LINKED OPEN AALTO
PROJECT PROPOSAL

Prof. Eero Hyvönen
Semantic Computing Research Group (SeC0)
Dept. of Media Technology Aalto University
http://www.seco.tkk.fi/
eero.hyvonen@aalto.fi

ABTRACT

Linked Open Aalto *is a research project aiming at developing a semantic web approach for creating and publishing interlinked educational, research, and managerial contents produced at different communities, schools, departments, research groups, and persons in Aalto. By using semantic Linked (Open) Data principles, technologies, and open datasets available, Aalto contents can be interlinked with related teaching and research materials in Finland and internationally. By aggregating and combining local contents from separate incompatible data silos and systems, the end-user can be provided with a global, cross-disciplinary perspective to knowledge produced in Aalto and other universities. For example, a web page describing a course can be interlinked automatically with related research results, publications, projects, Wikipedia pages, research groups, researchers, internationally available video lectures, open course materials, events in Aalto, conferences, blog discussions, and so on.*

BACKGROUND

LINKED OPEN DATA: INTERNATIONAL HOT TOPIC AND MOVEMENT

Web of Data, also called the Giant Global Graph (GGG), is a recent megatrend of web research based on Tim Berners-Lee's original vision of the semantic web (Heath, Bizer, 2011). The idea and international collaboration initiative, supported by the World Wide Web Consortium (W3C), aims at creating,   inside of the World Wide Web (WWW), a web of machine "understandable" semantic metadata, using the principles of Linked Data and Linked Open Data datasets (http://linkeddata.org). In this way, virtually all kinds of cross-domain contents from public sector, companies, academic research, social media etc. could be made semantically interoperable, and used in next generation "intelligent" web services and applications based on vast amounts of data. The core LOD cloud, built around DBPedia extracted from Wikipedia, has at the moment already 30 billion semantic connections (triples) between data items, and 500 million interlinking connections across datasets. The idea of Linked Open Data has already been adopted as the basis for data publication in various countries on a national level, e.g. in the U.K. (http://data.gov.uk/).

AALTO IS WELL-POSITIONED

In Finland, these ideas have been developed and deployed e.g. in the various application demonstrators of the national FinnONTO-project (2003-2012). FinnONTO is lead by the Semantic Computing Research Group (SeCo) at Aalto and the University of Helsinki. The project is funded currently by some national 40 organizations, and is well-know also internationally in the Semantic Web community. The FinnONTO ontology infra, ONKI ontology services (15,000 unique monthly users), and tools provide a strong basis on which the proposed Linked Open Aalto –project can be grounded (Hyvönen et al., 2008). The project will be supported by the new national FinnONTO follow-up project "Linked (Open) Data Finland" (2012-2014) lead by SeCo, too.

RESEARCH PROBLEM

A university such as Aalto is a "knowledge factory", producing new scientific knowledge, knowhow, and data for students, researchers, scientific communities, companies, and the society in the large. At the moment, knowledge and data production and publication is based on using **separated data silos** of different domains, such as journal series, seminars and conferences, internal publications, curricula and course descriptions, web pages, demo systems, news feeds, email lists etc. Different communities typically have created *local* data publishing systems and forums of their own, without connections to related communities and information systems. A problem from a student's, researcher's, and manager's viewpoints is how to get a cross-domain, *global* understanding of all activities about a topic (s)he is interested in. Organizational boundaries between scientific disciplines, communities, university departments, schools, universities, countries, and languages make this very hard indeed.

USE CASES

For example, consider the very topic of this application "Linked Open Data" and the following use cases:

1. A **student**/researcher is considering, whether (s)he should to focus on studying this topic.

2. A **researcher** is planning a new research project in this field, and is establishing a context and collaboration network for this, and looking for colleagues in Aalto schools to collaborate with.

3. A **research director** would like to evaluate whether a research project in this field should be funded.

4. A **journalist** has got an invitation to a seminar and now looks information and research about the topic and for researchers in Aalto to interview them.

THE PROBLEM OF SEMANTIC DATA AGGREGATION

Following pieces of information about Aalto activities are then needed, are available somewhere but in different data silos, or is otherwise difficult to find using current web sites and systems:

- **What is the value of this topic?** An authorative short article, e.g. in Wikipedia or Freebase would be helpful, with references to key literature, news, and other relevant information.

- **Who are doing related research in Aalto?** Related research is carried out in different schools of Aalto, at least in School of Science (Dept. of Media Technology, semantic web research.; Dept. of Computer Science, web data mining, description logics; Dept. of Information and Computer Science, semantic web programming,), in School of Applied Arts (semantic cultural heritage and art), School of Economics (business models of open data and open source).

- **What projects related to the topic are there in Aalto?** Information about current projects is in principle there in web pages and project registries, but are in practice difficult to find, and the project situation is changing rapidly in time.

- **What courses about the topic are available?** Again information is there in Noppa and other Aalto systems, but is not easy to gather.

- **What kind of related events are there in Aalto?** Various conferences, seminars, and other occasions are organized in Aalto, but the information cannot be connected automatically to other pieces of knowledge.

- **What international collaborations are there?** For example, EIT is funding related collaboration networks, there is a new Masters Program related to the topic, etc.

- **What publication have Aaltoers produced related to the topic?** Again, there are registries but they are in silos of their own and not connected to other pieces of knowledge in Aalto. A separate information systems has to be used, again.

- **What software and datasets have Aaltoers produced related to the topic?**

- **What about contents available form other universites?** The questions above apply to other universities. For example, MIT publishes course material openly, there are online courses in Stanford, University of Helsinki and other national universities in Finland has many activities, teaching material etc. available on the web. However, let lay in their own silos.

- **How to search material in different languages?** Education and scientific contents are produced in different languages, but search engines do not support multilingualism. Semantic Linked Data is indexed based on language independent semantic identifiers, making cross-lingual search possible.

We envision that finding new answers to these use cases and problems is essential when developing new, more usable web services for Aalto, evaluated recently in end-user studies (Haahtela, Sainio, 2011; Korhonen, 2011). A challenge found out in this work was, that it is difficult to inform readers of the aalto.fi site and students about research in Aalto, because this information is published in various home pages of departments and research groups.

SOLUTION APPROACH

The idea of Linked Open Data provides a novel approach to these problems. SeCo-group has already developed several related research demonstrators, such as Museum Finland (http://www.museosuomi.fi) and HealthFinland (http://www.tervesuomi.fi), that both were awarded with the Semantic Web Challenge Prize by the international research community. A special focus in our research is how to deal with massive datasets of semantically heterogenous data originating from different multilingual sources (Mäkelä et al., 2011). These ideas and the underlying methods and technology could be developed further in the proposed Linked Open Aalto project.

The idea is to **develop methods and technology** for harvesting and publishing Aalto data silos using Linked Data principles in RDF-form. Based on the global repository, a **demonstrator** of the next generation semantic Aalto portal is created. The pilot system result is **evaluated** with end-users in order find out the values and problems of the approach, and to decide on next steps towards the vision of the Linked Open Aalto. The topical focus is in the beginning on computer science that is related to many other disciplines in Aalto.

The Linked Data approach makes it possible to connect Aalto contents not only within Aalto, but also internationally to the emerging international Linked Open Data cloud of universities, that has quite recently started to evolve (http://linkeduniversities.org). For example, the Open University in London recently published a semantic dataset about 14,000 lecture videos harvested from 27 different web sites and YouTube channels (Fernandez et al., 2011). By using semantically interoperable Linked Data such repositories could be connected to Aalto materials and portal, and conversely, Aalto materials disseminated and used by other universities in their information systems.

In our vision, the Linked Open Aalto portal has web pages for Persons, Research Groups, Courses, and Projects (these are a starting point). By semantic data aggregation, a persons web page is automatically linked with e.g. courses and research projects (s)he is at the moment involved in, publications, web pages of colleagues, conferences in which (s)he is presenting or reviewing papers, etc. A course page in "Linked Open Noppa" could automatically suggest further information about the topic from DBPedia/Freebase, related educational materials

from Aalto and other universities, a synopsis of research and publications in different departments of Aalto about the topic, information about the course's relation to modules and other courses, links to topically related events in Aalto, and so on. To accomplish this, ontologies about curricula, topics, persons, departments, etc. need to be devised, as well as metadata schemas for knowledge representation. A key research challenge will be, how semantically unambiguous metadata creation can be accomplished as automatically and accurately as possible.

Collaboration Network

In Finland

SeCo and the Aalto Univerity IT department have discussed collaboration, and there is a consensus for starting a small pilot. The IT department is able to budget some resources for this (e.g. APIs or data dumps of current IT systems are needed); resources are needed especially for the research part of SeCo from Aalto. SeCo is also discussing collaboration with Aalto University Communications, Marketing and Events. The project is also of potential interest to research projects in Aalto dealing with publication and use of scientific and educational data, e.g. projects on information retrieval research related to large publication databases conducted at HIIT.

SeCo has researchers not only from Aalto but from the University of Helsinki, too, making UH a natural choice for collaboration in Finland. Also other Finnish data sources will be considered, focusing on ready-to-use materials.

International Collaboration

SeCo group has already discussed collaboration with the Open University in London (prof. Enrico Motta) related to this project, and has a post doc researcher in the University of Muenster, one of the first universities that have joined the Linked Universities initiative (and was involved in organizing the first workshop about the topic at the ISWC 2011 conference recently). If the project starts, Aalto will join the Linked Universities network.

Resourcing and Time Table

The project is done as part of PhD research supervised by prof. Eero Hyvönen and in collaboration with Aalto University IT. Collaboration with other departments is possibly also needed later, in order to get datasets from silos. In the initial phase in 2012, one person year with travel and additional costs (50 000 EUR) is needed in order to create a more complete plan of the Linked Open Aalto, as well as small limited demos for evaluating the possibilities and challenges. Based on the results, a larger project plan is devised.

The ultimate goal of our research is, in addition to top level academic publications and results, that the research prototype solutions created in the research projects will finally be imported into operational Aalto portals by the Aalto university IT, Aalto University Communications, Marketing and Events, and companies developing the Aalto web sites. Separate additional funding is of course needed for deploying the solutions in practice. This was the case e.g. with our former prize-winning HealthFinland (Suominen et al., 2008), that was deployed/commercialized by two Finnish companies and the National Health Institute in Finland.

References

Heath T., Bizer C.: Linked Data: Evolving the Web into a Global Data Space. Claypool & Morgan, CA, 2011.

M. Fernandez, M. d'Aquin, E. Motta: Linking DataAcross universities: An Integrated Video Lectures Dataset. Proceedings of ISWC 2011, Springer-Verlag, 2011.

Marjo Haahtela, Mikael Sainio: Aalto.fi-sivusto ja Inside-intranet. Käytettävyystesti. Agage Ltd, Finland, 2011.

Eero Hyvönen, Kim Viljanen, Jouni Tuominen and Katri Seppälä: Building a National Semantic Web Ontology and Ontology Service Infrastructure--The FinnONTO Approach. Proceedings of the European Semantic Web Conference ESWC 2008, Springer-Verlag, 2008.

Hanna Korhonen: Intranet Insiden käyttäjäkyselyn (toukokuu 2011) tulokset. Communications, Marketing and Events, Aalto-Univeristy, 2011.

Eetu Mäkelä, Eero Hyvönen and Tuukka Ruotsalo: How to deal with massively heterogeneous cultural heritage data – lessons learned in CultureSampo. Semantic Web Journal, 2011. Accepted for publication.

Osma Suominen, Eero Hyvönen, Kim Viljanen and Eija Hukka: HealthFinland - a National Semantic Publishing Network and Portal for Health Information. Journal of Web Semantics, vol. 7, no. 4, pp. 271-376.