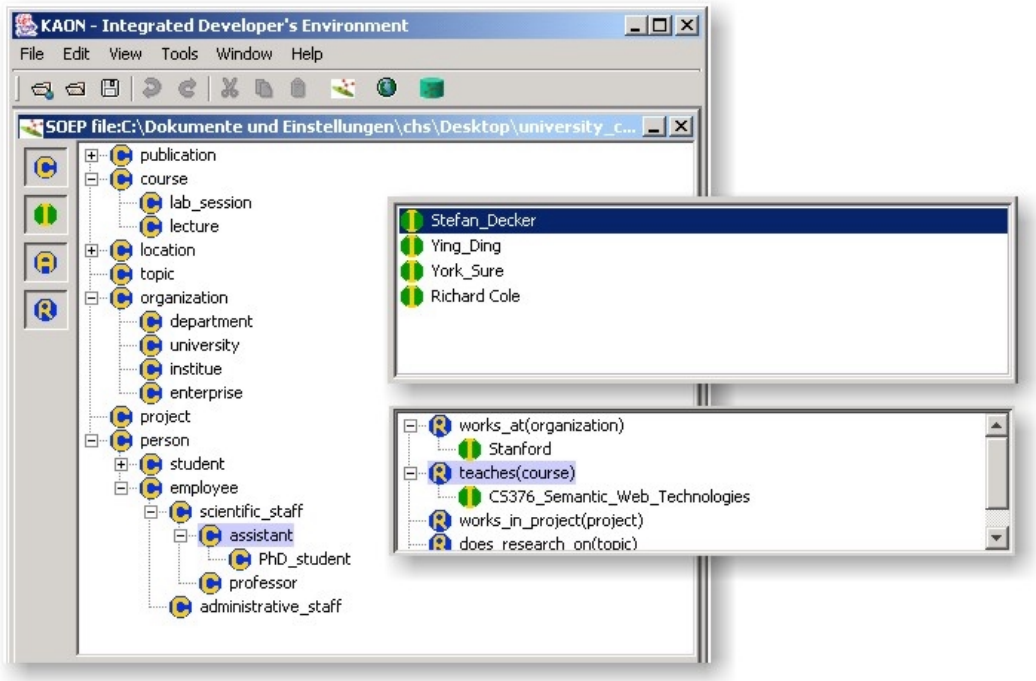


**Figure 1:** An ontology about research and teaching topics.

Ontologies are a means of specifying the concepts and their relationships in a particular domain of interest. Thus, they facilitate the sharing of knowledge between actors (Staab et al 2001). A small example of a learning ontology is shown in Figure 1. The left side shows the concept hierarchy. 'Assistent' is for instance a sub-concept of 'scientific_staff'. The upper right box shows the instances of this concept (e. g., 'Stefan_Decker'), and the lower right box shows the relationships the highlighted instance has to instances of other concepts (e. g., 'Stefan_Decker teaches semantic_web_technologies'). As we will see below, this formalization will help to also find the homepage of Stefan Decker when one is searching for course material related to the Semantic Web.

The Courseware Watchdog aims at finding and visualizing relevant educational material on the WWW and in the peer-to-peer network according to the user's needs, and detecting trends and changes within the field of interest. The Courseware Watchdog consists of the following components (see Figure 2):

- A *focused crawler* will find related web sites and documents that match the user's interests. The crawl can be focused by checking new documents against the user's preferences as specified in terms of an ontology.
- Integration into the *Edutella framework* enables the user to annotate educational resources with metadata to support exchange in the Edutella peer-to-peer network.
- As different users may have different points of view, *subjective clustering techniques* are used to generate subjective views onto the documents.
- *Visualization and intelligent browsing* capabilities for the results of the crawling and clustering steps will be provided.
- These tasks are supported by an underlying ontology. In order to reflect changes and trends within the field of interest, ontology learning methods will be employed to facilitate *ontology evolution* ( Stojanovic et al 2002).

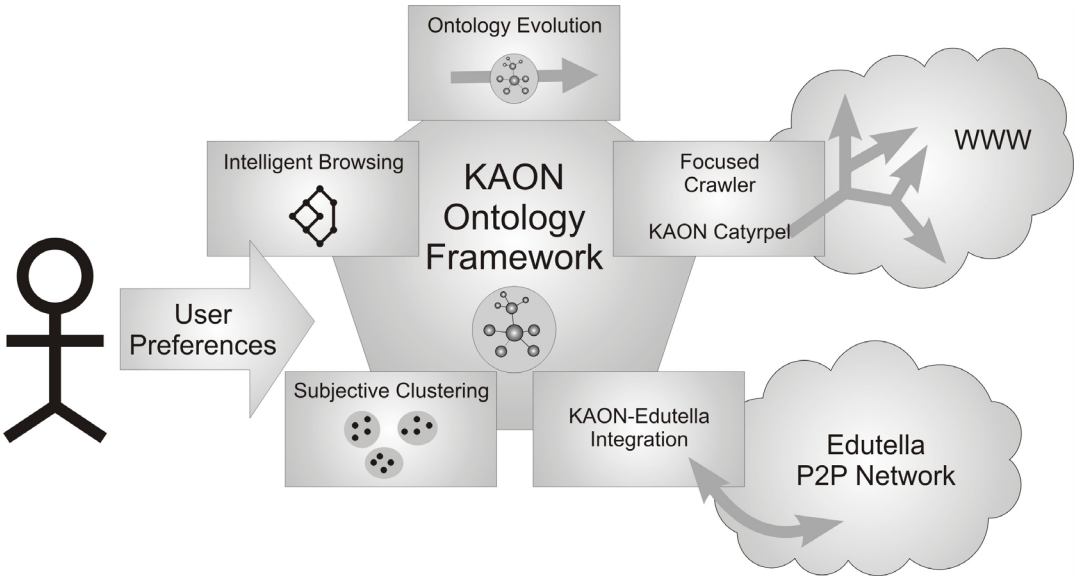


**Figure 2:** The components of the Courseware Watchdog.

These components are connected through the Karlsruhe ontology framework KAON (Handschuh et al 2001). KAON provides a plug-in infrastructure for ontology-based tools and applications.

The components are discussed in detail in the remainder of the paper, and are illustrated by a running example: consider a lecturer who has to prepare a new course about the semantic web. Although our lecturer has no course material in that field, she is confident that she will find relevant material either on the WWW or from colleagues working on that domain.

The rest of the paper is organized as follows. The next two sections discuss the information sources: the first presents the focused crawler and the second the Edutella peer-to-peer network. The following three sections discuss the treatment and interaction with the courseware collected and address the following components: subjective clustering, intelligent browsing, and the ontology evolution. We then conclude the article.

## Focused Crawler

A web crawler is a program that collects data from the web automatically by following links extracted from web documents. Thus, a portion of the web is traversed in a breadth-first manner, usually without regarding the relevance of the collected documents with respect to the user's needs. In order to restrict the traversal to material relevant to the user, the crawling process can be *focused*. Focusing in this context means preferring those links in the crawling process that appear to be pointing to relevant documents.

KAON Catyrpel (Maedche et al 2002a) is an ontology-based focused crawler embedded in the KAON environment for ontology-based tools (Handschuh et al 2001). With Catyrpel, the user can specify topics of interest in terms of an ontology. The user's preferences, i. e. entities in the ontology, and the ontology itself are then used to compute the relevance of documents and hyperlinks. After simple linguistic preprocessing (HTML tag removal, stemming etc.), lexical entries of the ontology are matched against the text and a relevance score is computed. Through the use of the ontology structure a useful measure of relevance can be evaluated, even if no exact match can be found. Several relevance measures can be used, which compared favourably against breadth-first search or simple keyword-based measures.

When searching for "Semantic Web", the focused crawler may for instance come across the web pages of Stefan Decker at Stanford University, who has already annotated his pages with RDF statements saying that he is giving a course 'Semantic Web Technologies'. The crawler will add the instances to the knowledge base used by the ontology as shown in Figure 1, and will return the homepage with a certain score.

## Integrating the Edutella Peer–to–Peer Network

The Courseware Watchdog includes the possibility to access the Edutella peer-to-peer network (Nejdl et al 2002a). Peer-to-peer (P2P) networks are decentralised networks, which allow for publishing and searching resources based on direct collaboration between its nodes.

The Edutella network applies the P2P paradigm to the exchange of structured information about available learning resources.[2] A common data model facilitates the integration of data sources such as relational database systems or XML and RDF repositories. Thus, all of these can act as Edutella peers, either as repositories or as consumer of the network resources.

For instance, our lecturer may query Edutella. She will be then able to locate relevant material on the P2P network which has been provided by other Edutella users. This way, she may for instance come across further lectures and tutorial about the topic. Moreover she will be able to control and share her own resources by annotating her resources with the annotating tool contained in the KAON framework (Nejdl et al 2002b).

## Subjective Clustering

In order to detect trends and tendencies in collections of courseware material, 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. Conventional clustering techniques provide a first answer for this purpose.

Typically, however, clustering is 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. We have recently developed clustering mechanisms that allow to provide *subjective views* onto documents (Hotho et al 2001), which are based on an underlying ontology.

For instance, one view may concentrate on differences and similarities of the content of learning material, while another view may concentrate on its presentation form, or on the levels of skills and experiences needed. The lecturer can then use the first view to select the material which addresses the topics which are most relevant to her planned course. She might also use the second view in order to see how the material is distributed over different types of material like presentation slides, exercise sheets, or online demonstrations.

## Browsing of Watchdog Data

The results from the crawling and clustering process need to be visualized in order to provide easy access for the user. The browsing component of the watchdog is designed to be based on techniques from Formal Concept Analysis and Open Conceptual Hypermedia in order to visualize the similarities and differences between documents and clustering results through lattices.

---

[2] http: //edutella.jxta.org

Formal Concept Analysis (Ganter & Wille 1999 & Stumme & Wille 2000) is a conceptual clustering technique which provides a hierarchy of concepts. When applied to learning material, the multiple inheritance within this hierarchy provides a rich conceptual landscape for navigating and retrieving the educational media. In particular, it allows to combine the different views created in the subjective clustering step; for instance to find first the most relevant clusters of documents, and second among them the slide presentations to set up the course.

We have recently implemented a Conceptual Email Management system (CEM) which supports exactly this navigation in collections of emails (Cole 2000). Our next step is to extend this approach to courseware material and other documents. This follows the vision that in the near future file, document, email, and web browsers will become tightly integrated in one unique user interface.

While Formal Concept Analysis focuses on the hypernym hierarchy on the concepts, Open Conceptual Hypermedia also considers arbitrary relations between the concepts (Goble 2001). It aims at enriching a web-based open hypermedia link service with semantic knowledge provided by ontologies. By combining both Formal Concept Analysis and Open Conceptual Hypermedia in the user interface of the courseware watchdog, we support thus navigation along both hierarchical and non-hierarchical conceptual relationships. For instance, the lecturer can continue his navigation from the cluster of relevant courseware material as obtained above by following the link 'created_by' to the creators of that material, and then follow on to the projects 'led_by' these researchers. There she may find additional presentation material to illustrate her lecture.

## Ontology Evolution

The Courseware Watchdog as presented in this paper so far builds heavily on a proper ontology that reflects what the user is interested in. However, over time such interests will invariably change together with the teaching/learning subject itself. Therefore, the ontology and the topics represented therein need to be updated and one must deal with several requirements incorporated in such updates:

*Introducing new concepts:* The first requirement is about (*i*) recognizing that a new concept (e.g. a new topic) has appeared in the course material available in the network or on the Web, (*ii*) inserting this concept into the right place of the taxonomy, and (*iii*) linking it via further relations to other concepts.

For instance, Web Services are today an emerging topic, and will probably have to be included in future courses on the Semantic Web. Hence 'Web Services' has to be recognized as a term that denotes a new concept. It must be inserted into the concept hierarchy (e.g. as a subarea of computer science) and it must be related to other disciplines (e.g. to business process modeling and E-Business). Only if this has been done the concept can be used by the courseware watchdog. Some example methods for this are proposed, e.g., in (Maedche et al 2002b).

*Recognizing concept drift:* The subject of "Semantic Web" is currently in flux and the understanding of it will evolve over time (such as that of other disciplines). In order to use the concept appropriately it will be necessary to adapt the concept to its current use ( Klenner 1994).

*Versioning ontologies:* Even when there are ontologies at different points in time, it should be possible to relate them, e.g. in order to explain why "Artificial Intelligence" in the 1970ies included computer algebra, but does not anymore in 2002. For this one needs comprehensive means for versioning ontologies ( Heflin & Hendler 2000).

*Process:* Eventually, updates, drift and versioning will not happen completely automatically. Rather, they will require an organizational process that safeguards critical changes (e.g., deletions of large parts of the ontology). Thus, one needs a process model and tools to handle the process. Some first models exists ( Maedche et al 2001, Staab et al 2001) but they need to be refined and adapted to E-Learning needs.

The combination of the different requirements will allow to exploit the benefits of the courseware watchdog, the central objective of which is to facilitate that the lecturer remains up-to-date about the courseware he is responsible for.

## Conclusion

The Courseware Watchdog is a comprehensive approach for supporting the learning needs of individuals in fast changing working environments, and for lecturers who frequently have to prepare new courses about upcoming topics. Being based on the Edutella peer-to-peer network the courseware watchdog provides personalized and flexible access to learning material that is offered somewhere on the Web. This functionality is achieved by tying together focused crawling, subjective clustering, flexible conceptual browsing, and ontology evolution.

The components of the envisioned courseware watchdog need further improvement. For instance, focused crawling has to be improved by offering further measures for computing the relevance of documents based on the ontologies and available metadata, and ontology evolution needs further techniques for better reflecting changes in the underlying learning material. Morever, we are also working on a better integration of the diverse functionalities to combine the diverse functionalities of the components.

Nevertheless, the courseware watchdog indicates how a Semantic Web based approach is better able to meet the needs of today's and future learning scenarios.

## Acknowledgements

We thank all members of our group for intensive discussions and contributions, especially Marc Ehrig and Andreas Hotho.

This research is partially funded by the German Federal Ministry of Education and Research (bmb+f).

## References

Richard Cole and Gerd Stumme (2000). CEM - a Conceptual Email Manager. In Bernhard Ganter and Guy W. Mineau, editors, *Proc. ICCS 2000*, volume 1867 of *LNAI*, Springer, 438–452.

Gerhard Fischer (2000). Lifelong learning – more than training. *Journal of Interactive Learning Research*, 11(3/4): 265–294.

Berhard Ganter and Rudolf Wille (1999). *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin – Heidelberg.

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

Siegfried Handschuh, Alexander Maedche, Ljiljana Stojanovic, and Raphael Volz (2001). KAON – the Karlsruhe ontology and semantic web infrastructure. `http://kaon.aifb.uni-karlsruhe.de/white-paper`.

Jeff Heflin and James A. Hendler (2000). Dynamic ontologies on the web. In *Proc. of AAAI/IAAI 2000*, 443–449.

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

Manfred Klenner and Udo Hahn (1994). Concept versioning: A methodology for tracking evolutionary concept drift in dynamic concept systems. In *Proc. of ECAI 1994*, pages 473–477. J. Wiley.

Alexander Maedche, Marc Ehrig, Siegfried Handschuh, Raphael Volz, and Ljiljana Stojanovic (2002a). Ontology-focused crawling of documents and relational metadata. In *Proceedings of the Eleventh International World Wide Web Conference WWW-2002*, May. (Poster).

Alexander Maedche, Viktor Pekar, and Steffen Staab (2002b). Ontology learning part one — on discoverying taxonomic relations from the web. In: Ning Zhong et al. (eds) *Web Intelligence*. Springer, to appear.

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

Hermann Maurer and Marianne Sapper (2001). E-learning has to be seen as part of general knowledge management. In *Proc. of ED-MEDIA 2001 World Conference on Educational Multimedia, Hypermedia & Telecommunications*, USA AACE.

Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjörn Naeve, Mikael Nilsson, Matthias Palmér and Tore Risch (2002a). EDUTELLA: A P2P Networking Infrastructure Based on RDF. In *Proc. of the 11th International World Wide Web Conference*, Honolulu, USA

Wolfgang Nejdl, Boris Wolf, Steffen Staab and Julien Tane (2002b). EDUTELLA: Searching and Annotating Resources within an RDF-based P2P Network. In Proc. of the Semantic Web Workshop, 11th Intl. WWW Conference, Honolulu, USA.

Wolfgang Nejdl (2001). Learning repositories - technologies and context. In *Proc. of ED-MEDIA 2001 World Conference on Educational Multimedia, Hypermedia & Telecommunications*, USA, AACE.

Steffen Staab, Hans-Peter Schnurr, Rudi Studer, and York Sure (2001). Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1), pages 26-34.

Liljana Stojanovic, Alexander Maedche, Boris Motik, Nenad Stojanovic (2002): User-driven Ontology Evolution Management. Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management EKAW, Madrid, Spain, 2002.

Liljana Stojanovic, Steffen Staab, and Rudi Studer (2001). E-Learning Based on the Semantic Web. In *Proc. of WebNet2001 - World Conference on the WWW and Internet*, Orlando, Florida, USA.

Gerd Stumme and Rudolf Wille, editors (2000). *Begriffliche Wissensverarbeitung — Methoden und Anwendungen*. Springer, Heidelberg,.