The Ontology Outreach Advisory

10G730, 2 Pleinlaan 1050 Brussels, Belgium
info@ontology-advisory.org
www.ontology-advisory.org

HR-Semantics Roadmap

The Semantic challenges and opportunities in the Human Resources domain

Document Identifier:	OOA-HR/2007-08-20/v025
Version:	Ver 0.25
Date:	August 20, 2007
State:	Pre-Final
Distribution:	Public
Editors	Mustafa Jarrar, Luk Vervenne, Diana Maynard
Contributors	Adil hameed, Andreas Schmidt, Christophe Roche, Clementina Marinoni, Christine Kunzmann, Ernesto Damiani, Francky Trichet, Jeroen Hoppenbrouwers, Karsten Lundqvist, Marc Van Coillie, Marco Ronchetti, Michael Brown, Pieter De Leenheer, Serge Ravet, Simon White, Stijn Christiaens.

Executive Summary

Purpose: This document identifies the challenges and opportunities in applying the ontology technology in the Human Resources domain.

Target users: 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 HR and ontology experts, 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.

Contributing organizations:

- 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

Contents

1. Introduction and Motivation

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.

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.

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).

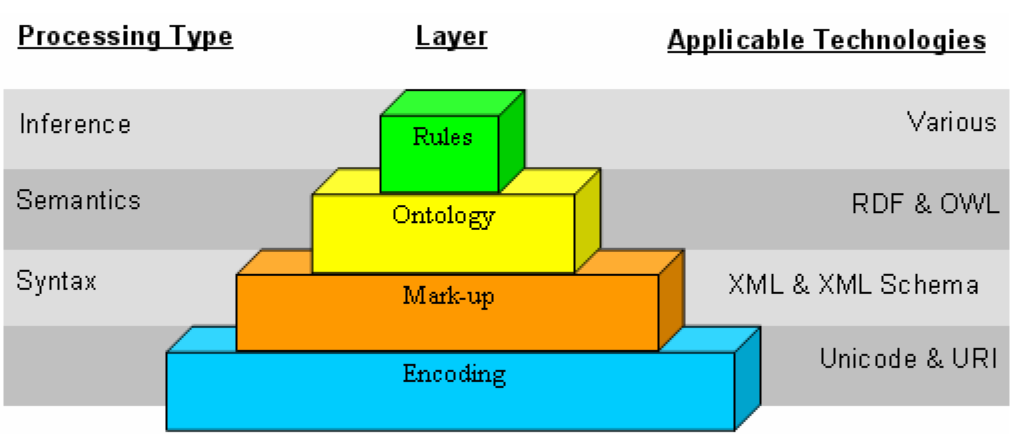


Figure 1 Simplified version of Tim Berners-Lee’s “layer cake”

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 [1] (Semantically Enabled Knowledge Technologies), which aims to develop and exploit the knowledge technologies underlying Next Generation Knowledge Management, and in particular, NEON [2] (Lifecycle Support for NEtworked ONtologies), which aims to support the whole ontology engineering lifecycle by developing a reference architecture and a concrete toolkit and

---

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

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

developing appropriate methodologies for the contextualised construction and use of networked ontologies and associated metadata. Projects such as TAO[3] 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 IST-2004-026460, http://www.tao-project.eu/

2. The HR Application Scenarios and the Need for Ontologies

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[4] and Job Portals[5] 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 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

[4] http://www.job-search.com

[5] http://www.aspanet.org/solutionstemp/jobport.html

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[6] 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.

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

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[7], the most popular international job-search and job-CV matching service

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

presently available, Job-search[8] from the US, CEN-Marketplace[9] from the UK, and Borsa Lavoro Nazionale[10] 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

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

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

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

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.

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[11] from the UK and the O*NET[12] from the US. The European Commission has also launched standardisation programmes such as Europass[13] – concerning CV and qualifications standardisation -- and the eSkills forum[14] initiatives - addressing the ICT skills and competences standardisation with the forthcoming European eCompetence framework.

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.

2.3 Competency-based assessment of employees

If we take for granted that e-assessment can represent an advantage for

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

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

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

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

both assessors and examinees[15], 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)[16];
- 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;

---

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

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

- 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)[17], developed in the UK, can be an excellent example of this preliminary task of defining

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

meanings according to a specific, reference context. In 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)[18].

An example of a full standardisation system between qualifications, occupational standards and assessments, comes from the UK again, with their National Qualification Framework[19]. 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.[20]

We could complete this system by linking the companies' assessment systems, too. However, concept standardisation is not enough to ensure

---

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

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

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

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.

2.4 Competence-based Learning 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 closely connected for efficient competence development strategies. These two management levels are explained in the following two sections.

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.

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]).

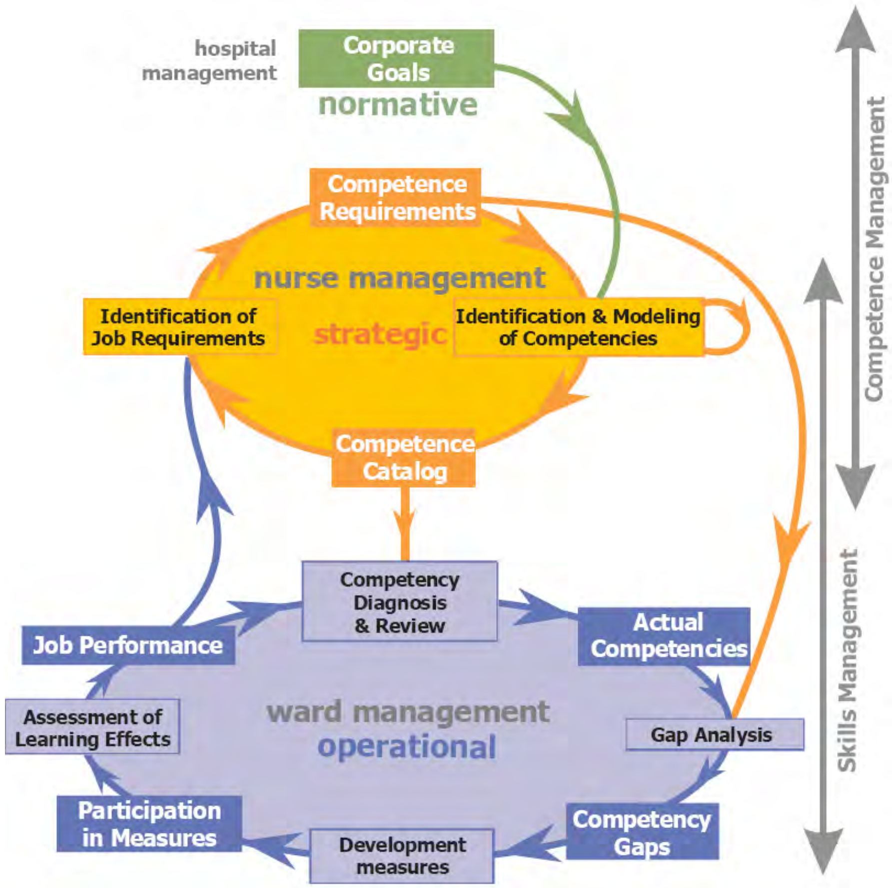


Figure 2: Connecting strategic and operational issues

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.).

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.

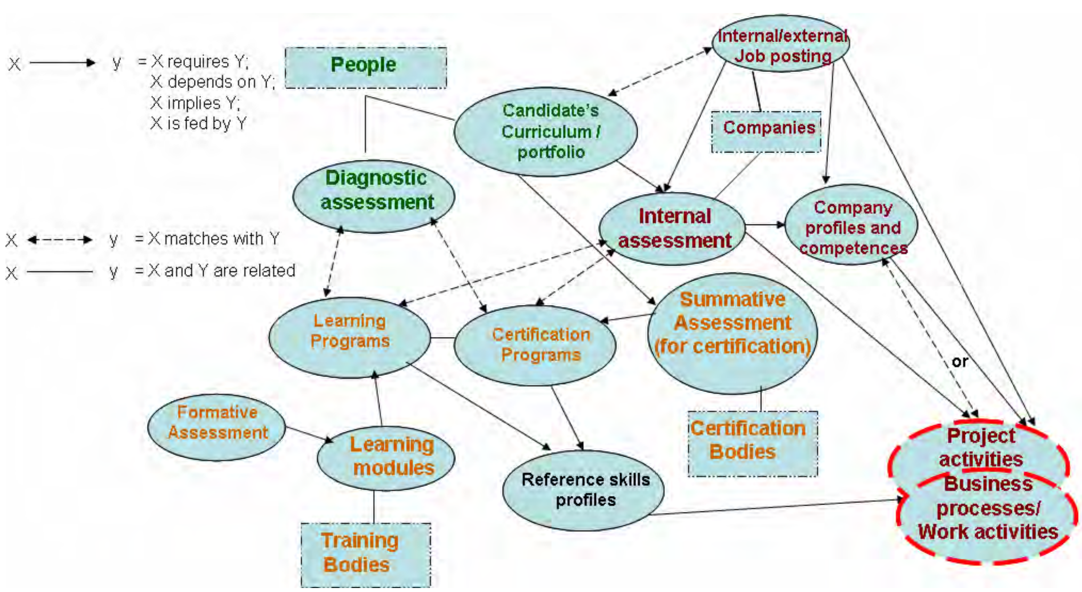


Figure 3: HR Domain network scenario

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. Challenges in the HR Domain

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.

3.2 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 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.

3.4 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 [21] incorporates certain logical inferences and relationships

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

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.

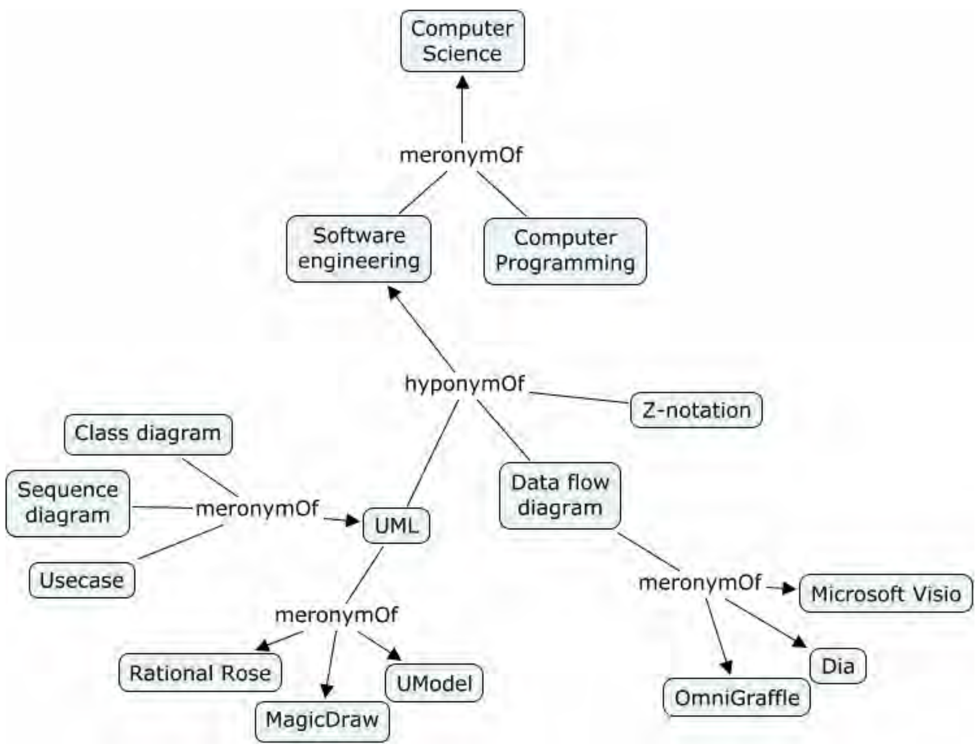


Figure 4: An example of a prototype ontology

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 differently. Ontologies from the Dublin Core Metadata Initiative[22] and the Friend of a Friend (FOAF) Project[23] could be used to support the automation of these processes.

[22] http://dublincore.org/

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

4. Existing Tools, Technologies and Methodologies

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).

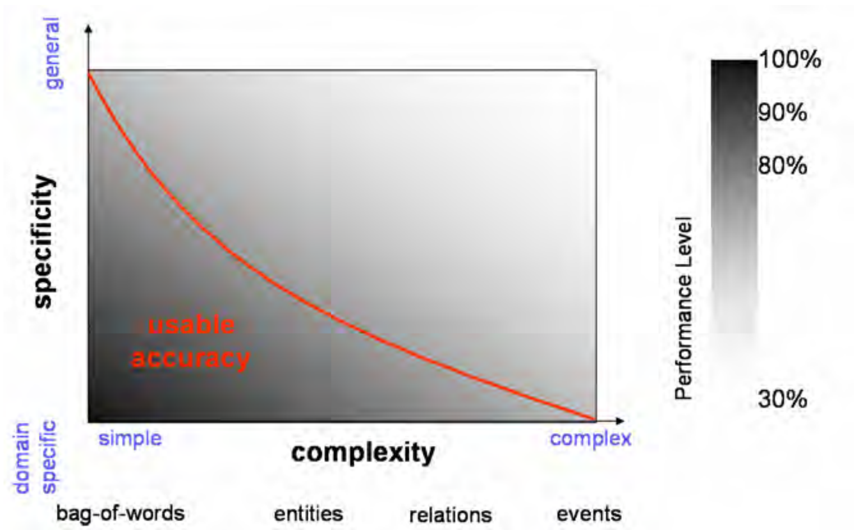


Figure 5: Tradeoff between specificity and complexity for language processing tasks

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].

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.

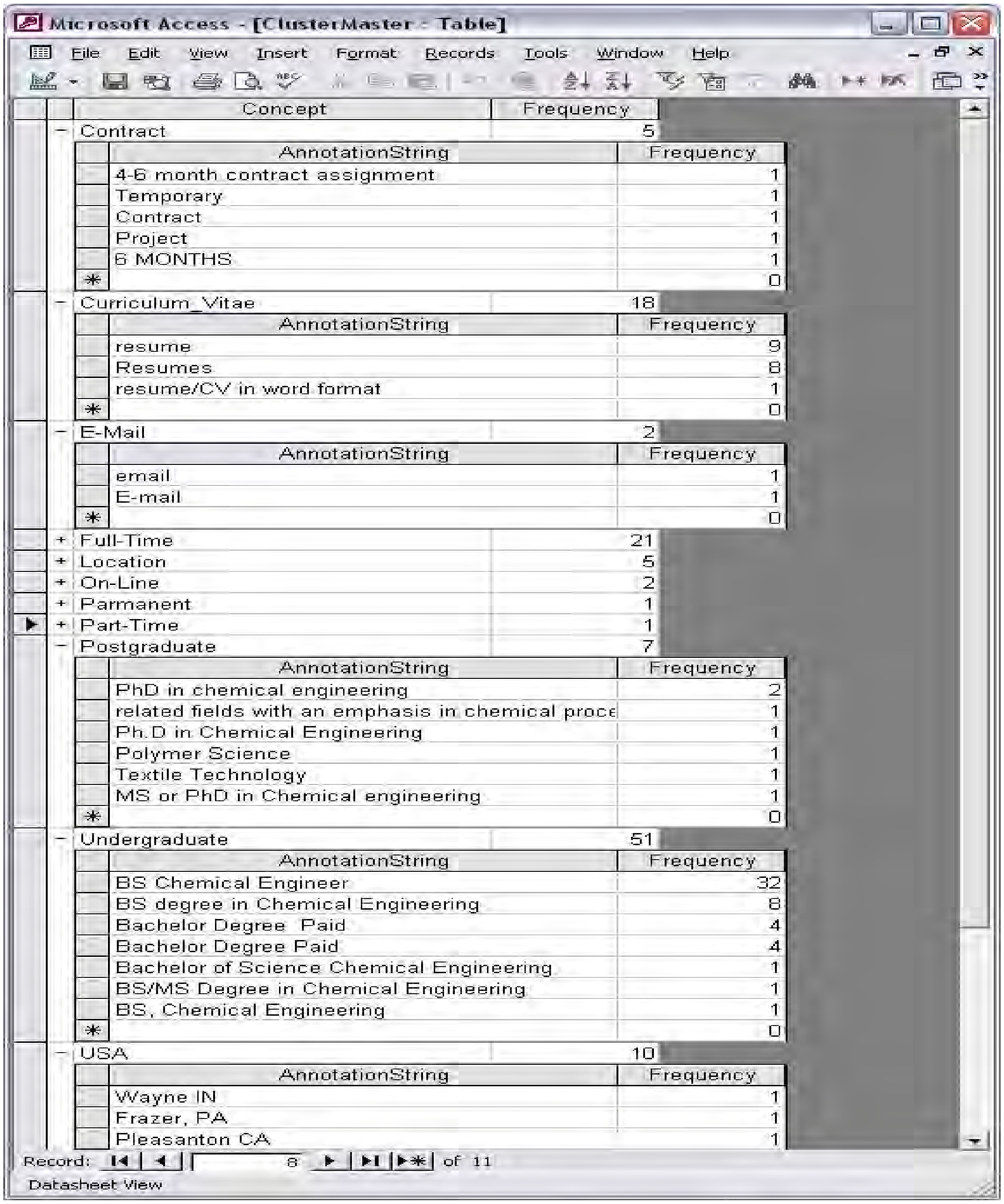


Figure 6: Screenshot from the h-Techsight application

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).

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[24] (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.

[24] 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.

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).

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[25]. 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 [26] 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

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

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

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.

4.5 Learning in Process: Context-Steered Learning Framework

Within the EU project LIP (Learning in Process)[27], 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] http://www.learninginprocess.com/

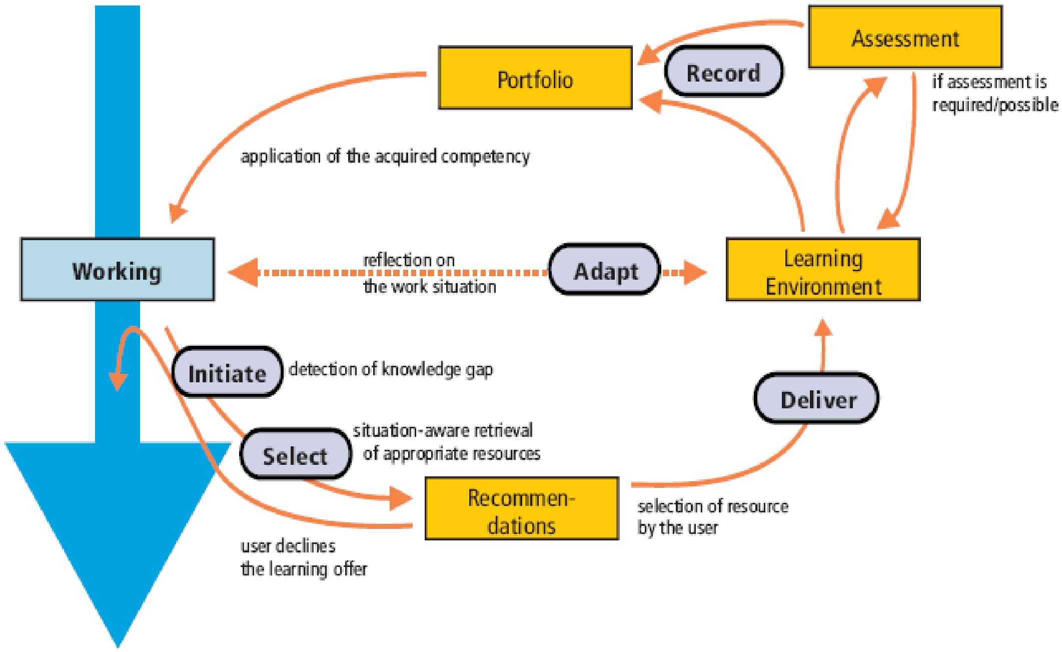


Figure 7: Context-Steered Learning [Schmidt 2006a]

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.

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.

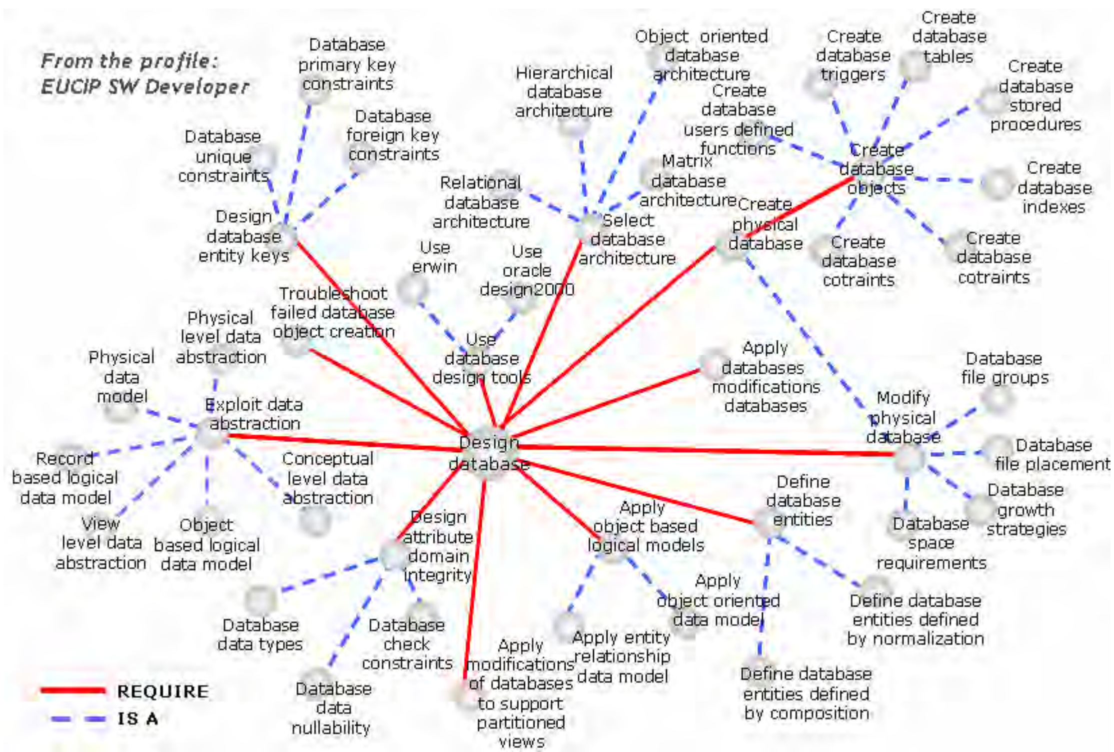


Figure 8: Example of a semantic network of knowledge objects

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.

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[28].

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.

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

6. Existing projects and initiatives

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.

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.

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.

6.3 PoCeHRMOM

The PoCeHRMOM[29] (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

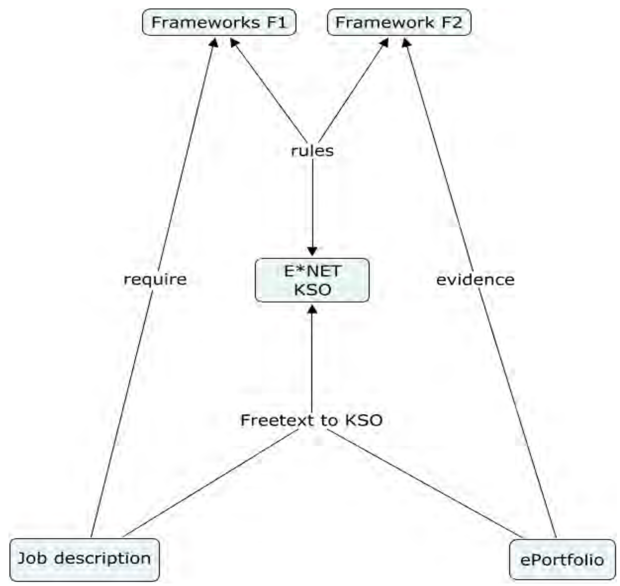
[29] PoCeHRMOM is an acronym for the Dutch phrase: **P**roject **o**mtrent **C**ompetenties en functies in **e**-**HRM** voor technologische toepassingen op het Semantisch Web door **O**ntologie en **M**eertalige terminologie.

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.

6.4 TRACE

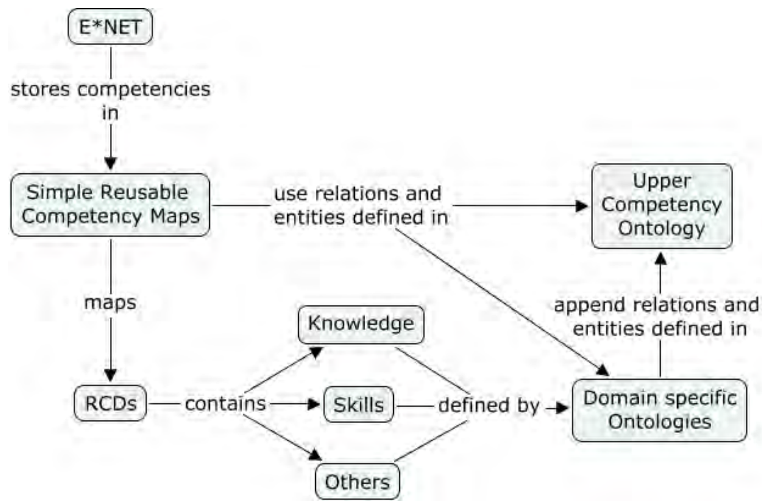
The central aim of the EU Leonardo sponsored project TRACE is to improve the transparency of competences across EU states, regions and sectors. TRACE is investigating the current state and use of existing competence systems and, from this investigation, is developing a methodology and technique to create 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.

7 Existing Ontologies

7.1 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.

7.2 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.

7.3 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.

7.4 "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.

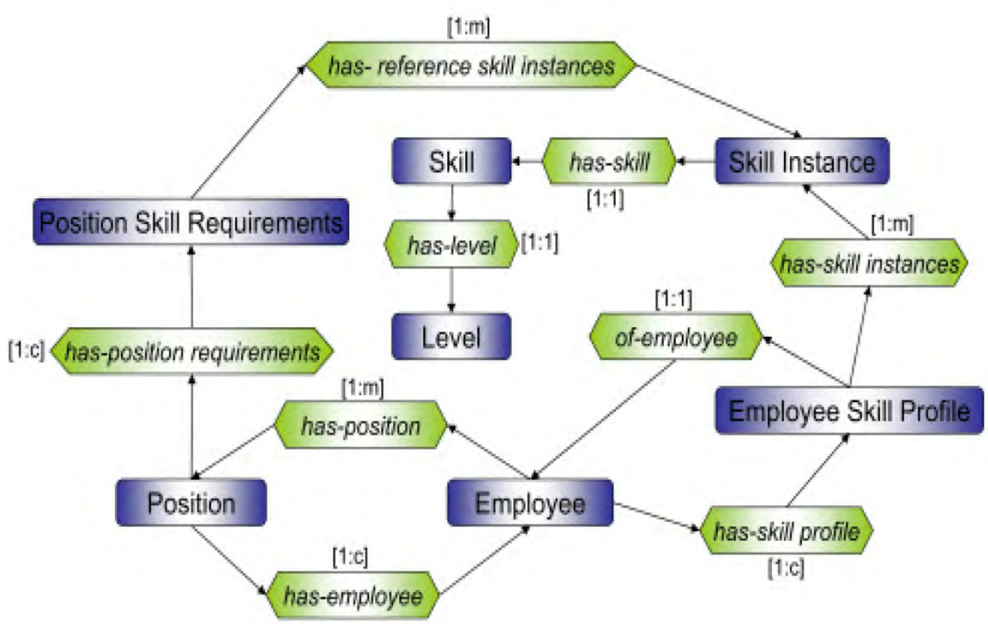


Figure 9: ePeople Ontology

7.5 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.

7.6 CommOnCV

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.

7.7 TOVE

(Toronto Virtual Enterprise Ontologies)[30] 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.

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

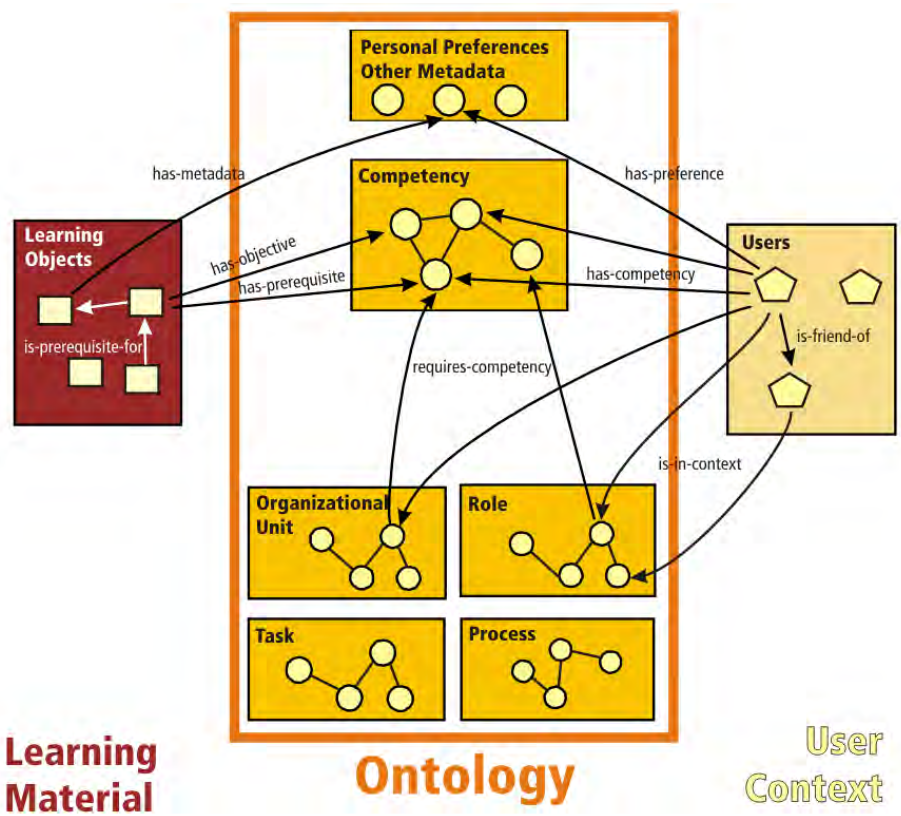


Figure 10: Schematic graphical representation of the LIP Ontology

7.8 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

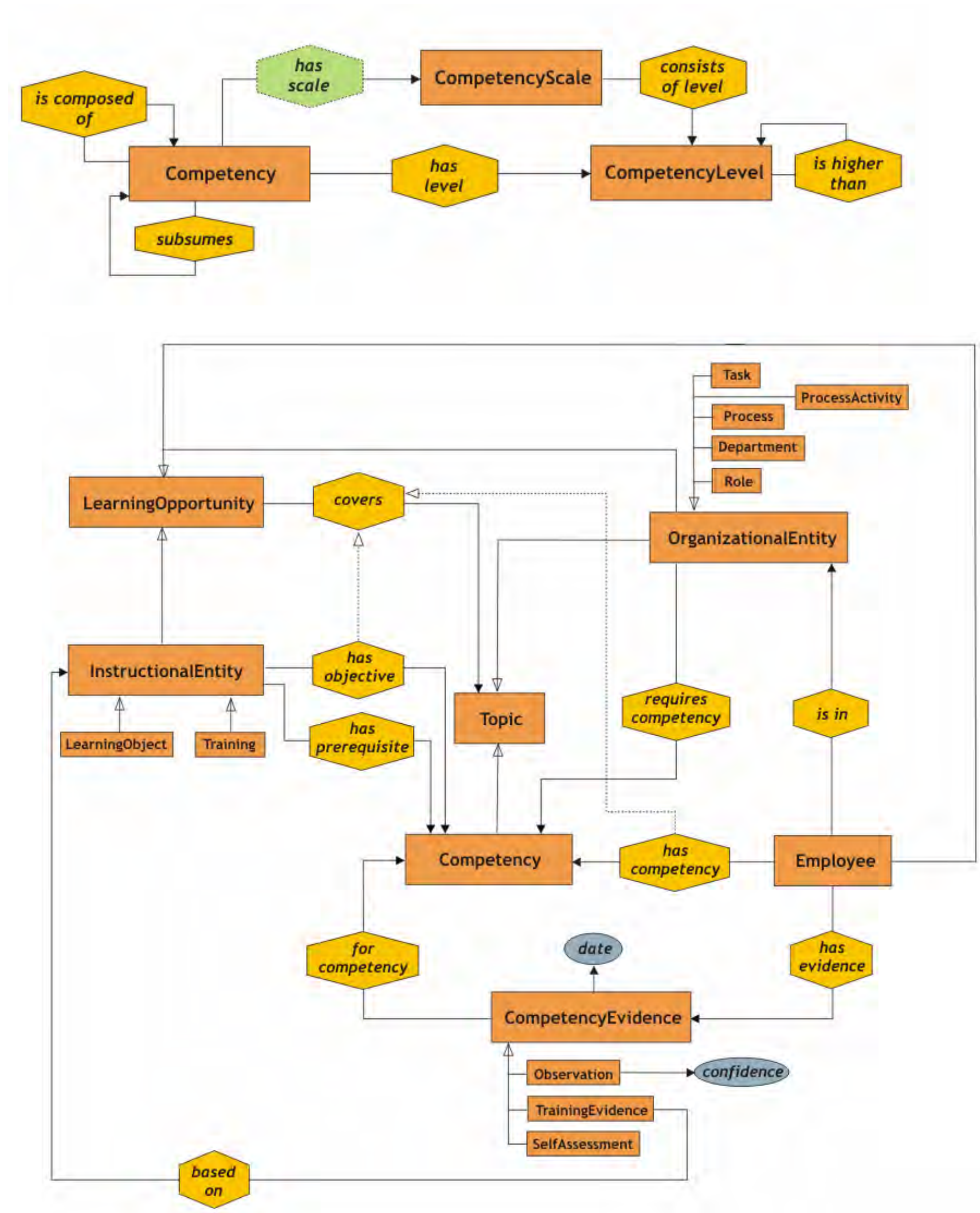


Figure 11: Professional Learning ontology

7.9 PROTON (PROTO-Ontology)

PROTON 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/.

7.10 COKE

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

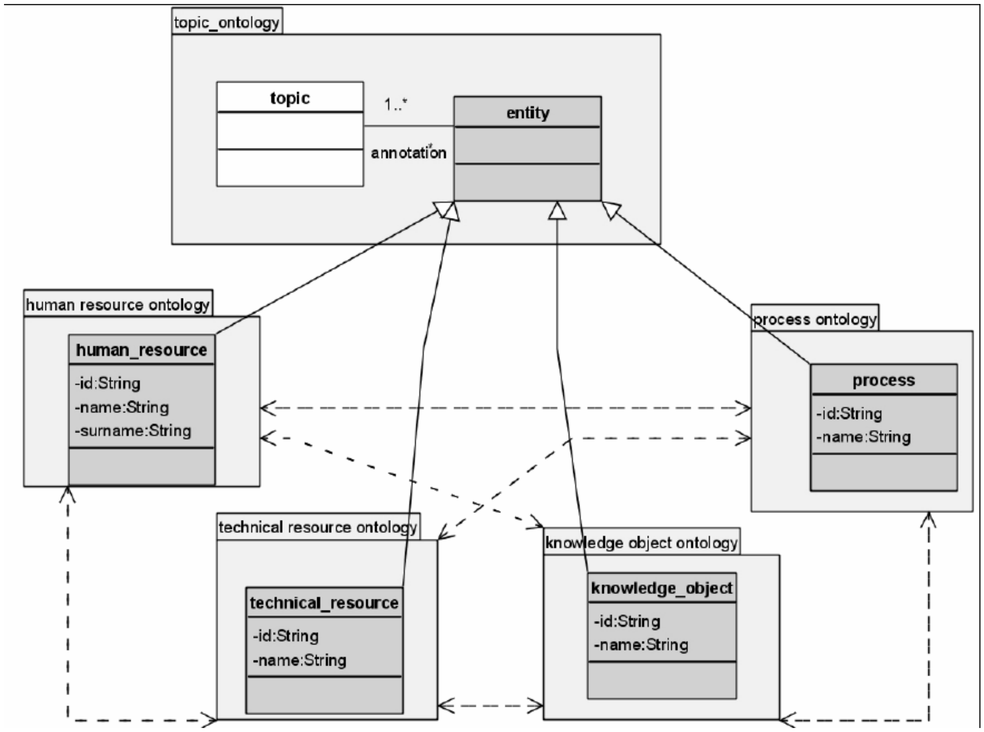


Figure 12: COKE ontology

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.

8 Success Stories

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.

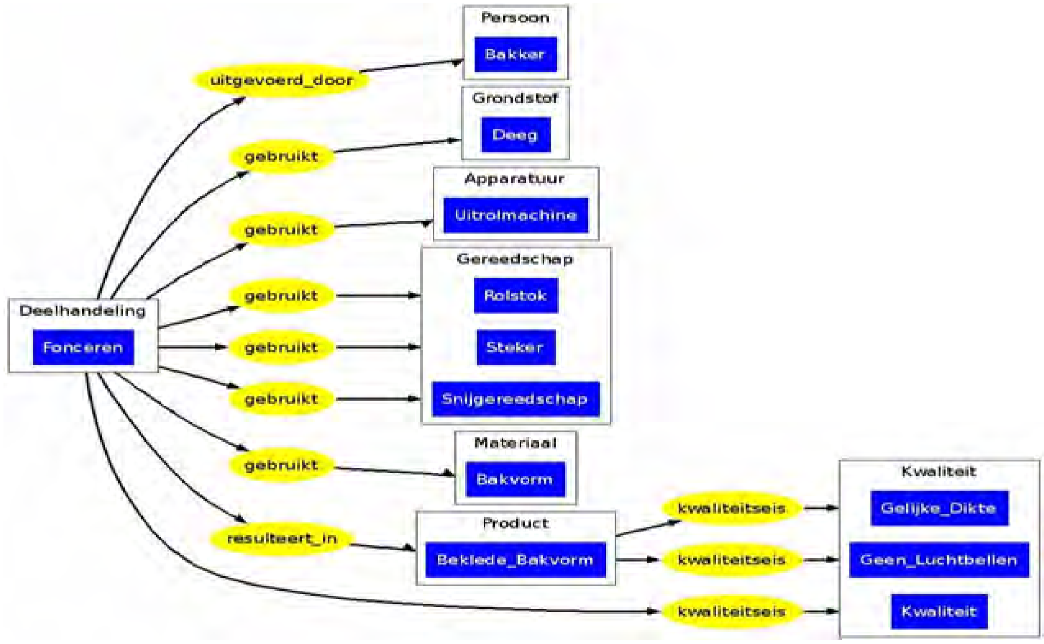
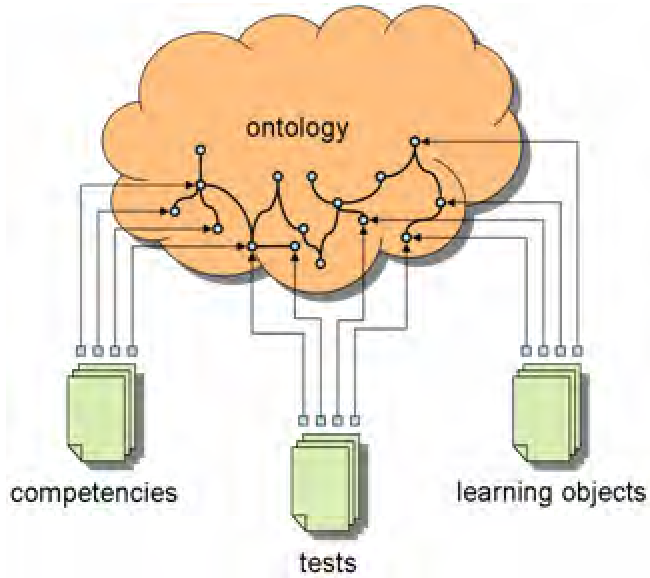


Figure 13: Example of a baker-created definition

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.

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.

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)