D 1.1.2 Prototypical Business Use Cases

Lyndon Nixon & Malgorzata Mochol (FU Berlin)

with contributions from:
Alain Leger, François Paulus, Laurent Rocuet (France Telecom),
Matteo Bonifacio, Roberta Cuel (Uni Trento),
Mustafa Jarrar, Pieter Verheyden (VUB),
Yiannis Kompatsiaris, Vasileios Papastathis, Stamatia Dasiopoulou (CERTH),
Asunción Gómez Pérez (UPM)

Abstract.
EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB
Deliverable D1.1.2 (WP1.1)
This document provides a set of business cases on how Semantic Web technologies have solved or could hypothetically solve concrete business issues which are relevant in respect to some strategic industries. It also considers future developments in research and industry which are of relevance to the deployment of Semantic Web technologies in business cases. From this an analysis of the potential of and the needs for deployment of Semantic Web solutions in industry is carried out and a few key use cases are identified.

Document Identifier:	KWEB/2004/D1.1.2/v1.0
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	V1.0
Date:	January 31, 2005
State:	Final
Distribution:	Public

Knowledge Web Consortium

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

University of Innsbruck (UIBK) – Coordinator
Institute of Computer Science,
Technikerstrasse 13
A-6020 Innsbruck
Austria
Contact person: Dieter Fensel
E-mail address: dieter.fensel@uibk.ac.at

École Polythechnique Fédérale de Lausanne (EPFL)
Computer Science Department
Swiss Federal Institute of Technology
IN (Ecublens), CH-1015 Lausanne.
Switzerland
Contact person: Boi Faltings
E-mail address: boi.faltings@epfl.ch

France Telecom (FT)
4 Rue du Clos Courtel
35512 Cesson Sévigné
France. PO Box 91226
Contact person : Alain Leger
E-mail address: alain.leger@rd.francetelecom.com

Freie Universität Berlin (FU Berlin)
Takustrasse, 9
14195 Berlin
Germany
Contact person: Robert Tolksdorf
E-mail address: tolk@inf.fu-berlin.de

Free University of Bozen-Bolzano (FUB)
Piazza Domenicani 3
39100 Bolzano
Italy
Contact person: Enrico Franconi
E-mail address: franconi@inf.unibz.it

Institut National de Recherche en Informatique et en Automatique (INRIA)
ZIRST - 655 avenue de l'Europe - Montbonnot
Saint Martin
38334 Saint-Ismier
France
Contact person: Jérôme Euzenat
E-mail address: Jerome.Euzenat@inrialpes.fr

Centre for Research and Technology Hellas / Informatics and Telematics Institute (ITI-CERTH)
1st km Thermi – Panorama road
57001 Thermi-Thessaloniki
Greece. Po Box 361
Contact person: Michael G. Strintzis
E-mail address: strintzi@iti.gr

Learning Lab Lower Saxony (L3S)
Expo Plaza 1
30539 Hannover
Germany
Contact person: Wolfgang Nejdl
E-mail address: nejdl@learninglab.de

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

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

Universidad Politécnica de Madrid (UPM)
Campus de Montegancedo sn
28660 Boadilla del Monte
Spain
Contact person: Asunción Gómez Pérez
E-mail address: asun@fi.upm.es

University of Karlsruhe (UKARL)
Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB
Universität Karlsruhe
D-76128 Karlsruhe
Germany
Contact person: Rudi Studer
E-mail address: studer@aifb.uni-karlsruhe.de

University of Liverpool (UniLiv)
Chadwick Building, Peach Street
L697ZF Liverpool
United Kingdom
Contact person: Michael Wooldridge
E-mail address: M.J.Wooldridge@csc.liv.ac.uk

University of Manchester (UoM)
Room 2.32. Kilburn Building, Department of Computer Science, University of Manchester,
Oxford Road
Manchester, M13 9PL
United Kingdom
Contact person: Carole Goble
E-mail address: carole@cs.man.ac.uk

University of Sheffield (USFD)
Regent Court, 211 Portobello street
S14DP Sheffield
United Kingdom
Contact person: Hamish Cunningham
E-mail address: hamish@dcs.shef.ac.uk

University of Trento (UniTn)
Via Sommarive 14
38050 Trento
Italy
Contact person: Fausto Giunchiglia
E-mail address: fausto@dit.unitn.it

Vrije Universiteit Amsterdam (VUA)
De Boelelaan 1081a
1081HV. Amsterdam
The Netherlands
Contact person: Frank van Harmelen
E-mail address: Frank.van.Harmelen@cs.vu.nl

Vrije Universiteit Brussel (VUB)
Pleinlaan 2, Building G10
1050 Brussels
Belgium
Contact person: Robert Meersman
E-mail address: robert.meersman@vub.ac.be

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to writing parts of this document:

Centre for Research and Technology Hellas
France Telecom
Freie Universität Berlin
Universidad Politécnica de Madrid
University of Trento
Vrije Universiteit Brussel

Changes

Version	Date	Author	Changes
0.01	26-04-2004	Lyndon Nixon	Proposed bare structure
0.02	28-05-2004	Lyndon Nixon	Revised bare structure
0.1	01-06-2004	Lyndon Nixon	Added first use case
0.2	23-06-2004	Lyndon Nixon	First draft with introduction and text on emerging business issues/Semantic Web research
0.3	30-06-2004	Lyndon Nixon	Interim version: added three more use cases and conclusion
0.4	07-07-2004	Lyndon Nixon	Added fifth use case, edited document
0.45	16-07-2004	Lyndon Nixon & Malgorzata Mochol	Revised diagrams & text
0.5	19-07-2004	Lyndon Nixon	Added sixth use case
0.6	19-10-2004	Lyndon Nixon	Reformatted four use cases
0.75	16-11-2004	Lyndon Nixon	Added new and reformatted use cases for a total of eleven
0.8	19-11-2004	Lyndon Nixon	Edit of introductory and concluding text following Berlin meeting discussions
0.9	02-12-2004	Lyndon Nixon	Pre-final edit of text with inclusion of industry sector analysis and new version of chapter 3
0.95	15-12-2004	Lyndon Nixon	Addition of 2 new use cases Added analysis of use cases Revision of text for consistency
0.97	21-12-2004	Lyndon Nixon	Addition of 14th use case
0.99	22-12-2004	Lyndon Nixon	Addition of 15th and 16th use cases
1.0	07-01-2005	Lyndon Nixon	Revision of text for legibility and understandability; editing as a result of the quality assessment
	31-01-2005	Lyndon Nixon	Final minor changes (and conversion of graphs to B&W) as a result of the quality control

Executive Summary

The KnowledgeWeb Network of Excellence has as one of its key goals the transference of ontology-based technologies (often referred to when deployed over the Web infrastructure as the Semantic Web) from the field of academia to industry.

As part of achieving this goal, the Industry Area of KnowledgeWeb is tasked with, among other things:

- Forming an industrial board of firms interested and active in the application of Semantic Web technologies
- Specifying business needs scenarios
- Identifying problems in industry that can be successfully treated with the Semantic Web
- Identifying the knowledge components and processing mechanisms that Semantic Web applications will need
- Showing value of ontology-based applications in key business areas.

This deliverable is the result of task 2 of this work package, the "system requirements analysis". This task analyses the typical information systems of organizations, their current work practices and the system requirements which arise from those practices. It aims at forming a growing body of knowledge about typical business problems which will prospectively benefit from the application of Semantic Web technologies.

KnowledgeWeb, through its partners and industry board, provides a set of business cases on how Semantic Web technologies have or could have (hypothetically) solved concrete business issues which are relevant in respect to strategic industries. This set is complemented by a review of where Semantic Web technologies are going in the future, and how these can fit emerging business issues.

At this stage we have collected a set of 16 use cases. This set will be extended in the next phase of the project as the industry board is finalized and KnowledgeWeb partners continue to actively collect use cases. From this we provide a basic overview of the benefits of Semantic Web technologies to industry as well as the key research problems that need to be solved for industry.

The collected use cases span 9 industry sectors. They are intended to be indicative of a broad potential usage of Semantic Web technologies in solving different business problems. We analyzed EU and OECD statistics for industrial economic growth and investment in research and development and identified 4 of the 20 industry sectors used in Knowledge Web as key sectors for the promotion of Semantic Web technologies. Selecting 4 of our use cases that come from these sectors, we aim to use these use cases in particular in Industry Area dissemination activities while focusing our further use case collection on business problems occurring in these industrial sectors. We will continue to seek to maximize the potential impact of our dissemination activities in migrating semantic technologies to EU industry.

Contents

1. Introduction

1.1 Overview of Document

It is well known that a major barrier between industry and academia is that the former speaks in terms of problems and solutions while the latter in terms of technologies and research issues. A business use case is basically a story that relates a business problem to a solution and a solution to a technology, and as a consequence to a research issue. Business use cases are thus good language tools to facilitate understanding between an organisation and a research group since they set a common context and simple natural language for interaction.

The KnowledgeWeb Network of Excellence has as its main and major goal the transference of ontology-based technologies (often referred to when deployed over the Web infrastructure as the Semantic Web) from the field of academia to industry. This is intended to be realised through technological support for industry as a result of research into Semantic Web and ontology tools, interfaces, languages and models, as well as the development of high-class education in the area of the Semantic Web and ontologies.

A significant factor for the uptake of Semantic Web and ontology technologies in the industrial sector is relevance. Cost, both in terms of time and money, must be justified in clear and practical benefits for business operations. While the results of KnowledgeWeb may promote consideration of Semantic Web technologies by industry and offer opportunity for savings and new market opportunities in its deployment, the business decision to deploy a Semantic Web or ontology-based solution will depend firstly on demonstrable feasibility and advantage. Hence it is important for the KnowledgeWeb effort to initially analyse industrial requirements so that activities aimed at research and education for industry have relevance to the actual needs of the industry.

In order to support the communication between academia and industry to their mutual benefit, the KnowledgeWeb Network of Excellence has formed an Industry Board. This board acts as representative of industry as a current or prospective benefactor from Semantic Web technologies. A clear case of establishing mutual benefit for both groups is in the exchange of experience and knowledge, from the industry partners that of actual business operations and requirements, and from academia the current state of the art in Semantic Web and ontology research. The aim of this exchange will be that academia can be better aware of industrial needs and hence direct research and education in the Semantic Web field towards those needs, and that industry can then consequently benefit from research and education efforts tailored to the needs of their business.

The purpose of this document is to report on a collection of real business use cases from the industry partners. These use cases are necessary for the industrial requirements analysis of the KnowledgeWeb partners. This deliverable is the initial presentation of use cases collected in the first phase of the KnowledgeWeb Industry Board. Industry partners

of the Industry Board were contacted for the collection of a representative set of use cases. The process of the use case collection is explained in more detail in the following section and the use cases themselves can be found in Chapter 2 and in the Industry Portal[1] of the KnowledgeWeb website.

Furthermore it is important to the aims of KnowledgeWeb to evaluate where Semantic Web technologies are going to in the future, which research topics are in the academic agenda and how these can fit emerging business issues. In this case KnowledgeWeb can also help to prepare academia and industry for future prospective mutual co-operation. An analysis of this has been prepared by the KnowledgeWeb Network of Excellence and is found in Chapter 3.

This document acts as an informative input to the research and educational activities of KnowledgeWeb. In order to provide an authoritative guideline to those activities the collection of business use cases will be subject to an analysis of the intended solutions for which semantic technologies are being identified and the technological locks preventing those solutions. We also consider the strategic importance of the different industry sectors in order to select a few key use cases to exemplify the potential benefits of Semantic Web technologies. The results of this analysis are given in Chapter 4.

1.2 Selection of Use Cases

Business use cases were collected over the first 12 months of the KnowledgeWeb project by the various KnowledgeWeb partners. The acquisition of use cases was aided by a simple questionnaire which industrial participants were requested to fill out (see Appendix A) and was acquired during face-to-face meetings between representatives of the KnowledgeWeb partners and industry. Through this process a communication process could also be started between many academic institutions and industrial partners.

The Industry area activities of KnowledgeWeb is a continuation of the OntoWeb project's SIG4[2], which formed an industrial network for the stimulation of technological transfer to concrete business cases. Arising out of the experience gained from this network, KnowledgeWeb is aiming for a balance between technology push and industry pull (i.e. the developers and users of Semantic Web based systems) in its Industry Board.

The KnowledgeWeb Industry Board will consist by the end of the first year of the project of up to 50 companies from various different industry sectors such as aerospace, automotive, banking, distribution, energy, health, manufacturing, telecommunication and transportation. This industry board will be consolidated by the 18th month of the KnowledgeWeb project (June 2005). It has been primarily the members of the industry board who have been approached by KnowledgeWeb partners and who have contributed to the business use cases. As industry partners they will benefit directly from their contributions through their direct communication path to the research and education

[1] http://knowledgeweb.semanticweb.org/o2i
[2] http://sig4.ago.fr

partners of the KnowledgeWeb project and early access to project deliverables and research results.

As a result of contacts between KnowledgeWeb partners and Industry Board members, 16 use cases were collected in the first phase (January – December 2004). These use cases are given in Chapter 2. This represents only the beginning of a continued process of use case collection, as new members join the Industry Board and contacts from research institutions and industry are built up and strengthened. Future use cases will be published directly on the KnowledgeWeb Industry Area portal[3], which represents the Web-based communication medium between the KnowledgeWeb Industry Area activities and the Industry Board members. Future trends which we expect to draw the attention of industries, and which benefit potentially from the inclusion of Semantic Web technologies, are given in Chapter 3, indicating possible areas for future business use cases. A simple analysis of the use cases according to the solutions that are being sought and the problems encountered forms an initial basis for identifying:

- What concrete business problems semantic technologies are being applied to solve
- What technological issues exist that are preventing semantic technologies from realizing their potential to solve these business problems

Also, based on the comparative weight of the business sectors (in terms of their economic importance and level of R&D expenditure) we conclude the document by identifying key sectors and use cases which we suggest the Industry Area should focus upon in order to achieve its aim to disseminate Semantic Web technologies within industry.

[3] http://knowledgeweb.semanticweb.org/o2i

2.1 Business Case: Recruitment

1 Overview

Challenge
Facilitate efficiently filling open job vacancies with qualified suitable candidates
Solution
Matching between job offers and job seekers
Why a Semantic solution
Semantics support richer matching based on expressed relationships between characteristics of jobs and candidates
Key Business Benefits
Employee recruitment is increasingly being carried out online. In Germany, for example, over 50% of recruitment is expected to result from an online job posting. Finding the best suited candidate in the fastest time leads to cost cutting and resource sparing with regard to recruitment and can potentially bring firm into contact with a higher level of quality of candidates.
Business Partners
Recruitment agencies

Keys components

Existing Software
Web-based User Interface
Database systems
Research and Development
Wrapper
Semantic query
Matchmaker
Metadata crawler
Technology locks
Query engines
Ontology matching
Storage and retrieval systems

The recruitment of employees is an important practice for any business. While other channels remain available and are utilised by businesses, e.g. newspaper advertisements, trade fairs, human resource advisors or internal recommendations, the Internet has evolved into a primary channel for recruitment. In the near future over 50% of recruitments made in Germany are expected to be the result of an online job posting [1].

The use of Information Technology in the recruitment process has been advantageous for businesses, both in terms of cost cutting and the efficiency of finding a suitable candidate for the post. In this use case we uncover open issues in the recruitment process that raise new system requirements and propose the Semantic Web as a technological solution.

This use case has been considered in the framework of the "KnowledgeNets" project[4], funded by the German Ministry of Research BMBF.A ontology for recruitment has been developed within the project and a prototype system using this ontology is being implemented in order to evaluate the use of Semantic Web technologies.

[4] http://www.inf.fu-berlin.de/inst/ag-nbi/research/wissensnetze/

2 Current Practices and Technologies

2.1 Typical business practices

IT systems are now taking on the role of mediating between firms and jobseekers. They provide a repository for the publishing and discovery of job postings and applicants with the advantages of online data stored in a computer system (e.g. access from any Internet-connected device, search functionality, online application procedure).

Two scenarios can be considered in this use case, one from the point of view of the jobseeker and the other from the point of view of the employer. In both, the goal is to fill a job vacancy with the best qualified candidate as quickly and efficiently as possible.

(1) A jobseeker creates an electronic profile of herself, which gives not only the common contact information but also a record of her qualifications, work experience, skills, interests and other potentially relevant details such as foreign languages and possession of a driving license. She seeks in the database of vacancies for jobs to apply for. This search is guided by the profile she has given so that vacancies for jobs which best match her profile are highlighted. She can manually review the vacancy and if she is interested in the position, request that her application data (drawn from the profile) is automatically forwarded to the firm.

(2) A firm has an open position which it wishes to fill as quickly as possible, and wishes to find the best suited candidate for that position. It creates an electronic description of the position, highlighting the qualifications, work experience, skills, interests and other relevant details wished of an ideal candidate. It publishes this description to a database of vacancies, and requests a list of available candidates (i.e. registered jobseekers in that system) who best suit the position. An e-mail is sent to the selected candidates inviting them to look at the position and apply for it. Applications to the position can be pre-filtered according to applicants' relevance, with only the best candidates being forwarded to a manual (e.g. job interview) phase.

2.2 System requirements Analysis

From the scenarios described in this use case, we can derive the following system requirements:

In the publication of the jobseekers' electronic profile or the firms' job description,

- that published content is unambiguous and able to be understood by a seeker,
- that published content is validated for inconsistencies,
- that published content can be identified as being well suited to an open position or an applicant.

In the searching of the recruitment system for vacancies or applicants,

- that search results are relevant,
- that search results can be filtered,
- that search terminology is consistent,
- that search results can be tailored to the seeker.

Finally, in the case that the job application is carried out online,

- that a job application can be prepared from the published data,
- that a job application can be validated before being sent,
- that job applications can be filtered by matching applicants' data to the vacancy.

Unlike existing systems, we conclude that these requirements can be best met by the use of semantic technologies for the processing, validation and matching of applicant profiles and job descriptions.

2.3 Review of the current systems

Existing IT systems for recruitment tend to be in the form of either job portals set up by state job centres (e.g. the German Federal Employment Office or the Swedish National Labour Market Administration) or those which are privately run and are financed by publication fees (e.g. Monster or Jobpilot). Additionally job postings are often made public on organisations' own Web sites and in other, relevant, locations (e.g. mailing lists and other community portals in the domain of the job). The different organisations (as "suppliers") and jobseekers (as "consumers") are fragmented in the online recruitment market (Figure 1.1). This means both firms and jobseekers must duplicate efforts across the recruitment market if they are to maximize their visibility to relevant applicants and vacancies.

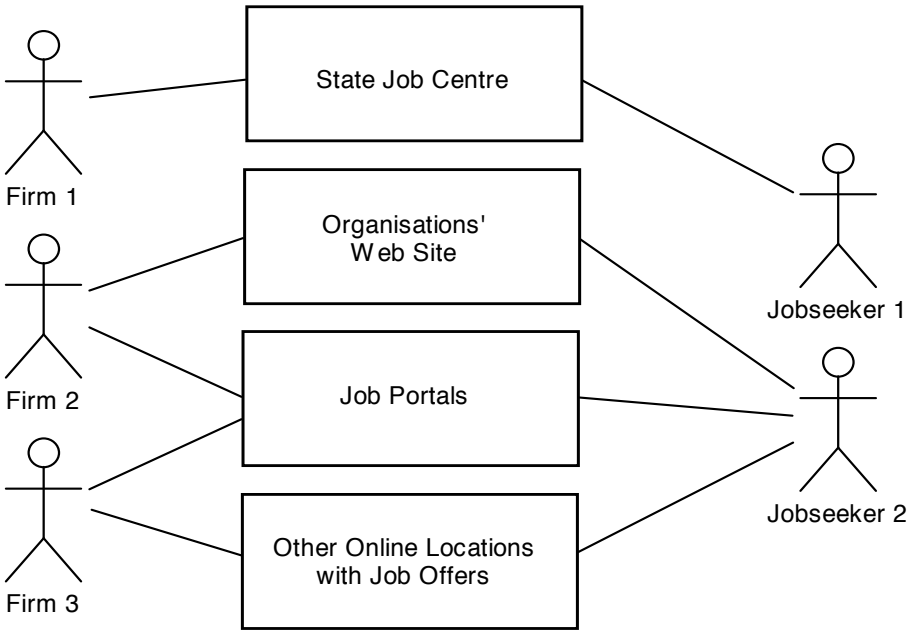

Figure 1.1 – Fragmentation of online job market

Current recruitment systems such as Jobpilot (http://www.jobpilot.de, see Figure 1.2) provide facilities to publish and search vacancies, as well as post applicant CVs which can be searched by employers. Their navigational structure is based on occupational domains and search on specifiable conditions (e.g. job location, domain, …)

Figure 1.2 – Job Pilot website

The German Federal Employment Office (http://www.arbeitsagentur.de, see Figure 1.3) aims to promote greater visibility in the German job market by integrating data from multiple sources into a central recruitment platform. To support data uniformity and integration, all data in the system is represented using the HR-BA-XML standard, which is a XML-based vocabulary for the human resources domain.

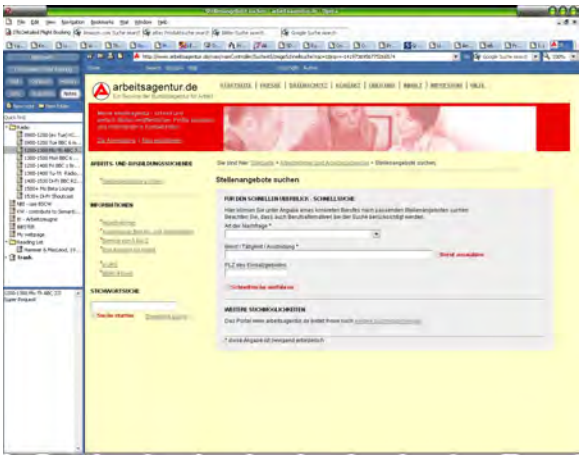

Figure 1.3 – German Federal Employment Office website

References

[1] Institut für Wirtschaftsinformatik der Johann Wolfgang Goethe-Universität Frankfurt am Main, Monster Deutschland & TMP Worldwide: „Recruiting Trends 2004“ (2003), 2. Fachsymposiums für Personalverantwortliche.

2.2 Business Case: Multimedia Content Analysis and Annotation

1 Overview

Challenge
The development and maintenance of large multimedia databases is a difficult task due to the challenges of organising, finding, re-using and distributing multimedia content
Solution
Multimedia is annotated in terms of knowledge extracted from it so that both humans and machines can process large bodies of multimedia more easily
Why a Semantic solution
The Semantic Web offers machine-processable data models for annotating the media and supports semantic search, navigation and reasoning functionalities
Key Business Benefits
Multimedia production and storage is a growing market, particularly among consumers. However, uptake is stalled by a lack of tools to ease media creation and management. Such tools can reduce costs for producers and encourage uptake of media services
Business Partners
Multimedia equipment manufacturers
Network operators
Content providers

Key components

Existing Software
Multimedia ontology management
Security mechanism
Research and Development
Query engine
Media annotation tool
Personalization
Semantic reasoning
Constraint reasoning
Technology locks
Knowledge extraction
Multimedia modelling
Ontology-based reasoning

To address the problems of developing and maintaining large multimedia databases, it is not sufficient to just develop faster hardware or to design more sophisticated algorithms. Rather, a deeper understanding of the information at the semantic level is required. This results in a growing demand for efficient methods for retrieving **semantic information** and **extracting knowledge** from such content, since these are key enabling factors for the management and usability of multimedia content. The goal is to bypass the ineffective and time-consuming process of manual searching and retrieval of multimedia content and use computers to make the content easy to be found and accessible to other parties.

The aceMedia "Integrating knowledge, semantics and content for user-centred intelligent media services[5]" 6th FP Integrated Project focuses on generating value and benefits to end users, content providers, network operators, and multimedia equipment manufacturers, by introducing, developing and implementing a system based on an innovative concept of knowledge assisted, adaptive multimedia content management, addressing user needs. The main technological objectives are to discover and exploit knowledge inherent to the content in order to make content more relevant to the user; to automate annotation at all levels; and to add functionality to ease content creation, transmission, search, access, consumption and re-use. In addition, available user and terminal profiles, the extracted semantic content descriptions and advanced mining methods will be used to provide user and network adaptive transmission and terminal-optimized rendering.

[5] http://www.aceMedia.org

2 Current Practices and Technologies

2.1 Typical business practices

In the following scenario, it is shown how the aceMedia project exploits knowledge inherent to multimedia content and associated textual information to enable new services and user experiences. The example is of a school student who aims to create a retrospective in the form of a multimedia presentation. In order to achieve this goal, she needs the functionalities of multimedia retrieval, annotation and automated organization. Additionally some content involves digital rights and must be paid for. The student's typical use of the system functionality is:

- She wants to create a multimedia presentation including images and videos about her uncle's tennis career. She looks through the textual information associated with local media footage of her uncle to create an overview of his life. From this information, the aceMedia system can select keywords and find further matching media on the Web.
- After paying for and downloading media, the aceMedia system uses the annotations of both the local and purchased media to sort the collected media into themes reflecting different event's in her uncle's tennis career.
- The system's proposed themes can be corrected by the student and the system can use the corrections to propose new themes based on low-level media similarities with the current themes.
- The student performs textual and visual queries to select particular media examples.
- Through built-in concept learning functionality, aceMedia builds up a personal profile of her areas of interest, based on the queries she makes. For instance, facial recognition matching identifies other professional tennis players in the media and the system suggests new themes where her uncle plays against these famous players.
- AceMedia subjects this set to media adaptation, to make it fit on the storage capacity of her aceMedia-enabled mobile terminal, and it is then downloaded onto it.
- On the way to the National News Archives, she transmits media examples of her uncle to the Archive so that automatic searching can start before she arrives. Results scaled down for her terminal are returned and she makes some relevance feedback on them.
- On arrival, she chooses to view the most relevant media results, selects some and purchases them. AceMedia adapts the media for storage on the mobile terminal, and automatically annotates and classifies it. She sends the purchased media to her father who selects some key frames and adds audio and text remarks.
- Finally, she creates a new subset of material based on more specific queries such as "only tennis games mentioned in the press" or "miraculous passes of the ball

over the net". From this subset she can now create her presentation and give it at her school.

2.2 System requirements Analysis

Considering the identified use cases in this scenario, we can derive the system requirements:

Annotating Content

The system shall be able to both automatically produce annotations of media from the basis of low level feature analysis and manually allow annotation of the media by the user with the selection of descriptive terms from controlled vocabularies in order to avoid ambiguity and support system processing. Hence we require

- means to derive concepts from low-level media features, both textual and non-textual
- means to determine the domain of selected media
- means to acquire and use a controlled vocabulary from that domain
- means to suggest to the user appropriate terms from the vocabulary for the annotation

Retrieving Annotations

The system shall be able to retrieve the annotations produced and stored with the media on the basis of queries modelled on the vocabulary of those annotations and present them to the user organized according to the organization of those terms within the controlled vocabulary. Hence we require

- the storage of references between annotations and the media they are annotating
- the extraction of annotations from the content store separate from the media they annotate
- an user interface which supports the user in expressing their query in the vocabulary of the annotations, or a means for the system to interpret a query in respect to that vocabulary
- a presentation component which can represent the relationship of terms from the domain to the user

Retrieving Content

The system shall be able to retrieve media on the basis of a query referring to the annotation of that media, or on the basis of low-level feature analysis, or a mix of both approaches. Media retrieval may be performed on external distributed sources (i.e. on the Web) where annotations may be unavailable or expressed using a different vocabulary. It will also need to take into account the different devices on which retrieved media will be accessed and the need to preserve bandwidth by giving users the option to avoid the retrieval of large media files. Hence we require:

- the internal capability to map between terms in different vocabularies, or from terms to low-level features of media, or from low-level features of media to terms
- automatic low-level feature analysis on media which is found to have no annotation
- the performance of media adaptation to ensure retrieved media can be delivered and displayed to the user on his or her device
- provision of a preview of retrieved media (e.g. thumbnails, brief description, short clips) in order to allow the user to make a pre-selection of desired content before downloading/accessing full media files

Paying for Content

The system shall be able to determine if certain media is copyright protected and hence has some stated restrictions on user access e.g. normally requiring digital payment before it can be downloaded or viewed. It shall also be able to negotiate to have access to the media by meeting the stated access requirements e.g. by making the payment to the copyright owner. In order to avoid any problems, the user shall be kept in the loop and shall have to give permission before the system carries out any operations which affect the user e.g. before payment from the users credit card or accessing restricted content. Hence we require:

- support for digital rights management (DRM), correctly interpreting different access mechanisms and rules possibly expressed in different formats
- the permission of the user to be given before specific access mechanisms are executed e.g. digital payment
- functionality for the support for (secure) communication between different, possibly heterogeneous, digital payment and content access systems

Organizing Content

The system shall be able to use a personal profile of the user, determined (semi-) automatically from the users actions e.g. queries, to carry out a personalized organisation of a set of selected media so that it can be better managed by the user. This management could take place as part of the retrieval functionality (presentation of results) or independently by a request from the user. The content management has the aim of promoting the media most relevant (or of most interest) to the user and to supporting the navigation and location of media through collecting related media into thematic groups. It may also be able to make proposals to the user in terms of new thematic groups of media. Hence we require:

- functionality to determine a user's personal profile through user actions
- means to use that profile to organize media – this means being able to determine relationships between terms in the user profile and terms in the media annotation
- the interpretation of media annotations based on the domain of that media in order to organize the media in a conceptually-related (and hence intuitive to the user) form

- the facility to make proposals to a user of new approaches to organizing media e.g. the creation of thematic sub-folders when there is too much media in a single folder

2.3 Review of the current systems

Current systems that could be reviewed include PC-based media organization tools or the software packaged with PVRs for managing recorded audio and video. However existing systems are not in a position to offer the level of semantic-based functionality described in this scenario.

2.3 Business Case: Peer-to-Peer eScience Portal

Figure 3.1 – The e-science portal S2S

1 Overview

Challenge
EAI (principally search) in an University setting (low budget, high heterogeneity). A high need for automation is present in this setting.
Solution
The RDF-based integration of IT systems including "homegrown" text-based systems. Dynamic decentralised semantic-enabled search.
Why a Semantic solution
The semantic interfaces to information allow the client greater re-use of existing information, are flexible, allowing easy expansion of the system and lower costs of administration for this reason. Automated support for all integrative processes is vital. The bottom line is that searches using semantic technology are more accurate and deliver better information.
Key Business Benefits
A new product in a high-growth sector.
Other business partners
A consultancy to handle teaching issues
A user partner

This business case deals with the provision of a peer-to-peer network for facilitating information exchange across an university setting in the field of science, overcoming the challenges of heterogeneity in enterprise applications and their data.

Universities, as large institutions consisting of various departments, faculties and administrative offices, store their data in a distributed heterogeneous manner yet for typical university-wide activities based on information pooling and sharing they require low cost solutions to a single entry point access to this data.

In this case, we consider the need in the scientific community within and across universities to share their research and publications, making results easily findable to other researchers. We take as a basis the paradigm of P2P (peer-to-peer) networking to enable a decentralised means to share documents across a scalable network of participating nodes, and of EAI (Enterprise Application Integration) to enable

communication between heterogeneous, possibly legacy systems which host these documents.

Keys components

Existing Software
EAI solutions
Data repositories
Peer to peer networking

Research and development
Autonomous systems/self-organisation
Semantic Grid
Semantic Web Services
Rights and security
Ontology mapping

Technology locks
Artificial Intelligence issues
Semantics in P2P networks/EAI

In this context we use metadata as a basis for the efficient and accurate search and indexing of the documents on the network in order to ensure users find quickly and correctly the material they require.

2 Current Practices and Technologies

2.1 Current business practises

At the communication level, Enterprise Application Integration (EAI) solutions are commonly used within enterprises to connect heterogeneous IT systems. Web Services are the most current evolution in EAI, using HTTP and XML as a standardized communication protocol and message structure respectively for integrating systems over the Internet. This integration is point-to-point and the messages exchanged are syntactically heterogeneous.

At the data level, data can be pooled in a repository, though this leads to latency problems (access to data in the central repository while it is being changed locally). This data pooling requires hard-coded data translation (from the source to target format) which grows exponentially as both users and providers expand in terms of formats used by their systems and proves to be unscalable.

A possible solution being examined by the eScience portal S2S[6] (see Figure 3.1) is the use of peer-to-peer networking to store documents in a distributed manner and with a standardized client to ensure consistent access to heterogeneous data on heterogeneous systems. It uses a text- and metadata-based search to allow users to locate the documents they seek on the network. The metadata is generated when documents are placed onto the network in a semi-automatic basis and using this metadata new or changed documents are identified each time the client logs into the network and a local search index is updated.

To provide for integration at the communication level, JXTA is used to provide for platform-independent peer-to-peer communication. At the data level, the heterogeneous documents are considered generically as objects on the network and their information for search functionality made accessible through a consistent metadata representation of each document.

2.2 System requirements Analysis

[6] http://s2s.neofonie.de

The approach of S2S still comes up against problems of P2P and EAI in terms of data and application heterogeneity. It is difficult to bend EAI solutions onto P2P, so that the peer-to-peer contact occurs only on the basis of a specific client that must be downloaded and installed on each machine. Semantics would be a possible solution to enabling the automatic integration of new peers into the network based on metadata about their communication needs. Likewise self-organisation of the network would be facilitated by a standardized means to express information about the network and its peers, such that changes can be propagated along the peers and each can react correctly to that change.

Likewise, messages between peers must be based on the same structure and syntax so that clients can understand one another. Mappings based on different domain vocabularies would allow peers to communicate which seek different languages. Extraction based on such vocabularies would improve the text-based analysis to provide metadata for the documents. Finally, metadata extraction could be further improved and automated so that data in legacy systems can be introduced to the network without great effort and new documents can be indexed effectively. Rights and security need also to be taken into account on the network infrastructure.

2.3 Review of the current systems

S2S exists in a research field which contains other activities in P2P and EAI for example:

- Alvis (http://www.alvis.info)
- EduTella (http://edutella.jxta.org)
- NeuroGrid (http://www.neurogrid.net)
- JXTA (http://www.jxta.org)

Unlike such pure research projects, S2S seeks to apply these innovative ideas practically in the implementation of a new type of application, in which information vital to research is effectively made available and an overview of research efforts in a given area is made possible. There are other commercial implementations of P2P search (e.g. www.grub.org; www.jigle.com) that differ from the approach of S2S in that they exhibit some of the following features:

- Search is based on a very narrow metadata set, extracted e.g. through music interpretation or video
- The user is restricted to generating content with a specific tool
- The data must exist locally to the user or the user can not choose which data will be indexed
- Users can not form into (thematically related) groups

S2S on the other hand exhibits these features, which could be then extended and improved through semantics:

- Search can be based on full text or metadata fields
- User is free to generate his data in whatever format he prefers
- User can select which data to be shared is stored locally and which is found on accessible FTP or Web servers

2.4 Business Case: News Aggregation Service

1 Overview

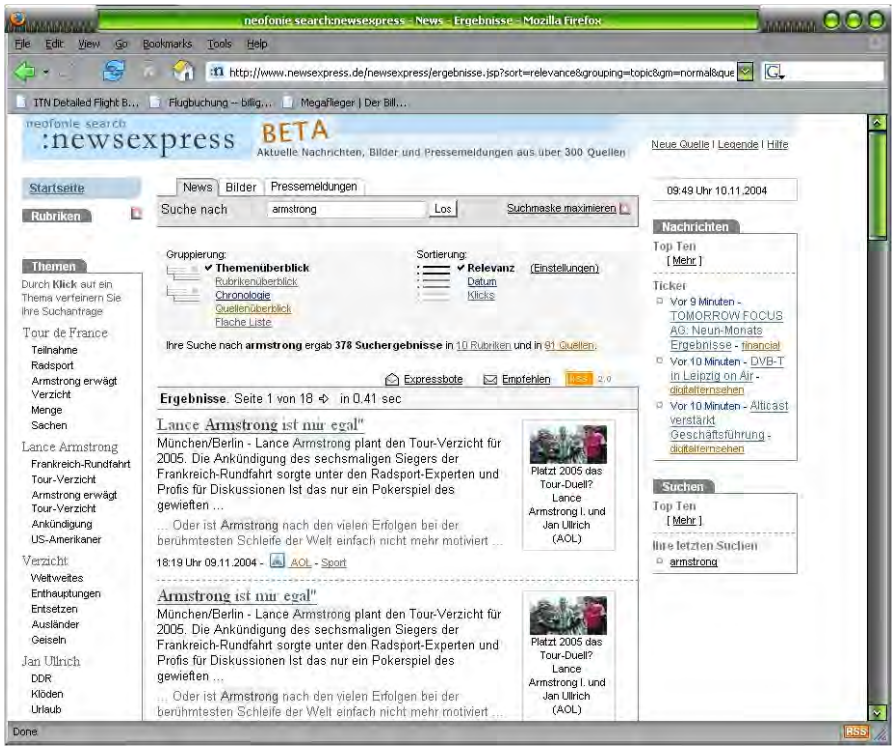

Figure 4.1 – The NewsExpress website

Challenge
To integrate and accurately classify news articles, dynamic web pages, press releases and feeds into a news service
Solution
At the moment a semi-automated integration takes place: (1) Template creation using neofonie search:webstract, (2) Automatic processing of the pages and feeds.
Why a Semantic solution
Semantic technology includes ontology mapping which is becoming increasingly important. As more information providers turn to semantic technology, the returns on automatic solutions increase. However to remain leaders in the market support must still be maintained for legacy systems. Integrating semantic technology in current solutions is therefore a must.
Key Business Benefits
Greater returns on diminishing labour costs. The possibility of concentrating development on new applications instead of on integration.

This business case deals with the provision of an aggregated news service which is able to provide accurate search, thematic clustering, classification of news stories, and e-mail notification of stories of user interest.

There is a business interest in following news in specific categories, including economics, science and IT, or on specific subjects such as particular companies, or developments in the business area. To do this, human resources need to be allocated which is of course a costly business in terms of both time and money. Presently a human will need to manually sort through a large number of different sources to find the news of interest to the business. This involves selecting not just news stories from the main news feeds and media

outlets but also press releases, announcements on websites and other 'alternative' sources.

Keys components

Existing Software
Aggregration system
Natural Language Processing

Research and development
Semantic annotation of documents
Ontology mapping
Semantic extraction of themes
Fault tolerance
Security and trust

Technology locks
Artificial Intelligence issues

While existing developments in providing aggregation services as feeds or Web portals aim to bring information more quickly to the user or users more quickly to the desired information, there remains a significant effort in terms of sorting and selecting the relevant stories. The systems operate by analysing simple metadata tags in the source documents (which are largely not standardized, variable and inexact) and the natural language syntax which is an ambiguous and inexact science.

We believe that the combination of use of semantics in the source documents and a semantically-aware aggregation system can be a major step in reducing the significant effort that still exists in finding the news stories of interest among the information swell existing on the Internet.

2 Current Practices and Technologies

2.1 Current business practises

The news aggregation service http://www.newsexpress.de from neofonie GmbH[7] (see Figure 4.1) is representative of a new class of news aggregation system which can be used to provide a targeted view of the news to users through insertion of its Web-based user interface into client Web sites, whether in an Intranet or Internet environment.

Its current implementation is a hybrid solution, made up of

(1) the manual creation by a source expert of a XSLT template for each news source
(2) the automatic processing of that news source through a thematic clustering algorithm (NLP) and classification with category mappings.

The template is able to convert the semi-structured source data (HTML or XML) into the internal format of the aggregation service and to classify the news story at different granularities within certain themes and (pre-defined) classifications using a mix of element analysis (e.g. value of HTML META tags) and content analysis (NLP on the story text).

As a result heterogeneous sources are integrated and made available through a single interface. Views of the integrated news data is available based on the traditional chronological model (i.e. latest news first) or on the basis of user selection (i.e. number of

[7] http://www.neofonie.com

clicks). Additionally searches can be made based on classification or theme (the clustering of subjects which occur often together in a story).

2.2 System requirements Analysis

The aim of the further development of newsexpress is to increase the accuracy and automatisation of the service. In seeking to achieve these aims, the following issues stand out:

(1) The use of a consistent and clearly understood vocabulary in metadata by the source documents to enable a correct classification
(2) The mapping between different vocabularies and a core system vocabulary to ensure semantic integration over the distributed and hence heterogeneous character of the Web
(3) The extraction of semantics from the source documents in order to remove the need for a fully manual metadata authoring by source document authors (which discourages the creation of metadata in the first place)
(4) Further development in Natural Language Processing and particularly the use of controlled vocabularies in relating NLP results consistently to specific subjects
(5) Addition of domain knowledge to searches to ensure more intelligent and accurate results. This could be seen as an extension of the thematic clustering which is already a step towards modelling relations between concepts within a particular topic.
(6) The need for security from the point of view of source providers (that their content will not be manipulated or inappropriately used) and trust from the point of view of the aggregation service (that news is accurate and not falsified)
(7) The need for fault tolerance in the integration process. A syntactic process breaks when the source syntax is changed, until the template is manually altered.
(8) Duplicate recognition. By modelling the subject of news stories and comparing models, stories which repeat the same issues can be filtered out while stories on the same topic offering a different slant on the story can be used to complement results.
(9) Performance. As semantic technologies are introduced, the tasks of knowledge extraction, concept mapping and semantic-based search must demonstrate sufficient performance capabilities to not bottleneck the system.

It is also recognised that adding semantic data to sources is still not well enough understood by users. The supporting technology doesn't make it simple enough to add metadata (also ontology based) to information. This is a UI issue with the various tools that are currently being used to produce content.

2.3 Review of the current systems

Current systems can be categorised in two broad categories

(1) Feed services, e.g. RSS-based

(2) Web portals, e.g. Google News

Feed services such as those based on the use of RSS or Atom (e.g. Syndic8 http://www.syndic8.com/) offer a flat integration of news from different sources (i.e. all stories exist on the same level). There is a heterogeneity of RSS versions and even their use by authors which has caused integration problems. The content offered by RSS is generally very basic (e.g. title, author, link to full story) and not suitable for any intelligent searching or organising.

Portals such as Google News (http://news.google.de) offer a large body of information that can be processed (i.e. sheer system power) and the organisation of that information through text processing, user clicks, etc.

newsexpress seeks to offer a cleaner solution in focusing on a more elegant semantic approach to news story organisation, being faster to publish breaking news stories and providing stronger means to find and follow news stories of interest (subject classification, thematic clustering, different views on the same stories). It is also implemented to support further development in the use of semantic technologies.

2.5 Business Case: Product lifecycle management

1 Overview

Challenge
There is a high cost associated with the development and maintenance of product catalogues throughout the product lifecycle
Solution
Explicitly model product knowledge according to an agreed, shared terminology for the product domain
Why a Semantic solution
The use of semantics in modeling product knowledge permits the use of open (Semantic Web) standards and the easing of knowledge maintenance over the product lifecycle. With reasoning, different views can be offered upon the same products.
Key Business Benefits
There is a high entry cost associated with creating a product catalogue. Open standards and tools built for semantically-based product lifecycle management will lower this barrier, especially for middle-sized companies.

Keys components

Existing Software
Product configuration system

Research and development
Ontology visualisation
Ontology engineering
Ontology maintenance
Semantic rules
Collaborative work tools

Technology locks
Ontology development

Companies have large product portfolios which they configure and maintain electronically using expensive and complex tools. The use of such product portfolios is to enable companies to manage the entire **product lifecycle**.

Within a product portfolio there is a lot of knowledge and rules about the products and their relationship to one another, yet this is hidden in the data without it being easily possible to extract and use in supporting tools. Hence typical applications upon product catalogue data (supply chain, marketing, research & development, B2C e-commerce) are implemented in a costly hard-coded approach in which catalogue data is imported and manipulated as it suits the particular context of use.

Due to the structure of product catalogue data, in which characteristics and relationships are not explicitly expressed in some sort of standardized terminological way, there is a high cost involved in developing and maintaining this data, in re-interpreting it for a different context of use or for a new catalogue structure, and in working with it in a collaborative environment or sharing it between departments (each of which may have a different understanding of its purpose and meaning).

As a result of this, ontologies are a viable approach to improving the development of product catalogues and their maintenance over the entire product lifecycle in that they offer a consistent terminology for products throughout their lifecycle and the possibility to generate different views for different contexts on the same products.

2 Current Practices and Technologies

2.1 Current business practises

Current product lifecycle management approaches are based on the manual development of a product catalogue in which product knowledge and relationships are defined by the developers at the implementation stage, and later understanding of these definitions is based on the clear understanding of their practices, which requires good documentation. Such approaches are extremely specialised: it requires a (well-paid!) consultant to ensure a properly customised system and approximately 3 months to learn how to model the products in that system.

Product knowledge is stored internally within the system in a proprietary standard, making it difficult to extract and use in different contexts.

2.2 System requirements Analysis

In this field there is a need to reduce entry costs, which is particularly significant for middle-sized businesses which have a large product catalogue but can not afford the expense of the product configuration systems. We see a requirement for ontologies in that:

- The use of a consistent terminology for product knowledge through the entire lifecycle makes the data easier to maintain and better suited to sharing between departments which then can have a shared understanding of the data's meaning
- In combination with tools using this terminology the development process can be made simpler, reducing development time and costs
- By basing the ontology on an open standard like OWL, we move away from proprietary standards, facilitating application development and data exchange
- Ontologies support reasoning, so that product knowledge described using an ontology is also available for being reasoned upon so that e.g. different views on the same product can be offered to different users

The use of ontologies in product lifecycle management makes the following demands:

- The development and availability of good product ontologies, or the provision of best practises and guidelines for the development of such ontologies
- Tools which aid the creation of product ontologies by being able to communicate the correct understanding of the ontology to the developer, feedback on best practises and guidelines, and provide ontology visualisation
- Collaborative work tools supporting ontologies
- Ontology engineering and maintenance practices and tools
- Rules to describe product relationships not expressible in OWL

2.3 Review of the current systems

Product lifecycle management is currently supported by expensive, resource-intensive tools such as the SAP Product Configurator, Matrix10 and MAPICS.

Integrity provides Integral (www.integrity.vc, see Figure 5.1), a SemTalk add on, as a tool to support product configuration in the Microsoft Visio environment. Integral is a tool which allows for a graphical representation of product configurations as Visio diagrams, in which the diagram components are also tied to descriptions using a common terminology. The current tool is SAP-based, however it could be extended to support Semantic Web standards such as OWL in the future.

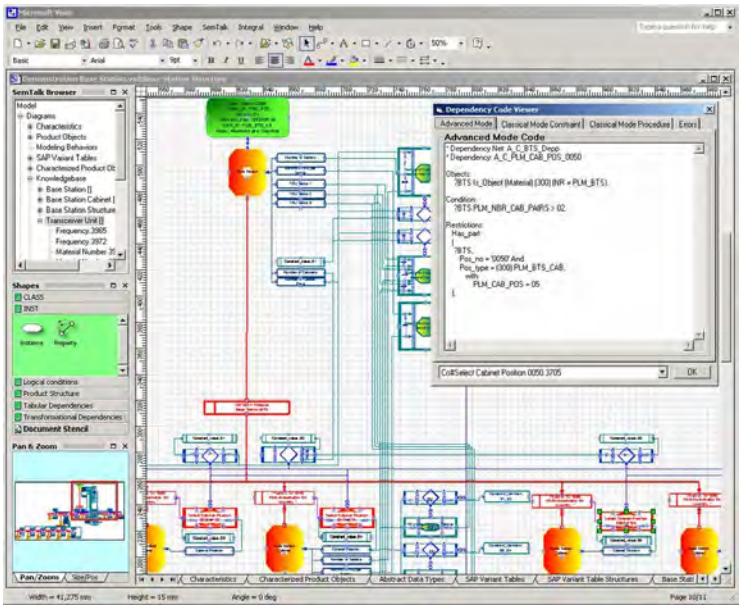

Figure 5.1 – Integral, a product configuration tool

2.6 Business Case: Data warehousing in healthcare

1 Overview

Challenge
A large health insurance company combines its data in a data warehouses to ensure data integration and consistency

Solution
Introduce a common terminology for healthcare data and wrap all legacy data in this terminology

Why a Semantic solution
The provision of a common terminology allows for a data integration and consistency checking solution, in which a computer system can determine the relationship between data items and identify contradictions

Key Business Benefits
There is a cost in time and money involved in a hard-coding of system and data integration, as well as costs to business where integration or consistency is not being handled (i.e. the costs of not having access to data, or having inconsistent data)

Keys components

Existing Software
Data warehousing system

Research and development
Ontology-based integration
Ontology-based proofs
Ontologies for measures
Semantic rules

Technology locks
Data integration
Logical consistency checking
Ontology development/re-use

Companies often administer their data, due to its sheer volume, using data warehousing solutions. A data warehouse is an isolated database which is used across an enterprise to combine data from different data stores and serve all business task-supporting systems with a unified view of the business data. It is characterised by a strict separation of operational and decision-making data and systems.

In this use case, a large health insurance company in Germany uses a Cognos data warehousing solution to administer its data. The business data is being stored in various computer systems, does not share the same data formats and is not integrated with one another. As a result, searching for particular data involves individual searches with the possibility that data is not found because the correct system was not searched. A data warehouse is seen as a solution to this.

Another problem is that data may be changed or updated in one system which may have an affect on data in another system, yet through the lack of any explicit associations between the data, this new inconsistency may not be detected.

As a result, this company has chosen to implement a data warehouse system with a data integration and consistency checking solution. In proposing such a solution, the Semantic Web can be considered.

2 Current Practices and Technologies

2.1 Current business practises

A typical data warehouse architecture consists of a data pool, an archiving system, and a metadata store. The metadata store contains figures, dimensions, reports, cubes and rules. To model a data warehouse system, a modelling tool must take a decentralised approach,

offer good visualisation capabilities, be intranet/Internet based, be simple to use, support namespaces and consistency.

Semtation GmbH offers the SemTalk (www.semtalk.com) tool which is an add-on to Microsoft Visio. It offers a decentralised modelling on the basis of a central model, an object-oriented approach, and support for Semantic Web technologies.

An important aspect of the data warehouse modelling is the description of rules. The types of rules that can be described are data reduction, data format simplification, the combination and separation of attribute values, calculations, data correction and consolidation. Dimensions and operational figures can be defined centrally, including rules governing their use. Through the description of such rules in the metadata of the data warehousing system, data integration and consistency checking can be included.

2.2 System requirements Analysis

At present the rules in the data warehousing system model are not very user-friendly, i.e. they use a very unreadable syntax that will not likely be understood by others apart from the rule developer. The introduction of ontologies is a basis for describing the rules in an open standard using a common, shared terminology (in this case, for the healthcare company and its peer companies). This helps make rules understandable, re-usable, and maintainable and ties them to a formal semantic model to support reasoning and consistency. It also forms the basis for more complex search queries and more accurate search results e.g. “How many leg fractures were there in Zehlendorf last year?” will only return a correct result if the system understands all diagnoses which match ‘leg fracture’ and can correctly identify which occurred in the district of Zehlendorf. Open issues in considering this approach (represented in Figure 6.1) are:

- The description of rules based on an ontology
- The definition of measures in ontologies

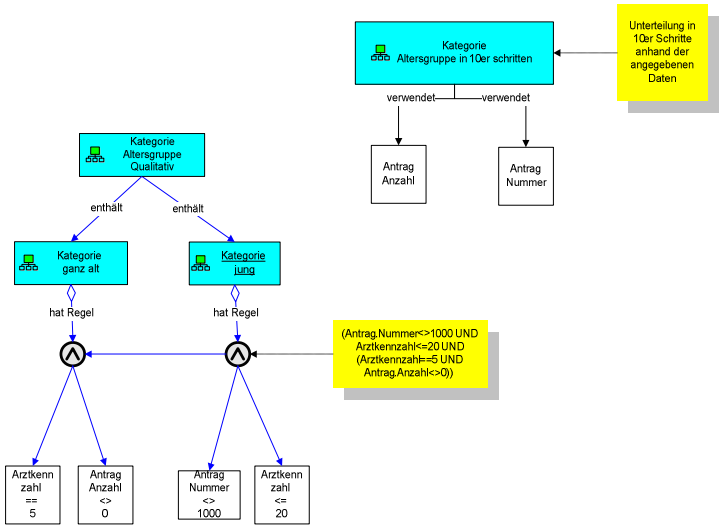

Figure 6.1 – Measures and rules in the data warehousing system (in German)

2.3 Review of the current systems

There are many data warehouse systems on the market, e.g. SAP Business Information Warehouse, SAS 9.1 and IBM DB2 Business Intelligence.

There are also commercial tools for metadata management in data warehouse systems. This includes Ardent MetaStage, IBM DataGuide, Microsoft Repository, Sybase WCC and Viasoft Rochade.

Current data warehouse environments exhibit no or insufficient support for consistent and comprehensive metadata management [1]. Metadata models are proprietary and typically in a warehouse a number of independent and heterogeneous repositories will co-exist.

In fact, metadata solutions are a key need of data warehousing [2], above all those which offer business process modelling and the integration of the heterogeneity from the operational source to the user interface.

References

1. Do, Hong Hai; Rahm, Erhard. "On Metadata Interoperability in Data Warehouses", Technical Report 1-2000, Institute for Computer Science, University of Leipzig, 2000.
2. Do, Hong Hai; Stöhr, Thomas; Rahm, Erhard; Müller, Robert; Dern, Gernot. „Evaluierung von Data Warehouse-Werkzeugen", Proc. Data Warehousing (DW) 2000, Friedrichshafen, Nov. 2000

2.7 Business Case: B to C marketplace for personalized tourism offerings

1 Overview

Challenge
Offer on line personalized tourism packages
One stop purchasing shop of the tourism package
(Train, plane, B&B or hotels, river excursions)
Solution
Dynamic exploitation of content, service providers and personalized data.
Commercial partnership setting
Why a Semantic solution
Dynamic exploitation of c/s providers makes full use of semantic web technology (brokering, mediations, discovery, composition, orchestration and invocation)
Key Business Benefits
Regional tourism offers today very heterogeneous resources that cannot be easily exploited and packaged. In France, the tourism market was evaluated 32 Billion euros whose river tourism represents a turnover greater than 250 Million euros.
Business Partners
Tourism content providers
Tour operators
Regional tourism councils

Keys components

Existing Software
Domain ontology
Geographical Information System
Web Services (1st generation)
e-Business solutions
Research and Development
Semantic data integration
Natural Language Processing
Semantic exploitation of content
Secured payment services
Technology locks
Semantic Web-Services
Semantic aggregation of content
Trusted services

Online sales and information on tourist offers have become prototypes of B2C (Business to Customer) processes on the Net and make up an important part of its commercial activity.

This is for several reasons:

- The consultation of offers can be spread out over time, with no pressure, unlike in specialized travel agents,
- It is possible to consult rival sites directly on line,
- It offers centralized reservation and payment services, just like an agency.

However:

- The customer does not benefit from the expertise of a professional, who can guide him through the ins and outs of the offers and the opportunities to grab hold of. His visibility of offers and services depends on his skills in using a search engine and browsing Internet sites,
- The offers valued on the existing sites are those of which the integration is valued by a minimum turnover: group offers with progressive prices according to the period, and both "standard" and "faraway" destinations (the Mediterranean, Central America, the Pacific).

The current sites essentially suggest package offers based on some information combinations (destination, period, price…), which form the framework of the promotional "package". The business objective is thus to complete the commercial perimeter of current holiday package offers, with some dynamically packaged solutions to meet the customers' expectations (holidays, weekend, all leisure services).

2 Current Practices and Technologies

2.1 Typical business practices

Weekend in Brittany Use Case (Figure 7.1):

A businessman away on business wishes to make the most of his weekend by visiting the Brittany region. He connects up to his favourite travel portal and enters the following information, speaking normally: "next weekend I'm going to Brittany". The portal suggests offers on hotel rooms, plus enjoyable walking routes and the possibility of refining his geographical, activity and affinity requests if he is not satisfied with the suggestion.

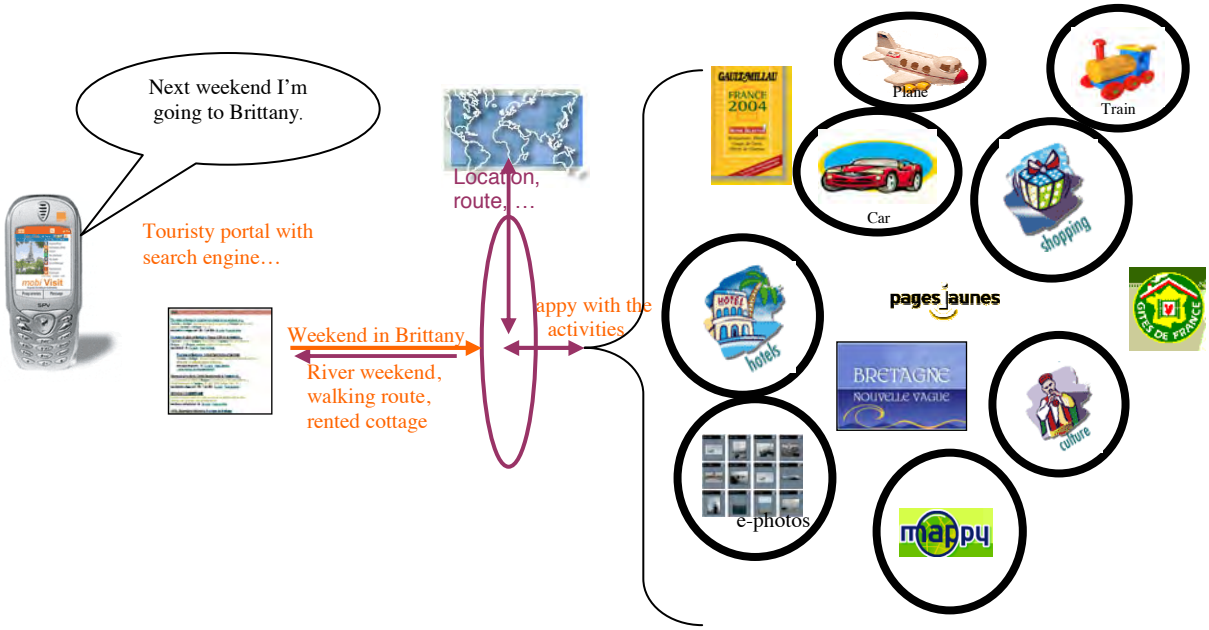

Figure 7.1 – Weekend in Brittany use case

The object of this Use Case is to suggest services for a marketplace for online holiday information and purchasing, which may not be simply based on a promotional push, but gives:

- A sufficiently user-friendly interface and ways of accessing information (by natural phrase, geo-localization…) to entice the customer towards a more interactive tourist search,
- Varied offers (this variety being more limited by the integration costs of the content on the site). In particular, access to regional offers that are currently available on institutional sites, but rarely valued on the marketplace commercial sites,
- A more realistic value with a standardized view and comparison of offers.

The second idea of this scenario is to use public photo albums (aggregation of photos of the same topic coming from different personal albums). There is a double interest:

- First, customers can visualize through photos and aggregated information spots corresponding to their requests.
- Second those public albums are published on a community portal and with time, this portal will get richer and richer.

The regional choice of the scenario is based on the acknowledgement that the offers (hotel rooms, walking or cycle routes, river boat trips...) linked to regional tourism are rarely developed on the Web.
They are essentially non-packaged and non-packageable on the principle of current tourist site offers, as they cause some distinct actors that do not increase the value of their offers to intervene. They would, on the other hand, be recyclable by a new generation of platforms based on the semantic Web.
Practically, our businessman expresses his request speaking normally: "Next weekend I'm going to Brittany and I would like to walk at the seaside". The last precision allows not only to restrict or to be more precise to refine customer tastes but also to underline capabilities of this portal. Indeed, whatever customer tastes, if the portal owns available photos about this topic, it will be able to compose a public album and also to publish it. Afterwards, the customer selects lovely spots that they would like to visit. The portal recovers them and transmits them to a broker in charge of the realization of weekend proposals. This broker has access to different tourism services providers through web services and composes them in order to produce weekend proposals fitting customer choices. We can also extend functionalities of this portal with on-line purchasing and booking services.

2.2 System requirements Analysis

The use cases are centred on the creation of a research and content aggregation Web platform with a sufficiently generic and modular technical solution for integrating additional services.

These Use Cases make the following problems stand out:

- With some existing solutions:

1. Geo-localization

The needs in terms of geo-localization, geo-coding, routes, etc. require the integration of a GIS (Geographical Information System). These days, this type of problematic gives rise to solutions that are already industrialized.

- With some potential semantic type solutions available in few months time:

2. Semantic data integration
3. Natural Language Processing
4. Exploitation of personal data

5. Secured payment services

- With the problems that we do not yet have solutions for:

6. Semantic Web-Services
7. Semantic aggregation of content
8. Trusted services

2.3 Review of the current systems

Dedicated sites for regional tourism become more fashionable thanks to content management capabilities but remain limited in providing personalized tourism package offers. Those sites take into account these pitfalls http://www.maine-anjou-rivieres.com/, http://www.tourismebretagne.com/.

For example research on http://www.tourismebretagne.com for a "Week-end in Brittany" proposes only information pages and no tourism offers (Figure 7.2):

Figure 7.2 – Search results at the Brittany tourism website

The choice of research criteria on B2C marketplace for tourism sites is limited. The responses are oriented to and reflect the commercial partnership of the given tourism marketplace. The offers are pre-packaged based on the prior knowledge of typical customers (see Figure 7.3).

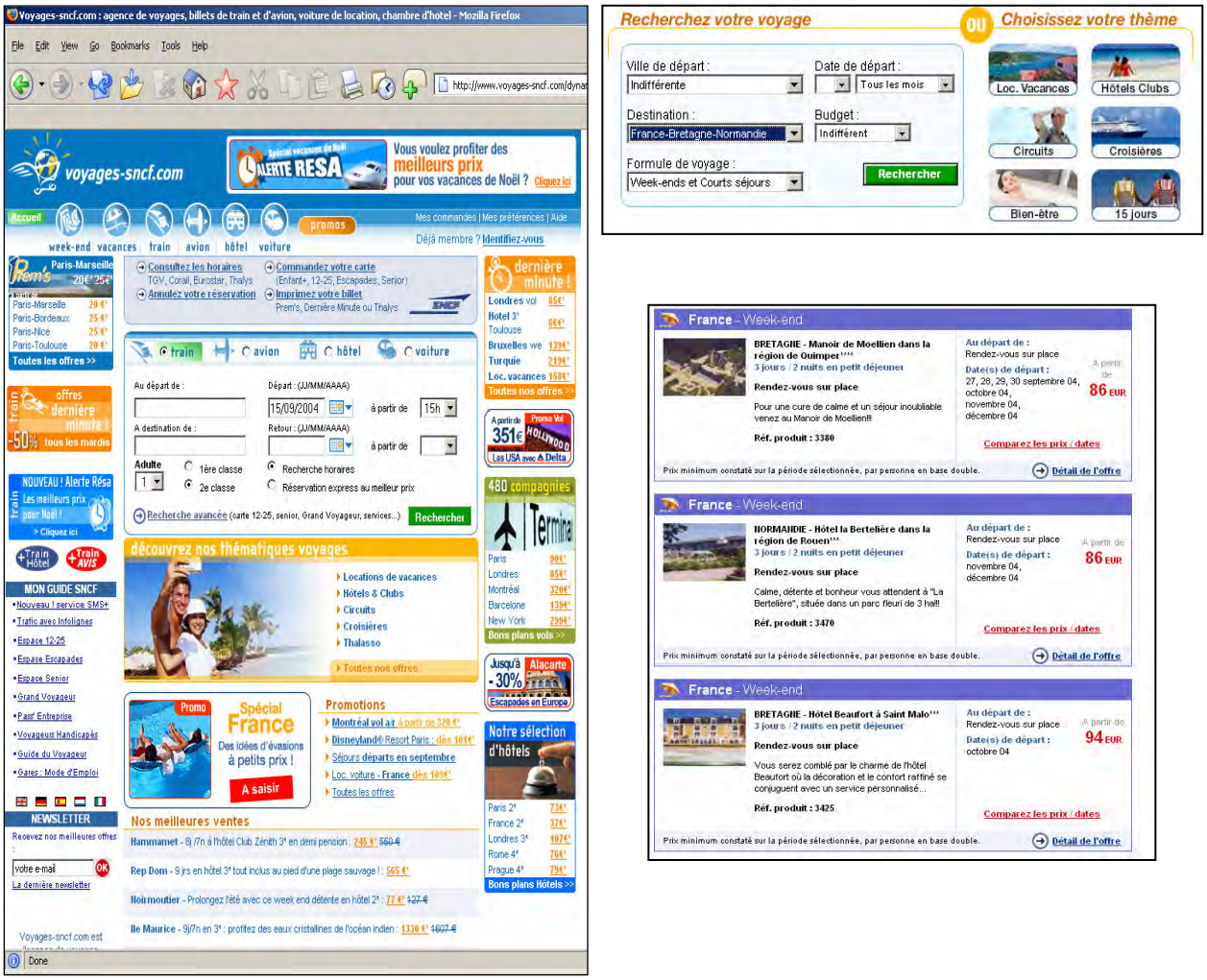

Figure 7.3 – Examples of packaged holidays found currently on websites (clockwise from left: front page of site, search for Brittany holiday, pre-defined packages as result)

2.8 Business Case: E-Photo album automation services on a portal

1 Overview

Challenge
To provide photo album and personal data services via portals.
Solution
Semi-automated aggregation and publication of personal content, digital camera and mobile phones offering today content annotations capabilities (localisation, date..)
Why a Semantic solution
Meta Data extraction and management, content provider mediation are promising areas for semantic portal
Key Business Benefits
B2C enhanced paying services, partnership with the actors of the photo publishing
Business Partners
Orange, Wanandoo, Photo publishing actors

Keys components

Existing Software
GIS, transformation voice/text, Pattern Recognition in digital image
Domain ontology

Research and development
Semantic meta-data fusion
Semantic data integration
Natural Language Processing

Technology locks
Semantic directory (e.g. UDDI)
Dynamic discovery of Content Providers
Dynamic integration of Content Providers in a directory
Automated annotation

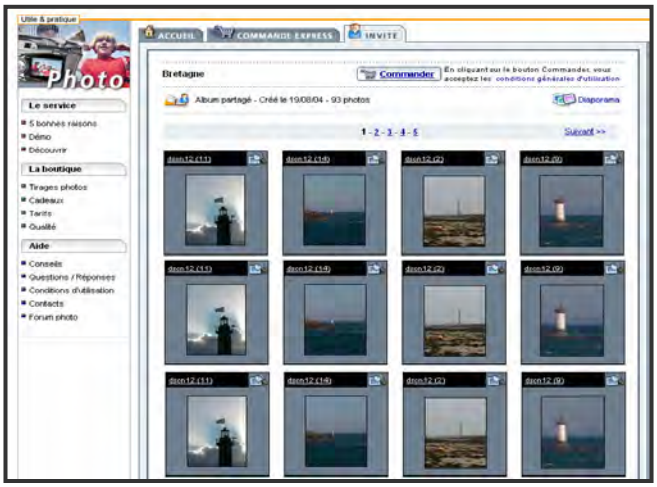

Figure 8.1 – E-photo album service

The management of personal data is changing due to new service portals and the appearance of a new generation of customer centred on multimedia (e.g. ILife solution from Apple). These services are limited today to a classic management of data: create/modify/delete photos, photo albums/musical library. They only exploit the superficial potential, which is involved in aggregation and in content inference (Figure 8.1).

The aim of this use of space is therefore to shed light on this potential, to identify associated problems and to class them into categories: the problems, for which the package solution exists or successful research work and those, which necessitate an investigation by R&D (laboratories, specialist companies, etc…).

Example of the application of new services dedicated to the management of photo albums via a portal:

- Mass manipulation (a digital camera can contain up to 5000 digitalized photographs) and semi-atomisation of the transferral of contents and their organisation,

- Semantic aggregation of contents (photos with other photos, photos with relevant content, etc.)
- Dynamic access to suppliers of external contents via the portal,
- Process of organisation and enhancement of contents (e.g. transformation voice/text, Recognition of form, Geographical location, etc.)

These services could equally be applied to other content types, for example the availability directory of a music library or all other content, since outside of specific treatment of a target category of the multimedia content it is intrinsically applied to more general problems:

- Due to the explosion of production of content, their exploitation needs very efficient, personalized, intuitive and automated tools and services,
- Rapid access and structure of the published contents on the web via inference solutions.

2 Current Practices and Technologies

2.1 Typical business practices

A tourist with a digital camera, under the spell of Brittany, takes photos of landscapes, of his/her friends, of his family, of all what he wants…
Back home, he loads his photos on his personal computer and stores them inside directories. Afterwards, directories constitute very often the only mean of accessing photos. Generally if the names attributed to directories seem to be temporarily suggestive, this feeling vanishes quickly with time when you are seeking in your past photos.

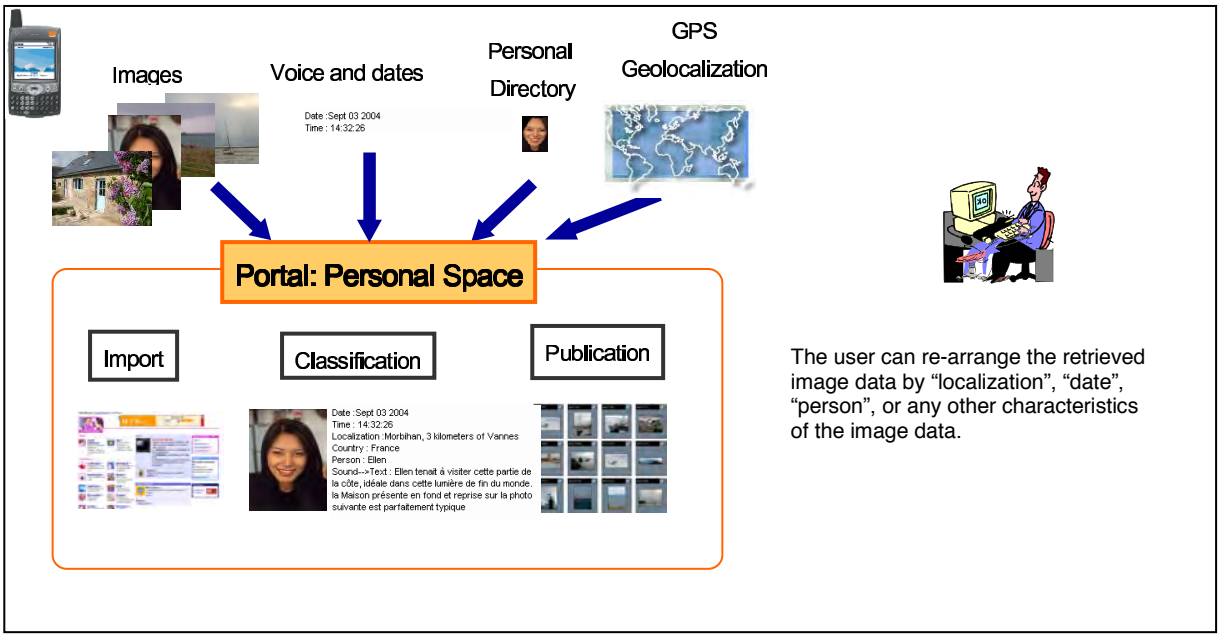

Figure 8.2 – Proposed semantic-enhanced photo album

The tourist loads his photos on the portal and more accurately on his personal storage space (Figure 8.2). A Personal storage space is dedicated to one customer and can contain any personal data that this customer wants to entrust to the portal and especially we can imagine that our tourist entrusted his personal contacts (like in mailboxes) to the portal. A contact is made up of a first name, last name, and a photo. This is the relevant information relative to a contact that the portal will need later as part of this scenario.
Anyway, the portal recovers photos and associated meta-data (Date and time, localization, vocal message describing photos). The portal first applies treatments to the provided meta-data:

- It turns a physical localization (x, y coordinates) into a geographical one (a town…) by using a distant geographical information system.
- It turns vocal messages into text by applying natural language processing. Afterwards, the portal indexes the text.
- Next, the portal makes those meta-data complete:

Based on indexed words, the portal is able to retrieve external relevant information coming from content providers in order to enrich photos. For instance, we can imagine that our tourist takes a photo of a lighthouse (Figure 8.3) and enters the following information: "I took a photo of the lighthouse of the Cordouan". Afterwards, the portal indexes words "lighthouse" and "Cordouan". Next, the portal identifies the appropriate content providers and through a mediation services, retrieves the relevant information relative to this lighthouse.

Figure 8.3 – Integration of relevant external data from keywords

If our tourist says in a vocal message that he took a photo of his wife, the portal is able to learn and to store this information by associating indexed words and personal contacts. Otherwise, the portal applies algorithms of pattern recognition based on photos of contacts in order to detect the presence of his wife.

2.2 System requirements Analysis

To make the search of photos easier for customers, we need metadata. It is relatively obvious to define useful meta-data for photos (localization, date and time, persons present

on the photo…) It is less easy to define when and how to associate those meta-data with photos. Concerning the moment, we thought that is preferable to associate meta-data with photos at the same time that the tourist takes photos. Concerning how to associate them, a digital camera is already able to associate automatically meta-data with photos. Indeed, a digital camera determines the date and time of the snapshot. Some digital cameras linked with a GPS can supply the exact localization, but we can also imagine, that tomorrow some cameras will allow our tourist to associate manually meta-data like vocal messages describing the photos he took.

These Use Cases make the following problematic stand out:

- With some existing solutions:

1. Geo-localization
   The needs in terms of geo-localization require the integration of a GIS (Geographical Information System). These days, this type of problem gives rise to solutions that are already industrialized.
2. Transformation voice/data and pattern recognition
   The problems of transformation from voice to data and text indexation obtained in this way such as based on recognition of patterns are resolved and we can use existing solutions for a rapid service deployment.

- With some potential semantic type solutions available in few months time:

3. Semi-automated integration of photos to the portal whatever the structure or syntax of the sources,
4. Aggregation of external and heterogeneous contents,
5. The availability of a user-friendly interface with efficient and personalized research capabilities

- With the following open research topics:

6. The dynamic discovery of new content providers,
7. Dynamic integration of content providers inside a semantic directory (a la UDDI),
8. The fusion of data requires prioritisation amongst the content providers for a choice of equivalent semantic information. Example: The same lighthouse gives information to 3 different content providers. A possible solution that is conceivable for this prioritisation is supported by trusted indications attributed to each content provider.

2.3 Review of the current systems

There exist current solutions like PBase http://www.pbase.com or e.photo http://photos.wanadoo.fr/ but they are limited. They provide only simple loading publication capabilities.

2.9 Business Case: Geosciences Project Memory

1 Overview

Challenge
In a KM context to give semantic access to multiple project documents and data (software, subsurface models) to practitioners
Solution
Automated on the fly semantic annotation of documents produced by the projects
Semantic search engine for ontology based queries
Automated summary capability
Why a Semantic solution
The present search engine technology is not efficient enough on full text search
Key Business Benefits
Avoid duplication of work and ensure quality of work in future projects by access to the best available information from the organizational memory (Capitalization and knowledge sharing)
Business Partners
IFP together with its projects partners (energy industry)

Keys components

Existing Software
Documentum, Verity full text search engine
Company reference taxonomy

Research and development
Automated Semantic annotation
Ontology driven user friendly query
Semantic search engine
Semantic data integration
Natural Language Processing

Technology locks
Large domain ontologies building and maintenance
Automated on the fly annotation
Automated Summary

Figure 9.1 – Knowledge management system

This business case deals with the management of large collection of project documents in the field of geosciences. IFP is involved in several such projects with a number of partners in the energy industry. These projects deliver various kinds of documents, in the form of texts, geological maps, software, subsurface models, data bases, etc. (Figure 9.1)

Our goal is to access the documents produced by these projects in a structured manner, so that a new project can make best use of the results produced by previous projects.

We believe that semantic annotation together with semantic search engines can be of great help in this context. The usage scenario is the following:

- Before or during the project, a domain ontology is developed either from scratch or by re-using existing relevant ontologies.
- During the project, or at the end of the project, the documents produced are annotated with semantic markers from the ontology. This annotation must better take place "on the fly", that is, while producing the document or just after its

finalisation. It must be automated as much as possible (e.g. 95% automated) so that the project contributors do not spend too much time in annotation.

- It should be emphasised that we intend to annotate not only textual documents, but other project documents as well i.e. software (description of input/output and internal processes), models (description of models variables, parameters, geo-localisation, hypothesis, usage mode…), data bases (structure of database, tables, fields…).
- After the project, a new user needing to retrieve information about a specific domain from past project work would query the geosciences projects semantic memory to find relevant documents. The query should be guided by using the domain ontology, navigating, selecting objects of interest, and making the query.

2 Current Practices and Technologies

2.1 Current business practices

Our current practice relies on a document management system, Documentum, and the Verity full text search engine. Our project memories are built by storing the documents in the knowledge capture and sharing system, annotating them with metadata about the document types, authors, project information, confidentiality level and other coarse grain annotations.

Storing documents in the corporate project memory system is an integral part of or Quality Assurance process related to the handling of project information.

There are no ways to retrieve documents in an "intelligent" way, as queries can only match words selected by the user. In addition, interoperation of different corporate memories is not possible or very difficult because of different referencing systems.

2.2 System requirements Analysis

These Use Cases make the following problematic stand out:

1. Building and maintaining large domain ontologies in complex evolving technical domains
2. On the fly annotation of documents of various kinds, using the domain ontologies. This automation needs to be highly automated as users will not want to spend too much time on this subject. Typically, 95% of the documents should be automatically annotated, even leaving 5% of this task to users would be quite challenging for them.
3. User friendly ontology driven query. When preparing a query to the projects memory, the user will navigate among ontology elements. This navigation needs to be user friendly, as there is no chance that an average user will

want to look at the internals of the ontologies. It should be possible to find the relevant concepts and relations in an easy and quick way.

4. Semantic search engine. This is a familiar problem but the semantic search should scale up well with the amount of project data and of possible queries i.e. a system like Google brings meaningful results in a few seconds, this is probably the target for a semantic search engine on large collections of documents.
5. Semantic data integration between different kinds of documents. It should be possible to retrieve a text about a map, together with the map, and present together with the links between them taken into account.
6. Defining quality assurance procedures for semantic annotation and retrieval: if we want to base our QA process on this new framework we need to demonstrate that the properties are at least better than current practice.

2.3 Review of the current systems

There exist current solutions like: Goldfire Innovator http://invention-machine.com, specialised in access to technical solutions by looking for subject-verb-object triplets in patents, or Arisem (Thales) http://www.arisem.com , based on semantic networks.

2.10 Business Case: R&D Support for Coffee

gruppo illy

illycaffé research & technology development

1 Overview

Challenge
In a KM context, workers have personal attitudes and aims at managing documents and other knowledge artefacts, in the way that better suits their needs. KM solutions should enable personal knowledge management and should allow knowledge sharing among workers.
Solution
A semi-automatic system of annotation and concepts representation, which expresses a personal point of view using ontology based tools.
Semantic search engine for ontology based queries
Why a Semantic solution
The current approach in KM is to organize knowledge in a standardized way, through a corporate knowledge system. This approach doesn't take into account personal needs and autonomous representations of knowledge. The traditional search engine technologies are not efficient enough on full text search in personalized and context dependent knowledge bases.
Key Business Benefits
Allow networking and knowledge sharing processes among workers who manage their personal knowledge in autonomous ways. Enable innovation through perspective matching
Business Partners
Technology solution provider together with its projects partners, networked laboratories, and Universities

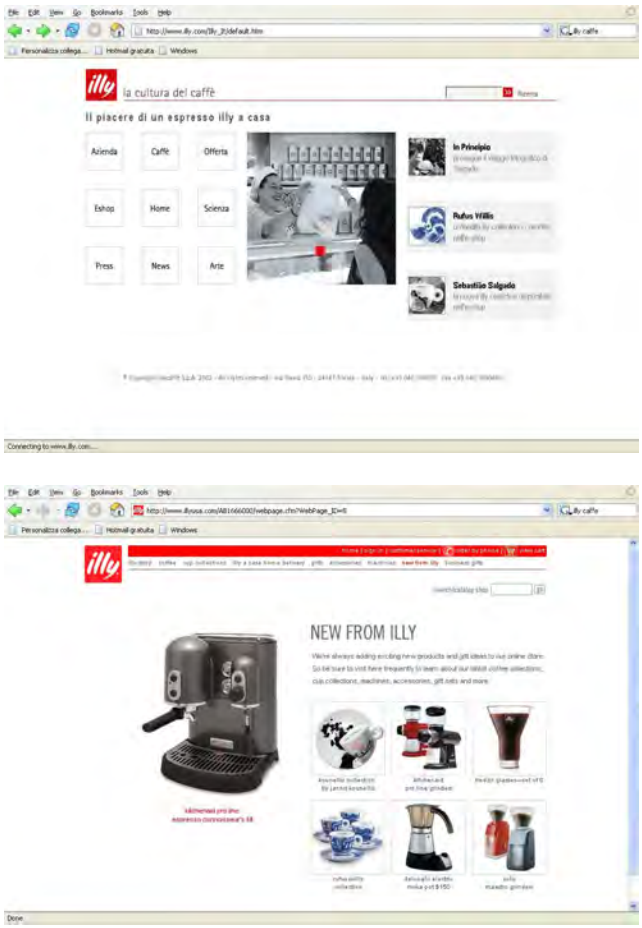

Figure 10.1 – Unstructured corporate knowledge on the Illy website

This business case deals with the management of large collection of documents in Illy Caffè (Figure 10.1), in which high level of quality and innovation in products and services are considered strategic assets. The R&D unit is composed by researchers and networked laboratories such as Aromalab, and Sensorylab. They are specialized in different research areas (botanic, physics, math, etc.) and they actively contribute to Illy's experiments. In particular, a lot of experiments and projects such as Illy Bar Concepts and Easy Serving Espresso are carried out. In this scenario various kind of documents, in the form of texts, designs, biological tests, software and hardware recommendation, etc. are delivered.

Keys components

Existing Software
Company reference taxonomy
Shared directories
Personal Database and archives.

Research and development
Semi-automated ontology/context creation
Semantic search engine
Semantic matching

Technology locks
Corporate domain ontology building and maintenance

The goal of the firm is to enable personal knowledge creation allowing autonomous management of knowledge and coordinating it among workers. The R&D group believes that semantic annotation together with semantic search engines can be of great help in this context.

The usage scenario is to promote the management of personal knowledge through personal representation and to promote peer to peer networks among workers.

From a managerial standpoint, a peer to peer approach to KM poses fundamental challenges on the changing nature of the managerial function. Peer to peer settings work out new roles (for instance the role of broker) which determine new competencies and abilities for a system of networked "knowledge". Managers should avoid personal or group behaviours of competitiveness and detention of knowledge. They have to promote knowledge sharing and coordination among single experts and small groups who handle all the core knowledge of the firm. These processes can be facilitated by both the introduction of new organizational roles (knowledge managers and brokers) and the creation of a culture (using wage incentives, group bonus, etc.) that allows people to identify themselves within the company as part of a whole, and to share knowledge for a common real gain. This could be facilitated by the introduction of peer to peer tools.

2 Current Practices and Technologies

2.1 Current business practices

Workers manage their personal knowledge using their personal systems of artefacts as notes, books, archives, document management systems, etc. Even if people use the same kind of artefacts, people appropriate and manage them in an autonomous way.

Although each document management system is used in a different way, workers support knowledge sharing by their personal interactions. This interaction is enabled by a good attitude of communicating, asking and chatting. Any time that people need to share knowledge in a group or in a team, they create a common and shared system of directories in which all the documents of the group are stored, they use phone calls and phone conferences, e-mail, e-mail attachments, etc. Moreover they use an Intranet in which workers can upload any document referring to their project. This Intranet (composed by a content management system organized according to a system of categories) is not much used within the organization because workers cannot modify its structure. They consider it oppressive and irrelevant, as a bureaucratic system which people have to deal with. Nowadays, Illy's practice of knowledge retrieval relies on

personal capacity to remember the time in which old projects have been developed and on the ability to retrieve related documents. There is no way to retrieve documents in an "intelligent" mode, and only common search engines are used. These search engines (e.g. vivisimo http://www.vivisimo.com, google http://www.google.com, etc.) support only lexical matching processes, search only through keywords and eventually cluster documents.

2.2 System requirements Analysis

From a technological point of view: **(i)** Each member (peer) should create and manage her/his personal system of documents and knowledge. The peer's document management system is based on syntactical indexes (indexes of labels), and contexts (a system of concepts managed by a context editor and a context browser) which allow the user to organize and represent her/his knowledge through an explicit complex system of categories, taxonomies or contexts. **(ii)** Each peer may join a community, which is considered as a unique subject in the system. The community's knowledge relies on the system of documents that each single peer shares according to a common representation and organization of knowledge. In other words peers participate and negotiate a common representation of knowledge (a taxonomy, an ontology, etc.) and make available their documents according to this representation. **(iii)** Peers are allowed to manage their document according to their personal and local systems of knowledge representation (as personal ontologies and context), and exchange documents through semantic tool, permitting coordination and negotiation processes. **(iv)** Searching processes are carried out by seekers. Users can search documents according to three different types of queries: lexical, conceptual, and semantic. **(v)** The provider peers match the queries and give back results, and, when it is possible, give back suggestions on who else can have the same kind of information.

This Use Case makes the following problematic stand out:

1. Personalized knowledge representation should be managed by local ontologies, taxonomies, etc. In computer science several languages and tools exist to help final users and system developers automatically, semi-automatically or manually create good and effective ontologies. Therefore a strong challenge unveils in building and maintaining a constellation of domain ontologies in a complex evolving environment, using very simple tools and methods.
2. These tools and methods should guarantee a high level of participation. The navigation and the management of personal perspective need to be user friendly, as there is no chance that an average user will want to look at the internals of the ontologies. It derives that an effective human and computer interface should be created.
3. Semantic search engine should scale up effectively with the amount of knowledge perspective developed by an increasing number of peers and communities.
4. Negotiation processes should be developed with the aim at allowing peers to participate and share knowledge within a community. Peers participate and negotiate meanings and artefacts (documents) and create common knowledge representation. This allows them to constitute a creative knowledge base which sustains common themes and interests, and allows them to achieve new goals and innovative perspectives.

2.3 Review of the current systems

The existing current solutions should be integrated to properly support the R&D office and the knowledge workers involved in the projects.

KEEx: it is a Peer-to-Peer (personal) document sharing system, which provide users with document management and retrieval functionalities based on lexical and semantic algorithms. Such algorithms allow users to perform both full text keyword search, and conceptual one. A matching algorithm based on natural language and mathematics processing functionalities (http://www.dthink.biz) is the fundamental core process of the tool.

2.11 Business Case: Co-ordination of real estate management

1 Overview

Challenge
In a geographically distributed and complex organization the coordination of the buying and managing processes of real estate is a relevant and strategic asset.

Solution
Matching information among descriptions of designs, maintenance techniques, estate descriptions etc. All of this information is emerging from different communities of workers: Order Managing, Research and Technology, Testing, Maintenance and Plant Engineering groups.

Why a Semantic solution
Semantics support a richer matching among concurrent descriptions of rolling stocks and railway materials. Semantics allow an effective design, maintenance, and transmission of train estate and materials during their whole life cycles.

Key Business Benefits
Coordinate 5.000 employees, according to their skills, over ten workshops (called Business Units, each is specialized on a different kind of rolling stock) which are geographically distributed in Italy and in charge of periodical maintenance of Trenitalia's rolling stock and materials.

Business Partners
Trenitalia and all the complex organizations that need to share knowledge among autonomous units, aimed at estate managing and buying.

Figure 11.1 – Trenitalia website

Semantics problems, dealing with communication and collaboration between different groups, are strategic in medium and big organizations where group heterogeneity is expected (such as network of organizations and networked organization). They can occur also in the case of companies merging, or in the case of growing companies, or when different points of view are considered strong values that should be managed.

Keys components

Existing Software

1. GAD (Gestione Archivio Disegni): the repository of the official projects' designs. It is accessed by some communities of the firm that are allowed to write, check, validate and review new Maintenance Technical Procedures (MTP) referring to train's components.

2. SICUESE: it is an Access database, dedicated to the storage of documents dealing with maintenance activities, including MTPs.

3. RFI web site: RFI (Rete Ferroviaria Italiana) is the Trenitalia's web site in which the railway infrastructure is managed

4. RSMS: it is a SAP system where all the operations performed on each rolling stock item are recorded.

Research and Development

Peer to peer document sharing and reviewing systems; semantic information retrieval tools; Social networking that supports the exchange of processes between communities.

Technology locks

Query engines; Semantic matching;
System for social creation and reviewing of documents, MTP, etc.

The project is focused on providing a complete knowledge management solution to Trenitalia (Figure 11.1), in particular to the engineering department. The KM problems we pointed out in this Trenitalia department are strongly related to semantic heterogeneity, since the department is made up of several groups, each with its own point of view, own language and own way of doing things. This heterogeneity often forms a barrier to performing those organizational tasks where collaboration and communication is required. For that reason, the solution we designed includes also a tool able to face semantic issues, automatically matching different knowledge schema (ontologies), and then supporting the collaborative work while respecting the single group's semantics.
Buying and managing real estate is a relevant and strategic process for several organizations and companies. Briefly, this process is made up of the following phases:

- Definition of estate features (documentation for the setting up of a call for bids)
- Negotiation phase (the selection of the right provider according to the required features)
- Order managing and checking (the process of monitoring the provider's work)
- Running (estate's managing and maintenance)
- Leaving (estate's dismissing and, eventually, selling)

Such a process is even more complex for public companies, due to European constraints and rules. The Business case described is one of the core processes in the UTMR department (Unità Tecnologie Materiale Rotabile) of Trenitalia s.p.a.

2 Current Practices and Technologies

2.1 Typical business practices

The headquarters of UTMR are based in Florence, and its core business deals are working on new rolling stock designing project and on the acquisition of new railway equipment.

The department consists of 5.000 employees, divided according to their skills, over ten workshops, called Business Units. Each one of them specializes in a different kind of rolling stock. The Business Units are geographically distributed in Italy and in charge of periodical maintenance of Trenitalia's rolling stock and equipment. Moreover UTMR includes different specialized communities. Some of them are focused on function specializations: Gestione Commesse GC (Order Managing), Tecnica e Ricerca TR (Research and Technology), Sperimentazione S (Testing), Ingegneria degli Impianti e della Manutenzione IIM (Maintenance and Plant Engineering). Others are focused on projects crosscutting the whole organization: Community by component (workers spontaneously aggregate in order to share a common professional interest); Community by order (workers involved in a specific order, such as ETR50 or Pendolino share knowledge according specific projects).

Communities are involved in several UTMR activities, for instance the process of writing a Technical Procedure of Maintenance (Norma Tecnica di Manutenzione). A Technical Procedure of Maintenance (MTP) is a document that workshop people access in order to find all the information about the maintenance activities required for specific trains or railway equipment. The writing process of this document involves several UTMR workers, belonging to different communities. Their job consists in searching, merging and validating different sources (both internal and external to UTMR) writing the specific MTP. Currently, no technological system supports this practices (except for using shared folders in file system), that leads to a quite complex and difficult situation for the workers involved. In Figure 11.2, the processes of MTP construction are depicted:

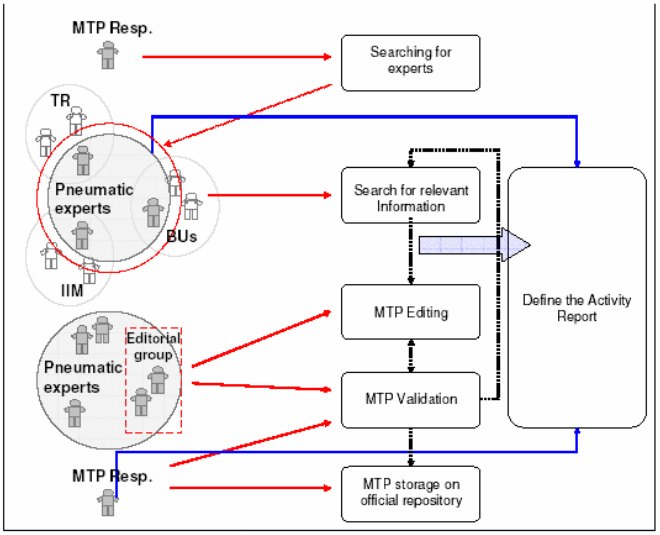

Figure 11.2 – Process of MTP construction

2.2 System requirements Analysis

From the scenarios described in this use case, we can derive the following system requirements:

Searching for experts and relevant information about MTP

The system should provide retrieval tools able to find expert workers and documents dealing with the specific theme treated by the MTP.
This means both tools supporting a full-text keyword search, but also tools able to match different categorization structures, since we can expect several different UTMR communities to use different ways of interpreting and referring to documents about the same subject. Obviously this introduces critical issues of semantics which can be solved by using semantic-based algorithms. Results could be then organized (and also ranked) according to the used search parameters.
System requirements: Expert maps, Temporary communities creation, Search by keyword, and by semantics, Search on different sources (distributed search), Document indexing, Document sharing, Communication among groups and communities members (ex. chat), Documents upload on official repositories, Results ranking

Publishing the new MTP in the official company repository

The final phase of the MTP writing process occurs when the responsible person stores the MTP in the company official repository.
The system should allow the responsible person to upload the final document in a categorization structure which has been previously created by the system administrator. The MTP is now available to whoever needs it.
System requirements: Information storing by multiple views; Information searching by keyword and by attributes; Navigate repository's categorization structures

2.3 Review of the current systems

Unlike existing systems (GAD; SICUESE; Rete Ferroviaria Italiana web site, and RSMS) we conclude that these requirements can be best met by the use of semantic technologies. Moreover, there are no technological systems supporting UTMR in performing all of the described business practises. Therefore, a particular technological architecture has been proposed. This architecture includes three different applications which are integrated to properly support all process phases:

1. **KEEx**: it is a Peer-to-Peer document sharing system, which provides users with document management and retrieval functionalities based on lexical and semantic algorithms. Such algorithms allow users to perform both full text keyword search, and conceptual search thanks to a matching algorithm based on natural language and mathematics processing functionalities (http://www.dthink.biz).
2. **Verity K2 Enterprise**: it is a quite complex suite of content and information retrieval tools (like, for example taxonomy builder, social network tools, document and content retrieval tools, expert location tools) based on a proprietary search technology. In the proposed architecture it has been included as just the lexical search engine (www.verity.com).
3. **FileNet P8**: it is the FileNet's Enterprise Content management (ECM) platform (www.filenet.com).

2.12 Business Case: Hospital Information System

1 Overview

Challenge
Data in a healthcare information system is dispersed and heterogeneous in a setting where speed of access and common presentation are important
Solution
Integration and subsequent mediation of medical databases at the semantic level
Why a Semantic solution
Solutions focused on integrating databases tend to ignore the underlying meaning of the data and its structure so that an intelligent consolidation and presentation of data is not possible
Key Business Benefits
Speed of access to the required data is vitally important in a healthcare setting as well as a common view on the data by different users

Keys components

Existing Software
Middleware for database integration

Research and development
Data wrapper (RDBMS -> ontology)
Query mediation
Semantic matching
Ontology engineering

Technology locks
Database to ontology mapping algorithm
Intuitive graphical mapping interface
Ontology model mapping

This use case deals with the issue of database integration in the domain of healthcare. An expanded description of the use case can be found at [1].

Health care organisations such as hospitals may have several dispersed data sources containing interrelated information. For example, there may be a central repository which contains administrative information of all patients registered at the hospital. Additionally, each division holds additional (or even the same) information about the diagnoses and treatment of the patients that they have dealt with. As information stored about a patient in one division may be relevant to a (para-)medical professional seeking information from another division, an unified search is highly desired.

Two further challenges in this case are that:
(1) data may be stored in very different ways, from totally unstructured text (e.g., notes written by a physician) to highly structured repositories (e.g., medical relational databases),
(2) access must be achieved within an efficient time frame.

2 Current Practices and Technologies

2.1 Typical business practices

The typical approach is based on the syntactic coupling of relational databases based on their structure, and the exchange of information through messaging. This is however only a partial solution as no use is made of semantics, i.e. the meaning of the information.

Language and Computing N.V. (L&C) and VUB STARLab collaborated to integrate dispersed relational databases using ontologies as the central conceptual knowledge schema. In this case the ontology is one from the medical domain containing over 2 million medical concepts interrelated by over 5.3 million relations. In order to enable this,

L&C wants to extend its ontology management server LinKFactory® (see review of current systems) with a component to integrate information from external relational databases. VUB STARLab developed a suitable language able to define a coupling between a relational database and an ontology in a natural manner and provided a methodology for creating such a coupling.

2.2 System requirements Analysis

From the analysis and work done so far the following requirements stand out:

- Greater accuracy in the matching of database to ontology concepts in order to reduce the amount of human involvement in the initial mapping generation
- Better visualisation approaches for data and ontology structures and their mappings
- A tighter semantic coupling between the database and the ontology
- Methodology for interpreting an ontology and expressing its relationship to a non-ontological data structure
- Resolution of differences in ontological models
- Handling of very large and complex ontologies and application databases

2.3 Review of the current systems

L&C's ontology management system LinKFactory®[8] is a framework designed for building, managing, and maintaining large and complex ontologies. It adopts a typical 3-tier architecture:

- LinKFactory® Workbench: a client application to manage the LinKBase ontology (client tier)
- LinKFactory®: the server interface, receiving and answering user requests, holding the business logic and requesting data (application-server tier)
- The data layer accessing the underlying database. This database contains all information the LinKFactory® server needs to operate (user information, ontology content and maintenance information) (data-server tier)

As a preliminary step, L&C has already developed the MaDBoKS system (Mapping Databases onto Knowledge Systems) as an extension to LinKFactory® (see also Figure 12.1, screenshots reproduced from [1]):

- A module for the LinKFactory® Workbench providing GUIs and features in order to conveniently and semi-automatically map (relational) databases onto the LinKBase® ontology (or any other ontology contained within the data-server tier).
- A mediation layer extending the data access layer of the ontology server and provides the necessary mediator agents and database wrappers.

[8] http://www.landcglobal.com/pages/linkfactory.php

- Modules for the LinKFactory® Workbench enabling the graphical browsing/querying of the semantically mapped databases at ontology level.

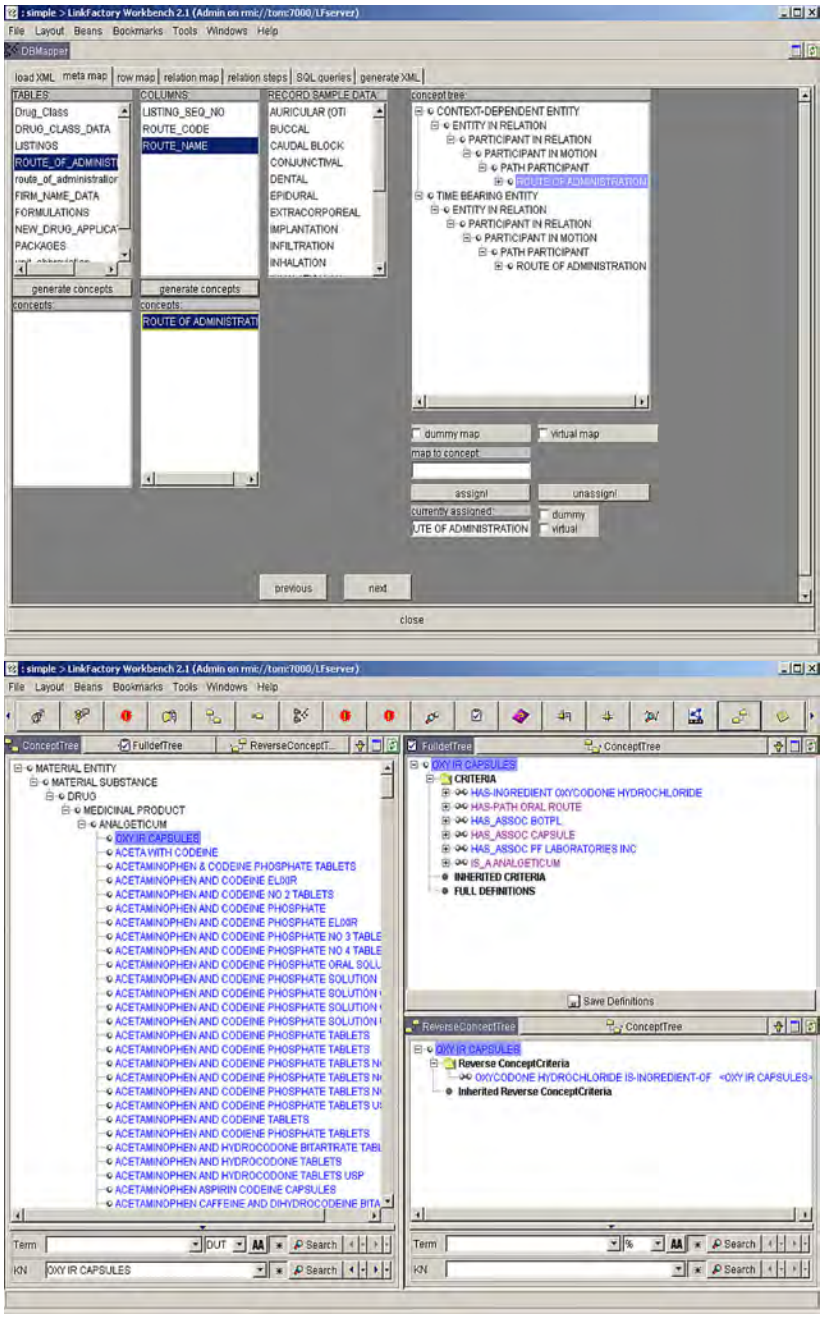

Figure 12.1 – Screenshots of LinKFactory® Workbench

References

[1] Deray, T. and Verheyden, P. (2003) Towards a Semantic Integration of Medical Relational Databases by Using Ontologies: a Case Study. In *Meersman, R., Tari, Z. et al. (eds.), On the Move to Meaningful Internet Systems 2003: OTM 2003 Workshops*, LNCS 2889, Springer-Verlag, pp. 137-150.

2.13 Business Case: Agent-based System for an Insurance Company

1 Overview

Challenge
The traditional way of handling insurance claims involves different experts in different departments using different approaches, making the process very costly.
Solution
Chain integration for insurance claim handling can make the entire process more economical
Why a Semantic solution
Ontologies are required to formalize the vocabularies of companies from different domains such as insurance and car repair.
Key Business Benefits
Reduce errors in the interactions between companies due to inconsistency in their databases. Ease the addition of more participating companies in the interaction chain.
Business Partners
Insurance agencies
Damage repair companies

Keys components

Existing Software
Local databases
Local data input, processing and distribution
Research and Development
Agent-based systems
Ontology development
Ontology mapping
Technology locks
Gap between agent technology and ontology technology
Semantic interoperability issues including ontological translations
Working with (for industry fairly new and therefore not industry-strength supported) languages such as RDF and OWL

For one of the largest insurance companies in the Netherlands, Acklin BV has built an agent-based system to handle all communication related to car damages with a damage repair company.

Insurance claims handling involves a costly process where each aspect of a claim is examined by different experts in different departments using different approaches. One of the processes within the claim handling process is repairing car damages.

The agent-based system requires an extension with ontology technology as a means to optimize interactions between two companies that have different processes and employ technologies of various degrees of sophistication. In designing and implementing the system, some requirements could be identified that raise the need for such a semantic solution.

2 Current Practices and Technologies

2.1 Typical business practices

The insurance market hugely relies on a traditional way of claims handling. Every aspect of a claim will often be dealt with by a different specialized person working in a different department of the company. The input, processing and distribution of data are treated by each part of the organization in their own traditional way, making the process being very costly. Nowadays, the insurance market is looking more and more for ways to economize the process of claims handling. Because the process of claims handling involves many different parties such as the victim(s), witnesses, surveyors, lawyers, insurance companies, middlemen and doctors there is a growing need for chain integration.

Acklin solved the business case using agent technology. The system is composed of two agents. One at the insurance company (the insurance-agent) and one at the damage repair company (the repair-agent). The insurance-agent sends data about insurance policies and car information to the repair-agent. The repair-agent on its return sends information related to performed jobs and invoices to the insurance-agent. When a car needs to be repaired, the repair-agent has to ask for permission at the insurance-agent. The agents communicate with each other using a peer-to-peer technique.

2.2 System requirements Analysis

When designing and implementing the system several (unexpected) issues delayed the progress of the project. The most interesting issue was the use of different vocabularies by the two companies, especially the use of the word "cause". The repair company used the word "cause" to indicate the reason of car damage e.g. (freely translated) "car has been hit by another car", "car ran into an obstacle" or "car broke down". The insurance company used the word "cause" to indicate why they had to pay for the repair of a damage, e.g. (freely translated) "repair a part of the car" or "replace a part of the car". The approach taken to make a translation between the different vocabularies was to write down three tables: one with the vocabulary of the insurance company, one with the vocabulary of the damage repair company and one translation table. The insurance company complained about faults in the table of the repair company and vice versa. After several meetings, the companies started to understand that their view on the world was not unique. From there, they also learned the concept of ontology and the technique of ontology translation.

However, there seems to be a gap between agent technology and ontology technology. Engineering intelligent agents typically involves dealing with distributed environments, including complex message passing in interactions and autonomy. Although, some pieces of ontology technology enable simple client-server architecture, most ontology technology seems to be engineered around a centralistic concept having central databases and one major ontology server. In this case, new pieces of technology have to be built in order to have agents working with ontologies, distributed databases and concept translations.

Connecting two or more systems to each other will always raise interoperability issues. Although agreeing on using the same transport protocol (e.g. FTP, HTTP) and content language (e.g. fixed width, XML) is hard, a lot of companies are not aware of semantic interoperability issues. Furthermore, companies are not (yet) aware of techniques and methods available for handing semantic interoperability issues, such as translations. Companies do understand (sometimes after a while) that they use different vocabularies in their processes and databases. The technical languages used in these companies are database schemas in SQL and XML. Therefore, most ontology techniques and methods are unavailable for these companies, because ontology research tends to focus on working with (for industry fairly new and therefore not industry-strength supported) languages

such as RDF and OWL. Of course these languages are needed, however the methods to reason with knowledge stored in these language should also be needed in the less expressive languages, such as SQL and proprietary XML, because today these are the languages most used.

2.3 Review of the current systems

AcklinQ is a commercial product aimed at supporting agent-based cross-organization, technology and process information logistics.

For more information see http://www.acklin.nl/products/acklinq/acklinq.pdf

2.14 Business Case: DaimlerChrysler Semantic Web Portal

1. Overview

Challenge
Finding data in a large institution can be difficult and time consuming
Solution
Knowledge management solutions allow to classify data and improve precision and recall.
Why a Semantic solution
This classification can be made according to a shared conceptualization of the domain, i.e. an ontology. Additional expressability through rules allows for more powerful search and retrieval based on reasoning.
Key Business Benefits
Employees find relevant or required data more quickly, which leads in total to significant time and cost savings for the enterprise.

Keys components

Existing Software
Underlying storage e.g. relational database

Research and development
Ontology development
Semantic portal technology
Access rights
Rules engine

Technology locks
Ontology creation tools supporting good ontology modelling
Integration of semantic technologies within the business IT infrastructure
Rule creation and testing

Daimler Chrysler is a large enterprise with departments spread worldwide and over 360 000 employees in total. A significant volume of data is being generated daily through business activities and this data may not only be relevant to the originating department at the time of generation but also to other sectors of the enterprise at any time in the future.

Knowledge management is a requisite for employees to locate the data that is relevant to their search. However currently much time is spent in locating that data from the large body of corporate data that is available. There is a need for higher precision and better recall so that employees find the required data more quickly.

As a result, an internal Semantic Portal is being implemented which organizes the data according to a shared conceptualisation of the domain, i.e. an ontology, of the different aspects of the enterprise (e.g. employees, products, processes).

2. Current Practices and Technologies

2.1 Typical Business Practices

The DCVD Semantic Portal is implemented on the company intranet as a research prototype. Later it will be available to employees of Daimler Chrysler to insert, navigate and search for corporate data.

It contains a navigation structure for manual location of data based on the enterprise ontology as well as a semantic search facility. Navigation is along the taxonomic structure of the ontology (sub/superclasses) and for a selected class it lists the known instances. Individuals can be selected and examined in terms of their properties and relations to other individuals. The search returns both direct matches as well as 'related' matches. The semantic matching and ranking is based on a direct match to classes or individuals and then a ranking based on subsumption (i.e. other relevant search results are returned based on their relationship to a superclass or subclass of the queried concept). In addition, rules have been modelled to describe knowledge that is not expressible at the ontological level. These rules form the basis for additional inferences so that other relevant matches are determined for the query.

2.2 System requirements Analysis

The portal implementation involved the building of an enterprise ontology from scratch. There is a need for better ontologies in the sense of maintaining consistency and following modelling guidelines, which in turn requires ontology development tools which are able to guide the ontology developer through the process. Correspondingly, the effort required to produce ontologies needs to be reduced.

Another requirement identified from the portal development is the need to integrate semantic technologies transparently into the business IT infrastructure. Rather than requiring a new set of technologies, existing systems should be extended as transparently as possible to use semantic technologies with the end user benefiting from the introduction of the use of semantics without being required to learn new tools or processes.

A key need is for access rights. This must be integrated into the business infrastructure in both semantic and non-semantic systems. Semantics could be very useful in this case in that the appropriate access of a user to some data could be inferred from the user's position and the data's status e.g. a project manager would reasonably be given read and write access to project data while other users who are involved in similar or related projects may be granted read access. Such 'intentional rights' would also need to take into account new access rights when a user changes their position within the enterprise (transferred to another department, taken off a project, etc).

The determination of relevant matches needs to be extended to take into account the complexities of the real world domain. For example, as matches on superclasses can be included within the bounds of relevant matches, a query on who is responsible for a particular problem may return not only the individual directly responsible but also his boss or the bosses' boss – yet it is clear that actually the big boss is unlikely to be interested in this small problem!

These complexities can be further modelled using rules. However there is a need for an easier means to model such rules for non-logicians and better tools to test and debug rule bases (collections of rules applied to some knowledge).

2.3 Review of the current systems

The current corporate Intranet has some structure for the navigation to data and a traditional search engine. In its current form data is linked to at most one category meaning that the user's location of the desired data is heavily dependant upon choosing the right category and other (relevant) data which has been placed into other categories is not found. Correspondingly, a search for data which transcends a simple flat categorization is limited to the traditional keyword search with its ambiguities and inconsistencies (e.g. when a keyword is input but spelt incorrectly). Finding e.g. 'all data related to Mercedes-Benz' is restricted to those documents which specifically mention the car make, and excludes others which are related but do not directly mention the car make.

2.15 Business Use Case: Specialized Web Portals for Businesses

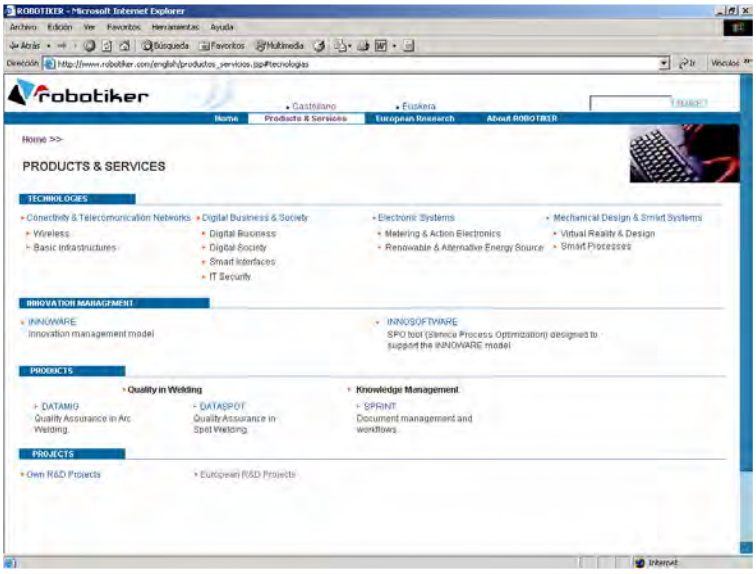

Figure 15.1 – The Robotiker Web Portal

1. Overview

Challenge
Robotiker, as a Technology Centre, develops hardware and software in different areas. We only have one general Website in which you have to navigate through different sections to get to the desired area. Individual Websites need to be created so that clients can access their desired information directly.
Solution
To generate specialized Web portals automatically using the information about the company stored in different storage media..
Why a Semantic solution
The use of metadata and ontologies will enable to model the company organization and to relate these concepts to the company information. With additional software development this can be used to generate the Web portals automatically.
Key Business Benefits
We have specialized Web portals for our clients generated automatically and with the update information at every time.
Business Partners
None.

The creation of specialised Web portals is becoming necessary in the enterprise world as well as in the administration area.

This is the case of ROBOTIKER, where we develop hardware and software in different areas (energy, telecommunications, automotive, and so on). We only have one general Web site (Figure 15.1) in which you have to navigate through different sections to get to the information about one area. We find it necessary to create specialised Web sites (one Web about telecommunications, and so on) so that our clients can access to the information they need directly.

Keys components

Existing Software
Commercial Web portal building tools

Research and development
Authoring tool based on ontologies
Web portal generation from metadata, ontologies and data repositories.
Repository access modules.

Technology locks
Ontology development
Knowledge extraction

There already exist tools that allow in some way to create Web portals automatically with pre-selected contents. Those tools generally are complementary to specific content management systems, so that previously you must acquire those systems. Those tools are very expensive, and can only be acquired by big companies or national administrations. This is the case of the most popular content management systems providers as Interwoven, Vignette, BroadVision, Documentun. All of them provide the Web portal generation tool, together with their content management systems. For example, Vignette offers Application Portal and Interwoven TeamPortal.

The SME-s (Small and Medium Enterprises) and regional or local administrations have the same information publishing needs, but the cost of these tools is too high for them.

The proposed system based on semantic technologies would offer the following advantages:

- Efficient look to the world, because of the speed of their presence in Internet.
- Prestige, because of offering updated contents, removing the expired and inappropriate contents.
- Time and resources saving. The enterprise is dedicated to its business, and not in worrying about its presence in Internet.

2. Current Practices and Technologies

2.1 Typical business practices

ROBOTIKER has a general Web site in which you have to navigate through different sections to get to the information about one area. It would take us time and money to create specialised Web portals for our clients and then maintain them up to date.

The ideal situation would be that these specialised portals were created automatically and dynamically extracting the information from the different information sources in Robotiker. The Web administrator should only select the concepts that must appear in each Web portal and its appearance.

2.2 System requirements Analysis

To get to the ideal situation described, firstly an ontology of our organisation should be created, as well as the metadata to associate the concepts of the ontology with the information sources in Robotiker.

Then, some Web components should be designed and developed that enable the automatic generation of Web portals depending on the concepts of the ontology selected.

Another requirement would be to develop several connectors to the different information sources to access to the information to show on the Web.

2.3 Review of the current systems

Nowadays to generate specialised portals from a general Web portal, or to create specialised portals starting from the contents, most of the software providers offer software tools based on the use of their content management system. As an example, the following three content management systems providers offer the following tools to publish the contents on the internet:

- Interwoven: Team Portal.
- Vignette: Application Portal
- Broadvision: One-to-one Portal, One-to-one Commerce, One-to-one Content.

Microsoft offers Microsoft SharePoint Portal Server that allows the users to search, share and publish information on their corporate, departmental and group portals, and the possibility to customise these portals using HTML code modules called Web Parts. This is based in the powerful search machine developed within Microsoft Content Management Server.

Red Hat Enterprise Portal Server is an open source software, but not free software, that allows to join local and remote content in a configurable platform that supports multiple languages for its user interface, and different devices as WAP, XHTML and VoiceXML. The portals can be built and oriented for specific persons, working groups, people with a common interest and for big organisations. There is also some open source software, and with no cost, such as XOOPS and OpenCMS.

The Institute AIFB of the University of Karlsruhe has also published several papers about a framework called SEAL for building semantic portals [1,2].

References

[1] A. Maedche, S. Staab, N. Stojanovic, R. Studer, and Y. Sure: SEmantic portAL - The SEAL approach. In: *Spinning the Semantic Web*. D. Fensel, J. Hendler, H. Lieberman, W. Wahlster (eds.), MIT Press, Cambridge, MA., 2003, pages 317-359

[2] A. Maedche, S. Staab, R. Studer, Y. Sure and R. Volz. SEAL - Tying Up Information Integration and Web Site Management by Ontologies. In: IEEE Computer Society Data Engineering Bulletin, Special issue on "Organizing and Discovering the Semantic Web", Vol. 25, No. 1, pp. 10-17, March 2002.

2.16 Business Case: Integration of Biological Data Repositories

1. Overview

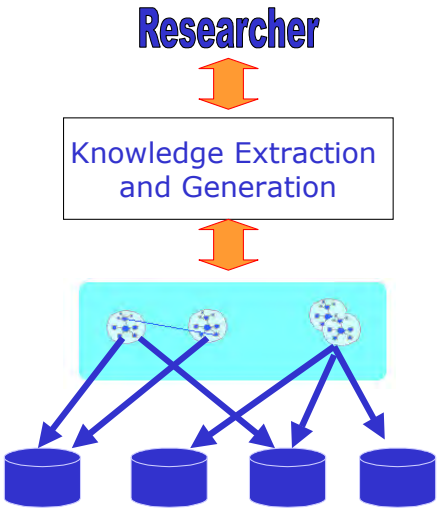

Figure 16.1 – Integration of Biological Data Repositories

Challenge
To provide an unified point of access to different biological data repositories accessible through the Internet, corporate databases, results of experiments, health cards, medical literature sites and so on.
Solution
Application of semantic technologies to solve the inherent features of the biology field: huge quantity of dispersed, distributed and autonomous data with great difficulties to be integrated due to differences in terminology, syntax and semantics.
Why a Semantic solution
Ontologies describe the vocabulary of the data stored at each repository. Annotations describe the data and link it with a corresponding ontology. Ontology merging and mapping techniques allow integration of repositories in a consistent and unified way.
Key Business Benefits
Aid to the researchers in the biological field, providing a unique point of access to biological data. For example, when a researcher wants to compare the results of an experiment with the genome annotation database.
Business Partners
Life science companies.

Currently, a great diversity of biological data exists in repositories: databases accessible through Internet, corporate databases and experiment results among others. Equally there exists a great diversity of ontologies for modelling this data. Therefore the situation that the researchers has to face with is a lot of dispersed data and different disconnected and non-user-friendly tools to access such data, therefore the researches have to confront great difficulties to aggregate all the data to carry out the research tasks in an integrated way (Figure 16.1).

Up to now ontologies in biology were considered as mere guides for data structure, with their only purpose being to access the more adequate documents and articles to the researchers' interests. This new vision will allow for combining and associating existing ontologies in the biological field and an integrated modelling of the biological data sources (genomics, proteomics, metabolomics and systems biology). Once modelled, the

Keys components

Existing Software
No existing software.

Research and development
Study and selection of ontologies already existing in the biological field.
Merging and mapping of ontologies
Annotation using selected ontologies

Technology locks
Knowledge extraction
Ontology based reasoning

annotations, intelligent agents, semantic web agents and the semantic grid will offer a centralised access point to extract and generate knowledge from the biological data repositories.

2. Current Practices and Technologies

2.1 Current business practices

Current practices by biologists seeking data are guided by the available tools and their capabilities, as given in section 2.3.

2.2 System requirements Analysis

Some requirements identified from the use case:

- Generation and extraction of knowledge from biological data by means of ontologies, ontologies existing in the knowledge domain, combining them (ontology merging) and/or associating them (ontology mapping) is to be exploited by means of annotations, intelligent agents, semantic web services and/or semantic grid.
- Using standards for the semantic web providing a unified entry point to different biological data repositories in the most automated way possible.

Ontology Mapping allows the determination of which concepts in some ontology A are the same as in another ontology B. Detecting common concepts is allowing the "jump" between ontologies. The ontology merging could be used by a company that wants to use de facto standard ontologies associating them to specific company ontologies. In this way, proprietary repositories can be "linked" with public ones.

Ontology Merging allows to sum one ontology C with another ontology D to obtain a more complete ontology. Due to the fact that a protein could be implied in cell signaling as in a biological process, summing up two ontologies, one describing cell signaling and other describing biological processes can give us a general overview of a protein function.

Both mapping and merging are shown in Figure 16.2 below.

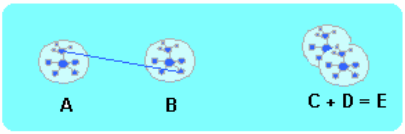

Figure 16.2 – Ontology mapping and merging

Once we have merged or mapped the ontologies in the above task we have to be capable to link these resulting ontologies with public or proprietary data repositories through semantic annotation. Deep annotation is a framework taken into account at this stage (Figure 16.3).

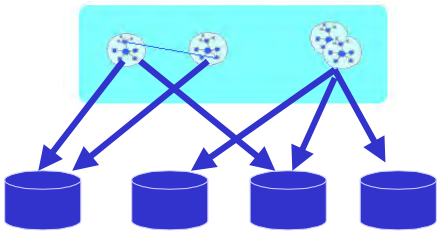

Figure 16.3 – Deep annotation to link data repositories to ontologies

The owners of the databases have the knowledge, so they are the ones that should annotate the databases with the appropriate ontology.

If the biological data repositories are linked to the ontologies by annotation, now we have to offer to the researchers appropriate tools to extract knowledge from these sources. In this task existing semantic and artificial intelligence technologies applicability and requirements detection will be carried out.

2.3 Review of the current systems

Currently there are several Data Mining tools focused to solve the immediate and concrete problems that the researchers in the biological field face day by day.

Entrez is a database search engine that provides a search that combines documents containing nucleotide or protein sequences, 3D structures and their respective references in MedLine; its power resides in the numerous cross references that it offers between the different databases, along with a computerised system for similarities between documents, that allows to provide the documents set most similar to the required one.(http://www.ncbi.nlm.nih.gov/Entrez/).

At the SRS tool there is no limit to the number of databases and applications that can be accessed. It also allows creation and saving of the users' own intuitive views for displaying data which can be made up of as many different databases and fields as wished. The most frequent database queries can be set up to be available every time the user logs into the system. To make results more meaningful, filtering of application results is possible using predefined queries. Work can be published to SRS for all appropriate colleagues to access in a read-only mode.

TAMBIS aims to aid researchers in biological science by providing a single access point for biological information sources round the world. The access point will be a single interface (via the World Wide Web) which acts as a single information source. It will find appropriate sources of information for user queries and phrase the user questions for each

source, returning the results in a consistent manner which will include details of the information source.

With the current tools it is still needed to understand the structure of the different databases, as with SRS, so the inherent problems of terminology, syntax and semantics are still present. It is difficult to know if a table in one database called Organisms is the same table called Species in another database. With TAMBIS it is necessary to define a wrapper service for each database to translate the queries so no automation is provided. To design the wrapper it is necessary to study the database before including it into the system.

3. Future Trends and Business Applicability

In the use cases given in this document we have gained an overview of typical business problems in various industry sectors and been introduced to the applicability of semantic technologies in resolving those problems. However, both the research agenda surrounding the Semantic Web and the business issues arising within industries are in a continual state of development. Hence while there is value in considering both factors in terms of the current state of the art it is also advisable to consider future trends and potential, which is the aim of this chapter.

3.1 Review of Key Research Trends

We list here some key research trends which, although existing currently mostly within the research field, are being promoted to industry as means to solve known business problems. Hence we expect to see some or all of these research trends migrating to industry in the same way as the Semantic Web, and identify the potential role of Semantic Web technologies within those trends.

Social networking – A social network is a map of the relationships between individuals. In the Internet, social network applications have become a popular means of expressing connections between friends, business partners or other individuals. Some social network tools are aimed at business application, such as Ecademy (forming trusted business networks, http://www.ecademy.com) and ReferNet (enhanced business networking through referral, http://www.refernet.net). Social networking is seen as a means to organize enterprise knowledge in terms of the connections between different departments and individuals and to bring disparate parts of the enterprise into contact on mutual areas. Social networking also forms the basis for one of the earliest widespread uptakes of Semantic Web technology, namely the Friend of a Friend (FOAF) vocabulary, which offers a standardized means to model personal information and relationships. In terms of community management, FOAF-based information could be used in many different ways – one introductory article [1] names:

- Augmenting e-mail filtering by prioritizing mails from trusted colleagues
- Providing assistance to new entrants in a community
- Locating people with interests similar to yours

Grid computing – The aim of Grid computing is to harness the untapped resources of computers on a network. At any time across a large network, e.g. a business intranet, some if not all machines connected to the network are not being used to their full potential. The computational resources of these machines is effectively being wasted. The operation of a computer Grid is analogous to that of the electricity grid: users can access power from distributed power generators without needing to be concerned with the source of the energy or its location. The use of Grid computing is expected to support large-scale data-intensive computational activities and enable businesses to efficiently make use of their total computational resources. Some Grid applications are already active in the pharmaceutical, medical and physics fields:

- Molecular modelling for drug design (http://buyya.com/papers/vlab-drug-design.pdf)
- Brain activity analysis (http://buyya.com/papers/neurogrid-ccpe.pdf)
- High energy physics (http://lcg.web.cern.ch/LCG/)

The potential of Grids to business are being examined, with the feeling being that Grid computing can offer "new business value chains and boost innovation capabilities by supporting the creation of collaborative working environments across multidisciplinary scientific, engineering and industrial communities"[9]. Fundamentally, some see the Grid as forming "virtual organizations" where dynamic collections of individuals and institutions share resources in a flexible, secure and coordinated manner [2].
Outstanding research issues are being tacked, for example in the EU project Next Generation Grid Computing, which outlines a research plan for 2005-10. One area of interest in the research field is the extension of the current Grid with well-defined meaning in order to support ease of use and seamless automation, the "Semantic Grid"[10].

Semantic Web Services – Businesses spend around 40% of their overall IT budget on system integration, according to recent estimates[11]. Enterprise Application Integration (EAI) solutions are vital to business operations, and in recent years Web Services have been promoted across industry as a simpler, more flexible, open, cheaper and loosely coupled means to achieve system integration in comparison with earlier approaches. While Web Services uptake by businesses will continue, researchers are already working on an extension of Web Services technologies designed to deal with their shortcomings and bring closer the possibility of full automation of business processes. This extension incorporates Semantic Web technologies, and hence this extension to Web Service technologies has been named 'Semantic Web Services'. By providing well-defined descriptions of Web Services and their functionality, it is expected to support the (semi-) automated discovery, invocation, co-ordination and composition of distributed and heterogeneous Web Services. An overview of the state of the art in Semantic Web Services is available to Knowledge Web members in an internal deliverable [3]. The DIP project (a member of the SDK cluster along with Knowledge Web) has produced a document on potential application areas for Semantic Web Services [4], identifying as the high potential areas business process management, content syndication, enterprise collaboration, search/data mining and social networks.

Multi-modal interfaces – Access to computer systems – the data they contain and the functions they can perform – is often limited by the means of access: the traditional method of typing and reading from a screen can be inefficient, time consuming and inappropriate in many cases. Multimodal interfaces is an emerging research area encompassing all forms of natural interaction and communication between computers and human users with the aim that individuals can perform actions with computers efficiently, conveniently and using everyday skills. Research areas include speech recognition, natural language understanding, image processing, computer vision and pattern

[9] IST workshop on Grid Economy and Business Models
http://www.cordis.lu/ist/grids/grid_economy_and_business_models.htm
[10] See http://www.semanticgrid.org
[11] Quoted from http://www-128.ibm.com/developerworks/ibm/library/i-holo/

recognition. Businesses encounter problems with integrating computer usage into different business practises where the use of computational resources is beneficial but the means of that use is not (e.g. customer support or mobile access to enterprise knowledge[12]). They could also benefit from the incorporation of such interfaces into their products or services (e.g. into automobiles[13]).
Some research issues in multimodal interfaces are the description of users and devices (which determine appropriate interaction forms), natural language processing from oral media and identifying objects and events from visual media. These fields all include the examination of semantic technologies as an enabler of multi-modal systems.

3.2 Review of Emerging Business Issues

The fast pace of change in society, also to a great extent caused by the rapid development in information technologies, is a challenge and an opportunity for industry. New revenue streams are made available through the Web and the knowledge-based economy, and Information Technology has enabled business activities that were previously difficult or impossible to achieve. Societal changes are taking place in how people interact with technology and use it in their everyday lives. Industry members are aware of the potential to gain ground when being the first to react to changes in their sector and taking advantage of new technological possibilities ahead of their rivals.

The following emerging business issues require technological solutions and support for industry to deploy those technological solutions. The Semantic Web is a possible solution. Where we may lack concrete use cases at this stage to be able to evaluate the benefits of Semantic Web technologies to the given emerging business issues it is possible for us to analyze potential benefits by considering the technical requirements of the business issues:

- **Multimedia services** on a small scale do not need semantic technologies. However on a large scale, where there is a developing need for automatisation, multimedia content must be retrieved and delivered by computer systems. In this case semantics can guide the retrieval and delivery process. Another aspect of multimedia which can benefit from semantics is content adaptation for different content delivery contexts.
- **Mobile services** again on a small scale do not need semantic technologies. However personalization is a major selling factor and this can be facilitated by semantics. Another factor which may be facilitated is that of location-based services.
- **Intra-enterprise application services** are being implemented presently using Web Service technologies without semantics. However their limitations, particularly in automating the discovery, composition, co-ordination and invocation of Web Services, have led to the research effort in Semantic Web Services.

[12] Some cases of business application of multimodal solutions are given in http://www.sarit.ch/mirror/scsc04.vptt.ch/docs/MultiModal_SCSC04_Nuance_Logictree.pdf
[13] http://www.verivox.de/News/ArticleDetails.asp?aid=1165&pm=1

- **Collaborative platforms** (knowledge management, social networks) are being developed without semantic technologies. However there is also a growing field for content management systems. Such systems can use semantics to adapt the user's view of the content to his/her needs, and to improve search and retrieval aspects. Shared workspaces are also established technologies without semantics. The potential of semantics here is again the personalization to the user as well as computational support of the collaboration e.g. automated user guidance or resource integration. A key functionality within such systems is the semantic mapping between different terminologies or ontologies.
- **Digital content access, rights and security** are being examined through a variety of specifications and standards, which up to now have not been considering the role of semantics. The potential role of semantics would be to allow for richer interpretation of security rules to ensure correct access to and usage of digital content. Security and trust issues are also important aspects of content and service sharing on the Web which require a common understanding of specified rules and restrictions, which could be achieved through the use of ontologies.
- **E-business and e-commerce** function today on the Web and on distributed computer systems (such as Intranets). They rely on electronic messaging with previously agreed message syntax. Semantics can extend this approach to allow for a wider scope of operation, language independence and richer modeling of relationships between concepts in an ontology. Semantic Web Services should be a major contribution to the demonstration and take-up of semantically based e-business and e-commerce.
- **Electronic retail** includes electronic payment systems, supply chain management, logistics and product tracking. Customer and product information management requires data synchronization which means standard data formats, classification schemas and information-exchange protocols. Ontologies could form the basis for this standardization. For example, Supply Chain Management is primarily concerned with the efficient integration of suppliers, factories, warehouses and stores so that merchandise is produced and distributed in the right quantities, to the right locations and at the right time, and so as to minimize total system cost subject to satisfying service requirements. The increase in use of SCM at all levels of decision making, including its need to handle real time data, integrate with user's information systems, collaborate both within the company and across different companies, and co-ordinate decision making makes it a key client of Semantic Web technologies.

References

[1] "XML Watch: Finding friends with XML and RDF" by Edd Dumbill. http://www-106.ibm.com/developerworks/xml/library/x-foaf.html
[2] "The Anatomy of the Grid: Enabling Scalable Virtual Organizations" by Ian Foster, Carl Kesselman and Steve Tuecke. http://www.globus.org/research/papers/anatomy.pdf
[3] "State of the art of current Semantic Web Services initiatives" by Lyndon Nixon and Elena Paslaru. Knowledge Web internal deliverable 2.4.ID1. Available from the Knowledge Web portal http://knowledgeweb.semanticweb.org

[4] “Report on current usage of Web Services and Semantic Web” by Joachim Quantz and Thorsten Wichmann. DIP internal deliverable 12.1. DIP members can access the document from http://dip.semanticweb.org/deliverables.html

4. Conclusion

4.1 Aim of the document

The collection of use cases in this document function not only as illustrative examples of actual or hypothetical deployment of Semantic Web technologies in concrete business cases, but also serve as the basis for an analysis of the potential benefits of Semantic Web technology deployment to strategic industry sectors and the selection of a few key use cases which will be prioritized for demonstrating the value of the technology on the more mature application fields in an unambiguous way.

It is important for all participants in society – Education, Industry, Research – to have a clear awareness of what benefits can be enjoyed in which cases if the Semantic Web is to successfully move out of the research labs and into our everyday practises. We can justify this thus:

- Educational activities should be focused on the right target groups and with appropriately prepared materials so that uptake and use of Semantic Web technologies can be promoted effectively and efficiently,
- Academic research should be producing work which will have practical benefit and not just spending money and time on things that can already be done well without the Semantic Web or that do not need to be done,
- Industry should be applying Semantic Web technologies where they offer benefits and conversely not applying them where they do not: early adoption of Semantic Web technologies for the wrong tasks will lead to industry-wide negative opinions of these new approaches and discourage interest in correct solutions where industry stands to benefit.

Consequently this analysis is a vital and valuable indicator for the correct direction of Semantic Web technology promotion, education, research and deployment. It benefits from being based on a significant collection of use cases which have been provided by industry and which represent actual business cases. This collection represents a wide range of industry sectors and reflects the current trends in business practices and technology use. Furthermore, by selecting a smaller number of key use cases that exemplify the application of Semantic Web technologies in the more mature application fields, we can offer some evidence of the benefits of the technology and have an ideal summary of this technology application for publicizing and disseminating more widely.

At this stage we have only just begun to form the Industry Board and collect use cases. The collection stands at 16 contributed use cases. From this, we will analyse which semantic technology-based solutions are being sought by industry and what problems currently exist in realising these solutions. From this analysis, we can identify the key areas where solutions are sought and the key locks to their realisation. These are extracted from the use case descriptions and are listed in Appendix C. A study of the relative weights of the different industrial sectors will then be used to select 4 key use cases.

4.2 Analysis of use cases

Here we present the findings of the analysis (Figures C.1 and C.2).

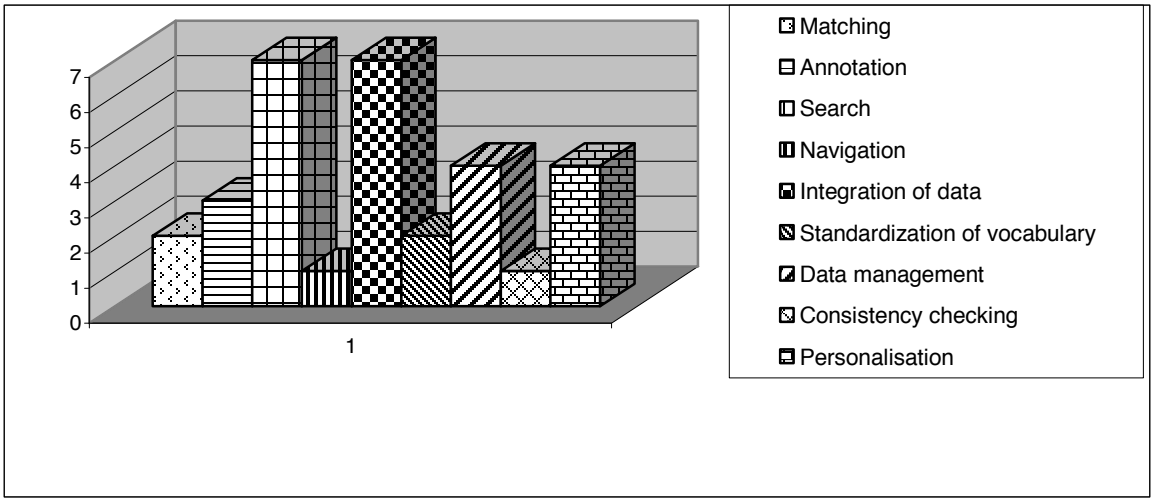

Figure C.1 – Semantic solutions sought in use cases

In Figure C.1 we see that the two areas in which nearly half of our use cases were seeking solutions are that of **data integration** and **semantic search**. Other areas which were mentioned in a quarter of use cases were **data management** and **personalization**. However the issues of search and integration are clearly significant application domains for semantic technologies.

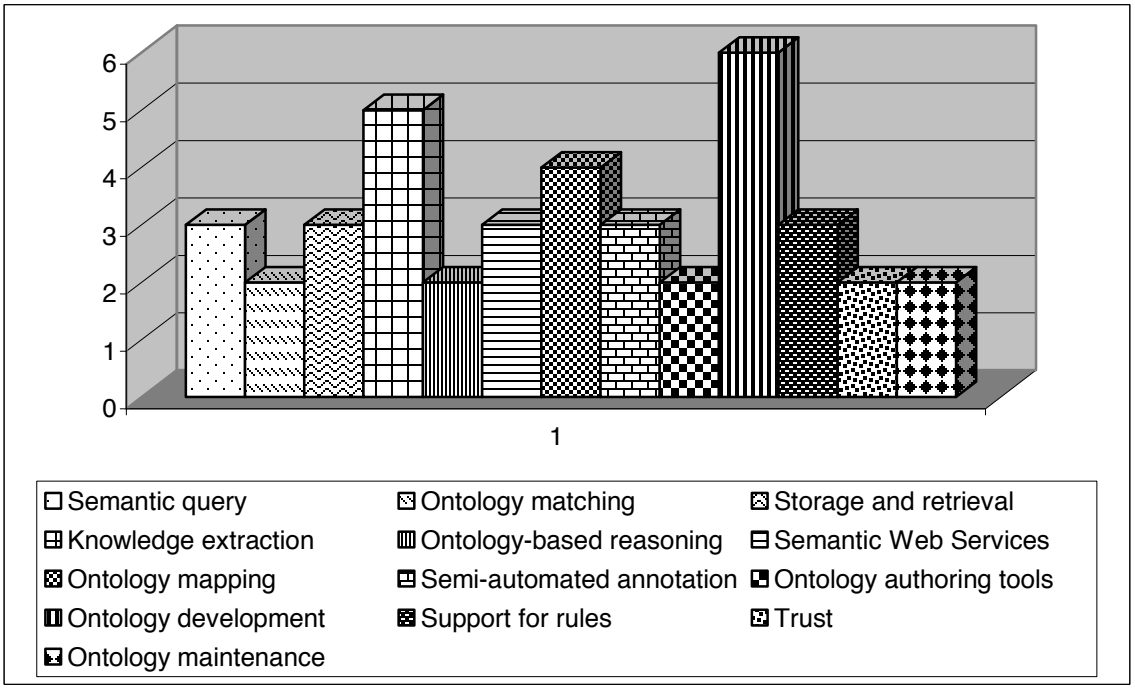

Figure C.2 – Technological locks identified in use cases

For legibility, we consider in Figure C.2 only the technology locks which were mentioned at least twice. Here we see three locks which stand out, occurring in 4 to 6 of our use cases respectively: these are **knowledge extraction**, **ontology mapping** and **ontology development**. While the relatively even spread of technology locks indicates that issues to be resolved in using semantic technologies in an industrial setting are broad (we can understand this as indicating that the entire Semantic Web technology framework requires further standardization, stabilization and support by enterprise-ready tools), industry is particularly needing solutions to the extraction of knowledge from existing sources (acknowledging the heavy cost and time commitment needed to generate knowledge from scratch), creation of ontologies – whether from scratch or re-use of existing ones, with intuitive tools and clear modeling guidelines – and the mapping between them – defining how knowledge defined by one ontology could be represented using a different ontology. The latter relates well to the issue of data integration and semantic search, as heterogeneous sources will likely use different ontologies yet functionalities like search need to know the relationships between those ontologies.

4.3 Selection of Key Use Cases

An effort has been made to focus on selecting use cases from industry sectors that is reflective of the strategic importance of those sectors to the European economy and the level of their activity in seeking Research and Development results. For a basic analysis of the industrial sectors, statistical publications from the OECD and EU were consulted[14]. A summary of our findings is:

- The growing role of knowledge is reflected in economic performance. Trade in high-technology goods such as aircraft, computers, pharmaceuticals and scientific instruments accounted for over 25% of total trade in 2000-1, up from less than 20% in the early 1990s.
- The services sector now accounts for 70% of OECD GDP while manufacturing accounts for about 18%. Part of this increase reflects growing demand in the manufacturing sector for services, particularly through outsourcing.
- In a survey of 9 European countries, entry rates for enterprises are substantially higher in the dynamic service sectors such as business services or ICT-related industries.
- R&D expenditure is rising steadily, with 28% of the world total being spent in the EU. Most of the rise is due to higher business investment in the services (23% of world total) and high-technology industries (47% of EU manufacturing R&D).
- 64% of enterprises in computing/technology, and 58% in banking and finance are actively innovating.
- The average annual growth rate in EU R&D expenditure over 1998-2002 was strongest in the sectors 'IT hardware', 'automobiles & parts' and 'pharma & biotech'. In IT and pharmaceuticals EU firms still spend comparatively much less than their US counterparts.

[14] Sources were the OECD "Science, Technology and Industry Scoreboard 2003" and the EUROSTAT publications "Innovation in Europe" and "Science, Technology and Innovation – Key Figures 2003-4"

The KnowledgeWeb Industry Area activity has identified 20 business sectors which are listed on the Industry Area portal[15]. In terms of the findings above, we can select which sectors are economically the most important for the EU (in bold) and which are innovatively the most active and interested (in italics).

Aerospace	*Automobile*	*Banking & Finance*	***Computing & Electronics***
Consumer Goods	Distribution	Energy & Public Services	Environment
Food Industry	Government	Industry & Construction	Internet
Luxury Goods	**Media & communications**	***Pharmaceuticals & health***	***Service industry***
Sports	**Technology providers**	Telecommunications	Transport & logistics

As a result, we identify 4 key business sectors which combine strong economic performance and high levels of expenditure on research and development. We consider these sectors as key areas for the examination of Semantic Web technologies to further support their innovation and growth. These sectors are printed in red in the above table.

In total, 16 use cases have been contributed in this time period by KnowledgeWeb partners. The complete collection of use cases is given in Appendix B. It can be seen (Figure C.3) that the business cases relating to the following 9 industry sectors have been contributed:

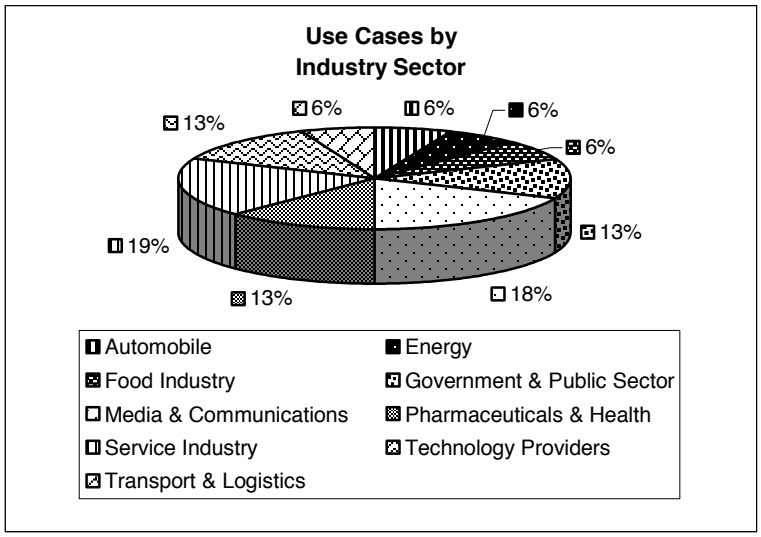

Figure C.3 – Breakdown of contributed use cases by industry sector

From these 16 use cases, 4 come from the key business sectors of Pharmaceuticals & Health and the Service Industry.

1. Recruitment from FU Berlin / ‚KnowledgeNets' project (p.11)
2. Data warehousing in healthcare by FU Berlin / Semtation (p.30)
3. B2C marketplace for tourism by France Telecom (p.33)
4. Hospital information system by VUB / L&C (p. 53)

[15] http://knowledgeweb.semanticweb.org/o2i

It is also interesting to note that three of these use cases cover the key semantic solution of data integration, and the final case also covers the issue of search (to be fair, the first use case on recruitment also has a future integration component where external recruitment data would be crawled, and all the cases would include a search functionality). Also, in terms of the key technology locks, ontology development is an identified lock in the second case and ontology mapping a lock in the fourth case.

We would like to use these cases in particular for further dissemination of the potential solutions of Semantic Web technologies to industry as well as the identified pitfalls to be overcome to research through the KnowledgeWeb Industry Area activities. On the other hand, we should focus in the next phase on also collecting use cases in particular from the Aeronautical and the Computing & Electronics sectors.

4.4 Concluding Remarks

This document is intended to introduce a collection of business use cases for the actual or hypothetical application of Semantic Web technologies to business problems and an analysis of the potential benefits of Semantic Web based approaches in meeting particular business problems as well as the current technological issues in terms of their application in strategic industries, exemplified by a smaller number of key use cases.

We have introduced how the business use cases were collected and what methodology was used to produce the final text. From the text, we could extract the sorts of business problems for which a semantic technology-based solution was being sought, and found that the two most significant application areas for semantics in industry are in search and data integration. We also extracted the identified technology locks, finding that while industry needs in general a matured Semantic Web framework of standards, practices and tools, three main locks to Semantic Web technology uptake are knowledge extraction as well as ontology development and mapping. These findings and further analyses carried out by Knowledge Web are intended to focus and direct the Semantic Web research effort to industry needs.

Industrial sectors were identified and analyzed in terms of their significance for Semantic Web technology introduction, using as criteria both the economic importance of the sector and the extent of its expenditure on research and development. Four key use cases were selected which best exemplify the application of Semantic Web technology on the more mature application fields in an unambiguous way. These could be used to more widely publicize and disseminate to industry the potential benefits of Semantic Web for their business problems, consolidating our focus on strategic industry areas.

Finally, use cases will continue to be collected to further populate the Knowledge Web industry portal alongside the use cases listed in this document.

Invitation to contribute to business case studies

Alain Léger and Robert Meersman (Co-chairs, Industrial Area)

Dear Colleague,

Thank you for joining the KnowledgeWeb Network of Excellence as an Industry Partner. We sincerely hope this will be the beginning of a fruitful exchange of knowledge and experience which will benefit us all in building up the Semantic Web and realising its potential for real world business use cases.

In order to support the co-operation of industry and research partners, one of the first tasks of KnowledgeWeb is to collect business case studies, illustrating the potential of Semantic Web technologies in meeting challenges in the industry environment – whether hypothetically or in concrete cases (including prototypical implementations).

We would like to hear about your concrete needs and possible challenging vision[16] that exists in your business domain that can be met by semantic technologies. We include a simple questionnaire which you can use to outline your case study. We also invite you to take the opportunity to meet with a local research partner to further discuss your case study and explore how ongoing Semantic Web research can better meet your business needs.
From the collected case studies we will carry out an analysis of the IT system and knowledge processing requirements for Semantic Web deployment in industry, feeding these results to the research activities in the KnowledgeWeb network. **As an Industry Partner, you will have exclusive early access to these results.**

We invite you to contribute your case studies to the KnowledgeWeb Network of Excellence. By doing so, you can know:

- You are participating in an EU-wide analysis of Semantic Web needs for industry;
- Your business needs and interests will act as an input to the further research and development of the KnowledgeWeb research partners;
- Future Semantic Web research will be better positioned to generate viable benefits for your business;
- You will have early access to the research results from KnowledgeWeb.

[16] Note also that in any case we can fully understand your need to not expose your internal strategy in a public document, and so in agreement with you we will fully respect this position.

Business case study contribution

Answers need only be a few lines in length.
Please refer to the example case study enclosed as a guideline.
(Your input will be given in the first box in the case study. The detailed text will be drawn up by a research partner following further discussion with you)
Thank you for your time and your contribution.

Company Name

Industry economical sector

OVERVIEW

Describe a concrete needs and challenge which you encounter in your business activities.

What is a proposed solution to meet this challenge?

Why does this solution need or involve the use of semantic technologies?

What are your expected key business benefits from realising this solution?

Besides your company, what other type of business partners would be necessary or desirable to realise this case study?

Do you agree to a visit from a local Knowledge Web research partner to further discuss your case study with you?

GENERAL

Links to further material on the case study (URLs or literature references to demos, mock-ups, slides, articles, etc.):

Existing technological solutions

Proposed/prototypical technological solutions (e.g. demos, mock-ups, slides, articles)

Technology locks (that you can anticipate)

Contact person for case study

Telephone number:

E-mail:

Please send the completed questionnaire:
By e-mail to nixon@inf.fu-berlin.de
Or by fax to +49 30 838 75220
Thank you.

Appendix B: Collected Use Cases

Page	Title	Contributing KW partner	Contributing industry partner	Industry sector
11	Recruitment	FU Berlin	Wissensnetze project	Service Industry
15	Multimedia content analysis and annotation	CERTH	AceMedia project	Media & comms
20	Peer-to-peer eScience Portal	FU Berlin	Neofonie GmbH	Gov't & Public Sector
23	News aggregration service	FU Berlin	Neofonie GmbH	Media & comms
27	Product lifecycle management	FU Berlin	Semtation GmbH	Tech providers
30	Data warehousing in healthcare	FU Berlin	Semtation GmbH	Pharma/health
33	B2C marketplace for tourism	France Telecom	France Telecom	Service Industry
38	Digital photo album management	France Telecom	France Telecom	Media & comms
42	Geosciences Project Memory	France Telecom	IFP	Energy
45	R&D Support for Coffee	Uni Trento	Illy Caffe	Food industry
49	Co-ordination of Real Estate Management	Uni Trento	Trenitalia S.p.a.	Transport & logistics
53	Hospital Information System	VUB / text by FU Berlin	L&G	Pharma/health
56	Agent-based System for an Insurance Company	UPM / text by FU Berlin	Acklin NV	Service Industry
59	DaimlerChrysler Semantic Web Portal	FU Berlin	Daimler Chrysler	Automobile
61	Specialised Web Portals for Businesses	UPM	Robotiker	Tech providers
65	Integrated Access to Biological Data	UPM	Robotiker	Gov't & Public Sector

Appendix C: Analysis of Use Cases according to Solutions and Locks

Page	Title	Semantic Solutions	Technology Locks
11	Recruitment	Matching	Semantic query, ontology matching, storage and retrieval
15	Multimedia content analysis and annotation	Annotation, search and navigation	Knowledge extraction, multimedia modeling, ontology-based reasoning
20	Peer-to-peer eScience Portal	Integration, search	Semantic Web Services, ontology mapping, knowledge extraction
23	News aggregration service	Integration	Ontology mapping, knowledge extraction, semi-automated annotation
27	Product lifecycle management	Standardization, data management	Ontology authoring tools, ontology development, support for rules
30	Data warehousing in healthcare	Integration, consistency, personalisation	Ontology development, support for rules, support for measures
33	B2C marketplace for tourism	Integration, personalisation	Semantic Web Services, semantic integration, trust
38	Digital photo album management	Annotation, search, data management	Semantic Web Services, semi-automated annotation, trust
42	Geosciences Project Memory	Annotation, search	Ontology development, ontology maintenance, semi-automated annotation, semantic query, storage and retrieval
45	R&D Support for Coffee	Search, personalisation	Ontology development, ontology maintenance
49	Co-ordination of Real Estate Management	Matching, data management	Semantic query, ontology matching, storage and retrieval
53	Hospital Information System	Integration, search	Ontology mapping, visualization
56	Agent-based System for an Insurance Company	Standardization, integration	Agent systems, ontology mapping
59	DaimlerChrysler Semantic Web Portal	Search, data management	Ontology authoring tools, ontology development, support for rules
61	Specialised Web Portals for Businesses	Personalisation	Ontology development, knowledge extraction
65	Integrated Access to Biological Data	Integration	Knowledge extraction, ontology-based reasoning