Corporate Semantic Web Report IV

State of the Art Analysis
Working Packages in Project Phase II

Technical Report TR-B-11-07

Adrian Paschke, Gökhan Coskun, Ralf Heese,
Radoslaw Oldakowski, Mario Rothe, Ralph Schäfermeier,
Olga Streibel, Kia Teymourian and Alexandru Todor

Freie Universität Berlin
Department of Mathematics and Computer Science
Corporate Semantic Web Group

20 July 2011

INNOPROFILE

UNTERNEHMEN REGION
Die BMBF-Innovationsinitiative
Neue Länder

Corporate Semantic Web: Report IV

Adrian Paschke, Gökhan Coskun, Ralf Heese,
Radoslaw Oldakowski, Mario Rothe, Ralph Schäfermeier
Olga Streibel, Kia Teymourian and Alexandru Todor

Freie Universität Berlin
Department of Mathematics and Computer Science
Corporate Semantic Web Group
Königin-Luise-Str. 24-26
14195 Berlin, Germany

paschke,coskun,heese,oldakowski,mrothe,
schaef,streibel,kia,todor@inf.fu-berlin.de

20 July 2011

Abstract

In this report, we introduce our goals and present our requirement analysis for the second phase of the Corporate Semantic Web project.

Corporate ontology engineering will improve the facilitation of agile ontology engineering to lessen the costs of ontology development and, especially, maintenance. Corporate semantic collaboration focuses the human-centered aspects of knowledge management in corporate contexts. Corporate semantic search is settled on the highest application level of the three research areas and at that point it is a representative for applications working on and with the appropriately represented and delivered background knowledge.

Contents

Chapter 1

Vision of the Corporate Semantic Web

Nowadays, companies seek more capable approaches for gaining, managing, and utilizing knowledge as well as for automating dynamic services and agile business processes. The Corporate Semantic Web (CSW) offers promising solutions here. The Corporate Semantic Web vision aims at bringing semantic technologies to enterprises (see figure 1.1).
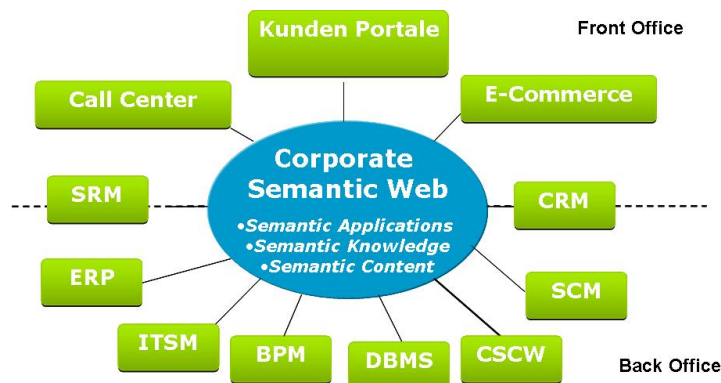
Figure 1.1: Corporate Semantic Web for Semantic Enterprises

Corporate Semantic Web (CSW) deals with the application of Semantic Web technologies (in particular rules and ontologies) within enterprise settings. It address the technological aspects of semantic engineering and managing semantic enabled IT infrastructure to support (collaborative) workflows, communication, knowledge management, and (business) process management in enterprises. The core building blocks of the Corporate Semantic Web for semantic enterprises are:

1. Semantic Content

2. Semantic Knowledge

3. Semantic Applications

Semantic content can be created from existing non-semantic data by automated extraction of the underlying semantics from, e.g., text, multi-media, (IT and business) events, (user) activities, etc. This allows for transforming the data into structured semantic content such as linked data and for (inductively) learning ontologies and rules. New semantic content can be produced in a manual semantic engineering and enrichment process, using, e.g., semantic annotation tools such as Loomp, and semantic content creation tools such as semantic editors and Ontology/Rule engineering tools like Leone and Prova.

The semantically enriched content and newly created semantic metadata can be used for **semantic knowledge** management, where so called semantic organizational memories (e.g., Linked Data clouds, semantic Wikis, semantic Content Management Systems, semantic Desktop Systems, etc.) store relevant enterprise knowledge for, e.g., multiple re-use in different application contexts, reducing information search times, faster training of staff, improvement of task and labor allocation, etc. This benefits the semantic integration and transformation of heterogeneous data from different sources into machine-readable information and company-relevant knowledge (see figure 1.2).
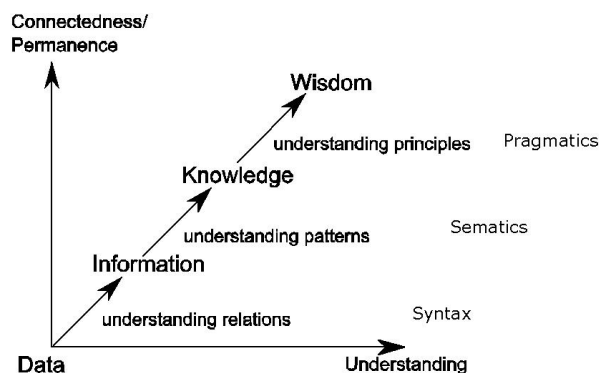
Figure 1.2: Semantic Web and Pragmatic Web Knowledge Transformation

It facilitates the automated analysis of the ex-post data by semantic data mining/trend mining, semantic business intelligence and the processing of real-time data by semantic complex event processing and situation detection, in order to make the implicit knowledge semantically explicit, so that it can be used in real-time enterprise information systems and agile (business) processes.

The semantic content and semantic knowledge is used to build **semantic tools and applications** with high levels of *semantic intelligence* (pragmatic wisdom). They support the three application domains of the Corporate Semantic Web approach (see figure 1.3).

- **Semantic Engineering**
  Methods and tools with which ontologies and rule bases for accurate and high quality corporate information and (reaction/decision) processes can be created and maintained.

- **Corporate Semantic Search**
  Solutions for semantic search in controlled enterprise information repositories.

4

- **Corporate Semantic Collaboration**
  New semantic platforms with which the business units can collaborate
  in order to create, collect, use, and manage information, processes and
  knowledge.

Figure 1.3: Corporate Semantic Web Approach

Finally, the vision of the Corporate Semantic Web also addresses the **pragmatic context** of actually using Semantic Web technologies in enterprises. This includes learning and training aspects as well as economical considerations - i.e. corporate in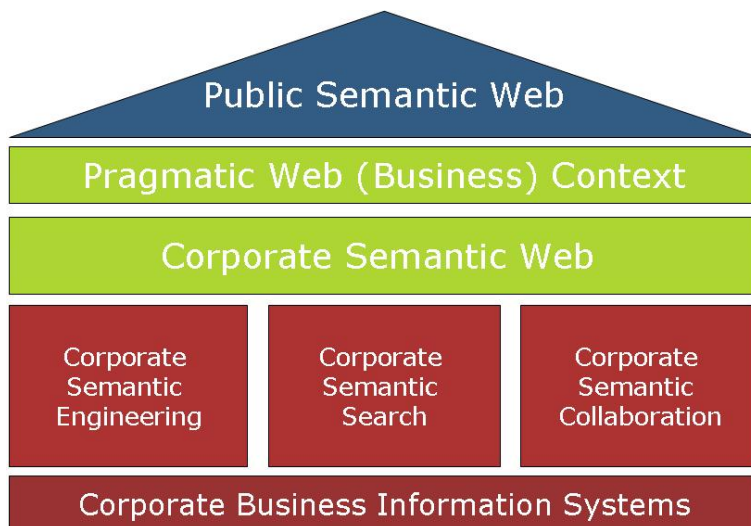 the sense of entrepreneurial activities. Incentives need to be provided to encourage in-house adoption and integration of these new Corporate Semantic Web technologies into the existing enterprises information systems, services and business processes. Decision makers on the operational, tactical, and strategic IT management level need to understand the impact of this new technological approach, its adoption costs, and its return on investment based on cost models for engineering the semantic content and knowledge.

In the first phase of the BMBF funded InnoProfile project *Corporate Semantic Web* parts of the CSW vision have been realized - see [35, 91, 90]. The second phase of the CSW project addresses several working packages which research and develop advanced methods and tools for personalized and multimedia search, distributed collaborative knowledge engineering and evolution, as well as the pragmatic aspect of quantifying the use of ontologies in terms of engineering cost models and qualifying their quality in terms of evaluation methods. (see appendix Work Packages A). This report describes the underlying research problems and analyses the state of art of existing problem solutions. The further report is structured along the three pillars of the CSW approach - *semantic engineering*, *semantic collaboration*, *semantic search* - and describes the state of art analysis for each working package in the second phase of the CSW project.

Chapter 2

Corporate Ontology Engineering

2.1 Cost Models for Ontologies (AP11)

In recent years, the adoption of Semantic Web technologies has gained significant momentum throughout the open Web.

For the open Web, the Linked Open Data initiative can be considered to be one of the main drivers [11], allowing service providers to publish content in a unified, machine readable data format and consumers to process this data or integrate it into their own web sites. The number of published data sets increases constantly (see figure 2.1). The web of documents is about to turn into a web of data, fulfilling one part of Tim Berners Lee's and Jim Hendler's initial visions of the Semantic Web [10]. Companies have also realized the advantage of standardized data exchange formats, as the recent adoption of schema.org[1] by the largest three search companies has shown.

However, companies are still hesitant when it comes to investing in the development of larger complex models for their data, hindering them to take full advantage of the possibilities of the Semantic Web. We believe that the lack of standardized, yet dynamic and agile processes for ontology building and maintenance is one of the major obstacles. Furthermore, investors have no means for estimating the costs for such an endeavor. This is why we consider cost models for agile ontology development and maintenance life cycles as crucial for the adoption of full-blown Semantic Web technologies.

2.1.1 State of the Art

While cost estimation has been a subject of study in the realm of Software Development for decades, very little research in that domain has been undertaken w.r.t. Ontology Engineering yet.

The efforts in Ontology Engineering cost estimation and project planning have strongly been influenced by the correspondent research endeavors in the Software Engineering discipline. Therefore, we start with pointing out the relevant work in that domain before we elaborate on the state of the art of cost

---

[1] http://schema.org/
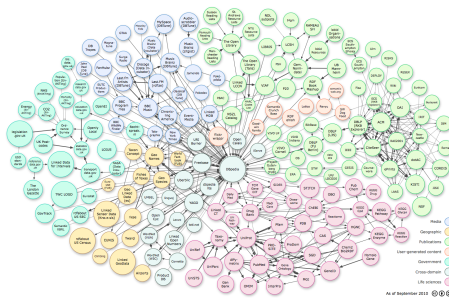
Figure 2.1: The LOD dataset cloud as of September 2010

estimation in the domain of Ontology Engineering and the results that have been achieved so far.

Cost Estimation Models from the Software Engineering Domain

Research on cost estimation models for the Software Engineering domain has a history dating back to the mid-1960s, when the System Development Corporation conducted an extensive survey in 169 US governmental software projects in order to identify cost defining attributes of the Software Development process [84]. Based on this work, first cost estimation models have been established.

Over the decades, Software Engineering methods have evolved, and so have cost estimation models.

In this section, we give an overview over the different cost estimation techniques and existing implementations in the form of cost models.

Boehm at al. classified Software Engineering cost estimation models using the following six gropus [17]:

- model-based cost estimation,

- expertise-based cost estimation,

- learning-oriented cost estimation,

- dynamics-based cost estimation,

- regression-based cost estimation, and

- composite cost estimation.

In this work, we give an overview of the relevant work in this area. Since learning- and dynamics-based cost estimation is outside of our focus, we confine ourselves to the remaining estimation models.

Model Based Cost Estimation

The earliest parametric model, SLIM (Software LIfecycle Management), was introduced by Putnam in 1978 [100].

It is based on the observation by Norden, Lee and others that the costs of Software Development projects typically follow the cumulative Rayleigh distribution [71]:

7

$$\text{Cost} = d(1 - e^{-at^2}),$$

where $Cost$ is the total effort expected, t is time, d is the scale factor of the distribution, and a is a parameter affecting the shape of the curve.

Putnam interprets the peak of the curve as the total development time and observes that the cost in person years can be captured by the model function

$$\text{Cost} = \left[\frac{\text{LOC}}{P \cdot t^{4/3}} \right]^3 \cdot B,$$

where

- Cost is the effort in person years,

- LOC is the product size in terms of effective lines of code (LOC),

- P is the process productivity,

- t is time in years, and

- B is a scaling factor.

The process productivity is defined as the fraction of software size and defect rate, and has to be estimated.

One problem with this model is that it only yields accurate results for large software projects (in terms of years rather than months).

Checkpoint is a proprietary, commercial Software Development cost estimation model developed by Jones [65].

It takes function points as its input, i.e., software projects participants estimate the size of a project in terms of functional units instead of mere code lines [4].

Another model-based, parametric cost estimation model is PRICE-S [88]. PRICE-S consists of three submodels:

- *The aquisition submodel*: Includes issues such as reengineering, code generation, spiral development, rapid development, rapid prototyping, object-oriented development, and software productivity measurement.

- *The sizing submodel*: Aids in estimating the project size. Supports LOC, Function Points, and Object Points, a project size measure adapted to object oriented software projects [81].

- *The life-cycle cost model*: Extends and makes use of the aquisition submodel in order to predict maintenance and support costs.

COCOMO (COnstructive COst MOdel) [16] and its successor COCOMO II [18] are probably the most popular parametric cost estimation models for Software Development processes.

COCOMO operates at three different modes, depending on size and complexity of the project under study:

- *Organic Mode*: Targets small to mid-sized projects. Requirements towards process formalization, i.e. specification and documentation, are minimal. Team members possess experience in the same kind of project, and estimated communication overhead is small.

- *Semidetached Mode*: Mid-sized projects. Team members have at least some experience in the same kind of project. Specification and documentation needs are stronger than in organic projects. Emphasis on communication routines is stronger than in organic mode.

- *Embedded Mode*: Estimates costs for large, complex projects or projects with complex requirements, e.g. security related systems. Assumes a rigid project structure w.r.t. specification, documentation, and communication needs.

The cost in Person Months (PM) is then calculated using the following formula:

$$\text{PM} = m \cdot \text{KDSI}^n,$$

where

- KDSI means thousands Delivered Code Instructions, excluding those for testing or documentation purposes as well as those that are removed prior to end product delivery.

- m is a factor depending on the operation mode. $m = 2.4$ in organic mode, $m = 3.0$ in semidetached mode, and $m = 3.6$ in embedded mode.

- n is also dependent on the operation mode. $n = 1.05$ in organic mode, $n = 1.12$ in semidetached mode, and $n = 1.20$ in embedded mode.

In addition to the three modes, COCOMO introduces numeric *cost drivers*.

Cost drivers can consist of system related issues as well as team related factors. Each cost driver has to be estimated by the project stakeholders and appears as a weighting factor in the cost estimation equation:

$$\text{PM} = m \cdot \text{KDSI}^n \cdot cd_1 \cdot cd_2 \cdot \ldots \cdot cd_x,$$

where each of $cd_1 \ldots cd_x$ is a cost driver.

Other model based cost estimation techniques include ESTIMATICS [103] and SEER-SEM [64].

Expertise-Based Cost Estimation

Expertise cost estimation techniques are useful when hard empirical data is not available. They rely on the experience of a software engineer in a project similar to the project under study. The following two methods have been developed with the aim to overcome the risk of misjudgment.

The delphi method [2] employs repetitive rounds of expert consultation on a specific matter, such as the estimated effort for a software endeavor. During the first round, the experts are asked to state their judgement without consulting

each other. In the subsequent rounds, they are asked to do so again, but after presenting them the opinions of the other experts. The observation here is that after a small number of rounds, the experts will find a consensus.

Another technique that accomplishes the Delphi method is the Work Breakdown Structure [7]. The idea behind this technique is that breaking down the workload necessary for a software projects into small units, like requirements, tasks, or modules, makes cost estimation easier. The role of the expert here is to specify these units based on his or her experience.

Cost Estimation Models for Ontology Engineering

ONTOCOM

At the time of this writing the only existing model for Ontology Engineering cost estimation is ONTOCOM (ONTOlogy COst Model) [93]. The first version of ONTOCOM was based on empirical data from 36 Ontology Engineering projects. In a second round, this data set has been extended to 148 projects [117]. The results have been calibrated using elimination of outliers, multivariate regression and bayesian analysis, and ANOVA analysis.

Like COCOMO, the ONTOCOM model takes into account a number of ordinal cost drivers which appear as weighting factors in the cost function. The results from the calibration suggest that from 11 cost drivers only six explain most of the behavior of the model. These are:

- *Domain Analysis Complexity (DCPLX)*: account for those features of the application setting which influence the complexity of the engineering outcomes,

- *Evaluation Complexity (OE)*: account for the additional efforts eventually invested in generating test cases and evaluating test results,

- *Ontologist/Domain Expert Capability (OCAP/DECAP)*: account for the perceived ability and efficiency of the single actors involved in the process (ontologist and domain expert) as well as their teamwork capabilities,

- *Documentation Needs (DOCU)*: state for the additional costs caused by high documentation requirements,

- *Language/Tool Experience (LEXP/TEXP)*: measure the level experience of the project team w.r.t. the representation language and the ontology management tools, and

- *Personnel Continuity (PCON)*: mirror the frequency of the personnel changes in the team.

2.1.2 Conclusion

ONTOCOM assumes a sequential ontology life cycle consisting of domain and requirement analysis by domain experts, conceptualization by domain and ontology experts, implementation by ontology experts, and evaluation by ontology and domain experts [117].

On the other hand, interviews with the industrial partners of the Corporate Semantic Web project revealed that small and mid-sized companies seek lightweight and dynamic processes for ontology development and maintenance with minimal need for ontology experts involved in the process [76]. With our agile life cycle model COLM [35] we explicitly address the need for usage-oriented agile ontology evolution [122, 123] (see figure 2.2).

We expect that applying a usage-oriented agile life cycle model to the Ontology Engineering process has at least a significant effect on the cost drivers Evaluation Complexity, Language/Tool Experience, and Personal Continuity. Ontology Evaluation in usage-oriented ontology evolution scenarios and in classical engineering life cycles differ dramatically. While the former exclusively relies on expert opinion and strict evaluation metrics, the latter is driven by explicit and implicit user feedback to an important extent. Since tool support for non-experts in Ontology Engineering plays an important role in our perception of agile ontology maintenance, which we addressed with an ontology editor for non-ontologists [90], language and tool experience will play a less important role. Since ontology evolution is driven by user feedback, again explicit or implicit, in a usage-oriented agile life cycle, personnel continuity might play a less important, but still significant role.

Taking the ONTOCOM approach as a starting point and adapting the existing cost drivers to the the agile ontology life cycle will provide future stakeholders ready to engage in agile ontology development with the appropriate means for forecasting the effort and risks.

2.2 Ontology Evaluation (AP 12)

The corporate context poses new challenges for ontology evaluation which is the process of measuring the quality of an ontology. While evaluating an artifact requires intuitively an in-depth analysis with keeping the application environment in view, the corporate context requires cost-sensitive and efficient processes. In Corporate Ontology Engineering the process of ontology evaluation is tackled in such a way that the influence on the existing and running enterprise systems and business processes should be as low as possible. Previously, we worked on the processes of ontology versioning, modularization and integration and argued that reusing existing ontologies in a modular way is important to avoid high investment costs. Efficient versioning on the other hand is essential to improve the first version of an adopted ontology continuously. In this regard, ontology evaluation is a part of the ontology selection and modularization process for creating the first version and on the other hand an important part of the continuously ontology improving process.

2.2.1 Evaluating Ontologies

An ontology is a complex artifact which has different aspects to consider during an evaluation process. In [134] the following six different aspects are identified.

- Vocabulary is the set of all names used in an ontology (e.g. URIs or literals).

- Syntax is the serialized representation of an ontology, either directly or described as a graph as done in RDF/XML.

- Structure of an ontology means always the structure of the graph representation of an ontology.

- Semantics is the most important aspect of an ontology, it is its content.

- Representation is the aspect of an ontology that connects the semantics with the structure.

- Context is the environment or application domain in which the ontology is used.

Besides the aspects of an ontology which have to be considered, evaluation needs criterias to have fix points what should be aimed at evaluation. There are different sets of criterias stated in the literature, all defining their own principles for a good ontology. Asunción Gómez-Pérez [54] defines consistency, completeness, conciseness, expandability and sensitiveness as criterias for ontology evaluation. Gruber [55] lists the criterias of clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment. Obrst [87] developed the set of coverage,intelligibility, validity and soundness, consistency, completeness,adaptibility and mappability for ontology evaluation. Grüniger and Fox [56] defined the single criteria of competency and introduced formal and informal competency questions to measure this criteria. Lozano-Tello [74] defined a detailed set of 117 criterias, describing the aspects of the formal language, the contents of the ontology, the costs of using the ontology and the available tools. Other sets of criterias are stated by Gangemi [47] and Vrandečić [134] summed them up.

2.2.2 State of the Art

Frameworks

In the last years different frameworks were developed aiming at different goals and approaches for the ontology evaluation.

The first kind of Frameworks handles the ranking of ontologies throught sorting a set of ontologies bases on some criteria which is usually a search term. Such ontology search engines like Swoogle [41], SWSE [59], Watson [36] or FalconS [32] ranks and displays ontologies in the order of the evaluation results of that search term.

Another kind of Framework aims towards ontology selection, which is a more specialized form of ontology ranking. Therefore only the best result for given criterias is selected and reused in some other process. D'Aquin and Lewen [37] uses such an approach in the Cupboard system for selecting single-axioms.

A third kind of Framework is presented by Gangemi et al[48] that covers ontology evaluation and validation. They integrate different evaluation methods using a model that consists of a metaontology called O^2 which characterises ontologies as semiotic objecs. Based on structural, functional and usability-profiling measures they use a ontology of ontology validation which chooses the best set of criteria to get the best ontology in the context of a given project.

Vrandečić [134] modified this framework and uses reification to consider more aspects of ontology evaluation.

Burton-Jones [28] proposes an approach for weighted criterias. The ontology is therefore evaluated for every critera, a weight is assigned for every single one, and the overall score is then computed as the weighted sum of its single criterias. He further groups single criterias into metric suites like syntax,semantics,pragmatics and social and uses the weights on this groups.

An nearby research field of ontology evaluation is the one of information and data quality as the criterias for good data match the ones for ontology evaluation. Juran and Godfrey [66] states that data is defined as having a high quality "if they are fit for their intended uses in operations, decision making and planning"

Methods and Approaches

Brank et al. [20] classified the evaluation methods into following four approaches:

1. comparing the ontology to a golden standard

2. task-based ontology evaluation

3. data-driven ontology evaluation

4. human-driven evaluation against a set of predefined criteria, standards or requirements (see [74])

Similarity-based approaches, also called the comparison of an ontology to a golden standard is a measure that compares an ontology with an existing ontology that serves as a reference.The higher the similariy of the ontology compared to the reference, that much better is the ontology in the evaluation process. Ehrig et al. [44] uses a similarity function that given two ontologies as an input, it calculates an real-valued similarity score between 0 and 1. Or formally:

$$sim : \mathcal{O} \times \mathcal{O} \rightarrow [0,1]$$

Such a function can be defined on different aspects and elements of an ontology. On the vocabulary level, Levensthein distance can be used to get the mean difference between the strings of the elements of an ontology. Also the structural aspects of two ontologies can be compared throught several measures, like the semantic cotopy of two hierachies (see [78]). Using precision and recall measures, the semantics aspects of ontologies can be evaluated with the golden standard [121]. The best usage can be found in the context of the evaluation of automatically learned ontologies to this golden standard ontologies.

Another approach for evaluating ontologies is to evaluate it in the context of the task and utility where it is used. Such a **Task-driven-approach** has the advantage that well developed software engineering methods can be applied for evaluation. As ontologies are never used directly in applications, but are a part of it, the performance of the ontology raises and falls within the application. Vrandečić[134] lists disadvantages of this kind of approach:

- there is no way to generally say if an ontology is good or bad, because the evaluation is done for a specific task

- the ontology could be only a small component of the application and its effect on the outcome may be relatively small

- if evaluating a large number of ontologies, they must be sufficiently compatible that the application can use them all

Porzel and Malaka [99] describes such an utility-based evaluation in an speech recognition problem. Haase and Sure [58] presented an evaluation function in an environment where ontologies are used for search tasks. The function describes how costly it is to determine relevant individuals in the class hierachy and therefore are an indicator how good this ontology performs. Sabou [106] creates custom ontologies from the formulation of a task and evaluates them afterwards.

An approach can be also to evaluate an ontology by comparing it to existing data about the domain of this ontology, using e.g. text documents. This **Data-driven-approach** has been used in different fields. Jakulin and Mladenić [63] describes an ontology grounding process on the data of individuals that shows users the errors which can be used in an refinement process. To archive domain completeness Natural Language Processing techniques can be used to a domain specific text-corpus to compare the ontology to the coverage of a complete domain corpus [130]. Patel [94] proposed an approach to determine if the ontology refers to a particular topic, and to classify the ontology into a directory of topics. Brewster [22] suggested the usage of a data-driven approach to evaluate the degree of structural fit between an ontology and a corpus of documents. In the case of ontologies that contains a lot of factional information Vrandečić [134] states that the corpus of documents could be used as the source of facts about the external world. The evaluation measure would be the percentage of these facts that can also be derived from the ontology.

2.2.3   Ontology Evaluation in Corporate Context

In a corporate context a pragmatic approach to ontology evaluation for reuse of existing ontologies is neccessary. Because what matters from the business point of view is the improving effect on the enterprise systems and the business processes. Being a semantic model utilized in the corporate context the focus of this work is on the two aspects semantics and context. We assume that RDF/XML is the standard syntax and we make use of the structure for appropriate visualization, which is neccessary to comprehend the content, the semantics.

The Corporate Ontology Lifecycle (shown in 2.2) represents the lifecycle of an ontology utilized in a company. It explicitly contains a phase called Evaluation in which feedback from the usage is analyzed to increase the ontology quality. We propose to enrich the structure visualization with usage data from the Feedback Tracking and Reporting phases. By doing so the strength or weakness of the structure as well as the mostly used concepts can be uncovered. This is important to understand how the ontology is used in the context, which part is more important for the application and to improve the ontology.

In fact, there is an additional evaluation step which is implicitly part of the ontology selection phase. Candidate ontologies need to be analyzed and evaluated whether they are approapriate or not. In this regard we propose visualization to simplify the analysis and comprehension of the content. Central
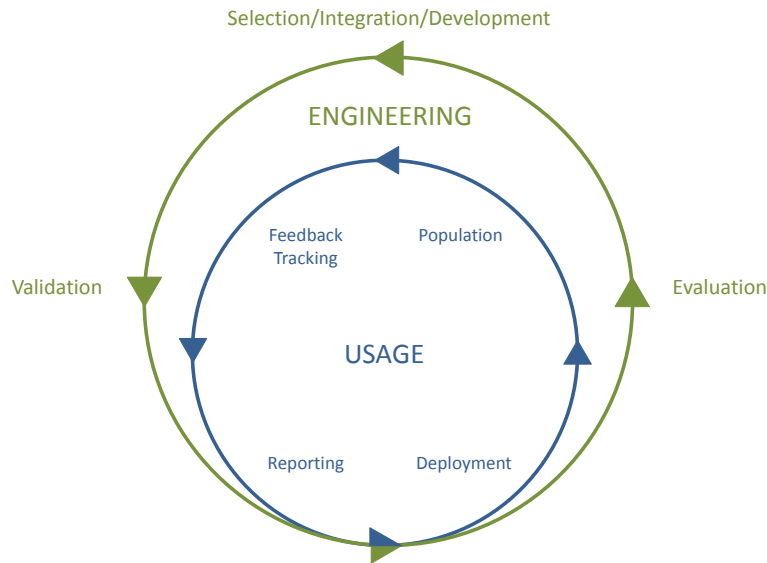
Figure 2.2: Corporate Ontology Lifecycle - COLM

concepts as well as the different subdomains need to be identifiable easily in visualization and compared to the expected concepts. The subdomains can be identified by concept grouping techniques which are based on the ontology structure.

According the methods and approaches to be utilized for ontology evaluation in a corporate context we propose to use a combination of task-based, data-driven and human-driven evaluation. The gold-standard evaluation requires an existing reference model. But it is not realistic to expect a company to have such a model. Thus, a task-based approach based on the corporate applications together with a data-driven view which considers the existing data executed by ontology engineers is necessary. In order to decide if the evaluation methodology should be more data-driven or task-based it is neccessary to have a look at the targeted system. If the ontology is needed to improve the corporate management systems by representing the management knowledge the focus will be on data-driven view in which management related documents have to be considered. On the other hand if the ontology is required to improve the companies IT products a more task-based approach will be applied which consideres the specifications and processes of the IT products.

Chapter 3

Corporate Semantic Collaboration

In a corporate environment knowledge appears and is used in various situations, e.g., as external information sources, organizational memory, or in automatic processing of events. Semantic technologies (e.g., ontologies and rules) are often the method of choice to represent knowledge. Knowledge emerges by dynamic and collaborative processes within organizations. These processes are in the focus of this chapter.

In Section 3.1 we have a look at knowledge that emerges from crowds. Considering social networks, for example, you could follow on Facebook, Twitter, and Flickr the political development in North Africa in 2011. In this situation, they were the most trusted information source. This is only a small example of emerging trends that are reflected by texts. At the moment, researches have to page through technology reports to find emerging trends. The question is if this process can be supported by trend mining algorithms.

In Section 3.2, we focus on the processing of event streams which may also utilize background knowledge to make a decision on how to handle an event. Many applications working on event stream require real-time processing of the events which may conflict with time-consuming reasoning on ontologies. We present an approach how pre-processing of complex event queries allow a semantic event processing engine to meet the real-time requirements.

As a matter of fact, organizational memory changes over time since new products are developed or new customers are acquired. Thus, ontologies backing up the knowledge base also evolve. The process of ontology evolution is collaborative work with many participants, e.g., ontology engineers, domain experts, or ontology users. In Section 3.3 we look at the status of current methodologies and tools of maintaining ontologies and point out some challenges.

3.1 Dynamic Access to Distributed Knowledge (AP 7)

The development of the Web with its technologies influences continuously many aspects of the corporate world. Collaboration and knowledge exchange in a com-

pany happens in many ways: email, IM, wikis, etc. More and more components of widely accepted communication technologies from the Web start playing an important role for the collaboration and knowledge exchange in companies: organizing and classifying information and knowledge through tagging, sharing knowledge through microblogging, networking through social networks. On the social software matrix[1] an overview of current and popular social software for companies is given. According to this overview, MS Sharepoint 2010[2] is perfectly fitting to the company's need for collaboration and knowledge exchange. Followed by Google App Pro+Wave[3], IBM Lotus Connections[4], Jive Social Business Software[5], Confluence 3.0[6] that are presented as with eighty percent of compatibility for the same needs. With the increasing importance of social networks on the Web, i.e. Facebook or Twitter, the need for an equivalent networking and microblogging services in companies emerged. Motivated by this fact in the last 12 months miKrow[7] and Yammer[8] have been and they seem to be a promising solution for knowledge management, sharing and collaboration in a corporate world. Many companies participate in social Web life maintaining their Facebook pages[9] and take the advantage of involving their potential customers and co-workers directly into the company's every day life.

Collaboration is a principle needed for knowledge sharing. The bigger a collaborating group, the more knowledge it can share. Knowledge on the Web shared by Web users allows for emerging of collective knowledge through constant information exchange, knowledge sharing and networking over the Web. This collective knowledge produced every day by crowds is of a huge importance also for companies.

In the following Sections we describe the different sources of knowledge and concentrate on the problem of deriving trends from these sources while focusing on knowledge integration.

3.1.1 Accessing Different Knowledge Sources

Access through Social Networks

Social networks like GooglePlus[10], Delicious[11], Diaspora[12], Facebook[13], Flickr[14], LinkedIn[15], Twitter[16], Xing[17], YouTube[18] etc. have become very popular

---

[1] `http://socialsoftwarematrix.org/` visited in July 2011
[2] `http://socialsoftwarematrix.org/microsoft-office-sharepoint-server/` visited in July 2011
[3] `http://socialsoftwarematrix.org/google-wave/` visited in July 2011
[4] `http://socialsoftwarematrix.org/ibm-lotus-connections/` visited in July 201
[5] `http://socialsoftwarematrix.org/jive/` visited in July 2011
[6] `http://socialsoftwarematrix.org/confluence/` seen in July 2011
[7] `http://mikrow.isoco.net/about/` visited in July 2011
[8] `https://www.yammer.com/` visited in July 2011
[9] `http://de-de.facebook.com/NeofonievisitedinJuly2011`
[10] `https://plus.google.com/` visited in July 2011
[11] `http://delicious.com/` visited June 2011
[12] `http://joindiaspora.com` visited June 2011
[13] `http://www.facebook.com` visited June 2011
[14] `http://www.flickr.com` visited June 2011
[15] `http://www.linkedin.com` visited June 2011
[16] `http://www.twitter.com` visited June 2011
[17] `http://www.xing.com` visited June 2011
[18] `http://www.youtube.com/` visited June 2011

among users on the Web. In recent years, Facebook attracted hundred millions of users worldwide, increasing its membership from over 100 million in 2009 to over 500 million in 2011[19]. Around 175 million[20] of Web users in 2010 had a Twitter account. Everyday there are 95 million[21] of tweets worldwide and "more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) each month" shared on Facebook[22]. Owing to these novel forms of communication, everybody could follow the developments during the flood in Rockhampton in Australia 2010/2011 since residents of this town created a public Facebook group reporting in real-time about the flood [23]. In mainstream media, the political events in Iran in 2009 have been described as Twitter-revolution[24] since many people communicated about these events using the microblogging service Twitter. Furthermore, the political developments, and revolutions, in North Africa beginning from January 2011 could be followed on Facebook, Twitter, Flickr, Bambuser, etc.. During this period, social networks became nearly the only trusted source of information. In Egypt, the media blackout and manipulation of facts led millions of users to extract information from several social network sources to be informed of what was really happening. Two big groups were formed in Twitter (with tag #jan25) and Facebook (R.N.N group)[25] to receive direct feed from the demonstrators in the streets.

Public Facebook status updates, tweets, bookmarks, and pictures represent immediate knowledge about our world, generated by Web users. Among this content, many reports on breaking news emerge in real-time together with the valuable knowledge about content relevant to the corporate world.

Access through Games with Purpose

Although computers and computation have been and still are being heavily developed and improved, there are still a lot of tasks computers are unable to solve [131]. Particularly, when it comes to semantics and meaning of things algorithms are dependent on human input. In the Semantic Web we have to deal with many tasks that computers cannot yet solve by themselves. This includes tasks like tagging of resources, particularly non-textual multimedia resources like images, audio and videos, locating objects in videos and the creation or alignment of ontologies [132, 118]. Briefly speaking tasks that require creativity.

The authors of [133] propose to solve such tasks by "channeling of human brainpower through computer games". Their suggestion is to wrap tasks that need "human computation" in a computer game. This way users can have fun playing the game while - as a side-effect - they solve tasks computers are unable to perform. The authors coined the term "Games With a Purpose" (GWAPs) for this class of games. GWAPs have evolved into an own area of research and can be seen as a part of the recently becoming more and more important "Gamification" field of study. According to Deterding et al. Gamification "is

---

[19]http://www.facebook.com/press/info.php?factsheet visited June 2011

[20]http://twitter.com/about visited June 2011

[21]as for September 2010

[22]http://www.facebook.com/press/info.php?statistics visited June 2011

[23]http://www.facebook.com/pages/Rockhampton-And-CQ-Floods-2010/184869738205940 visited June 2011

[24]http://www.washingtontimes.com/news/2009/jun/16/irans-twitter-revolution/ visited June 2011

[25]https://www.facebook.com/RNN.World?ref=ts visited in June 2011

an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement"[39].

Up to now there are already few implementations of GWAPs ranging from tasks of tagging images[132], videos[118], audio[70] or people[12] to folding proteins[34].

Access through Tagging

Tagging is a simple way of associating "things" on the Web, i.e. relating Twitter messages to topics through the hashtags, classifying Web pages on Delicious through word tags, or giving a meaning to pictures on Flickr with the description tags. Tag tagging is one step further in tagging[125]. It is a way of relating, classifying and giving meaning to tags itself. However, the relevance of a given tag regarding the thing that has been tagged with it, changes in time.

3.1.2 Knowledge and Trends

Trend Examples

The reports on political developments in North Africa in the period of January to February 2011 were dominated by the breaking news on protests and revolutions: started by reports of people from Tunisia overthrowing their government, followed by news via social networks from people taking part in protests in Egypt and by emerging social network updates and reports on protests in Libya and plans for protests in Algeria. Clearly, the trend for a political change from the old political system, named regime by the North African nations, to the democracy-based system in Northern African countries and part of Middle-East was raising. Every person interested in political events worldwide noticed at

Figure 3.1: A timeline-based visualization of selected reports on protests in North Africa in January to February 2011.

some time point in January 2011 that the amount of news reporting on the Egyptian situation is growing and that the events in Egypt are growing in importance. One could read from different sources the reports on unrest in Egypt and there were different opinions on how the situation is developing and what will be the consequence of this development. Owing to the novel forms of communication, Web users participating in social networks and interested in politics, could follow the events described in Twitter and Facebook posts di-

19

rectly[26][27] from the people taking part in the unrest in Egypt. Starting from January 25th 2011 Twitter was dominated by the trending hashtag[28] #jan25 untill February 13th. #jan25 was used as a tag in every message relating to Egyptian revolution. However, #jan25 that referred to January 25 2011, the "Day of Revolt" as named in the Wikipedia article[29] in Egypt started to be used also in many breaking news reports on blogs and news sites (i.e. BBC[30], AlJazeera[31]). Immediately, a Wikipedia[32] article has been written, explaining the chain of events related to the unrest taking place in Egypt starting from January 25th 2011[33]. News articles, blogs, tweets[34], posts come from different sources but are mostly texts written in a natural language. Most of them are public accessible on the Web that emerges as a constantly growing, most important information and knowledge platform . The revolution in Egypt that took place from January 25th till February 13th followed by the dimision of the former Egyptian president, is, among other trends appearing in the same period, an example of a topic that increased in interest and political relevance. Considering the 2 month period in the beginning of 2011, Egyptian revolution is an example of a (political) trend. Other common understandable examples of trends are *the financial crisis* and *insolvent companies* as emerged in 2008 in business news.
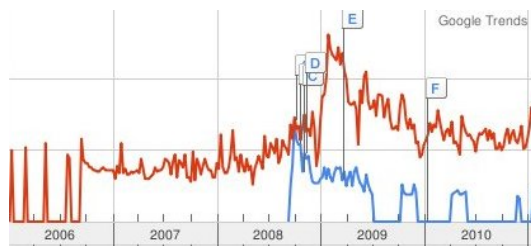
Figure 3.2: Example of search trends that shows search volume index for terms *financial crisis* (blue curve) and *insolvent* (red curve) in region Germany in time period 2006-2011. Source: GoogleTrends http://www.google.com/trends)

Derivation and Integration of Knowledge

However, trends in texts can be found in several more information fields and trend mining in textual data is useful in any application field based on temporal analysis of textual information, not only in stock exchange or politics. In many scenarios of human's daily work detection and analysis of emerging trends in

---

[26]Facebook RNN group: https://www.facebook.com/RNN.World?ref=ts visited May 2011
[27]Twitter public news lists, i.e.http://twitter.com/#!/democracynow visited in June 2011
[28]http://hashtags.org/ visited March 2011 free service in 2010/2011
[29]http://en.wikipedia.org/wiki/Timeline_of_the_2011_Egyptian_revolution_up_to_the_resignation_of_Mubarak#25_January_.E2.80.93_Day_of_Revolt
[30]http://www.bbc.co.uk/
[31]http://english.aljazeera.net/ visited June 2011
[32]http://en.wikipedia.org/ visited June 2011
[33]http://en.wikipedia.org/wiki/Egyptian_Revolution_of_2011 visited March 2011
[34]A tweet is a post from the microblogging service Twitter that has been the most popular microblogging service on the Web in the beginning of 2011. Tweets are 140 characters long messages published online.

texts is of a significant importance, i.e. in technology research, market analysis, journalism, etc. The examples of researchers paging through online technology reports in order to find emerging trends in their particular research field, market analysts reading business news and online forums in the search for trends in a given market, journalists monitoring user blogs worldwide in order to spot an astonishing development in the area of their interests - all are based on similar task; the task of identifying trends in the field of interests by reading online texts that report on this particular field of interest and appear one after another in a limited time period. Most problematic about this task is the fact that the Web continues to grow as a source of varying forms of online news and the task of reading the relevant texts on a large scale and identifying trends is hard to cope with for human. As the Web continues growing, it changes. Interesting developments appear: Web users while contributing to news using social networks (i.e. Facebook status updates), blogs(i.e. Blogger[35] community blogs) and microblogging services(i.e. Twitter messages) also contribute to classification of news articles, blog posts or microblogs, bookmarking (i.e. Delicious bookmarks) and tagging them (i.e. Twitter hash tags or blog posts tags). Additionally, Wikipedia emerges as an important source of collective knowledge on the Web.

On the other hand, the idea of Semantic Web and LinkedData initiative help in making information on the Web machine-readable and interlinked, i.e. research project Dbpedia "translates" Wikipedia information boxes into RDF data and allows for expressing of human knowledge in machine-readable format; RDFa[36] allows for easy description of information embedded in HTML and many ongoing research projects concern with schemes and ontologies that can be use on the Web (i.e. the FOAF[37], SKOS[38]). The very recent launch of the common schema[39], an initiative of Google, Bing[40], and Yahoo![41] for the structured data markup on web pages is a promising step in the realization of the Semantic Web vision. Considering Web as most important source of texts and regarding its current developments lead us to a novel viewpoint on trend mining research. Taking a knowledge-based perspective on trend mining, we propose to deliberate on the problem of knowledge integration in the process of trend detection.

Impact of Trend Research

Knowledge about emerging trends is relevant and useful for many different application fields, i.e. medical diagnosis, opinion mining, predictive analysis, market research, financial market analysis and social analytics. Regarding Gartner hype cycle from August 2010 as illustrated in 3.3, two of mentioned application fields: predictive analysis and social analytics can be found on the so called mainstream adoption timeline. Especially, social analytics arises as a technology trigger with the perspective of being adopted as a mainstream technology in 2-5 years according to Gartner. Predictive analytics is located in the "slope of enlightenment" part of Gartner curve and it is expected that it reaches the

---

[35]http://www.blogger.com
[36]http://www.w3.org/TR/xhtml-rdfa-primer/
[37]http://www.foaf-project.org
[38]http://www.w3.org/TR/skos-reference/
[39]http://www.schema.org visited in July 2011
[40]http://www.bing.com visited in July 2011
[41]http://yahoo.com visited in July 2011
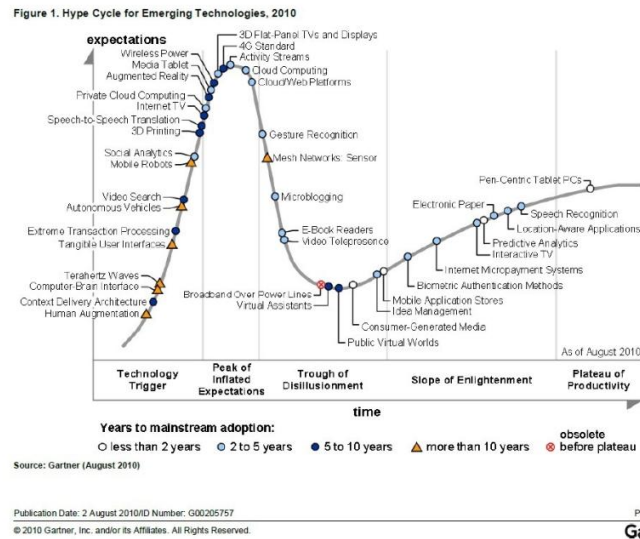
Figure 3.3: A hype cycle for emerging technologies by Gartner in August 2010

mainstream adoption before 2012, according to Gartner. Both technologies can be hardly pursued without respective trend mining approaches from information retrieval and knowledge discovery research. And, regarding trend mining approaches in the respective research fields, there is still lots of need for improvement.

3.1.3 Conclusion

In this Section we deliberated on the aspects of knowledge distribution on the Web and focused on the dynamics of knowledge: its change in time and the related problem of trend detection. In this project package we are concentrating on time-based aspects of knowledge and on trends. We are continuing experiments with the eXTS[42] tool developed in the earlier stages of the project and focusing on knowledge integrating trend mining approach.

3.2 Semantic Complex Event Processing (AP5)

This Section provides an overview about Complex Event Processing (CEP) and describes our initial concepts for the Semantic enabled CEP (SCEP). The main motivation behind this research work is to enable more intelligent event processing. The reality in many business organizations is that many complex events can not be used in process management because they are not detected in the workflows and the decision makers could not be informed about them. Detection of events is one of the critical factors for the event-driven systems and business process management. Because of current successes in business process management (BPM) and enterprise application integration (EAI), many organizations

---

[42]http://slup.imp.fu-berlin.de/exts/ visited in July 2011

know a lot about their own activities, but this huge amount of event information cannot be used in the decision making process. The permanent stream of low level events in business organizations needs an intelligence real-time event processor. The detection of presence of complex events in organization can be used to optimize the management of business processes.

The combination of event processing and semantic technologies can lead to novel semantic-rich event processing engines. These intelligent event processing engines can understand, what is happening in terms of events and can (process) state and know what reactions and processes it can invoke, and furthermore what new events it can signal. The identification of critical events and situations requires processing vast amounts of data and metadata within and outside the systems.

In cooperate setting different user activities can be considered as events, e.g., sending an email, a message in a SAP/R3 system, an RFID sensor data which informs about incoming of vehicle to the company site. The combination, aggregation and derivation of complex events from raw, simple user activities can build up complex knowledge. This real-time knowledge can be extracted by monitoring the user activities and processing of the captured event data in real time.

3.2.1 Complex Event Processing

An event is anything that happens, or is contemplated as happening.[1] Real-world occurrences can be defined as events that are happening over space and time [82, 79]. An event instance is a concrete semantic object containing data describing the event. Events can be a compound of other events that build the complex events. An event pattern is a template that matches certain sets of events [75]. CEP is about the detection of complex events from a cloud of events which are partially temporal ordered by matching complex event patterns against event instance sequences. Streams of events should be processed in one or more event processors to detect the complex events based on their syntax, sequence and their senses and semantics. Complex events can be composed or derived from other events called its members. Detection of complex events is required to be able to define and trigger reactions to the complex events.

A survey and requirements analysis about the event processing methods is provided in [113]. The existing methods for event processing can be categorized into two main categories, logic-based approaches and non-logic-based approaches. Based on these mechanisms several research prototypes are implemented for complex event processing. Some of the event processing approaches are based on the formalizations such as finite state automata [53], event-graph [30] or petri nets [51]. One of the logic-based approaches is introduced in [92] which proposes a homogeneous reaction rule language for complex event processing. It is a combinatorial approach of event and action processing, formalization of reaction rules in combination with other rule types such as derivation rules, integrity constraints, and transactional knowledge. One of the logic-based approaches is introduced in [89] which proposes a homogeneous reaction rule language for complex event processing. It is a combinatorial approach of event and action processing, formalization of reaction rules in combination with other rule types such as derivation rules, integrity constraints, and transactional knowledge.

3.2.2 Semantic Extension of CEP

At present, the areas of event processing and rules/ontologies are addressed by different research and development communities. The existing event processing approaches are dealing primarily with the syntactical processing of low-level signals, constructive event database views, streams, and primitive actions. They typically reach higher levels of scalability in computing complex events, but provide only inadequate expressiveness to describe the ontological semantics of events, processes, states, actions, and other concepts that relate to change over time. They also do not provide adequate description methods for the complex decisions, behavioral logics including expressive situations, pre- and post-conditions, complex transactional (re-) actions, and work-flow like executions. All of these are needed to declaratively represent many real-world domain problems on a higher level of abstraction.

Using semantics of events is one of the promising approaches for detection of real-world complex events. Knowledge about event types and their hierarchies i.e. specialization, generalization, or other forms of relations between events can be useful. Semantic (meta) models of events can improve the quality of event processing by using event metadata in combination with ontologies and rules (knowledge bases). Event knowledge bases can represent complex event data models which link to existing semantic domain knowledge such as domain vocabularies / ontologies and existing domain data. Semantic inference is used to infer relations between events such as, e.g., transitivity or equality between event types and their properties. Temporal and spatial reasoning on events can be done based on their data properties, e.g., a time ontology describing temporal quantities. Specific domain, task and application ontologies need to be dynamically connected and integrated into the respective event processing applications, which also leads to a modular integration approach for these ontologies. Capturing domain-specific complex events and generating complex reactions based on them is a fundamental challenge. The following shortcomings for current event processing approaches can be observed:

- **Lacking Knowledge Representation Methods:** Event processing needs a knowledge (metadata) representation methodology. The current event processing systems do not provide any knowledge representation methods for events, and there is no precise logical semantics about events and other related concepts. There is a need for methods which can include ontological semantics of events, processes, states, actions, and other concepts to the event processing without affecting the scalability and real-time processing. The questions here are: How should knowledge about events and event patterns be represented? Is it possible to represent events, actions, states, situations in background ontologies which can build up a knowledge bases about them?

- **Limited Processing Methods:** The processing approach of current event processing engines often rely on processing of simple event signals. In the existing event processing engines, events are merely implementation issues. They do not implement any usage of metadata about events and other related in the application domain. The main questions here are: Would description logic expressiveness be some kind of limitation

for real-time processing or scalability of the system? What OWL[43] sub-language/generation can we address? How can events be represented and processed with required expressiveness without effecting the real-time processing or scalability? Are real-time processing and expressiveness two conflicting goals and what is the nature of this trade-off?

3.2.3 Knowledge Representation for CEP

Ontologies play an important key role in the Semantic Complex Event Processing (SCEP). Ontologies should be the conceptualization of the application domain to allow reasoning on events and other non-event concepts. A number of different event ontologies for modeling of events and their relationships have been proposed. Shaw R. et al. provide a comparison [114] of existing event ontologies with an analysis based on their main constituent properties like, type, time, space, participation, causality and composition.

Table 3.1 lists some of the existing ontologies for representing of events. Each of these ontologies are developed for different proposes and use cases, e.g., Event ontology [139] is developed to describe events in conjunction with other music-related ontologies. Each of the event ontologies are develop to describe events in different application domains, like historical events in museums, digital libraries, musical events, agent-based system, etc. They describe some of the common concepts of events, and vary in other aspects based on the target application area. To the best of our knowledge none of the existing ontologies are specially developed and used for complex event detection and processing or for the conceptualization of complex event processing area in general.
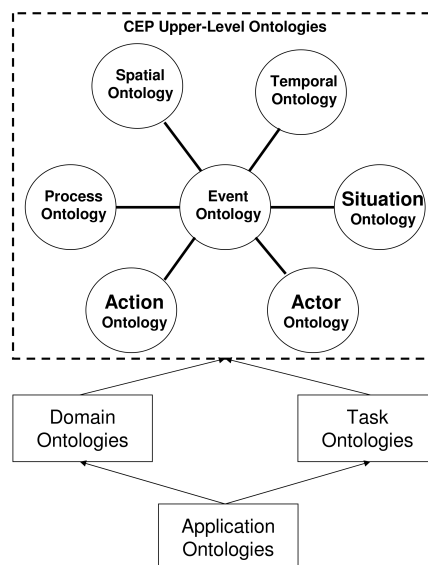

Figure 3.4: Modular Top Level Ontologies

Event Models	Description of Application Area / URL
CIDOC CRM [42]	Events in museums and libraries `http://cidoc.ics.forth.gr/OWL/cidoc_v4.2.owl`
ABC Ontology [69]	Conceptualization of events in digital libraries `http://metadata.net/harmony/ABC/ABC.owl`
The Event Ontology [139]	Description of events in digital music and in general `http://motools.sourceforge.net/event/event.html`
DOLCE+DnS Ultra-lite [49]	Upper ontology for modeling event aspects in social reality `http://www.loa-cnr.it/ontologies/DUL.owl`

Event-Model-F [112]	Enabling interoperability in distributed event-based systems
	`http://events.semantic-multimedia.org/ontology/2008/`
	`12/15/model.owl`

VUevent Model	An extension of DOLCE and other event conceptualizations `http://semanticweb.cs.vu.nl/2009/04/event/`
IPTC. EventML	Collecting and distributing structured event information `http://iptc.org/`
GEM [138]	The Geospatial Event Model
Event MultiMedia[43]	Event model for multimedia
LODE [114]	Modeling for publishing records of events as Linked Data
CultureSampo [104]	National Publication System of Cultural Heritage

EventsML-G2	A standard for conveying event information in a news industry environment
	`http://www.iptc.org/site/News_Exchange_Formats/`
	`EventsML-G2/`

Table 3.1: An Excerpt of the Ontologies for Conceptualization of Events.

Upper-Level Ontology Modules for CEP

We propose that event processing domain should be described by a modular and layered ontology model which can be reused in different application areas. Figure 3.4 shows the upper-level ontology modules for event processing domain, which can capture very general top concepts such as event, situation, process, temporal, spatial, action, agent concepts. Event ontology includes the core concepts and attributes of event and connection between modules can be realized with, e.g., ε-connection[57], Distributed Description Logics[19]. A temporal ontology captures all of the concepts related to the happening time of the event and can be seen as time ontology module. The spatio ontology includes the concepts related to the geological information of the events and answers questions like, where the event did happen, or which locations are related to the event.

Beside a general top event ontology, also specific domain-ontologies and task/application ontologies are required, e.g., to describe an event about ”stock quote price change” different attributes are needed than describing an event about ”a state change of a Web service”. It would never be a complete approach which covers all potential application domains. Events can have different relationships to each other in different domains but a simple hierarchical relationship between events is not satisfying. The reasons for modular top ontologies for CEP are the same general reasons for ontology and database modularization like, scalability, complexity management, understandability, context-awareness, personalization and reuse [124].

3.2.4 Event Processing Approach

The existing methods for event processing can be categorized into two main categories, logic-based approaches and non-logic-based approaches. Based on these mechanisms several research prototypes are implemented for complex event processing. Some of the event processing approaches are based on the formalizations such as Finite State Automata [53], Event-Graph [30] or Petri Nets [51]. For the integration and aggregation of the background domain knowledge with the incoming event stream and the timely processing of the whole knowledge a highly scalable and real time processing approach is required. Specially the inferencing on the huge amount of the background knowledge can badly effect the processing time. In the following, We discuss the different possible processing approaches and describe their different pros and cons. At the end we introduce our approach for the processing of events.

- **Storage-based Realization** The basic and more naive approach might be the storage of the incoming event data on a database and steady querying and pulling of the database. The events can be processed first after their storage on a database. The main disadvantaged of this approach is that the processing is only after storage possible and database are pulled with each new incoming event. This approach can work for use cases which do not have high event throughput and huge amount of background knowledge to process. The advantage of this approach is that a complete reasoning on the whole knowledge inventory is possible. The scalability and real time processing are the problems of this approach which makes it impossible to use it for time-sensitive use cases like algorithmic trading, or fraud-detection systems. The usage of distributed databases can improve the scalability but it can badly effect the performance.

- **Rule Engines** The most existing approaches for event processing use a rule engine for the processing of events without the permanent storage of events, the storage of historical event data is only optional for other purposes. The event data stream can easy be moved throughout the system without any necessary storage. The rule-based event processing engines can process the events in real time, because they can handle the whole facts and rules in the main processing memory. But these approaches can not achieve high scalability and high performance, when they have to process huge amount of domain background knowledge and use the available knowledge for the event detections. The main problem here is that they have to keep the whole knowledge base in the main processing memory, and this will be impossible when the background knowledge are huge, e.g., all of the background knowledge about the companies traded on the stock exchange market world wide.

Semantic Enrichment of Event Stream (SEES)

One of the possible approaches is to use the existing rule-based event processing engines and do a normal syntactic processing on events, but enrich the event stream with new derived events. These derived events are generated from the raw events and are just processed for the internal usage, e.g., if the price of OPEL company is changed the system can generate several internal events like,
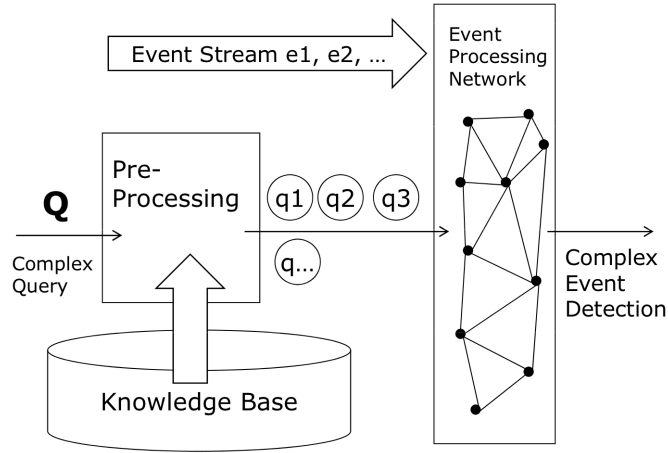
Figure 3.5: Complex Event Query Pre-Processing.

price of an automobile company is changed, price of company who produces in Germany is changed, and so on. Knowledge can be used by event mapping agents (EMAs) to generate the derived events by reasoning on the knowledge base. The mapping agents can be replicated to achieve better scalability. In the next step, the enriched event stream can be monitored by several event processing agents which have the complex event query and can process the complex query. The main disadvantage of the semantic enrichment of events can be the management of huge amount of the derived event data which are produced by incoming of each new event and should be processed by the final event processing agent to match the complex query. This can badly effect the performance of the system.

A similar initial effort is done on publish subscribe systems[97] using a taxonomy of related concepts and events are mapped to super types of other events for the semantic matching to the user subscriptions. But in this approach there is no description logic reasoning on events or other non-event concepts.

Complex Event Query Pre-Processing

In this Section, we describe our approach for the realization of SCEP engine by pre-processing of complex event queries. We propose Event Query Pre-Processing (EQPP) which means that the complex query is pre-processed before the query is executed against the incoming stream event data. The original complex event query Q_a is pre-processed under the usage of a knowledge base and divided into a set of simple event queries like $\{q_1, ...q_n\}$. A simple query is here a query which can be processed only with the information from the event stream and there is no need for using background knowledge. The extracted queries are not similar to the queries which are produced in query expansion approaches, because by the query expansion other new queries are derived which are not given in the main query. But in EQPP the new are only a partial queries of the complete query, the answer of each of the new queries might not be the result of the whole query.

The complex query Q_a can be considered as a propositional formula which

can be converted to conjunctive normal form (CNF) $Q_a \leftarrow q_1 \wedge ... \wedge q_n$, i.e., if all of the simple queries are given, then the complex event query is satisfied. The pre-processing is done by a processing agent which can access the knowledge base and divide the complex query to several simple queries. Event query can be divide into several concrete queries which includes the names of companies but not use the pattern which includes ontology specific concepts. The complex query Q_a can also be mapped in disjunctive normal form (DNF) $Q_a \leftarrow q_1 \vee ... \vee q_n$, i.e., if one of the simple queries are given, then the complex event query is satisfied. Complex query might also be mapped to a combination of both disjunctive and conjunctive of simple queries.

One important issue is that there is a need for an algorithm which can divide and partition the complex query to simple queries. Givens to this algorithm are the available domain knowledge stored in the knowledge base, the user query and the pre-defined raw event types, e.g., key/value types of events. This algorithm should consider the query structure and be able to divide it to an optimized level of granularity for the distribution over several EPAs. The algorithm should provide different sets of query partitioning, the best optimized partition can be then chosen based on other related attributes and models, e.g., cost models.

In the next step, these simple event processing queries are distributed on multiple EPAs which are connected in an Event Processing Network (EPN)[75]. One or more Mediator Agents (MA) are initiated and monitor the notifications from different EPAs and if all of the queries $(q_1 \wedge ... \wedge q_n)$ are fired within the given time window then the main complex query Q_a can be fired. These mediators can do the job of joining the partial matches in tree format similar to the concepts in distributed query processing in databases. Based on the throughput of the raw events the joins can be places in different tree levels of the matching to achieve low matching costs. The communication delays between EPAs and MAs can be considered to be very low, but these communications should also be organized in an optimized form.

Any changes in the knowledge base requires the re-processing of all or some of the complex event queries which depends on this change, and so it will require an update on some of the simple queries which are distributed over EPAs. In many use cases the knowledge base is not updated frequently, and so we can consider that such an updates can be acceptable for use cases such as algorithmic trading. For example the directory of company is not changed every minute or a company will not discharge lots of employees every hour.

The proposed approach for EQPP, their EPAs and MAs can be deployed on a network of logical hosts which can communicate to each other using a communication network. From a user perspective, all of the hosts and networks are deployed on the cloud, and a user client has a simplified communication with the cloud. The cloud of event processing network can scale up and down and be highly scalable and elastic.The main advantage of this approach is its highest scalability level because of distribution of EPAs and its high performance because of simplicities of the final queries. One disadvantage might be delay on update of simple event queries when the knowledge base is changed.

Table 3.2 provides a comparison of the above described event processing methods. The four important parameters, performance, scalability, elasticity and reasoning are compared. Performance of the event processing engine is its ability to process events in real time. Scalability of the event processing engine can have too dimensions, the first dimension is the scalability of the

	DB-Based	Rule Engine	SEES	EQPP
Performance	low	high	limited	high
Scalability	limited	limited	limited	high
Elasticity	no	no	high	high
External Reasoning	no	no/limited	high	high

Table 3.2: Comparison of Different Event Processing Approaches.

event throughput and the other one is the scalability of the fact base or the background knowledge base of the system. It can scale up with a very high throughput or with a huge amount of facts on the KB. Elasticity is needed because the throughput of event or the fact base can be very dynamic and can increase and decrease. The EP system should be an elastic system. Reasoning on knowledge base means the ability of the system to do reasoning on the reference data and not only a simple integration of reference or historical data such as what the state-of-the-art systems can already do. Each incoming event may start an inferencing process on the KB, and the results are integrated back by to the event during the event processing. But in EQPP the background reasoning is done in advance.

3.2.5 Conclusion and Outlook

In this Section, we described the main research challenge of our research on semantic complex event processing and introduced our initial work on semantic event processing approach. The main idea of our concept is to use the ontological knowledge in combination with traditional event processing and stream processing to achieve more intelligent detection and processing of events.

Our future steps are to work on more details of knowledge representation for events, situations, actions, and other related concepts. We have to work on semantic of event processing languages, and define which semantics can be an adequate semantic for event processing. Furthermore, there is need to find out adequate algorithms for mapping of complex queries to simple queries which can be processing by syntactic event processing engines.

3.3 Ontology and Knowledge Evolution through Collaborative Work (AP 8)

The employees of a company are important assets because they have know-how about processes and workflows within the company as well as knowledge about customers. Often the know-how and knowledge is only accessible by one person – the person who gained the experience. In literature the authors refer to knowledge that is difficult to transfer between individuals as *tacit knowledge*. Knowledge management (KM) is the technological and organizational approach to counter this effect. Its purpose is to make implicit knowledge explicit and to make it usable for other people.

In [72] Liebowitz compiled several definitions of knowledge which are relevant to the topic of knowledge management. In the context of the project Corporate Semantic Web we will use the definitions of Woolf and Turban:

- Knowledge is organized information applicable to problem solving. [137]

- Knowledge is information that has been organized and analyzed to make it understandable and applicable to problem solving or decision-making. [128]

In the 90s Ikujiro Nonaka started establishing the discipline of knowledge management by writing about *knowledge creating companies* [85, 86]. The focus of his work was on analyzing the dynamics of knowledge creation and especially on converting tacit knowledge into explicit knowledge. A few years later the well-known book of Davenport and Prusak [38] appeared in which they describe an alternative model of KM focusing on the design of organizational processes that allow for knowledge generation, codification, and transfer. Both models agree that groups and teams sharing a common purpose and beliefs facilitate the creation and utilization of knowledge most effectively.[44] This coincides with the requirements of today's large organizations where people have to work together in teams to benefit from the synergy of their joint knowledge. Since these organizations are often geographically spread over the world and therefore may be accustomed to different terminologies it is important to have a shared understanding of the application domain and knowledge base.

Sanchez presents in his paper [108] two major approaches for transferring knowledge within organizations: the tacit knowledge approach and the explicit knowledge approach. The fundamental thought in the first approach is that some or all of individuals' knowledge will remain tacit because it they are not able to describe it in word or to write it down. Thus, the best way of disseminating knowledge is to transfer people as knowledge carriers from one part of an organization to another. This approach supports the research results of Nonaka and Davenport. In the explicit knowledge approach the researches belief that any knowledge can be explained by individuals – even though some effort my be required. After the knowledge has been expressed explicitly as knowledge assets, e.g., in form of documents, drawings, or manuals of best practice, it can be disseminated in an organization using some kind of information system. Both approaches have its advantages and disadvantages (see [108]). Although Sanchez states that organizations without a KM concept should start with the tacit approach, he comes to the conclusion that in the long term organization should create a hybrid design for their knowledge management.

From the presented perceptions of knowledge management we conclude that organizations will have to use some approaches to represent the corporate knowledge explicitly. To ensure that the knowledge assets are understood easily they also have to ensure a shared understanding of their application domain.

Semantic technologies provide means to support the goal of creating a shared understanding, e.g., ontologies, standardized data model for representing and exchanging information as well as rule languages. Over time the knowledge represented with ontologies and rules have to be updated because the understanding of the application domain will change and evolve, e.g., a company will develop new products, acquire new customers, or establish new workflows. Further reasons for changing knowledge bases are individuals interpreting knowledge assets and discussing them among each other. In this section we will focus on the evolution of ontologies caused by collaborative work in a corporate environment. In

---

[44]For a more detailed comparison of the two models we refer the interested reader to [33] because it is not within the scope of this article.

the following we first describe requirements on knowledge management systems from the user's perspective and from a technical viewpoint. Since ontologies play an important role in the context of KM we then have a look on ontology evolution. Finally, we conclude and give and shortly describe our planned contributions.

3.3.1 Knowledge Evolution

In [8] the authors transfer the Lamarck's theory on biological evolution to the evolution of knowledge systems and characterized the evolution of knowledge systems. Using their hypothesis as foundation we describe the evolution of knowledge as follows:

- Evolution of knowledge is driven by new needs or constraints in its environment.

- New solutions are integrated into the knowledge by keeping prior characteristics and adding some new ones. One could say that they inherit acquired characteristics.

- Solutions persist as long as their objectives and goals exist. In the other case, they are integrated to the system history. As solutions are not deleted but archives, researchers coined the term *corporate memory*.

Considering the explicit knowledge approach these points basically mean that existing knowledge assets are adjusted, archived, or deleted and new one are created. Accordingly, the idea of evolution can also be applied to knowledge represented by semantic technologies such as ontologies and rules. They also are modified according to the needs and constraints of their environment, e.g., changing facts and rules.

3.3.2 Requirements on Knowledge Management

Looking at knowledge management we can distinguish between requirements considering the tasks and processes of managing knowledge and those considering the technical realization of KM systems. Based on the list compiled by Frank in [46] the first group include the following requirements:

Emphasis on concepts and reasons. Concepts and their relationships needed for the description and analysis of the corporation should be defined within the KM system.

Re-use of existing knowledge. To improve the quality of the KM system and to reduce the effort for creation and maintenance it should re-use existing knowledge, e.g., documentations of relevant causal concepts and relationships.

Support of multiple perspectives. Since the context of users (e.g., their current task, their role within an organization) determines the needed knowledge and its level of detail, a KM system should provide various perspectives on the knowledge. We

Visualization. Although visualization is closely related to the previous item we want to emphasize its importance. The KM system has to present its content in an appropriate and comprehensible way.

Integration with information. Knowledge and information[45] add value to each other. Therefore, a KM system should be able to combine knowledge with pieces of information, for example, to generate new knowledge.

Support of awareness. A KM system should provide means that users can get notified about changes in the knowledge base. For example, they could subscribe to certain types of knowledge that is relevant to them.

From a technical perspective other requirements may be important as Uren et al. used to compare KM systems in [129]. Their focus is more on knowledge representation and interoperability as the following list of requirements shows:

Standard formats. As a matter of fact KM system have to use standard formats for providing access to the knowledge base because it allows to integrate future, not envisaged tools and services can easily be integrated. Moreover, it also allows to deploy the KM system in a heterogeneous environment. Nevertheless, the system has to support well-known document formats because they are already widely used in organizations.

Evolution of ontologies and documents. Ontologies as well as documents are subject to changes. On the one hand the system has to ensure that all concepts relationships used to annotate documents are valid and consistent with the current release of the ontology. On the other hand, it has to monitor documents and adjust its annotations after it has been modified.

Annotation storage. In the Semantic Web, documents and their annotations are stored separately. This allows for querying annotations easily. However, it also leads to the challenge of linking document content and annotations. In an organizational setting, they favor integrating annotations into documents.

User centered design. Creating annotations is a resource intensive task but few organization have the capacity to employ professional annotators. Thus, simplification the annotation process should be the goal of designing its user interface allowing for creating annotations efficiently in a high quality.

Support collaboration. As mentioned before collaboration is a driver for creating new knowledge from existing resources and for disseminating knowledge. As a consequence a KM system needs to facilitate the collaboration between users.

Automation. Another important requirement is the provision of facilities to automatically annotate content or even large document collections. The quality of automatically generated annotations is a key issue in this context.

Uren et al. analyzed a couple of annotations frameworks with respect to the requirements identified by them. They came to the conclusion that most frameworks work on top of native Web formats like HTML and XML which tends to divorce the annotation process from the process of document creation. Most of the reviewed tools support a propagation of changes in ontologies to annotated documents poorly or not at all. Moreover, they do not allow changing the ontology from the user interface, e.g., to suggest new concepts based on documents or split a concept. Since the frameworks are not aware of the effects of ontology maintenance it does not surprise that provenance is currently not an issue. Vice versa, a coordinated approach is needed to tackle the issues of versioning annotations as documents evolve. Considering the storage of annotations

---

[45]In contrast to knowledge information usually refers to general data (e.g., expressed in numbers, strings, images, . . .) and can be transported and shared more easily.

both approaches are used by annotation frameworks; some keep annotations and documents separately and some integrate annotations into the document. However, we think that a hybrid approach is more desirable: Annotations should be integrated into documents and be extracted and stored in separately on save. It combines the advantages of both approaches as annotations get updated on document changes and they can be queried separately.

3.3.3 Semantic Technologies for Knowledge Evolution

Ontologies are a fundamental technology when applying semantic technologies to knowledge management. As already mentioned adapting ontologies to changing needs and constraints is an important task within KM. During the past years researchers have published methodologies to describe the lifecycle of ontologies. In [116] Simperl and Luczak-Rösch give an overview of current methodologies, methods, and techniques for collaborative ontology engineering.

In general, participants having different levels of expertise participate in the process of ontology engineering (e.g., domain experts, knowledge engineers, ontology engineers, and end-users). Furthermore, they distinguish two key roles of participants: contributors and editors. Depending on their role participants are assigned different access policies on the ontologies. For example, while contributors are limited to reporting issues and ideas and discussing them, editors are allowed to modify ontologies (e.g., integrate contributor's feedback or maintain them).

Especially in a collaborative environment, an ontology lifecycle has to consider these different user groups, roles, and access policies to keep ontologies consistent and of high quality and to keep track of modifications. In this context, a proper versioning of ontologies is essential. This is a challenge because general-purpose versioning systems as used in software engineering can hardly be applied as they compare files line by line. However, the semantics of the a single ontology can be serialized differently. From a contributor's viewpoint communication between all participants in the engineering process is important. Hereby, a key issue is to maintain a direct link between discussions and the content of the ontology they refer to. It guaranties for examples that joining participants can understand and follow on-going discussions more easily. But Simperl and Luczak-Rösch discovered that this issues is hardly addressed by any of the analyzed tools.

In [116] the authors also analyzed the characteristics of methodologies for ontology engineering. Most of them suggest an iterative process for ontology evolution which also considers feedback of ontology users. DILIGENT and HCOME are the most prominent examples of such methodologies which are based on the IBIS argumentation model (see [67]). To a limited extend collaborative approaches for ontology evolution have found wide adoption in some domains to develop ontologies.

3.3.4 Conclusion

Despite some missing features, from a technological viewpoint current methodologies and tools can be used to create and maintain knowledge and ontologies collaboratively. However, in a corporate environment we rarely find them in use. In our opinion the reason is that individuals of organizations have to study

semantic technologies to understand and to operate these methodologies and tools as they have rather complex user interfaces providing too many and unneeded functionalities. Organizations need easy-to-use tools which fit to their needs.

Thus, we will not develop new knowledge management system in the working package. Instead, we develop new concepts of interaction with semantic tools to reduce the complexity of creating semantically enriched content and managing ontologies. As a use case we consider the creation of GoodRelation descriptions of the member of the Xinnovations e.V. which are backed up by an ontology. The ontology is currently maintained by Ontonym GmbH but actually it should be refined in a collaborative manner by the members of the association.

Chapter 4

Corporate Semantic Search

4.1 Searching Non-Textual Data (AP3)

In recent years, with the continuous growth of broadband penetration and the appearance of multimedia-centric sites such as Youtube and Hulu, Flickr and Picasa and countless other similar sites, we have seen a regular explosion of non-textual data on the web. Due to the massive amounts of data these sites hold and it's non-textual nature, search and retrieval of this information has become a serious challenge. The main difference from traditional search trough textual data is that the information is not directly interpretable by machines, resulting in a search that is solely limited to the available metadata. In order to overcome these impediments complex approaches are needed for the recognition and annotation of the individual concepts present in multimedia data. These approaches are a combination of multimedia indexing techniques used for the identification of various features and concepts contained in the multimedia files and a series of metadata vocabularies which provide a predefined way to annotate the data.

4.1.1 State of the Art

Metadata Vocabularies

A series of metadata standards have been developed in order to cope with the need of annotating binary multimedia files. Each of these standards focus on a specific type of multimedia contents such as image, video or audio files and describe only certain aspects of the information contained. The most widely metadata vocabularies are EXIF and IPTC for image files, ID3 for audio files and MPEG-7 which has been mainly used for video files but which can also cover audio and image content.

- The Exchangeable Image File Format (EXIF)[1] is a metadata standard for both image and audio file formats. It is mainly used by digital cameras in order describe a series of structural image information such as resolution, orientation, exposure time etc. It can also be used to describe information with a semantic significance such as the date and time when the photo was

---

[1] http://www.exif.org/

taken or the geographic location. The main file formats that support EXIF are JPEG and TIFF, in which it can be found in the form of embedded information, other file formats such as PNG and GIF do not support this standard.

- THE IPTC Header Standard [2] was proposed due to the widespread adoption of the practice of ADOBE products to insert IPTC metadata headers into various image files edited with said products. It is similar to EXIF in that it focuses solely on images and that the information is embedded directly into the file. However, IPTC header enables the annotation with a wider range of keywords and is therefore better suited to capturing more varied metadata.

- ID3[3] is a de-facto metadata standard for audio files that allows for the annotation of these files with related information such as title, artist, track number, album, lyrics, genre, band etc. However, it is only usable in conjunction with mp3, wav and mp4 file formats, other formats such as ogg are not capable of embedding id3 tags and have their own metadata vocabularies.

- MPEG-7[4] is the most comprehensive standard for the annotation of multimedia content. It allows for the definition of rich content descriptions that enable the efficient retrieval of specific multimedia content. Contrary to the previous standards, MPEG-7 doesn't focus on a single aspect of multimedia files such as image or audio but covers the whole multimedia domain. MPEG-7 is composed of a series of modules called "tools", these include Descriptors(Ds) which represent a singular syntactic or semantic features ; Description Schemes (DSs) which are formed by a series of Descriptors and/or Description Schemes and the relationships between them; other components include Description Definition Language (DDL) whose main purpose is to provide a mechanism for the definition of new Description Schemes and Descriptors and System Tools which deal with various technical issues related to binary file inclusion, synchronization and transport.

Due to the proliferation of these overlapping standards, there is no agreed upon best practice approach for the annotation of semantically rich multimedia files. As noted in [105], the existing vocabularies are severely limited in their application domain, focus solely on a single file type, display a high degree of ambiguity and have limited to no interoperability.

Multimedia Ontologies

An efficient approach to annotation multimedia files needs to consider the numerous aspects of multimedia data that have to be annotated such as: people, places, objects, temporal criteria, frame sequences, image regions as well as the underlying semantics that represent the relationships between different the different concepts. Classical annotation technologies such as XML can capture

---

[2]http://www.iptc.org/site/Photo_Metadata/
[3]http://www.id3.org
[4]http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

various structural aspects of multimedia files, however they are ill-suited for representing complex relationships, and can only annotate information trough simple keywords and not trough associations with specialized domain ontologies. Due to this aspect a series of Ontologies have been developed that try to cover various aspects of the multimedia domain.

Mpeg-7 Ontologies

MPEG-7 is a complex standard that deals with aspects of all multimedia files and is therefore best suited for multimedia annotation. However it has 2 major drawbacks as presented in [6]:

- It's not capable of using existing Semantic Web resources such as Domain Ontologies for annotation purposes, and therefore necessitates the laborious redefinition of those concepts each time an annotation is necessary.

- Due to it's XML nature, and the way it's XML schema has been defined, it defines multiple tags for the same purpose. Furthermore it allows application to freely choose the way they use those tags, resulting in descriptions that are incompatible between applications.

In order to address these problems and to provide a richer semantic model for multimedia annotations a series of Ontologies have been developed that try to lift the MPEG-7 Standard from it's XML-based limitations to a semantic web format such as RDFS or OWL.

The first such ontology [62] was developed as part of the Harmony International Digital Library Project [5]. The Harmony ontology tries to evaluate the feasibility of using semantic-web technologies such as RDFS in order to express the semantics of the XML-based MPEG-7 descriptors and descriptor schemas. Due to the inherent complexity of the MPEG-7 standard, they apply a manual approach and model just part of the XML schema into a RDFS ontology, adhering as closely as possible to the semantics of the original schema. In order to improve the extendability of the Harmony ontology and therefore to enable the interoperability with other more specialised ontologies, Hunter has proceeded to align the Harmony ontology with the ABC Upper Ontology [68]. In order to enable more complex semantics the Harmony MPEG-7 ontology has been ported to DAML+OIL and at a later time to OWL [6]

During the execution of the aceMedia project[7], a set of two ontologies has been developed in order to model various aspects of the MPEG-7 standard. The first of these ontologies is the Multimedia Structure Ontology(MSO) whose main concern are the structural aspects of the multimedia content described by the MPEG-7 standard, whilst the Visual Descriptor Ontology (VDO) describes the low-level features of the specific multimedia content. As is the case with the Harmony ontology, the aceMedia ontology makes use of an Upper ontology in order to facilitate the integration with other ontologies such as specialized domain ontologies. However, in contrast to the ontology proposed by Hunter, the aceMedia ontology follows a modular approach by separating the structural and

---

[5] http://metadata.net/harmony/
[6] http://metadata.net/mpeg7/mpeg7.owl
[7] http://www.acemedia.org/aceMedia

visual aspects of MPEG-7 in two different ontologies. Furthermore, it introduces a series of new classes in order to disambiguate some key concepts present in MPEG-7, which are modeled only as generalised classes in the Hunter ontology.

The SmartMedia ontology[120] is similar to the previous ontologies, but is the firs to focus on the use of richer semantic in order to enable a more efficient query and reasoning functionality. In order to achieve this it makes use of SmartSUMO an upper ontology developed trough the merger of 2 of the most widely know upper ontologies, namely DOLCE and SUMO. It also differs slightly from the previous ontologies in the way the authors chose to model some aspects of the MPEG-7 schema such as the Segment class.

THE DS-MIRF ontology[127], is the first to fully capture the semantics of MPEG-7 in OWL-DL. The development of this ontology has been done manually, however the authors used a methodology that follows closely the XML schema structure and transfers it's semantics to OWL DL.

Based on the same approach of "lifting" a XML-Schema to OWL DL, Garcia et all developed the Rhizomik MPEG-7 Ontology[50]. In their paper the authors even went one step further and automated this approach by creating a tool called ReDeFer[8], which automatically converts XML Schema to OWL DL.

The DS-MIRF and Rhizomik ontologies, although being very extensive and covering almost all respectively all of the MPEG-7 specification, suffer from some of the same problems as MPEG-7 itself. These problems result from the specific modeling approach applied in their creation, which follows the MPEG-7 too closely, resulting in large, monolithic ontologies with limited extendability and reusability.

The Core Ontology for MultiMedia (COMM)[6] is one of the most recent multimedia ontologies and tries in contrast to the previous ontologies, not only to cover the MPEG-7 specification but to introduce a whole ontological framework for addressing the problems posed by multimedia annotation. In [6] the authors identify 5 major requirements for multimedia ontologies. The first of these requirements is "MPEG-7 compliance" due to the widespread existing standard. "Semantic and Syntactic Interoperability" are other 2 important requirements regarding the reuse of annotations between different applications and the meaning and structure of those annotations. Due to the mix of structural information with semantic descriptions in the MPEG-7 Standard, the authors mention the necessity of the "Separation of concerns" and try to achieve this trough a modular ontology design approach. Furthermore "Modularity" allows them to more easily and efficiently reengineer the MPEG-7 standard and also achieve a high degree of "Extensibility" in their design. In practice the ontology is designed around the DOLCE upper ontology and implements the Descriptions and Situations(D&S) and Ontology of Information Objects (OIO) design patterns from DOLCE. Furthermore the authors extended these design patterns and developed 5 new patterns in order to address the peculiarities of multimedia annotation: 1) the decomposition pattern which deals with the issue of identifying individual segments of interest in multimedia data and localizing them. 2) the content annotation pattern which specifies the way in which metadata is attached to multimedia data 3) the media annotation pattern deals mainly with the description of the physical characteristics of the specific multimedia files/segments. 4) the semantic annotation pattern enables the annotation of

---

[8]http://rhizomik.net/html/redefer/

multimedia content with concepts from domain specific ontologies. 5) Algorithm Pattern which provides a common framework for the different algorithms that need to be applied to multimedia data in order to extract meaningful information from it. In contrast to other ontologies, the COMM ontology covers almost the entire MPEG-7 spectrum while being developed in a manual, well-conceived way instead of using automatic means. It provides a rich semantic model, that has been designed with the best ontology development practices in mind. Furthermore, a series of open-source software has been developed such as an API[9] and an annotation framework called KAT[10].

Other Ontologies

Recognizing the problem posed by the different multimedia annotation formats and standards, the W3C Media Annotations Working Group[11] proceeded to develop a simple ontology with the main scope to iterate all these competing formats into a single data model. For this purpose they have developed the Media Resource Ontology (MRO)[12], which integrates over 23 different formats and standards. The main properties described by MRO are: identification, creation, content description, relational, copyright, distribution, fragments and technical properties. MRO can also be called a mapping-ontology, since its main purpose is data-integration between different formats, however it's semantic model is relatively poor and is not well-suited for more advanced tasks such as complex reasoning. The main advantage delivered by MRO is simplicity and ease of use coupled with the importance of the W3C standardisation process, MRO currently being a W3C candidate recommendation.

The Multimedia Metadata Ontology (M3O)[105] is a modular ontological framework based on the DOLCE+DnS Ultralight upper ontology. In order to create a general framework that can accurately represent MPEG-7 as well as other existing formats and standards, the authors of the COMM ontology decided to create a new ontological framework. Similarly to the COMM ontology, M3O provides multiple design patterns. The most important new designed pattern introduced with M3O is the Information Realization Pattern, which focuses on the fact that the same multimedia object can be found in different realizations (resolutions). Another new pattern is the Annotation Pattern which subclasses the Content, Media and Semantic Annotation Patterns of the COMM ontology. The third main design pattern is the Decomposition pattern which is similar role with the one in the COMM ontology. The main benefit brought by the M3O ontology is the ability to map other formats and standards except MPEG-7. For this purpose the authors developed an alignment methodology and provided predefined alignments for EXIF, XMP, Dublin Core, Yahoo!'s Search Monkey Media, ID3, the Ontology for Media Resources and of course the COMM ontology. Furthermore, the tools developed for COMM have been extended and adapted to work with the M3O ontology, and a new M3O API[13] has been developed.

---

[9]http://www.uni-koblenz.de/FB4/Institutes/IFI/AGStaab/Research/comm/API/
[10]https://launchpad.net/kat
[11]http://www.w3.org/2008/WebVideo/Annotations/
[12]http://www.w3.org/TR/mediaont-10/
[13]http://launchpad.net/m3o

4.1.2 Multimedia Search in the Corporate Context

Searching for non-textual data is an important factor in various corporate use cases. Any large company operating in the multimedia branch accumulates large amounts of non-textual data that needs to be sorted and found easily for later use. For example, a publishing company has a huge database of images and video clips. When a new article needs to be written, they require a series of images and videos to accompany the article. However, in a database that contains millions of images, audio and video clips it is impossible or highly difficult to identify all pictures of a specific person or location. Furthermore, if the requirements rise and we don't want to only find a person withing a picture, but want to ask a more complex query such as a person in a specific place at a specific event, current formats and standards for multimedia annotations cannot cope with it. Another important factor is the technological shift towards the multimedia web. With the increasing broadband speed and the wide adoption of new technologies such as HTML5 video, more and more content on the web such as news is moving from text towards video or mixed multimedia presentations. In such cases where information that used to be textual can now be found mostly in large video files, it is important for publishers to provide users the possibility to find the content they are interested in, in a fast and efficient manner.

Multimedia ontologies offer the possibility of not only attaching some keywords to a file, they allow us to express the relationship of the existing concepts found in multimedia files. In addition, more advanced ontologies such as M3O allow companies to integrate and align their existing metadata repositories, and provide ready-made tools for this purpose.

4.1.3 Conclusion and Outlook

In this section we described the chalanges faced when trying to perform searches trough non-textual data. We then described some of the existing metadata annotation formats and standards developed over time and presented their advantages and disadvantages. In order to find the best solution for this problem we presented some of the most important approaches towards ontological annotation of multimedia data.

In future steps we will proceed to implement multimedia search systems based on ontological annotation in combination with linked data resources for various use-cases provided by our partners. This will allow us to evaluate semantic multimedia search approaches based on real world scenarios.

4.2 Personalization and Contextualization of Search (AP4)

Since the explosion of information on the Web users are confronted with a huge information overload making it increasingly difficult to find relevant information for knowledge-intensive tasks or to make an optimum choice from vast amounts of alternative resources such as, for example, products or services. This problem is addressed by adaptive and adaptable software systems which aim at providing personalized and context-aware access to huge amounts of information [83].

The application of personalization and context-aware search techniques delivers the greatest benefits in environments characterized by user diversity with respect to their preferences, knowledge, goals, environmental context, etc. Such conditions can clearly be observed within business enterprises where personalization and contextualization can be targeted at internal (employees) and external (customers or business partners) users [35].

The most prominent examples of personalized and context-aware systems can be found in the field of adaptive and adaptable hypermedia systems [31] as well as recommender systems [40]. Since the former often face the problem of re-usability in a system-independent manner and mostly only work well on a fixed set of documents defined at the design time of the system, this section mainly focuses on the latter, especially due to their prevalence in e-commerce applications such as online-shops, thus making them highly relevant from the business perspective.

In the remainder of this section we first concentrate on issues regarding user modeling, also taking into account various kinds of contextual information, followed by an overview of different recommendation approaches, briefly describing their advantages and shortcomings. This section concludes with a discussion of possible improvements to classic recommender systems through the application of Semantic Web technologies as well as references some examples of implemented Semantic Web recommender systems.

4.2.1 User and Context Modeling

One of the central components of every adaptive or adaptable system is the user model. It represents information about individual users required by the system in order to provide the adaptation effect [24]. The process of creating and maintaining the user model is referred in the literature as user modeling. Depending on the information beeing modelled, we can identify models representing features of users and models that are rather concerned with the context of the user's work or search scenario. Brusilovsky and Millán provide a comprehensive overview of the most popular and useful user features and context related information relevant in user modeling [24]:

- **Interests/Preferences** are, in general, the most important, and in most cases the only, part of a user model in adaptive information retrieval and filtering systems as well as in (content-based) Web recommender systems in particular, where they are referred to as user profiles. The most common representation of user profiles, still up to this day, is a weighted vector of keywords extracted from textual data [3]. In contrast to this approach, concept-based user profiles represent user interests as an (weighted) overlay of a domain model, for example in form of an ontology [115]. Concept-based models are generally more powerful than keyword-based models due to their ability to represent user interests in a more accurate way (thus avoiding common problems associated with term ambiguity). Additionally, semantic links in the underlying domain ontology enable interest propagation onto related concepts which can be utilized to address the problem of sparsity in large overlay models. Gauch et al. [52] deliver a detailed comparison of different variations of the aforementioned user profile representations and discusses several methods for explicit and implicit

collection of user information.

- **Knowledge** as a user feature enjoys the most significance in Adaptive Educational systems, often beeing the only feature modelled. The simplest representation of user knowledge is the scalar model which expresses the degree of knowledge in a particular domain (regarded as a whole) on a predefined scale of either quantitative (e.g. from 0 to 5) or qualitative (e.g. excellent, good, average, etc.) kind. The more advanced structural model, in contrast, divides the body of domain knowledge into fragments (e.g. indicated by ontology concepts) and estimates user's knowledge level for each fragment. An example of this model implemented in the Human Resource domain is presented in [126].

- **Background** of a user relates to a collection of features regarding previous experience outside the core domain of a particular system and may include, for example, profession, certain role within a corporation, work experience in related areas, demographics, language, etc. As argued in [24], most systems do not require detailed information about user background, therefore the common way to model user background is a simple stereotype model.

- **Goals and Tasks** represent the user's immediate purpose for the interaction with an adaptive system. Especially in search scenarios, goals and tasks may also be viewed as context of a given query, which has a great impact on the quality of results delivered by (recommender) systems. For example, a user might be buying items for his or her personal use (1), items which are work related (2) or intended as a gift (3). In those cases user's personal interests have diminishing impact (from 1 to 3) on the quality of recommendations. Consequently, Anand and Mobasher [5] propose a more complex user model, distinguishing long-term interests from short-term goals, which takes this kind of contextual information better into account. The current goal of a user can be modeled as an overlay of a predefined goal catalogue of independent goals. More advanced approaches utilize a goal/task hierarchy decomposing top-level goals into sub-goals at lover hierarchy levels and/or introduce additional relations between goals/tasks in form of an ontology [73]. Due to the short-lived character of user goals as well as the difficulty and impreciseness of goal recognition most system rely on explicit specification of the current user goal.

- **Context.** In general, context can be described as additional mainly short-term information about the circumstances, objects, or conditions surrounding a user (cf. [101]). In this sense, the border between traditional features of a user model described above and context is not always clearly defined. Furthermore, user and context modeling are interrelated, since many user models incorporate context features and similar techniques are applied for modeling [24]. There also exist integrated frameworks for modeling of both context and user features - for instance the general user model and context ontology GUMO [60] represented in OWL. In particular, most commonly used categories of contextual information refer to:

User platform. Especially since the wide proliferation of various kinds of mobile devices, early context-aware systems were mainly concerned with platform adaptation [13]. Rendering content differently for desktop and mobile devices based on screen size or bandwidth are examples of the application of platform-oriented context.

Physical context includes such factors as current location and time. User location is usually represented in a coordinate-based or zone-based manner, depending on the location sensing. In context-aware adaptive systems this kind of information is used for finding nearby objects of interest. Time-related factors, such as weekday or opening hours, may be used to impose additional search constrains. The most prominent examples of applications utilizing physical context can be found in the domains of tourism and visitor guides [9] as well as cultural heritage and museum guides [141].

Human Dimension includes personal and social context. Example features of personal user context are health, mood, affective state, etc., which determination, however, greatly depends on the appropriate sensory input or explicit specification by the user. Social user context may be represented, for example, by people accompanying the user during interaction with the adaptive system (e.g. while looking for a restaurant for a group of people) or by the user's social network. Especially the latter has increasingly been analyzed within the research community exploring social graphs for improved recommendations [77].

4.2.2 State of the Art of Recommender Systems

Recommender Systems address the problem of information overload by reducing the search space to items or resources of interest to the user. Since people use different strategies to make choices about what to buy, business investments, leisure activities, etc., recommender systems aim at automating some of the decision making strategies by providing personal, affordable and high-quality recommendations [40]. Nowadays, they are an integral part of many e-commerce applications such as online-shops (e.g. Amazon[14] as the prime example) increasing customer satisfaction and retention thus leading to raise in sales [111]. A user model constitutes a central component of every recommender system, however, the way it is created and employed differs depending on the particular recommendation approach. User related information, for example, can be acquired through explicit specification by users or implicitly by monitoring their behavior. The most common recommendation techniques discussed in the literature (cf. [26], [3], [40]) are collaborative filtering, content-based filtering, and knowledge-based recommendations.

Collaborative Filtering

Collaborative filtering (CF) can briefly be defined as the process of filtering or evaluating items based on the opinions of other people [110]. The main assumption is that people who had similar preferences in the past will have similar tastes in the future. Consequently, CF mainly focuses on the social context

---

[14]http://www.amazon.com/

of users. User profiles in collaborative filtering consist of a collection of item ratings. Based on this information the similarity among users can be computed as the cosine of the angle between two rating vectors [21] or using the Pearson correlation coefficient [102]. A given user is then provided recommendations of previously unseen items which have been highly rated by those similar users.

Collaborative filtering has been the most widespread recommendation approach in large-scale e-commerce sites due to its ability to handle a very large rating matrix [109]. Another advantage of CF arises from the fact, that it does not depend on the description of items, hence it can be applied for recommending a wide range of objects like web sites, products, or multimedia resources. Moreover, CF facilitates serendipitous discoveries of items unexpected but relevant from the user's perspective. The limitations of this approach arise when new users or new items are introduced into the system. In the former case there exist not enough user ratings for the computation of similar users, in the latter situation new items cannot be recommended unless they have been rated by a substantial number of users. Furthermore, collaborative recommender systems require a critical mass of users in order to avoid the problem of rating sparsity (e.g. when an item is rarely rated, or a single user has unusual tastes) [3].

Content-based Filtering

The content-based filtering (CBF) approach generates recommendations based on characteristics of items and the specific preferences of a user [95]. To perform this task, content-based recommenders require a representation of item features as well as features the user is interested in. Both can already be represented in a structured format like database records, xml, RDF or must be extracted from textual data (e.g. texts, e-mails, Web pages). In the last case, keyword analysis techniques from information retrieval are commonly applied to remove stopwords (such as "a", "the", etc.) and to reduce variations of the same words to their root word[15]. The documents are then typically represented as a vector of TF-IDF [107] weights computed for each extracted keyword. In a similar way, a user profile is generated, based upon features extracted from items previously seen or explicitly rated by the user, and can be also be represented as a weighted feature vector. When the vector space model is used as feature representation, the cosine similarity metric [21] is applied to generate recommendations. Other approaches based on statistical learning and machine learning techniques are discussed in [40] and [3].

Unlike CF, CBF does not face the new item limitation[16] which makes it suitable for domains characterized by a short life-span of items such as news articles. The drawback of CBF, however, is that feature extraction, especially as far as multimedia items (without meaningful annotations) are concerned, is not always possible. Acquisition of subjective item qualities based on user impression is another challenging task. Moreover, on the one hand, the new user problem still applies to CBF relying on user monitoring. On the other hand, if the user interests have already been well established, CBF tends to only recommend items similar to those already seen or rated. This problem, known as overspecialization, is often addressed by introducing some randomness or by filtering out items which are too similar [14].

---

[15] using, for example, Poerter's suffix-stripping algorithm [98]
[16] since it does not rely on a critical number of item ratings

Knowledge-based Recommender Systems

The distinctive characteristic of this third type of recommender systems is that they use explicitly formalized knowledge about users and items as well as apply reasoning about what item features meet user's requirements [25]. In contrast to the approaches described above, the knowledge-based recommendation process is highly interactive: a user must specify the requirements for which the system tries to find an optimum solution through a guided feedback process. The two basic types of classic knowledge-based recommender systems (KBR) are case-based [23] and constraint-based [45] systems. The former focus on the retrieval of similar items based on different kinds of similarity measures, whereas the latter depend on explicitly defined recommendation rules.

KBR do not face the cold start problem associated with CF or CBF approaches, and are more sensitive to preference changes. Moreover, KBR systems are able to utilize complex relations between user needs and item features (e.g. preference for a romantic location may be satisfied by a restaurant with a sea view). The interactive process combined with the ability to generate explanations allows users to explore and understand the information space. The main challenge of KBR systems, however, is the creation of the knowledge base, which is a costly and time-consuming task, especially since the knowledge has to be formalized in the language of the system and in most cases can only be applied in a particular use case. Furthermore, the interactive acquisition of user preferences can lead to a problem referred to as stonewalling [119], which occurs when after a user-feedback-loop no or just few results were found and the determination of the filter criterion to be relaxed is difficult.

Hybrid Recommender Systems

The recommendation approaches described above exploit various sources of information and follow different recommendation paradigms. Each of those methods has its pros and cons making them suitable for different application domains. Hybrid recommender systems combine several algorithm implementation with the goal of overcoming their weaknesses (especially cold start and data sparsity) thereby improving the overall quality of recommendations. Burke [27] has classified several ways to combine different recommendation methods as follows:

- **weighted** - numerical combination of different recommendation components

- **switching** - selection of a single recommender based on the recommendation situation

- **mixed** - presentation of recommendations of different components side-by-side in a combined list

- **feature combination** - inputs features of one component (e.g. CF) into the algorithm of another approach (e.g. CBF)

- **feature augmentation** - generation of new features for each item by using recommendation logic of the contributing domain

- **cascade** - a secondary recommender is used to break ties in the scoring of the primary one

- **meta-level** - uses a model learned by one recommender as input for another one

The noteworthy results of Burke's empirical analysis[17], comparing 41 different hybrid recommenders, are: firstly, that cascading hybridization turned out to be a very effective means of combining recommenders of different strengths, and secondly, the utility of a knowledge-based component as a contributing component.

Semantic Web Extensions to Recommender Systems

Since the emergence of the Semantic Web [10] there has been a growing interest in the application of Semantic Web technologies to the recommendation task. Peis et al. [96] considers semantic (web) recommender systems (SWR) as any system that relies its performance on a knowledge base, defined through conceptual maps (such as taxonomy or thesaurus) or an ontology, that uses Semantic Web technologies. Considering this definition, semantic recommenders (SWR) can be viewed as a sub-category of the aforementioned knowledge-based systems. However, unlike KBR, in which the underlying knowledge-base is hard-coded into the system, SWR benefit from the improved interoperability of Semantic Web technologies thereby overcome the problem of costly and time-consuming generation of the underlying knowledge base. From another perspective, SWR can also be classified as hybrid recommenders combining CF or CBF approaches with a knowledge-based component.

The Semantic Web Technologies in recommender systems can be applied, for instance, to:

- address the challenges of feature extraction in CBF by providing formal representations of recommendable items (e.g. GoodRelations [61]) and user related information (e.g. GUMO [60])

- integrate missing or additional item features from distributed sources, e.g. Web of Linked Data [15]

- overcome the problem of data sparsity through preference spreading mechanism based on semantic relations between ontology concepts [29]

- introduce item features into computation of user similarity in CF [135]

- overcome the problem of overspecialization in CBF [80]

- address the new user problem in CBF through recommendation based on related domain concepts [136]

- introduce the use context into the recommendation process [140]

Experimental results of implemented SWRs showed, that Semantic Web technologies not only helped to address many weaknesses of traditional recommender systems but also contributed to the overall increase in recommendation quality in terms of precision and recall.

---

[17]we will not discuss the results in detail for brevity, cf. [27]

4.2.3 Conclusion and Outlook

In this section, we focused on various aspects of personalized and context-aware access to large amounts of Web resources. First, we concentrated on the user model which is the central component of every adaptive or adaptable system. We provided an overview of several kinds of user-related information with a strong emphasis on user context, and discussed various aspects of modeling. Second, we explored the state-of-the art research in the field of recommender systems, which are the primary example of systems utilizing knowledge about users and their context. We delivered an overview of both classic recommender approaches as well as discussed possible extensions to those methods, which result from the application of Semantic Web technologies. We argued that Semantic Web technologies not only provide recommender systems with a more precise understanding of the application domain, formalized in ontologies, but can also be used for a richer representation of user related information.

Since the tasks of personalization and contextualization are highly interrelated, our future research on contextualization of search will build on methods and tools developed in the first stage of the CSW project. In close cooperation with our industrial partners, we will be pursuing further development of approaches for semantic web recommender systems, extending them with context-aware features.

Chapter 5

Conclusion and Outlook

In the first phase of the BMBF funded Innoprofile Corporate Semantic Web project core methods, technologies, and tools for the realization of the Corporate Semantic Web approach have been researched and developed (see [35, 91, 90] and [1]). This report addresses the second phase of the CSW project in which further advanced CSW application domains such as multimedia content, distributed systems and knowledge, and pragmatic (business) context will be analyzed and researched. For each working package in the second phase of the CSW project, we have identified the significant problems in the field of research, outlined the current knowledge of the three problem domains of the CSW approach (semantic engineering, semantic collaboration, and semantic search), as well as surveyed the state of existing solutions.

In close collaboration with the project industry partners, we will now work on our proposed new CSW solution approaches in the next milestones of the project. The applied research methodologies will build on the results achieved in the completed working packages of the first phase of the project. Until the end of the CSW project we will contribute to the problem solutions with several new semantic design artifacts which will help to implement the CSW approach so that the described CSW vision of networked Semantic Enterprises can become a reality.

---

[1]http://www.corporate-semantic-web.de

Appendix A

Work Packages

Work package 3	**Searching non-textual data (multimedia search)**	02/11-01/13
WP 3 Task 3.1	Survey on common methods of annotating multimedia data and evaluation of their implementation in the corporate context	02/11-04/11
WP 3 Task 3.2	Conception of a method for knowledge retrieval from non-textual corporate data	05/11-07/11
Work package 4	**Search personalization**	02/11-01/13
WP 4 Task 4.1	Survey on common methods of modeling user context (location, time, device, current task, etc.) and adaption for the application in the corporate context	02/11-04/11
WP 4 Task 4.2	Conception of a method for modeling user (co-worker) context and contextualization of search results	05/11-07/11
Work package 7	**Dynamic access to distributed knowledge**	02/11-01/13
WP 7 Task 7.1	Analysis of different sources for knowledge acquisition	02/11-04/11
WP 7 Task 7.2	Conception of a method for (i) integrating knowledge from distributed heterogeneous sources and (ii) derivation of new knowledge, including identification of trends, corporate structures, or potential problems	05/11-07/11
Work package 8	**Ontology and knowledge evolution through collaborative work**	02/11-01/13
WP 8 Task 8.1	State-of-the-art survey on ontology and knowledge evolution; adaption of ontology and knowledge evolution principles and methods for the application in the corporate context	02/11-04/11

WP 8 Task 8.2	Design of a semantic method for the semi-automated evolution of ontologies or knowledge bases by analysing collaborative work	05/11-07/11
Work package 11	**Ontology cost models for enterprises**	02/11-01/13
WP 11 Task 11.1	Survey on existing cost models and analysis of their applicability to ontology engineering methods	02/11-04/11
WP 11 Task 11.2	Design of a statistical method for cost estimation of ontology development processes, ontology deployment and maintenance within corporate structures	05/11-07/11
Work package 12	**Ontology evaluation**	02/11-01/13
WP 12 Task 12.1	Analysis of existing methods for ontology evaluation	02/11-04/11
WP 12 Task 12.2	Design of a methodology for ontology evaluation with regard to usage criteria relevant for enterprises, reusability, and adaptation	05/11-07/11

Appendix B

Acknowledgement

Bibliography

[1] Event glossary 1.1, event processing technical society. `http://www.ep-ts.com`, 2008.

[2] Harold A. Linstone and Murray Turoff. *The Delphi Method : Techniques and Applications*. Addison-Wesley, 1975.

[3] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 17(6):734–749, 2005.

[4] Aj Albrecht. Measuring Application Development Productivity. In I. B. M. Press, editor, *IBM Application Development Symp.*, pages 83–92, October 1979.

[5] Sarabjot Singh Anand and Bamshad Mobasher. From web to social web: Discovering and deploying user and content profiles. chapter Contextual Recommendation, pages 142–160. Springer-Verlag, Berlin, Heidelberg, 2007.

[6] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. Comm: designing a well-founded multimedia ontology for the web. *The Semantic Web*, pages 30–43, 2007.

[7] B.F. Baird. *Managerial Decisions Under Uncertainty: An Introduction to the Analysis of Decision Making*, volume 4. Wiley-Interscience, 1989.

[8] Francoise Barthelme, Jean-Louis Ermine, and Camille Rosenthal-Sabroux. An architecture for knowledge evolution in organisations. *European Journal of Operational Research*, 109(2):414–427, September 1998.

[9] Christian Becker and Christian Bizer. Exploring the geospatial semantic web with dbpedia mobile. *Web Semant.*, 7:278–286, December 2009.

[10] T. Berners-Lee, J. Hendler, and O. Lassila. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5):34–43, 2001.

[11] Tim Berners-Lee, Christian Bizer, and Tom Heath. Linked Data-The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

[12] Michael Bernstein, Desney Tan, Greg Smith, Mary Czerwinski, and Eric Horvitz. Collabio: a game for annotating people within social networks. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 97–100, New York, NY, USA, 2009. ACM.

[13] Silvia Berti, Giulio Mori, Fabio Paternò, and Carmen Santoro. An environment for designing and developing multi-platform interactive applications. In *Proceedings of HCITALY2003*, pages 7–16. University of Turin, 2003.

[14] Daniel Billsus and Michael J. Pazzani. User modeling for adaptive news access. *User Modeling and User-Adapted Interaction*, 10:147–180, February 2000.

[15] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst*, 5(3):1–22, 2009.

[16] Barry Boehm. *Software Engineering Economics*. Prentice-Hall, Englewood Cliffs, N.J, 1981.

[17] Barry Boehm, Chris Abts, and Sunita Chulani. Software Development Cost Estimation Approaches - a Survey. *Annals of Software Engineering*, 10(1-4):177–205, January 2000. ACM ID: 590690.

[18] Barry W. Boehm, Bradford Clark, Ellis Horowitz, J. Christopher Westland, Raymond J. Madachy, and Richard W. Selby. Cost Models for Future Software Life Cycle Processes: COCOMO 2.0. *Annals of Software Engineering*, 1:57–94, 1995.

[19] Alex Borgida and Luciano Serafini. Distributed description logics: Assimilating information from peer sources. *Journal of Data Semantics*, 1:2003, 2003.

[20] J. Brank, M. Grobelnik, and D. Mladenic. A survey of ontology evaluation techniques. In *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD)*, Ljubljana, Slovenia, 2005.

[21] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. pages 43–52, 1998.

[22] C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks. Data driven ontology evaluation. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, 2004.

[23] Derek Bridge, Mehmet H. Göker, Lorraine McGinty, and Barry Smyth. Case-based recommender systems. *Knowl. Eng. Rev.*, 20:315–320, September 2005.

[24] Peter Brusilovsky and Eva Millán. The adaptive web. chapter User models for adaptive hypermedia and adaptive educational systems, pages 3–53. Springer-Verlag, Berlin, Heidelberg, 2007.

[25] Robin Burke. Knowledge-based recommender systems. In *Encyclopedia of Library and Information Systems*, volume 69, 2000.

[26] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12:331–370, November 2002.

[27] Robin Burke. The adaptive web. chapter Hybrid web recommender systems, pages 377–408. Springer-Verlag, Berlin, Heidelberg, 2007.

[28] A. Burton-Jones, V.C. Storey, V. Sugumaran, and P. Ahluwalia. A semiotic metrics suite for assessing the quality of ontologies. In *Data and Knowledge Engineering*, 2004.

[29] Iván Cantador, Alejandro Bellogín, and Pablo Castells. Ontology-based personalised and context-aware recommendations of news items. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, pages 562–565, Washington, DC, USA, 2008. IEEE Computer Society.

[30] S. Chakravarthy and D. Mishra. Snoop: an expressive event specification language for active databases. *Data Knowl. Eng.*, 14:1–26, November 1994.

[31] Sherry Y. Chen. *Adaptable and Adaptive Hypermedia Systems*. IRM Press, 2005.

[32] Gong Cheng, Weiyi Ge, and Yuzhong Qu. Falcons: searching and browsing entities on the semantic web. In *Proceeding of the 17th international conference on World Wide Web*, WWW '08, pages 1101–1102, New York, NY, USA, 2008. ACM.

[33] Chun Wei Choo. Perspectives on managing knowledge in organizations. *Cataloging and Classification Quarterly*, 37:205–220, 2003.

[34] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popovic, and Foldit players. Predicting protein structures with a multiplayer online game. *Nature*, 466:756–760, August 2010. 10.1038/nature09304.

[35] Gökhan Coşkun, Ralf Heese, Markus Luczak-Rösch, Radoslaw Oldakowski, Ralph Schäfermeier, and Olga Streibel. Towards corporate semantic web: Requirements and use cases. Technical Report B 08-09, Freie Universität Berlin, August 2008.

[36] M. d'Aquin, C. Baldassarre, L. Gridinoc, M. Sabou, S. Angeletou, and E. Motta. Watson: Supporting next generation semantic web applications. In *Proceedings of the WWW/Internet conference, Vila real, Spain, 2007*, 2007.

[37] Mathieu d'Aquin and Holger Lewen. Cupboard - a place to expose your ontologies to applications and the community. In Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvnen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Paslaru Bontas Simperl, editors, *ESWC*, volume 5554 of *Lecture Notes in Computer Science*, pages 913–918. Springer, 2009.

[38] Thomas H. Davenport and Lawrence Prusak. *Working Knowledge: How Organizations Manage What They Know.* Harvard Business School Press, Cambridge, MA, 1998.

[39] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. Gamification. using game-design elements in non-gaming contexts. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 2425–2428, New York, NY, USA, 2011. ACM.

[40] et al. Dietmar Jannach. *Recommender Systems: An Introduction.* Cambridge University Press, 2011.

[41] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R. Scott Cost, Yun Peng, Pavan Reddivari, Vishal Doshi, and Joel Sachs. Swoogle: a search and metadata engine for the semantic web. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, CIKM '04, pages 652–659, New York, NY, USA, 2004. ACM.

[42] Martin Doerr. The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI Mag.*, 24(3):75–92, 2003.

[43] Max J. Egenhofer, Christian Freksa, and Harvey J. Miller, editors. *Geographic Information Science, Third International Conference, GIScience 2004, Adelphi, MD, USA, October 20-23, 2004, Proceedings*, volume 3234 of *Lecture Notes in Computer Science.* Springer, 2004.

[44] Marc Ehrig, Peter Haase, Mark Hefke, and Nenad Stojanovic. Similarity for ontologies - a comprehensive framework. In *ECIS*, 2005.

[45] Alexander Felfernig and Robin Burke. Constraint-based recommender systems: technologies and research issues. In *Proceedings of the 10th International Conference on Electronic Commerce*, ICEC '08, pages 1–10, New York, NY, USA, 2008. ACM.

[46] Ulrich Frank. Knowledge management systems: Essential requirements and specific design patterns. In Waleed W. Smari, N. Melab, and K. Yetognon, editors, *Proceedings of the International Symposium on Information Systems and Engineering: ISE"2001*, pages 114–121. CSREA Press, 2001.

[47] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. Ontology evaluation and validation - an integrated formal model for the quality diagnostic task. Technical report, LOA , ISTC-CNR, 2005. Available at: http://www.loa-cnr.it/Files/OntoEval4OntoDev_Final.pdf.

[48] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. Modelling ontology evaluation and validation. In York Sure and John Domingue, editors, *ESWC*, volume 4011 of *Lecture Notes in Computer Science*, pages 140–154. Springer, 2006.

[49] Aldo Gangemi and Peter Mika. Understanding the semantic web through descriptions and situations. In *CoopIS/DOA/ODBASE*, pages 689–706, 2003.

[50] R. Garcia and O. Celma. Semantic integration and retrieval of multimedia metadata. In *2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media*, pages 69–80. Citeseer, 2005.

[51] Stella Gatziu and Klaus R. Dittrich. Detecting compositeevents in active database systems using petri nets. In *Fourth International Workshop on Research Issues in Data Engineering: Active Database Systems*, 1994.

[52] Susan Gauch, Mirco Speretta, Aravind Chandramouli, and Alessandro Micarelli. User profiles for personalized information access. In *The Adaptive Web: Methods and Strategies of Web Personalization*, chapter 2, pages 54–89. 2007.

[53] Narain H. Gehani, H. V. Jagadish, and Oded Shmueli. Composite event specification in active databases: Model & implementation. In *VLDB '92: Proceedings of the 18th International Conference on Very Large Data Bases*, pages 327–338, San Francisco, CA, USA, 1992. Morgan Kaufmann Publishers Inc.

[54] Asunción Gómez-Pérez. Ontology evaluation. In *Handbook on Ontologies*, pages 251–274. Springer, 2004.

[55] T. R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. *Int. Journal Human-Computer Studies*, 43(5/6), 1995.

[56] Michael Gr'uninger and Mark S. Fox. Methodology for the design and evaluation of ontologies. In *International Joint Conference on Artificial Inteligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995.

[57] Volker Haarslev and Ralf Möller. Racer system description. In *IJCAR '01: Proceedings of the First International Joint Conference on Automated Reasoning*, pages 701–706, London, UK, 2001. Springer-Verlag.

[58] Peter Haase and York Sure. Usage tracking for ontology evolution. SEKT Deliverable D3.2.1, Institute AIFB, University of Karlsruhe, 2005.

[59] Andreas Harth, Aidan Hogan, Jürgen Umbrich, and Stefan Decker. Swse: Objects before documents! In *Proceedings of the 7th International Semantic Web Conference, Semantic Web Challenge 2008*, volume 5318. Springer, October 2009.

[60] Dominik Heckmann, Eric Schwarzkopf, Junichiro Mori, Dietmar Dengler, and Alexander Kröner. The user model and context ontology gumo revisited for future web 2.0 extensions. In Paolo Bouquet, Jérôme Euzenat, Chiara Ghidini, Deborah L. McGuinness, Luciano Serafini, Pavel Shvaiko, and Holger Wache, editors, *C&O:RR*, volume 298 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2007.

[61] Martin Hepp. Goodrelations: An ontology for describing products and services offers on the web. In *Proceedings of the 16th international conference on Knowledge Engineering: Practice and Patterns*, EKAW '08, pages 329–346, Berlin, Heidelberg, 2008. Springer-Verlag.

[62] J. Hunter. Adding multimedia to the semantic web-building an mpeg-7 ontology. In *International Semantic Web Working Symposium (SWWS)*. Citeseer, 2001.

[63] Aleks Jakulin and Dunja Mladenić. Ontology grounding. In *Proceedings of 8th International multi-conference Information Society IS-2005*, pages 170–173, 2005.

[64] R. Jensen. An Improved Macrolevel Software Development Resource Estimation Model. In *5th ISPA Conference*, pages 88–92, 1983.

[65] C. Jones. *Applied software measurement: global analysis of productivity and quality*. McGraw-Hill, 2008.

[66] J. M. Juran and A. Blanton-Godfrey. *Juran's Quality Handbook*. McGraw-Hill, Columbus, OH, U.S.A., 5ª edition, 1998.

[67] Werner Kunz and Horst Rittel. Issues as elements of information systems. Working Paper 131, Institute of Urban and Regional Development, University of California, Berkeley, California, 1970.

[68] C. Lagoze and J. Hunter. The abc ontology and model. *Journal of Digital Information*, 2(2):77, 2001.

[69] Carl Lagoze and Jane Hunter. The abc ontology and model. In *DCMI '01: Proceedings of the International Conference on Dublin Core and Metadata Applications 2001*, pages 160–176. National Institute of Informatics, Tokyo, Japan, 2001.

[70] Edith L. M. Law, Luis Von Ahn, Roger B. Dannenberg, and Mike Crawford. Tagatune: A game for music and sound annotation. In *Proceedings of the 8th International Conference on Music Information Retrieval*, 2007.

[71] D. Lee. Norden Raleigh Analysis: A Useful Tool for EVM in Development Projects. *The Measurable News*, 2002.

[72] Jay Liebowitz. *Knowledge Management Handbook*. CRC Press, Inc., Boca Raton, FL, USA, 1999.

[73] Yun Lin and Arne Sølvberg. Goal annotation of process models for semantic enrichment of process knowledge. In *Proceedings of the 19th international conference on Advanced information systems engineering*, CAiSE'07, pages 355–369, Berlin, Heidelberg, 2007. Springer-Verlag.

[74] A Lozano-Tello and A Gomez-Perez. ONTOMETRIC: A method to choose the appropriate ontology. *JOURNAL OF DATABASE MANAGEMENT*, 15(2):1–18, APR-JUN 2004.

[75] David Luckham. *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*. Addison-Wesley Longman, Amsterdam, 2002.

[76] Markus Luczak-Rösch. The Methodology, Methods and Tools for Agile Ontology Maintenance - A Status Report. In Ruzica Piskac and Elena Simperl, editors, *Proceedings of the Doctoral Consortium of the Future Internet Symposium 2010*, volume 623 of *CEUR Workshop Proceedings*. CEUR-WS, September 2010.

[77] Ashwin Machanavajjhala, Aleksandra Korolova, and Atish Das Sarma. Personalized social recommendations - accurate or private? *PVLDB*, 4(7):440–450, 2011.

[78] Alexander Maedche and Steffen Staab. Measuring similarity between ontologies. In Asuncin Gmez-Prez and V. Richard Benjamins, editors, *EKAW*, volume 2473 of *Lecture Notes in Computer Science*, pages 251–263. Springer, 2002.

[79] Chandy K. Mani and Schulte W. Roy. *Event Processing: Designing IT Systems for Agile Companies*. McGraw-Hill, Inc., New York, NY, USA, 2009.

[80] Stuart E. Middleton, Nigel R. Shadbolt, and David C. De Roure. Ontological user profiling in recommender systems. *ACM Trans. Inf. Syst.*, 22:54–88, January 2004.

[81] A. Minkiewicz. Measuring Object Oriented Software with Predictive Object Points. Technical report, PRICE Systems, 1998.

[82] Gero Mühl, Ludger Fiege, and Peter R. Pietzuch. *Distributed event-based systems*. Springer, 2006.

[83] Peter Brusilovsky; Alfred Kobsa; Wolfgang Nejdl, editor. *The Adaptive Web: Methods and Strategies of Web Personalization*. Springer Verlag, 2007.

[84] E. A. Nelson. Management Handbook for the Estimation of Computer Programming Costs. Tech Memo, System Development Corporation, 1967.

[85] Ikujiro Nonaka. The knowledge-creating company. *Harvard Business Review*, pages 96–104, November-Dezember 1991.

[86] Ikujiro Nonaka and Hirotaka Takeuchi. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, New York, NY, 1995.

[87] Leo Obrst, Werner Ceusters, Inderjeet Mani, Steve Ray, and Barry Smith. The evaluation of ontologies. In Christopher J.O. Baker and Kei-Hoi Cheung, editors, *Revolutionizing Knowledge Discovery in the Life Sciences*, chapter 7, pages 139–158. Springer, 2007.

[88] R. Park. The Central Equations of the PRICE Software Cost Model. Technical report, 4th COCOMO Users' Group Meeting, 1988.

[89] Adrian Paschke. A homogenous reaction rule language for complex event processing. In *In Proc. 2nd International Workshop on Event Drive Architecture and Event Processing Systems (EDA-PS*, 2007.

[90] Adrian Paschke, Gökhan Coşkun, Dennis Hartrampf, Ralf Heese, Markus Luczak-Rösch, Mario Rothe, Radoslaw Oldakowski, and Ralph Schäfermeierand Olga Streibel. Realizing the corporate semantic web: prototypical implementations. Technical Report TR-B-10-05, Freie Universität Berlin, 2010.

[91] Adrian Paschke, Gökhan Coşkun, Ralf Heese, Markus Luczak-Rösch, Radoslaw Oldakowski, Ralph Schäfermeier, and Olga Streibel. Realizing the corporate semantic web: Concept papers. Technical Report TR-B-08-09, Freie Universität Berlin, April 2009.

[92] Adrian Paschke, Alexander Kozlenkov, and Harold Boley. A homogeneous reaction rule language for complex event processing. *CoRR*, abs/1008.0823, 2010.

[93] E. Paslaru Bontas Simperl, C. Tempich, and Y. Sure. ONTOCOM: A Cost Estimation Model for Ontology Engineering. *The Semantic Web-ISWC 2006*, pages 625–639, 2006.

[94] Chintan Patel, Kaustubh Supekar, Yugyung Lee, and E.K. Park. Ontokhoj a semantic web portal for ontology searching, ranking, and classification. In *Proc. 5th ACM Int. Workshop on Web Information and Data Management*, pages 58–61, New Orleans, Louisiana, USA, 2003.

[95] Michael J. Pazzani and Daniel Billsus. The adaptive web. chapter Content-based recommendation systems, pages 325–341. Springer-Verlag, Berlin, Heidelberg, 2007.

[96] E. Peis, J. M. Morales del Castillo, and J. A. Delgado-López. Semantic recommender systems. analysis of the state of the topic. *Hipertext.net*, 6:(online), 2008. Author: E. Peis; J. M. Morales-del-Castillo; J. A. Delgado-López Location in hierarchy: Home - ¿ Number 6, May 2008 -¿ Recommender URL of this page: http://www.hipertext.net/english/pag1031.htm Citation: E. Peis; J. M. Morales-del-Castillo; J. A. Delgado-López. Semantic Recommender Systems. Analysis of the state of the topic [on line]. "Hipertext.net", num. 6, 2008. ¡http://www.hipertext.net¿. ISSN 1695-5498.

[97] Milenko Petrovic, Ioana Burcea, and Hans-Arno Jacobsen. S-topss: semantic toronto publish/subscribe system. In *Proceedings of the 29th international conference on Very large data bases - Volume 29*, VLDB '2003, pages 1101–1104. VLDB Endowment, 2003.

[98] M. F. Porter. *An algorithm for suffix stripping*, pages 313–316. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997.

[99] R. Porzel and R. Malaka. A task-based approach for ontology evaluation. In *Proceedings of the 16th European Conference on Artificial Intelligence*, Valencia, Spain, 2004.

[100] L. H. Putnam. A General Empirical Solution to the Macro Software Sizing and Estimating Problem. *IEEE Transactions on Software Engineering*, 4:345–361, July 1978.

[101] Anand Ranganathan and Roy H. Campbell. An infrastructure for context-awareness based on first order logic. *Personal Ubiquitous Comput.*, 7:353–364, December 2003.

[102] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer supported cooperative work*, pages 175–186, New York, NY, USA, 1994. ACM.

[103] H.A. Rubin. Macro-Estimation of Software Development Parameters: The Estimacs System. In *SOFTFAIR Conference on Software Development Tools, Techniques and Alternatives, Arlington,*, pages 4–16. IEEE Press, 1983.

[104] Tuukka Ruotsalo and Eero Hyvönen. A method for determining ontology-based semantic relevance. In Roland Wagner, Norman Revell, and Günther Pernul, editors, *DEXA*, volume 4653 of *Lecture Notes in Computer Science*, pages 680–688. Springer, 2007.

[105] C. Saathoff and A. Scherp. Unlocking the semantics of multimedia presentations in the web with the multimedia metadata ontology. In *Proceedings of the 19th international conference on World wide web*, pages 831–840. ACM, 2010.

[106] Marta Sabou, Jorge Gracia, Sofia Angeletou, Mathieu d'Aquin, and Enrico Motta. Evaluating the semantic web: A task-based approach. In *ISWC/ASWC*, pages 423–437, 2007.

[107] Gerard Salton. *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[108] Ron Sanchez. ïtacit knowledgeï versus ïexplicit knowledgeï approaches to knowledge management practice. IVS/CBS Working Papers 2004-01, Department of Industrial Economics and Strategy, Copenhagen Business School, 2004.

[109] Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.

[110] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative filtering recommender systems. In Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors, *The adaptive web*, pages 291–324. Springer-Verlag, Berlin, Heidelberg, 2007.

[111] J. Ben Schafer, Joseph A. Konstan, and John Riedl. E-commerce recommendation applications. *Data Min. Knowl. Discov.*, 5:115–153, January 2001.

[112] Ansgar Scherp, Thomas Franz, Carsten Saathoff, and Steffen Staab. F–a model of events based on the foundational ontology dolce+dns ultralight. In *K-CAP '09: Proceedings of the fifth international conference on Knowledge capture*, pages 137–144, New York, NY, USA, 2009. ACM.

[113] Kay-Uwe Schmidt, Darko Anicic, and Roland Stühmer. Event-driven reactivity: A survey and requirements analysis. In *SBPM2008: 3rd international Workshop on Semantic Business Process Management in conjunction with the 5th European Semantic Web Conference (ESWC'08)*. CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073), June 2008.

[114] Ryan Shaw, Raphaël Troncy, and Lynda Hardman. Lode: Linking open descriptions of events. (Paper 2009-036), August 2009.

[115] Ahu Sieg, Bamshad Mobasher, and Robin Burke. Web search personalization with ontological user profiles. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, CIKM '07, pages 525–534, New York, NY, USA, 2007. ACM.

[116] Elena Simperl and Markus Luczak-Rösch. Collaborative ontology engineering: A survey. *The Knowledge Engineering Review*, 26, 2011. to appear.

[117] Elena Simperl, Igor O. Popov, and Tobias Bürger. ONTOCOM Revisited: Towards Accurate Cost Predictions for Ontology Development Projects. In Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl, editors, *The Semantic Web: Research and Applications*, volume 5554, pages 248–262. Springer, Berlin, Heidelberg, 2009.

[118] Katharina Siorpaes and Martin Hepp. Games with a purpose for the semantic web. *IEEE Intelligent Systems*, 23:50–60, 2008.

[119] Barry Smyth. The adaptive web. chapter Case-based recommendation, pages 342–376. Springer-Verlag, Berlin, Heidelberg, 2007.

[120] D. Sonntag and M. Romanelli. A multimodal result ontology for integrated semantic web dialogue applications. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, pages 24–26. Citeseer, 2006.

[121] P. Spyns. EvaLexon: Assessing triples mined from texts. Technical Report 09, Star Lab, Brussels, Belgium, 2005.

[122] Ljiljana Stojanovic, Nenad Stojanovic, Jorge Gonzalez, and Rudi Studer. Ontomanager – a system for the usage-based ontology management. In Robert Meersman, Zahir Tari, and Douglas Schmidt, editors, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, volume 2888 of *Lecture Notes in Computer Science*, pages 858–875. Springer Berlin / Heidelberg, 2003.

[123] Nenad Stojanovic and Ljiljana Stojanovic. Usage-Oriented Evolution of Ontology-Based Knowledge Management Systems. In Robert Meersman and Zahir Tari, editors, *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, volume 2519 of *Lecture Notes in Computer Science*, pages 1186–1204. Springer Berlin, Heidelberg, 2002.

[124] Heiner Stuckenschmidt, Christine Parent, and Stefano Spaccapietra, editors. *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, volume 5445 of *Lecture Notes in Computer Science*. Springer, Berlin, 2009.

[125] Vlad Tanasescu and Olga Streibel. Extreme tagging: Emergent semantics through the tagging of tags. In *International Workshop on Emergent Semantics and Ontology Evolution*, volume 292, pages 84–94, 2007.

[126] Robert Tolksdorf, Malgorzata Mochol, Ralf Heese, Rainer Eckstein, Radoslaw Oldakowski, and Christian Bizer. Semantic-web-technologien im arbeitsvermittlungsprozess. *Wirtschaftsinformatik*, 1, 2006.

[127] C. Tsinaraki, P. Polydoros, and S. Christodoulakis. Interoperability support between mpeg-7/21 and owl in ds-mirf. *IEEE transactions on knowledge and data engineering*, pages 219–232, 2007.

[128] Efraim Turban and Louis E. Frenzel. *Expert Systems and Applied Artificial Intelligence*. Prentice Hall Professional Technical Reference, 1992.

[129] V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, and F. Ciravegna. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1):14–28, 2006.

[130] P. Velardi, R. Navigli, A. Cucchiarelli, and F. Neri. Evaluation of ontolearn, a methodology for automatic learning of ontologies. In P. Buitelaar, P. Cimmiano, and B. Magnini, editors, *Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press, 2005.

[131] Luis von Ahn. Games with a purpose. *Computer*, 39:92–94, 2006.

[132] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, pages 319–326, New York, NY, USA, 2004. ACM.

[133] Luis von Ahn and Laura Dabbish. Designing games with a purpose. *Commun. ACM*, 51:58–67, August 2008.

[134] Denny Vrandečić. *Ontology Evaluation*. PhD thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany, June 2010.

[135] R.Q. Wang and F.S. y Kong. Semantic-enhanced personalized recommender system. In *Proc. of the Int. Conference on Machine Learning and Cybernetics*, volume 7, pages 4069–4074, 2007.

[136] Yiwen Wang, Natalia Stash, Lora Aroyo, Peter Gorgels, Lloyd Rutledge, and Guus Schreiber. Recommendations based on semantically enriched museum collections. *Web Semant.*, 6:283–290, November 2008.

[137] H. Woolf, editor. *Webster's new world dictionary of the American language.* G. and C. Merriam, 1990.

[138] Michael F. Worboys and Kathleen Hornsby. From objects to events: Gem, the geospatial event model. In Egenhofer et al. [43], pages 327–344.

[139] S. Abdallah Y. Raimond. The event ontology. `http://motools.sourceforge.net/event/event.html`.

[140] Zhiwen Yu, Yuichi Nakamura, Seiie Jang, Shoji Kajita, and Kenji Mase. Ontology-based semantic recommendation for context-aware e-learning. In Jadwiga Indulska, Jianhua Ma, Laurence Tianruo Yang, Theo Ungerer, and Jiannong Cao, editors, *Ubiquitous Intelligence and Computing, 4th International Conference, UIC 2007, Hong Kong, China, July 11-13, 2007, Proceedings,* volume 4611 of *Lecture Notes in Computer Science,* pages 898–907. Springer, 2007.

[141] Andreas Zimmermann, Marcus Specht, and Andreas Lorenz. Personalization and context management. *User Modeling and User-Adapted Interaction,* 15(3-4):275–302, August 2005.