Dutch Semantic Web day
16 March 2009

http://esw.w3.org/topic/DutchSemanticWebGettogether
VU University Amsterdam, MF G513, van der Boechorststraat 7–9, Amsterdam

09:30–09:45	welcome
09:45–10:30	Ivan Herman: updates on POWDER, RIF & SPARQL
10:30–10:45	*coffee*
10:30–11:30	speeddating
11:30–12:30	lightning talks & discussion
12:30–13:15	*lunch*
13:15–14:15	lightning talks & discussion
14:15–14:30	*coffee*
14:30–15:30	lightning talks & discussion
15:30–16:30	speeddating and drinks

Henk Matthezing ANP
henk.matthezing@kb.nl

Alia Amin CWI
a.k.amin@cwi.nl

Comparison search: Interest in linking heterogeneous semantic data has been growing rapidly in the past several years. Despite this interest, however, user-centered exploration on novel applications that would benefit from semantic web technology in the real world setting remains small in numbers. In this poster, we discuss LISA, a comparison search application that enables users to compare similarity and differences across heterogeneous cultural heritage repositories. We conducted a user-centered design process that included a field study, design and implementation and user evaluation. Preliminary user feedbacks on the functionality & interface of the application are positive. We have identified comparison search use cases from cultural heritage experts. We are currently in the middle of the implementation phase and we plan to conduct an evaluation phase in the near future.

Michiel Hildebrand CWI
michiel.hildebrand@cwi.nl VU
Jacco van Ossenbruggen
jacco.van.ossenbruggen@cwi.nl

We investigate to what extent explicit semantics can be used to support end users with the exploration of a large heterogeneous collection. In particular we consider cultural heritage, a knowledge-rich domain in which collections are typically described by multiple thesauri. We focus on three types of end user functionality. First, searching for terms within multiple thesauri to support manual annotation. Second, keyword search, as it has become the de-facto standard to access data on the web. Third, faceted browsing as it has become a popular method to interactively explore (image) collections. For these three tasks we question the role of explicit semantics in the search algorithm, the result organization and visualization and how to evaluate the added value of for end users. We investigate these questions by the implementation and evaluation of three prototype systems on top of large and real wold data collections.

Ivan Herman CWI
ivan@w3.org W3C

This presentation will give a short overview of some of the technologies that are actively developed at W3C today, namely POWDER, RIF, and the work of the newly started SPARQL groups. The slides are available at http://www.w3.org/2009/Talks/0316-Amsterdam-IH/

Davy van Deursen Ghent
davy.vandeursen@ugent.be U.
Sam Coppens
sam.coppens@ugent.be
Erik Mannens
erik.mannens@ugent.be

With the growing amount of multimedia content in the current multimedia landscape, it is important to have an efficient, accurate, and automatic metadata creation process for the multimedia content. However, algorithms for automatic feature extraction and interpretation of multimedia content are usually developed by people working in the signal processing research domain; they are less familiar with metadata modeling (i.e., formatting the output of their algorithms according to a specific metadata scheme). Moreover, it is a non-trivial task to use the outcome of feature extraction algorithms (usually low-level features such as colors and patterns) for the assignment of high-level semantics (e.g., a car) to multimedia content (this problem is related to the semantic gap). Therefore, we believe that a collaborative effort between the Semantic Web and the signal processing community could serve as a basis to develop a structured and well-defined generic framework for the automatic generation of high-level, formally described metadata.

Paul Groth ISI
pgroth@isi.edu

The emergence of Linked Data is exciting as it provides the basis for easily creating new data driven applications (see http://election.windycitizen.com/candidates/all). Additionally, as tools that leverage structured data to help end-users create mash-ups (PotLuck, Taverna) become widely available; there will be a proliferation of new applications generating even more data. This would be great! However, in such environments the need for determining provenance becomes even more important as it's difficult to ascertain whether one should trust what is being produced by these semantic mash-ups. Practical work in understanding provenance (in particular interoperability) is going on in the scientific workflow community under the heading of The Open Provenance Model, and Provenance Challenge. I believe this can help inspire provenance solutions for broader Semantic Web applications.

Hennie Brugman hennie.brugman@mpi.nl Véronique Malaisé vmalaise@gmail.com Luit Gazendam luit.gazendam@telin.nl	MPI VU Telin	The Documentalist Support System: a Web-Services based Tool for Semantic Annotation and Browsing The Documentalist Support System (DocSS) is developed to suite novel needs of documentalists working within the Dutch archive for Sound and Vision, broadcasters working outside of Sound and Vision and people interested in the Cultural Heritage value of the archive, who want to perform search in context. The documentalists (and to some extent the other users mentioned) need an environment in which they can view and manipulate multiple types of information (documents and metadata), receive annotation suggestions from their controlled vocabularies, create catalogue descriptions and browse for semantically similar documents within their collection. The open architecture, the publicly published annotation format and the SKOS representation requirement for the controlled vocabularies make the generated annotations interoper- able with other annotation databases and the DocSS usable for any documentalist annotating material with controlled vocabularies, for which a digital textual representation of the data exists.
Lloyd Rutledge lloyd.rutledge@ou.nl	OU	The Semantic Forms extension for Semantic MediaWiki provides interfaces for displaying annotations as infoboxes and entering them with forms. We aim to make Semantic Forms "friendlier" to both users and the Semantic Web simultaneously by adding Semantic Web functionality and processing it for facilitating form-based annotation, improving factbox displays and extending Semantic Web export. This effort centers around two academic activities at the Open Universiteit Nederland: a Bachelors team project and a Masters Semantic Web course. Course students carry out overlapping projects that populate a semantic wiki. Simultaneously, the team project researches, designs and implements user- and semantics-friendly Semantic Forms extensions. The team project will deliver planned incremental to the course wiki for evaluation by its students in the context of developing their projects. The project provides the technology and the course the case study and evaluation for developing and populating ontologies as a community using a facilitating wiki interface.
Gijs Geleijnse gijs.geleijnse@philips.com	Philips	
Jan Jellema jan.jellema@tno.nl	TNO	Ik zou graag wat willen vertellen over onze activiteiten om tot een wereldwijde geowetenschappelijke thesaurus te komen. (Multi-lingual Thesaurus for the Geosciences) Daarnaast hebben we ook diverse tools ontwikkeld om de thesaurus te integreren met onze kennisbank, Hier kan ik ook voorbeelden van laten zien,
Geert-Jan Houben g.j.p.m.houben@tudelft.nl Philipp Cimiano p.cimiano@tudelft.nl Kees van der Sluijs k.a.m.sluijs@tue.nl Pieter Bellekens p.a.e.bellekens@tue.nl	TUD TU/e	We argue why it is necessary to associate linguistic information with ontologies and why more expressive models, beyond RDFS, OWL and SKOS, are needed to capture the relation between natural language constructs on the one hand and ontological structures on the other. We discuss that in the light of such tasks as ontology-based information extraction (i.e., ontology population) from text, ontology learning from text and natural language generation from ontologies, currently available datamodels such as RDFS, OWL and SKOS are not sufficient as they only allow to associate atomic terms without linguistic grounding or structure to ontology elements. Towards this end we have developed a model called LingInfo which allows to associate complex lexical and linguistic information to OWL ontologies. This model is based on previous efforts such as LingInfo and LexOnto and uses the Lexical Markup Framework to glue these models together in a principled way, building on standard developed in the linguistics community.
Yiwen Wang y.wang@tue.nl Natalia Stash natalia.stash@gmail.com Lora Aroyo l.m.aroyo@cs.vu.nl	TU/e VU	The use of thesaurus or ontologies provides various semantic relations, such as broader, narrower and related. For content-based recommender systems, these relations enables more sophisticated concepts to be recommended. However, not all related concepts are interesting for end users. In CHIP, we identify nine basic patterns of semantic relations, which are within one vocabulary or across two different vocabularies. We test the CHIP Art Recommender with end users by applying these patterns and comparing the recommendation results in terms of measurements such as precision, frequency, coverage, novelty and serendipity
Rogier Brussee rogier.brussee@telin.nl	Telin	The subject I am working on relating directly to the semantic web is on the one hand relating texts to more formal ontologies and thesauris and the use of semantic web techniques for archives in the cultural and public sector. However this being an informal meeting I am tempted to write an informal paper about one of my pet semantic web peeves (which I still have to do) with a title along the lines "RDF and XML, a bad marriage" with the central tennet that layering RDF on top of XML and URNs and the resulting horrible serialisation is the single biggest obstacle for the large scale adoption of RDF.
Raymond Franz r.franz@trendlight.nl	Trendlight	

Hans Nederbragt hans@trezorix.nl Ger Senden ger@trezorix.nl Marco Vink marco@trezorix.nl Jacco Drenth jacco@trezorix.nl Sander van der Meulen sander@trezorix.nl	Tresaurix	Launched in 2008, the Sterna project is an eContentPlus best-practice network that aims to contribute to the further development of the European Digital Library initiative as a portal to the riches of Europes cultural and scientific heritage. Sternas participants, mostly European institutions that are concerned with collecting and managing content on biodiversity, wildlife and nature in general, join forces to explore new ways of providing their content to the public. The project was initiated by the Netherlands natural history museum Naturalis and major technical contributor Trezorix.
Scott Marshall marshall@science.uva.nl Andrew Gibson a.p.gibson@uva.nl	UvA	As steadily more organizations build data and knowledge repositories, there is a growing interest in information extraction and knowledge capture technologies. Researchers are developing components for knowledge management that will enhance the ability perform computational experiments and share knowledge. In the Virtual Laboratory for e-Science project (VL-e), researchers in Adaptive Information Disclosure (AID) have combined information retrieval, machine learning, and Semantic Web technologies to create the basis for a knowledge management system in the AID Application Toolkit (AIDA). Applications of AIDA components are now available as Taverna workflows, a web interface, and a plugin to a grid resource browser called VBrowser. Some general principles and challenges to knowledge sharing will be described, along with some applications, including use cases that demonstrate resource management and knowledge discovery in biology and food informatics.
Marco Roos m.roos1@uva.nl	UvA	The general topic of my research is to apply e-Science technologies, such as Semantic Web and Workflow, for a biological application. In particular we would like to support hypothesis generation for the elucidation of biomolecular mechanisms that control DNA (chromatin) structure, a determinant of gene expression. Generating hypotheses is an empirical process in which knowledge from literature plays a significant role. We progress towards an application that supports the task of generating a hypothesis about biomolecular mechanisms using Semantic Web technologies and a workflow to carry out text mining in a service-oriented architecture. The methodology can be used to bootstrap the process of human-guided construction of semantically rich biological models using the results of knowledge extraction processes.
Ronald Cornet r.cornet@amc.uva.nl	UvA	Over the last ten years, the dept of Medical Informatics has developed a line of research in the area of (bio)medical knowledge representation. Topics include: development of terminological systems in medicine, applying terminologies in practice, auditing of terminological systems in medicine, integration of terminological models and information architecture, integration of ontologies and classifications, large-scale reasoning. Publications can be found at: http://kik.amc.uva.nl/home/rcornet/publications.html
Maarten de Rijke mdr@science.uva.nl	UvA	We are a group of around 20 people working on intelligent information access, with an emphasis on Information Retrieval, Language Technology and Semistructured Information. During the past 6 months we have set up a number of pilots on large-scale semantic annotation and linking data. This includes web services for entity recognition and normalization as well as semantic search services for all Dutch political data. We'd like to present a poster on our information extraction web services.
Rinke Hoekstra hoekstra@uva.nl Saskia van de Ven svandeven@leibnizcenter.org	UvA	The corpus of legal information is the web's little sister. It consists of a huge volume of heterogeneous, closely inter-linked documents. References between texts are not always absolute, typically point to parts of documents, and often import an externally defined meaning of a term. Consolidation of such semantic references is a significant maintenance issue, as law adopts an intricate versioning scheme: the meaning of terms in law imposes an ordering on entities in reality that can change over time, but stays applicable to older cases. A representation, for annotation, consistency checking, or reasoning, should take these dynamic and structural properties into account. We represent definitions of concepts in OWL in such a way that their semantic interpretation mimics the structure and applicability of the texts. This includes means to scope definitions to parts of a text, as in e.g. deeming provisions, temporal validity, and jurisdiction.
Hajo Rijgersberg hajo.rijgersberg@wur.nl	WUR	Quantitative research is part of most scientific disciplines. At present, the computer support of quantitative research is inadequate. Different, unlinked systems are required and a lot of information remains implicit and is therefore not accessible for automatic processing. The aim is to disclose quantitative information in a semantic way. An application is, for example, Tiffany, a research supporting system in development. In this research 1) we develop a quantitative vocabulary, 2) apply this vocabulary in supporting tools to be developed, and 3) evaluate the use of these tools with researchers (the target group). We aim for generic services such as unit and dimension consistency checking of formulae and converting units and quantities on the basis of mathematical models. Also, we develop an infrastructure that manages the communication of quantitative knowledge between numerical tools such as Matlab.

Jan Top jan.top@wur.nl	WUR	Food industry is increasingly aware of the long-term benefits of precompetitive cooperation in research. This requires a proper knowledge management policy, implemented in organizational processes and supported by dedicated e-science tools. We have developed a generic RDF/OWL-model of the scientific workflow and a web application called Tiffany for accessing this information. We are facing two important issues. First, SW-technology is often used for sharing of public data. However, in many real life applications it is often desirable that some of the data is shielded for some groups of people. In Tiffany we have built a dedicated security authorization mechanism, with roles and different kind of permissions. However, an integrated security and authorization mechanism would be most welcome. Second, an RDF/OWL data store gives a lot of freedom and flexibility to store data. However, in the GUI most data is fixed (users want to know what they can expect). How best to combine these opposite properties?
Don Willems don.willems@wur.nl	WUR	The Emerging Risk Detection Support System is a reasoning system that uses information from various sources to find emerging risks in the food chain. To enable reasoning over the different domains that might influence food safety, we needed to develop a new ontology that is able to describe the different concepts from biology, topography, psychology, logistics, and other domains. New information extracted automatically from the web, or by hand is, validated, added to an RDF repository and then used by a forward chaining rule engine to find new emerging risks. The experts from the food-safety authority who evaluate the newly found emerging risks, also need to be able to assess the reasoning steps that lead to the detection of the emerging risk. In our prototype application, the expert user can evaluate each step in the reasoning process and thereby determine whether an emerging risk needs acting upon.
Nicole Koenderink nicole.koenderink@wur.nl Roos Groeneveld roos.groeneveld@wur.nl Mark van Assem mark@cs.vu.nl Marielle Timmer marielle.timmer@wur.nl Mari Wigham mari.wigham@wur.nl	WUR VU	Interview-based knowledge acquisition process can be used to obtain multi-domain, multi-expert task-specific knowledge. Interview-based methods have some disadvantages: they take a lot of time; it is difficult for the domain expert to give a full overview of his knowledge; a model is typically created from scratch, even if reusable sources exist. We propose to add a reuse-based ontology construction component (ROC) to our knowledge acquisition toolkit that gives domain engineers a prominent role in the acquisition process. This way, we aim to make the knowledge acquisition process more efficient. The ROC-component supports domain experts in identifying relevant knowledge, for ROC incorporates a prompting process that offers the expert terms associated with earlier identified terms. This helps the domain expert to better cover all aspects of his knowledge. For the prompting process, existing knowledge sources are used. Thereby, we benefit from already existing knowledge sources.
Stefano Bocconi stefano.bocconi@gmail.com	VU	Due to the OKKAM project I am currently working on, from the Semantic Web research area I am particularly interested in several issues regarding entities, namely: the concept of context-free entities (is identity between two entities a matter of context?), entity-based data integration (how to integrate different knowledge sources once it is known that they refer to the same entity), and entity identifiers (how to guarantee persistency of identifiers when new knowledge causes the view on a particular entity to change). The domain where I want to apply this research is scientific publishing (particularly in Biology) and news publishing.
Borys Omelayenko b.omelayenko@few.vu.nl	VU	AnnoCultor is a tool for porting XML datasets or databases to RDF-based Semantic Web applications. We specially developed AnnoCultor to convert collections and vocabularies of Louvre, Rijksmuseum, Tropenmuseum, and many others. It is quite likely that a developer making a converter for another institution would run into exactly the same problems as we faced and solved in this work. Using AnnoCultor such a developer would gain a conversion infrastructure, a rich set of structural conversion rules, and rules for term lookup in vocabularies for data enrichment.
Guus Schreiber schreiber@cs.vu.nl Dan Brickley danbri@danbri.org	VU	NoTube is a new European project applying Semantic Web technology to explore approaches that combine television and the Web. This talk will outline the approach we're taking to describing content and users, some of the challenges and opportunities, and why it's a project worth watching.
Chide Groenouwe chide@few.vu.nl	VU	Towards a Constitution Based Game for Fostering Fluency in "SemanticWeb Writing": the Semantic Web (SW) is still far from realising its full potential, partly because it is still lacking enough high quality SW representations of information. We argue that a step in the right direction is fostering peoples capability to fluently create high quality SW representations of the information they generate during problem solving processes. To foster such a capability, we propose a game in which teams compete in creating the best translations of texts into SW representations. Although playing the game is in itself already a way to foster such a capability, we moreover pursue learning from the game which are the most successful translation strategies (embodied by "constitutions") so that they can also be used by people outside a game setting.

Ronald Siebes rm.siebes@few.vu.nl	VU	RDF management inspired by the UNIX GIT versioning system: on the E-culture server more and more vocabularies are stored, made by different people and becoming more complex to handle. Some problems were always there and some start to arise: what namespace to choose for each vocabulary, how to deal with scalability, what if somebody else makes a vocabulary using (stealing) the same namespace, how to do version management, how to prevent any central mechanism to play God, how to keep the system robust by redundancy, how to make our management approach be generic enough that it not only works for the VU group, how to keep our management approach not to interfere with the standards, keeping it backward compatible?
Shenghui Wang swang@few.vu.nl Antoine Isaac aisaac@few.vu.nl Lourens van der Meij lourens@few.vu.nl Stefan Schlobach schlobac@few.vu.nl Balthasar Schopman baschopm@few.vu.nl	VU	In the STITCH project, we investigate deployment of automatic ontology alignment technology which can really be benefit to Cultural Heritage contexts, such as the Dutch National Library. To this end, we have investigated several application scenarios, such as re-indexing of books, and confronted the results of various alignment techniques to them. We confirm that different scenarios will work better with alignment that focuses on specific dimensions of concepts to align, such as their extensions. We have developed and applied further extensional alignment methods in a European context (TELplus project). These efforts have made obvious that still many obstacles remain for properly deploying alignment techniques in realistic context, such as the scale of vocabularies and collections, and multilinguism.
Eyal Oren eyal@cs.vu.nl	VU	Within the LarKC project we work on efficient techniques for analysing and processing large datasets. Many Semantic Web problems are difficult to solve through common divide-and-conquer strategies, since they are hard to partition. We present MaRVIN, a parallel and distributed platform for processing large amounts of RDF data, on a network of loosely-coupled peers. We present our *divide-conquer-swap* strategy and show that this model converges towards completeness. We evaluate performance, scalability, load balancing and efficiency of our system. See http://larkc.eu/marvin.

Frank van Harmelen
frankh@few.vu.nl

Christophe Guéret
cgueret@few.vu.nl

Zhisheng Huang
huang@cs.vu.nl

Willem Robert van Hage
wrvhage@few.vu.nl

Szymon Klarman
sklarman@few.vu.nl

Ruud Stegers
ruud@stegers.info

Laura Hollink
laurah@cs.vu.nl

Anna Tordai
atordai@few.vu.nl

Jan Wielemaker
j.wielemaker@cs.vu.nl