

Semantic Web Meta-data for e-Learning

– Some Architectural Guidelines

Mikael Nilsson <mini@nada.kth.se>
Matthias Palmér <matthias@nada.kth.se>
Ambjörn Naeve <amb@nada.kth.se>

Knowledge Management Research Group*
Centre for User Oriented IT Design
Department of Numerical Analysis and Computer Science
Royal Institute of Technology
Stockholm, Sweden.

Keywords: meta-data, e-learning, knowledge community, P2P

14th March 2002

*<http://kmr.nada.kth.se/>

Abstract

Meta-data is the fundamental building block of the Semantic Web. However, the meta-data concept is too loosely defined to provide architectural guidelines for its use. This paper analyzes important uses of meta-data in the e-learning domain, from a pedagogical and philosophical point of view, and abstracts from them a set of fundamental architectural requirements for Semantic Web meta-data. It also describes some flexible generic techniques for working with meta-data, following these requirements. Finally, the paper describes a Semantic Web-based e-learning architecture based in these requirements and techniques currently under development at the Knowledge Management Research Group at CID (Centre for user oriented IT Design) at KTH, the Royal Institute of Technology in Stockholm. This architecture builds on Edutella, a peer-to-peer meta-data exchange network, and a technique called *conceptual modeling* using the Conzilla concept browser, a new kind of knowledge management tool for conceptual navigation and exploration. The architecture provides an inquiry-based e-learning system that fits into the Semantic Web philosophy, and is based on a pedagogical framework called the *knowledge manifold*.

1 Introduction

The e-learning community are quickly embracing many modern Web technologies, including XML, XML Schema, P3P, and other Web technologies from the W3C and elsewhere. The educational technology standardization movement has also grown to become a significant force, including such organizations as IMS Global Learning Consortium [23], IEEE [21], Dublin Core [17], ISO [25], ADL [2], which are standardizing important base technologies for e-learning applications. Examples include meta-data, content structure, digital repositories, and many more.

A good example of the level of acceptance these e-learning standards are meeting is the recent MIT Open Knowledge Initiative [30], an effort to bring most of the courses offered by MIT online. The OKI is being developed in close cooperation with these standardization movements. Many, if not most, e-learning applications follow the same track, and are either compliant to these standards, or will soon be [19].

At the same time, it has become increasingly evident that the educational community will not be accepting Semantic Web technology for meta-data very quickly, although the potential benefits are many. For example, only recently has the popular IEEE LOM (learning object metadata) been expressed in RDF [29], and in spite of this, most implementors and researchers remain with XML Schema-based technology for meta-data.

Additionally, many e-learning applications are highly monolithic and seriously lacking in flexibility [38, 49]. The kind of intelligent computer support enabled by Semantic Web descriptions, such as software agents and self-describing systems, is not taken into account in the design.

In short, we have reached the somewhat surprising and perhaps paradoxical situation that the e-learning community is lacking in knowledge representation technology. For this reason, Semantic Web technology has not been extensively used and studied for educational applications, and there is therefore a need for a detailed analysis of the needs of the e-learning community concerning Semantic Web infrastructures.

This paper is an attempt to close the gap by documenting our experiences from building e-learning applications using Semantic Web technology. Our research group, the KMR (Knowledge Management Research) [44] group at CID (Centre for user oriented IT Design) [11] at KTH, the Royal Institute of Technology in Stockholm, is involved in several e-learning projects making use of Semantic Web technologies and paradigms. The projects reach from low-level RDF schema design and database interfacing, via distributed architectures to various forms of end-user tools for content management, navigation and querying.

Even though this paper focuses on the specific benefits Semantic Web technologies bring to e-learning, and the demands e-learning puts on Semantic Web technologies, it seems to us that many of the lessons we have learned are applicable to Semantic Web implementations in other fields as well.

Section 2 describes some of our more philosophical lessons regarding the interpretation of the meta-data concept. Section 3 describes some techniques for working with meta-data in a way that follows the guidelines we have developed. Section 4 gives an overview of the tools and infrastructures we are developing, and gives an e-learning scenario that exercises these tools and some of the principles discussed in this paper.

2 Semantic Web Semantics

The accepted definition of meta-data is "data about data" [5]. However, it still seems that most people use the word in different and incompatible meanings, causing many misunderstandings. In the course of implementing meta-data in e-learning applications, we have encountered objections of varying kinds to the concept of meta-data and its use. It seems to us that many of those objections stem from what we regard as misconceptions about the very nature of meta-data. Some of the features that are often attributed to meta-data, and that are involved in these misconceptions, include:

- meta-data is objective data about data.
- meta-data for a resource is produced only once
- meta-data must have a logically defined semantics.
- meta-data can be described by meta-data documents.
- meta-data is the digital version of library indexing systems.

- meta-data is machine-readable data about data.

The confusion about the meaning of meta-data is slowing the adoption of Semantic Web technology. What is missing in order to clear up the present confusion is a *meta-data semantics*. We need to sort out what we mean by meta-data, and better define how it is intended to be used. It turns out that the statements above provide a good background to sort out some of these issues, so we will tackle these statements one by one, from an e-learning point of view.

2.1 The Objectivity of Meta-Data

The first misconception about meta-data is the image of meta-data as being objective information about data. This misconception is tied to the fact that most meta-data aware systems only contain indisputable information such as title, author, identifier, etc. (you will note that most Dublin Core Elements are of this kind). When other kinds of meta-data enters the picture, such as the type of granularity of objects, pedagogical purpose, assessments and learning objectives, etc., many implementors raise skeptical voices. The reason for this skepticism is that such properties do not represent factual data about a resource, but rather represent *interpretations* of resources. When meta-data is viewed as authoritative information about a resource, adding descriptions of such features becomes not only counter-productive, since it excludes alternative interpretations, but also dishonest, forcing an subjective interpretation on the user. The ongoing debate is creating conflicts and is seriously hindering the adoption of meta-data technologies.

When meta-data descriptions are instead properly annotated with their source, creating meta-data is no longer a question of finding *the* authoritative description of a resource. Multiple, even conflicting descriptions can co-exist. This amounts to a realization that meta-data descriptions are just as subjective as is any verbal description. In fact, we *want* people to be able to express personal views on subjects of all kinds. It is a simple fact of life that consensus on these matters will never be reached, and the technology must support that kind of diversity in opinion, not hinder it.

In RDF, support for information about meta-data, or meta-meta-data, is built-in via the *reification* mechanism. In essence, reification makes a meta-data statement into an ordinary resource which can be annotated with regular RDF descriptions. Reification and meta-meta-data are thus of fundamental importance for a meta-data architecture¹.

Naturally, the problem of supporting this fundamental subjectivity in queries is not trivial. But the built-in support in RDF for meta-meta-data will make this task surmountable. Imagine, as a simple example, adding a link called "Who said this?" to each query answer. Another possibility is to add functionality to search using only trusted sources. This example emphasizes the need for trust networks and digital signatures of meta-data, in order to ensure the sources of both meta-data

¹See [27] for a discussion on reification.

and meta-meta-data. Supporting trust will be an absolutely fundamental part of the Semantic Web infrastructure if it is ever to gain acceptance².

One related philosophical point regarding authorities, that has played an important role in our choice of Semantic Web technologies, is related to the democratic ideals of the Internet. The Internet was originally designed as a peer-to-peer network where anyone can connect to anyone, and that is still one of the main reasons for its success. In the same way, the success of HTTP and the modern hypertext concept is fundamentally dependent on a peer-to-peer model, where anything may link to anything. This creates a democratic web, where there is no single point of control, no middle man in control of the network³.

However, the web has developed into a predominantly client-server based system, which mainly relies on centralized information handling, something that really defeats the purpose of Internet technology. Peer-to-peer networks is a way out of that trap. RDF is also deliberately designed as a peer-to-peer architecture, where anyone can say anything about anything [5], so it naturally fits into the democratic network philosophy. In a democratic network, objectivity is defined by consensus, not by authority. Meta-data needs to be a part of that consensus building process.

2.2 The Meta-data Eco-system

The second misconception regarding meta-data is related to dynamics. A popular view of meta-data is that it is something you produce once, often when you publish your document or resource, and which remains with the resource for its lifetime. This is the way meta-data is implemented in most systems supporting it.

This conception is related to the conception of meta-data as being authoritative, objective information consisting of facts that do not change. The problem with implementing meta-data support in this way is that it efficiently hinders subjective opinions and context-dependent meta-data.

One problem that immediately arises is how you can describe a resource if you don't know what its intended use is. For example, a single piece of media like a photograph can have different meaning when used in a History context than when used in a Photography context. These contexts may very well not be known when the resource is published, and new uses of resources may arise long after publication.

Instead, meta-data needs to be handled as a continuous work in progress, where updating and modifying descriptions is a natural part of the meta-data publishing process.

Treating meta-data as a continuous work in progress and allowing subjective meta-data leads to a new view of meta-data. Meta-data is information that evolves, constantly subject to updates and modifications. Competition between descriptions is encouraged, and thanks to RDF, different kinds and layers of context-specific

²W3C is promoting the "Web of Trust" via the use of digitally signed RDF.

³See, for example, [13]

meta-data can always be added by others when the need arises. Any piece of RDF meta-data forms part of a global network of information, where anyone has the capability of adding meta-data to any resource.

In this scenario, meta-data for one resource need not be contained in a single RDF document. Translations might be administrated separately, and different categories of meta-data might be separated. Additional information might be added by others. Consensus building becomes a natural part of meta-data management, and meta-data can form part of the ongoing scientific discourse. The result is a global *meta-data eco-system*, a place where meta-data can flourish and cross-fertilize, where it can evolve and be reused in new and unanticipated contexts, and where everyone is allowed to participate.

This provides support for the conceptual calibration process in a bottom up fashion [31], which builds consensus in the same way as it is done between people.

2.3 Layers of schemas

The third misconception relates to the use of RDF for expressing both simple meta-data (like Dublin Core) and for expressing RDF Schemas, ontologies (DAML, OIL) and query languages (Edutella [15, 18]). The semantics of information expressed in either of these formats is not derivable from the semantics of RDF itself, and thus needs to be specified independently. From a formalistic point of view, this means these are all different languages, and that data from several of them cannot be mixed.

However, each such new language will to large extent be similar to RDF with a slightly different semantics. Consequently you will end up with many different Semantic Webs that need to be kept apart from the original, and from each other, to avoid misinterpretations. Is this really what we want?

For one thing, we can be sure that there is no language to capture all the possible meanings we might want to encode on the Semantic Web. The complex world of human intention is too large for that. We must allow languages with different expressivity to co-exist. The semantics of pure RDF is very limited. A small vocabulary is predefined to allow for the semantics of instantiation, collections and reification. With RDF Schema additional terms for specific classes and predicates are introduced to allow specification of inheritance between classes and predicates.

Note that the semantics of RDFS is not derivable from its expression in RDF. For example, the transitivity of the predicate `subClassOf` has to be expressed elsewhere. Similarly, when new schemas are defined with the help of RDFS the semantics will only be partly there. E.g. there is no way to express that the predicate `title` in DC should be used for displaying a title rather than used in searches as a keyword.

This should not surprise us, if we consider how natural language works. In everyday life we do have a preferred language, with a basic set of terms that we agree on. When one day we find that we need to talk about new things, we express them by combining existing terms, and possibly inventing some new primitive terms in

our language. These terms are now equipped with new semantics, but they can still be mixed with terms we already know. Thus, language standardizes how to talk, not what to say. The same technique should work equally well on the Semantic Web: we should allow new vocabularies to be introduced at any time, and the terms to be mixed with data we already have.

Humans prefer to define their semantics whenever they need it, and this semantics only to some extent captures the true meaning of the defined terms. In this perspective RDF/RDFS is a nice compromise which gives you the opportunity to define your own vocabulary. If you want to, you can reuse vocabulary defined elsewhere. In other words, it is a small toolbox for allowing reuse of already defined semantics. Eventually, there may be very broad successful vocabularies in RDF that allows you to express nearly everything. But such vocabularies will most probably be patchworks of many small defacto vocabularies, that are developed in small steps by people who need them for specific tasks. This is very similar to how natural languages evolves, never reaching any final form, but rather changing continuously, reflecting the needs and thoughts of those using it. It's evolution, and it is a natural part of the meta-data eco-system.

We therefore argue that meta-data needs a flexible, extensible, layered architecture based on RDF [28, 6].

2.4 Meta-data Instances, or The Great RDF vs. XML Battle

The fourth misconception relates to XML, and has its roots in the popularity of XML as a document format. Describing meta-data in XML (based on XML schema, not on RDF), naturally leads to a document-oriented view of meta-data for a resource. This document is often referred to as the *meta-data instance* describing a resource^x. Most of the learning technology specifications for meta-data define meta-data using XML document instances. This includes, for example, IMS, IEEE, and ADL⁴. One notable exception is Dublin Core, which uses RDF as the primary encoding for both the basic Elements, the Qualifiers, and the Educational elements.

Thus, the learning technology specification community is building on XML technology, especially the advanced features of XML Schema, to enable extensibility and flexibility. The needs to combine many schemas and to precisely define vocabulary interrelationships is becoming greater, but the basic philosophy is still to use XML-based meta-data instances. We have noted several problems with this approach:

- RDF descriptions and XML meta-data documents are fundamentally different. An XML document is essentially a labeled tree containing text. An RDF description, by contrast, consists of a simple statement: a subject, a predicate and an object. Many such statements can be combined to form a set of connected statements (in the form of a graph), but each RDF statement can, in principle, be independently distributed. An XML meta-data document

⁴See, for example, the IEEE LOM standard [20] and IMS [24].

cannot be arbitrarily inserted into another XML meta-data document. For this very reason, XML is significantly less flexible for expressing meta-data, which by its very nature is subjective, distributed and expressed in diverse forms. RDF descriptions, while simpler, are flexible enough to support these principles.

- Defining schemas using RDF Schema and XML Schema are fundamentally different activities. XML Schemas describe the syntactic structure of XML documents. The interoperability work that is being done using XML Schema works with so-called application profiles, which are essentially XML Schemas that describe how to combine parts of different XML schemas. The result is a specification for XML meta-data documents that contain descriptors from several schemas.

By contrast, when defining meta-data schemas using RDF Schema, you do not define how instances will be expressed, but rather provide a vocabulary to use for describing certain features of the data. As described in the previous section, an RDF Schema describes the semantics of a vocabulary that can be reused in any setting. The important difference is that combining two vocabularies is not any more difficult than using two different vocabularies in natural language. You only need to follow the grammar for each statement (subject, predicate and object in RDF), and make sure the semantics is meaningful. Apart from that, any descriptions using any vocabularies can co-exist without explicitly declaring them.

The problem with defining meta-data application profiles using XML schema is that each application profile defines precisely which schemas you are allowed to use. Therefore, for each new meta-data vocabulary you need to support, you will need to define a new application profile. This automatically puts a stop to the use of alternative meta-data descriptors, and results in an authoritarian limit on meta-data expressions. When using RDF, meta-data using unknown vocabularies can be present without disturbing supported meta-data.

- The semantics of XML Schemas and RDF are fundamentally different. XML Schemas have a primarily syntactic interpretation, restricting the set of XML documents that can be produced. RDF, on the other hand, has a primarily semantic interpretation. While XML Schemas are used for modeling XML documents, RDF is used to model knowledge, where tree-based representations are not enough. This has important consequences for all applications needing semantic information.

The difference can be formulated in this way: XML/XML Schema is a data modeling language, and designing an XML meta-data instance is a purely syntactical activity [7]. RDF is a meta-data modeling language, and RDF schema design requires modeling some of the semantics of the terms used. Modeling meta-data as XML documents severely restricts its flexibility [36].

It has become clear that the move away from meta-data instances towards the globally connected knowledge eco-system is the central step in realizing the full potential of meta-data, as it enables precisely the subjective opinions and dynamic descriptions that is needed for a vital meta-data architecture.

2.5 New Uses of Meta-data

The fifth misconception about meta-data is that it is a digital replacement for library indexing systems. Meta-data obviously fulfills that role, but it is much more than that. Some of the important uses of RDF meta-data include:

description Since a resource can have uses outside the domain foreseen by the author, any given description (meta-data instance) is bound to be incomplete. Because of the distributed nature of RDF, a description can be expanded, or new descriptions, following new formats (schemas), can be added. This allows for new creative uses of content in unforeseen ways.

certification There is no reason why only big organizations should be able to certify content - individuals such as teachers may want to certify a certain content as a quality learning resource that is well suited for specific learning tasks. How to handle this kind of certification will be an important part of the Semantic Web, as discussed above.

annotation Everything that has an identifier can be annotated. There are already attempts in this direction: Annotea [4] is a project where annotations are created locally or on a server in RDF format. The annotations apply to HTML or XML documents and are automatically fetched and incorporated into web pages via a special feature in the experimental browser Amaya [3].

extension Structured content (typically in XML format) will become common. Successive editing can be done via special RDF-schemas allowing private, group consensus or author-specific versions of a common base document. The versioning history will be a tree with known and unknown branches which can be traversed with the help of the next generation versioning tools.

reuse RDF is application independent. As the meta-data is expressed in a standard format, which is independent of the underlying schemas, even simplistic applications can understand parts of large RDF descriptions. If your favorite tool does not support the corresponding schemas, it can at least present them in a rough graph, table or whatever standard form it has for describing resources and their properties. If more advanced processing software is available (such as logic engines), more advanced treatment of the RDF descriptions is possible.

and more Apart from these uses, you can invent new schemas describing structures, personalization, results from monitoring and tracking, processes and interactions that can enrich the learning environment in various ways.

In short, meta-data can be used to do many new and fascinating things. Limiting meta-data to only perform indexing would unnecessarily restrict its potential.

2.6 The Conceptual Web

The last in our list of misconceptions is that meta-data is only about machine-understandable data. The stated goal of the Semantic Web is to enable machine understanding of web resources. The rationale behind the development of the Semantic Web has been that deriving meaning from contemporary HTML or other web resources is nearly impossible due to the lack of a common meta-data framework for describing resources. In fact, most resource descriptions today are in the form of natural language text embedded in HTML. While such semantic descriptions are meaningful only to the human reader, the Semantic Web will provide such descriptions in machine readable format.

However, it is not at all evident that such machine readable semantic information will be clear and effective for human interpretation. The hyper-linked structure of the current web presents the user with a totally fluid and dynamic relationship between context and content, which makes it hard to get an overview of the conceptual context within which the information is presented. As soon as you click on a hyperlink, you are transferred, helplessly, to a new and often unfamiliar context. This results in the all too well-known "surfing-sickness" on the web, that could be summarized as "Within what context am I viewing this, and how did I get here?" [32, 33, 10] The conclusion we draw is that extracting usable meaning from web pages is often as difficult for a human reader as it is for a machine. This strongly suggests that there is a need for a human-understandable semantics for web resources as well.

This form of semantics becomes even more important within the emerging field of e-learning. In a learning context, the conceptual structure of the content is an essential part of the learning material. Losing the contextual information of the content means more than just "surfing-sickness". It means that you will not be able to contextually integrate the concepts that you are trying to learn, which is vitally important in order to achieve an understanding of any specific subject area.

The semantic web initiative, as it looks today, does not provide such a semantics. It provides descriptions of web resources, but no way to present them to the user in a contextually clear way. There are initiatives, such as topic navigation and visual history browsers, that try to address this problem, but they fail miserably in giving the necessary overview of the conceptual context.

In order to solve this problem, we are working on ideas to extend the Semantic Web in order to provide not only semantic information for the machine, but also conceptual information for the human user. This form of extended semantic web, which we call the Conceptual Web [35], is a long-term vision with many components, some of which are described in the next section.

Thus, it is important to realize that meta-data is not only for machine consumption. In the end, computers are a medium for human-to-human communication, and

conceptual meta-data that is understandable for both the human and the machine can definitely form an important part of that communication.

2.7 Conclusions

To summarize the discussions in this chapter, the Semantic Web needs a meta-data architecture that is

- subjective and non-authoritarian, supporting different views of the same resource.
- evolving, supporting a dynamic meta-data eco-system.
- extensible, allowing introduction of new vocabulary with new semantics.
- distributed, supporting descriptions by anyone about anything, anywhere.
- flexible, supporting unforeseen uses of resources.
- conceptual, supporting the evolution of human knowledge.

We have motivated some of these requirements by referring to problems encountered in the e-learning domain, but we believe our requirements to be generally applicable to the Semantic Web as a whole.

3 Flexible techniques for working with meta-data

We have seen above that there are several misconceptions about what constitutes meta-data. One of the consequences of these misconceptions has been unnecessary thresholds when working with meta-data, reflected by methodologies and tools. We now try to sketch a new work process and suitable tools for supporting it to unleash the true power of the Semantic Web, based on the principles discussed above.

There are basically three modes in the meta-data work process, *creation*, *publication* and *retrieval*⁵. The modes are not necessarily distinct or performed in a certain order. For example, at creation time you should look for similar meta-data to avoid duplication, inconsistencies, etc., while in retrieval you may want to publicize some metadata about yourself to allow some system to figure out what meta-data you need.

⁵Note that this is a low level view of the work process, higher level views may be better described as annotating, certifying, assessing, investigating, etc. The low level view is still valid though.

3.1 Creation

To create an XML document containing the one and only meta-data instance for a resource is naturally a difficult and cumbersome task that only domain specialists can feel comfortable doing. Subjective meta-data, on the other hand, can be added by anyone. Additionally, when metadata within authoritative documents is replaced with a patchwork of meta-data instances from different sources, it's easier to add meta-data in small chunks. Supportive tools need to be able to combine existing read-only meta-data sets with a creation time meta-data process. That is, adding translations, extensions, comments on the meta-data of others, etc., will put high demands on your meta-data editor. There will be a need for several different editors, e.g. graphically oriented for conceptual meta-data⁶ and classification meta-data (such as ontology editors)⁷ and text orienteds⁸ for meta-data in the form of property lists (e.g. Dublin Core meta-data).

3.2 Publication

Classical web publishing involves administrative decisions such as where to physically and logically put material, i.e. which server and what URL should be chosen. This works as long you deal with material of substantial size where authoring remains the largest part of the work. But for meta-data, especially if it is added in small chunks and not in large finished sets, the administrative task can easily grow exponentially.

Another problem is that in web publishing the physical location imposes restrictions on the logical location (URI). This is due to the fact that the locator (URL) is the same as the identifier (URI). The power of this delusion is proved by the fact that even SiRPAC, the W3C Java parser for RDF, fails to load models that do not have a URL as a system identifier. Clearly the solution is to have mechanisms for finding meta-data that separates the identifier from the locator. This separation allows people to change the location of their meta-data without having to alter anything. They may want to keep it to themselves, store it at some provider or maybe distribute it on a peer-to-peer network.

Storing meta-data on a central server is one form of publication, where you will need a special account or use some publicly available storage, which often imposes restrictions on accessibility and identity. Since by its very nature, the Semantic Web is decentralized, subjective and in a state of change, peer-to-peer environments are much more suitable than a central server approach [9]. There is at least three ways to publicize material on peer-to-peer networks:

- Your peer acts as a provider for your material and the network may retrieve and maybe choose to replicate your meta-data.

⁶E.g. A conceptual Browser such as Conzilla[37]

⁷E.g. Protege [46]

⁸E.g. the IMS compliant meta-data editor ImseVimse[43].

- The network is used as a more convenient way to access some specific storage, i.e., you know the location and can equally well connect directly to your storage location.
- You push material out on the network and hope someone will take care of your material.

The first alternative will allow providers, organizations as well as individuals to keep their original meta-data within reach, which has both psychological and practical aspects. The administrative overhead is kept to a minimum, as only logical identifiers need to be considered and maybe some initial configuration of the peer. The second alternative is really a central server approach with specific access methods. Defining a specific end provider protocol could be beneficial for easy uploading.

The last alternative presents a real challenge to avoid flooding of the network. Probably there is no practical limit on the total amount of storage space that could be publicly available, but selection processes will be requested for keeping relevance, quality and efficiency standards high. Meta-data that isn't accessed for a long time, outdated by newer versions or just broken will probably be in the danger zone of not being renewed. Hence the model for individual peers will probably be to provide low priority storage with little or no guarantee for persistence. Successful meta-data however will automatically achieve persistence by being stored in several places and be replicated on access. If authors of meta-data are dissatisfied with how their meta-data survives, they either have to re-inject their meta-data on a regular basis, act as a provider mentioned in the first alternative above or try to change the selection process somehow. This is essentially Freenet without cryptography [13]. It is important to note that such a network would provide an evolutionary environment which fits very well with the previously mentioned concept of a meta-data ecosystem.

Automatic construction of logical identifiers and storage on peer to peer networks via the pushing technique may allow virtually anonymous meta-data. This technique would support the ideas of a democratic network where anyone may say anything about anything.

Peer-to-peer systems are not famous for being reliable. Peers may be down temporarily and client peers will go up and down very often. Redundancy can be achieved by replicating meta-data on other peers. The difference to the pushing technique described above is that the party choosing to replicate does so on the basis of minimizing or easing network traffic, not for providing storage for homeless meta-data. Client peers may want to use replication as a cache for offline work.

The down side is that replication of data will result in inconsistencies, and renewal of the copy will be needed on a regular basis.

3.3 Retrieval

A strong incentive for publication is of course a large need for consumption of meta-data. Exposing meta-data to other humans is a much stronger incentive than the fact that machines will profit (and result in better human-computer communication). Adding material on the web, straightforward publication, does not necessarily lead to better exposure, as it risks drowning in the flood of all other material.

We very seldom know the location of meta-data we search for. Neither do we know a logical identifier for the meta-data / meta-data set. What we often do know is some pattern in the meta-data, e.g., it should be a description of a book written by a Swedish author or a lecture certified by a teacher you know. Searching for such meta-data is a pattern matching process that can be represented in a query-formalism. If you want to search for this pattern on the ordinary web, it amounts to an enterprise which takes several months. Luckily for us several big search engines have already performed this search and allow you to query against their index instead. However, since such an index is basically constructed for matching of keywords plus maybe some heuristics, searching for more complex relations or patterns is not really feasible.

Since the Semantic Web is a structured graph, it is much more straightforward to write an algorithm doing the matching against a fixed set of meta-data. However, since the Semantic Web will be spread out much like the original web, there is no obvious place to ask your question unless some company similar to Google replicates the whole Semantic Web (continuously) and provides an interface to a set of clustered computers doing the matching.

Another solution would be to use peer-to-peer technologies for routing queries to the meta-data end providers. This requires a routing procedure depending on the pattern of queries, i.e. meta-queries are used for registering what queries a peer can answer. This has to some extent already been tested with JXTA-search [48]. However, since JXTA-search only works on tree structured data encoded in XML, the processes concerning routing and matching have to be modified to work on RDF-graphs.

A yet unsolved problem is how the registrations in one peer should be further registered in other peers. In other words, should all the registrations be forwarded as they are, with the new peer as registrator or should a general 'least common denominator registration' be constructed instead? An even worse problem is how to know which questions that could be answered without generating large sets of questions and then testing them out. It is also problematic to know when to update your meta-query registration, e.g. whenever new meta-data is added, on regular intervals, etc. After a meta-query is generated it still needs to be spread out in the peer-to-peer network for updating purposes. If those problems are not solved with care, there is a serious risk of the network being flooded with updates and other administrative communication.

Hence a query service has to include a query formalism for exchange of queries, a routing procedure, an end provider resolving procedure and a collection proce-

dures for sending answers back to the initiator.

4 Some Tools supporting the Conceptual Web

At the KMR group, we have developed several tools that fill some of the roles that we feel are necessary for the realization of the meta-data architecture requirements for a Conceptual Web as described above. This chapter describes how these tools interrelate and how they can be used to implement important scenarios for e-learning. It describes only part of a complete architecture, but reflects the goals of some of the work we are doing.

4.1 Overall Architecture of our Conceptual Web Tools

RDF and RDF Schema provide the underlying model and representation. We also use standard RDF vocabularies such as Dublin Core and IMS/IEEE LOM for expressing some meta-data. The addition of ontology layers is of course also a fundamental part of resource description on the web, and is being considered for inclusion.

4.1.1 Edutella

The KMR group at CID is participating in an international collaboration project called PADLR[15], whose driving vision is a learning web infrastructure which will make it possible to exchange/annotate/organize and personalize/navigate/use/reuse modular learning resources, supporting a variety of courses, disciplines and universities. Within this project, we are collaborating with research groups at the universities of Uppsala, Stanford, Hannover and Karlsruhe in order to develop Edutella [18], an infrastructure and a search service for a peer-to-peer network that will facilitate the exchange of educational resources. Edutella, which will be a set of services implemented within the JXTA system [41, 48], is aiming (among other things) to solve the problems of meta-data retrieval described above. The envisioned services will include searching, mapping and replication. Searches will be routed to anyone who has registered a matching answering capability. Mapping will enable translation between schemas. This will allow very flexible reuse of information, since an application will not need to adapt to competing or more capable schemas because these schemas can be mapped to something that the application already understands. There will be no closed formats. Replication will allow meta-data about learning resources to be spread across the web, which will simplify the discovery of the corresponding resources.

4.1.2 Conceptual Modeling and Knowledge Manifolds

The fundamental building block of our idea of the Conceptual Web is *conceptual modeling*, which provides a human-understandable semantics for both abstract

ideas and concrete resources. We use a technique called Unified Language Modeling (ULM) for conceptual modeling, which is a modified version of UML (Unified Modeling Language) [40] that better supports modeling how we speak about things that has been developed at the KMR group by Ambjörn Naeve [32, 33, 34]. UML provides a well-proven and standardized vocabulary for conceptual modeling. Unfortunately, the relationship between RDF and UML is still rather unclear. We strongly support the forces that try to refactor UML in order to achieve a more precise meta-model [45], as well as the efforts to merge/combine RDF and UML [14, 12]. We regard these strategic efforts as necessary prerequisites for building the Conceptual Web.

Using the above technologies, we are designing the Conceptual Web as a *knowledge manifold*. A knowledge manifold is an educational architecture, developed at the KMR group, that provides an overall strategy for the construction, management and use of well-defined contexts for distributed content [31, 34].

4.1.3 Conceptual Browsing with Conzilla

One of the fundamental tools of the conceptual web is a new type of knowledge management tool which we call a *concept browser* [32, 33]. This tool allows the user to browse conceptual contexts in the form of context maps (typically ULM diagrams) with rich annotations. Thus the full power of visual modeling is combined with the distributivity and universal annotation property of RDF into a hyper-linked web of conceptually clear material. This combination gives the user a clear overview of the subject area (= context), while at the same time allowing the exploration of its various forms of content. Incorporating web resources as content is done by associating concepts with occurrences in resources. This has the important benefit of a clear and browsable visual overview of the context while viewing the content in, for example, an ordinary web-browser. Combined with our form of visually configurable query/search/filter engines [39] built using ULM and interfacing with Edutella, this results in a new and pedagogically revolutionary web experience.

Our first incarnation of a concept browser is called *Conzilla* [37], and has been developed as an open source project at the KMR group over the last couple of years. It is proving to be a very valuable tool for providing an overview of complex web-based material. Using Conzilla, several instances of knowledge manifolds are presently under construction at the KMR group, e.g. within the fields of mathematics, e-administration, IT-standardization and interoperability between different systems for e-commerce. Conzilla also has the potential to become a very useful and visually pleasing presentation tool for any kind of RDF data with a conceptual content.

4.1.4 Digital Portfolios

For content management, we are using a digital portfolio implementation designed and developed by the KMR group [44]. A digital portfolio is a personal online repository of information, which is used in e-learning scenarios by both teachers and students for publishing and storage. We have designed this portfolio implementation to use RDF descriptions of both meta-data and structure, using the IMS meta-data and content packaging [22] standards for that purpose. When equipped with an Edutella peer interface, a portfolio suddenly becomes a content management system allowing not only publishing of documents, but also dissemination of meta-data about documents and the structure of courses, as well as subjective annotations of online resources.

4.1.5 Application Independence: Semantic 3D and VWE

An added benefit of using the Semantic Web as a basis for the Conceptual Web is application-independence. Just as the Semantic Web gives the machine (software agents and applications alike) a sort of "sixth sense" about the meaning of web resources, the conceptual web gives the human user a sixth sense about the conceptual context and the underlying meaning of the current situation, which is independent of the currently used application. We are therefore studying ways to introduce the Conceptual Web into other environments [42, 26]. Apart from their usage on the ordinary web, we are investigating the fascinating possibility of introducing conceptual structures in 3D environments. A 3D environment filled with semantics and conceptual structures would present a fundamentally different experience, enabling for the first time a virtual reality full of meaning, and not only packed with dead 3D objects whose meaning is defined by the graphics engine. This semantics could even be accessed from outside such an environment, making the 3D environment fully semantically transparent.

Another application framework where we are introducing RDF for interoperability is the so-called Virtual Workspace Environment, VWE [47], which has been developed under the supervision of Fredrik Paulsson of the KMR group. VWE is a distributed Learning Management System, which is designed to support the construction of customizable learning environments by enabling the composition of learning resources. In fact, VWE is a small configurable operating system that can run in a web browser, which allows you to access your own learning environment from everywhere. Enabling RDF-based integration of VWE application enables semantic tool interoperation.

4.2 An e-learning scenario

Imagine you are studying Taylor expansions in mathematics. Your teacher has not provided the relevant links to the concept in Conzilla, so you first enter "Taylor expansions" in the search form in Conzilla. The result list shows that Taylor expansions occurs in several contexts of mathematics, and you decide to have a look

at Taylor expansions in an approximation context, which seems most appropriate for your current studies.

After having dwelled a while on the different kinds of approximations, you decide you want to see if there are any appropriate learning resources. Simply listing the associated resources turns out to return too many, so you quickly draw a query for "mathematical resources in Swedish related to Taylor expansions that are on the university level and part of a course in calculus at a Swedish university". Finding too many resources again, you add the requirement that a professor at your university must have given a good review of the resource. You find some interesting animations provided as part of a similar course at a different university, where it has been annotated in the personal portfolio of a professor at your university, and start out with a great Quicktime animation of Taylor expansions in three dimensions. The movie player notes that you have a red-green color blindness and adjusts the animation according to a specification of the color properties of the movie which was found together with the other descriptions of the movie.

After a while you are getting curious. What, more precisely, are the mechanisms underlying these curves and surfaces? You decide you need to more interactively manipulate the expansions. So you take your animation, and drag it to your graphing calculator program, which retrieves the relevant semantic information from Conzilla via the application framework, and goes into Edutella looking for mathematical descriptions of the animation. The university, it turns out, never provided the MathML formulas describing the animations, but the program finds formulas describing a related Taylor expansion at an MIT OKI site. So it retrieves the formulas, opens an interactive manipulation window, and lets you experiment.

Your questions concerning Taylor expansions multiply. You badly feel the need for some deeper answers. Asking Edutella for knowledge sources at your own university that have announced interest in helping out with advanced Calculus matters, you find a fellow student and a few math teachers. Deciding that you want some input from the student before talking to the teachers, you send her some questions and order your calendaring agent to make an appointment with one of the teachers in a few days.

A week later you feel confident enough for changing the learning objective status for Taylor expansions in your portfolio from 'active, questions pending' to 'resting, but not fully explored'. You also mark your exploration sequence, the conceptual overviews you produced in discussion with the student and some annotations, as public in the portfolio. You conclude by registering yourself as a resource on the level 'beginner' with a scope restricting the visibility to students at your university only.

In this scenario, we see some of the important points being exemplified:

- Distributed material and distributed searches.
- Combinations of meta-data schemas (for example, personal information and content descriptions) being searched in combination.

- Machine-understandable semantics of meta-data (calendar info, finding the right kind of resources).
- Human-understandable semantics of meta-data (contexts, persons, classifications)
- Interoperability between tools. Any tool can use the technology.
- Distributed annotation of any resource by anyone, in this case using digital portfolios.
- Personalization of tools, queries and interfaces, affecting the experience in several ways.
- Competency declarations and discovery for personal contacts.

For another scenario, see [8].

5 Conclusions and future work

Learning, as well as other human activities, cannot be confined within rigidly defined boundaries such as course systems [16]. Moreover, a learning environment has to support trust building [26] and rich forms of communication between teachers and learners as well as between learners. In order to be powerful, the environment must be inspiring and trigger curiosity for the learning task. We believe Semantic Web technologies form a basis for realizing a multitude of fascinating e-learning visions. But without the proper meta-data semantics, the visions will not be implementable.

Although much of the present development within e-learning is driven by the so-called knowledge economy, there are more fundamentally important issues for the future; namely, how to provide access to knowledge for people who can not afford to pay. Our research work is driven by the overall vision of a global knowledge community, where relevant information and effective support for the knowledge construction process is freely available for all.

References

- [1] *Proceedings of the 2nd European Web-Based Learning Environment Conference (WBLE)*, Lund, Sweden, Oct. 2001.
- [2] Advanced Distributed Learning Network. <http://www.adlnet.org/>.
- [3] Amaya, W3C's Editor/Browser. <http://www.w3.org/Amaya/>.
- [4] Annotea Project. <http://www.w3.org/2001/Annotea/>.

- [5] T. Berners-Lee. Metadata Architecture. <http://www.w3.org/DesignIssues/Metadata.html>, Jan. 1997.
- [6] T. Berners-Lee. Semantic Web Roadmap. <http://www.w3.org/DesignIssues/Semantic.html>, Sept. 1998.
- [7] T. Berners-Lee. Why the RDF Model is Different from the XML Model. <http://www.w3.org/DesignIssues/RDF-XML.html>, Sept. 1998.
- [8] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, May 2001. <http://www.sciam.com/2001/0501issue/0501berners-lee.html>.
- [9] D. Brickley. The Power of Metadata. <http://www.openp2p.com/pub/a/p2p/2001/01/18/metadata.html>, Jan 2001.
- [10] M. J. Carnot, B. Dunn, A. J. Cañas, P. Gram, and J. Muldoon. Concept Maps vs. Web Pages for Information Searching and Browsing. <http://www.coginst.uwf.edu/~acanas/>.
- [11] Centre for User Oriented IT Design. <http://cid.nada.kth.se/>.
- [12] W. W. Chang. A discussion of the relationship between rdf-schema and uml. <http://www.w3.org/TR/NOTE-rdf-uml/>, Aug. 1998.
- [13] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A Distributed Anonymous Information Storage and Retrieval System. In H. Federrath, editor, *Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability*, LNCS2009, New York, 2001. Springer. <http://freenet.sourceforge.net/>.
- [14] S. Cranefield. Networked Knowledge Representation and Exchange using UML and RDF. *Journal of Digital information*, 1, Feb. 2001. <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Cranefield/>.
- [15] S. Decker, C. Manning, A. Naeve, W. Nejdil, T. Risch, and R. Studer. Edutella - An Infrastructure for the Exchange of Educational Media. Part of the PADLR proposal to WGLN, Mars 2001.
- [16] J. D. Douglas. Only Freedom of Education Can Solve America's Bureaucratic Crisis of Education. *Policy Analysis*, 155, 1991. <http://www.cato.org/pubs/pas/pa-155.html>.
- [17] Dublin Core Metadata Initiative. <http://dublincore.org/>.
- [18] Edutella Project Homepage. <http://edutella.jxta.org/>.
- [19] S. Gnagni. Building Blocks - How the Standards Movement plans to revolutionize Electronic Learning. http://www.universitybusiness.com/0101/cover_building.html, 2001.

- [20] IEEE Learning Object Metadata. <http://ltsc.ieee.org/wg12/>.
- [21] IEEE Learning Technology Standards Committee. <http://ltsc.ieee.org/>.
- [22] IMS Content Packaging Specification. <http://www.imsglobal.org/content/>.
- [23] IMS Global Learning Consortium. <http://www.imsglobal.org/>.
- [24] IMS Meta-data Specification. <http://www.imsglobal.org/metadata/>.
- [25] ISO/IEC JTC1/SC36 - Information Technology for Learning, Education, and Training. <http://jtc1sc36.org/>.
- [26] C. Knudsen and A. Naeve. Presence Production in a Distributed Shared Virtual Environment for Exploring Mathematics. In *Proceedings of the 8th International Conference on Advanced Computer Systems (ACS 2001)*, Szczecin, Poland, 2001.
- [27] J. McCarthy. First Order Theories of Individual Concepts and Propositions. *Machine Intelligence*, 9, 1979. <http://www-formal.stanford.edu/jmc/concepts.html>.
- [28] S. Melnik and S. Decker. A Layered Approach to Information Modeling and Interoperability on the Web. Technical report, Database group, Stanford University, Sept. 2000. <http://www-db.stanford.edu/~melnik/pub/sw00>.
- [29] Mikael Nilsson et al. IMS 1.2 meta-data RDF binding. <http://www.imsproject.org/rdf/index.html>, April 2001.
- [30] MIT Open Knowledge Initiative. <http://web.mit.edu/oki/>.
- [31] A. Naeve. The Garden of Knowledge as a Knowledge Manifold - A Conceptual Framework for Computer Supported Subjective Education. Technical Report CID-17, TRITA-NA-D9708, Department of Numerical Analysis and Computing Science, KTH, Stockholm, 1997. http://kmr.nada.kth.se/papers/KnowledgeManifolds/cid_17.pdf.
- [32] A. Naeve. Conceptual Navigation and Multiple Scale Narration in a Knowledge Manifold. Technical Report CID-52, TRITA-NA-D9910, Department of Numerical Analysis and Computing Science, KTH, Stockholm, 1999. http://kmr.nada.kth.se/papers/ConceptualBrowsing/cid_52.pdf.
- [33] A. Naeve. The Concept Browser - a New Form of Knowledge Management Tool. [1].
- [34] A. Naeve. The Knowledge Manifold - an Educational Architecture that supports Inquiry-Based Customizable Forms of E-Learning. [1].

- [35] A. Naeve, M. Nilsson, and M. Palmér. The Conceptual Web - our research vision. In *Proceedings of the first Semantic Web Working Symposium*, Stanford, Jul. 2001.
- [36] M. Nilsson. The Semantic Web: How RDF Will Change Learning Technology Standards. <http://www.cetis.ac.uk/content/20010927172953/viewArticle>, Sept. 2001.
- [37] M. Nilsson and M. Palmér. Conzilla - Towards a Concept Browser. Technical Report CID-53, TRITA-NA-D9911, Department of Numerical Analysis and Computing Science, KTH, Stockholm, 1999. http://kmr.nada.kth.se/papers/ConceptualBrowsing/cid_53.pdf.
- [38] M. Palmér, A. Naeve, and M. Nilsson. E-Learning in the Semantic Age. [1].
- [39] D. Pettersson. Aspect Filtering as a Tool to Support Conceptual Exploration and Presentation. Technical Report TRITA-NA-E0079, Department of Numerical Analysis and Computing Science, KTH, Stockholm, Dec. 2001.
- [40] J. Rumbaugh, I. Jacobson, and G. Booch. *The Unified Modeling Language Reference Manual*. Addison Wesley Longman Inc., 1999.
- [41] Sun Microsystems, Inc. Project JXTA: An open Innovative Collection. <http://www.jxta.org/project/www/docs/OpenInnovative.pdf>, April 2001.
- [42] G. Taxén and A. Naeve. CyberMath - Exploring Open Issues in VR-based Learning. In *SIGGRAPH 2001 Conference Abstracts and Applications*, pages 49–51, 2001. SIGGRAPH 2001 Educators Program.
- [43] The IMS Editor Vimse. <http://imsevimse.sourceforge.net/>.
- [44] The Knowledge Management Research Group. <http://kmr.nada.kth.se/>. CID/NADA/KTH.
- [45] The precise UML group. <http://www.cs.york.ac.uk/puml/>.
- [46] The Protégé-2000 Project. <http://www.smi.stanford.edu/projects/tege/>.
- [47] Virtual Workspace Environment. <http://www.vwe.nu/>.
- [48] S. Waterhouse. JXTA Search: Distributed Search for Distributed Networks. <http://search.jxta.org/JXTAsearch.pdf>.
- [49] S. Wilson. The next big thing? - Three architectural frameworks for learning technologies. <http://www.cetis.ac.uk/content/20010828163808/viewArticle>, Aug. 2001. CETIS, Centre for Educational Technology Interoperability Specifications.