The Next Wave:
CETIS interviews Mikael Nilsson about the Edutella project

Scott Wilson, Mikael Nilsson
Feature article, Centre for Educational Technology Interoperability Standards (CETIS)

CID, CENTRE FOR USER ORIENTED IT DESIGN

The Next Wave:
CETIS interviews Mikael Nilsson about the Edutella project

Scott Wilson, Mikael Nilsson.
Feature article, Centre for Educational Technology Interoperability Standards (CETIS)

Scott Wilson, Mikael Nilsson.

The Next Wave: CETIS interviews Mikael Nilsson about the Edutella project
Feature article, Centre for Educational Technology Interoperability Standards (CETIS)
Report number: CID-200
ISSN number: ISSN 1403 - 0721 (print) 1403 - 073 X (Web/PDF)
Publication date: Sept 2001
E-mail of author: mini@nada.kth.se

The next wave: CETIS interviews Mikael Nilsson about the Edutella project

Scott Wilson, CETIS staff
September 27, 2001

We are in the midst, it seems, of a quiet revolution in computing. The technologies that built the internet, with its giant server farms and gigabyte databases, are giving way to a new wave of distributed technologies. After the furore over music network Napster dies down, we are seeing a new generation of projects in the education field taking up the peer-to-peer challenge of building distributed, "democratic" learning technology.

One of the most interesting new projects out there is Edutella, an attempt to create a distributed web for learning metadata making use of emerging standards. We talked to Mikael Nilsson about the project, and what the future holds.

CETIS: First, can you tell me something about yourself, and how you came to get involved in this project?

MN: I am working as a PhD student within the KMR (knowledge management research) group headed by Ambjörn Naeve at CID, the Center for User-Oriented IT-design at the Royal Institute of Technology in Stockholm, Sweden. Behind me, I have MSc titles in Mathematics and in Computer Science. My main interests are in using emerging Web technologies to enhance the learning experience in a significant way, with a special focus on mathematics.

Three years ago, I started working on a project at CID called Conzilla, which is a new kind of knowledge management tool which we call a concept browser, for conceptual exploration and overview. I quickly became engaged in meta-data issues, as we were using IMS meta-data in this project. At about the same time, we were starting to look into RDF as a powerful encoding format for the kind of information we were navigating with Conzilla. This way, I got involved in the metadata effort at IMS, and coordinated the effort to produce an RDF binding, which was a necessary requirement for us to use RDF.

> **❝ peer-to-peer meta-data architecture just seemed like an idea too good to be ignored ❞**

About a year ago, we started a collaboration with Learning Lab Lower Saxony in Hannover, Germany, on distributed learning repositories and personalization. The idea of a peer-to-peer exchange network for educational meta-data came from them, but fitted perfectly into our vision of navigation and presentation of a distributed knowledge landscape.

We are working on the design of interactive learning environments on several levels. We have a philosophical framework called the 'knowledge manifold' for designing learning tools and environments. We are working on the pedagogical as well as the technical design of flexible learning frameworks, including tools supporting such frameworks.

From this perpective, combining RDF and a peer-to-peer meta-data architecture just seemed like an idea too good to be ignored.

CETIS: So who else is involved? Who are the main sponsors of the project?

The project is financed by the Wallenberg Global Learning Network, and is carried out as a cooperation between CID (Ambjörn Naeve et al.), the
Uppsala Learning Lab (Tore Risch et al.), the Learning Lab Lower Saxony (Wolfgang Neidl et al.), and Stanford Infolab (Stefan Decker, Chris Manning et al.). Recently, the collaboration has been expanded to include University of Karlsruhe (Rudi Studer, Steffen Staab et al.) and the Business University of Vienna/EU-IST UNIVERSAL project Gustaf Neumann et.al.)

CETIS: What is the main purpose of Edutella? Who do you see using it?

MN: The overarching goal of Edutella is to facilitate the reuse of globally distributed learning resources by creating what I like to picture as an eco-system for meta-data, a place for meta-data to flourish and cross-fertilize.

More concretely, Edutella aims to produce an open-source, standards-based peer-to-peer architecture for the exchange of RDF-based meta-data. This architecture will include services such as advanced distributed queries, semantic mappings between schemas, replication of meta-data, distributed annotation etc.

The providers of meta-data will be anyone with content they want to make available. This includes anything from individual teachers and students to universities and other educational institutions.

When it comes to the users of Edutella, we hope that Edutella will become a technology that benefits students and teachers in their every-day use of computers for learning and teaching. The applications that may benefit from Edutella support are many and diverse. Some of the applications are being studied within the WGLN project funding Edutella.

Within the KMR group at CID we are working, among other things, on designing an open, layered framework for distributed learning applications. Edutella will play a central role in that research.

CETIS: So how easy do you envisage it being for someone to set up an Edutella peer?

MN: The vision is to make it very easy to set up the simplest peers, while allowing for advanced peers to announce their capabilities. The simplest client-only peers could be small applets executing in a PDA or other mobile device, or perhaps software agents of different kinds.

The simplest peers that would be active, in the sense of presenting services to the network, should be possible to set up using, for example, a standard web-server with an Edutella plug-in.

On the other hand, advanced peers could for example be backed by a complex database of learning resources or acting as a mediator to a whole network of query services.

> **The simplest client-only peers could be small applets executing in a PDA or other mobile device, or perhaps software agents of different kinds.**

CETIS: Could you describe a few scenarios of how people might make use of Edutella?

MN: Imagine you are looking for resources related to Taylor expansions in mathematics. You open your Edutella navigator and 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, you add the requirement that a professor at your university must have given a good review of the resource. You find some interesting animations, 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.

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 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 site. So it retrieves the formulas, opens an interactive manipulation window, and lets you experiment.

Your questions concerning Taylor expansion multiply. You badly need some deeper answers. Asking Edutella for knowledge sources at your university that have announced interest in helping out with Calculus matters, you find a fellow student and a few math teachers. Deciding 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.

This scenario is not a complete fantasy. We are designing tools to enable this kind of learning experience via Edutella right now at CID and the other institutions involved, and we are researching ways to make them even better. We see some of the important points of Edutella being used in this scenario:

- 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 (calendaring info, finding the right kind of resources).
- Tool interoperability. Any tool can use the technology.
- Distributed annotation of any resource by anyone.
- Personalization of tools, queries and interfaces, affecting the experience in several ways.

CETIS: A lot of the discussion around "digital repositories" for education has centred around client-server systems, with the bulk of the material sat in relational databases. So what prompted you to look at a peer-to-peer solution?

MN: One philosophical motivation that has played an important role 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 the is no single point of control, no middle man in control of the network.

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, as is turns out, is also deliberately designed as a peer-to-peer architecture, where anyone can **say** anything about anything, so it really fits into that philosophy. Edutella makes it possible for anyone, even with very limited technical and financial resources, to participate in the exchange of learning resources.

> **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**

Of course most large institutions will use databases even in the future, in varying forms and with varying content. A second motivation for the peer-to-peer solution is the problem of getting those systems to talk to each other, to interoperate in an active manner, and on a global scale. Simply having a database and perhaps a web interface does not solve the problem of distributing information about your material globally.

Peer-to-peer networks have, by contrast, proven themselves to be effective propagators of information. One of the main disadvantages of classical peer-to-peer networks is the lack of rich meta-data -- often only supporting searching on a file name. Harnessing the distributive power of peer-to-peer networks for the dissemination of RDF-based meta-data seems very natural once the idea has been formulated.

A third motivation lies in the overall system architecture. Peer-to-peer networks have the advantages of simple setup, extensibility in provided services, and scalability.

CETIS: How will the network of Edutella peers join up? Will there be a central server that lists the peers available? Also, when searching Edutella, will you have to search a particular node, or will you be looking across all the available peers?

We are designing a system where peers will announce their querying capabilities within the network so that searches can be directed to the right peers in the network. Some peers may act as powerful mediators to simpler search services, thus enhancing the searches. When you want to search, you distribute your query within Edutella and the query is routed to the appropriate peers and mediators.

One important strength of the Edutella query design is its powerful transparency. Answering queries may involve combining partial results from a set of disparate information sources. This form of intelligent combining occurs within the network in a way that is transparent to the querying peer.

We are trying hard to avoid dependencies on central servers. This is not a question of avoiding legal issues, but a decision taken from a scalability and reliability standpoint.

That said, the query and routing protocols are still under heavy discussion. The design has not yet been set in stone.

CETIS: Will users without an Edutella peer application be able to make use of the Edutella network, using a web search engine for example?

MN: Sure! Web interfaces with Edutella backends are an interesting possibility, and will probably be one of the first demo applications. This is just a question of implementing an Edutella peer behind the scenes on that web server. In fact, many other kinds of "stand-in" peers are imaginable, such as in the middle-tier logic in three-tier architectures, enabling very leight-weight clients, or why not software agents.

CETIS: Unlike other peer-to-peer distribution systems (like Napster for example), you describe Edutella as a service for exchanging metadata rather than content. Was this prompted by copyright issues, or have you always

thought of Edutella as being essentially about metadata?

MN: Edutella has always been about meta-data. Access to educational content is not always as simple as downloading a file -- it might include logging in to a web service, starting a certain application with specific parameters, etc.

Copyright issues are of course important, and difficult for this kind of network. Sometimes the meta-data itself is subject to copyright restrictions, which means that copyright meta-data possibly should be included in the system. On the other hand, we are not planning to include any copyright **enforcement** schemes into Edutella.

Another related point is digital signatures to certify meta-data, which will probably have to be included at some point.

CETIS: What kinds of metadata will Edutella hold? There are minimalist approaches like Dublin Core, or the IEEE-LOM/IMS/SCORM structural metadata. Plus you have IMS subsets like CanCore. Which are you thinking of supporting?

We are aiming for a layered architecture. The meta-data will build on Dublin Core at the bottom layer. The second layer will be formed by the RDF binding of IMS/IEEE-LOM meta-data that was produced under our coordination, and which has recently been included in the IMS meta-data specification. This new IMS/IEEE-LOM RDF binding is Dublin Core-compatible, so the two layers will coexist nicely.

We are also working on producing an RDF binding of the IMS Content Packaging specification (which forms a large part of the SCORM Content Structure Format), and this will form the third layer.

RDF allows you to easily extend your meta-data descriptions with other schemas (such as AICC, the second part of SCORM). We are designing Edutella to be schema-agnostic, so that information in any schema can be included. However, Edutella will probably have additional built-in support for the above three layers, which we view as the fundamental building blocks.

CETIS: How do you envisage metadata getting onto Edutella peers? Will you be able to collect metadata from existing repositories spider-fashion? Or do you see repositories as using Edutella to publicise their resources?

Edutella meta-data providers provide query services, which take responsibility for certain queries and for the routing of queries to other services. Thus, the totality of content of repositories is never explicitly announced, but only available through the query service.

Many peers will have a database backend, and it might be that certain peers interface to legacy databases or crawlers, but that is outside the scope of Edutella proper. It is a peer implementation issue.

> **" adding personal annotations or professional evaluations to any content is extremely important, and is easily done within Edutella "**

CETIS: Will users be able to provide their own "alternative metadata" for existing resources? For example, if the publisher of a resource describes it using their own metadata, can the users of that resource publish alternative metadata on Edutella giving their own view of how that resource should be described? If this is the case, how will searches deal with these multiple descriptions? (This presupposes that metadata is itself subjective)

This is one of the fundamental reasons why we are using RDF, as it allows distributed annotation in an intelligent way. As a simple example, managing translations of resource descriptions (those that are textual, that is) separately from the "master" meta-data is definitely feasible. In the same way, the possiblity of adding personal annotations or professional evaluations to any content is extremely important, and is easily done within Edutella. Another important possibility along the same lines is the possibility of describing the same resource according to different schemas.

You will note that this connects in a natural way to the democratic vision of Edutella.

This also shows why the monolithic XML documents that have become a de-facto standard within LOM/IMS/SCORM do not fit the meta-data concept very well, except for meta-data instance interchange. By their very nature, resource descriptions are subjective, distributed and expressed in diverse forms, and this must be reflected in the encoding language. Therefore we have started a campaign for IMS, IEEE and SCORM to start the transition to RDF for meta-data.

Naturally, the problem of supporting this fundamental subjectivity in searches is not trivial. But the built-in support in RDF for meta-meta-data will make this task surmountable, we believe. Imagine, as a simple example, adding a link called "Who said this?" to each query answer. Another possilibity is to add functionality to search using only trusted sources.

CETIS: How will users get from the metadata for a resource to the resource media itself?

MN: As we are using RDF, each resource must have a URI. The way from that URI to the resource itself is not determined by Edutella. It may be an HTTP URL, so that your Edutella-aware application can point your browser in the right direction. Or it might just be a URN, uniquely naming the resource but not locating it, and you must go through some sort of lookup service to find it. I imagine that information about lookup services and retrieval methods will be distributed within Edutella as well.

The main point is that describing a resource and accessing it are two separate activities. Edutella focuses on descriptions, which of course may include describing the location, but not on the retrieval process. That being said, there have been some discussions on allowing simple forms of content to be transported within Edutella, but that is a future project.

CETIS: As well as metadata describing single learning objects, do you also envisage aggregated metadata for content packages being used? So that, for example, a user can group together metadata identifying resources for a particular course, and then create a 'package' containing references to those resources plus metadata about the package?

If so, how do you see this working? Would you use the IMS Content Packaging specification, or some other scheme?

Yes, this is an extremely important capability in order to be able to define courses and related content etc. This is why we are working on including an RDF binding of IMS Content Packaging as a fundamental part of Edutella.

On the other hand, we will not prohibit the use of other schemas to describe such structures, even though that would create interoperability problems.

We are designing a schema translation methodology and service to be included in the Edutella architecture that will be helpful in situations where different schemas say similar things.

CETIS: One of the big arguments in learning technology is "what are learning objects?". We have conflicting views of the type of granularity of objects, whether they must have a specified (pedagogic) purpose, contain assessments and learning objectives, whether they are recursive in nature, and so on. How does this debate affect Edutella?

MN: It does very much affect us. This again shows that most meta-data descriptions are subjective in some sense, and underlines the importance of supporting the RDF meta-meta-data constructs in Edutella.

When meta-data descriptions are properly annotated with their source, creating meta-data is no longer a question of finding **the** objective description of a resource. Multiple, even conflicting descriptions can co-exist. In fact, we **want** people to be able to express personal views on these subjects. Otherwise, the technology will be completely lifeless. 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.

❝ meta-data can evolve and be reused in new and unanticipated contexts ❞

CETIS: One issue relating to learning objects is context; that is, how can you describe a learning object if you don't know what its intended use is? For example, a single piece of media like a photograph has different meaning as part of a History context than when used in a Photography context. Will Edutella tackle this kind of issue or will you keep your learning objects context-neutral?

MN: Note that the flexibility of **content** depends on the content provider and not on Edutella. Thus, some content might never be possible to reuse in new contexts -- even though this is the goal.

The context neutrality of **meta-data** is a totally different matter. This initially depends on the meta-data provider. Some may choose to publish neutral meta-data, and other may choose to publish context-specific. The levels of meta-data descriptions provided by SCORM is an important guide in the process of choosing the right meta-data descriptions.

Thanks to RDF, different kinds and layers of context-specific meta-data can always be added by others when the need arises. This is a perfect example of the learning meta-data eco-system in action, where meta-data can evolve and be reused in new and unanticipated contexts.

CETIS: How do you plan to develop Edutella? Have you decided on platforms, tools, or languages?

MN: We plan to build Edutella as a service layer on top of the open-source JXTA peer-to-peer architecture (http://www.jxta.org/), initiated by Sun Microsystems. In fact, it will be a collection of interrelated JXTA services. JXTA is platform- and language-agnostic, and is based on XML, in order to insure maximum interoperability. Such an open architecture has been a priority since day one.

We hope to make Edutella a fundamental part of the central JXTA services, such as JXTA Search. Enabling RDF searches within JXTA have clear benefits for the whole JXTA project, and they have shown great interest in Edutella.

CETIS: I notice you already have 5 developers registered with the project. Are you looking for more people to get involved, or are you going to keep a smaller team for now? If the former, how do people get involved?

MN: Actually, we are many more involved in the design discussions, as can be seen on the mailing lists. But on the other hand, a project like this cannot remain closed. We encourage any kind of participation! I just want to add that the first phase will be heavily system design oriented, while actual implementation will start somewhat later.

Edutella is now an official JXTA project, and development can be followed on http://edutella.jxta.org/.

CETIS: Whats the timescale involved? When will we be able to see the first pieces of Edutella?

MN: We hope to be able to produce a working framework within the coming year, with a first prototype including simple demos and mock-ups by the end of 2001. The first pieces of the design are being prepared right now, and need to be finalized quickly, so some pieces will hopefully be available in the coming months. But we are not a commercial organization, so any timelines are subject to change.

CETIS: So just how big a difference do you see Edutella making?

MN: I think it has great potential! It is a realization of some of the visions being discussed within the Semantic Web community, such as distributed, intelligent queries and the meta-data ecosystem, so I believe it will have a large impact within that community. As the semantic web seems to be the future of web technologies, the future looks promising.

On the other hand, the need for a global infrastructure for learning resource interchange has not yet been fulfilled, in spite of the many standards and projects within this domain. I think Edutella has a chance of affecting e-learning projects in a large way thanks to this global perspective on learning resources, a perspective shared with important projects such as the MIT Open Knowledge Initiative. It is important to realize that the project is not primarily driven by technology, but by a vision of a global network of learning.

For more information on Edutella, visit http://edutella.jxta.org

Mikael Nilsson can be contacted by email at mini@nada.kth.se