D 1.3.6

Ontology Outreach to Industry

-Report on the OOA Activities

Contributors:

Mustafa Jarrar (Vrije Universiteit Brussel)

Diana Maynard, (University of Sheffield)

Jeroen Hoppenbrouwers (Vrije Universiteit Brussel)

Zhisheng Huang (Vrije University of Amsterdam)

EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB

Deliverable D1.3.6 (WP1.3)

Abstract. This document reports on the activities of the Ontology Outreach Advisory (OOA), a legal body recently founded by KnowledgeWeb consortium members to assure extension of KnowledgeWeb results and activities into the future. The OOA has been legally registered according to the Belgian law as an international non-profit association (VzW). Currently, there are 28 registered members, and 16 organisations expressed their interest in becoming a member. The OOA mission is to develop strategies for ontology recommendation and standardisation, and promote the ontology technology to industry. The OOA is implementing this mission specially in two domain sectors: Human resources and eHealth. *Several activities have been carried out by the OOA in these two domains, including: the successful OOA-HR kick-off workshop, a new OOA-HR roadmap, in-reaching and out-reaching workshops, the HR summit, contribution to HR standards, a repository of eHealth uses cases, quality guidelines for ontology authoring, OnToContent06, and OntoContent'07.*

Document Identifier:	KWEB/2007/D1.3.6/v0021
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	Ver 0021
Date:	August 8, 2007
State:	Final
Distribution:	KnowledgeWeb Consortium Only

Knowledge Web Consortium

This document is part of a research project funded by the IST Programme of the Commission of the European Communities as project number IST-2004-507482.

University of Innsbruck (UIBK) – Coordinator

Institute of Computer Science,

Technikerstrasse 13

A-6020 Innsbruck

Austria

Contact person: Dieter Fensel

E-mail address: dieter.fensel@uibk.ac.at

École Polythechnique Fédérale de Lausanne (EPFL)

Computer Science Department

Swiss Federal Institute of Technology

IN (Ecublens), CH-1015 Lausanne.

Switzerland

Contact person: Boi Faltings

E-mail address: boi.faltings@epfl.ch

France Telecom (FT)

4 Rue du Clos Courtel

35512 Cesson Sévigné

France. PO Box 91226

Contact person : Alain Leger

E-mail address: alain.leger@rd.francetelecom.com

Freie Universität Berlin (FU Berlin)

Takustrasse, 9

14195 Berlin

Germany

Contact person: Robert Tolksdorf

E-mail address: tolk@inf.fu-berlin.de

Free University of Bozen-Bolzano (FUB)

Piazza Domenicani 3

39100 Bolzano

Italy

Contact person: Enrico Franconi

E-mail address: franconi@inf.unibz.it

Institut National de Recherche en Informatique et en Automatique (INRIA)

ZIRST - 655 avenue de l'Europe - Montbonnot Saint Martin

38334 Saint-Ismier

France

Contact person: Jérôme Euzenat

E-mail address: Jerome.Euzenat@inrialpes.fr

Centre for Research and Technology Hellas / Informatics and Telematics Institute (ITI-CERTH)

1st km Thermi – Panorama road

57001 Thermi-Thessaloniki

Greece. Po Box 361

Contact person: Michael G. Strintzis

E-mail address: strintzi@iti.gr

Learning Lab Lower Saxony (L3S)

Expo Plaza 1

30539 Hannover

Germany

Contact person: Wolfgang Nejdl

E-mail address: nejdl@learninglab.de

National University of Ireland Galway (NUIG)

The Open University (OU)

National University of Ireland
Science and Technology Building
University Road
Galway
Ireland
Contact person: Christoph Bussler
E-mail address: chris.bussler@deri.ie

Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA
United Kingdom.
Contact person: Enrico Motta
E-mail address: e.motta@open.ac.uk

Universidad Politécnica de Madrid (UPM)

Campus de Montegancedo sn

28660 Boadilla del Monte

Spain

Contact person: Asunción Gómez Pérez

E-mail address: asun@fi.upm.es

University of Karlsruhe (UKARL)

Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB

Universität Karlsruhe

D-76128 Karlsruhe

Germany

Contact person: Rudi Studer

E-mail address: studer@aifb.uni-karlsruhe.de

University of Liverpool (UniLiv)

Chadwick Building, Peach Street

L697ZF Liverpool

United Kingdom

Contact person: Michael Wooldridge

E-mail address: M.J.Wooldridge@csc.liv.ac.uk

University of Manchester (UoM)

Room 2.32. Kilburn Building, Department of Computer Science, University of Manchester, Oxford Road

Manchester, M13 9PL

United Kingdom

Contact person: Carole Goble

E-mail address: carole@cs.man.ac.uk

University of Sheffield (USFD)

Regent Court, 211 Portobello street

S14DP Sheffield

United Kingdom
Contact person: Hamish Cunningham

E-mail address: hamish@dcs.shef.ac.uk

University of Trento (UniTn)

Via Sommarive 14

38050 Trento

Italy

Contact person: Fausto Giunchiglia

E-mail address: fausto@dit.unitn.it

Vrije Universiteit Amsterdam (VUA)

De Boelelaan 1081a

1081HV. Amsterdam

The Netherlands

Contact person: Frank van Harmelen

E-mail address: Frank.van.Harmelen@cs.vu.nl

Vrije Universiteit Brussel (VUB)

Pleinlaan 2, Building G10

1050 Brussels

Belgium

Contact person: Robert Meersman

E-mail address: robert.meersman@vub.ac.be

University of Aberdeen (UNIABDN)

King’s College

AB24 3FX Aberdeen

United Kingdom

Contact person: Jeff Pan

E-mail address: jpan@csd.abdn.ac.uk

Document Change Log

Version	Date	Author	Changes
0001	16-04-2007	Mustafa Jarrar	1st draft
0009	19-06-2007	Jeroen Hoppenbrouwers	Further elaborations, ready for review
0010 0011	06-07-2007 7-07-2007	Diana Maynard, Jeroen Hopp. Zhisheng Huang	English language check; restructuring. Chapter 4
0011	10-07-2007	Jeroen Hoppenbrouwers	edits.
0012	17-07-2007	Mustafa Jarrar	Quality assurance + improvements
0013	1-08-2007	Zhisheng Huang	Quality Control
0020	3-08-2007	Mustafa Jarrar	Pre-Final edits
0021	8-08-2007	Mustafa Jarrar	Final edits + Submission

Executive Summary

This document reports on the activities of the Ontology Outreach Advisory (OOA), a legal body recently founded by KnowledgeWeb consortium members to assure extension of KnowledgeWeb results and activities into the future. Chapter 1 gives an overview of the **founding activities**:

- The *OOA has been legally registered* according to the Belgian law as an international non-profit association (VzW). All administration and internal procedures, memberships, mission, and workplan are prepared. Currently, we have 28 registered members and 16 prospective members.
- A new *OOA website* was launched www.ontology-advisory.org

Chapter 2 and chapter 3 present the **OOA activities in the HR domain**:

- *Out-reaching and in-reaching workshops*. Several top HR experts from industry had been invited to attend and interact with several K'Web general assemblies. (See section 2.1.1 and 2.1.2)
- *The OOA-HR Kick-off in Oxford*. This was a successful and strategic outreaching event: 15 demo and position presentations were given, some by leading world experts in the field; 50 participants from industry; 17 applications to join the OOA, and most importantly the production of the OOA-HR roadmap. (See section 2.1.3)
- *OnToContent'06 and OnToContent'07.* This is the OOA annual scientific event. The idea is to give special attention (within the research community) to ontology content issues in the HR and eHealth domain. (See section 2.1.4)
- *The HR summit*. The OOA has been invited to co-organize an HR summit (Oct.2007) in cooperation with key organizations in the HR domains (ePortfolio, EIfEL, HR-XML, TENCompetence, IEEE-LTSC, and Prolix). The summit aims at both an in-depth and broad exploration of the HR issues, 700 participants from industry are expected to attend. The OOA has been given the responsibility of the "HR semantic interoperability" during this summit. (See section 2.1.5)
- *Contribution to standards*. Many OOA-HR members are actively contributing the IEEE-LTSC conceptual modeling framework, which is being standardized. The initial proposal of this framework (incorporating semantics) has been submitted by the OOA-HR chair. Furthermore, the OOA has been invited to introduce a semantic layer to underpin the HR-XML standard. (See section 2.1.6)
- *The OOA-HR roadmap* (called "Semantic challenges and opportunities in the HR domain"). This is the first OOA product/publication, to be used as a reference for both the HR and the ontology communities. Also, to be used as a roadmap for the OOA itself, within the HR domain. (See chapter 3)

Chapter 4 and chapter 5 present the **OOA activities in the eHealth domain**:

- *Collection and promotion of eHealth use cases*. This activity (led by the Vrije University of Amsterdam) aims to collect successful use cases and promote them to industry. The collected use cases are published at the OOA website *regularly*, and promoted at the OOA annual industrial events. There are 7 use cases published at the moment, and several others are underway (in cooperation with WP1.1). (See chapter 4)
- *The "OOA Quality Guidelines for ontology authoring"*. This is an ongoing activity in close collaboration with NIST, ECOR, and NCOR. The idea is to identify and *recommend* a set of guidelines that contribute directly to ontology quality, or indirectly by pursuing ontology reusability and adoption. The target users of these guidelines are tool developers, ontology modellers, and the OOA itself for future ontology evaluation. (See chapter 5)
- *Other activities*: Mapping the syntax of the OBO (Open Biomedical Ontologies) into OWL, and mapping the ORM (Object Role Modeling) into OWL. This is a joint effort between the OOA and the university of Manchester. The idea is to enable ontologies represented in OBO and ORM to be converted into OWL and to be benefit from its reasoning services. (This activity is not reported in this deliverable, but can found in deliverable D1.3.8 called "Accessible syntax for OWL semantics").

Table of Contents

1 Overview of the OOA activities

The Ontology Outreach Advisory (OOA) has been playing an active role in outreach to industry activities within KnowledgeWeb since 2004, and has been officially registered as an independent legal entity in January 2007 in Belgium. Although it is legally independent, the legal seat of the OOA is currently located at STARLab at the Vrije Universiteit Brussel. The OOA has already organized many outreaching activities, particularly in the eHealth and the HR domains. This chapter reports only the foundation and administration activities. The OOA activities in the HR and the eHealth domains are presented in the other chapters.

1.1 The OOA Legal Foundation


Bijlagen bij het Belgisch Staatsblad - 23/01/2007 - Annexes du Moniteur belge

"ONTOLOGY OUTREACH ADVISORY"
Vereniging zonder Winstoogmerk
1050 Brussel, Pleinlaan 2, geb. G-10
0886 345 517

BRUSSEL
12 -01- 2007

07013979

After many preparations, for which we refer to the previous KnowledgeWeb deliverable D1.3.3, the Ontology Outreach Advisory was legally founded in January 2007 as a not-for-profit association under Belgian law (VZW/ASBL). This entails a few legal requirements, of which the relevant are summarised here:

- The purpose of the association is not a commercial activity,
- There is no will from the members to take some material profit from their participation to the association,
- There are at least three members (being Belgian or foreigners),
- In case of liquidation, the remaining balance has to be used for a not-for-profit goal.

1.1.1 The OOA Mission and Objectives

From the Charter, it is the OOA mission:

To develop strategies for ontology recommendation and standardisation,

and promote the ontology technology to industry.

This mission has been expanded into the following OOA objectives:

- To promote science and research on ontology engineering, with a particular focus on ontology content;

- To develop and promote quality (and methodological) guidelines for ontology content and tools, including evaluation and recommendation strategies;
- To evaluate, recommend, and/or standardise ontology content and tools;
- To provide publication infrastructure for ontology content.
- To become a representative forum for experts within this field;
- To organise chapter committees and working groups on subjects related to its purpose;
- To distribute, and make available, knowledge about ontologies through publications, the organisation of seminars, courses and conferences, and any other means that serve this purpose.
- To collaborate with scientific and other institutions, organisations and societies, as well as with industrial companies, national and international bodies with similar or related purpose.

1.1.2 OOA Board

The OOA has three board members at present. They are responsible for the daily business and coordinate all OOA activities with OOA members.

Prof. Dr. Robert Meersman: The president of the OOA and the director of STARLab. His research interests are ontology engineering, database semantics, interoperability, and uses of databases in applications such as enterprise knowledge management and the Semantic Web. Prof. Meersman was appointed as Full Professor at VUB in 1995. Earlier positions include the University of Antwerp (UIA, 1975-78), Control Data Corp. (Data Management Lab, Brussels, Belgium, 1978-83). Held chairs and founded the InfoLabs at University of Limburg (Belgium, 1983-86) and at Tilburg University (The Netherlands, 1986-95). Member and Past Chairman of the IFIP WG2.6 and IFIP TC 12, current Chairman of TC 2 (Software Theory and Practice).

Mr. Theo Mensen: A board member of the OOA, and currently senior advisor of both the CWI (the Public Employment Services of The Netherlands) and AKA (the Labour Market Information and Policies organisation, The Netherlands). His main interests includes Matching on Competencies, European HR-XML standards, EUROPASS portfolio and the exchange of data on jobs and CVs, transparancy related to Skills and Mobility, and Transitions in Education and Labour Market.

Dr. Mustafa Jarrar: The secretary and board member of the OOA. Dr. Jarrar earned his PhD degree at Vrije Universiteit Brussel in 2005. He currently is a senior researcher at STARLab and has served as Manager/workpackage leader in more than 8 European research projects. His main research interests include: ontology engineering, logic, lexical semantics and multilinguality, e-learning, and knowledge management. He is a full member of the IFIP2.6 and the IFIP2.12, the IEEE Learning Standards Committee, and the CEN/ISSS ICT Skills and Curricula. Dr. Jarrar has chaired more than 10 international workshops, and has been a PC member of over 50 international conferences.

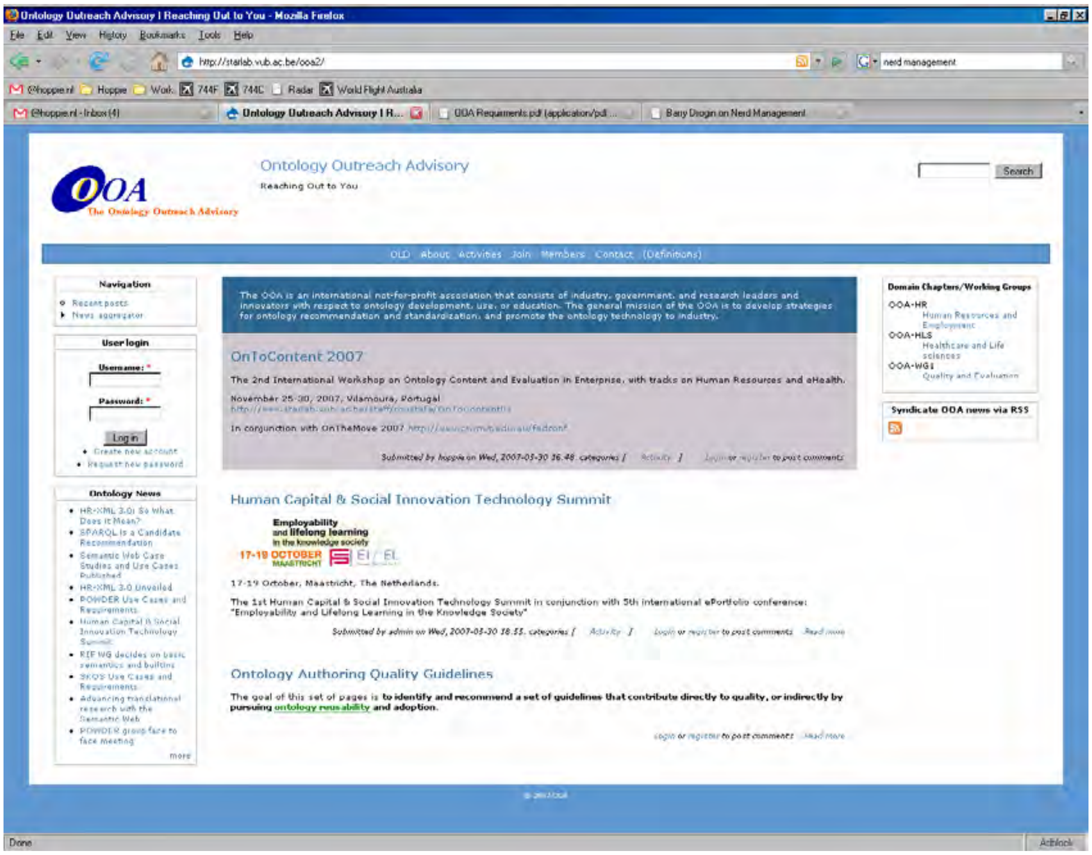
For a list of current OOA members and their contributions, please see Section 1.3.

1.1.3 OOA Initial Awareness Activities

In line with the legal foundation, a domain name was registered (ontology-advisory.org) and a few neighbouring domains were registered but not activated, to prevent domain hijacking (ontology-advisory.net, ontology-advisory.com). Further domains were not considered important enough to register at the moment.

The OOA web site has been set up at **www.ontology-advisory.org,** which contains a first representation on the web; this address should be considered the canonical web address of the OOA.

After the simple static web site had been put online, additional activities were deployed to provide the infrastructure for a real OOA community. An OpenSource content management system was selected and put into operation, initially populated with the static text of the original web site. *At the time of this writing, the new web site is in partial production, but has not taken over the static web site yet. If you read this and the canonical web address still leads you to the static web site pictured above, please click on the* ***Quality Guidelines*** *link which will lead you to the new dynamic site.*



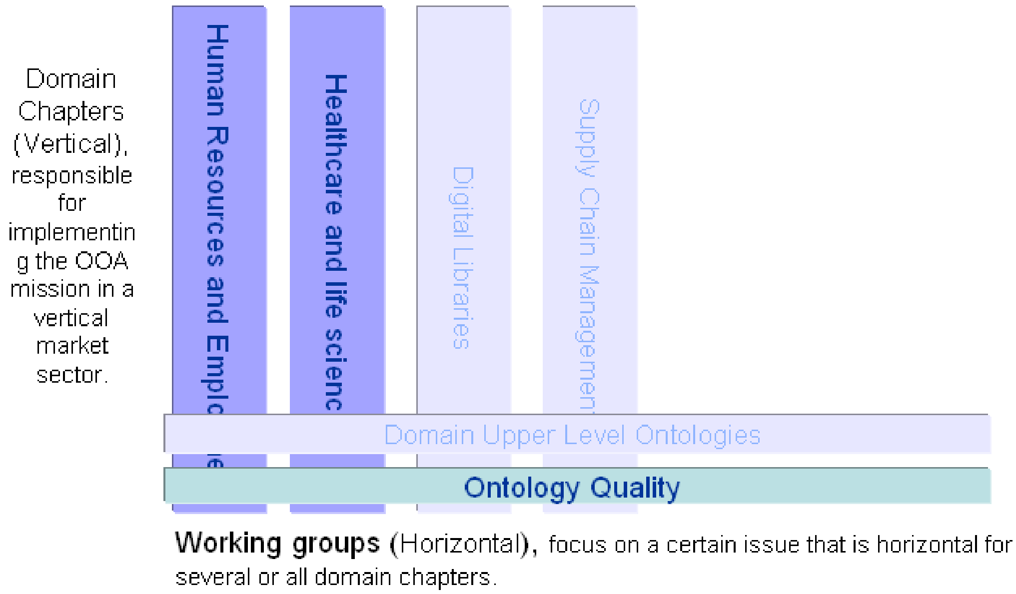
With the OOA ambitions reaching far beyond being just a dissemination organisation, the web site should reflect an active, committed community which uses the web site as a primary means of communication, both within the community and to the outside world. To support such as web site, a content management system is required. Good experiences and already existing local installations of Drupal (www.drupal.org) led us to this platform. At the moment of writing, the site looks as the image shows. Note that many page objects can and will change over time, as content is updated. The OOA might also commission a specific OOA custom style in the future.

Essential for a dynamic web site is that there is a community of people feeding it. As the OOA cannot rely on a permanent staff of editors, care has been taken that content maintenance does not require any particular skill beyond common practices. With this in mind, a light selection process to allow people write access to the web site has been designed, with a small group of stakeholders keeping an eye on the content as it gets posted. The OOA web site is not a fully open system which allows ‘anybody’ to post, as it has to represent at least a sizeable majority of OOA members.

Active search for relevant news and announcements to be posted on the OOA web site will become a routine operation of the OOA members. The web site should facilitate and simpify the process of announcing something to such an extent that sending bulk mail to a distribution list becomes less attractive. In particular, the archiving and rich annotation/comment facilities should be key here. Combined with the RSS features, which are already operational, all infrastructure is in place to start featuring the OOA web site on well-known portal sites in the relevant domains and industries.

1.2 The operational plan

The OOA, as not-for-profit association, will have just a (very) small staff to provide operations support. For this reason the operational plan has been designed to maximise the available resources from the OOA members. Where possible, existing activities have been bundled and reorganised into a matrix organisation which caters for all participants without needing many additional resources.



Vertically, several Domain Chapters provide for domain-specific activities which are of interest to well-defined subgroups of OOA members. This structure assures that sufficient critical mass will be available to reach the OOA's targets while not drawing too much on external resources. For the moment, Domain Chapters are in place for the Human Resources and Employment domain and the Healthcare and Life Sciences domain. Additional Chapters in Digital Libraries and Supply Chain Management are foreseen.

Horizontally, broad areas of interest are projected which appeal to much wider audiences. Upper Level Ontologies in general (including those for specific domains) and Ontology Quality Guidelines are the first topics being addressed. It is the intention that the overlaps between the horizontal and vertical fields (the matrix cells) focus on applications of the broad area of interest to the specific domains.

This matrix structure provides both sufficient resources and focus for industrial partners to join and stay interested, while the OOA can also push the edge in broader problem areas which otherwise would not receive sufficient attention from industry.

1.3 OOA Members and Membership

As an Association, the OOA has members. Membership is subject to several considerations. To start out, many members were solicited from the existing industrial and research community forming around KnowledgeWeb.

1.3.1 OOA Founding Members

There 25 officially registered members form the OOA General Assembly until now (see the table below). In addition, there are 16 other organisations that have declared their intentions to become a member but have not formally signed the membership document yet The list of 16 members (not shown the table below) includes: LOA-CNR, DERI- Galway, The Free University of Amsterdam, University of Aberdeen, Asemantics, HR-XML, EIFEL, NCOR, NIST OBO, BioVista, and many others.

In this foundation phase, most organizations are formally member through one single representative person, as this considerably sped up the registration process. It is the intention of the OOA to convert these personal commitments to organisation commitments as soon as practical.

	Robert Meersman, Mustafa Jarrar Vrije Universiteit Brussel, Belgium, Belgium		Jean-Claude Xheneumont, Le Forem, Belgium
	Werner Ceusters, SUNY at Buffalo NY, USA		Clementina Marinoni, Fondazione Politecnico di Milano, Italy
	Luk vervenne, Synergetics, Belgium		Francky Trichet, LINA - University of Nantes, France
	Theo Mensen, CWI public employment services, Holland		Claude Ostyn, IEEE-LTSC, USA
	Marti Malcik, Ostravska univerzita		Christophe Roche, University of savoie, France
	Michael Brown, Skillsnet Enterprises, USA		Keith Baker, University of Reading, UK

	Marco Ronchetti, University of Trento, Italy		Angelika Buehler, University of the Arts Berlin, Germany
	Jaime Bonilla, Tecnológico de Monterrey, Mexico		Ernesto Damiani, University of Milan, Italy
	Mathias Brochhausen, Institute for Formal Ontology and Medical Information Science, Saarland University, Germany		Andreas Schmidt, FZI Research Center for Information Technologies, Germany
	Cornel Mueller, OE GmbH, Switzerland		Diana Maynard, University of Sheffield, UK
	Margarita Perez, Menon network, Belgium		Simon White, Catalysoft, UK
	Enrico Franconi, Free University of Bozen-Bolzano, Italy		Simon Grant, UK
			Rob Bieleveld, Holland

1.3.2 Membership Fees

The proposed OOA membership fees per annum are:

- €3,000 for a large organisation
- €1,000 for a medium-size organisation
- €200 for a small organisation
- €30 for individuals

The membership fees were waived for all members in 2007. The fee structure has been designed to reflect current practices in industry.

1.3.3 Membership benefits

OOA members may play a prominent role in their domain of interest as far as the direction and acceptance of ontology technology is concerned. Their recommendations, through the OOA, influence the domain, and may generate new business and customers. These recommendations include explicit proposals for global standards via standardisation bodies such as ISO and IEEE (which is a member). OOA members are the first to know about new standards and developments, and have full access to the OOA developments, voting, etc.

It is expected that the OOA will become an important organisation to shape the Human Capital & Social Innovation Technology Summit. This provides OOA members with access to one of the largest conference clusters in the field of E-Learning and HR practice.

2 The OOA-HR chapter

The OOA plans to have several Chapters, each of which addresses the specific issues of a single domain. Human Resources and Employment has been a crucial domain for the industry for a while, and the OOA recognises this by having a HR Chapter from the start.

The mission of the OOA HR Chapter:

To promote the semantics technology into HR/e-learning

standards and applications.

The domain of Human Resources is characterised, from a knowledge perspective, by a few aspects that together create a particularly difficult automation problem.

HR tries hard to quantify and qualify the common knowledge that people must have to have meaningful conversations about human resources. The common language of jobs, function descriptions, and vacancies is well worked out. Task-level descriptions, where multiple tasks combine into one job or vacancy, are becoming standard as well. However, below the task level, there is the competency level which at least for now looks like the level of the smallest useful building brick. Further division of competencies seems pointless right now. If we could work with competencies all over the HR domain, it would be a natural granularity to describe, plan, assess, and train human resources and exchange these details all over the industry.

The current focus topics of the OOA-HR chapter include:

- Semantic interoperability
- HR upper level concepts
- Semantics of HR-XML
- Semantic metadata for HR applications
- Semantics in job matching
- Semantics in learning technologies
- Modelling and representation of: Jobs, CVs, Competencies, Skills, Employees, People, Organizations, Social Events, etc.
- Multilinguality in human resources ontologies
- Best practice and semantic patterns in ontology modeling and evaluation.

The current understanding of the problem of semantics in HR is fragmented and only partial solutions exist. How HR applications could benefit from research and development in the field of semantics is generally not understood. The HR community in general and the HR-XML community of practice in particular is not familiar with this field of research, while the ontology research and engineering community neither fully appreciates the opportunities offered by the HR community, nor fully understands its requirements and specificities.

The OOA will bring the main stakeholders of both the HR and the ontology research and engineering communities together to discuss and reach a common understanding of the most critical issues that need to be solved in the real world of employment, and the potential solutions that semantics could provide. Finding common ground, understanding common concepts and issues, and documenting these in a formal way that can be automated, is the first step towards a broad semantic underpinning of the whole HR domain.

2.1 OOA activities in the HR domain

Until now, the semantic web community and the Human Resources (HR) communities have been largely disparate. While it seems clear that the HR domain is well aware of the existence of ontologies, it is not apparent that they have the expertise to make best use of them, and so tend to stick to tried and tested methods of storing information such as relational databases and high-level simplified tree structures.

Unlike the mission of the OOA in the eHealth community, where ontologies are well-understood, the mission of the OOA in the HR domain is focused on promoting the ontology technology to the HR stakeholders. This implies that our activities in the HR domain are more concerned with awareness, networking, roadmapping, and basic demonstrations.

In this chapter we present the past and future activities in the HR domain. The next chapter (chapter 3) is dedicated to the OOA-HR roadmap, which has been built (and is being used) by the OOA-HR community.

2.1.1 OOA Meeting in Crete

A one-day joint meeting with industry was organized during the Knowledge Web general assembly in Crete (June, 2005). Three top HR experts -Theo Mensen (from the Dutch employment agency), Luk Vervenne (Synergetics.be, HR-XML competency), and Claude Ostyn (the IEEE-LTSC chair) - were invited to this meeting to discuss and give presentations about the need for semantics in the HR domain. Knowledge Web partners presented their experience in previous/ongoing research projects related to HR. During the discussion panel, the HR experts expressed the need for a semantic layer topping the HR XML standards, and the need for scalable solutions. The OOA HR chapter was initiated during this meeting.

2.1.2 OOA Seminar in Trento

Claude Ostyn (a worldwide HR expert) was again invited to give a seminar during the Knowledge Web general assembly in Trento (January, 2006). Claude presented very well the main challenges of modelling, representing, and using competence ontologies. The seminar ended with a lively discussion with all WP leaders, area managers, and other Kweb partners. It was clear how much the subjectivity and the dynamics of competency definitions are difficult to model, represent, and exchange.

2.1.3 OOA-HR Kick-off in Oxford

This was a successful and strategic outreaching event.

The OOA-HR inaugural workshop was held on 11 October 2006 as part of the e-Portfolio conference[1], the main industrial HR conference in Europe, with 300 participants, mainly technology uptakers and provider companies. This was the ideal territory for the workshop, whose aim was to bring together people from the HR and ontology communities, in order to discuss semantic challenges and opportunities in the HR domain.

The OOA-HR workshop had 50 participants and consisted of 15 demo and position presentations, some by leading world experts in the field. More details about the workshop can be found on the OOA website. The workshop included a general introduction to the OOA, sessions on Semantics and Competency Frameworks, Interoperability Challenges, a Demo session, and a Discussion session. A participant survey was also conducted during the workshop, focusing on the attitudes of participants towards the needs and challenges in the community. Combined with the discussion panel session, this led to strategic collaborations with the HR community.

The outcomes of the workshop were most encouraging, with a number of objectives met. One of the main aims was to link the two communities together, and this was achieved via much discussion between HR and ontology experts, between industry and academia, and between technology uptakers and providers, leading to the beginnings of a real OOA-HR community, as demonstrated by the 17 applications to join the OOA as a direct result of the meeting. Thanks to its success and to the interest of so many participants, the workshop was given an invitation to participate again the following year in ePortfolio 2007 (see the planned activities below). Also, interest was shown by various participants in putting together a new international/European proposal for a new project, which is currently underway.

Most importantly, a comprehensive document has been produced as a direct result of the workshop: *Semantic Challenges and Opportunities in the Human Resources Domain (we call it the OOA-HR*

1 http://www.eife-l.org/news/ep2006/

roadmap), with contributions from most of the participants as well as some leading researchers in the field. This roadmap (presented in chapter 3) is the first OOA product. It is being used as a reference for both the HR and the ontology communities, as well as. as a roadmap for the OOA itself.

2.1.4 OnToContent Workshop

Ontology content and evaluation in Enterprise
(With two tracks on Human Resources and eHealth)

This is the OOA annual scientific event. The goal of this workshop is to give a special attention (within the research community) to ontology content issues in two industrial sectors: human resources and employment, and healthcare and life sciences.

OnToContent 2006.

This workshop was organized as part of the federated conference OnTheMove 2006, October 2006 in Montpellier. We received 35 submissions, and we accepted only 12 papers (acceptance rate 34%) . The quality and the OOA-relevancy of the papers were very high indeed. For the program of this workshop please visit at the OOA website. The most important outcome of this a collaboration with the University of Milan, Italy (Prof. Ernesto Damiani'), and with the FZI Research Center for Information Technologies, Germany, (Andreas Schmidt). Both institutions became active members in the OOA-HR chapter.

OnToContent 2007

For 2007, the OnToContent workshop will be repeated as part of the same federated conference, November 2007, in Algarve, Portugal. For more information about this workshop, please visit the OOA website.

2.1.5 Human Capital & Social Innovation Technology Summit

Following the success of the OOA kick-off workshop (see 2.13) and production of the OOA-HR roadmap (see chapter 3), the OOA received an invitation from the ePortfolio conference organize another workshop on “HR semantic interoperability”. The workshop is turned recently into an HR summit[2] where many organizations are involved: EIfEL (European Institute for E-Learning), HR-XML, Prolix (align learning with business processes), TENCompetence (infrastructures to support individuals, groups and organisations in lifelong competence development), the IEEE-LTSC working group meeting, and the OOA. It is expect that this summit will be attended by 700 participants. The summit is planned 16-19 October, Maastricht, The Netherlands.

The summit aims at an in-depth and broad exploration of the societal, organisational and individual issues linked to digital technologies - and result in real outcomes. Past events have resulted in the creation of national and international networks, contribution to policies, the launch of the ePortfolio for All campaign, progress in interoperability through the organisation of Plugfests, transnational projects and numerous publications.

The role of the OOA in this summit is not only to present and disseminate the Kweb results and use cases to the HR community (during its workshop), but also the OOA is invited particularly to contribute to the HR-XML schema that is being developed by the HR-XML members. The OOA is give the responsibility to introduce a semantic layer underpinning this schema.

2.1.6 Contribution to the IEEE-LTSC and the HR-XML standards

The OOA is actively involved in the IEEE-LTSC working group and the HR-XML Competency group. The main OOA-HR members who are active in these activities are: Luk vervenne (the OOA-HR chair), Claude Ostyn, Andreas Schmidt, Clementina Marinoni, and Mustafa Jarrar.

The IEEE-LTSC group is currently standardizing a Competency Modeling Framework. The OOA-HR members (who are involved in this group) have submitted a proposal of this framework that incorporate semantics, have contributed with precise definitions of some important notions such as “competency”,

[2] http://events.eife-l.org/HCSIT2007/HCSIT2007

“evidence”, “task”, etc. We expect the standard to finalized within 6 months. The unfortunate death of our most active colleague Claude Ostyn may delay the release of this standard.

HR-XML is developing and standardizing an XML schema for representing and exchanging HR records (CVs, job offers, digital identities, competencies, etc.). The need for semantics was (1 year ago) not realized at all by the HR-XML members. Thanks to Luk Vervenne who contributed with several presentations showing that a semantic layer is key requirement to enable semantic interoperability. As result, the HR-XML consortium has invited the OOA this year and give the responsibility to introduce a semantic layer underpinning the HR-XML schema. The OOA (see 2.1.5) shall be presenting and discussing this layer at the HR-XML meeting during the summit.

3 Semantic challenges and opportunities in the Human Resources domain

This chapter presents the first OOA product/publication: "The OOA-HR Roadmap".

Purpose: an OOA publication, to be used as a reference for both the HR and the ontology communities. Also, to be used as a roadmap for the OOA itself, within the HR domain.

Background: During the discussion panel at the OOA kick-off workshop, which was attended by more than 50 participants (see section 2.1.3), the need for this roadmap was realized. It was obvious that the current understanding of the problem of semantics in HR is fragmented and only partial solutions exist. People from both the HR and the ontology communities speak different languages, have different understandings, and are not aware of existing solutions.

Contributors: Many people were actively contributing to this roadmap, including 12 non-Kweb members. In the following we list most active:

- *Catalysoft, UK*
- *Christine Kunzmann, Germany*
- *Eifel.org*
- *HR-XML Europe*
- *Fondazione Politecnico di Milano, Italy*
- *FZI, Germany*
- *Skillsnet Enterprises, USA*
- *Synergetics, Belgium*
- *University of Milan, Italy*
- *University of Nantes, France*
- *University of Sheffield, UK*
- *University of Reading, UK*
- *Université de Savoie, France*
- *University of Trento, Italy*
- *Vrije Universiteit Brussel, Belgium*

Version: Ver 1.4

Publication: OOA website, This version shall be presented and disseminated during the HR summit (16-19 October, Maastricht, The Netherlands).

Further reading: For a roadmap about the Semantic Web Technology in general please refer to: http://wiki.ontoworld.org/wiki/Category:Topic_Knowledgeweb_technology_roadmap

3.1 Introduction and Motivation

3.1.1 The HR Domain

Knowledge based automation in the domain of Human Resources faces some particularly daunting challenges. Information technology scientists and practitioners involved in the Human Resources domain have to quantify and qualify the common knowledge that underlies meaningful conversations about human resources. They must also implement the operational processes and data stores that exploit and capture that knowledge to further the enterprise's strategic objectives. The common language used to describe jobs, functional roles and staff vacancies is generally well understood and formalised, at least within specific enterprise domains or regional scopes. Models and emerging standards for the description of tasks and responsibilities have been used with various degrees of success. Various standardisation efforts also support capturing the combination of tasks and responsibilities that make up a typical job description or job vacancy.

However, performance in the accomplishment of tasks and the fulfilment of a role requires competence. While competence, as a highly individual and context-specific quality, is very difficult to define in operational terms, there is general agreement that competence is to a large extent the product of a number of specific competencies. Such competencies must in turn be specified at levels of granularity that support specific operational processes such as targeted assessments, staffing, training or performance support.

What makes the situation even more difficult is the level of controversy over the term "competency" itself, and the inordinate amount of time that is still being wasted in battles over terminology. Many people use "competency" as a generic term that encompasses skills, knowledge, attitude and abilities, and even facets of a given competency such as cognitive, psychomotor and affective facets. Most job functions require a mix of those. On the other hand, some other people hold that those aspects of competency should not be included in the definition of the term. Often, what is labelled as "competency" by one group is labelled "standard", "behavioural indicator", or any number of other terms by other groups.

A pragmatic approach might leave the terminology battles to philosophers and focus on what can be made to work. If, for practical purposes, we can agree to use the term "competency" in its broader, more encompassing sense, the problem of automation support for competency related information and processes the Human Resources domain appears to be quite tractable. One useful approach is to consider a loose framework within which coexist several information and processing models. Used together, these models support the operational requirements of Human Resources. The framework must enable automation when appropriate, because only automation can help make sense of the myriads of competencies required in the operation of an enterprise, something that no single individual can hope to achieve. Automation also provides the speed of data exchanges and processing that are required for effective implementation of many competency related processes. At the same time, however, the framework must be realistic about the need to subject automation to human judgement and innovation. Also, a practical framework must be able to coexist with existing processes, beliefs and values at least until a critical mass of data can be captured in the framework.

One useful model uses competency definitions as building blocks in competency information modelling and related operations. In this model, a competency definition uses natural language to capture information about a particular competency. This information is captured in a form most useful for human readers, but is not in itself sufficient to support automation. The competency definition is however captured in a data record that can be referenced in various operations, just as one can reference an existing book in various academic or business transactions without having to rewrite it or even open it every time it is used. For example, it becomes possible to create compact competency records for different people, with all the records referencing the same competency definition rather than duplicating the information in each record. Such a competency definition can capture the competency definitions found in many existing competency models or "competency standards". It can also be more or less specific or contextual. Obviously, a less context-specific definition can be useful for reuse in more contexts than a highly specific one.

The natural language of Human Resources often invokes competencies along with terms like "human capital". Recently, many systematic attempts have been made to try to formalise the meaning of competencies in this context. These efforts aim at capturing and representing competencies in a way that supports actionable capabilities and requirements, such as the ability to describe, plan, assess, and train human resources. Ideally, it should be possible to exchange information about common competency definitions throughout an industry or even within an enterprise: we can say that *competencies are becoming the common currency of the labour market*.

Competency models that support automation for related competencies in the appropriate context to enable practical applications is the next logical step. This requires a model to capture competency semantics in ways that support automation. Competency definitions, for all their usefulness, are not sufficient for this purpose. So another, richer model that works alongside inventories of competency definitions is required.

Competencies are never defined or applied in a vacuum. In reality, they are always acquired, assessed or applied toward successful performance in a particular context. Tasks and responsibilities, administrative classifications, professional license requirements and of course job descriptions are examples of contexts in which competency definitions at various levels of granularity become operational. However, there are still significant obstacles to overcome. Reading even the simplest competency definition expressed in natural language, such as “can effectively defuse conflicts with co-workers”, one quickly discovers that it implies a lot of sub-skills and related competencies, some of which are highly dependent on a specific context such as specific tasks or the type of work environment.

So, formally capturing the semantics of competencies requires semantic models that are specific to particular contexts. The domains covered by such models may be more or less broad, such as a trade or profession, or a task that is performed by many people in a particular context. In such a formal semantic model, the basic competency definitions are essentially “building blocks” that contain the human readable descriptions. In other words, the smallest level of granularity at which competencies are defined in semantic modelling for competencies is a competency definition expressed in natural language. We saw above that competency definitions can be useful at any level of granularity. A semantic model can also be associated with a higher level competency definition. In that case, it is used to describe formally the components and implications of that higher level competency in a particular context.

These days, competencies are typically described in a natural language which cannot be properly processed by machines. This means that competency information is usually not available in any formal notation that supports automation, especially for operations like semantic matching to find similarities and relations between definitions from different sources or which are expressed in different human languages. Experience has shown that attempting to do this kind of matching without some formal underlying semantic model is problematic at best. For example, two enterprises may use the same competency title to describe quite different competencies which imply different sets of constituent sub-competencies.

The sheer number of competencies that exist in people’s minds and in more or less idiosyncratic models is staggering. To enable automation, all these must be captured in such a way that their semantic information can be used for filtering, processing, aggregation and matching. Natural language resources are not sufficient to support automated processing of the information. They are also not sufficient to support exchange of information about the meaning of competencies with other entities within enterprises, with trading partners and with entities concerned with workforce development, job placement and education.

As we saw above, traditional databases and inventories of competencies that rely on string search and string matching are not sufficient for this task. Systems that understand the knowledge aspects of competency information are required. Such systems must be able to manage millions of competencies and interrelationships, and to support operations on the knowledge itself, rather than just the massaging of fragments of text. This is the only way to manage the massive amounts of information required in the more complete forms of competency modelling. It is also the only way to manage the massive, unavoidable real world changes that affect any competency model. By constructing semantic models and leveraging both competency definitions (mostly for humans) and semantic models (mostly for machines), automation becomes possible and can provide effective results that the human stakeholders will feel they can adopt.

Competency information changes all the time. Change affects not only the sets of competencies that are relevant to a job, but also the global competency set for the broader domains. In any domain, some competencies appear, some disappear, and some become obsolete. Using formal competency definitions to capture this information allows for a measure of sanity. For example, if one accepts the idea that when a competency changes, the previous definition should not change but should be replaced with a new one, it becomes possible to manage the impact of the change on models and processes that reference those definitions. When only natural language resources are used to document competencies, they tend to quickly fall behind reality, since change tracking cannot be automated without an understanding of the dependencies and implications of the competencies involved. In all but the simplest models, this is simply too difficult for humans to process.

A third kind of model seems to be needed to automate operations that are more restricted in time and scope and to represent competency models in a format that is somewhat less daunting for normal human beings than formal semantic models. Unlike semantic models, which can be unapproachable for most people, this simple competency model approach uses hierarchical models to represent a simplified "snapshot" of a more elaborate competency model as it exists at a particular moment in time in a particular context. Such a simple competency model can be used to implement certain operations like summation of measures for related sub-competencies, or to present a useful fragment of a model to human readers in a familiar format. Many existing competency models used today in enterprises and government funded programs are basically lists or hierarchies of competency definitions. A simple competency model that is hierarchical is immediately understandable by anyone who is familiar with those models. Since those existing models can be often be mapped into a standard simple competency model, this can facilitate the capture of information to be fed into a more elaborate semantic model.

It is impossible to discuss competencies without mentioning assessment and measurement. In theory, competencies can be seen as "predispositions of human behaviour" and as such cannot be directly measured. However, in the real world competencies are assessed all the time. A number of known assessment methods do exist; those vary depending on the type of competency, the credibility requirements, the form of evidence that can be assessed, and operational constraints such as time and cost. Assessments basically imply heuristics to deduce measures of proficiency for various competencies from observable human performance [21]. (e.g. in work processes, or in exams). In the real world, assessment results are often viewed as a more or less credible predictor of job performance. Most organisations define, more or less formally, various proficiency reference levels. A proficiency reference level is typically contextual. It may be associated with, or even take the name of a particular job or administrative classification. Often, when an individual is assessed for proficiency in a particular competency, the result is a measure relative to the proficiency reference level. If the measure meets or exceeds the reference level, the individual will be considered "competent" for that particular competency. This may in turn guide operational decisions such as hiring or training, which is why some measure of assessment credibility is often critical.

In conclusion, without standardisation of modelling approaches, the meaningful exchange of competency information for specific competencies is impossible. While some enabling standards are already emerging, others still need to be created. It is unrealistic to expect that a single standard would be sufficient for the whole HR domain. Such a single all-encompassing standard would also not be desirable, because of its sheer size and complexity, not to mention the difficulty of ensuring relevance over time and for the many different HR applications and processes. Such a standard would be obsolete before agreement could be reached among all the stakeholders about its features. It is therefore reasonable to expect the emergence of an evolving collection of interoperable standards; smaller, more targeted standards rather than massive standards will be the norm, not the exception. Some of the standards might be foundations on which broader standards can be built, while others might be profiles that build on foundation standards to support specific domains or applications. Some standards might be of interest only to very specialised stakeholders, while others will be broadly applicable.

3.1.2 A Brief History

In order to understand the importance of semantic web technology in the HR domain, it is useful to have an overview of the brief history of computing in business.

At first, Information and Communication Technology (ICT) was used to automate existing administrative processes. These islands of administrative automation were limited to one company or even part of the company, while all communications to other companies or parts of the same enterprise were still done on paper. The best one could hope for in this scenario was *data integration* which meant that two administrative systems were merged into one system. If paper data streams were involved, human interpretation was used to translate from system to system.

The next step was value chain integration or enterprise support, commonly called Enterprise Resource Planning (ERP) by then. Here it is not data integration, but *service integration* which drives the process. Although in many cases there still is one (large) system, the processes have been nailed down much better and the industry now pushes for service-level cooperation between systems. For this to work, humans still need to interpret service agreements and other contractual issues.

Human resources are still sparsely supported by ERP systems, in our opinion partly because the level of interoperability required for any useful exchange between systems goes beyond data and services, to actual *knowledge*. HR is a domain *par excellence* where tacit, dynamic knowledge is essential. This knowledge is extremely detailed and becomes overwhelmingly large if this level of detail is formalised. So there is a natural barrier, which we might call the *semantic barrier*, which prevents ERP-like systems from successfully moving into the HR domain beyond flat administration or vacancy exchange. A good example is that the majority of documents exchanged in HR are in a pure word processor format, in other words, without any associated meaning. Formalised data items are trivial and come straight from the *islands of administrative automation* era.

Semantic technologies, including (standardised) ontologies, are a prime candidate for equipping the HR world with formalised but useful small knowledge blocks, namely *competencies*, which may serve to exchange knowledge about people and requirements between unrelated systems.

3.1.3 Introduction to Ontologies and Their Role in the Semantic Web

The term 'ontology' originates from philosophy, where it concerns the nature of being and the essence of things. In the early 1990s, the same term was reused by Artificial Intelligence researchers to describe high-level specifications of data structures that make it easier for computers to co-operate and share knowledge to answer questions and solve problems. When building an information system, it is desirable to separate the descriptions of things that exist in the real world from the mechanisms that are necessary to make the system work. An ontology is a set of descriptions of real-world things - particularly when they refer to classes of things rather than individual items. The ontology is a declarative *specification* of the representations that will be embedded in the system, but it has the advantage that it can be inspected and refined independently of the system. This makes it far easier for computers, or humans for that matter, to share a common understanding of domain terms and reuse the same set of terms in different projects.

The best-known definition of ontology is from[17]:

> *"An ontology is an explicit specification of a conceptualisation"*

To briefly explain this concise definition: an ontology is *explicit* because it is external to the system that uses it; it is a *specification* because it describes the knowledge representation without being the implementation of it (the implementation could be in the data structures of a computer program or the schema of a relational database); and it is a *conceptualisation* because it concerns the representation of concepts. In practical terms, an ontology is both a controlled vocabulary of things in the real world and a networked knowledge structure that captures the relationships among them. It is also a *model* of the domain of discourse. The terms modelled may be 'things', concrete or abstract, or processes (examples are `Person`, `Project`, and `Interview`). As there is generally more than one way of modelling domain concepts and their relationships, we usually speak of an ontology meaning a particular model of that domain.

A model is an ontology if it:

- is a declarative, explicit representation of a domain; that is, it is possible to inspect the domain representation independently of the system(s) that use it;
- is consensual, containing the combined knowledge of more than one domain expert;
- can be used to solve more than one problem in the domain;
- can be used in multiple applications;
- is stable (i.e. changes little over time) and long-lived.

The benefits of ontologies can be seen in three main areas [36]:

- Communication;
- Systems Engineering;
- Interoperability.

Ontologies aid communication because they expose perceptions that might otherwise remain hidden in the minds of developers or the inner workings of an information system. By explicating and sharing these perceptions, different views of the domain are properly discussed and agreed upon, and such discussions do not require deep technical knowledge. Such discussions help to highlight problems early in a project's lifecycle before the mistakes become too costly to repair. In other words, the construction of a new ontology (or the validation of an existing ontology) helps to elicit the requirements of the system. Ontologies can aid systems engineering in other ways too: the requirement for ontologies to be reusable typically provides greater insights into the dependencies among concepts and the assumptions they make. Such insights can lead to a superior design for the resulting system. Ontologies can also prove useful when devising test cases for the implemented system. Ontologies help to provide interoperability among systems because they embody a shared understanding of the domain. Systems can interoperate by committing to the same ontology, and can exchange information even if the low-level representations of domain concepts are different (for example, because they are implemented in different programming languages).

Ontologies are already being widely applied to scientific disciplines ranging from biology and medicine to geoscience and astronomy. For example, scientists at NASA/JPL have developed a semantic framework called SWEET (http://sweet.jpl.nasa.gov/sweet) for the exchange of earth science information; and several biological ontologies are listed at the Open Biological Ontologies site (http://obo.sourceforge.net/).

Tim Berners-Lee named ontologies as one of the key technologies in his vision of the semantic web. In the semantic web, shared meanings are ascribed to terms used in published semantic web documents, thus enabling resources on the Internet to be processed by machines (as well as humans), instead of resources being 'merely' human-readable text. Figure 1 shows the layers in the architecture of his vision and how they fit together (adapted and simplified from the diagram at http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html).

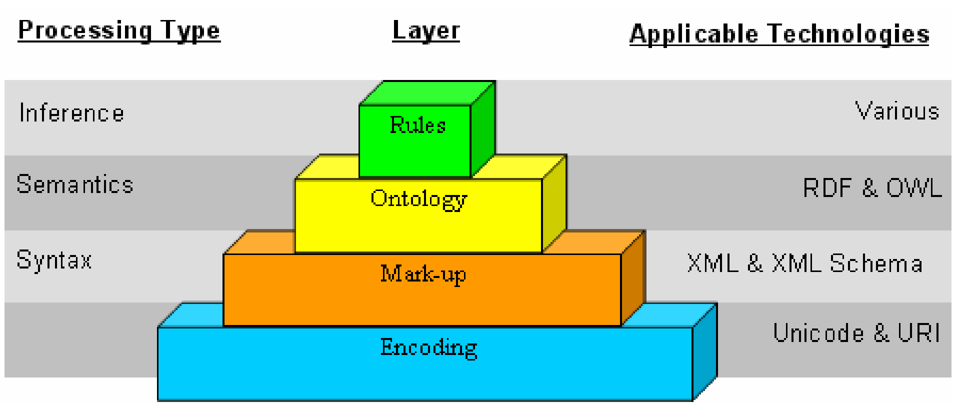


Figure 1 Simplified version of Tim Berners-Lee's "layer cake"

Starting from the bottom, the encoding layer contains a mapping from numbers to visible character glyphs; the mark-up layer contains text organised into structured elements through the addition of mark-up tags; the ontology layer constrains the meaning of those elements by specifying the relationships among them; and the rules layer provides for the ability to automatically derive property values, prove properties of elements, or assess the trustworthiness of a description.

As ontologies contain a shared understanding of the domain of interest, they are absolutely fundamental to the acquisition and distribution of knowledge. They act as enabling technologies for large scale knowledge management, providing the semantic interface for different sites to act as knowledge providers or knowledge acquisition agents. Note that while databases can provide a powerful repository for the storing and retrieval of information at individual nodes of a ‘knowledge network’, it is the ontologies that enable the interoperability among heterogeneous systems.

Typically, Semantic Web applications begin their lifecycle in a closed domain with simple ontologies, and may eventually evolve into networked and/or shared domains and into complex applications. Since Semantic Web research is still very much in its infancy, research tends to start from simple applications demonstrating proof of concept, but one of the major problems is scaling these up to be of practical use in the real world. This problem is currently at the forefront of the development of applications for the Semantic Web, hence the existence of EU research projects such as SEKT[3] (Semantically Enabled Knowledge Technologies), which aims to develop and exploit the knowledge technologies underlying Next Generation Knowledge Management, and in particular, NEON[4] (Lifecycle Support for NEtworked ONtologies), which aims to support the whole ontology engineering lifecycle by developing a reference architecture and a concrete toolkit and developing appropriate methodologies for the contextualised construction and use of networked ontologies and associated metadata. Projects such as TAO[5] are focusing on transitioning legacy systems to open semantic service-oriented architectures, enabling semantic interoperability between heterogeneous data resources and distributed applications, at low cost, for both SMEs and large enterprises. The time is now right for the creation of infrastructures to aid transitioning of legacy applications by means of ontologies and refactoring, thereby enabling companies to take up these new developments without having to reimplement their applications. While TAO does not focus specifically on the HR domain. the same principles apply throughout: for example, the bottleneck of semi-automatic ontology creation, automatic methods for semantic metadata creation, creation of distributed heterogeneous repositories, and so on.

In 2002 the Gartner Group predicted a massive usage of ontologies for business application integration in the timeframe 2005-2010, foreseeing a roadmap starting with lightweight ontologies or taxonomies evolving into strong knowledge representations in 80% of application integration projects within this timeframe [14]. The use of ontologies is likely to expand enormously in two particular ways: **data browsing, search and retrieval** (moving from retrieving documents to retrieving specific data, and enhancing search technologies with semantics), and in terms of **inferencing**. However, the widespread use of ontologies in real world business applications is largely impeded by factors such as scalability, interoperability and usability. We shall discuss these issues in greater detail in Section 2, where we look specifically at the need for ontologies in the domain of HR, and in Sections 3 and 4, where we investigate the current problems and potential solutions.

3.2 The HR Application Scenarios and the Need for Ontologies

3.2.1 Introduction

The field of Human Resources (HR) is a generic domain into which a great deal of effort in terms of knowledge management tends to be placed, because every company, organisation and business unit must encounter it. HR departments often have an eye open for knowledge management in order to monitor their environment in the best way, and many recruitment consultancy companies have watchdogs to monitor and alert them to changes. Among the multiplicity of online portals there exists a variety of job search engines (portals) which already use knowledge management extensively to link employees and employers, e.g. JobSearch[6] and Job Portals[7] The growing pervasiveness of Knowledge Management (KM) in industry marks an important new watershed. KM has become embedded in the strategy, policy and implementation processes of institutions and organisations worldwide. The global

3 IST-2003-506826, http://www.sekt-project.com/

4 IST-2004-27595, http://www.neon-project.org/

5 IST-2004-026460, http://www.tao-project.eu/

6 http://www.job-search.com

7 http://www.aspanet.org/solutionstemp/jobport.html

KM market has more than doubled in size since 1991 and exceeded US$8.8 billion in 2005. KM applications are expected to save Fortune 500 companies around $31 billion, and the broader application cost has similar projected forecasts.

The HR domain has many facets. But one particular 'view' on HR is fast becoming the motor for serious change. Competency-centric HR is not only reaching a great number of traditional HR processes, it is also the start for a new wave of change in the labour market as a whole. Programmes such as 'Matching on Competencies' (MoC, replacing the traditional Job-CV matching), Mobility@Work and competency driven qualifications form the basis for competencies to readily become the currency of the European labour market.

Clearly, therefore, the combination of KM and Human Resources has enormous implications for the growth and dispersion of such new technologies to industry as a whole. Tools and resources such as next generation Knowledge Management platforms pave the way for such developments, by leading to interesting and useful acquisitions of knowledge that save time and money and benefit real users in industry. Examples of such systems are the h-TechSight Knowledge Management Platform[24] and Ontotext's JOCI (Jobs and Contacts Intelligence – Recruitment Intelligence through Semantic Web Technologies). Companies such as Innovantage[8] have recently been established which make use of such technology to provide information such as vacancies, contact information and biographies harvested direct from company, academic and government websites via tangible business intelligence tools for the recruitment market.

A "conditio sine qua non" for such projects, tools and systems to make a real impact, however, is a meaningful way to exchange competency data between industry, education and public and private employment services. Europe therefore needs a 'semantic' standard topping the existing internationally accepted HR and Learning syntactic standards. HR is far from alone in this respect.

In the rest of this section, we describe the main application scenarios (i.e. categories of applications) in the HR domain, and their demands for a semantic component.

3.2.2 Job Search Engines and Job-CV Matching

During the last decade, job-search engines and job-CV matching devices have multiplied and at present, many websites are available both at local and at international levels. Most of them offer job search as well as job-CV matching, and some of them use business intelligence tools such as agents to find, classify and structure job information.

In general, using agents means that software is made accessible to automatically search the information needed, rather than relying on humans. The more agents are able to work automatically, and to learn and adapt, the more they can be considered intelligent. Using agents implies that a user has to instruct his/her agent to search information in place of them. Instructing an agent implies that the user has to declare what he/she wants to be searched.

At present, the most advanced job-search and job-CV matching services make agents available to the users. Hence, agents relieve users (both job seekers and employers) from the tedium of searching among many possible postings and applications on the internet. This saves time and may also produce a more extensive and thorough search.

However, the current agents available work by word matching only. Moreover, in most cases, in order to instruct an agent to perform a job search, users can only choose items of information from pre-defined sets of menus. From the job seeker's perspective, these menus usually include job location and job area; in some cases they can also consider job profile, business sector, and type of employment contracts. Sometimes the selection of at least one specific item is mandatory; in most cases, between 1 and 10 items can be selected for each menu. Keywords are also always requested and word matching usually follows the Boolean logic criteria. Examples of such services are Monster[9], the most popular

[8] http://www.innovantage.co.uk/

[9] http://www.monster.com/geo/siteselection.asp

international job-search and job-CV matching service presently available, Job-search[10] from the US, CEN-Marketplace[11] from the UK, and Borsa Lavoro Nazionale[12] from Italy.

In the case of job-CV matching, currently available services allow a user both to upload an existing CV and to create a new one according to a specific format. Here most services use agents just to recognise keywords inside a CV, which match the requirements expressed by companies.

Finally, with respect to most job-search and job-CV matching services, once the information required has been selected, the personal agent will start to search inside the reference job-search or job-CV matching service provider's database. Currently, only a few services (such as CVmatching™ and Innovantage) have started to use automated systems for natural language context-based recognition that goes beyond simple word matching, and hence are the only really advanced business intelligence tools.

Currently, most jobseekers need to make use of job-search and job-CV matching services, but nonetheless these tools appear still to be quite unreliable. The quality of such services is measured in terms of accuracy of information, fast and successful response as well as time and effort spent in arranging a job posting or application. Although responses are generally fast, the accuracy of the information received and the amount of time spent in job posting (both in terms of application and analysis of the responses) is still problematic. Using fixed menu and word matching systems to instruct one's own agent can make the subsequent search less effective, and the risk of missing important information is very high. When the items selected for the search are too generic, then the responses may be quite superfluous; on the other hand, when the items selected for the search narrow the search scope too much, then the information returned is insufficient. If a user makes a poor selection of items, the information differential can even overtake 100%.

Consequently, not only is information altogether unsatisfactory but the time saved during the search is then wasted in screening the answers received and in revising the items selected for the search. Moreover, users also spend a lot of time uploading their job postings or applications from one service to another. Apart from a small amount of basic information universally required, each service asks for different information and even when the information scopes are the same, the items inside change. Users have to change codification rules any time they shift from one service provider to another, or even with the same provider in a different country. Finally, building effective sets of keywords can be quite complicated.

The main problems related to job-search and job-CV-matching services are of two types:

1. Any service provider usually provides quite different items to be used for search or for the instruction of search agents. The reasons may concern their marketing strategy solutions and business customer features, but they are also quite strongly related to the lack of standardisation of concepts behind the words and phrases used.

2. Most job-search and job-CV-matching services currently still use rigid tools based on syntactic rules only (e.g. word matching and in some cases synonyms).

Consequently, the following can happen:

1. Users might select the wrong terms because there are no clear and explicit definitions behind them, i.e. there is a lack of transparency.

2. Job search even by agents has a limited range of actions defined by the words or phrases selected.

If the terms used are quite vague and imprecise, as in (1), and the search is only defined by these chosen terms, as in (2), then the probability of making an inaccurate search is very high. This margin of error can increase when there are more items to be selected. In fact, even if that seems to help users better define the range of search, it could on the other hand just become a restraint preventing a successful result. Moreover, users have to reformulate their search criteria any time they shift from one service provider to another. This means a waste of time and an increase of the possibility of error.

10 http://www.job-search.com

11 http://www.cambridge-news.co.uk/marketplace/jobs/

12 http://www.borsalavoro.it/wps/portal

These are clearly problems related to semantics and to semantic standardisation. Hence, in order to tackle these problems, it seems that we have to resort to semantic solutions.

Standardisation of concepts can answer the problems generated by the first question mentioned above. Actually, an agreement on the meaning of terms has always been the *conditio sine qua non* to effectively communicate. In this case, the reference subjects are employers/companies, applicants and job-search or job-CV matching service providers, but the list could extend to education and to other labour and recruiting agencies, too; the concepts to be defined and shared by these communities involved could be things like *competence*, *skill*, *job area*, *job profile*, *business sector*, etc. In this way, the transparency of job-search and job-CV matching services could be improved because reference words and concepts would become clearer and users could arrange their job postings and applications as well as resumés in unambiguous ways which are more easily reusable.

On the other hand, more advanced semantic solutions can answer the problems generated by the second question mentioned above. In fact, if agents were fed with sentences and whole documents containing examples of what is needed, and if they were able to learn from these, they could really avoid the problems related to word and phrase selection as described above.

Semantic web solutions using sound inference processes would allow the recognition of the relevant pieces of information without compelling users to make selections between fixed items. The search would be guided neither by sets of fixed words nor by mere matching rules; instead they would be performed by semantic-based deduction programs, which would be able to infer from given sentences other possible sentences. Hence, these solutions would make searches more powerful However, in order to implement such advanced semantic solutions, i.e. to build *shared* common languages, semantic standardisation is necessary.

Concept standardisation would allow users to maintain the same postings and application formats from one service provider to another, thus saving time, whereas sharing a full common language would allow employers to simply put their job posting on their websites without sending them to each reference service provider selected. In fact, in this way, the job-search and job-CV matching service providers would be able to autonomously crawl the companies' websites, extract the proper information identified, and make suitable inferences.

Furthermore, portals could also be created as interfaces between users (both job seekers and employers) and job-search or job-CV matching services' databases, so that the resumé postings, on one hand, and the information retrieval, on the other hand, could be addressed to all the reference services at once. This would allow users to save more time and the search results could be even more complete and effective.

For the same reasons, the use of the semantic web for job-search and job-CV matching tools could be also useful inside large companies for the internal mobility of people who have to move from e.g. one continent to another. In fact, internal personnel research according to specific competency-based criteria can be very difficult when a company has many offices all over the world, each of them defining terms in different ways and having a high complexity of information to be managed.

In conclusion, semantic standardisation and the related consensus processes among the communities involved (economic and education systems) appear to be the first steps required towards the enhancement of job-search and job-CV matching services. At present, the UK and the US have the most sophisticated job-search and job-CV matching services. In these countries, the standardisation process on concepts began about 20 years ago and now they have powerful and shared systems nationwide. See, for example, the National Occupational Standards - NOS[13] from the UK and the O*NET[14] from the US. The European Commission has also launched standardisation programmes such as Europass[15] – concerning CV and qualifications standardisation -- and the eSkills forum[16] initiatives - addressing the ICT skills and competences standardisation with the forthcoming European eCompetence framework.

---

[13] http://www.ukstandards.org/

[14] http://europass.cedefop.europa.eu/

[15] http://europass.cedefop.europa.eu/

[16] http://communities.trainingvillage.gr/esf

The challenge is to reach common standardised semantic systems nationwide and Europe-wide. For this reason, a strong cooperation between national and international institutions as well as the construction of multi-stakeholder partnerships is needed to facilitate and foster the labour market mobility and transparency.

3.2.3 Competency-based assessment of employees

If we take for granted that e-assessment can represent an advantage for both assessors and examinees[17], then just like the job-search and job-CV matching services presented in the previous paragraph, competency-based assessment tools would also be more effective and useful if:

- they could be linked to other services, e.g. e-portfolio (i.e. electronic CV)[18];
- their results could be communicated to diverse institutions in the same format, e.g., different training or business offices, labour agencies and market places;
- their contents should be shared among communities, e.g. tests for assessments.

In short, interoperability among different organisations could improve the effectiveness of assessment processes and the usage of assessment outcomes.

If we focus on large companies with offices all over the world, electronic data interchange and integration could also speed up some internal HR management processes and make them more effective. For example, electronic interoperability could be very useful:

- in the case of recruitment and internal mobility, to identify the right people wanted, by matching their assessment outcomes with job positions requested;
- in the case of multi-stakeholder projects, to define partners' mutual roles inside projects, by sharing common e-assessment results;
- in the case of competences gap-analysis, to identify competences to be enhanced and the related training paths to be carried out, by matching assessment results with reference learning modules.

In general, along the supply chain, if companies were able to mutually recognise and match up people's expertise (assessment outcomes) with job roles, then they could better:

- evaluate suppliers' competences;
- sell their own competences;
- determine prices, rates, fees and compensations more transparently.

Nevertheless, some requirements must be met, in order that assessment data inside a company can be really connected with competences and job roles, on the one hand, and with other companies' assessment data, on the other hand.

Just like job-search and job-CV matching services, assessment also first needs concept standardisation. It is therefore necessary to define:

- the objects to be assessed, typically expressed in terms of outcomes;
- the competences to which assessment outcomes may be related, according to the reference business contexts. Note that it is not necessary to define job roles if it is assumed that job roles are combinations of competences.

The means of assessment, i.e. the methods adopted to evaluate the outcomes identified, should also be defined just for transparency and mutual trust).

The Skills Framework for the Information Age (SFIA)[19], developed in the UK, can be an excellent example of this preliminary task of defining meanings according to a specific, reference context. In

[17] http://www.jiscinfonet.ac.uk/

[18] http://ferl.becta.org.uk/display.cfm?resID=13337

[19] http://www.sfia.org.uk/

fact, they have been standardising skills and competences inside the ICT business process, just to refer to a common set of items (skills and competences) really usable to define job roles, assessment outcomes, skills requirements, etc.

Furthermore, if we also want to link assessment results to learning paths, a standardisation of qualification models behind them is needed. A first pan-European attempt to find a common basis for qualifications standardisation has been made by the expert group who have just elaborated the EQF (European Qualification Framework)[20].

An example of a full standardisation system between qualifications, occupational standards and assessments, comes from the UK again, with their National Qualification Framework[21]. It describes the structure of national qualifications and groups them into three categories (General, Vocational and Occupational) with nine competence/learning levels (from entry level to level 8). The related database allows the user to view all the national qualifications and to choose between them. For each codified qualification, they can find the associated accredited awarding body, the specific competence/learning level, and the related codified competence/ learning modules and unites, to be assessed by the reference awarding body. Each awarding body has a website describing the assessment method and the links to the Education Institutions managing the reference qualification paths. The latest development is the e-assessment system to be also related to the e-portfolio.[22]

We could complete this system by linking the companies' assessment systems, too. However, concept standardisation is not enough to ensure that assessment results match competences or job roles required, match training modules fulfilling them, and feed into e-portfolios.

As with job-search and job-CV matching services, searching the most suitable assessment outcomes fitting the reference competences required, or searching the most suitable learning modules for the assessment outcomes, as well as integrating assessment outcomes into an e-portfolio, may require complex inference processes besides simple word matching. That means a full common language (meaning of terms plus semantic rules definition) should be shared.

According to this scenario, cooperation between the different actors, i.e. companies in these examples, again is a must. Reference frameworks and interoperable electronic tools can be really useful and effective only if actors want to become partners forming business communities. Multi-stakeholder partnerships are needed not only in order to allow the frameworks and e-tools be constructed but also in order to make them effective.

3.2.4 Competence-based Learning Management and Competence Management

Recently, competencies have been on a road to success as a suitable abstraction for individual and collective human behaviour/performance and respective requirements from an organisational/market point of view. Competencies are largely superseding the concept of knowledge in many places, as competencies provide a more holistic point of view [24]. This is especially true for the domain of human resource development, training, workplace learning etc., which has often been alienated by oversimplified views of knowledge management and the neglecting of the complexity of learning processes.

The importance of competencies for the HR domain is reinforced by the fact that major players in the HR domain (both specialists and Enterprise Resource Planning (ERP) vendors) have incorporated some form of support for competencies (e.g. SAP, Peoplesoft, META4, ExecuTrack, Dexter). These products provide facilities for recording employee competencies and for analysing the competency status (e.g., for monitoring human resource development activities). However, these products currently lack the interconnection of different levels and aspects: Competency records of employees are rarely used to personalise the learning offer on the one hand, or for automatic aggregation into competency and HR development controlling on the other hand. Specifically, we can distinguish two levels of learning and education management: the *micro-level* (where we provide individual learning offers to employees) and the *macro-level* (where we want to ensure and foster organisational competence), which have to be

[20] http://ec.europa.eu/education/policies/educ/eqf/index_en.html

[21] http://www.qca.org.uk/

[22] http;//www.qca.org.uk/6877.html

closely connected for efficient competence development strategies. These two management levels are explained in the following two sections.

3.2.5 Micro-Management Issues: Learning on Demand

Traditional training programmes are increasingly being questioned in the light of the increased rate of change and the individualization of educational processes. The key to more efficient human resource development is learning on demand methods, providing fine-grained learning offers just when they are needed. In such learning on demand scenarios, learning micro management becomes so complex that we need to automate it to a large extent to keep it manageable at all. We need to capture the work context/situation, derive from it the requirements and deduce a competency gap that needs to be overcome by some means or other.

As already shown in [24], ontologies are promising instruments for capturing the work situations in terms of organisational structures (processes, roles, departments) and in terms of personal tasks. Integrating competencies as descriptions capabilities of an employee, the requirements of situational aspects and learning goals of learning resources can yield a comprehensive framework that is both more fine-grained and more manageable.

Extending learning activities to informal learning also brings “expert finder” applications back into focus. Based on competency descriptions of employees, learning support systems can recommend colleagues who can provide help for the current situational need. The other source for informal learning is knowledge management activities where ontology-based approaches can provide a smooth transition.

3.2.6 Macro-Management/Competence Management Issues

On the macro level, competence management has been established as an important element of value-oriented management practice. It is still rather problematic, however, to know how to connect the strategically-oriented competence management with its organisational and market perspective using learning micro management.

Ontologies offer the possibility that we can use the same competence catalogue on both levels – although probably at different levels of abstraction (see Figure 2, taken from [19]). This allows for coherence on goals and results over the different levels and ensures efficient communication. Furthermore, automated aggregation methods from fine-grained individual competencies and situational requirements to organizational competencies and requirements can improve the agility of competence management approaches drastically (see [19]).

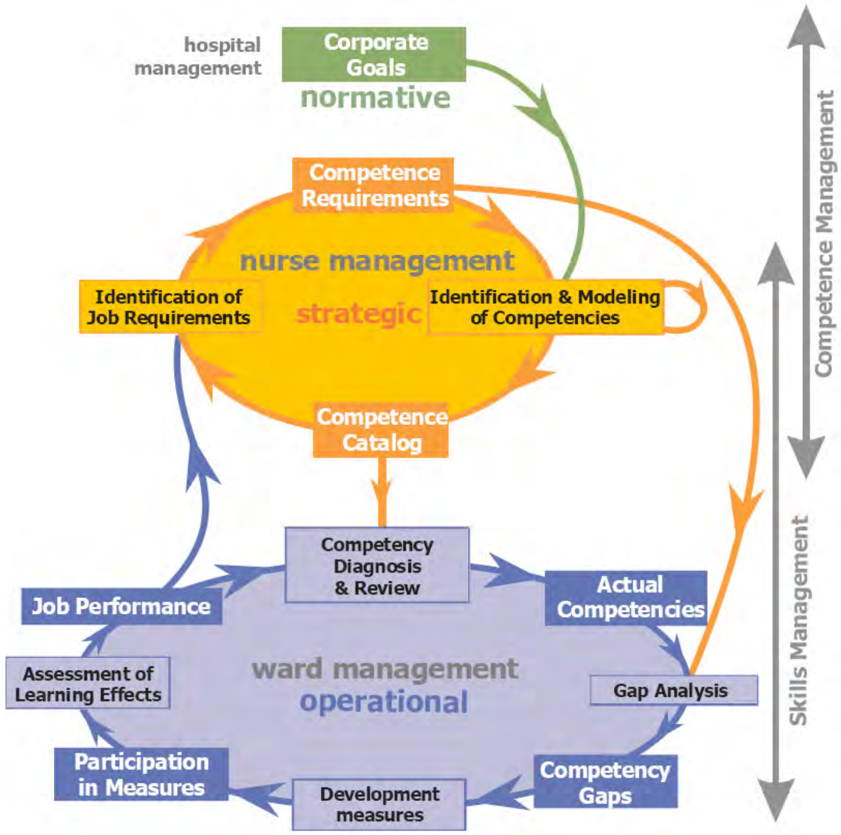


Figure 2: Connecting strategic and operational issues

3.2.7 Potential of Ontology-enabled Approaches

The main benefit of ontologies in this scenario is the automation of processes that would otherwise not be manageable at all or would require a substantial reduction in the level of detail. So what kind of algorithms/methods do we need? The following two cases have emerged:

1. Profile matching with similarity measures. The most frequently cited case is the matching of a individual's competency profile with a requirements profile.

2. Finding learning opportunities with knowledge gap analysis and competency subsumption.

Whereas in the first case, the result is the degree to which a person fits a requirement, another important use case is the identification of suitable learning opportunities that can be proactively recommended. In order to realise this, a knowledge gap needs to be calculated by comparing the requirements profile with the current competency profile, yielding missing competencies. One important aspect that needs to be taken into account here is the issue of competency subsumption, i.e., we cannot simply rely on direct comparison, but need to consider that a competency can be subsumed by another competency (e.g. higher competency level, generalisation, composition).

Furthermore, semantically coherent models foster the tighter integration of different levels (operational, strategic, normative) and different functions (training and e-learning, knowledge management, management-by-objective, organisational competence management etc.).

3.2.8 HR Domain connections

The schema below (Figure 3) summarises some of the most relevant links between the main objects belonging to the HR domain, already mentioned in the previous paragraphs of this chapter:

- People's curriculum – portfolio
- Companies' job postings

- Assessments
- Certification programmes
- Learning programmes, learning modules, learning materials
- Competence and job profiles frameworks
- Project activities, business and work processes.

This picture aims to underline some of the main scopes and to help focus on some possible prospective relationships. In fact, as already illustrated in the previous paragraphs, some possible relationships between the HR domain objects can create new social and business opportunities and challenges if well manifested and managed by new ICT tools such as ontology-based systems and semantic web applications. The simultaneous visualisation of all these links between curricula and job posting, between assessments and learning or certifications, and so on, provides an overview of the whole HR domain and its current and potential connections.

Particular attention should be paid to *competence and job profiles frameworks*, on one hand, and to *projects, business processes and work activities*, on the other hand. In fact, they can be a basic reference for building the other objects such as assessments, learning and certification programmes, and job postings as well.

With regard to the scope of projects, business and work processes, if they were well-defined, we could imagine verifying or even anticipating possible competences by deduction (using a well-formed inference system). ERP (Enterprise Resource Planning[34]) in the form of systems standardising business processes, as well as continuous improvement approaches to business and work processes, like Lean Six Sigma[18], could become a starting point to build two-tier languages shared by the HR communities.

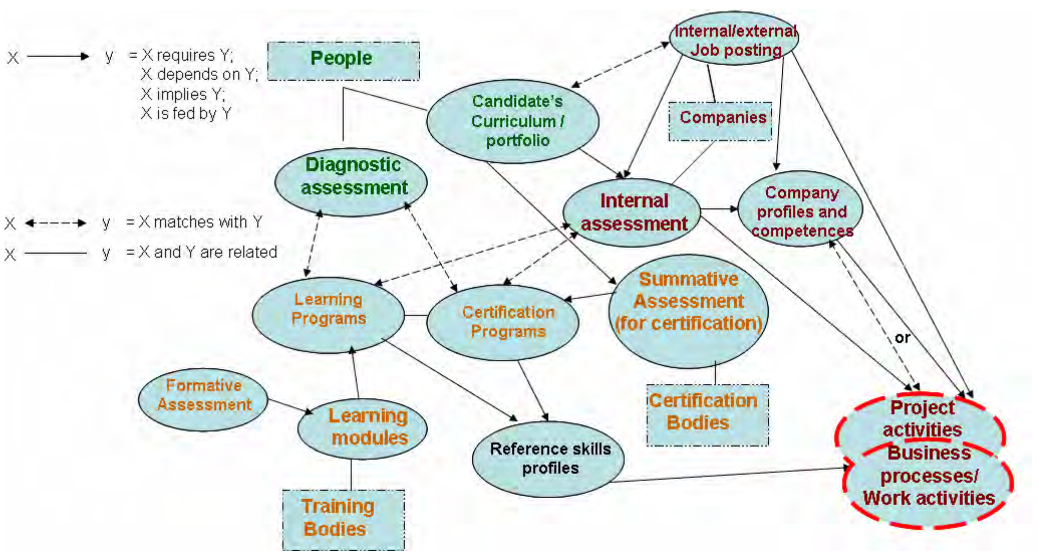


Figure 3: HR Domain network scenario

In the diagram, each node is linked to other nodes in various ways. For the reasons explained in the previously, if all nodes could speak the same language (concept definitions and semantic rules), building and exploiting these relations would be much easier and make the system more effective.

In this scenario it would be interesting to represent business and work processes, or even project activities too, to which competences, profiles, assessments, learning programmes, etc. can refer. In the schema the reference nodes are circled in red to indicate that it is a critical area to be explored. It would also be interesting to infer or verify possible profiles and competences from complex processes for which the margins for standardisation are very fine. For example, it could be a way of finding out new prospective options in terms of competences to develop, which could even become key success factors and competitive advantages for companies.

On the other hand, concerning big projects, it could be more effective to share a project formalisation helping to better define competences and job roles among partners and along the supply chain.

3.3 Challenges in the HR Domain

3.3.1 The Challenge of Modelling Competence Ontologies

Before assessing the ontology engineering process for the particular problem domain, there are some important considerations to be made.

The ontology is supposed to be a shared understanding of the domain in which the involved stakeholders are interested. Usually, an ontology is conceived as a set of concepts (e.g. entities, attributes, processes), their definitions, and their relationships; this is referred to as a conceptualisation.

In the intended HR processes, the (currently) smallest and most important artefact we can identify is a (human) competency. For processing purposes, a competency is supposed to be measurable, therefore it is crucial to define it very precisely. Competencies, however, are usually acquired through experience. This is called tacit knowledge. Polanyi used the phrase ``we know more than we can tell" to describe what he meant by tacit knowledge. Tacit knowledge is a kind of knowledge which is difficult to articulate with formal language because it is either too complex or simply because it is informally internalised in people's minds. Yet it is shared and exchanged in normal social interaction.

Furthermore, suppose tacit competency knowledge took an explicit form found in written statements, documents or metaphors As currently there exists no standard, the interpretation requires reflection among individuals which is subjective, hence making it useless for machine processing.

In order to tackle these problems, the competency elicitation process should be considered as a collaborative where domain experts gradually build and manage increasingly complex versions of competency artefacts and their diverging and converging interrelationships, in order to ultimately come to a unique shared understanding. This constantly evolving wealth of knowledge requires persistent versioning support.

Divergence is the point where domain experts disagree or have a conflict about the meaning of some competency. Divergent knowledge positions appear as a natural consequence when people collaborate in order to come to a unique common understanding. Divergence arises because of differences among individuals. Individuals' experiences, personalities, and commitment become the potential for conflicts. However, conflicts should be seen as an opportunity to negotiate about the subtle differences in interpretation, revise their positions, and finally come to a shared understanding disposed of any subjectivity.

To summarise, in the HR domain, competencies are the principal knowledge artefacts that are to be formalised. The main problem with this kind of knowledge is that these artefacts are to be elicited from tacit knowledge from individuals, or from more explicit forms such as written statements or documents. This tacit and explicit knowledge, however, is subjective due to the personal experience and background of the individuals, and the fact that currently no standard exists for explicating competencies. A way to get rid of subjective definitions is to conduct the competency elicitation process as a collaboration between domain experts. Conflicts that arise where experts disagree should be considered as an opportunity to negotiate about the conflicts, and ultimately come to a unique shared understanding.

3.3.2 The Challenge of Representing and Exchanging Competence Definitions

With the increased focus on lifelong learning and the development and assessment of competencies in the labour market, the learning and knowledge domain is slowly beginning to provide functional bridges between education, industry and public employment. This societal realignment of interoperability requirements poses new challenges for the design and implementation of technology standards relating the learning and human resources domains. As computing environments evolve from self-enclosed, proprietary, monolithic systems toward a service-oriented architecture (SOA), the challenge involves developing XML standards to support these functional bridges.

One of the most prominent areas in need of standardisation for such cross-domain communication and functional synergy is that of *competencies*. There are currently several standard specifications which each originated in their own community of standards.

The IMS (http://www.imsglobal.org) Reusable Definition of Competency or Educational Objective (RDCEO) specification was based on a draft from the IEEE Learning Technology Standards

Committee (LTSC). Released in 2001, this specification was aimed primarily at the learning management domain. In a completely different arena, the HR world, the HR-XML consortium (http://www.hr-xml.org) developed a specification for competency records, aimed primarily at the recruiting and employee selection domain.

Fast forward to 2005. The RDCEO specification is back in the accredited IEEE standardisation process, where the LTSC is using it as the basis for the Reusable Competency Definitions (RCD) standard draft P1484.20.1 The HR-XML consortium has established a liaison with the eLearning community and is developing new specifications for XML encoding of competency information that incorporate the Reusable Competency Definitions concept and that can be extended to the areas of assessment and learning management. Various national and European initiatives have worked on profiles for the RDCEO model, or built competency inventories that are compatible with this model.

There are also many 'out-of-band' competency approaches implemented by the European Public Employment services, which currently are slowly moving towards HR-XML compliance and pressuring the HR-XML group in addressing their requirements.

Being a semantic rich environment, the RCD and related standards have been hampered by the lack of a semantic underpinning. When it is considered in a perspective of lifelong learning, and the attendant requirements for a lifelong competency framework, this situation is likely to get worse. Take for instance the concept of a Curriculum Vitae (CV). In learning technology, an embodiment of this concept can be found in the IMS ePortfolio specification. In the HR world, HR-XML has two different specifications that use a CV-like concept (or resumé), but even within that consortium those specifications are not aligned properly.

What semantic technology could offer is a conceptual layer that bridges these related standards specifications, and to which they could all ontologically commit. As such, the semantic specification model would add meaningful interoperability to the other specifications that are based on simpler data models and instances that focus on the syntax of the data for exchange.

3.3.3 The Sustainability Challenge

One of the key challenges for human resources development is how to prepare employees for changing requirements in time. The increased dynamics of change has put a lot of pressure on HR, and ontology-enabled HR promises to be better prepared for this accelerated world. But changes also pose severe challenges to ontology engineering processes. Usually, these processes are designed as more or less one-time processes, and not as continuous processes drawing immediately from practical experiences. If semantic technologies are to succeed in solving problems in the HR domain, they have to deal with sustainability issues, among which the following appear to be the most crucial issues:

- *Embed modelling into business processes*. Methodologies for modelling competencies and semantic relationships need to be aware of business processes and need to take into account that model maintenance has to be work-integrated.

- *Close the loop*. What is probably even more important is a closed loop approach in which the adequacy of models and their impact on business performance is measured in a differentiated way, so that it is possible to detect areas of improvement or even trends.

- *Consider different levels of formality*. Semantic technologies usually rely on formal models that are machine processable. But in reality, conceptualisations like competencies evolve from informal descriptions to formal definitions only gradually. Methodologies, but also tools need to consider this, e.g. in the area of competence modelling.

This challenge can thus only be met through interdisciplinary approaches combining technical and organisation issues.

3.3.4 The interoperability challenge (data exchange across applications)

When exchanging competency and other HR information between disparate applications, a new challenge arises. Standards describing competencies have been established -- these facilitate data exchange between applications (see above). However these do not allow the different applications to perform automated tasks on the data. This is due to the fact that the RCD standard is purely a syntactic device, which provides a structure with which users can describe and define the competency in natural language. This enables applications to display the competency consistently, but does not provide semantic information. The standard is prepared for statements or references to outside repositories, which could be ontologically mapped; however, this is not formally specified in the standard.

The proposed but not yet finalised Simple Reusable Competency Mapping[23] incorporates certain logical inferences and relationships between different RCDs, but still does not specify a uniform way to create relationships between the semantic information contained "inside" the RCDs. Relating this to Figure 1, it signifies that the two upper layers of the Tim Berners-Lee layer cake are missing, hence semantic relation building and rule based inferences are not possible. This highlights the need for an intermediate competency / HR ontology standard, which can be used to map the semantic information of the data structures to enable automated comparisons and inferences.

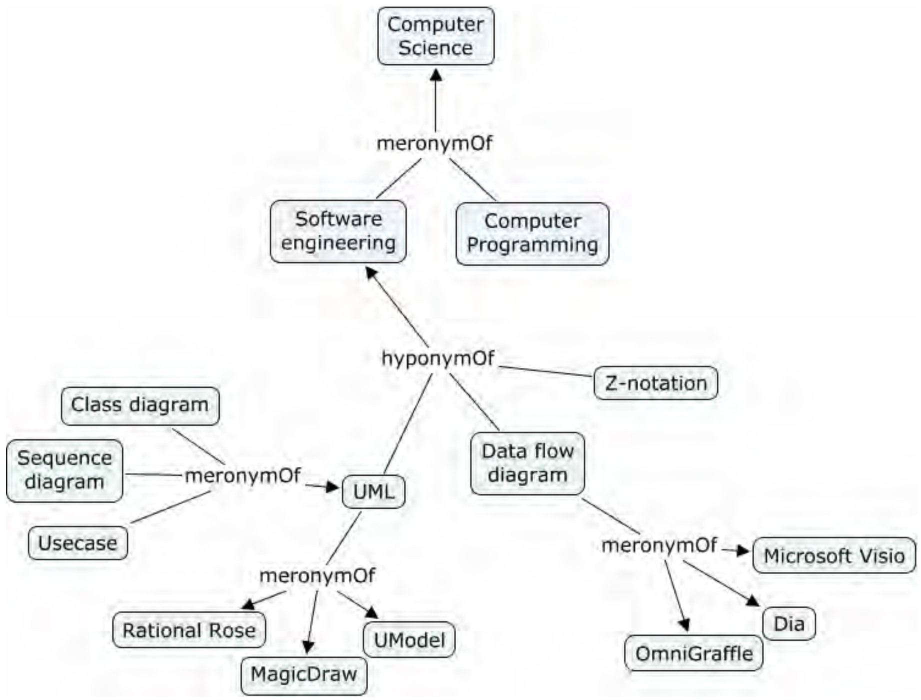


Figure 4: An example of a prototype ontology

The type of automation that would be possible can be exemplified by using the small prototype ontology in Figure 4. Let us imagine that an RCD semantically contained the definition of what "use cases" are and how to use them, and another RCD contained the skill of using "Rational Rose". Then an ontology-driven application would be able to automatically infer that the "Rational Rose" RCD would semantically cover the "use cases" RCD, because "UML" is part of "Rational Rose" (via the linguistic relation *meronym*) and "use cases" are part of "UML". Hence in a skill gap analysis or comparison of two different competency profiles, this relationship could be explored creating a more "intelligent" and useful result.

This problem is especially evident when considering semantically overloaded information such as competencies. However it is also a problem with relatively simple HR data such as names, addresses and former employers, when this information is encapsulated in data structures which are not ontologically mapped. For instance, if there is a need for exchanging information between one type of CV and another, then this cannot be done automatically because the fields might be described

[23] http://www.ostyn.com/rescompetency.htm#props

differently. Ontologies from the Dublin Core Metadata Initiative[24] and the Friend of a Friend (FOAF) Project[25] could be used to support the automation of these processes.

3.4 Existing Tools, Technologies and Methodologies

3.4.1 HLT-based Applications

Human Language Technology (HLT) applications such as automatic metadata creation, data annotation, ontology creation and population can play an important role in the integration of ontologies in the HR domain. Semantic metadata forms one of the key mechanisms through which data and ontologies can interact, by linking instances in the text with concepts in the ontology. Most of the research in human language technology in the last decade has focused on reuse of data resources, especially since the rise to the fore of machine learning techniques for natural language processing tasks, which require very large volumes of training data in order to achieve high performance levels. Clearly the enormous increase in the volume of data available on the internet has also played an important role, making it much easier to reuse data and to avail oneself of large training corpora and reference data (such as lexicons, dictionaries, ontologies, etc.).

In recent years, the increasingly large-scale and technologically significant nature of language processing science has placed increasing burdens of an engineering nature on research and development workers seeking robust and practical methods in the field of HLT. Similarly, the increasingly collaborative nature of research in this field puts a large premium on software integration and interoperation. As a result of this, the last decade has seen a number of significant systems and practices being developed in the field of architectures for HLT (see for example [6]). It is notoriously difficult to build conventional software systems in an explicit and systematic way compared with other kinds of engineering tasks[30].

Despite these advantages, the domain of HLT suffers from a major drawback: the difficulty in making the transition between research prototype and real world application suitable for use in an industrial setting. This is largely because of the nature of language itself and ensuing inherent difficulties with natural language processing tasks, such as incompleteness, language change and so on. As discussed in [25] language processing tasks only become really accurate and usable when they are tightly focused and restricted to particular applications and domains. Figure 5 below shows a three-dimensional tradeoff graph between generality vs. specificity of domain, complexity of the task, and performance level. From this we can see that the highest performance levels are achieved in language processing tasks that are focused on a specific domain and that are relatively simple (for example, identifying named entities is much simpler than identifying events).

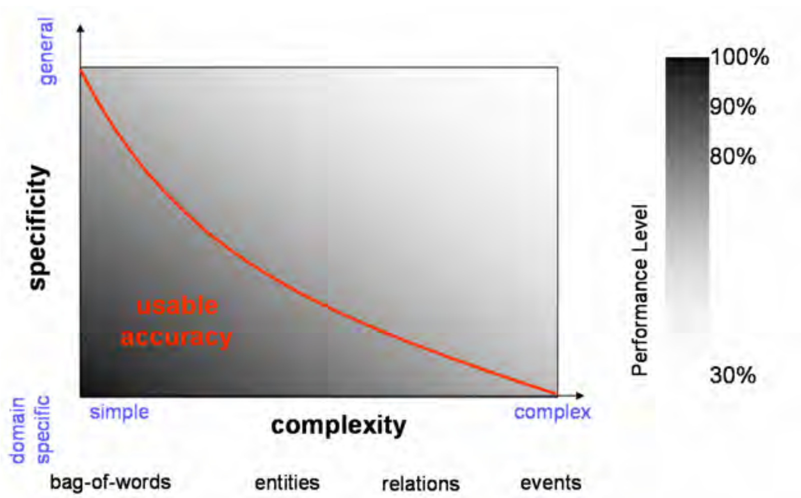


Figure 5: Tradeoff between specificity and complexity for language processing tasks

[24] http://dublincore.org/

[25] http://www.foaf-project.org

In order to make feasible the integration of semantic web applications, there must be some kind of understanding reached between ontologists and HR experts as to what constitutes a reasonable expectation. For example, applications involving HLT may not be realistically usable in the real world as standalone automatic systems, unlike other kinds of semantic web applications. Most HLT applications are designed to *assist* a human user rather than to perform the task completely autonomously. There is often a tradeoff between the amount of autonomy that will most benefit the end user. For example, information extraction systems enable the end user to avoid having to read in detail hundreds or even thousands of documents in order to find the information they want. On the other hand, the user has to bear in mind that the system will probably not be 100% accurate, and it is important for the design of the system to be flexible in terms of the tradeoff between precision and recall. For some applications, it may be more important to retrieve everything, although some of the information retrieved may be spurious; on the other hand, it may be more important that everything retrieved is accurate, even if some things are missed. Similarly, the user must be aware that such systems may require some human interaction, for example in the form of post-editing the results or of tweaking the system appropriately. This may require specialist domain, linguistic or computational knowledge. It is therefore of paramount importance that for all kinds of application (not just those in HLT), the applications should be designed where possible with a particular goal and a particular usage scenario and type of user in mind and that end-user/developer interaction should take place throughout the system development. *Communication* here is the key to successful integration.

This leads us to the idea of *collaborative* development. There has been focused research in some recent EU projects on tasks such as collaborative ontology development, ontology alignment and conflict resolution, and evaluation of ontology content. A recent overviews of the state-of-the-art in ontologies and related methodologies can be found in [15]. The recently-started NEON project in particular focuses specifically on problems such as interoperability of ontologies, collaborative development, and the ontology lifecycle as a whole.

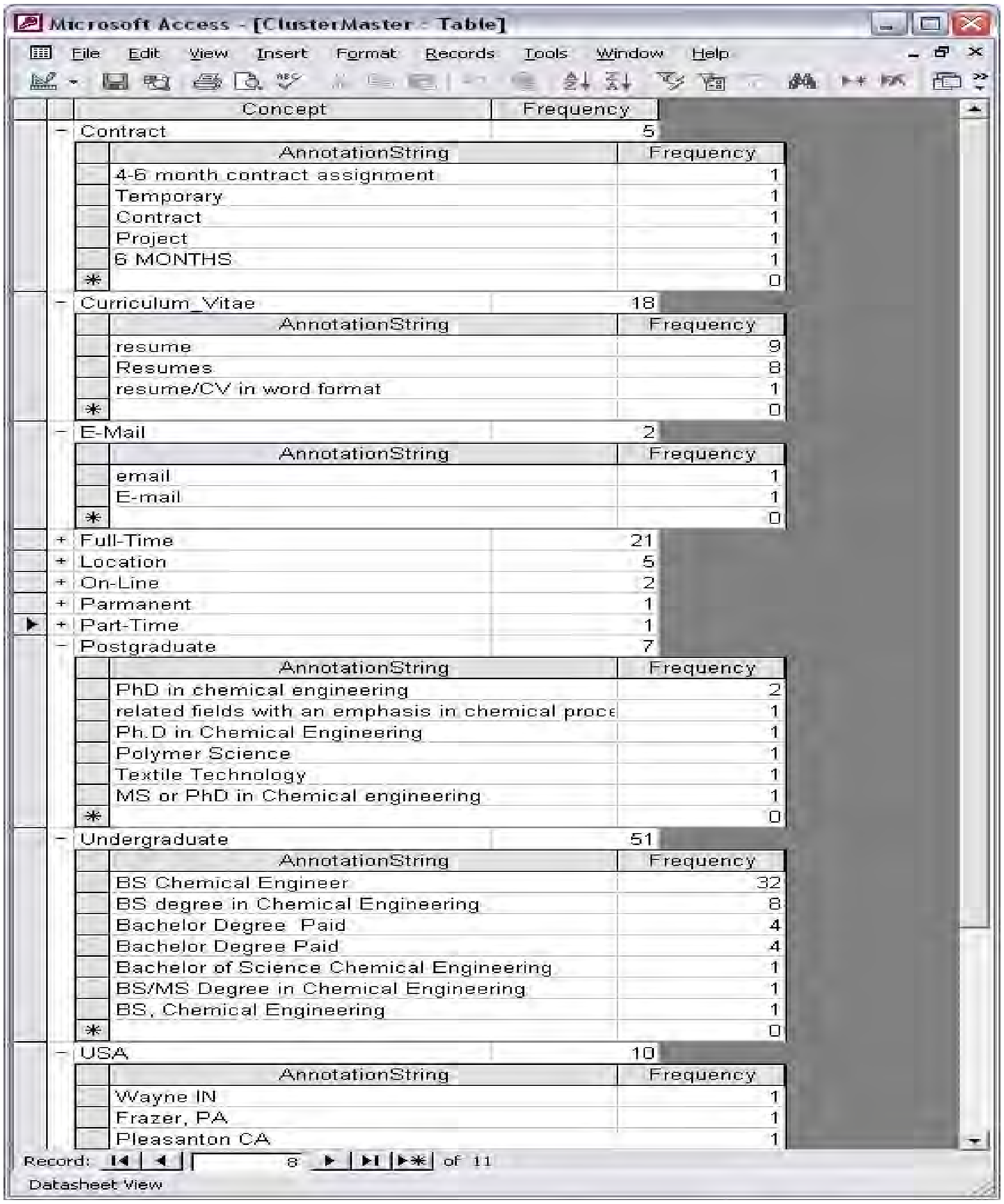
Current problems include:

- the integration of results of ontology learning and collaborative ontology development;
- determining the specific roles of evaluation in the different phases of the lifecycle and respective appropriate methods of ontology quality assessment;
- proper placement and utilisation of ontology alignment services within the lifecycle.

Further discussion of best practices in the HR domain can be found in [25].

3.4.2 GATE

GATE is an architecture for language engineering developed at the University of Sheffield [7]. It is freely available for research purposes and is used by thousands of users worldwide. GATE comes with a default set of tools for language analysis, ANNIE, and also many plugins suitable for processing ontologies and texts. In the context of the Semantic Web, it contains ontology-based tools and plugins that enable a user to manually or automatically add annotations to texts in the form of concepts or instances from ontologies, and to populate ontologies with instances from texts.



Microsoft Access - [ClusterMaster : Table]

Concept	Frequency	AnnotationString	Frequency
Contract	5		
		4-6 month contract assignment	1
		Temporary	1
		Contract	1
		Project	1
		6 MONTHS	1
		*	0
Curriculum_Vitae	18		
		resume	9
		Resumes	8
		resume/CV in word format	1
		*	0
E-Mail	2		
		email	1
		E-mail	1
		*	0
Full-Time	21		
Location	5		
On-Line	2		
Parmanent	1		
Part-Time	1		
Postgraduate	7		
		PhD in chemical engineering	2
		related fields with an emphasis in chemical proce	1
		Ph.D in Chemical Engineering	1
		Polymer Science	1
		Textile Technology	1
		MS or PhD in Chemical engineering	1
		*	0
Undergraduate	51		
		BS Chemical Engineer	32
		BS degree in Chemical Engineering	8
		Bachelor Degree Paid	4
		Bachelor Degree Paid	4
		Bachelor of Science Chemical Engineering	1
		BS/MS Degree in Chemical Engineering	1
		BS, Chemical Engineering	1
		*	0
USA	10		
		Wayne IN	1
		Frazer, PA	1
		Pleasanton CA	1

Record: 8 of 11

Datasheet View

Figure 6: Screenshot from the h-Techsight application

In recent years, GATE has enabled the development of various applications in the HR domain. The core technology of GATE and ANNIE is used for the information extraction components of the JOCI system used by Innovantage, as described earlier. A recent EU project, h-TechSight, comprised an ontology-based information extraction system developed in GATE to identify key components from online job advertisements such as skills required, working conditions (pay scales etc), job locations, etc. in order to monitor automatically different facets of the job market over time. For example, a user could select to watch the appearance of particular types of jobs in a selected geographical region, or to monitor what kinds of qualifications were being demanded for a particular kind of job, or simply the ratio of jobs being offered by different companies over time. Figure 6 depicts a screenshot of a database produced from analysing some job advertisements in the chemical engineering domain over a particular month. It shows the concept types from the ontology, such as "Contract", "Postgraduate" (qualifications), etc., and the annotations (instances) found in the advertisements which have been linked to the correct concept by the system, together with their frequency of occurrence. These figures can then be added to a larger temporal database and monitored over time. For example, in the month shown, there were 7 mentions of postgraduate qualifications: 3 for a PhD in Chemical Engineering, 1 for an MSc or PhD in Chemical Engineering, 1 for a qualification in Polymer Science, and so on. The system also allowed similar instances to be grouped together, for example “email” and “E-Mail”.

GATE is particularly relevant to the problems of interoperability and reusability, by providing a common framework within which applications can be developed, and a set of core resources and plugins which can be reused and/or extended as necessary. It is also fully Unicode-compliant and addresses very well the problem of multilinguality (see for example [23]) as it enables integration of components in different languages, and its core components are either non-language specific (such as gazetteer lookup) or can be easily be adapted to new languages with minimal effort (e.g. tokenisation and part-of-speech tagging components).

3.4.3 DOGMA

DOGMA (Developing Ontology Guided Methods for Applications) is STARLab’s research initiative, where various theories, methods, and tools for ontology engineering are studied and prototypes developed. DOGMA articulates an ontology in two distinct layers, which is called the principle of double articulation [39].

The first layer is the lexon base and consists of context-specific plausible fact types called lexons. A lexon is described as a 5-tuple <V, term1, role1, role2, term2>, where V is the context identifier and term1 plays the role role1 w.r.t. term2, while conversely term2 plays the role role2 w.r.t. term1. An example will make this easier to understand: <Human Resources, Secretary, requires, are essential for, Good typing skills>. Each (context, term)-pair lexically identifies a unique concept, e.g. (Human Resources, Secretary) points to the concept SECRETARY[26] (stored in the Concept Definition Server and described unambiguously by a gloss). This link to a concept transforms a lexon into a language- and context independent meta-lexon.

The second layer defines specific (i.e. application-dependent) interpretations and is called the commitment layer. It mediates between the lexon base and the applications. The commitment layer consists of a finite set of axioms that specify which lexons are to be interpreted, how the application maps its vocabulary to the ontology, and which rules and constraints govern the use of these lexons by the application in this interpretation. An example of such a rule/constraint would be: a Secretary **must have at least three** Language skills.

DOGMA has three major benefits. First, it is an approach to ontology engineering which is not restricted to a specific ontology language (e.g. RDF or OWL). Once the elicitation process is finished, and the ontology formalised, the DOGMA tools (DOGMA Studio Server, DOGMA Studio Workbench) can output the information to the requested language. Existing ontologies can be converted from their representation language to DOGMA, so they can be maintained and updated using the DOGMA Studio toolset.

The second benefit is DOGMA’s grounding in the linguistic representations of knowledge. In this way, domain experts and knowledge engineers can use ordinary language constructs to communicate and capture knowledge. As the main input for ontologies is domain knowledge, this is very important. Domain experts should not have to tackle language issues or learn to think in a new paradigm: the complexity of just capturing knowledge is difficult enough already.

The third DOGMA benefit is its strict separation between conceptualisation (i.e. lexical representation of concepts and their relationships) and axiomatisation (i.e. semantic constraints). This separation results in higher re-use possibilities and design scalability. It also eases ontology engineering, as the complexity is divided and agreement can be more easily reached.

DOGMA tackles the challenges for interoperability (because ontologies are essential to solve interoperability issues), multilinguality (by using the conceptualization in the Concept Definition Service on top of the lexical representation) and reusability (by providing ways to build and capture domain knowledge over two layers, which increases the potential for reusability).

[26] Note that we try to make the distinction as a concept by using capitals. The real conceptualization comes from the reader, who can recall a mental image of what a secretary is and does.

3.4.4 DOGMA-MESS

DOGMA-MESS (DOGMA Meaning Evolution Support System) is STARLab's technology (and tool) to support inter-organisational or community ontology engineering [10]. The main focus in DOGMA-MESS is how to capture relevant inter-organisational commonalities and differences in meaning. It provides a community grounded methodology to address the issues of relevance and efficiency.

In DOGMA-MESS, there are three user roles: (1) Knowledge Engineer, (2) Core Domain Expert and (3) Domain Expert. The task of the Knowledge Engineer is to assist the (Core) Domain Experts in their tasks. Most of the knowledge is captured by the Domain Experts themselves. The Core Domain Expert builds high-level templates in the so-called Upper Common Ontology[27]. The Domain Experts specialise these templates to reflect the perspective of their organisation in their Organisational Ontologies. The Domain Experts are shielded from complexity issues by assigning specific tasks in the elicitation process (e.g., specialize the "Subtask" template for "Baking"). In every version of the process, common semantics are captured in the Lower Common Ontology[28] while organisational differences are kept in the Organisational Ontologies. Information in the Lower Common Ontology is distilled from both the Upper Common Ontology and the Organisational Ontologies using meaning negotiation between (Core) Domain Experts. The Lower Common Ontology is then used as input for future versions in the process. Initial user tests of DOGMA-MESS showed promising results [10][5] in the first version of the methodology and the tool.

The importance of DOGMA-MESS is that (1) it allows the domain experts themselves to capture meaning, (2) relevant commonalities and differences are identified and (3) every version in the process results in a usable and accepted ontology.

DOGMA-MESS tackles the following challenges; (1) competency exchange, because it allows ontology-based annotation of these competences, which makes them formally specified, (2) interoperability, for the same reason and (3) reusability, again for the same reason and also because it creates involvement from many different stakeholders (people from different involved organisations). This involvement will improve the reusability of the generated content.

3.4.5 Learning in Process: Context-Steered Learning Framework

Within the EU project LIP (Learning in Process)[29], a methodological and service-oriented technological framework was developed that is geared towards integrating learning activities into work processes [29][26]. For that purpose, the system observes the (potential) learner's work activities, while they interact with their everyday applications. The system deduces from its domain knowledge and the learner's knowledge potential knowledge gaps. For these gaps, the system can compile small learning programs from available learning resources and recommend them to the learner, who can decide whether to learn now, to postpone it, or to discard the recommendation completely. This type of learning support is called context-steered learning, which is between the extremes of self-steered and course-steered learning (see Figure 7).

[27] Upper because it must be specialised further, Common because it specifies meanings used in all the different organisations.

[28] Lower because it has a specialised meaning, as opposed to the more general meaning in the Upper.

[29] http://www.learninginprocess.com/

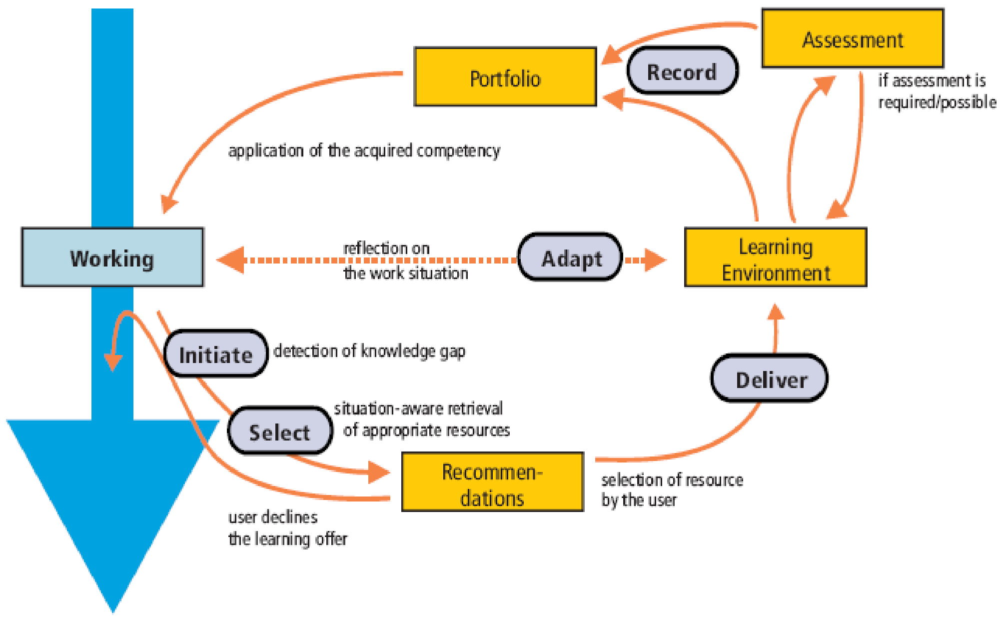


Figure 7: Context-Steered Learning [Schmidt 2006a]

The conceptual and technical enablers for this methodology are the following:

1. a thorough competency-oriented modelling approach (see description of LIP ontology below) that allows for connecting the worlds of business process, knowledge management and e-learning;
2. a flexible user context management infrastructure that captures from various sources the work processes and performances of the user;
3. a set of added-value services for competency-based operations (like gap analysis, on-demand compilation of learning programs from fine-grained learning objects etc.).

Through the specifically developed ontology-centred service-oriented architectural style, interoperability on user context and competencies has been achieved between a wide range of applications, ranging from learning management systems, communication servers, desktop applications up to HR and ERP systems.

3.4.6 The eCCO system

eCCO (eCompetences and Certifications Observatory) is a project which commenced in 2004 and is promoted by the two largest Italian ICT associations (AICA and Federcomin) and the Technical University of Milan (Fondazione Politecnico di Milano), under the aegis of the then Italian Ministry of Technology and Innovation. It aims at satisfying the needs of transparency, comparability, information and guidance expressed by the European Commission and claimed by several local players with regard to ICT competences and job profiles; in fact, no common reference ICT competence and qualification systems had been developed yet nationwide.

The eCCO Information System is an eCompetence Management Tool Based on Semantic Networks, built on the awareness that an ontology defines a common vocabulary and semantic rules for communities who need to share complex sets of information in a domain, In this scenario, the system allows users to identify ICT professional profiles that better fit their competences and to help them create a personal profile.

The system is based on the concepts of knowledge object, skill, competence, profile and semantic network as follows:

- **Knowledge:** the set of know-what, know-how and know-why;
- **Knowledge Object (KO):** a “small enough”, self consistent set of knowledge (with respect to specific areas of analysis, targets, objectives, etc.);
- **Skill:** KO put into action, KO + Action Verb (AV): to be able to do something;
- **Competence:** a skill in a specific context (Cx) of activity, KO + AV + Cx: to be able to do something with respect to a specific context of activity;
- **Performance:** a set of observable behaviours producing an objective result;
- **Job profile:** a set of competences related to the concept of key performances, expected results; in the eCCO tool it is represented as a sequence of knowledge objects and skills.

The semantic network depicted in Figure 8 links knowledge objects and skills that belong to different profiles. The nodes are KO and KO+AVs, the arcs are “IS-A” and “requirements” relations. A dictionary makes the network stronger, with the possibility to choose both words and verbs, from not only synonyms, but also different languages.

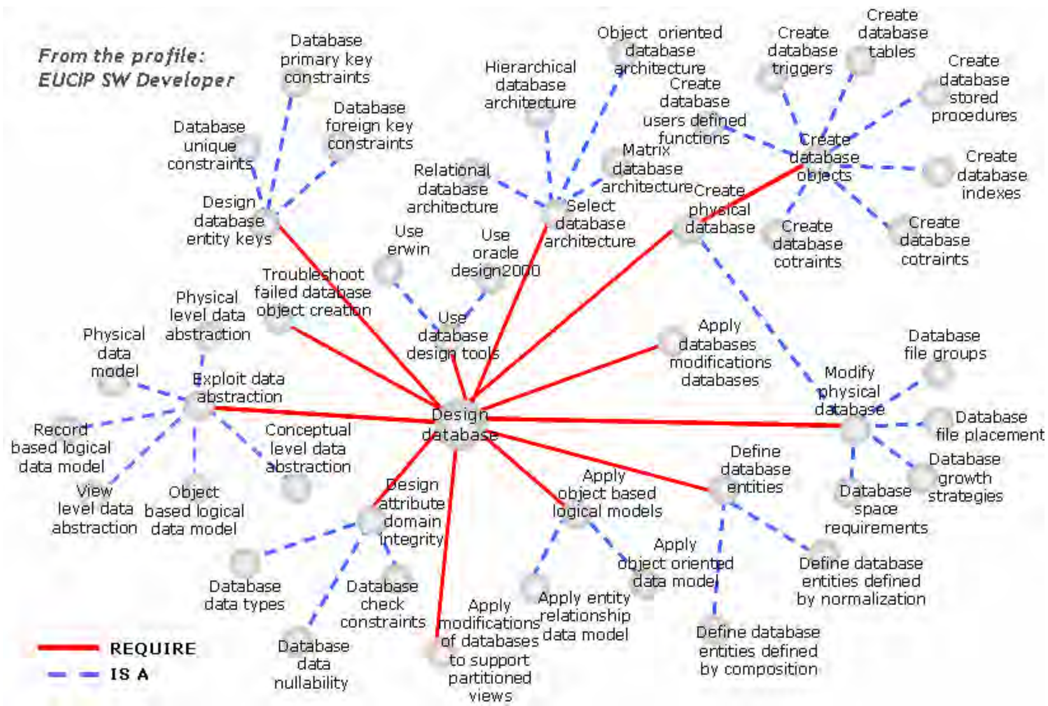


Figure 8: Example of a semantic network of knowledge objects

In order to satisfy the needs of flexibility and integration, the eCCO system allows the construction of different job profiles, starting from the same network of knowledge, skills and competences. Moreover, it provides both top-down and bottom-up approaches for new knowledge, skills and competence identification: that is, knowledge, skills and competences can be detected starting from the business processes analysis by expert teams as well as from the experiences declared by individual users of the system. In fact, users are allowed to add into the network their skills and competences not found in the system and to make connections between them. The network administrators will further validate the items and links suggested by users. In this way, new competences already informally grown inside ICT communities of practice can be input into the network stream.

At present, the system contains EUCIP profiles and also profiles coming from Europe (AITTS, SFIA, CIGREF), so it is possible to make comparisons between local and international frameworks.

The pilot eCCO system is currently used by companies for defining the mutual roles inside a project and to transparently decide what competences any partner (vendor and buyer) can make available, hence for evaluating suppliers’ competences; selling own company competences to clients and determining an objective competence-based quotation of human resources at clients; understanding new emerging ICT job profiles.

In the next months, the eCCO project will develop a tool for the interoperability of the eCCO system with other similar systems in order to build homogeneous ICT career paths. In order to satisfy that need, the European Leonardo da Vinci project “EURO ICT Lane - Towards a shared European language for ICT Qualifications and Competencies” is now in progress with the purpose to develop a shared model to read and to understand the different ICT qualifications offered by European countries and by the main ICT qualification suppliers; to provide a guide to compare and to evaluate each ICT qualification and to give ways of designing and performing new ICT qualifications.

3.5 Existing standards

HR-XML is the most widely accepted standard for information representation in the HR domain, and at least subsets of it are supported by major real-world HR applications. As an XML-based approach, it concentrates on information representation issues rather than conceptual issues, but its components provide a comprehensive overview of the administrative part of human resource management. The standard is actively being further developed by the HR-XML consortium[30].

The HR-XML consortium has also built up a library of more than 75 interdependent XML schemas which define the data elements for particular HR transactions, as well as options and constraints governing the use of those elements:

- HR-BA-XML was developed by the German Federal Employment Office and is a German extension of the international HR-XML standard. The categories defined in HR-XML were supplemented on the basis of German employer requirements.
- HR-XML-SE is a Swedish standard which consists of the original HR-XML parts (transformed from DTD's into schemas), to which schemas with Swedish extensions are added.

3.6 Existing projects and initiatives

3.6.1 Prolix

PROLIX (http://www.prolixproject.org/) is a 48 month research and development integrated project co-funded by the European Commission under the Sixth Framework Programme, Priority 2 "Information Society echnologies", which started on 1st December 2005. The objective of PROLIX is to align learning with business processes in order to enable organisations to improve the competencies of their employees more quickly, according to continuous changes of business requirements. To reach this goal, PROLIX develops an open, integrated reference architecture for process-oriented learning and information exchange.

PROLIX supports a complete learning process life cycle comprising:

1. the analysis of complex business situations;
2. the identification of individual and organisational learning goals;
3. the analysis of competencies and their matching with individual skills;
4. the definition of appropriate learning strategies and the simulation of competency-oriented processes;
5. the execution of improved learning processes;
6. the monitoring of learners’ performance according to the goals defined.

The PROLIX project mainly focuses on the following challenges: (1) modelling competence ontologies, (2) representing and exchanging competence definitions, (3) interoperability and (4) reusability.

[30] http://www.hr-xml.org

In order to tackle the first two challenges, a competency model has been developed. This model (mainly developed by Synergetics) allows for relations and groupings of competencies in order to construct competency ontologies. The model is an extension of the SRCM and RCD proposals. A great deal of focus has also been put on the reusability challenge, in order to facilitate reuse through construction from existing resources or from scratch. A collection of competences in one domain can be reused easily in another domain. The model provides semantic placeholders (ready for STARLab material) as well, where an ontological annotation can be stored in order to obtain formal description, and thus provide a solution for the interoperability challenge. An initial version of the model can be found in [4]. More recent descriptions are forthcoming.

The results from this competency modelling (backed by a context ontology) will be used by other partners in the PROLIX project in order to achieve the objectives state above.

3.6.2 CoDrive

The CODRIVE project (http://www.codrive.org/) is a competency elicitation project for vocational education. It aims to develop a new competency driven approach to knowledge in vocational education, which will facilitate and innovate interoperability and matching between Learning Content Management Systems (LCMS) and Public Employment Service Applications (PESA) through intelligent competency ontology design. The CODRIVE Project is part of the Leonardo Da Vinci Community Vocational Training Action Programme (Phase 2), an initiative by the European Commission Education and Culture DG.

DOGMA-MESS (see section 3.4.4) was mainly developed in this project. As such, this project tackles all the challenges that DOGMA-MESS tackles. One of the expected outcomes of the CODRIVE project is a domain (Bakery) ontology that will make a complete alignment possible of this domain. All competencies, learning objects, tests etc. can then be aligned by linking them to the domain ontology in order to support full interoperability.

3.6.3 PoCeHRMOM

The PoCeHRMOM[31] (http://cvc.ehb.be/PoCeHRMOM/Home.htm) project aims to provide small to medium-sized enterprises (SMEs) with competency management possibilities. Existing e-HRM applications (e.g. automatic translation of job openings) are mostly relevant for companies with an in-place competency management. The development of this knowledge is a tough job, which scares SMEs away from competency management. As a result, they cannot benefit from existing (and continuously improving) e-HRM applications. The main focus of this project is to develop a common database that SMEs can use to build their own competency profiles.

Contrary to similar existing initiatives, the PoCeHRMOM database will contain multilingual information (English, French and Dutch). This lexical information will be linked to a formal, standardised representation of concepts that point to occupations, general tasks and basic competences in an ontology platform. Different knowledge patterns (such as function terms and default sentences for competences) that point to the same concept in the ontology will receive the same formal identification code. This method allows for linguistic variants as the information is coupled with a formal, standardised vocabulary. The multilingual, ontological database will be made available in an exchange format. A test case with a relevant application will prove the usability of the ontology platform for the SME.

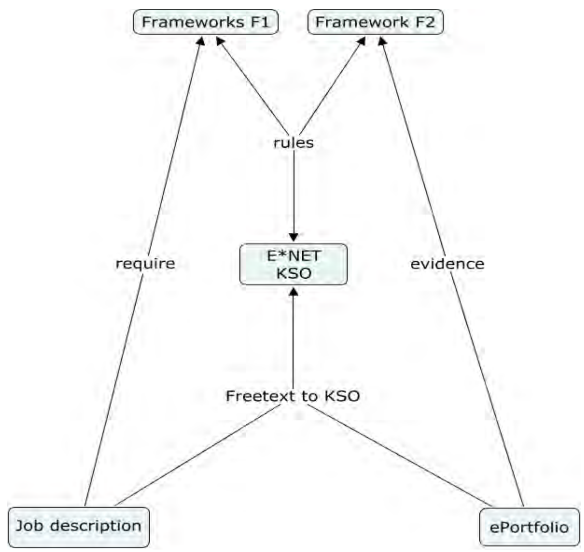
Currently, this project has collected information from several sources (O*NET, SOC classification, etc.) and linked this together into a multilingual competency ontology (challenges 1 and 4). Other challenges (interoperability and reusability) will also be tackled in this project: interoperability, as we want the data to be used by different applications; and reusability, as the content will be used and reused by several people in different organisations.

[31] PoCeHRMOM is an acronym for the Dutch phrase: **P**roject **o**mtrent **C**ompetenties en functies in **e-HRM** voor technologische toepassingen op het Semantisch Web door **O**ntologie en **M**eertalige terminologie.

3.6.4 TRACE

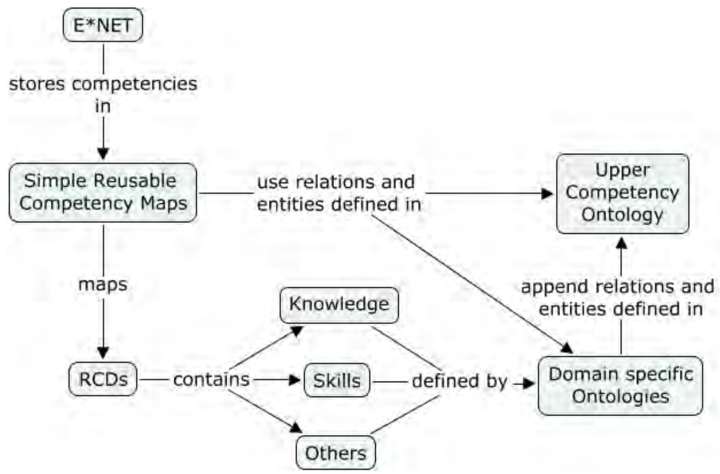
The central aim of the EU Leonardo sponsored project TRACE is to improve the transparency of competences across EU states, regions and sectors. TRACE is investigating the current state and use of existing competence systems and, from this investigation, is developing a methodology and technique to create interoperability between competence systems, especially between different competency frameworks. The project runs until December 2007, and has the following partners: University of Reading (promoting organisation), Scienter (co-ordinating organisation), Menon, Bitmedia, EIfEL, Helsinki University of Technology – Dipoli, Junta de Andalucia, , AEAE Andras, Scienter Espana and SkillsNet

One of the anticipated outcomes of TRACE is the definition of an intermediate competency "language" or description which will enable users to reference competency descriptions to a common repository of competencies, though within the scope of the TRACE project this will only be achieved for a subset of domains. The intermediate competency description has the working title E*NET, influenced by the American occupational framework O*NET. E*NET will provide a single point of reference which competency stakeholders can use when performing their tasks, whether it be transforming different frameworks or other kinds of competency descriptions. It will use common standards within the competency domain on the syntactic level, and develop an ontology of competencies on the semantic level, hence it is addressing the challenge of representing and exchanging competence definitions both on the syntactic and semantic level.



This intermediate competency description is being created using an ontological approach, where the common elements of competencies will be defined (knowledge, skills and other.) These will further be supplemented with domain-specific competencies with identified interrelations between the individual elements. Because of this approach it will be feasible to produce automated inference engines, which extends beyond simple comparison, such as tools for skill gap analysis, recruitment aids and job search guidance could be produced.

Another important feature is that anybody who wishes to extend the framework can do so, as long as the extensions are performed using the defined entities and interrelations in the upper competency ontology. Therefore it will provide an extendable basis for stakeholders.



The Reusable Competency Definition (RCD) standard will provide the syntactic level of transparency, and competency mappings will be using the RCDs as building blocks in creation of competency mappings. Tools should be created to allow users to create RCD based competencies with bindings to the common competency library, hence allowing the semantics of the library to persist even amongst user-defined competencies.

3.7 Existing Ontologies

- **ProPer Ontology.** Pioneering the use of ontologies for skills management, the ProPer ontology is probably the first (expressive) ontology for the HR domain, focusing on the issue of matching skill profiles with the help of ontological measures. It was developed at the University of Karlsruhe. The ProPer ontology is publicly available in OIL, DAML and F-Logic at http://ontobroker.semanticweb.org/ontos/proper.html; it is not being further developed, however.
- **KOWIEN Ontology.** This ontology was developed at the University of Duisburg-Esssen within the project KOWIEN, an German national project on cooperative knowledge management in engineering networks[11]. The KOWIEN ontology is not restricted to HR issues, but rather consists of a generic top-level ontology with a domain-specific profile for competence and skills management, allowing for representing and reasoning about statements about competencies of employees, mainly for the use cases expert finder and team staffing. Its strength is the formal foundation in F-Logic as an ontology formalism, however it is currently not publicly available and not being further developed.
- **Knowledge Nets.** This ontology was developed within the project Knowledge Nets at the Free University of Berlin [3]. It was based on the KOWIEN ontology and the German translation of HR-XML as well as national and international classifications for jobs and branches.
- **"ePeople".** This ontology was developed at DaimlerChrysler 2003-2006 in the context of the ePeople project (in cooperation with FZI Karlsruhe), aiming at established an integrated competence management system at Daimler Chrysler. It primarily represents Skills, Skill Profiles of Employees and Job Skill Requirements in order to allow for exploiting similarity measures on competency profiles for skill profile matching (see Figure 9, taken from [2]). The ontology was developed in KAON, an extension of the RDFS data model, in German and is not publicly available; it is not being further developed.

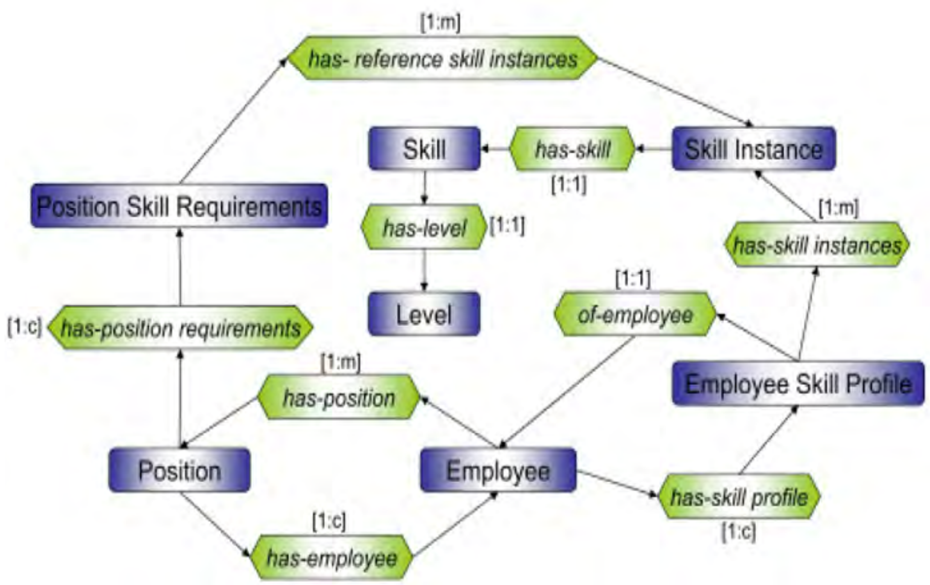


Figure 9: ePeople Ontology

- **LIP Ontology.** This ontology was developed within the EU project Learning in Process (2002-2004) in order to support embedding learning processes into work and business processes [26]. Its focus is on the automatability of on-demand learning support and is directed towards relating employees, their organisational context and relevant learning resources (which can range from learning objects up to immature documents or colleagues). It specifically aims at bridging the gap between e-learning, knowledge management, human resource development and performance support. The main idea, which is illustrated in Figure 10, is to have three major parts: learning objects (and their dependencies), a domain-specific ontology incorporating competencies and an organisational model, and users (and their social relationships). These three parts are connected via competency requirements and competency objectives respectively. The ontology was developed in the KAON extension of RDFS; it is publicly available.
- **CommOnCV** was a project concentrating on an ontological representation of CVs [30] for automatically extracting competencies from CV descriptions. The ontology was developed by the University of Nantes (France) and is not publicly available.
- **TOVE** (Toronto Virtual Enterprise Ontologies)[32] represents a set of integrated ontologies for the modelling of commercial and public enterprises It constitutes a classical and comprehensive enterprise ontology, for representing organizational structures and resources. It has a strong methodological background. The ontologies are developed by the University of Toronto; the ontology developed in first-order logic and implemented in PROLOG has extensive documentation; the machine-readable files are not publicly available.

[32] http://www.eil.utoronto.ca/enterprise-modelling/tove/index.html

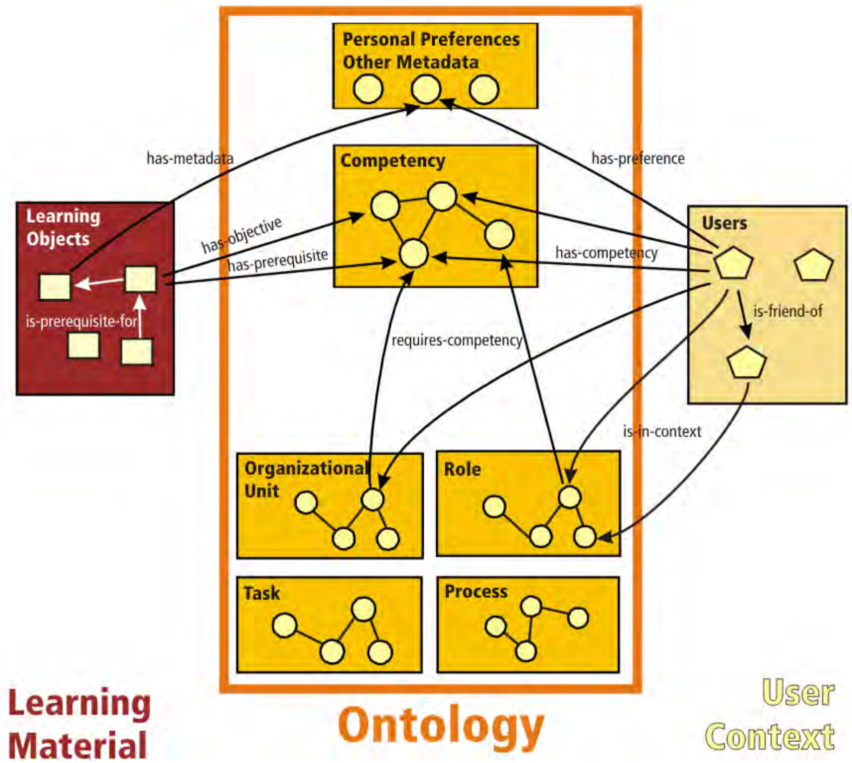
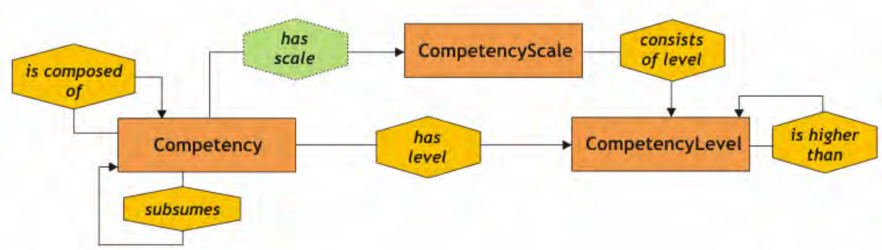


Figure 10: Schematic graphical representation of the LIP Ontology

- **Professional Learning Ontology.** This ontology [28] is the result of merging the LIP ontology with competence management approaches for improving training planning processes [19] and constitutes the successor of the LIP ontology. It tries to explicitly bring together different disciplines concerned with learning in organizations, especially knowledge management, competence management and human resource development. It furthermore tries to balance formal and informal learning. The ontology is particularly designed to distinguish between properties whose instances are expected to be explicitly collected and properties that are to be inferred (within the ontology formalism) or computed (via heuristics outside the formalism). OWL-DL has been chosen as a modelling formalism, but the major part of the ontology is also in OWL-Lite. The ontology, depicted in Figure 11, is freely available under a Creative Commons license from http://www.professional-learning.eu/competence_ontology.shtml and is actively further developed at FZI within activities like the project "Im Wissensnetz" (In the Knowledge Web), especially in the direction of community support and representation of social relationships for exploitation in informal learning activities



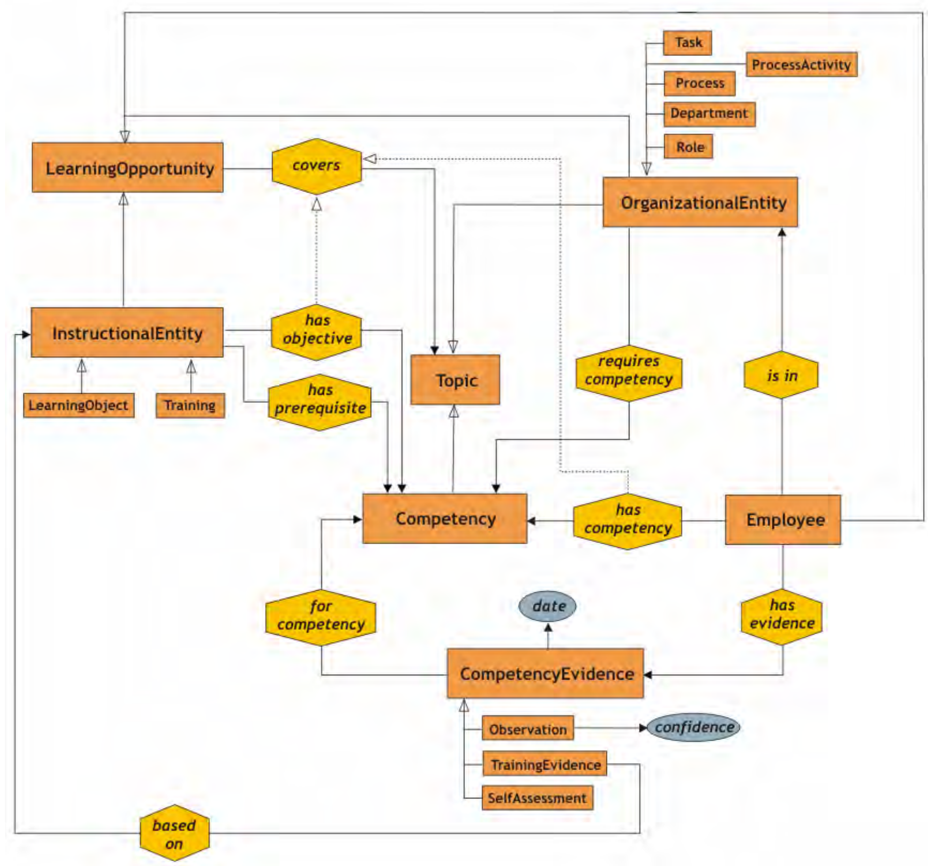


Figure 11: Professional Learning ontology

- **PROTON (PROTO-Ontology)** was developed by Ontotext Lab in the scope of the SEKT project as a light-weight upper-level ontology, which serves as a modelling basis for a number of tasks in different domains. The ontology was designed not for a fairly complete modelling of the domain, but rather for information extraction purposes for automated metadata extraction and other techniques. The ontology was developed in OWL-Lite and is publicly available from http://proton.semanticweb.org/.
- **COKE** is a three-level ontology containing a top-level Human Resources ontology (representing employees and their social groups), a middle-level Business Process ontology and a lower-level Knowledge Objects ontology [16] which are related to organizational entities. It tries to connect the organisational frame with individual knowledge objects. It is developed by the University of Calabria with DLP+ as a formalism. The ontology is not publicly available

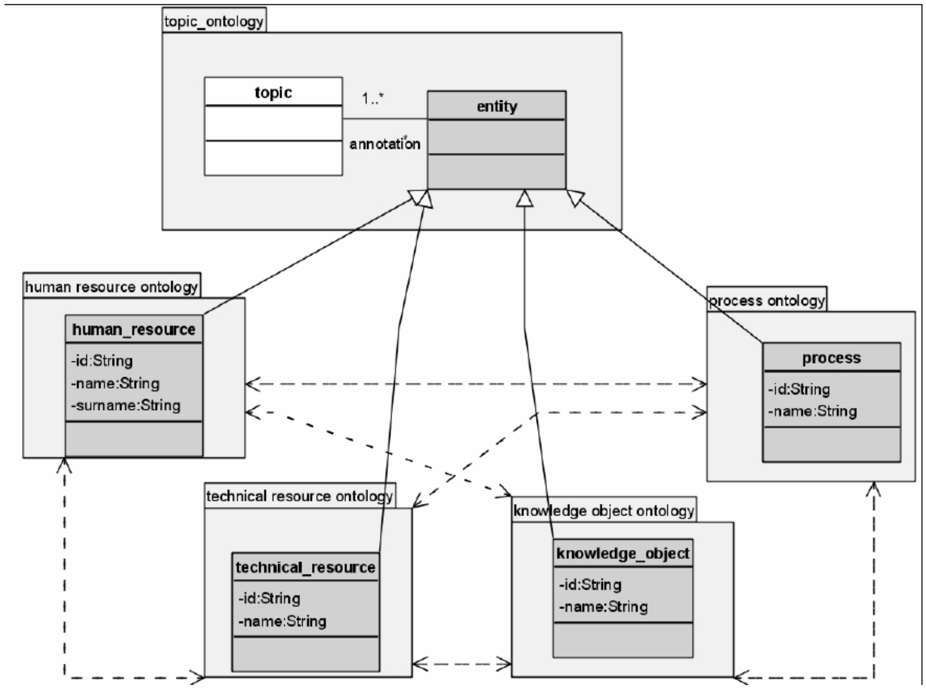


Figure 12: COKE ontology

In summary, there are several approaches to elaborating ontologies in the HR domain, each of them with a different focus. It seems promising to try to combine the strengths of different approaches like the Professional Learning Ontology, the TOVE ontologies, and the HR-XML initiative.

3.8 Success Stories

3.8.1 The CODRIVE Success Story

In this story, we report on the successful elicitation of a valid and accepted Human Resources ontology that has been built by the domain users themselves. The success in this story can be carried over to other projects for further exploitation. The system is available on demand.

The CODRIVE project is a competency elicitation project for vocational education. It aims to develop a new competency driven approach to knowledge in vocational education, which will facilitate and innovate interoperability and matching between Learning Content Management Systems (LCMS) and Public Employment Service Applications (PESA) through intelligent competency ontology design.

The CODRIVE Project is part of the Leonardo Da Vinci Community Vocational Training Action Programme (Phase 2), an initiative by the European Commission Education and Culture DG.

The success story in the CODRIVE project can be divided into two phases; namely (1) ontology creation and (2) competency annotation. We have achieved success in the first phase and are working hard to make the second phase a success as well.

In the first phase of this success story, we tackled the issue of how to obtain an ontology. The number of stakeholders is very large: all people involved in the bakery domain in the Netherlands (e.g., bakers, bakery students and teachers …). The domain knowledge is very specialised and not known by knowledge engineers themselves (as is usually the case). In order to handle the complexity, we created the DOGMA-MESS methodology and tool [10]. DOGMA-MESS divides the complexity of ontology engineering into different roles (Knowledge Engineer, Core Domain Expert and Domain Expert) in which the bulk of the work is done by the domain experts themselves (completely supported by the system). The domain experts have no need to understand complex logics or representation languages; they define and negotiate in their own domain and in their own language. The domain experts are guided and assisted by the core domain expert. This core domain experts deals with the templates (abstract knowledge patterns), while the domain experts fill these templates to create their definitions.

Figure 13 shows an example of a baker-created definition of the “Deelhandeling” (= Subtask) “Fonceren” (= Panning). The example is in Dutch as it is taken from actual data. The white (outer) boxes represent the template, built with more abstract concepts, such as “Persoon” (= Person), “Grondstof” (= Resource), “Apparatuur” (= Equipment). The blue (inner) boxes portray the actual definition, more specified than the template, e.g. “Bakker” (= Baker) and “Deeg” (= Dough). It is quite simple for the domain expert to state his knowledge in this manner. The knowledge engineer can also understand these simple facts, e.g. “Fonceren” gebruikt “Deeg” (= Panning uses Dough).

The collection of all these definitions and templates is the ontology for their domain. It becomes an interesting resource, as it is created by the domain experts themselves. It is easy to bring this kind of resource into implementation (applications, metadata, etc.) as it (1) represents correct and accepted knowledge and (2) results from and creates involvement of all stakeholders. An ontology that is created by a small group of knowledge engineers in splendid isolation and *forced* into reality and implementation has little chance of acceptance. Meetings with domain experts made it clear that *without* the ontology technology we used, the success would not be feasible to this level. An added benefit is increased understanding of their own domain for domain experts.

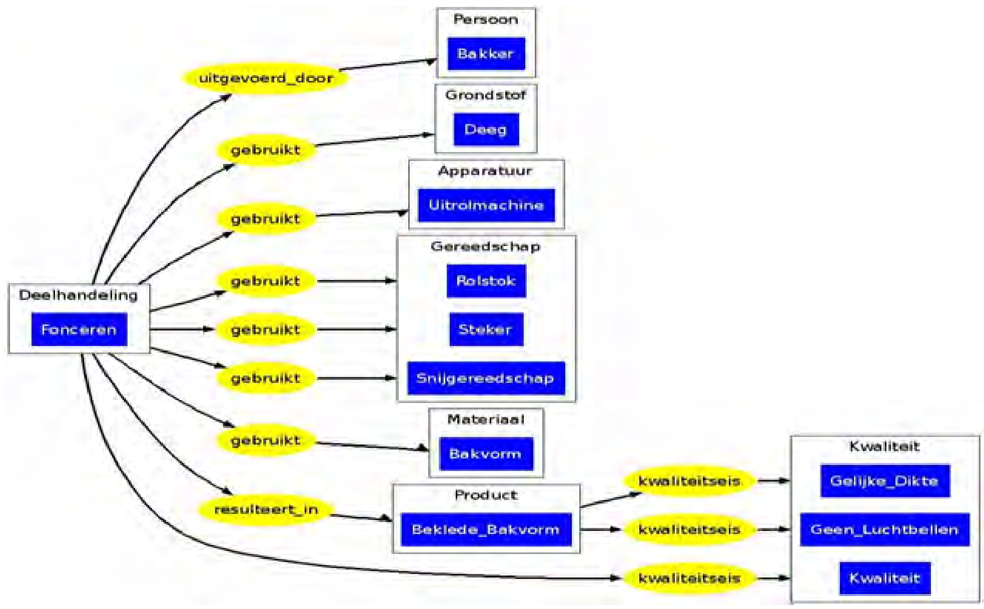
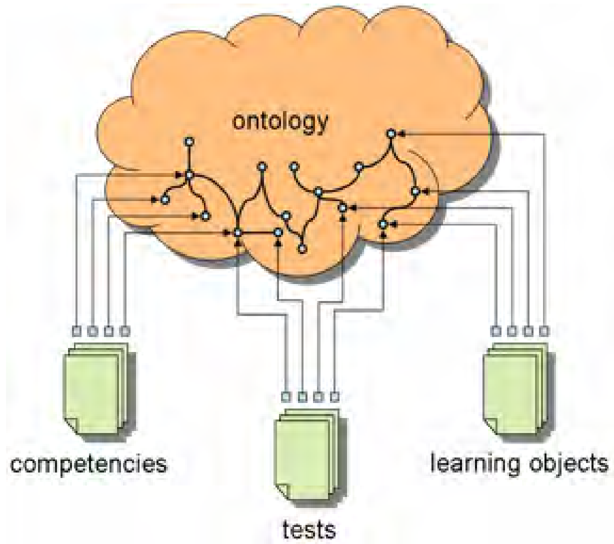


Figure 13: Example of a baker-created definition

In the second phase we need to make sure that the ontology provides results. We will use the ontology to facilitate interoperability and matching between competencies, learning objects and tests. Without the ontology, this would result in a serious linking problem, as competencies would have to be linked to learning objects and to tests, and learning objects to tests as well. This leads to a combinatorial explosion of links, which is hard both to create, and manage. Given the fact that new competencies, learning objects and tests are created all the time, and that they are subject to continuous evolution, we would have a situation that is simply not scalable.



We can solve this scalability issue, by circumventing the link problem through the ontology. This resource is much less dynamic and much less subject to evolution than competencies, learning objects and tests, since it represents the domain knowledge, and a domain is a slowly evolving entity. If competencies, learning objects and tests are linked through the ontology, the number of links decreases. This approach requires only one link between each object (competency, test, learning objects) and the ontology. A competency is then linked to a learning object because they are both linked to the same concept in the ontology (e.g. the concept of Panning). As a result, it is relatively easy to perform meaningful matching between these objects. For instance, if a competency C is linked to "Panning", and a learning object LO as well, the application can advise the user to study LO in order to obtain C.

3.9 Conclusions, Recommendations, and Future Directions

In this chapter, we have outlined the particular situation of the Human Resources domain relative to the Semantic Web. We have described why the HR domain is tough to handle, how it is moving towards a field evolving around competencies, and how information and communication technologies will introduce some major innovations in the domain. Subsequently we touched upon the subject of ontologies, essentially machine-readable knowledge descriptions which are required to bridge the gaps between organisations when they want to become much more interoperable than today (a target of the Semantic Web). A paramount aspect of ontologies is that they are shared over a relatively large domain and many organisations, which leads to a negotiation towards standardisation. Ontologies are also interesting because they remain useful even when there is no standard yet, unlike with previous attempts at the data level such as EDI.

With properly shared ontologies available, existing and newly created applications can be anchored to these ontologies in order to explicitly declare the meaning of certain elements in the application. This will relieve organisations of lengthy technical negotiations each time they want to deploy or modify interoperable systems, leading to more dynamic and streamlined business processes even for new business such as competency matching and competency bridging (E-learning) services.

While the HR domain evolves to its new competency-centric paradigm, it also needs to work hard on building the appropriate domain ontologies, and equally on keeping them up to date with the rapid changes to the HR domain itself. This is a major task which cannot be delegated to external entities, just as describing and tracking the HR domain should be left in the hands of HR experts. A number of attempts to create relevant ontologies have been made, as we have seen in previous sections. None of these, however, are directly applicable to support system interoperability, some of these are not available to the public, and many are not actively maintained. The issue therefore becomes one of negotiations between influential HR organisations to at least get some common ontology ground, moving towards a partial world standard registered with the ISO, the IEEE, or the HR-XML consortium.

In the meantime, individual organisations will need to anchor their existing applications to at least some form of existing ontology, which gives them a much better position to make their applications interoperable both among themselves and with external organizations, and to link up their knowledge frame with international standards as they emerge (knowledge sharing and re-use). Methodologies and tools, such as DOGMA-MESS, may help organisations with this work. Current work will not be wasted when a standard emerges, as the hard thinking required for ontology construction leads to business insight, not to lines of code embedded into legacy applications.

The Ontology Outreach Advisory will be a major player in this field, bringing the expertise together of many industry and research experts on both HR, ontology engineering, and the Semantic Web. With at least two international standards in progress, the OOA is uniquely positioned to guide and advance the state of the art in HR Semantic Web applications.

3.10 References

[1]E. Biesalski, A. Abecker. Similarity measures for skill-profile matching in enterprise knowledge management. In *8th International Conference on Enterprise Information Systems (ICEIS)*, 23 - 27, May 2006

[2] E. Biesalski. *Unterstützung der Personalentwicklung mit ontologiebasiertem Kompetenzmanagement*. PhD thesis, University of Karlsruhe, 2006

[3] C. Bizer, R. Heese, M. Mochol, R. Oldakowski, R. Tolksdorf, R. Eckstein, The Impact of Semantic Web Technologies on Job Recruitment Processes. *International Conference Wirtschaftsinformatik (WI 2005)*, Bamberg, Germany, February 2005.

[4] S. Christiaens, J. De Bo, R. Verlinden. Competencies in a semantic context: meaningful competencies. In: R. Meersman, T. Zahir, P. Herrero et al. (eds.), *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops (OnToContent06)*, Springer-Verlag 2006.

[5] S. Christiaens, A. de Moor. Tool interoperability from the trenches: the case of DOGMA-MESS. In *Proc. of the First Conceptual Structures Tool Interoperability Workshop (CS-TIW2006)*, Aalborg University Press, 103–118, 2006.

[6] H. Cunningham and D. Scott, editors. *Special Issue of Natural Language Engineering on Software Architecture for Language Engineering*. Cambridge University Press, 2004.

[7] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), 2002.

[8] P. De Leenheer, A. de Moor, R. Meersman. Context Dependency Management in Ontology Engineering: a Formal Approach. *Journal on Data Semantics* VIII, LNCS 4380, p. 26-56, 2006.

[9] P. De Leenheer, R. Meersman. *Towards a formal foundation of DOGMA ontology: part I*. Technical Report STAR-2005-06, VUB STARLab, Brussel, 2005.

[10] de Moor, A., De Leenheer, P., Meersman, R.: DOGMA-MESS: A meaning evolution support system for interorganizational ontology engineering. In: *Proc. of the 14th International Conference on Conceptual Structures, (ICCS 2006)*, Aalborg, Denmark. Lecture Notes in Computer Science, Springer-Verlag, 2006.

[11] L. Dittmann, S. Zelewski: Ontology-based Skills Management. In *Proc. of the 8th World Multi-conference on Systemics, Cybernetics and Informatics (SCI 2004)*, Vol. IV, 190-195

[12] F. Draganidis, P. Chamopoulou, G. Mentzas: An ontology based tool for competency management and learning paths. In: 6th International Conference on Knowledge Management (I-KNOW 06), Special track on Integrating Working and Learning, 2006

[13] A. Funk, B. Davis, V. Tablan, K. Bontcheva, and H. Cunningham. *Controlled language IE components version 2*. Deliverable D2.2.2, SEKT, 2006.

[14] Semantic Web Technologies Take Middleware to Next Level. 2002. http://www.gartner.com/DisplayDocument?doc_cd=109295

[15] A. Gomez-Perez, M. Fernandez-Lopez, O. Corcho. *Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer, 2004.

[16] Andrea Gualtieri, Massimo Ruffolo: An Ontology-Based Framework for Representing Organizational Knowledge, In *Proceedings of I-KNOW '05*, Graz, 2005

[17]T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.

[18]Bill Kastle, Michael L. George, Dave Rowlands, What is Lean Six Sigma, McGraw-Hill Professional, 2004

[19] C. Kunzmann, A. Schmidt: Ontology-based Competence Management for Healthcare Training Planning: A Case Study. In *Proceedings of I-KNOW 2006*, Graz, September 2006, Special Issue of the Journal of Universal Computer Science (J.UCS), ISSN 0948-695X, pp. 143-150

[20] T. Lau, Y. Sure: Introducing Ontology-based Skills Management at a large Insurance Company, In: *Modellierung 2002, Modellierung in der Praxis - Modellierung für die Praxis*, 123-134, 2002

[21] T. Ley, D. Albers: Identifying Employee Competencies in Dynamic Work Domains: Methdological Considerations and a Case Study. In *Journal of Universal Computer Science*, vol. 9, pp. 1500-1518, 2003.

[22] R. Maier, A. Schmidt: Characterizing Knowledge Maturing: A Conceptual Process Model for Integrating E-Learning and Knowledge Management In *4th Conference Professional Knowledge Management: Experiences and Visions (WM 07), Workshop on Convergence of Knowledge Management and E-Learning (CKME '07)*, Potsdam, 2007.

[23][Maynard03] D. Maynard and H. Cunningham. Multilingual adaptations of a reusable information extraction tool. *Proc. of EACL*, Budapest, 2003.

[24][Maynard05] D. Maynard, M. Yankova, A. Kourakis and A. Kokossis. Ontology-based information extraction for market monitoring and technology watch, ESWC Workshop "End User Aspects of the Semantic Web", Heraklion, Crete, 2005

[25] J. Pan, L. Lancieri, D. Maynard, F. Gandon, R. Cuel, A. Leger. Success Stories and Best Practices, KnowledgeWeb Deliverable D1.4.2v2, 2007.

[26] M. Polanyi. *The Tacit Dimension*. Garden City, NY, Doubleday, 1966.

[27] [Schmidt05] A. Schmidt. Bridging the Gap Between E-Learning and Knowledge Management with Context-Aware Corporate Learning Solutions. Professional Knowledge Management. *Third Biennial Conference WM 2005, Revised Selected Papers, Lecture Notes in Artificial Intelligence (LNAI)*, volume 3782, Springer, pp. 203-213, 2005.

[28][Schmidt 2005a] A. Schmidt. Knowledge Maturing and the Continuity of Context as a Unifying Concept for Knowledge Management and E-Learning In *Proceedings of I-KNOW 2005, Special Track on Integrating Working and Learning*, Graz, 2005.

[29][Schmidt06] A. Schmidt, C. Kunzmann: Towards a Human Resource Development Ontology for Combining Competence Management and Technology-Enhanced Workplace Learning In *Proceedings of OntoContent 2006 (in conjunction with OTM Federated Conferences 2006)*, Springer, Lecture Notes in Computer Science (LNCS), vol. 4278, pp. 1078-1087, 2006.

[30][Schmidt06a] A. Schmidt, S. Braun. Context-Aware Workplace Learning Support: Concept, Experiences, and Remaining Challenges In *Proceedings of the European Conference on Technology-Enhanced Learning (EC-TEL 06)*, Heraklion, October 2006.

[31][Somerville01] I. Sommerville. *Software Engineering*. Addison-Wesley, Reading, MA, 2001.

[32][Spyns02] P. Spyns, R. Meersman, and M. Jarrar. Data modelling versus ontology engineering. *SIGMOD Record: Special Issue on Semantic Web and Data Management* 31(4), pages 12-17, 2002.

[33][Sure00] Y. Sure, Alexander Maedche, and Steffen Staab: Leveraging Corporate Skill Knowledge - From ProPer to OntoProPer. In: D. Mahling & U. Reimer. *Proceedings of the Third International Conference on Practical Aspects of Knowledge Management (PAKM 2000)*, Basel, Switzerland, 2000,.

[34][Sure02] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, and D. Wenke. OntoEdit: Collaborative Ontology Development for the Semantic Web. In 1st International Semantic Web Conference (ISWC2002), Sardinia, 2002.

[35]Thomas, H., Davenport, *Mission Critical: Realizing the Promise of Enterprise Systems*, HBS Press, 2000; www.cio.com/research/erp/edit/erpbasics.html

[36][Trichet03] F. Trichet, M. Leclère. A Framework for Building Competency-Based Systems Dedicated to Human Resource Management. *ISMIS 2003*: pp. 633-639, 2003.

[37] M. Uschold and M. Gruninger. Ontologies: Principles, Methods and Applications. *KnowledgeEng. Rev.*, vol. 11 (2), 1996, pp. 93-155.

[38][Zelewski05] S. Zelewski; Y. Alan. Generische Kompetenzontologie für computerbasierte Kompetenzmanagementsysteme. In S. Zelewski, Y. Alan, A. Alparslan, L. Dittmann, T. Weichelt, (Hrsg.). *Ontologiebasierte Kompetenzmanagementsysteme – Grundlagen, Konzepte, Anwendungen.* Berlin, pp. 429-535, 2005.

[39][J05] Mustafa Jarrar: Towards methodological principles for ontology engineering. PhD Thesis. Vrije Universiteit Brussel. (May 2005)

4 The OOA-HLS chapter

Health and Life Sciences is one of the first domains in which the need for ontologies to advance science on the one hand, and to make better applications on the other hand, has been understood. The OOA-HLS chapter aims to develop, recommend and promote quality guidelines for ontology content and tools. The current focus topics of the OOA-HLS chapter include:

- Ontologies in Biomedicine and Bioinformatics.
- Ontologies of diseases, nursing, therapeutics, drug, etc.
- Upper level concepts of healthcare and life sciences ontologies.
- Semantic metadata for Clinical Data Interchange.
- Semantics of medical XML standards and vocabularies.
- Multilinguality in Biomedicine and bioinformatics ontologies.
- Best practice and semantic patterns in ontology modeling and evaluation

Please refer to deliverable D1.3.3 for a deep analysis of the role of ontologies in eHealth and how the OOA plans to position itself within this domain (where ontologies are well-understood, and many related initiatives already exist). In what follows we summarize the activities of OOA-HLS chapter.

4.1 OOA activities in the HLS domain

The OOA-HLS chapter is currently focusing on the following ongoing activities:

1- *Collection and Promotion of successful use cases*. This activity is led by Vrije university of Amsterdam. In the following section we present some of the collected use cases and our methods for outreaching them to industry.

2- *Developing and recommending a set of "quality guidelines for ontology authoring"*. This is a joint activity in a close cooperation with other international bodies in the eHealth domain, which are: The US National Center for Ontological Research (NCOR), the US National Institute of Standards and Technology (NIST), and the European Center for Ontological Research (ECOR). This activity will be presented in the next chapter.

4.2 Collection and Promotion of Use Cases

To properly describe the various ontological *challenges* in the eHealth domain and to better match these challenges with existing *solutions*, this activity aims to collect successful use cases from academia and promote them to industry, which is inline with the general mission of the OOA and with the sustainability actions of the KnowledgeWeb project. The idea is not only to collect use cases and publish them at the OOA website regularly but also to promote these use cases at the annual industrial events of the OOA. We believe this activity is not only sustainable, but also provides an efficient dialogue between industry and academia (as demos are better understood compared to research papers). Furthermore, an extra feature to the OOA website is being implemented to allow the community (i.e., anyone interested) to post comments on these use cases. This would probably bring another channel of interaction with the authors of the use cases.

Currently, there are 7 use cases published at the OOA website, and we are revising and packaging several other cases from KnowledgeWeb partners (collected in WP1.1) and others. The use cases collected in WP1.1 are originally written in a form/purpose of in-reaching research with industrial needs (i.e. testing research ideas against industrial requirements). The new version of these use cases will follow the structure of the use cases provided by the W3C Semantic Web Education and Outreach Interest Group[1]. Each use case description shall include a general description of the problem, a solution section, and the key benefits of the semantic web technology.

It is worth to note that this activity is led by Vrije University of Amsterdam (VU). VU is not only becoming an active member in this OOA chapter, but also, through its active involvement in several

W3C working groups, VU is providing an effective synergy between these groups and the OOA-HLS activities.

In the following, we present the 7 use cases collected so far. The first five cases have beencontributed by the researchers of Vrije University of Amsterdam and its research partners. Use case 6 and 7 have been provided by the Semantic Web Education and Outreach Interest Group[2].

1. Formalized Terminologies to support tasks at Intensive Care Units of Hospitals (DICE/I-Catcher). *Contributors: Michel Klein and and Ronald Cornet*
2. Using the OpenKnowledge System to Ease Re-use Algorithms in the Proteomics Domain. *Contributors: George Anadiotis, Paolo Besana, David Dupplaw, Dietlind Geldoff, Frank van Harmelen, Spyros Kotoulas, Adrian Perreau de Pinninck, Dave Robertson and Ronny Siebes*
3. Drug Ontology Project for Elsevier (DOPE). Contributors: *Anita de Waard, Christiaan Fluit, and Frank van Harmelen*
4. Protocure: Ontologies for Enriching Medical Protocols. *Contributors: Radu Serban, Annette ten Teije, and Frank van Harmelen*
5. Vague modeling for Evaluating Clinical Trials. *Contributors: Stefan Schlobach, Linda Peelen, and Michel Klein*
6. Semantic Web Technology for Public Health Situation Awareness by the School of Health Information Sciences, University of Texas, United States.
7. Semantic-based Search and Query System for the Traditional Chinese Medicine Community by Zhejiang University and China Academy of Chinese Medicine Sciences, China

4.2.1 HLS Case Study 1: Intensive Care

Case Study: Formalized Terminologies to support tasks at Intensive Care Units of Hospitals

Michel Klein, Ronald Cornet. June 2007.

General Description

Intensive Care (IC) is a complex, expensive form of care. Patient Data Management Systems (PDMS) containing information about patients and outcomes are now available that can partially support primary care and the evaluation of its quality. In the I-Catcher project we have investigated how formal terminological knowledge about IC diagnoses and semantic web technology can support the search, navigation and registration of diagnostic terms.

The problem

Registration of clinical data is of increasing importance for daily care practice and management task as well as for evaluative research. At most Intensive Care Units (ICU's), the reason for admission is registered using one ore more diagnostic terms from a controlled vocabulary. However, both management and evaluation require that *stratification* can be performed on the registered data. Stratification means separation of data into homogeneous subgroups based on certain characteristics. An example of this would be:

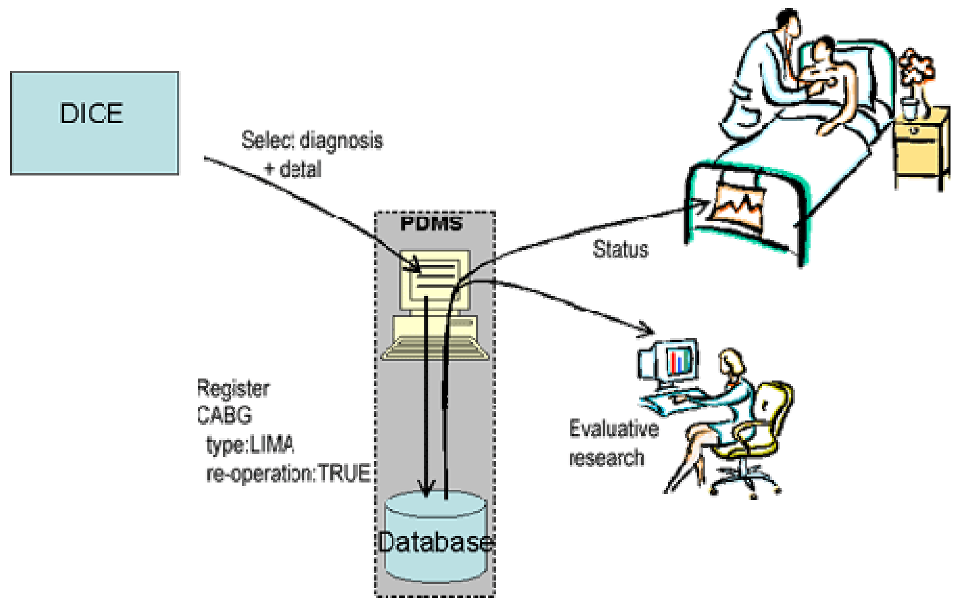
patient groups admitted to the ICU because of renal failure

In this example the selection criterium is 'renal failure'. For stratification the characteristics may be general as well as highly detailed. However, patient data registrations based on controlled vocabularies do not allow to make such specific selections. Moreover, data registrations based on different vocabularies can not be combined nor compared.

The solution

The solution to this problem is the usage of a formalized terminology system, i.e. an ontology, which represents all reasons for admission and their characteristics. The Medical Informatics group at the Academic Medical Center (AMC) in Amsterdam, the Netherlands, developed such an ontology. The ontology, called DICE, consists of 1460 'reasons for admission' and is formalized in OWL DL. In addition to the 'reasons for admissions' it contains another 2300 concepts, described by some 5000 lexical terms. These concepts are related to each other with some 4300 relational links of 50 different relation types. DICE mainly aims to cover concepts in the Intensive Care domain, and is structured in five different hierarchies (called “aspects” in DICE): abnormalities (255 concepts), medical procedures (55 concepts), anatomical locations (1512 concepts), body subsystem (13 concepts), and causes (85 concepts). Together, these five vocabularies are the main organisational structure of DICE. Each aspect has a domain of possible values, organized in a tree structured taxonomy. The concepts in the aspect taxonomies are labeled with a language attribute, in the current version either Dutch or English. If a concept is named with multiple terms, one of the terms functions as ‘preferred’ term-label, and the others as synonyms.The ontology has been improved over time by checking its consistency using the reasoning support that is available for OWL DL.

The ontology is used to register the reasons for admission for individual patients. Patients data is annotated with a class defintion from the ontology, either a pre-defined class, or a new class definition using class and property names from the ontology (so called post-coordinated terms). For example, the reason for admission of a specific patient could be described with the predefined classname Aneurysm_of_Aorta, or with the post-coordinated term CABG rdfs:subClassOf Re-Operation, dice:type hasvalue dice:LIMA (bypass surgery). On querying, the logical relation between the registered diagnoses and the diagnosis used as query is determined, and all patient data that have a more specific diagnosis is returned. This process is illustrated in the Figure below.



To compare the patient data of one hospital with the data of another, for example for quality assessment purposes, the controlled vocubulary in one hospital has to be expressed in terms of the DICE ontology. To provide a basis for this mapping ontology alignment techniques can be used. When the controlled vocabulary is mapped in this way, the DICE ontology can be used to query the data of the other hospital.

Key Benefits of Using Semantic Web Technology

Key benefits of Semantic Web technology for patient registration at Intensive Care Units include:

- The use of formal ontologies allow for dynamic selection of patient groups, which in turn facilitates daily care practice and management tasks.
- Ontology alignment techniques developed in the context of Semantic Web allow to combine and compare patient registrations from different hospitals.
- The reasoning support that is available for OWL DL ontologies help to verify the terminology system.

4.2.2 HLS Case Study 2: OpenKnowledge

Using the OpenKnowledge system to ease re-use algorithms written by biomedical and biological scientists working in the Proteomics domain

George Anadiotis, Paolo Besana, David Dupplaw, Dietlind Geldoff, Frank van Harmelen, Spyros Kotoulas, Adrian Perreau de Pinninck, Dave Robertson and Ronny Siebes . June 2007.

開

General Description

The number of different proteins that are present in a certain tissue (e.g. human liver cells) under certain conditions (e.g. after intake of alcoholic beverages) can easily reach several hundreds, or more. Characterising this contingent of proteins, i.e. identifying as many of the proteins present as possible and considering other information that is known about each of them, is crucial for biologists trying to understand the underlying regulation and adaptation of the respective biological system (here the human liver). A technologically advanced strategy to characterise proteins on a large scale involves fragmenting the proteins and the use of mass spectrometric analysis to determine the amino acid sequence of each fragment. This technique is often referred as "Proteomics" by biologist researchers. To accomplish an actual identification of each protein, the fragments are compared with the sequences stored in centrally maintained databases. This is undertaken either in-house or via WWW-servers and the chances for identification vary depending on the quality of the samples subjected to mass spectrometry. Given this limitation, proteomics experts express a great interest in gaining access to data resulting from their colleagues' research, to help them with their own analyses. Interestingly, even access to data that was of no further use to some researchers and was discarded, could be of interest to others. In this case study, we want to ease the cooperation between these researchers in the proteomics domain, by developing a system that allows to share, invoke and publish workflow descriptions and services in an open way.

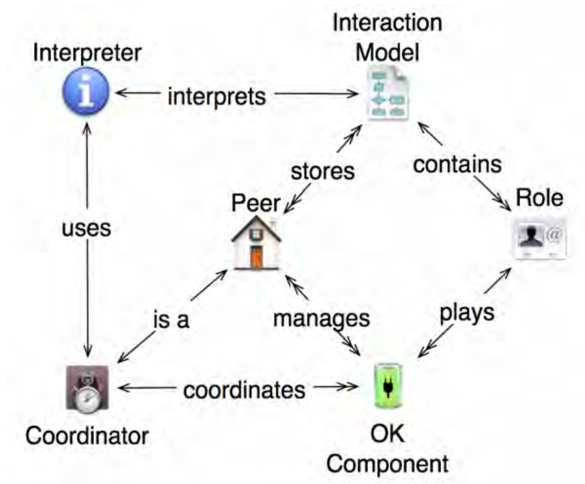
The problem

The pool of potentially available knowledge about proteomics on the Internet is huge. It is fed by the traditional Web: by application programs feeding data onto theWeb, by Web services accessed through various forms of application interface, by devices that sense the physical environment, and so on. It is consumed in a wide variety of ways and by diverse mechanisms (and of course consumers may also be suppliers). Proteomics experts express a great interest in gaining access to the knowledge resulting from their colleagues' research, to help them with their own analyses. Interestingly, even access to data that was of no further use to some researchers and was discarded, could be of interest to others. The current solutions are inadequate to enable easy re-use of the knowledge provided by peers in the field.

The solution

The aspiration of the FP6 funded OpenKnowledge project is to allow knowledge to be shared freely and reliably, regardless of the source or consumer. Reliability here is interpreted as a semantic issue. The Internet is in the fortunate situation that physical and syntactic reliability have been solved to satisfactory degrees, making semantic reliability the main challenge. Semantic reliability means that we want the meaning ascribed to knowledge that is fed into the pool, to be preserved adequately for the purposes of consumers.

Of course such "open knowledge sharing" is an aspiration that we know to be unattainable, in the strong sense where all knowledge supplied can be consumed with perfect freedom and reliability. Globally consistent common knowledge is impossible to guarantee in an asynchronous distributed system. The good news is that only a small proportion of the pool of available knowledge will be of use to any given consumer, since each must have an upper limit on how much knowledge it can process. A pragmatic aim of open knowledge sharing, then, is to obtain knowledge appropriate to the activities in which each consumer wants to engage, while maintaining free and (adequately) reliable connections between suppliers and consumers. By building a system, we demonstrate that in the proteomics scenario, sharing workflows and services at very low cost to consumers and suppliers is possible. The novelty of this system is that each interchange of knowledge is made in the context of the (shared) workflow descriptions. We then address the (unavoidable) tasks of ontology mapping, query routing, etc. using algorithms that are comparatively simple because they can (at no additional cost) use knowledge about the structure of the interaction and the ways in which it has been performed (successfully or unsuccessfully) within a peer group.



Key Benefits of Using Semantic Web Technology

Key benefits for the OpenKnowledge project include:

- Localisation of experts, Web-services and workflows in a distributed way by semantic enabled discovery algorithms
- A system in which people can develop, share and visualize via different user interfaces, the semantic descripitions of the resources (experts, services and workflows)
- Providing a shared point of access for all people interested in any field, including the proteomics domain
- Enabling re-use of mappings between terminologies used in different domains using Semantic Web technology.

4.2.3 HLS Use Case 3: Drug Ontology

Drug Ontology Project for Elsevier (DOPE)*Anita de Waard (Elsevier), Christiaan Fluit (Aduna) and Frank van Harmelen (Vrije Universiteit Amsterdam)*

General Description

Innovative research institutes rely on the availability of complete and accurate information about new research and development, and it is the business of information providers such as Elsevier to provide the required information in a cost-effective way. It is very likely that the semantic web will make an important contribution to this effort, since it facilitates access to an unprecedented quantity of data. However, with the unremitting growth of scientific information, integrating access to all this information remains a significant problem, not least because of the heterogeneity of the information sources involved - sources which may use different syntactic standards (syntactic heterogeneity), organize information in very different ways (structural heterogeneity) and even use different terminologies to refer to the same information (semantic heterogeneity). The ability to address these different kinds of heterogeneity is the key to integrated access.

Thesauri have already proven to be a core technology to effective information access as they provide controlled vocabularies for indexing information, and thereby help to overcome some of the problems of free-text search by relating and grouping relevant terms in a specific domain. However, currently there is no open architecture which supports the use of these thesauri for querying other data sources. For example, when we move from the centralized and controlled use of EMTREE within EMBASE.com to a distributed setting, it becomes crucial to improve access to the thesaurus by means of a standardized representation using open data standards that allow for semantic qualifications. In general, mental models and keywords for accessing data diverge between subject areas and communities, and so many different ontologies have been developed. An ideal architecture must therefore support the disclosure of distributed and heterogeneous data sources through different ontologies. The aim of the DOPE project (Drug Ontology Project for Elsevier) is to investigate the possibility of providing access to multiple information sources in the area of life science through a single interface.

This approach is sketched in figure 1 (the letters refer to the figure):

A. Elsevier's main life science thesaurus, EMTREE ©, has been converted to an RDF-Schema format.

B. Using EMTREE, several large data collections (5 million abstracts from the MEDLINE database, and about 500,000 full text articles from Elsevier's ScienceDirect have been indexed using Collexis Fingerprinting technology. In addition to the fingerprint (a list of weighted keywords assigned to a document) metadata about the document such as the authors and the document location are posted on the Collexis server.

C. The Collexis metadata have been dynamically mapped to an RDF model in two steps: the first transformation creates an RDF model, which is an exact copy of the data structure provided by the fingerprint server. The final model is a conceptual document model used for querying the system.

D. An RDF database, in this case implemented as a Sesame repository [1] using the SOAP protocol, communicates with both the fingerprint server and the RDF version of EMTREE.

E. A client application UI allows the user to interact with the document sets indexed by the thesaurus keywords, using SeRQL queries sent by HTTP.

F. The system is designed in a way can be extended by adding new data sources, which are mapped to their own RDF data source models and communicate with Sesame.

G. New ontologies or thesauri can be added, which can be converted into RDF-Schema, and which also communicate with the Sesame RDF server.

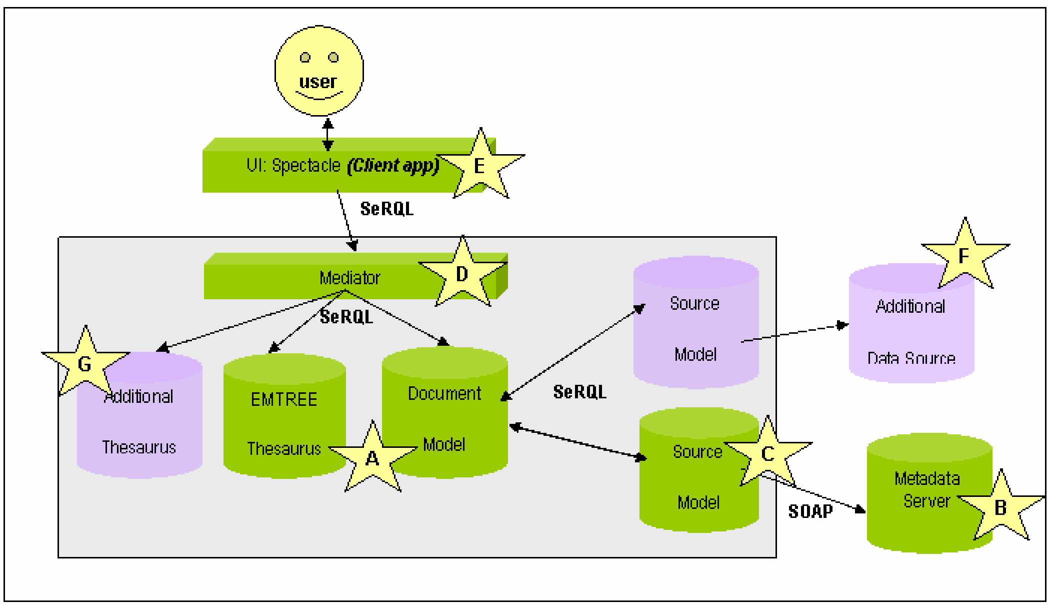


Figure 1: Basic Schematic of the DOPE architecture (protocols and data formats given in bold)

Technical Implementation

In order to provide the required functionality, a technical infrastructure is needed to mediate between the information sources, thesaurus representation and document metadata stored on the Collexis fingerprint server. Besides the technical integration, the representations of the different information sources have to be integrated on a syntactic and structural level.

In the DOPE prototype, this mediation is implemented using the RDF repository Sesame [1]. On a syntactic level we have achieved interoperability by converting all relevant sources to RDF [2]. In particular, we produced an RDF version of the EMTREE thesaurus. The hierarchy of the thesaurus is represented as an RDF schema class [3] hierarchy enabling us to use the reasoning abilities of Sesame to expand user queries to narrower keywords. The problem of structural heterogeneity between the different sources was addressed using transformations on the RDF representation of information. These transformations have been implemented using the Sesame query language SeRQL [4], which also supports queries that create an RDF model as output differing structurally from the queried model. These so-called construct-queries are used Â to communicate with the fingerprint server of the Collexis server [5]. The Collexis Server is a repository of information that is not equipped with RDF-based in- and output facilities. Therefore, an Extractor component is deployed which, through use of the Collexis SOAP interface, converts the available information in an RDF format that is a 1:1 mapping to the original information: the physical model (see Figure 2).

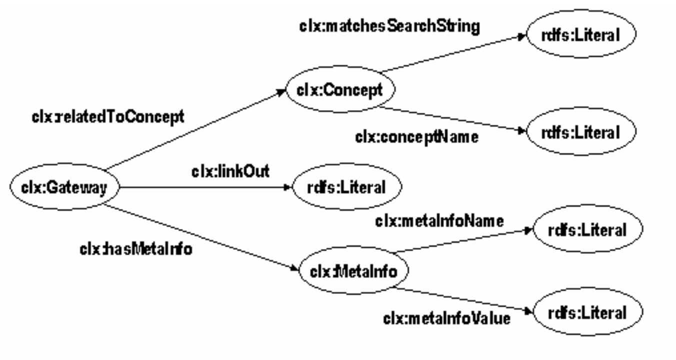


Figure 2 The physical model: an ontology in Collexis terminology

Although the physical model is already in RDF, it is not in the terminology in which the queries are formulated; furthermore, it is less suited to direct merging with different data sources. Therefore, the SeRQL query andÂ transformation language is used to transform the physical model into a logical model. The logical model is based on a subset of the OntoWeb ontology (http://ontoweb.aifb.uni-karlsruhe.de/Ontology/) that has been adapted to our purposes. In particular, the representation of author information has been simplified, and the model has been linked to the schema we use to represent the EMTREE thesaurus. This link can be seen in the lower part of figure 2: each publication is linked to an RDF schema class which represents a preferred term in the thesaurus, and which is further characterized by a label and a relation to similar search strings that is computed on the fly when a query is processed.

The DOPE Browser

A prototype of a user interface client called the "DOPE Browser" has been designed and created. It provides querying and navigation of a collection of documents using thesaurus-based techniques, while hiding much of the complexity of the back-end, such as the existence of multiple data sources, any thesaurus or ontology mapping that may take place, etc. In this system, the user sees a single virtual document collection made navigable using a single thesaurus (EMTREE). Typical document metadata such as e.g. title, authors and journal information is associated with each document. Due to this simplified view on the data, the user interface will be easily reusable on other data sets and thesauri. The DOPE Browser makes use of a thesaurus-driven, interactive visualization technology called the Cluster Map [6], developed by Aduna, for creating overviews of and getting insight in the available information.

One assumption made during the design is that the EMTREE thesaurus is too large for end users to navigate directly. Researchers typically focus their work on an area that can be described by specific terms nested deep inside a thesaurus. They may have difficulties finding their way to these terms. Apart from the cognitive load, manual navigation of the thesaurus may also be cumbersome simply because of its size. An approach has been followed where the user can quickly focus on a topically related subset of both the document collection and the thesaurus. First, the user selects a single thesaurus keyword. The system then fetches all documents indexed with this target keyword, as well as all the other keywords with which these documents are indexed. These co-occurring keywords are used to provide an interface in which the user can explore the set of documents indexed with the focus keyword.

Suppose a user wants to browse through the existing literature on aspirin. The string "aspirin" can be entered in the text field at the upper left of the figure. The system then consults Sesame for all keywords than are related to this string. It responds with a dialog showing four possible EMTREE terms, asking the user to select one. (This dialog is omitted when there is only one exact match with an EMTREE keyword.) Assuming that the user chooses the keyword "acetylsalicylic acid", which is the chemical name corresponding with the brand name, this becomes the new focus keyword. The system consults Sesame again and retrieves up to 500 most relevant documents about "acetylsalicylic acid", including their metadata fields (e.g. titles and authors) and the other keywords with which these documents are indexed. The co-occurring keywords are presented in the tree at the left hand side of the screen, grouped by their facet keyword (the most generic broader keyword, i.e. the root of the tree they belong to). The user can now browse through the tree and check one or more checkboxes that appear alongside the keywords. This action results in a visualisation of their relations and contents at the right hand side of the screen.

Figure 3 shows the state of the interface after the user has checked the terms "mortality", "practice guideline", "blood clot lysis" and "warfarin". The visualization graph shows if and how their document sets overlap. Each sphere in the graph represents an individual document, with its color reflecting the document type, e.g. full article, review article or abstract. The colored edges between keyword and clusters of spheres reveal that those documents are indexed with that keyword. For example, there are 25 documents about warfarin, 22 of them are only labeled with this keyword, two have also been labeled with blood clot lysis, and one is about warfarin, blood clot lysis and mortality. This visualization shows that within the set of documents about aspirin there is some significant overlap between the keywords blood clot lysis and mortality, and that 4 of the practice guidelines documents relate to these two topics as well.

Various ways exist to further explore this graph. The user can click on a keyword or a cluster of articles to highlight their spheres and list the document metadata in the panel at the lower right. Moving the mouse over the spheres reveals the same metadata in a tool tip. The visualizations can also be exported to a clickable image map that can be opened in a web browser.

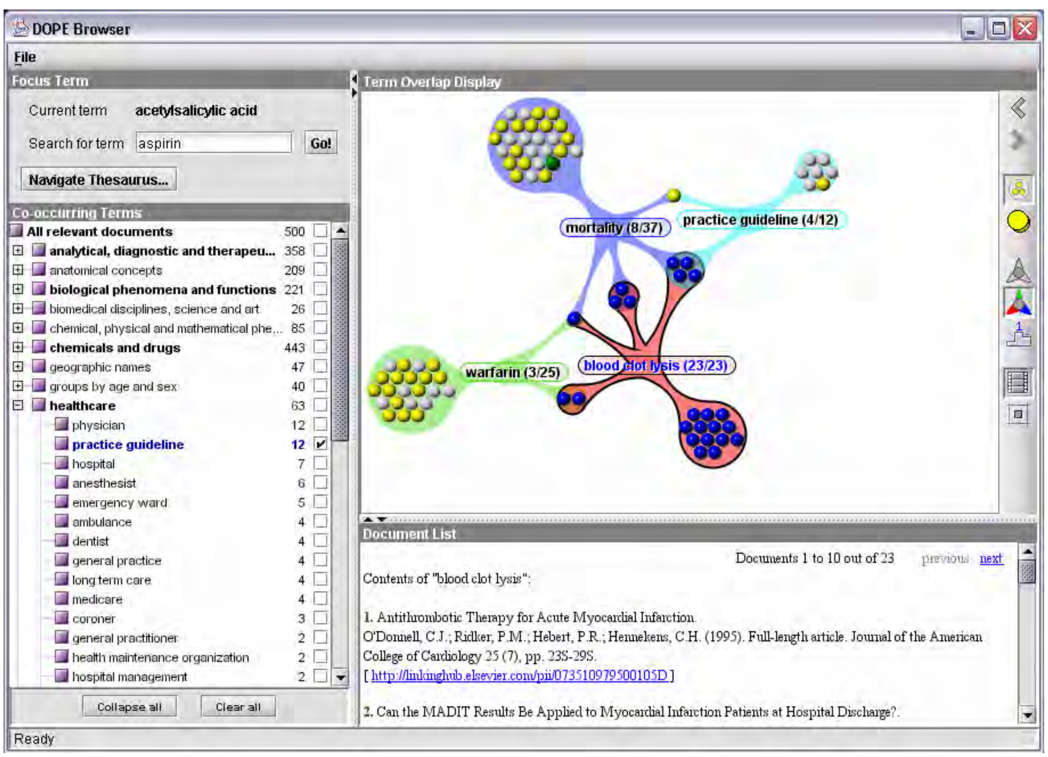


Figure 3: The DOPE Browser

Contributions of Semantic Web technology

Interoperability

The use of RDF-based datamodels and exchange syntax has greatly eased the integration of heterogeneous information sources. The EMTREE thesaurus, the Medline database and the Science Direct collection were all developed independently over many years, with separate separate datamodels and syntactic formats. By wrapping the indexed database as an RDF source and by transforming EMTREE into an RDF Schema structure it was possible to integrate these heterogenous sources.

Functionality

The semantics of the thesaurus is used in the following ways in the functionality of the DOPE system:

* Initial keyword queries by the user are disambiguated by detecting homonyms in the thesaurus
* Search results are hierarchically organised using the thesaurus
* Search results are graphically presented in clusters based on their location in the thesaurus
* Queries can be either widened or narrowed by navigating up or down the thesaurus hierarchy.

Conclusions

Discussions with users about the potential benefits supported the conclusion that the main benefit of the tool is the exploration of a large, mostly unknown information space rather than support for searching for concrete articles. Examples of beneficial applications mentioned by potential end users included: filtering material for preparing lectures about a certain topic, and supporting graduate students in doing literature surveys (e.g. using a "shopping basket" to collect search results). A more advanced potential application that was mentioned was to monitor changes in the focus of the research community. This however would require an extension of the current system with mechanisms for filtering documents based on date of publication as well as advanced visualization strategies for changes that happen over time.

Current Status

Currently, Aduna and Elsevier are discussing a more widespread adoption of the DOPE prototype in a corporate setting, for use to access various heterogeneous datasets with different, overlapping ontologies. The principal DOPE architecture will be used as a starting point for these investigations. The DOPE prototype which is currently offline will be rebuilt using insights from recent work from both parties.

References

[1] J. Broekstra, A. Kampman, and F. van Harmelen. Sesame: An architecture for storing and querying RDF data and schema information. Proceedings of the First International Semantic Web Conference ISWC 2001. See also http://www.openrdf.org

[2] O. Lassila and R. Swick. Resource description framework (RDF). Proposed recommendation, W3C, January 1999. http://www.w3c.org/TR/WD-rdf-syntax.

[3] D. Brickley and R.V. Guha. RDF vocabulary description language 1.0: RDF schema. Working draft, W3C, April 2002. http://www.w3.org/TR/2002/WD-rdfschema-20020430/

[4] J. Broekstra, A. Kampman. SeRQL Querying and Transformation with a Second Generation Language. Technical White Paper, Aduna/ Vrije Universiteit Amsterdam, January 2004.

[5] Van Mulligen EM, Van Der Eijk C, Kors JA, Schijvenaars BJ, Mons B., Research for research: tools for knowledge discovery and visualization. Proc AMIA Symp. 2002. 835-9.

[6] C. Fluit, AutoFocus: Semantic Search for the Desktop In Ninth International Conference on Information Visualisation (IV'05), 2005.

4.2.4 HLS Use Case 4: Medical Guidelines and Protocols

Integrating formal methods in the development process of medical guidelines and protocols.

Radu Serban1, Annette ten Teije, Frank van Harmelen

Vrije Universiteit, Amsterdam, The Netherlands

General Project Description

The Protocure II project (www.protocure.org, [Protocure-II]) has been a multidisciplinary project (2004 – 2006) bringing together medical doctors and computer scientists and aimed at improving the quality of computerized guidelines and funded by the European Commission under contract number IST-FP6-508794. The project addressed the topic of quality improvement of clinical guidelines and protocols by integrating formal methods of software engineering in the life cycle of guidelines development and maintenance.

The Problem

Medical guidelines and protocols describe the optimal care for a specific group of patients and are meant to improve the quality of patient care by providing better patient informedness and lower inter-practician variability. It has been proved that adherence to guidelines and protocols may reduce health-care costs up to a 25%. The effort spent on developing and disseminating the rather high number of medical guidelines and protocols developed within the last decade far outweighs the efforts on guaranteeing their quality. Indeed, anomalies like ambiguity and incompleteness are frequent in medical guidelines and protocols. Recent efforts have tried to address the problem of quality improvement. These approaches are not sufficient since they rely on informal processes and notations. As a result, many guidelines and protocols in practical use are still ambiguous or incomplete. A different approach, grounded on a formal representation, can answer these needs, as demonstrated in the Protocure I project. The approach taken in Protocure I views medical protocols as computer programs, and formal methods as means to assess and improve their quality. The Protocure II project extends this metaphor to guidelines and makes the analogy between “medical protocol development” and Software Engineering. Thus, Protocure II is aimed at integrating formal methods in the life cycle of guidelines, by developing techniques & tools to support the whole guideline development process. An important part of this effort is dedicated to defining re-usable relations and concepts to build a part of a model of a medical guideline, which can be applied to several guidelines.

The Solution

The research methods used to achieve the project goals are as follows:

1) Study the current guideline development process and augment it to support the use of formal methods and tools for assisting high-quality and up-to-date guidelines (living guidelines);

2) Develop a methodology and tools for supporting the transition from narrative to formal representation of guidelines;

3) Define a library of guideline components, in the form of re-usable patterns

4) Adapt and apply theorem proving and model checking techniques to the formal representation.

To address steps 2 and 3 mentioned above, the problem of linguistic guideline templates and the role of medical ontologies in identifying and maintaining these templates has been studied.

The text of medical guidelines often has a modular structure that is suitable for automatic translation into a formal representation. Patterns include:

- In the event of [MedicalContext] the treatment of choice is [Treatment].
- In the event of [MedicalContext], [Treatment] is recommended.
- [MedicalTargetGroup] [recommendation_op] receive [Treatment] with [MedicalGoal].

Such linguistic templates can be generated and instantiated with the use of an ontology of the medical domain. Producing meaningful linguistic pattern templates and translating them into a formal representation cannot be fully automated, but partial automatable steps can reduce greatly the effort spent in building and maintaining a model of a medical guideline.

To identify compositional linguistic patterns, one can learn an ontology of guideline terms by extracting key phrases from guidelines and assigning them to vocabulary categories (selected as a subset of semantic categories from existing medical thesauri). With the help of a guideline domain ontology that takes into account the guideline components (e.g., action, condition, goal, effect), medical knowledge, and relationships, the terminology specific to a specific disease can be decoupled from the terminology specific to the guideline.

The guidelines contain basically two kinds of pattern templates:

- templates that describe medical background knowledge;
- templates that describe control knowledge.

For the objectives of the Protocure II project, aimed at producing an operational model of a guideline, we have identified action-centric patterns which belong to the following classes:

- Effects of actions
- Associations action-goal
- Recommendations for specific actions for target groups
- Preferences for a specific medical intervention
- Background knowledge-centered patterns
- Associations disease-treatment.
- Associations disease-target group
- Combinations of actions, focused on hierarchical decomposition, sequencing and temporal relations between actions
- Action Sequencing
- Patterns for more complex action sequencing
- Patterns for high-level coordination of actions
- Search for the right dosage of medication

We have evaluated the usefulness of these patterns in guideline formalization, by performing the following steps: (1) a rough comparison (quantitative) of the amount of knowledge (automatically) identified by using patterns with respect to the knowledge modelled by (manual) knowledge acquisition; (2) an analysis (qualitative) of the utility of identified pattern instances, performed on specific fragments of the guideline; and (3) in connection with the two previous steps, an identification of the essential knowledge elements that current patterns overlook.

The conclusion was that the most useful patterns are the control patterns, which are the closest to the implementation and include:

- decomposition, ordering, and repetition, or describe constraints such as the fact that an action may be associated with a time frame, intention and medical effect;
- patterns produced by knowledge engineers to describe relations between medical terms.

Key Benefits of Using Ontology Technology

Existing medical ontologies/thesauri, such as [MESH], [UMLS], [NCIOntology], already contain a lot of background medical information that can be used when building a domain model for the medical guideline and for verifying the consistencies of the relations defined in a clinical guideline or protocol.

More than 15 semantic categories used in the UMLS thesaurus are included in a mini-ontology for the medical domain used in the linguistic template retrieval application used for the Protocure II experiments. By establishing mappings between existing medical thesauri and a custom-built guideline ontology, more medical background information can be extracted from existing medical knowledge sources and incorporated in a guideline model, reducing the risk of inconsistencies and the effort of (re)defining relations between medical terms ([AIIMJ2006, FCTC2006]).

References

[Protocure-II]Protocure 2 Project. Integrating formal methods in the development process of medical guidelines and protocols. URL: http://www.protocure.org. Accessed: 1 June 2006.

[AIIMJ2006]R. Serban, A. ten Teije, F. van Harmelen, M. Marcos, C. Polo-Conde. Extraction and Use of Linguistic Patterns for Modelling Medical Guidelines. Artificial Intelligence in Medicine Journal, Volume 39, Issue 2, Pp. 137-149, 2006.

[FCTC2006]Radu Serban, Annette ten Teije., Formalization of medical guidelines exploiting medical thesauri. Proceedings of the Workshop on Foundations of Clinical Terminologies and Classifications (FCTC 2006), Timisoara, Romania.

[MESH]Mesh (Medical Subject Headings (MESH)). URL:http://www.nlm.nih.gov/mesh/meshhome.html. Accessed: 1 June 2006.

[UMLS]Unified Medical Language System (UMLS). URL: http://www.nlm.nih.gov/research/umls/. Accessed: 1 June 2006.

[NCIOntology]National Cancer Institute (NCI) Ontology. URL: http://www.mindswap.org/2003/CancerOntology/. Accessed: 1 June 2006.

4.2.5 HLS Case Study 5: Clinical Trials

Vague modeling for Evaluating Clinical Trials

Stefan Schlobach, Linda Peelen, Michel Klein

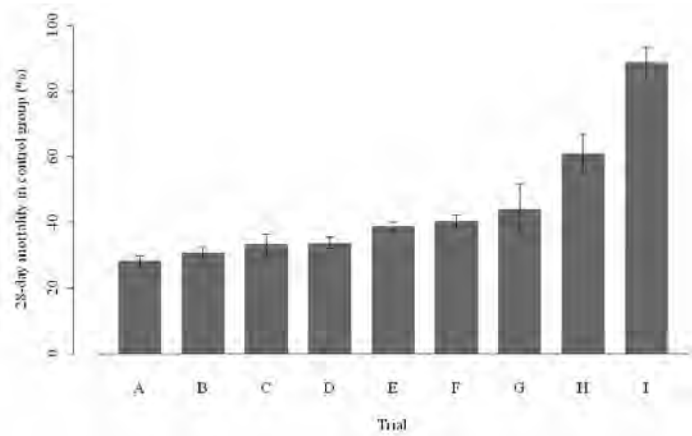
General Description

Clinical trials use entry criteria to select patients for the study. The choice of these criteria is an important step in clinical trial design. To be able to compare the results of the trial with those of other trials and to assess the generalizability of the results to daily clinical practice, the entry criteria have to be compatible with definitions used in comparable trials and the agreed standard definitions of disease. Description Logics, the logical foundations for ontology modeling are obvious candidates to model these entry criterias, to allow for declarative descriptions of classes of patients with particular symptoms. This is obviously complicated when no crisp disease definition exists. For this purpose we model clinical trials in an extension of Dls, so called Rough Description Logics, and use the semantics of these languages to study 9 different clinical trials about the sepsis condition.

The problem

Our current research was motivated by a recent study of the definitions for sepsis used in clinical trials. Before a medical treatment can be used in daily clinical practice, its effect and impact on the patient have to be investigated in a clinical trial. When several trials have been performed it is interesting to compare the results of those trials. Unfortunately, the nine different trials that were investigated in [Peelen et al. , 2005] showed too much variation in their definitions of severe sepsis patients to enable a fair comparison of trial results. Illustration 1 shows the mortality rate for these nine trials. Obviously, a proper comparison of the outcomes of different trials is almost impossible if the patient population is incomparable.

One of the reasons for the problems of defining valid trials in this domain is that there is no accepted general definition of the sepsis condition. Medicine is a typical domain where concepts cannot always be described in a crisp manner. E.g., the definition of a disease is not always clear-cut, especially if a single marker is lacking that distinguishes a patient with a disease from a patient without the disease. This is common in psychiatry and in diseases in which the underlying pathology of the disease is unclear. An example of the latter is sepsis.



Sepsis is a disease in which the immune system of the patient overreacts to an infection. Due to this reaction the patient becomes severely ill, which easily results in organ failure and eventually death. The cause and underlying cellular pathways of this disease are unclear, which hinders the precise characterization of the sepsis patient. Therefore, a consensus definition of sepsis was established in 1992 to define several stages of sepsis [Bone, R.C., 1992]. This definition does not provide a precise definition of sepsis, but gives the criteria for which there was a consensus that they should at least hold for a patient with severe sepsis. In this paper we focus on the patients with severe sepsis , but for brevity we will refer to these patients as septic . The consensus statement defines patients with severe sepsis as ‘patients having a confirmed infection with at least two out of four Systemic Inflammatory Response Syndrome (SIRS) criteria:

- temperature >38◦C OR temperature <36◦C
- respiratory rate >20 breaths/min OR $PaCO_2$<32 mmHg
- heart rate >90 beats/minute
- leucocyte count <4,000 mm3 OR >12,000 mm3

and organ dysfunction, hypoperfusion, or hypotension. From now on we refer to these criteria as the Bone criteria. Patients who have this combination of symptoms may have sepsis, however, this is not necessarily the case. We refer to these patients as being possibly septic . On the other hand, we can define a group of patients that are septic for sure, namely those who fulfill the Bone criteria and have severe multiple organ failure. We will refer to these patients as the definitely septic patients and define them as fulfilling the strict criteria: the Bone criteria plus at least three of the following symptoms of organ failure:

- pH < 7.30
- thrombocyte count < 80,000 mm3
- urine output < 0.5 ml/kg body weight/hour (provided the patient not on chronic dialysis)
- PaO_2/FiO_2 <· 250, and
- systolic blood pressure <90 mmHg OR vaso-active medication.

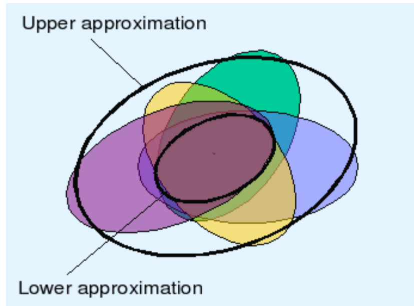
In order to be able to compare clinical trials about sepsis, we need to formalise this information in an ontology, for which we extended the standard Description Logics by rough operators, i.e. the possibility to define approximations of concepts.

The solution

Rough Description Logics (Rough DL) provides us with the possibility to describe diseases for which a crisp definition is lacking by defining lower and upper approxiamtions. In the spirit of Rough Set theory, two concepts approximate an underspecified, vague, concept as particular sub- and super-concepts, describing which elements are definitely , respectively possibly , elements of the concept.

In order to use Rough DL for patient selection we first translated the definition for each trial into a DL formula. We did the same for the Bone definition and the Strict definition of sepsis, thus building a

TBox with 11 definitions for septic patients. In addition we have translated a dataset from the Dutch National Intensive Care Evaluation (NICE) registry containing information on 71,929 patients into an ABox, using the terminology fromthe TBox. With the selection criteria for the different trials and the translated data, we used a DL-reasoner to select the patients that would be eligible for the different trials (thereby mimicking the patient selection process).



We now model the strict and Bone criteria mentioned above as lower and upper approximation of sepsis. We use Rough DL to formalise and compare sepsis definitions used in different trials. Describing sepsis through approximations enforces powerful semantic consequences. Rough DL turns out to be an appropriate logical representation language to model vague concepts and provide crisp answers to queries, and can thereby assist in the validation of existing and, ultimately, the construction of new trials.

Key Benefits of Using Ontology Technology

In our evaluation of medical trials about sepsis patients we have shown that modeling vague knowledge can help to answer important questions in the design of clinical trials. The validation of trials

- Based on their formal definitions is already an improvement over the usual data-based validation.
- When the validation done in a declarative way using Rough DL, the logical consequences of the semantics immediately reveals inconsistencies in the trial definitions, whereas several successive queries are necessary to do the same with standard DLs.
- Finally, we claim that Rough DL can be very useful when building new trials with vaguely defined medical conditions, as they enforce better models for the selection of patients.

References

[Bone, R.C., 1992] Bone, R.C. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Crit Care Med, 20(6):864 - 874, 1992.

[Peelen et al., 2005] L. Peelen, N.F. De Keizer, N. Peek, E. De Jonge, R.J Bosman, and G.J. Scheffer. Influence of entry criteria on mortality risk and number of eligible patients in recent studies on severe sepsis. Crit Care Med, 33(10), 2178 - 2183, 2005.

Summary

From the analysis of the use cases above, we summarize the key benefits of the Semantic Web and ontology Technologies in Health and Life Sciences as follows:

- **Interoperability**. The use of the semantic-web-based data models and exchange syntax has greatly eased the integration of heterogeneous information sources, as shown in Use Case 1 (Intensive Care) , Use Case 3 (Drug Ontologies), Use Case 6 (Public Health Situation Awareness), and Use Case 7 (Traditional Chinese Medicine).
- **Flexibility**. The use of formal ontologies allows for dynamic selection of medical data and knowledge. It allows for more advanced data analysis and integrative knowledge discovery based on the huge web of data., as shown in Use Case 1 (Intensive Care) , Use Case 2 (OpenKnowledge), Use Case 6 (Public Health Situation Awareness), and Use Case 7 (Traditional Chinese Medicine).

- **Better Modeling**. The use of ontology technologies enforces better models for medical data, guidelines and protocols, as shown in Use Case 4 (Medical Guidelines and Protocols) and Use Case 5 (Clinical Trials).
- **Knowledge Sharing**. The use of semantic enabled technologies allows distributed way for sharing of data, knowledge, and services. Ontology alignment techniques developed in the context of Semantic Web allow to combine and compare medical data from different data sources. This benefit is shown in Use Case 1 (Intensive Care), Use Case 2 (OpenKnowledge), Use Case 3 (Drug Ontologies), Use Case 6 (Public Health Situation Awareness) and Use Case 7 (Traditional Chinese Medicine).
- **Reasoning Support**. Robust reasoning support and various reasoning tools are available for ontologies and semantic enabled data. It allows for efficient way to check the inconsistency of medical knowledge or data. It increases the query expressiveness so as to retrieve more complete answers. This benefit is shown in Use Case 2 (OpenKnowledge), Use Case 4 (Medical Guidelines and Protocols), Use Case 5 (Clinical Trials), and Use Case 7 (Traditional Chinese Medicine).

5 The OOA Quality Guidelines

Developing good quality ontologies is an important goal in ontology engineering. The importance of quality is not only to build reliable ontologies; good quality also enables reusability, consensus, adoption, correctness in reasoning and prediction, performance in computation, and so on.

In close cooperation with The US National Center for Ontological Research (NCOR), the US National Institute of Standards and Technology (NIST), and the European Center for Ontological Research (ECOR), the OOA is working on a comprehensive set of ontology authoring quality guidelines that are intended to become a recommendation to the ontology community in general.

Remark: This activity has been initiated by the OOA in 2006. However, there are some preparatory issues delaying our production. As you will see in the text below, the call-for-guidelines, the terminology used, and the recommendation procedures have to be *intensively discussed* and agreed among the above organizations. This is in fact due to the nature of this activity (i.e., recommendation). It has been agreed that the call-for-guidelines will be sent out beginning of September.

5.1 The Ontology Quality Challenge

Achieving an agreed or principle-based set of criteria that can be generalised to assess ontologies remains a very difficult task. The overall value of an ontology (as a representational artefact supported by a logical theory which is adequate for the purposes the ontology has to serve) is clearly dependent on a number of very different factors such as its faithfulness to reality, and the relationship with the conceptualisation of reality built up by the ontology authors. Depending on the paradigm for ontology development chosen, relevant issues are for instance how many of the intended models are captured; the validity of its formal properties, e.g. whether all representational units denote something in reality, whether there are no axioms that imply each other. Further issues are the ontology's compliance with respect to a given standard specification language or syntax; to its applicability, i.e. how much a given ontology fulfils a given application's requirements; to its coverage, richness, granularity, formality level, and so on.

Most of these factors are not only very difficult to quantify or systematically measure in practice, but also an ontology which excels in one factor may provide little to no value in real applications if it does not meet minimal criteria in relation to other factors. Ontologies that are not good for certain applications or domains may be good for others.

Another, indirect, aspect of quality is how much an ontology is reused and adopted successfully in real life applications. Indeed, the repeated use of an ontology in distinct operational environments, rather than a single research environment, gives a good indication that it is accepted in a certain community, that there is a consensus about it, and that it has been adequately tested and improved. Some researchers and practitioners argue that the ultimate evaluation of an ontology is in terms of its adoption and successful use, rather than its consistency or coverage. The Gene Ontology is an example: while being clearly impoverished in many representational aspects, it is nevertheless a success story.

5.2 Goal for the Guidelines

The goal of this chapter is to identify and recommend a set of guidelines that contribute directly to quality, or indirectly by pursuing ontology reusability and adoption. These guidelines will be *promoted* *not only to ontology engineers but also to tool developers*. An ontology authoring tool can then be evaluated and scored, for example, based on how much it implements these guidelines. The idea is that enforcing these guidelines during the ontology development phases ensures a certain quality of the product, i.e. the ontology. Although the final recommendations are not intended to play a role of a gold standard for quality assessment, but as a first initiative in this regard, they are supposed to lead to better ontology content authoring. Furthermore, *the final results (guidelines) will be used by the OOA itself for future ontology evaluation services* (See deliverable D1.3.3).

5.3 Guideline Collection (Call for Guidelines)

Experienced ontologists are invited and encouraged to contribute to this recommendation, by submitting guidelines based on their best practice and research findings. A guideline is not necessarily a rigid assessment criterion or a theory, but can also be a methodological recommendation that guides ontology builders to achieve better quality, reusability and/or adoption. Not every guideline includes a formal technical specification, nor can every recommendation be embodied by support from authoring tools; each is intended to be directly comprehensible by industrial users and applicable in a wide range of industry settings. While some guidelines are interrelated, each is self-contained and can be followed independently of the others.

The collected guidelines will be refined to arrive at more agreement. Where there is eventual disagreement, the differing opinions will be clearly documented. All guidelines will be reformulated where appropriate to provide a coherent approach to ontology content. Each guideline should be easy to understand and apply by normal ontology engineers or non-technical domain experts.

Guideline collection will be performed online, using the web site of the OOA as platform. Specific guideline templates have been designed and provided in the content management system, and a review procedure has been set up as a work flow to allow a few cycles of comments (Delphi method) before a guideline will be released as ‘formally recommended’. However this procedure will not be closed until formal release. The transient status of a not-yet-released guideline will be made clear, but since the whole idea is to get comments and feedback, even preliminary guidelines will be accessible. The process will closely follow the well-known and established RFC process of the Internet Engineering Task Force.[33] This model calls for an open submission using quite strict format and procedure requirements, a review and revision by a small group of renowned specialists, and a cycle of publish-and-feedback. The OOA plans to have a continuous process in place, to avoid an unnecessary long interval between submission and approval/dissemination of a new guideline.

5.4 Remarks on Terminology

To gain more consensus, the notion of *ontology* in the context of this document is not required to be very formal. Ontology content can range from fine-grained axiomatized theories to structured collections of terms.

Any contributor to this document is encouraged to follow the terminology defined in the attached glossary. However, in case of disagreements on these definitions, a parallel definition should be provided.

5.4.1 Glossary

Conceptualisation: A cognitive representation of a portion of reality in a cognitive agent’s mind

Epistemology level: The level that deals with the knowledge structuring primitives (e.g. concept types, structuring relations, etc.).

Ontology reusability: the ease of using an ontology (or part of it) for several kinds of (autonomously specified) tasks.

Ontology usability: the ease of using an ontology in different applications that perform the same kind of task.

Representation is for example an idea, image, record, or description which refers to (is of or about), or is intended to refer to, some entity or entities external to the representation. Note that a representation

[33] http://www.rfc-editor.org/

(e.g. a description such as 'the cat over there on the mat') can be of or about a given entity even though it leaves out many aspects of its target.

Composite Representation is a representation built out of constituent sub-representations as their parts, in the same way in which paragraphs are built out of sentences and sentences out of words.

Representational Units; The smallest constituent sub-representations, examples are: icons, names, simple word forms, or the sorts of alphanumeric identifiers we might find in patient records. Note that many images are not composite representations since they are not built out of smallest representational units in the way in which molecules are built out of atoms. (Pixels are not representational units in the sense defined.)

Universals (in the vernacular also called 'types' or 'kinds'). A universal is something that is shared in common by all those particulars which are its INSTANCES.

Domain is a portion of reality that forms the subject-matter of a single science or technology or mode of study; for example the domain of proteomics, of radiology, of viral infections in mouse.

Taxonomy is a tree-form graph-theoretic representational artifact with nodes representing universals or classes and edges representing is_a or subset relations.

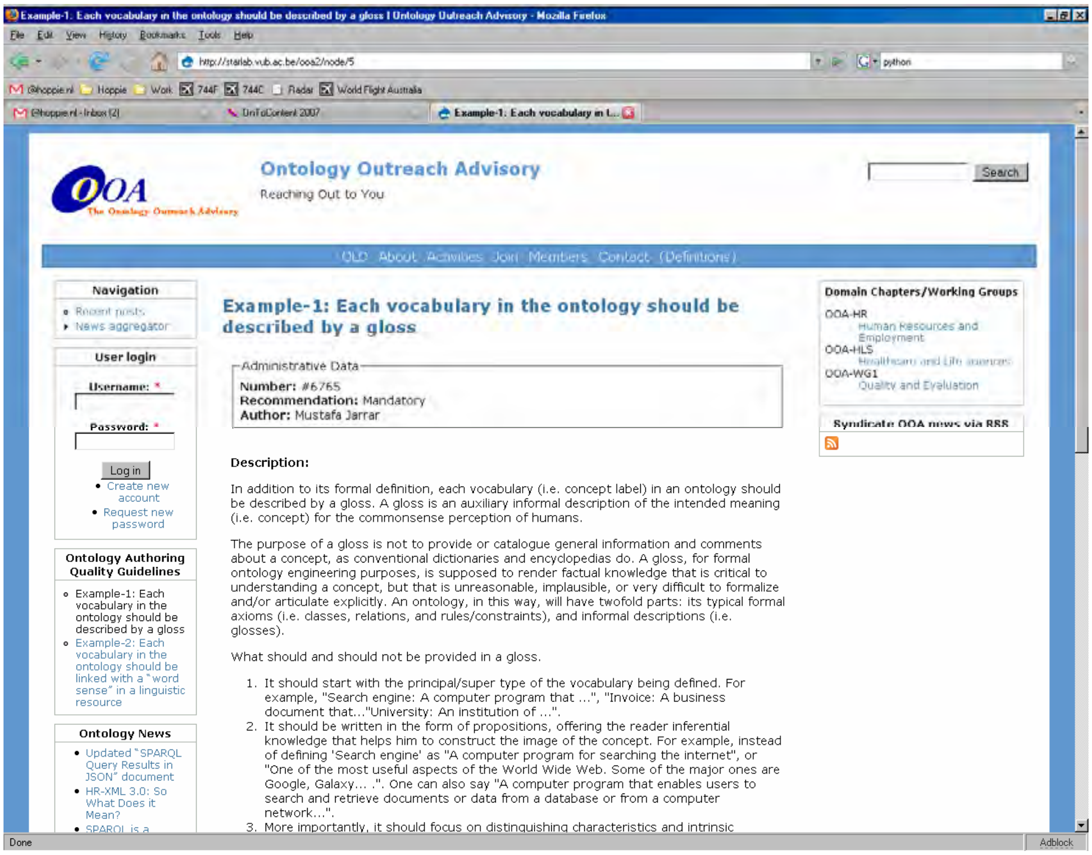
Terminology is a representational artefact consisting of representational units which are the general terms of some natural language used to refer to entities in some specific domain.

5.4.2 Guideline Template and Examples

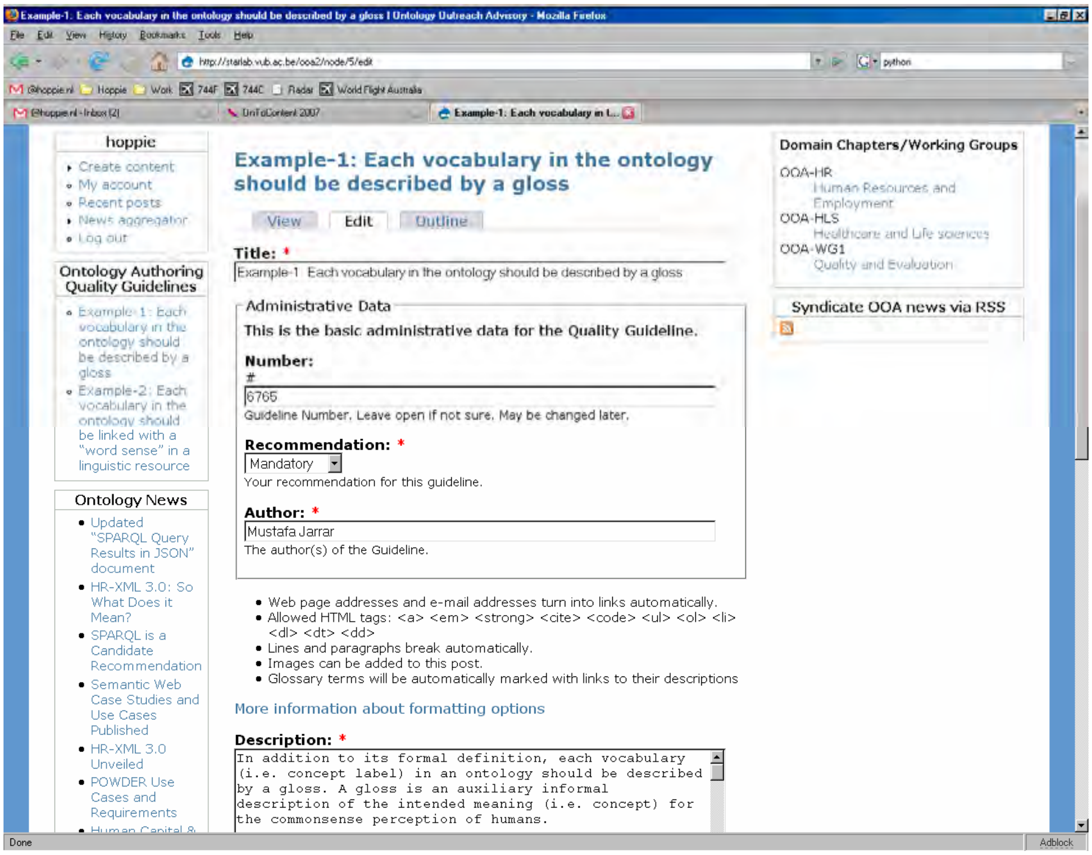
The template below is meant for reference only; actual Guideline submissions take place on the OOA web site.

Guideline Number:	**#**		**Recommendation**	**{Necessary| Mandatory | Encouraged | Optional}**
Author	*{the author of the guideline}*			
The Guideline	*{Summarize the guideline in 1-2 lines. We recommend your guideline to start e.g. each concept should…, each relationship should…, the scope of the ontology should,…}*			
Description	*{Describe the guideline in details, about one page. The description should be easy to understand by any ontologist, non-technical domain experts, etc. Your guideline should be easy to apply in industry}*			
Advantages	*{"Numerate" clearly the advantages of your guidelines, what do we gain in practice if we follow it? In which sense it contributes to the quality? Reusability? adoption of ontology content?}*			
References	*{Provide necessary references only, e.g. to find more details, implementation support, etc.}*			

After submission and during review, the Guidelines are visible in a strict format on the OOA web site. Their status is clearly indicated.



The Guideline Editors (including the original author) have access to the technical interface to update the Guideline. This interface is equipped with all standard Drupal elements to allow for a controlled work flow with strict versioning and access control. New submissions can be indicated via mail, RSS, and other standard ways to the team of editors. Discussions about the Guideline may take place in a forum or by means of attached comments.



Production of the Guidelines in hard copy is not foreseen, as new Guidelines will be released often. However, at relevant gatherings such as the HR Summit and scientific conferences, specific presentations about the current state of affairs, new ideas, and strong recommendations will be presented. People will be encouraged to contribute to the Guidelines in as many places as possible.