INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team EDELWEISS

Exchanges Documents Extraction Languages Web Ergonomics Interaction Semantics Servers

Sophia Antipolis - Méditerranée

THEME COG

Table of contents

1. Team

Head of project team
Rose Dieng-Kuntz [Research Director (DR) INRIA]

Vice-head of project team
Olivier Corby [Researcher (CR1) INRIA]

Administrative assistant
Patricia Maleyran [TR INRIA, till September]
Angela Ouvrier [since September]

Staff member Inria
Jean-François Baget [Researcher (CR2), INRIA]
Fabien Gandon [Researcher (CR1) INRIA]
Alain Giboin [Researcher (CR1), INRIA]

Staff member University
Michel Buffa [Assistant professor, I3S, UNSA, till August 31st]
Martine Collard [Assistant professor, I3S, UNSA, since October 1st]
Isabelle Mirbel [Assistant professor, I3S, UNSA, since September 1st]

External Research scientist (partner)
Catherine Faron-Zucker [Assistant professor, I3S, UNSA]

Project technical staff
Mohammed Bennis [Development Engineer]
Virginie Bottollier [Associate Engineer, till October]
Hacène Cherfi [Development Engineer]
Priscille Durville [Development Engineer]
Adil El Ghali [Development Engineer]
Emmanuel Jamin [Development Engineer, since April 1st]
Khaled Khelif [Development Engineer]
Leila Kefi-Khelif [Development Engineer, since July 1st]
Amira Tifous [Development Engineer]

Ph. D. student
Sylvain Dehors [INRIA, UNSA, till February]
Guillaume Erétéo [INRIA, UNSA, since December]
Phuc-Hiep Luong [AUF Grant/ INRIA, Ecole Nationale Supérieure des Mines de Paris, till December 14]
Noureddine Mokhtari [INRIA, UNSA]
Freddy Limpens [INRIA, UNSA, since September 1st]
Amel Yessad [University Badji Mokhtar, Annaba, Algérie, since September]

Visitor
Moussa Lo [AUF Grant/ Gaston Berger University, Saint-Louis, Senegal, from November 19 till December 1st]
Kalina Yacef [Invited Professor, University of Sydney, Australia, from April 1st till June 30]

Trainee
Ouassim Akabbal [Polytech'Nice-Sophia Antipolis, from from March 1st till June 30]
Carine Colombo [Polytech'Nice-Sophia Antipolis, from April 23 till September10]
Ibrahima Diop [University Gaston Berger, Saint-Louis, Senegal, from July 23]
Eleni Kottaki [University of Patras, Greecefrom March 26 till September 9]
Cheikh Niang [University Gaston Berger, Saint-Louis, Senegal, from July 23]
Bassem Makni [ENSI, Tunisie, from February 12 till August 10 and from October 22]
Nizar Ghoula [ENSI, Tunisie, from February 12 till August 9]
Yassine Mrabet [ENSI, Tunisie, from February 12 till August 9]

Minh Tuan Nguyen [IFI, Vietnam, from March 21 till September 9]

2. Overall Objectives

2.1. Introduction

2.1.1. Context and Objectives

Actors and interaction devices are becoming more and more mobile while knowledge sources, services and their networks are becoming ubiquitous. In this context we witness the emergence of communities of interest and/or practice, very light and agile structures that can be ephemeral and virtual. To assist the life-cycle of such communities we are interested in providing tools and methodologies supporting the interactions and the memories of these focused groups. Throughout its life time, a community uses, produces, exchanges, and shares resources materializing knowledge through various types of documents (that may be structured or not, textual, multimedia, etc.). A community may also rely on some services or programs available inside the community or outside. To ensure mutual understanding between community members, the exchanges inside a community rely on a common terminology and common concepts that may evolve throughout the life of the community. These exchanges can also use various media.
The context of the emergence of such virtual communities (inside organizations, across organizations or independently of any organization) is the use of the Web not only for information sharing but also for support to cooperation, the use of new interaction channels, the evolution of Web technologies (Semantic Web, social Web, Web services, mobile Web, ubiquitous Web).
Edelweiss aims at offering models, methods and techniques for supporting ergonomic, web-based, knowledge management and collaboration in virtual communities interacting with information resources through the Web. Edelweiss will thus perform research on graph-based, ontology-based, web-based knowledge representation and inferences for interacting with or through information resources.

2.1.2. Research Topics

The support to such a virtual community can be studied according to several viewpoints:

- The activities of the community consist of structuring, searching, retrieving, reusing, and composing the community internal or external resources / services. A support to these activities can be offered through a semantic-web-based approach, by processing annotations of such resources / services;
- Conceptual modelling of the interactions and collaboration among community members mediated by tools could enable us to propose ergonomic tools adapted to support such collaboration;
- To achieve the development of such supporting tools and methodologies, basic blocks are needed to represent knowledge and to reason and perform inferences on this representation: we choose to rely on a graph-based representation.

Therefore, we will study thoroughly three complementary research directions, corresponding to these three viewpoints:

- Research direction 1: Annotation of Information Resources
- Research direction 2: Interaction Design
- Research direction 3: Knowledge Graph-based Representation.

Our research topics can be decomposed as follows:

- Semantic Annotation of Information Resources
  - Ontology-guided annotation process
  - Semi-automatic generation of annotations from texts
  - Contextualization, Evolution, Heterogeneity of annotations
  - Reflective annotations and meta-annotations
  - Design and development of an Annotation toolkit
- Interaction Design of Semantic Systems
  - Supporting Human Interoperability in Semantic Activities through Articulating Functionalities and in Scenario Management Activities
  - Experimental evaluation of inferences for information retrieval and other tasks
  - Ontology-based intelligent interfaces
  - Interaction platform
- Knowledge-Graph-based Representation of the Semantic Web Knowledge
  - Semantic Factory and Knowledge Graph Toolkit
  - Scaling graph representations and operations
  - Ontology-based Model Driven Engineering
  - Inferences characteristic to graphs
  - Inferences characteristic to distributed Web sources.

2.1.3. International and industrial relations

We collaborate or collaborated with industrialists in the following fields: aeronautics (Aerospace, Dassault-Aviation, EADS), car industry (Renault, ItalDesign), telecommunications (CSELT, T-NOVA, Telecom Valley, France Telecom), service integration (Atos Origin), semi-conductors (Philips Semi-Conductors, now NXP), manufacturing (Estanda) and earth sciences (BRGM, IFP), and in biology (IPMC, Immunosearch). We also had collaborations with researchers in accidentology (INRETS) and in the field of healthcare (Nautilus), in civil engineering sector (CSTB)and in software engineering (ILOG). We had international relations with Griffith University and CSIRO (Australia), Parma University (Italia), and T-Systems Nova (Germany).

Our work was applied in the context of the IST project CoMMA and we took part in the OntoWeb thematic network. We take part in the Knowledge Web Network of Excellence, in the Integrated Project Palette and in the STREPS projects SeaLife and SevenPro.

2.2. Highlights of the year

The project-team Edelweiss following the Acacia project was created in 2007. We will adopt an Open Source strategy for Edelweiss software.

3. Scientific Foundations

3.1. Foundations

Keywords: *Annotation*, *Artificial Intelligence*, *Assistance to the User*, *Co-operation*, *Cognitive Sciences*, *Collaboration*, *Community of practice*, *Conceptual Graph*, *Corporate Memory*, *Corporate Web*, *Distributed Services*, *Documents*, *Ergonomics*, *Exchanges*, *Extraction*, *Graphs*, *Information Retrieval*, *Interaction Design*, *Intranet*, *Knowledge Acquisition*, *Knowledge Engineering*, *Knowledge Management*, *Knowledge Server*, *Knowledge representation*, *Languages*, *OWL*, *Ontology*, *RDF*, *Reasoning*, *Scenarios*, *Semantic Web*, *User interfaces*, *Virtual community*, *Web*, *Web architectures*, *XML*.

Knowledge Management (KM) is one of the key progress factors in organizations. It aims at capturing explicit and tacit knowledge of an organization, in order to facilitate its access, sharing out and reuse [78]. The considered organization can be an actual enterprise or a public organization, but it may also just consist of a given department or service; it can also be a group, or a community, or a virtual enterprise (made of members possibly stemming from different companies, but sharing a common interest).

The Acacia approach relied on the analogy between the resources of a corporate memory and the resources of the Web. We considered that a corporate memory can be materialized in a corporate semantic Web, that consists of [78], [79]:

- resources (i.e. documents in XML, HTML or non Web-oriented formats, people, services, software, materials),
- ontologies (describing the conceptual vocabulary shared by the different communities of the organization),
- semantic annotations on these resources (i.e. on the document contents, on persons' skills, on the characteristics of the services/software/materials), these annotations using the conceptual vocabulary defined in ontologies.

According to [104], communities of practice (CoPs) are groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis. CoPs can be found within businesses, across business units or across company boundaries [105], still they differ from business or functional units, from teams and networks: people belong to CoPs at the same time as they belong to other organizational structures. An effective organization comprises a constellation of interconnected CoPs, as these are privileged nodes for the exchange and interpretation of information. CoPs preserve the tacit aspects of knowledge that formal systems cannot capture. CoPs can be considered as a means by which knowledge is owned in practice. Indeed, such groups allow the functions of creation, accumulation and diffusion of knowledge in organizations.

The Edelweiss project-team extends this hypothesis to virtual communities and considers that a support to knowledge management and cooperative work in a community can also rely on a community semantic Web.

The underlying research topics are:

- Annotation of information resources:
  - How to offer a methodological and software support to the annotation process combining semi-automation and collaboration among humans through the Web?
  - How to improve linguistic techniques for semi-automatic annotation? How to offer semi-automatic annotation of textual of multimedia resources?
  - How to extract various contextual annotations from texts or graphics? How to represent and use contextual information?
  - How to handle the links among these multiple representations and to choose the most adapted according to the context and the user profile?
  - How to reason on multiple (coherent or possibly contradictory) annotations and how to reason about contextual, multi-viewpoint annotations?
  - How to extract, represent and reason about these meta-annotations?
  - How to evolve annotations?
- Interaction Design: We can identify two main families of interaction issues:
  - Man-machine interaction: where the issues are in the interaction between a machine and its users;
  - Man-man interaction: where the issues are in the computer mediation of human interactions.

Within these two families of interaction a number of precise problems can be identified:

- Problems due to the underlying conceptualization of a system that may not be familiar or even relevant to a user;
- Problems of jargons and languages used and other non textual representations;
- Problems of the appropriateness of an interface to the user's task;
- Problems of compatibility between an operation (e.g. approximate, categorize, etc) as envisaged by a human and the effective way it is performed by a machine;
- Problems of dedicated interfaces and the tension with underlying generic models and processes;
- Problems of integrating / articulating several tools and models involved in a task.

- Knowledge graph representation
  - How to offer a generic knowledge graph toolkit?
  - How to enable scaling graph representations and operations?
  - How to load and index large graphs while being able to apply operators to these graphs in a timely fashion?
  - How to reconcile on one hand graph indexing and compiling and on the other hand memory use?
  - How to study semantic Web software as an application of the MDE (Model Driven Engineering) approach?
  - How to study semantic Web operations from the point of view of model transformation?
  - How to exploit inferences characteristic to graphs in the framework of semantic Web?
  - How to provide relaxed graph matching relying on distances?
  - How to build a generic graph abstract machine?
  - How to enable inferences characteristic to distributed Web sources?
  - How to distribute graph-based algorithms in order to perform operations involving several semantic Web servers?

3.1.1. Semantic Web application development platforms

CoGITaNT [85] is a C++ library for representing and manipulating conceptual graphs (CGs) and related objects such as graph support and rules. It offers CG operations, an inference engine and a graph editor. CoGITaNT was developed first at LIRMM in Montpellier and then in Angers University.

Protégé (http://protege.stanford.edu) is a free, open source ontology editor and knowledge-base framework.The Protégé platform supports two ways of modelling ontologies with frames and OWL editors. Ontologies can be exported into several formats such as RDF(S), OWL, and XML Schema. Protégé is developed at Stanford and by a large community of developers.

Jena (http://jena.sourceforge.net/) is a Java framework for building Semantic Web applications. It provides an environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine. It offers persistency through database connectors. Jena is developed by HP labs at Bristol.

Most of the current contributions on semantic Web servers focus on low-level infrastructural problems and do not really address the new interaction means provided by the semantics added by the semantic Web.

KAON [102] is an open-source ontology management infrastructure targeted for business applications. It includes tools allowing ontology creation and management and provides a framework for building ontology-based applications. Its successor, KAON2, advertises a number of features: an API for programmatic management of OWL-DL, SWRL, and F-Logic ontologies; a stand-alone server providing access to ontologies in a distributed manner using RMI; an inference engine for answering conjunctive queries (expressed using SPARQL syntax); a DIG interface, allowing access from tools such as Protégé; a module for extracting ontology instances from relational databases. Thus KAON2 server capabilities are at much lower level than the one we consider here and focus on APIs and reasoning services.
RDF Gateway is a platform for the development and deployment of Semantic Web applications and InferEd is its associated authoring environment to navigate and edit RDF documents. RDF Gateway is both a Web server and a RDF database server to support the management of metadata on the Web. RDF Gateway is also at a lower infrastructural level i.e. it is closer to the back-end and does not address user interaction and ergonomics.
Likewise, Joseki is a SPARQL Server for Jena: an HTTP and SOAP engine that supports the SPARQL Protocol and the SPARQL RDF Query language. Closer to our concern is the semantic Web application development framework ODESeW applying the Model-View-Controller design pattern to semantic Web applications.
Finally, SIMILE is a joint project conducted by the W3C, MIT Libraries, and MIT CSAIL that seeks to enhance inter-operability among digital assets, schemata, metadata, and services. SIMILE includes in particular three tools relevant to our topic here: Longwell, Semantic Bank and Piggy Bank. Longwell relies on the RDF data model to propose faceted browsing and enable users to visualize and browse RDF dataset and to build a Web site out of these data. A Semantic Bank ensures remote persistence on a server to share data and lets you publish your information, both in RDF and regular Web pages through Longwell. Piggy Bank is an extension to Firefox letting users make use of existing information on the Web and in particular to merge information from several sites and recreate a unifying customized presentation. Thus the focus in SIMILE is on data integration and not so much on using the semantics to design new interfaces.

3.1.2. Semantic Annotations

As detailed in [100], annotation tools can be compared according to:

- The nature of suggestions of the annotation tool: instances of concepts, values of their attributes, instances of relations, metadata on named entities, etc.
- The storage of annotations: in the document source or in an annotation server.
- The language for representing annotations: XML, RDF...
- The language for handling the reference ontologies used for creating the annotations.
- The type of processing: manual process vs. automatic vs. semi-automatic process.
- The automation techniques used: Natural Language Processing (NLP) using linguistic techniques vs. machine learning.
- The fact that resources are internal or external.
- The nature of annotated resources: static Web pages, dynamic Web pages, multimedia resources (images, video, audio)
- The evolution of annotation according to resource evolution.

In the current state of the art, we distinguish the following approaches:

- Purely manual editors (such as OntoMat Annotizer, the Mangrove system or Cohse annotator [72]).
- Manual, collaborative editors such as or Annotea [94] the W3C project for collaborative annotation that proposes Amaya [47] and Vannotea.
- Use of manually written rules or wrappers in order to capture known patterns for the annotations: e.g. Melita enables the user to write rules based on regular expressions; Cafetiere is a rule-based system for generating XML annotations.

- Supervised systems learning from sample annotations marked up by the user: e.g. tools such as Ontomat, MnM and Melita include Amilcare [76] that learns how to annotate the documents by generalizing the users annotations.
- Unsupervised systems using various strategies to learn how to annotate without users supervision: e.g. Armadillo or Ontomat-Pankow relying on the Pankow approach [75].
- Annotation service providers, offering such services on demand for users browsing non-annotated resources: e.g. Magpie [81].

3.1.3. Interaction Design

Initially concerned with formal and technical aspects, the Semantic Web community recently acknowledged the necessity to take seriously into account uses and users of Semantic Web applications so that such applications can be accepted by users and their organizations. An indicator of this new concern is the emergence of scientific events such as the International Workshop on Interaction Design and the Semantic Web (2004), and its successors, the Workshops on End-user Semantic Web Interaction (2005, 2006, 2007). The aim of these workshops is to help Semantic Web application designers bring the power of the semantic Web to end-users, applying Interaction Design and more specifically Social Interaction Design. Interaction Design is the discipline of defining and creating the human interaction with digital, environmental or organizational systems. Interaction design defines the behaviors or interactions of an object or system over time with its users' population. Interaction designers create systems that are typically informed by research on users and their practices. Social interaction design accounts for interactions among users as well as between users and their devices. Social interaction design is practice-oriented. It is concerned with sign and symbolic value, social behaviors, etiquette and norms, groups and communities, structured interactions, and routines, sequencing, and temporal organization.

Interaction design is critical to a number of applications: an application may use state of the art algorithms; if it does not provide a usable interface, it will not be effective. For interactions to be supported efficiently in a community, supporting tools have to be designed taking into account the nature, the rules, the protocols, the context, etc. of these interactions. In particular, community-supporting tools must:

- help users to articulate their activities and the representations they handle during these activities;
- be able to assist or reproduce some of the inferences involved in the interactions and for instance involved in switching representations from some member to another;
- reduce the heterogeneity of information sources and interfaces and ease the integration of the multiple interaction channels used by community for its interactions. Assisting the cooperation within a community will raise issues of personalization, interface ergonomics, context-awareness and transversally it will also raise the issue of the links between semantics (as in knowledge representation formalisms) and semiotics (as in representations for user interfaces).

3.1.3.1. Interaction design and other human-in-the-loop approaches to semantic Web design

The Edelweiss project clearly fits the Interaction Design for the Semantic Web or End User Semantic Web Interaction trend, the goal of which is to optimize user interfaces for semantic Web applications, and user interactions with and through these applications. Interaction issues are sometimes expressed in terms of usability (see, e.g. [92]). Usable Ontology is thus a method and a set of software tools, aimed at supporting domain experts in populating a domain ontology and obtaining a shared consensus on its content.

More generally the Edelweiss project fits what can be called the human-in-the-loop approach to semantic Web design, the goal of which is to adapt semantic Web applications to users and communities of users. This approach encompasses the Interaction Design trend as well as other trends such as Human Semantic Web, Pragmatic Web [107], Sociosemantic Web [106], and Web 2.0 or Participatory Web [97], or Social Web.

The challenge of the human-in-the-loop approach is to effectively bring the power of the semantic Web to end-users. It requires interdisciplinarity and collaborations between Semantic Web researchers and Interaction/UI/HCI/CSCW researchers.

3.1.3.2. User interfaces to the Semantic Web

Interface languages
To hide the complexity of formal languages of the Semantic Web to the users, it has been proposed to provide the users with more understandable languages such as: (1) Natural language: see, e.g., AquaLog Q/A system or SemSearch engine; (2) Semi-structured languages: query interfaces which do not restrict the user with a formalistic language, but also impose some structure on the users input so as to be also domain-independent. (3) Graphical and Graph languages: In the Human Semantic Web concept browser Conzilla, RDF is combined with the more human-readable UML. Conceptual Graphs are sometimes considered as an interface language, but some authors contest its intuitiveness and propose alternative languages, like the Object-Process Methodology (OPM)-based dual graphic-textual representation, in Visual Semantic Web. In Edelweiss, we will not work directly on natural language, but on semi-structured, graphical and graph languages.

Visualizing ontologies
In this special case of the interface language issue, the goal is to make ontologies not only understandable to users, but also graspable when they are complex or large (see, e.g., the CropCircles graph visualization technique). The goal is also to make the ontology pervasive or transparent; that is, to shift end user's goal from visualizing ontologies or concepts per se (rather an ontologist's goal), to visualizing familiar objects referred to internally by ontologies or concepts. This last issue will be a major concern of Edelweiss.

Interface algorithms
Interface algorithms are those algorithms which allow, for example, to materialize users' viewpoints (e.g. view-hierarchies and multi-facet search paradigm), to support the user's query construction process (interpretation algorithms of user moves as query-construction steps), or to exploit the commonalities of the (search patterns, search terms, etc.) history of the current user with other users of the system. Collaborative filtering techniques are based on such commonalities.

Interface ontologies and ontology-based methods for designing user interfaces
Ontologies are now used to represent the world of interfaces (i.e., interface objects), or to ground methods for designing interfaces. The OntoWeaver approach, which provides high level support for Web site design and development, represents this use of ontologies very well.

Interface scenario-making and prototyping
Scenario-making [68], [74] and prototyping, in particular paper prototyping, are widely used methods in HCI design. They are often related to participatory design. Such methods begin to be used seriously in Semantic Web interface design. For example, to better understand user expectations in annotation activities and to inform the design of annotation mechanisms.

4. Application Domains

4.1. Panorama

Keywords: *Accidentology*, *Aeronautics*, *Automobile*, *Biology*, *Engineering*, *Health*, *Micro-electronics*, *Oncology*, *Telecommunications*, *Transportation*.

There are various application domains of the project: our work on technical memory or project memory has applications in engineering (aircraft industry and car industry). Our work on the knowledge servers also has applications in engineering, in the sector of telecommunications (for corporate memory, skills management and technological watch) and in the biomedical field. Edelweiss work on virtual communities have potential applications in medical field, in pharmacological field, in engineering, in earth sciences and in telecommunications.

4.2. Transportation: Accidentology

We collaborated with INRETS for the modeling of knowledge of several experts in road accident analysis (psychologists specialized in the driver's behavior, vehicle engineers, infrastructure engineers). This application of accidentology illustrates an example of (partial) corporate memory and moreover, served as concrete example for numerous works of the team: analysis of co-operation between experts during a collective problem resolution, analysis of explanatory dialogues, comparison between multiple expertise models via our MULTIKAT software, exploitation of CommonKADS method generic models, association of conceptual graphs to expertise documents via our CGKAT software, representation of the artificial agents associated to the experts and their COMMONKADS expertise models, exploitation of the C-VISTA model for the representation of multiple points of view of different experts. We developed the RESEDA system (Intranet Network for Detailed Study of Accidents) in XML and Java, in order to support INRETS for road accident analysis.

4.3. Transportation and Engineering: Automobile

In the context of the improvement of the vehicle design process control, we collaborated with Renault to develop a memory of problems encountered during vehicle projects, whose traces were stored in the corporate information system. The construction of this project memory relied on techniques of knowledge engineering and of linguistic analysis. SAMOVAR (Système d'Analyse et de Modélisation des validations Renault) system can be considered as a concrete example of corporate semantic Web.

4.4. Transportation and Engineering: Aeronautics

In the past, we had collaborated with Aérospatiale and Dassault Aviation on project memory. Recently, we collaborated with EADS Corporate Research Laboratory for building a Corporate Memory for an Industrial Research Laboratory.

4.5. Telecommunications

Our work on corporate memory, in particular the use of intelligent agents, ontologies and XML technology, is of particular interest for companies of the telecommunications sector. We thus collaborated with T-NOVA (Deutsche Telekom) and CSELT (Italian Telecom) in the framework of the CoMMA IST project. T-NOVA applied this work for the assistance to insertion of new employees and CSELT for the assistance to technological monitoring. We also collaborated with Telecom Valley and the GET (ENST and ENST-Bretagne) for our work on skills management in the RNRT KmP project. We finally collaborated with ENST-Bretagne for the CNRS Specific Action on "Semantic Web and E-learning". A new collaboration will start on December on communities of practice in telecommunications.

4.6. Civil Engineering Sector

Our work on corporate memory, in particular the use of intelligent agents, ontologies and XML technology, is also interesting for the construction industry: we thus collaborated with the CSTB (French Scientific and Technical Center for Building) within the framework of the CoMMA project for a scenario of technological watch. We continue a collaboration on the topics of technological watch.

4.7. Micro-electronics

We collaborated with Philips Semi-Conductors, now NXP, for an intra-firm skills management application.

4.8. Biomedical Domain

Our work on corporate memory, in particular our corporate semantic Web approach (ontologies and XML technology), is applied to several biomedical applications: use of linguistic techniques for building an experiment memory for transcriptome analysis (in the framework of the MEAT project in collaboration with IPMC), use of a medical ontology, viewpoints and CSCW for supporting collaborative work in a healthcare network (in the context of the ACI *Ligne de Vie* project in collaboration with the SARL Nautilus and SPIM (Service de Santé Publique et d'Informatique médicale de la Faculté de Médecine Broussais-Hôtel Dieu)). In the framework of SeaLife IST project, we work on a semantic browser for Life Sciences, with scenarios such as evidence-based medicine, or literature and patent mining. In Immunosearch project, our work on literature mining seems useful for supporting experiments aimed at studying harmlessness of the molecules used in perfumes, aromatics and cosmetics

4.9. Earth Sciences

We collaborate with IFP and BRGM on semantic portals enabling access to resources and services in Earth Sciences domain. Semantic portals will in particular assist geologists in discovering geological sites where storing carbon dioxide (CO_2) produced by power stations, so contributing to reductions in global Greenhouse Gas emissions.

5. Software

5.1. Corese

Keywords: *Conceptual Graph*, *Information Retrieval*, *OWL*, *RDF*, *RDFS*, *SPARQL*, *Semantic Web*, *XML*, *ontology*.

Participants: Olivier Corby [correspondant], Virginie Bottollier, Fabien Gandon.

5.1.1. Description.

Corese (COnceptual REsource Search Engine) is an RDF(S)-dedicated engine based on Conceptual Graphs (CG) http://www-sop.inria.fr/edelweiss/wiki/wakka.php?wiki=Corese. It enables us to load RDFS schemas and RDF annotations and to transform them into conceptual graph formalism. It then enables us to query the base of annotations thus created, by using the projection operator offered by the conceptual graph formalism.

Corese implements RDF, RDFS, some statements from OWL Lite and the SPARQL query language (Simple Protocol and RDF Query Language). Furthermore, Corese query language integrates original features such as approximate search, group, count, graph path. Approximate search consists of searching the best approximate answers to a query according to the ontology. Graph path enables to search the graph structure of RDF. Corese also integrates an RDF Rule Language based on the CG Rule model. The inference rule engine works in forward chaining.

Corese is a semantic Web engine that enables us to design and develop semantic Web applications; it is available for download. Corese is embedded in a Semantic Web Server (based on Tomcat) and called Sewese.

Corese benefited from an INRIA operation of software development (ODL) in addition to a software engineer, intended to improve quality of the implementation in order to support its diffusion and from an associate engineer. In 2007, we decided to distribute Corese as open source software under licence CeCILL-C.

5.1.2. Applications.

Corese has been applied in more than 20 applications at the INRIA. It is used as a search engine :

- for the RNRT KmP project on skills management,
- for the KmP-Drire project following this RNRT project with Telecom Valley,
- for the KmP-Philips project on intra-firm skills management with Philips Semi-Conductors (now NXP),
- in the *Ligne de Vie* project on healthcare network,
- in the MEAT project on experiment memory on transcriptome analysis,
- in our co-operation with EADS on corporate research laboratory memory,
- in QBLS system for e-Learning,
- in SweetWiki semantic wiki,
- in the SevenPro IST project,
- in the e-WOK_HUB ANR RNTL project,
- in the Palette IST project.

In the past, Corese was the cornerstone of four co-operations of the former Acacia team:

- the IST project, CoMMA (Corporate Memory Management through Agents) [83], [84],
- the SAMOVAR project with Renault [87], [89], [88],
- the co-operative research action ESCRIRE [95],
- the Color action Aprobatiom with CSTB.

5.1.3. Diffusion.

- Corese was registered at APP.
- Corese was made available to:
  - Renault,
  - ATOS Origin,
  - T-Systems NOVA (Deutsche Telekom),
  - CSTB,
  - CSELT (Telecom Italia),
  - LIRMM,
  - M@inline team at ESSI,
  - CETU (Centre d'étude des tunnels du Ministère de l'Equipement).
  - University of Santiago Chili,
  - ENST Bretagne,
  - Tech-CICO team at Université Technologique de Troyes (UTT),
  - Facultad de Informatica, LSIIS,
  - Zuhlke Engineering AG, CH,
  - W3C Group on the Social Meaning of RDF Graphs, Deltek Systems, Inc. USA,
  - Galaad team at INRIA Sophia Antipolis,
  - the partners of the SevenPro IST project: Semantic Systems, etc.
  - the partners of the e-WOK_HUB RNTL project: EADS, etc.
- The work on Corese was published in [32], [33],[4], [5], [3], [77].

5.2. Sewese

Keywords: *RDF*, *RDFS*, *SPARQL*, *Semantic Web Server*, *ontology*.

Participants: Fabien Gandon [correspondant], Priscille Durville.

5.2.1. Description

Sewese is a generic factory to design and develop semantic Web servers and portals [34]. It is designed to embed Corese as semantic search engine and is based on Tomcat.

Sewese enables us to design semantic forms in order to edit predefined queries. An XSLT (the Extensible Stylesheet Language Transformations) based compiler generates a JSP form and a JSP for processing the form. Sewese enables to process query results by means of XSLT stylesheets. It manages a set of compiled stylesheets to improve performance.

Sewese is based on Tomcat filters to build a semantic Web site including session, menu and content management. Sewese also includes a JSP Tag library for predefined processing such as graphic tags for RDFS ontology browsing and editing and for RDF annotation editing.

5.2.2. Applications

- ECCO, cooperative ontology editor,
- SweetWiki, semantic wiki.

5.2.3. Diffusion

- Sewese was registered at APP.
- Sewese was made available to:
  - the partners of SevenPro STREPS project,
  - the partners of e-WOK_HUB RNTL project.

5.3. SweetWiki

Keywords: *RDF*, *RDFS*, *SPARQL*, *Semantic Wiki*, *ontology*.

Participants: Michel Buffa, Guillaume Ereteo, Fabien Gandon [correspondant], Adil El Ghali, Amira Tifous.

5.3.1. Description

SweetWiki, a semantic wiki, supports all the common wiki features such as easy page linking using WikiWords, versioning, etc., but also innovates by integrating a WYSIWYG editor extended to support social tagging functionalities, embedded SPARQL queries etc., masking the OWL-based annotation implementation. Users can freely enter tags and an auto-completion mechanism suggests existing ones by issuing queries to identify existing concepts with compatible labels. Thus tagging is both easy and motivating (real time display of the number of related pages) and concepts are collected in folksonomies. Wiki pages are stored directly in XHTML or in JSPX format, embedding semantic annotations (RDFa and GRDDL) ready to be reused by other software.

5.3.2. Applications

- SweetWiki was distributed to several communities of practice in the framework of the Palette project.
- SweetWiki is used as wiki of the consortium of the e-WOK_HUB ANR RNTL project.
- A robotics (RoboSoft) company uses SweetWiki for parts of its intranet and parts of its public community site
- Informal tests: we have carried out experiments with children, teachers, students and secretaries asking them to use it for one of their daily tasks.

5.3.3. Diffusion

Sweetwiki will be distributed as open source software under CeCill-C licence.

5.4. ECCO

Keywords: *RDF*, *RDFS*, *SPARQL*, *Semantic Web*.

Participants: Priscille Durville [correspondant], Fabien Gandon, Alain Giboin.

5.4.1. Description

We haved designed and implemented a cooperative ontology editor, named ECCO, dedicated to support end-users with different profiles (domain expert, engineer, ontologist, ...) in a cooperative process of ontology construction and evolution.

5.4.2. Applications

ECCO is used in:

- the e-WOK_HUB ANR RNTL project,
- the Palette IST project.

5.4.3. Diffusion

It has been distributed to the partners of the e-WOK_HUB project.

5.5. KmP

Keywords: *RDF*, *RDFS*, *Semantic Web Server*, *ontology*, *skills management*.

Participants: Priscille Durville, Sémi Gaieb, Fabien Gandon, Olivier Corby, Alain Giboin.

5.5.1. Description

The KmP System is a semantic Web server based on Corese, and a real-scale application illustrating an "inter-corporate semantic Web", with 20 pilot companies. It comprises:

- KmP inter-firms ontologies,
- KmP ontology-based interfaces, developed by following the scenario method adapted to the design of usable ontology-based interfaces,
- a semantic clustering algorithm adapted to user representation of concept similarity.

The KmP system was evaluated by the pilot companies, through scenario-based approach. A pre-industrialization of KmP is planned.

5.5.2. Applications

KMP has been adapted:

- for the KmP-Drire project following the RNRT project with Telecom Valley,
- for the KmP-Philips project on intra-firm skills management with Philips Semi-Conductors (now NXP) for which a new version has been developed using Sewese.

The work on KMP was published in [86], [82].

5.6. MeatAnnot

Keywords: *Natural language Processing*, *Ontology*, *RDF*, *RDFS*, *Semantic Annotation*, *Semantic Web*, *Text-Mining*.

Participants: Khaled Khelif [correspondant], Rose Dieng-Kuntz.

5.6.1. Description

MeatAnnot is a software enabling the automatic generation of ontology-based semantic annotations: starting from a textual document, it allows us to generate a structured semantic annotation, based on a domain ontology, and describing the semantic contents of this text. MeatAnnot relies on several NLP techniques (e.g. modules of GATE (General Architecture for Text Engineering), RASP (Robust Accurate Statistical Parsing) parser and a relation extraction grammar we wrote in JAPE); it extracts information from text, instantiates concepts and relationships of the reference ontology and generates RDF annotations for the document.W

5.6.2. Applications

MeatAnnot was applied:

- on a corpus of scientific articles in biology for the MEAT project,
- on a corpus of patents in the PatAnnot system of patent mining,
- on the GeneRIF (Gene Reference Into Function) corpus in the framework of the Immunosearch project,
- on the Neli (National Electronic Library of Infection) Web site in the framework of the SUPROD system aimed at user profile detection ,
- for term extraction from design documents of Estanda and ItalDesign in the framework of the Sevenpro project.

5.6.3. Diffusion

MeatAnnot has been distributed to IRIT.

6. New Results

6.1. Annotation of Information Resources

Keywords: *Assistance to the User*, *Co-operation*, *Corporate Memory*, *Corporate Semantic Web*, *Evolution*, *Information Extraction*, *Knowledge Acquisition*, *Knowledge Acquisition from Texts*, *Knowledge Engineering*, *Knowledge Management*, *Language Technology*, *Natural Language Processing*, *Ontology*, *Semantic Annotation*, *Semantic Web*, *Text-mining*.

The objective of this research direction is to propose (1) a methodological guide for collaborative, semantic-Web-based, annotation process in a community of practice; (2) an ontology-based, service-oriented annotation toolkit offering a service of semi-automatic annotation from textual documents, a service of collaborative annotation and management of evolution of the annotations. The methodological guide and the toolkit will tackle complex contextualization of annotations, various kinds of Web-accessible external resources, reflective annotations and more complex types of heterogeneous resources and services.

6.1.1. Management of Corporate Semantic Web Evolution

Keywords: *Corporate Memory*, *Corporate Semantic Web*, *Evolution*, *Ontology Evolution*, *Semantic Annotation*.

Participants: Phuc-Hiep Luong, Rose Dieng-Kuntz.

This work is being carried out within the framework of Phuc-Hiep Luong's PhD thesis [21] that aims at solving some problems related to the life cycle and evolution of a Corporate Semantic Web (CSW): evolution of each component of a CSW (resources, ontologies and semantic annotations) as well as evolution of relations among these components.

We focused on the ontology evolution, its influence on semantic annotations expressed with the vocabulary provided by the underlying ontology and the evolution of these semantic annotations. We proposed a new approach for the evolution management of a CSW and we focused on two main scenarios of ontology evolution : (i) with trace and (ii) without trace of ontology changes which are carried out during its evolution [26].

Corresponding to these scenarios, we proposed a procedural approach and a rule-based approach in order to manage semantic annotations evolution and particularly to detect inconsistent annotations and to guide the process of resolution of these inconsistencies. We have established a set of ontology changes and all the possible solutions for each ontology change operation allowing users to select an appropriate way to repair inconsistencies in ontology or in semantic annotations.

In order to detect and to correct automatically inconsistent semantic annotations, we implemented in our rule-based approach some inconsistency detection rules: they enable to find different kinds of annotation inconsistencies (i.e. inconsistency on concept, property, domain, range or datatype) with respect to the new ontology. These detected inconsistencies will be solved with the help of correction rules and resolution strategies for semantic annotations. These propositions were implemented and validated in CoSWEM (Corporate Semantic Web Evolution Management) system. CoSWEM facilitates the evolution management of the CSW and particularly manages the propagation of the ontology changes to the semantic annotations depending on this ontology [39], [21].

This system has been developed as a Web-based application integrating the semantic search engine Corese and the Sewese library dedicated to semantic data manipulation (e.g. querying, modifying, updating) of ontology and annotation bases. CoSWEM enables to carry out automatically or semi-automatically some tasks such as comparison of different ontologies, inconsistency detection and correction of the semantic annotations, etc. This system was also experimented within the framework of the IST project Palette and of the ANR RNTL project e-WOK_HUB with a set of real and evolving data from these projects.

6.1.2. Word sense disambiguation

Keywords: *Natural language Processing*, *Semantic Annotation*.

Participant: Khaled Khelif.

In order to improve our term detection method [62], [25], we proposed a technique to solve the ambiguity problem confronting MeatAnnot results. The main idea of this method is to use the ambiguous word context to decide to which semantic type we can affect it. This context consists of the set of terms which occur with the ambiguous word in the same sentence or in the same paragraph.
So, if MeatAnnot affects several semantic types to the same candidate term, the disambiguation module tries to find the right semantic type. This module computes similarities between the semantic types affected to the ambiguous word and other semantic types affected to the neighbours of the ambiguous word in the text. The semantic type which has the highest similarity is then selected. The calculation of similarity between semantic types is based on Corese semantic distance.
The algorithm was tested on the WSD (Word Sense Disambiguation) test corpus (which is a standard collection for evaluating disambiguation methods [103]) and had a good results.
This work is applied to the Sealife IST project.

6.1.3. Patent Mining

Keywords: *Natural language Processing*, *Semantic Annotation*, *Semantic Web*, *Structured Document*, *Text-mining*.

Participants: Khaled Khelif, Nizar Ghoula, Rose Dieng-Kuntz.

In the context of patent mining (one of the use cases of Sealife IST project), we proposed a domain-independent method for annotation generation on patents [37], [59].

Our approach consists of generating semantic annotations by relying on structuring and building a semantic representation of patent documents. The generated annotation comprises three parts: a structure annotation, a metadata annotation and a domain-based annotation. These different annotations are merged into the so-called Patent Semantic Annotation.

To structure these annotations, we designed and implemented a modular ontology called PatOnto describing the different aspects we took into account : the structure and the content. The domain ontology is an existing biomedical ontology (UMLS) that is used to study the textual content of a patent.

6.1.4. User profile detection

Keywords: *Semantic Annotation*, *Semantic Web*, *User profile*.

Participants: Khaled Khelif, Yassine Mrabet, Rose Dieng-Kuntz.

One of the use cases in the IST project SeaLife consists of linking information on biomedical Websites to appropriate secondary knowledge (existing ontologies/terminologies, RSS feeds). This case study will demonstrate how to provide the user with additional information on resources he/she is viewing on biomedical Websites, using a semantic mapping to appropriate online portals and databases (called targets). In this purpose, the SeaLife browser must recognize the user profile in order to select the appropriate ontology and targets.
In this use case, we proposed a generic and domain-independent approach to detect Web user profiles. Profiles generated can be used for Web browsing personalization, document recommendation, professional activities discovery, etc. We implemented this approach in a system called SUPROD and we tested it on the biomedical domain through the Neli (National Electronic Library of Infection) Web site.
We developed (i) a biomedical profile ontology describing: classification of biologists and doctors, (ii) a generic log analyser, (iii) a profile classifier algorithm and (iv) a profile detector algorithm [42], [65].

6.1.5. Semantic Annotation Generation and Use: GeneRif corpus

Keywords: *Biomedical application*, *Natural language Processing*, *Semantic Annotation*, *Semantic Web*, *Text-mining*.

Participants: Leila Kefi-Khelif, Khaled Khelif, Rose Dieng-Kuntz, Olivier Corby.

We used MeatAnnot [25], an automatic system for the generation of ontology-based semantic annotations to annotate the GeneRIF (Gene Reference Into Function) corpus. GeneRif documents consist of a set of concise phrases, limited to 255 characters in length, describing a function related to a specific gene, supported by at least one PubMed ID. Actually, there are 214354 GeneRifs decribing 36052 genes. For each Generif, we generated a structured semantic annotation, based on three subontologies (1) UMLS (Unified Medical Language System) to describe the biomedical domain, (2) DocOnto to describe metadata about scientific articles, the structure of articles and to link documents to UMLS concepts and (3) GO (Gene Ontology) to add knowledge about the human genes.

We developed a navigation interface which allows the navigation in the Generif corpus. The search module is based on Corese [4]. By using SPARQL (Standard RDF query language implemented by Corese), the interface allows to perform searches and inferences on the annotation base for retrieving relevant information. User can find different information about a gene:
(a) Its synonyms.
(b) GO concepts it is attached to (such as carbohydrate binding, ATPase activity...).
(c) Documents about this gene.
(d) All entities it is associated to in the corpus.

We use the Gene Ontology to find (a) and (b) and the annotations we have made on the Generif documents using MeatAnnot to extract (c) and (d).

The information (d) is presented in a hyper graph which is a more interactive way to present this information. User can click on an edge of the hyper graph to see in which documents the relation between the two entities appears. He can also click on an entity of the graph to have more information about it (generation of a new query). Another graph, "Gene Cart", shows a group of genes in which the user is interested and the interaction between these genes of interest. It can be considered as the result of all the searches the user has done.

A filter will be added to the graphs to allow user to choose which relations and/or which type of entities he would like to store in the graphs. For example, if the user wants to know which entities affect the gene g, he can explicitly ask to hide all the relations that are different from e affects g (using a check box for example).

This work is applied to the ImmunoSearch project in the P.A.S.S (Parfums, Arômes, Senteurs, Saveurs) Competitivity Pole.

6.1.6. Annotation Processing for Earth Sciences

Keywords: *Earth sciences application*, *Natural language Processing*, *Ontology*, *Semantic Annotation*, *Semantic Web*, *Text-mining*.

Participants: Priscille Durville, Rose Dieng-Kuntz.

Some work have been done with e-WOK_HUB project partners in order to identify the significant terms for geological and CO_2 domains from a set of geological articles. One of the objectives of this work is to specify and implement a system that will be able to extract and generate significant annotations from geological and CO_2 articles in a semi-automatic way.

6.1.7. Document repository annotation

Keywords: *Natural language Processing*, *Semantic Annotation*, *Semantic Web*, *Text-mining*, *oNTOLOGY*.

Participants: Hacène Cherfi, Minh Tuan Nguyen, Ouassim Akabbal, Khaled Khelif, Noureddine Mokhtari, Rose Dieng-Kuntz.

We aim at developing a document annotator [49], [50] which allows engineers to:

1. highlight the important concepts and relations contained in the document text with respect to one or several ontologies of the domain;
2. mark-up texts according to given ontologies of the domain with accurate annotations in order to retrieve the text by querying a semantic search engine, namely Corese.

We have extended our Document ontology [51] to the so-called DocumentContents [48] dedicated to support text annotation process. The main modifications consist of separating the document (considered as an object which may contain text, audio, and video materials) from the text contents itself to be analysed. The DocumentContents ontology formalizes both structural and contents information. The original Document ontology is extended based on the document genre (norm, assembly instructions, etc.).

The Natural Language Processing (NLP) analysis of the documents is strongly related to knowledge resources available. In order to feed the SevenPro corporate repository, and to refer to the right document according to the specific engineering activity conducted, a precise study of the end-users ontology requirements is carried out. Additional knowledge resources (named entities, black lists, white lists, etc.) are necessary. We have emphasized the need of dedicated concepts/properties within end-user ontology in order to successfully map the terms/verbs present in the document and the concepts/properties present in end-user ontology, and consequently generate accurate annotations for the document. This task has led us to refine end-user ontologies.
Our annotator being based on GATE, an open-source platform for language technology developed by the University of Sheffield. We developed wrappers of two term extractors, FASTR and ACABIT, so as to integrate them into GATE.

We have tested the annotation generation on sentences coming from end user real-world texts. A number of features are designed for generating correct annotations from different grammatical patterns, including sentences which contain subordinate phrases. Far from being exhaustive, we aim at progressively increasing the complexity of sentences for which the text annotator can extract accurate annotations.
A precise evaluation will follow together with continuing the task of handling complex sentences. In particular, future test phase, with real-world end-user texts will show that capabilities of text annotation are sensitive to the text genre (i.e., the topic of the document). Actually, depending on the kind of document to annotate, one or several information elements are important. For example: assembly and instructions of product usage: verbs matter first, e.g., screw the bolt, open passenger car window, start engine, etc. For standard and norm documents, named entities are important. For test reports, contracts, commercial offers: quantities and figures matter. The first release of the text annotator will be tested during next year by end-users on real-world documents. Finally, we will explore the possibility to annotate Spanish/Italian texts. The process remains identical. However, we have to find out efficient NLP parsers for these two languages.
Last, we proposed a solution based on statistical calculations for automatic generation of semantic annotations from the texts associated to an image.

6.1.8. Extraction and Exploitation of Contextual, Evolving Semantic Annotations for a Virtual Community

Keywords: *Context*, *Natural language Processing*, *Ontology*, *Rhetorical Relationship*, *Semantic Annotation*, *Semantic Web*.

Participants: Noureddine Mokhtari, Rose Dieng-Kuntz.

This work is carried out within the framework of Noureddine Mokhtari's PhD.

The aim of this work is to propose a system of automatic extraction and exploitation of contextual semantic annotations from the text based on semantic web technologies.

Initially, we started by a state of the art on knowledge engineering, semantic Web (RDF/S, ontology) and various uses of context and their definitions. We considered the context of annotation as a semantic and a structural relationship (spatial, temporal and various) between annotations. Then, we concentrated on the approaches of knowledge extraction from texts and we compared some statistical and linguistic tools of natural language processing (NLP) in the aim to propose a method for extraction of semantic annotation (concepts and relationships of a reference domain ontology) from texts. Then, we proposed and we implemented automatic extraction algorithms of text structure to identify contextual structural relationship (successor of, proximity, belonging) between annotations. In addition, we propose to exploit contextual semantic relationships: for example, spatial relations (e.g. under, beside), temporal relations (e.g. since, during) and various complex semantic relations such as rhetorical relationship (e.g. moreover, in fact, so that, however, except).

We made a first experimentation of our approach on concept extraction from text and we obtained good scores (89,82% of precision). In addition, we obtained 75% and 69,71% of precision for respectively identification of contextual semantic relationships and their arguments. The precision corresponds to the rate of correctly extracted entities among the extracted entities.

The originality of this work consists of (a) the integration of the semantic annotations context which gives new ways of reasoning and more information based on both structure and semantic of text, (b) the use of several technologies such as NLP (GATE, JAPE, TreeTagger), Semantic annotations, Knowledge representation and semantic Web (RDF/S, OWL, Ontology, SPARQL, XQuery, Corese) to build a system of automatic extraction and exploitation of contextual annotation from texts.

6.1.9. EMail Mining

Keywords: *Natural language Processing*, *Ontology*, *Semantic Annotation*, *Semantic Web*, *eMail Mining*.

Participants: Bassem Makni, Khaled Khelif, Rose Dieng-Kuntz, Hacène Cherfi.

We proposed a methodology for semi-automatic creation of an ontology along with the subsequent annotation base extracted from a mailing list belonging to a community dedicated to computer assistance. This study raises many original issues which are unusual for NLP techniques because it starts from an email-list corpus. One of the aspects is how to deal with texts which are intended to be informal and generally grammatically erroneous. The challenging annotation extraction process from this email-list will feed frequently asked questions (FAQ) with typical questions and answers according to the domain knowledge.

In order to build the ontology specific to a community of practice exchanging mails about the problems encountered with ICT tools, we applied term extraction tools (ACABIT, FASTR and LIKES) on the corpus of their emails. Then, we designed a modular ontology containing four modules: the OeMail ontology (that describes a MIME message), the Problem Ontology, the Component Ontology designed from the WeboPedia online hierarchy (OntoPedia) and the O'CoP ontology. The Problem ontology was bootstrapped from terms extracted from the mails, using linguistic tools, such terms revealing component problems.

We also studied automatic attachment of new problem terms to the top level of Problem Ontology (Hardware Problem, Software Problem, etc.).

The messages received in the corpus of emails were then automatically annotated with the ontology. To extend the work for future messages of the community members, we developed an application which monitors an email server, detects new messages, and automatically generates their corresponding annotations.

Last, we developed a Web-based GUI allowing the members of the community to navigate through Problem and Component ontology in a hyperbolic manner and to query the semantic search engine Corese to get the mails annotated by the chosen concepts and their replies [40], [67].

6.1.10. Text and Data Mining for Knowledge Management

Keywords: *Data Mining*, *Ontology*, *Semantic Web*, *Text Mining*.

Participant: Martine Collard.

This work is carried out in the framework of Martine Collard's visit at INRIA. It will be based on her previous work in Data Mining, work devoted to model discovery, model evaluation and interpretation. Traditional data mining algorithms are applied on numeric or categorical structured data. But current evolutions tend to extend researches and applications to semi-structured and non-structured data like textual sources and other non-organized sources of explicit or implicit knowledge.

For instance, in biology, numerical expressions of genes in a DNA micro-array experiment cannot be interpreted without paying attention to some heterogeneous sources of information from the biological domain such as ontologies, scientific publications or research results of similar experiments. Among research interests of Edelweiss team, two have a closed relationship with Martine Collard's work:

- on one hand, the objective to provide solutions for knowledge representation, and particularly for ontology management,
- on another hand, the interest in text mining: the participation of Edelweiss in European projects like Sealife or Sevenpro or in the Immunosearch project is partly related to this topics.

As a consequence, Martine Collard's work in Edelweiss aim at studying the integration of data mining, text mining and knowledge management techniques. They are organized according the following objectives:

- extending current Edelweiss solutions on text mining (mainly based on linguistic tools) with a data mining point of view,
- using Edelweiss knowledge management solutions for extending data mining methodologies defined in her team at I3S laboratory.

These objectives are being pursued through collaboration in Sealife and Immunosearch projects.

6.1.11. Semantic Annotation of Usages and Persons

Keywords: *Communities*, *Semantic Web*, *Semantic annotation*, *Social Network*.

Participants: Fabien Gandon, Freddy Limpens.

This work is carried out in the framework of the PhD of Freddy Limpens. It is devoted to study how annotating the varied resources of the Web can help its users better interact. The use of metadata has grown among Web applications, be it in a free manner with tags, or grounded on ontologies. Moreover, the formal status of the Web switched from a global library to a virtual platform of interaction and exchange of different kinds of services and resources. In this context, it is possible to acquire valuable information from the usages, and to represent them through semantically enriched metadata. These metadata can then in turn be integrated and exploited by semantically enabled applications to enhance the global experience of the Web users.
To achieve this objective, we need to investigate the possibilities to obtain semantically rich metadata in an unobtrusive way, and then to find how they can be efficiently exchanged or retrieved by users or other Web applications. A growing number of Web services, labelled Web 2.0, propose their user to attribute a list of arbitrary keywords or tags in order to annotate their resources, thus creating folksonomies. This approach has the great benefit of requiring little efforts, and to rapidly and enthusiastically inject human's intelligence within the overwhelming flow of data constantly aired on the Web [90]. However, it revealed somewhat tough for computers to reason and make valid inferences on these extremely versatile, and most of the time ambiguous, structures of knowledge.
To tackle this problem a number of researchers tried to detect stable semantic structures within folksonomies [91], or to use ontologies to semantically constrain tagging [98]. Gruber [90] also launched a call to action and proposed to build collaboratively an ontology about the act of tagging, to account for all its varied aspects. Another approach is to adapt the technological structures that sustain the activity and exchange of data within communities with original methods to produce semantically rich metadata at the earliest stages of the process of data creation. In this prospect, the preliminary task is to model the interactions between users or group of users and the resources they manipulate. To date, a number of different research works are relevant to this aspect. Breslin et al. [73] proposed to model how data are exchanged among online communities. Thus, the Semantically Interlinked Online Communities(SIOC) framework provides an ontology that describes concepts and relations about online communities[1].

Other approaches focused on the communities of practice, a term coined by Wenger [104] naming groups of persons who share a common interest in a specific matter, and at the same time are aware of this common tie and actively collaborate. Research has been lead to establish appropriate typologies of this type of communities [80] and to provide them with suitable ontologies in order to sustain their activity and annotate their resources [101]. To complement this state of the art, we need to further investigate the cognitive and social phenomenons bound to the interactions with resources on the Web in general. Some contributions heading this direction can be noticed, such as Sinha's work on the process of tagging [99], but some more work needs to be done on evaluating how satisfying and helpful are collaborative tagging tools. For this purpose, a survey could be conducted among a targeted community in order to get feedbacks about the use of such systems.

6.2. Interaction Design

Keywords: *Assistance to the User*, *Co-operation*, *Cognitive Psychology*, *Cognitive Sciences*, *Communication*, *Community of Practice*, *Corporate Memory*, *Corporate Semantic Web*, *Human-machine interaction*, *Knowledge Acquisition*, *Knowledge Engineering*, *Knowledge Management*, *Ontology*, *Virtual Community*.

The objective of this research direction is to study various forms of human interoperability (e.g. search / annotation human interoperability, users' scenarios / developers' scenarios interoperability), so as to specify and to implement corresponding articulating functionalities. We will perform experimentations on different kinds of semantic distances. Moreover, we aim at extending the current interaction toolkit (SeWeSe) with new visualization techniques (e.g. statistical views on ontology-based representations) and new interaction channels (e.g. mails, IRC). We also intend to generalize ontology-based reasoning for smarter interfaces.

6.2.1. ECCO

Keywords: *Cooperation*, *Ontology*, *Semantic Web*.

[1] http://sioc-project.org

Participants: Priscille Durville [resp.], Fabien Gandon, Alain Giboin, Adil El Ghali.

We are implementing a cooperative ontology editor, named ECCO, dedicated to support end-users with different profiles (domain expert, engineer, ontologist, ...) in a cooperative process of ontology construction and evolution. This editor is a new version of the previous OntologyCreator editor and it is based on:

- the first users' feedbacks on OntologyCreator prototype,
- the new users needs that have appeared,
- a more friendly and integrated (relative to the future project's portal) user interface,
- the Semtags library.

6.2.2. SweetWiki

Keywords: *Community*, *Cooperation*, *Ontology*, *Semantic Web*, *Wiki*.

Participants: Michel Buffa [resp.], Guillaume Erétéo, Fabien Gandon, Adil El Ghali, Amira Tifous.

The idea of SweetWiki is to revisit the design rationale of wikis, taking into account the wealth of new standards available for the Web eleven years later to address some of the shortcomings identified through experience. SweetWiki relies on Web standards for the wiki page format (XHTML), for the macros included in pages (JSPX/XML tags), for the semantic annotations (RDFa, RDF), for the ontologies it manipulates (OWL Lite), etc. It improves access to information with faceted navigation, enhanced search tools and awareness capabilities, acquaintance networks identification, etc., [23], [24], [30].

It also provides a single WYSIWYG editor for both metadata and content editing, with assisted annotation tools (auto-completion, checkers for embedded queries or annotations). It comes with an embedded ontology editor and a reasoning engine. Finally it allows metadata to be extracted and exploited by other applications. By semantically annotating the resources of the wiki and by reifying the wiki object model itself, SweetWiki provides reasoning and querying capabilities. All the models are defined in OWL Lite schemata capturing concepts of the wikis (wiki word, wiki page, forward and backward link, author, etc.) and concepts manipulated by the users (users, folksonomy, external ontologies). These ontologies are exploited by an embedded instance of Corese allowing us to support the lifecycle of the wiki, e.g., restructure pages, to propose new functionalities, e.g., semantic search, user-profile-based monitoring and notification, and to allow for extensions, e.g., support for new medias or integration of legacy software.

In SweetWiki we have paid special attention to preserving the essence of a wiki: simplicity and social dimension. Thus SweetWiki supports all the common wiki features such as easy page linking using WikiWords, versioning, etc., but also innovates by integrating a WYSIWYG editor extended to support social tagging functionalities, embedded SPARQL queries etc., masking the OWL-based annotation implementation. Users can freely enter tags and an auto-completion mechanism suggests existing ones by issuing queries to identify existing concepts with compatible labels. Thus tagging is both easy and motivating (real time display of the number of related pages) and concepts are collected in folksonomies. Wiki pages are stored directly in XHTML or in JSPX format, embedding semantic annotations (RDFa and GRDDL) ready to be reused by other software.

6.2.3. Sewese: Semtags and Semservices libraries

Keywords: *Ontology*, *Semantic Web*, *Web service*.

Participants: Priscille Durville, Fabien Gandon [resp.].

Semtags and Semservices are two libraries that allow the use of semantic Web notions and Corese software in Web applications background. Semtags is a set of JSP tags dedicated to semantic Web. This tag library provides Web developers with tools like Corese administration tasks, ontology and annotation management tasks, and tools to send SPARQL queries. This library is written in Java language and uses the JSTL library (for JSP tags part) and Corese software (as the semantic search engine).

Semservices is a set of Web services relying on the same functionalities as Semtags but dedicated to Web services architecture. This set of services is written in Java language and currently deployed with the CXF Web service container (and Apache-Tomcat).

These two libraries are used in a European project (Palette), in an ANR RNTL project (e-WOK_HUB) and in applications developed by the Edelweiss team like CoSWEM.

6.2.4. Designing User-Adapted Semantic Web Applications

Keywords: *Assistance to the User*, *Co-operation*, *Cognitive Psychology*, *Cognitive Sciences*, *Community of Practice*, *Human-machine interaction*, *Scenario*, *Virtual Community*.

Participants: Adil El Ghali, Carine Colombo, Olivier Corby, Rose Dieng-Kuntz, Priscille Durville, Fabien Gandon, Alain Giboin [resp.], Amira Tifous.

Our goals are:

1. To propose methods and models to help the user-oriented design and evaluation of Semantic Web applications; in particular, importing and adapting methods from the Human-Computer Interaction (or software ergonomics) and CSCW communities to the Semantic Web community, esp. the ontology engineering community.
2. To study practices of users and communities of users to inform the design of Semantic Web applications.

6.2.4.1. Collaborative heuristics to evaluate collaborative ontology editors and collaborative Semantic Web applications

"Heuristic evaluation" [96] is a widely-spread method for evaluating the usability of human-computer interfaces. It consists of inspecting, or walking through, a given interface, with some user's scenario or task in mind, to assess if the interface complies a set of usability heuristics (or ergonomic principles) such as "Heuristic 9: Help users recognize, diagnose, and recover from errors". This technique however is more appropriate to evaluate "tools for single users" (or individual software) than to evaluate "collective tools" (or groupware). In other words, this method relies mainly on individual heuristics; it doesn't integrate collective heuristics. To overcome this limitation, Baker, Greenberg and Gutwin [70], [71], developed a preliminary set of groupware-specific heuristics (e.g., "Heuristic 7: Support people with the coordination of their actions"), based on the "mechanics of collaboration", i.e., "the small-scale actions and interactions that group members must perform in order to get a task done in a collaborative fashion".

We started an action aiming at contributing to the development of heuristics for evaluating collaborative ontology editors and, more generally, collaborative Semantic Web applications. This year, as part of the e-Wok HUB project, we used and adapted Baker et al.'s heuristics to evaluate the generic version of the collaborative ontology editor ECCO developed in Edelweiss [53]. Some adaptations were inspired by users' behaviors and comments during a user experiment comparing the use of ECCO and the use of a classical text editor to elaborate the hierarchy of concepts of a small ontology [52].

6.2.4.2. The « interactive document-questionnaire » technique

During the need analysis phase of any system development cycle, it is not always possible for designers/developers and users of the to-be developed system to interact face-to-face or synchronously. It is necessary to use asynchronous or distant communication modes. As part of the e-WOK_HUB project, we developed and used the "interactive document-questionnaire" technique for asynchronously performing the needs analysis of a collaborative ontology editor. The aim of this technique is to allow users and designers/developers to get a clear and shared vision of: (a) the kind of editor to design/develop, (b) the types of users of the editor, (c) the types of ontologies that the users intend to construct, (d) the ontology building process (i.e., tasks and scenarios) that the users want to follow, and (e) the kinds of functionalities and user interfaces that could be expected by the users. The interactive document-questionnaire is iteratively filled by the users in interaction with the designers/developers who can ask users new questions to refine and clarify the points considered as important [46].

6.2.4.3. State-of-the-art of collaborative ontology editors

For the most, existing ontology editors are not collaborative, or (as Protégé and OilEd) they provide a tiny support to the collaborative development of ontologies. As part of the e-WOK_HUB project, we performed a state-of-the-art of existing ontology editors which integrate collaborative aspects, and, to complement this state-of-the-art, we also reviewed other kinds of collaborative tools such as: collaborative editors of knowledge bases, collaborative text/graphic editors, collaborative learning tools, collaborative argumentation tools, and workflow management tools [46]. This state-of-the-art has been used as a need-analysis instrument. It was distributed to the e-WOK pilot users to help them specify the adaptations to be made to the generic version of the ECCO ontology editor, and to indicate the functionalities or interface components they want or do not want to see included in the editor [46]. One of the critical needs identified is that the tool allows a flexible (or not strictly sequential) workflow of ontology construction steps1.

6.2.4.4. Scenario-based design of semantic services and applications

In the framework of the Palette project, we contributed to the elaboration of scenario-based techniques (definition of scenario templates, construction of scenario-based evaluation questionnaires and procedures) for designing and evaluating CoP-dedicated services, especially KM services. We applied these techniques to elaborate and evaluate the Palette CoP scenarios to be supported by the services [54]. Note that these techniques are part of a larger participatory design methodology.

6.2.4.5. Survey of graph-visualization and graph-edition user interfaces

As part of the Color action Griwes, we initiated a survey of graph-visualization and graph-edition user interfaces. The aim of this survey was to help specify the "User Interfaces" level of the graph platform Griwes.

6.2.5. Knowledge Management Services for Communities of Practice

Keywords: *Assistance to the User*, *Co-operation*, *Cognitive Psychology*, *Cognitive Sciences*, *Community of Practice*, *Human-machine interaction*, *Ontology*, *Semantic Annotation*, *Semantic Web*, *Virtual Community*, *Web service*.

Participants: Adil El Ghali, Amira Tifous, Rose Dieng-Kuntz [resp.], Alain Giboin, Bassem Makni, Kalina Yacef, Eleni Kottaki.

This work takes place in the framework of the Palette IST project, aimed at supporting communities of practice (CoPs), by offering them Knowledge Management services.

6.2.5.1. Ontology for Communities of Practice

Participants: Adil El Ghali, Amira Tifous, Rose Dieng-Kuntz, Alain Giboin.

We integrated the different subontologies related to the main concepts of Community, Actor, Competency, Resources so as to constitute the O'CoP ontology dedicated to communities of practice (see http://ns.inria.fr/palette/2007/07/ocop). The method for the collaborative building of the O'CoP ontology, elaborated by Edelweiss, as well as O'CoP main concepts and our experience feedback for each phase of the O'CoP building process are described in the Palette deliverable [66], and synthesized in [45], [44].

The ECCO tool that supports the ontology development method, has been used by some partners as well as some CoPs mediators - for the results' validation phases -. These uses enabled us to collect the users' feedbacks and improve ECCO interface and functionalities.

6.2.5.2. Basic Knowledge Management Web services

Participants: Adil El Ghali, Priscille Durville.

Knowledge Management Services rely on a set of basic operations, that were implemented in the top of the `Sewese` (see section 5.2) library.

These implemented services offer the following functionalities :

- Ontology management: to create, remove and modify ontologies in RDF/S and OWL. http://argentera.inria.fr/semservices/OntologyEdition?wsdl
- Annotation management: to manage annotations http://argentera.inria.fr/semservices/Annotation?wsdl
- Retrieval: to perform semantic search using the SPARQL Query Language, and based on the semantic search engine Corese. http://argentera.inria.fr/semservices/Query?wsdl
- Administration: to factorize some administrative operation needed by all the other services. http://argentera.inria.fr/semservices/Admin?wsdl

The `semservices` was deployed and tested. They can accessed at: http://argentera.inria.fr/semservices. A detailed presentation of the semantic services is included in the deliverable [55].

6.2.5.3. Collaborative knowledge creation services

Participants: Adil El Ghali, Amira Tifous, Alain Giboin, Rose Dieng-Kuntz.

The tool SweetWiki has been deployed for 9 CoPs involved in the Palette project. Some preliminary results and observations on its use by some of these CoPs are synthesized in [36] and described in the Palette deliverable [55].

Additional functionalities for the collaborative work and learning using SweetWiki were identified by Edelweiss in collaboration with the CoPs using it, they mainly concern:

- semantic awareness functionalities: enabling users to subscribe to the set of predefined queries that meet their respective needs, so as to receive notification mails with information about the changes on the wiki content and the statistics that they are interested in.
- and ease of use through the improvement of the tagging mechanisms and tag-based search, and an enhanced tag-management interface (ergonomics, drag & drop mechanism for structuring the ontology, enabled multiple inheritance, etc.).

6.2.5.4. Evaluation services

Participant: Kalina Yacef.

This work was carried out in the framework of Kalina Yacef's visit in the Edelweiss team. Within the Palette project and its Evaluation Service, we have explored how mining the data of CoPs using SweetWiki could support the evolution of this particular Knowledge Service, so that its functionalities and quality remain high at all times and that its maintenance is manageable. As a community-owned Web service, there are needs to maintain the evolving knowledge contained in SweetWiki (in the structural form of Webs, wiki pages, tags and ontologies as well as the more implicit form of thematic networks) as it is dynamic in nature and is subject to becoming redundant and obsolete. There are also needs to monitor the evolution of the community itself and its expertise network in order to provide better user adaptation.

We have identified four broad goals of support where data mining can contribute:

1. Support for semantic annotation and ontology maintenance;
2. Support for building thematic networks within the community;
3. Support for building user profile and identification of community experts;
4. Support for the evaluation of community resources.

To support these goals, we mainly selected the following data mining techniques and visualization techniques: clustering, association rules, frequent sequential pattern mining, standard data exploration techniques (such as histograms, statistics) and specific visualizations (a timeline tree-based visualization, a social network analysis-based visualization).

The analysis using real CoP data will be conducted at the University of Sydney in December 2007-January 2008.

6.2.6. Method Engineering

Keywords: *Community of Practice*, *Method Engineering*.

Participant: Isabelle Mirbel.

Researches in the field of method engineering aim at providing efficient solutions to built, improve and support the evolution of software development processes. Our background is more precisely in the domain of situational method engineering, which focuses on developing, customizing and configuring a situation-specific method from parts of existing methods.

In this research area whose main motivations are adaptability and reuse, we were more specifically interested by usage concerns.

In our previous work, we proposed a framework to combine method building and method configuration. This framework is based on a method chunk repository and a kind of ontology allowing to capitalize contextual information to refine method chunk and reuse situation while building or configuring methods.

In the continuation of this work, our current consideration focuses on means to spread and share knowledge about practices in software development. In other words, our new goal is to transform tacit knowledge about reflection on practice into explicit knowledge in order to support experience-based learning [41].

The concept of *community of practice* sounds very promising in this context as it refers to the process of social learning that occurs when people sharing common interest in some subject or problem collaborate over an extended period to share ideas, find solution and build innovation.

Due to the rise of the Web, several online professional communities dealing with software development have emerged. They are communities of people who organize themselves and interact primarily through the Web for work and knowledge sharing. Open source software communities are typical examples of such communities.

These communities generate huge amount of information as result of their interactions. This information is mostly structured in order to be quickly reused (mailing lists, forums, etc.). There is few means, (like FAQ for instance) to capitalize the information over a longer period of time and to turn it into knowledge (through a semantic FAQ for instance). Moreover, the capitalized knowledge mainly deals with the discussion topics (i.e., the system under development) and few about the members of the community.

In this context, our current goal is to take advantage of the semantic Web and the social Web concepts and technologies to improve the support provided to these communities in terms of long-term knowledge capitalization about the resources and the members of the community.

We plan to work on means to make members expertise, commitments and relationships among them explicit. One of our aim will be for instance to allow members to better understand who knows what in the community in order to improve coordination among members. Another one will be to explicit the expertises present in the community to let the members better understand how to contribute and to encourage them to do it.

We also plan to work on means to summarize and organize the knowledge in order to improve long-term capitalization support. One of our goal will be for instance to provide summary to discussion threads in forum as well as alternative means to organize and present discussion threads.

6.2.7. Ontology-driven Dynamic Course Generation for Web-based Education

Keywords: *Ontology*, *Semantic Annotation*, *Semantic Web*, *e-Learning*.

Participants: Amel Yessad, Rose Dieng-Kuntz.

Actually, the Web is becoming a de-facto standard platform for providing various kinds of educational resources to support teaching in a university or a technical training company. One of the greatest benefits of the Web is that the course material created to support a specific course no longer remains the only educational material that the students can use during the course. Thus, all these resources can be reused and shared in an e-learning context. However, it is very difficult for learners and even for formation responsible to identify the resource relevance to share and reuse it. In order to reduce this problem complexity, we work just with the Web repositories like ARIADNE, MERLOT, etc. which already offer a certain organization of learning resources to facilitate their selection and access.

In general, the learning resources, stored in Web repositories, are first subject to a pedagogy engineering work in order to give them reusable in the context of a particular formation. This activity is time and effort-consuming. In our work, we propose a different approach that consists of moving this engineering effort from the formation responsible/ expert to the software system. It consists of delivering adaptive courses directly and with a minimum human effort. This leads us to have recourse to semantic Web technologies.

We aim to make the formation process most efficient by using the semantic structure of the domain ontology and the resource annotations. So, we offer to the learner adaptive learning paths according to his level of knowledge, goal and time constraints. The learner is assisted to construct a correct representation of the domain knowledge. This representation, called an ACM (Adaptive Cognitive Map), is first calculated by applying filters/rules to the domain ontology and next revised by analyzing, in semi-automatically manner, the learner paths. This analysis can lead to the ontology or/and the resource annotations evolution.

Concretely, we have developed an adaptive learning organizer, called « Organisateur de parcours adaptatifs de formation» (OrPAF). It allows automatic generation of an individualized cognitive map taking into account a specific learning goal, the level of the learner's knowledge and his temporal constraints. Adaptive learning resources are searched from Web repositories like ARIADNE or created locally by our experts and are attached to the generated ACM. The idea of an ACM is to generate dynamically, for each learner, both an individualized course structure and course content by selecting the most optimal learning topics/concepts (knowledge) and materials (presentation, example, test or problem) at any moment. The optimal learning concept and material are selected to bring the learner closest to the ultimate learning goal. This approach is well suited for individual students taking a self-study distance-learning course. These students can be employees in an organization who have different experience and background knowledge, or students in an online university with different backgrounds and goals.

6.3. Knowledge Graph Representation

Keywords: *Conceptual Graphs*, *Corporate Memory*, *Graphs*, *Information Retrieval*, *Knowledge Engineering*, *Knowledge Management*, *Knowledge Representation*, *Knowledge Server*, *OWL*, *Ontology*, *RDF*, *Semantic Web*, *Semantic Web Server*, *Web services*, *XML.*

The goal of this research direction is to propose a framework to develop applications at the knowledge level i.e. a framework where data structures and processing can be designed relying on ontology-oriented models and ontology-based inferences. We want to provide functionalities (e.g. search, clustering, statistics, etc.) independent from the low-level implementation details (storage, distribution, provenance, etc). We aim at developing a family of graph-characteristic inferences for simulating semantic distances used in approximate searching, clustering, and suggesting. We will start abstracting graph structures, indexing and operations and generalize representations and operators.

6.3.1. Corese Semantic Web Engine

Keywords: *Conceptual Graphs*, *Graphs*, *Information Retrieval*, *Knowledge Representation*, *OWL*, *Ontology*, *RDF*, *SPARQL*, *Semantic Web*, *Semantic Web Server*, *Semantic search engine*, *Web services*, *XML.*

Participants: Virginie Bottollier, Olivier Corby [resp.], Fabien Gandon.

Corese relies on the conceptual graph model to implement RDF/S and SPARQL. It is now open source with CeCILL-C free software license.

The Edelweiss team received a second year grant from INRIA for an engineer to participate in the development of Corese. A new version of Corese has been released [2].

6.3.1.1. Path

We have introduced a convention to integrate path matching into SPARQL, following [69], [93]. We have developed the corresponding algorithm with regular expression matching. The query below matches any path between two resources:

```
?x $path ?y
```

The query below matches any path composed of relations that are subproperties of `c:SomeRelation`:

```
?x c:SomeRelation::$path ?y
```

In addition to path match, it is possible to enumerate the relations of the path by applying the SPARQL graph pattern on the path variable.

```
graph $path { ?a ?p ?b }
```

It is possible to restrict the length of paths using the `pathLength()` function.

```
filter( pathLength($path) >= 3 )
```

The path relations can match a regular expression. For example the regular expression below enumerates the elements of a RDF list, i.e. zero or more `rdf:rest` followed by one `rdf:first`.

```
?x $path ?y
filter(match($path,  star(rdf:rest) && rdf:first))
```

Regular expressions are made of property names combined with operators: concatenation, disjunction, negation and iterations.

6.3.1.2. Context

We have proposed a design pattern for contextual metadata. It enables to define hierarchy of states where one state inherits the triples from its ancestors. The design pattern uses a cos:subStateOf transitive relation.

We propose a slight syntactic extension to SPARQL that compiles to standard SPARQL.

```
state ?s  { PATTERN }
```

compiles to

```
graph ?si { PATTERN } ?s cos:subStateOf ?si
```

6.3.1.3. Misc.

We have introduced some extensions to SPARQL:

- Return the result of an assessable expression in the select clause.
  ```
  select function(?x) as ?y where { PATTERN }
  ```
- Group results by same values of variables.
  ```
  select * where { PATTERN }
  group by ?z
  ```

We have extended Corese RDF Rule language with a production rule version that enables function calls in conclusions.

Eventually we have designed an API and a factory for Corese.

6.3.1.4. Formats

In order to speed up RDF/XML loading we have developped a translator from XML/RDF format to NTriple format. Loading in NTriple format is now two times faster.

We have integrated into Corese an RDFa parser that enables to extract RDF triples from XHTML documents under the RDFa convention.

[2] http://www-sop.inria.fr/edelweiss/wiki/wakka.php?wiki=Corese

We have proposed an extension to RDF/XML that enables to specify the URI for named grahs. The convention consists of adding an cos:graph attribute, with an URI as value, in the RDF/XML markup. The inner triples of the tag with cos:graph belong to the URI that is the value of the attribute [35], [56].

```
<c:Person rdf:about='#John'>
  <c:address cos:graph='http://www.inria.fr/edelweiss/personal'>
    Sophia Antipolis</c:address>
  <c:address cos:graph='http://www.inria.fr/edelweiss/office'>
    Valbonne</c:address>
</c:Person>
```

6.3.2. RDFa Parser

Keywords: *RDF*, *RDFa*, *Semantic Web*.

Participant: Fabien Gandon.

In addition to our work on a GRDDL transformation for RDFa we embedded the resulting transformation into a java library to provide a Lightweight RDFa Parser in Java. This parser is available as freeware and under the non viral open-source licence CeCILL-C. It uses Java 1.5 and works only with well-formed XHTML documents allowing us to extract RDF triples embedded in XHTML Web pages. http://www-sop.inria.fr/edelweiss/wiki/wakka.php?wiki=RDFaParser.

6.3.3. RDF expressiveness

Keywords: *RDF*, *Semantic Web*.

Participants: Olivier Corby, Hacène Cherfi.

We have participated in the definition of primitives for expressing some features resulted from engineering activity demand. Using built-in capabilities in RDF(S) and SPARQL we have proposed some easy capabilities to express: order, quantity, units and various metadata on RDF triples and SPARQL result queries [31]. This work was carried out in the framework of the Sevenpro IST project.

6.3.4. Graph-based knowledge representation languages

Keywords: *Conceptual Graphs*, *Graphs*, *Knowledge Representation*, *OWL*, *RDF*, *SPARQL*, *Semantic Web*, *XML*.

Participant: Jean-François Baget.

We have been involved in extending graph-based knowledge representation languages such as conceptual graphs and the RDF/RDFS semantic Web language. A first extension has been the introduction of RDF-like datatypes in conceptual graphs. For added expressivity, we have also introduced computed relations, and have extended the language formal semantics to take into account these new objects [28]. We have studied the effects of different semantic variations on the computational properties of the deduction problem. This language has already been extended to a rules-based language, and can be considered as a first step towards the definition of formal semantics for the SPARQL query language. A second work, in collaboration with the EXMO project (INRIA Rhône-Alpes), has been the extension of both the RDF and the SPARQL query languages with regular expressions, allowing to use paths of unknown length in a query [22].

6.3.5. Knowledge Graph Platform

Keywords: *Conceptual Graphs*, *Graphs*, *Knowledge Representation*, *OWL*, *Ontology*, *RDF*, *SPAR@L*, *Semantic Web*, *XML*.

Participants: Jean-François Baget, Olivier Corby, Rose Dieng-Kuntz, Catherine Faron-Zucker, Fabien Gandon [resp.], Alain Giboin.

This work was performed in the framework of Griwes, a one-year COLOR INRIA project on graph-based representation and reasoning and corresponding programmatic interface with a special emphasis on semantic Web applications. This collaboratiove project between Edelweiss, RCR from LIRMM, and I3S from UNSA aims to design a shared and generic platform for graph-based knowledge representation and reasoning. We are interested in multiple languages of representation, such as conceptual graphs, RDF/S, and in various extensions of these languages.

To abstract and factorize mechanisms common to these various languages will allow the simultaneous development of extensions for all the languages defined on these core abstractions; rather than to develop transformations from one language to another, one will thus be interested in a generic language core. An important challenge here is not to sacrifice the algorithmic efficiency to generic designs. Current results of this project are available online (http://www-sop.inria.fr/acacia/project/griwes/wakka.php?wiki=ColorGriwes) and can be summarized as follows:

(1) a list of motivating scenarios against which we defined the scope of our shared model,

(2) a framework definition distinguishing three layers of abstraction and one transversal component for interaction:

- a structure layer: this layer gathers and defines the basic mathematical structures (e.g. oriented acyclic labelled graph and partial orders) that are used to characterize the primitives for knowledge representation (e.g. type hierarchy);
- a knowledge layer: that factorizes recurrent knowledge representation primitives (e.g. vocabulary base, rule base, fact base, etc.) that can be shared across specific knowledge representation languages (e.g. RDF/S, Conceptual Graphs);
- a language & strategy layer: this two sided layer is still being defined and will gather definitions specific to a language (e.g. RDF triple) and strategies that can be applied to each language (e.g. validation, completion);
- interaction interfaces: this transversal component will gather events (e.g. additional knowledge needed) and reporting capabilities (e.g. validity warning) needed to synchronize conceptual representations and interface representations.

We have proposed a first draft of the two first layers.

6.3.6. Management and Composition of Semantic Web Services

Keywords: *Knowledge Representation*, *Semantic Web*, *Semantic Web services*.

Participant: Mohamed Bennis.

The orchestration Web services architecture in e-WOK_HUB project is based on a semantic annotation of descriptors of services. First, we made a comparative study between the various proposals in the semantic Web service. We chose to describe services with Semantic Annotation Web Service Description Language (SAWSDL) which is a W3C recommendation [3]. In this case we developed a standalone tool about semantic Web service. This tool helps users to add semantic concepts to WSDL file, a standard language of description of Web service. The semantic concepts used are about the elements of Ontology files (RDFS, OWL).

In order to use these semantic capacities added in the description of the Web Service, we specified an approach of deployment of the services by using Corese, a semantic engine. Thus, we developed a XSL transformation of SAWSDL file towards RDFS ontology. This transformation is inspired from two specifications:

- The W3C specification on RDF Mapping of WSDL [4].
- The last specification of Web service description language (WSDL version 2.0) [5].

[3] http://www.w3.org/TR/sawsdl
[4] http://www.w3.org/TR/wsdl20-rdf
[5] http://www.w3.org/TR/wsdl20

This transformation makes possible to extract the data in RDF formal ontology starting from a SAWSDL document. Indeed, our approach is based on GRDDL specification (Gleaning Resource Descriptions from Dialects of Languages [6]). This mechanism introduces markup based on existing standards for declaring that XML document includes data compatible with the Resource Description Framework (RDF).

The final aim of this work is to help users to find out Web Services and build processes by an effective composition of the Web services.

6.3.7. Distributed Memory Servers

Keywords: *Distribution*, *Semantic Web*, *Web services*.

Participants: Cheikh Niang, Fabien Gandon.

The integration of the different sources of an engineering memory requires the development and deployment of semantic servers, holding the Corporate Knowledge Repository (generic ontologies, global models and semantic annotations), dispatching query results to requesting applications, and performing changes to ontologies on request. Such a component provides the infrastructure (API and server) to semantically query the repository of annotations and ontologies, and is accessed by other components for interacting with the knowledge sources. In a first stage we address two facets of this problem:

1. the creation of a generic Web application to provide:
   - a service wrapping the API of the search engine (Corese) as a Web Service
   - a service to initiate / manage a distributed search
   - a service to participate in a distributed search
   - an interface to register / declare peer servers
   - an interface to visualize the query solving process
2. the design of a protocol to propagate queries and results efficiently (a distributed SPARQL over SOAP or HTTP), by relying on: automatic content summary generation for the different bases and query routing based on these summaries. We are currently focusing on the structure and generation of these summaries.

6.3.8. Approximate Matching in an Engineering Memory

Keywords: *Approximate search*, *Corporate memory*, *Semantic Web*, *Semantic distance*.

Participants: Ibrahima Diop, Fabien Gandon.

In order to allow efficient reuse of engineering knowledge from past cases, we are developing approximate search mechanisms to retrieve design cases even if they are not perfect matches but only close enough to provide inspiration in finding engineering solutions [58]. To do so, ontology-based approximation inferences have been proposed and extended. We studied a first metric space: the taxonomy of types with a depth attenuated path-length distance. We identified a number of constraints for this to work properly and in particular we proposed a refinement of the end-user ontologies to avoid mixing different hierarchies (e.g. taxonomy, partonomy). The new taxonomy was tested on some annotations and the resulting schema and annotations were loaded and queried in Corese successfully.

We are proposing and experimenting with several semantic distances for ontology-based approximation. We consider:

- Different metric spaces (ontology graph, annotation graphs, statistics, etc.)
- Different formulas / algorithms for computing the distance
- Different combinations for distances (additive, average, min/max, etc. operators, weighted results etc.)

[6] http://www.w3.org/TR/grddl

We also consider the interface for exploring outputs, suggesting cases for helping with a current case or clustering views for navigating the memory as well as interfaces to capture customization (e.g. weights, relevant relations)

Our approach is in seven steps:

1. tests with current hierarchy-based distance and current data
2. state of the art on existing distances and the tasks they can be used for
3. revision of metric space (ontology) to ensure the quality of its structure
4. test on a real scenario with real data (CAD and other resource annotations)
5. prototype with recurrent queries for typical patterns
6. proposal of different distances
7. prototype of search with customization.

This work is applied to the Sevenpro IST project.

6.3.9. Virtual Reality Interaction based on knowledge-based reasoning

Keywords: *Corporate memory*, *Reasoning*, *Semantic Web*, *Virtual Reality*.

Participants: Emmanuel Jamin, Olivier Corby, Fabien Gandon.

The knowledge, mainly organized in an engineering knowledge base, is static; whereas an engineering system is dynamic. Therefore we need a strategy to reflect the dynamicity of a system by using reasoning with a knowledge-based system.

To solve this sort of interaction problem, an intermediate module had to establish communications between the inference engine (Corese) and the VR module (VRM). In the SevenPro project, this intermediate module is called "VR reasoning module" (VR-RM) [61]. It consists of a user friendly GUI, a number of reasoning methods and predefined VR scenarios and is a rule-based module, i.e. it contains condition specifications that enrich the knowledge base.

Through this intermediate module, Corese is interfaced with the VR scene. From an architecture viewpoint, we developed a Java API to share methods of each module and to have a bidirectional communication between them. One means to state communication is to retrieve object knowledge from the VR scene. To organize these different module communications, the Java API is composed of specific services:

- Query Service: this method set concerns the management of knowledge base and the retrieval process.
- Process Service: it is a method set to get information from current status of the VR scene (for example, the name of VR object that has to be assembled in the next step of assembling procedure).
- Interaction Service: it is a method set to have actions possibilities on the VR scene following reasoning results (for example, result of retrieved VR objects can be displayed directly in the VR scene).

Based on this API, a Corese GUI is implemented to edit, display results of modules communication and results of reasoning process.

In parallel to the communication service modules, reasoning methods are used to take benefits from the knowledge base according to VR user actions. One of these methods is Information retrieval based. The first method is a query-based reasoning approach. The RDF graph is a set of triples (resource, property and value). A triple can be related to many other ones in the RDF graph. These triples could be issued from different data files based on different RDFS ontologies. The goal is to scrupulously search the graph by means of the triples to obtain inferences about implicit knowledge.

The second approach is rule-based reasoning. The rule mechanism enriches the expressiveness of the knowledge base. Adding some kinds of relations (such as transitivity, symmetry, etc.) to the graph, for example, could be relevant. The principle of the rule-based mechanism is based on defining graph conditions under which the graph has to be transformed as well as defining the graph transformation content. The rule formalism is an XML format which includes SPARQL parts in two specific tags, e.g. <IF> for condition and <THEN> for transformation content.

A user mode for the co-action of semantic technology and virtual reality was developed to facilitate enhanced interaction/representation of product information/knowledge in a VR scene. It is implemented and tested. The answer time for the retrieval of so-called supplementary data is satisfactory. Control of object behavior in a VR scenario by the VR reasoning module is in its initial phases to state whether, or not, it performs accurately.

The application of different reasoning approaches shows an improvement in knowledge reuse and sharing. The reasoning transforms implicit knowledge to usable information for different applications. In the context of virtual reality, the dynamic aspect of reasoning is in its infancy and will need further research study.

6.3.10. Semantic Server Agent integration

Keywords: *Corporate memory*, *SPARQL*, *Semantic Web*, *Virtual Reality*, *Web service*.

Participants: Emmanuel Jamin, Fabien Gandon.

The main objective of the Semantic Server Agent (SSA) integration is to provide extended functionalities of retrieval to assist virtual reality (VR) uses. For that, Corese has to become a knowledge engine for knowledge integration. SevenPro project has participated in the definition of SPARQL queries to the Semantic Server Agent (SSA) for different modules of SevenPro (mainly VR and Engineering memory tool). Semantic-systems (Spain) provides a heterogeneous knowledge annotation repository (from documentations, industrial knowledge bases, etc.) which can be retrieved by Web Services. In order to implement this integration, Corese is considered as a mediation search engine between the VR module and the heterogeneous knowledge repositories. Hence, it can receive a SPARQL query from the VRM, ask the SSA to retrieve efficient knowledge (RDF triples), load this specific knowledge and return a generic result for the VRM. One of the difficulties is to specify appropriate queries according to user needs.

7. Contracts and Grants with Industry

7.1. KMP

Keywords: *Ontology*, *Semantic Web*, *Skills Management*.

Participants: Olivier Corby (co-resp), Priscille Durville, Fabien Gandon, Alain Giboin (co-resp).

The KmP (Knowledge Management Platform) projects: The KmP projects are a set of pluridisciplinary and user participatory projects aiming at designing systems for managing collective and individual competencies, which resulted in a prototype platform based on Corese and Sewese.

The KmP-2 project (Managing collective competencies). If this one-year project, aimed at pre-industrializing the KmP prototype platform, ended last year for the former Acacia team (now the Edelweiss team), it was still going on for the Telecom Valley Association, one of the partners with whom we elaborated the KmP prototype platform. This year, Telecom Valley officially launched the new KmP platform (http://kmp.telecom-valley.fr/). This year too, a KmP co-property agreement was signed by INRIA, CNRS and TelecomValley to allow the Telecom Valley association to continue the work done and to build on this work.

The KM2 project (Managing individual and collective competencies). This two-year project funded by NXP (formely Philips Semiconductors) aimed at designing and validating a prototype supporting the strategic management of individual and collective competencies within a firm, namely NXP. 2007 was the second year of the project. During this year, Edelweiss acted as a consultant, and collaborated with the NXP engineer in charge of making the KM2 prototype operational in the firm.

8. Other Grants and Activities

8.1. Regional Actions

8.1.1. Laboratory of Usages at Sophia Antipolis

Participant: Alain Giboin.

We take part in the Laboratoire des usages (Laboratory of Usages) of Sophia Antipolis, Group "Usages" (leaded by Marc Relieu, ENST) and in the Association Use Age (Sophia Antipolis) (http://www.use-age.org/).

8.1.2. Competitivity Poles

- The e-WOK_HUB ANR RNTL project has been labelled by the Solutions Communicantes Sécurisées (SCS) competitivity pole.
- The Immunosearch project has been labelled by the PASS competitivity pole.

8.1.3. Griwes

Keywords: *Conceptual Graphs*, *Graphs*, *Knowledge Representation*, *OWL*, *Ontology*, *RDF*, *SPAR@L*, *Semantic Web*, *XML*.

Participants: Jean-François Baget, Olivier Corby, Rose Dieng-Kuntz, Catherine Faron-Zucker, Fabien Gandon [resp.], Alain Giboin.

Griwes is a one-year COLOR INRIA project on graph-based representation and reasoning and corresponding programmatic interface with a special emphasis on semantic Web applications. It is a collaboration between Edelweiss, RCR from LIRMM, and I3S from UNSA to design a shared and generic platform for graph-based knowledge representation and reasoning. We are interested in multiple languages of representation, such as conceptual graphs, RDF/S, and in various extensions of these languages. More details on the work performed can be found at 6.3.5.

8.2. National Actions

8.2.1. ANR RNTL project e-WOK_HUB

Keywords: *Evolution*, *Ontology*, *Semantic Annotation*, *Semantic Web*, *Web service*.

Participants: Mohamed Bennis, Olivier Corby, Rose Dieng-Kuntz [resp.], Priscille Durville, Fabien Gandon, Alain Giboin.

e-WOK_HUB is a 3-years ANR RNTL project, coordinated by the Edelweiss team, with IFP, BRGM, EADS, ENSMP, ENSMA and CRITT as partners. e-WOK_HUB aims at building a set of communicating portals (the e-WOK Hubs), offering both: (a) Web applications accessible to end-users through online interfaces, and (b) Web services accessible to applications through programmatic interfaces. As applicative objectives, e-WOK_HUB aims at enabling management of the memory of several projects on CO_2 capture and storage, with use of results of technological watch on the domain.

Edelweiss is responsible for WP2 on Generic Tools and Services [46] and worked on Support to ontology creation (see section 6.2.1), Management and composition of semantic Web Services (see section 6.3.6) and Annotation processing (see section 6.1.6).

8.2.2. ImmunoSearch

Keywords: *Biomedical Application*, *Natural Language Processing*, *Ontology*, *Semantic Annotation*, *Semantic Web*.

Participants: Leila Kefi-Khelif, Rose Dieng-Kuntz, Olivier Corby, Khaled Khelif.

This work is carried out in the framework of the P.A.S.S (Parfums, Arômes, Senteurs, Saveurs) Hub. It is a collaborative project with ImmunoSearch SARL, Institute of Molecular and Cellular Pharmacology (IPMC-CNRS/UNSA), I3S (UMR CNRS-UNSA) and industrialists (IRISPHARMA and Skinethic, the perfumers, l'Oréal, etc.)

The objective of ImmunoSearch is to design biomarkers for controlling the harmlessness of the molecules used in perfumes, aromatics and cosmetics. The purpose of this research is to conduct comparative studies of in vivo and in vitro test models on the skin (irritation, allergy) and to propose alternative methods defining the new norms applicable in this field. In this context, we aim at proposing methodological and software support for capitalization and valorization of knowledge resulting from experiments and techniques to preserve and reuse data. We rely on the semantic Web technologies (semantic annotations, ontologies, RDF, SPARQL...). More details on the work performed can be found at section 6.1.5.

8.2.3. Working Groups

Members of the Edelweiss team take part in several working groups:

- Rose Dieng-Kuntz is member of:
  - the board of the GRACQ (*Groupe de Recherche en Acquisition des Connaissances*) http://www.irit.fr/GRACQ.
  - the TIA Group (*Terminology and AI*) http://tia.loria.fr.
- Alain Giboin is member of:
  - Group « Psychologie ergonomique » of the Département Recherche de la Société française de Psychologie (SFP). Founder member. Secretary of the group (until September 2007). Webmaster of the group Website (http://www-sop.inria.fr/acacia/gtpe/) (until June 2007). In collaboration with Aline Chevalier (Université Paris X), the new Webmaster: transfer and integration of the group website to the Website of the SFP (http://sfpsy.org/spe-grape/).
  - Research Network "Psycho Ergo". Member of the Conseil de groupement. Coordinator (with Pascal Salembier, IRIT Toulouse, UTT Troyes) of the Thematic Group "Coopération homme-machine et Coopération homme-homme".

8.3. European Actions

8.3.1. Knowledge Web

Keywords: *Evolution*, *Ontology*, *Ontology Alignment*, *Semantic Annotation*, *Semantic Web*.

Participants: Olivier Corby, Sylvain Dehors, Rose Dieng-Kuntz (resp.), Fabien Gandon, Alain Giboin, Phuc-Hiep Luong.

We take part in the Knowledge Web Network of Excellence. Our work concerned the workpackages WP2.2 Heterogeneity (cf.ontology alignment [27]), WP2.3 Dynamics (cf. ontology evolution [26], [39], and WP3.2 and WP3.3 on Education (cf. semantic Web and e-learning [20]).

8.3.2. Palette

Keywords: *Assistance to the User*, *Co-operation*, *Collaboration*, *Community of practice*, *Ergonomics*, *Interaction Design*, *Knowledge Engineering*, *Knowledge Management*, *Ontology*, *Scenarios*, *Semantic Annotation*, *Semantic Web*, *User interfaces*, *Virtual community*, *Web Service*.

Participants: Olivier Corby, Rose Dieng-Kuntz [resp.], Adil El Ghali, Fabien Gandon, Alain Giboin, Eleni Kottaki, Bassem Makni, Amira Tifous, Kalina Yacef.

Palette is a 3-years long integrated project, coordinated by ERCIM and EPFL, with as partners the University of Fribourg, CTI (Greece), Centre de Recherche Public Henri Tudor (Luxembourg), University Abou Bekr Belkaid (Algeria), University of Liège, EM Lyon, Groupe d'Analyse et de Théorie Economique (GATE CNRS), Center for Study of Education and Training (CSET) (Lancaster), ePrep, Nisai, MindOnSite - Integral Coaching SA, LICEF Téluq (Canada), INRIA.

The Palette project aims at facilitating and augmenting individual and organisational learning in Communities of Practice (CoPs). Towards this aim, an interoperable and extensible set of innovative services as well as a set of specific scenarios of use will be designed, implemented and thoroughly validated in CoPs of diverse contexts. Palette will thus offer information services, knowledge management services (based on an ontology dedicated to communities of practice) and mediation services for communities of practice (CoPs). Eleven pilot CoPs are involved in the participatory design of Palette services. These CoPs, located in various European countries (Belgium, France, Greece, Switzerland, UK), belong to three different domains: (i) teaching, (ii) management, and (iii) engineering. Their size varies from less than ten members to more than a hundred of members.

Edelweiss is leader of the WP3 aimed at designing ontologies and ontology-based Services for Knowledge Management in Communities of Practice.

In 2007, we focused on:

- Construction of the O'CoP ontology (see section 6.2.5.1),
- Scenario-based design of semantic services (see section 6.2.4.4),
- Basic Knowledge Management Web services [55] (see section 6.2.5.2),
- Collaborative knowledge creation services (see section 6.2.5.3),
- Semi-automatic annotation and semantic retrieval/dissemination services based on EMail Mining (see section 6.1.9).

We were editors of two deliverables [66], [55] and contributed to [54].

8.3.3. SeaLife

Keywords: *Biology*, *Life Sciences*, *Medicine*, *Natural Language Processing*, *Ontology*, *Patent mining*, *Semantic Web*, *Semantic annotations*, *Text-mining*, *Web services*.

Participants: Olivier Corby, Rose Dieng-Kuntz [resp.], Fabien Gandon, Nizar Ghoula, Khaled Khelif, Yassine Mrabet.

SeaLife is a 3 year-long STREPS project, coordinated by Dresden University, with Edinburgh University, London College, Manchester University, Scionics as other partners; it started on April 2006.

The objective of SeaLife is the design and realization of a semantic Grid browser for the Life Sciences, which will link the existing Web to the currently emerging eScience infrastructure. The SeaLife browser will allow users to automatically link a host of Web servers and WebGrid services to the Web content they are visiting. This will be accomplished using eScience growing number of WebGrid Services and its XML-based standards and ontologies. The browser will identify terms in the pages being browsed through the background knowledge held in ontologies. Through the use of Semantic Hyperlinks, which link identified ontology terms to servers and services, the SeaLife browser will offer a new dimension of context-based information integration.

This SeaLife browser [60] will be demonstrated within three application scenarios in evidence-based medicine, literature and patent mining, and molecular biology, all relating to the study of infectious diseases. The three applications vertically integrate the molecule/cell, the tissue/organ and the patient/population level by covering the analysis of high-throughput screening data for endocytosis (the molecular entry pathway into the cell), the expression of proteins in the spatial context of tissue and organs, and a high-level library on infectious diseases designed for clinicians and their patients.

In this project we take part in 6 among the 7 work packages and we are coordinator of the text mining and natural language processing work package [63], [64]. We worked on:

- Word sense disambiguation (see section 6.1.2),
- Patent Mining (cf. the use case called 'literature and patent mining') (see section 6.1.3),
- User profile detection (see section 6.1.4),
- Analysis of gene interactions relying on information extracted automatically from GeneRif annotations.
- Development and specification of Web Services to publish our techniques (patent search, profile detection and textmining) [47].
- Improvement of our method of annotation generation from text and integration of external sources (Uniprot) to enrich annotations.
- Taking into account the context in semantic annotation generation process.

8.3.4. SevenPro

Keywords: *Corporate Memory*, *Natural Language Processing*, *Ontology Design for Products*, *Reasoning Engine*, *Semantic Annotation*, *Semantic Web*, *Virtual Reality*.

Participants: Ouassim Akabbal, Hacène Cherfi, Olivier Corby, Rose Dieng-Kuntz [resp.], Ibrahima Diop, Fabien Gandon, Alain Giboin, Emmanuel Jamin, Minh-Tuan Nguyen, Cheikh Niang.

SevenPro (Semantic Virtual Engineering Environment for Product Design) is a European STREPS project. The SevenPro project develops technologies and tools supporting deep mining of product engineering knowledge from multimedia repositories and enabling semantically enhanced 3D virtual reality (VR) interaction with product knowledge in integrated engineering environments. It aims at helping an engineer to design new objects by providing a 3D viewing of the object designed, informations on each part of the object (suggestions of other objects with similar or close properties could be performed) and information about repetitive design processes.

SevenPro is coordinated by Semantic Systems (Spain), and involves partners from industrial and academic areas. SevenPro project is carried out during 34 months, starting from January 2006, by a consortium composed of partners from five different EU countries:

1. Three research centres: Edelweiss (France), Technical university of Prague (Czech Republic), Fraunhofer institute (Germany).
2. Two IT companies (Semantic Systems (Spain), and LivingSolids (Germany).
3. Two users: Estanda (Spain) and ItalDesign (Italy).

The components and the whole engineering environment that are to be developed will be integrated and tested continuously in an iterative and incremental strategy. Edelweiss is leader of WP02 on Knowledge Engineering and WP04 on Semantic Annotation of Corporate Repositories.

Edelweiss is WP leader on the Document repository annotation (WP04) [50], [49] (see section 6.1.7). We also worked on:

- Virtual reality (VR) interaction based on knowledge reasoning [61] (see section 6.3.9),
- Engineering memory tool (see sections 6.3.7, 6.3.8) [58]
- Semantic server agent (see section 6.3.10).

8.4. International Actions

8.4.1. W3C

Participant: Fabien Gandon.

We participate to several W3C working groups and interest groups. We are respectively editors and co-authors of two documents of the GRDDL working groups (a mechanism to extract RDF from XML dialects):

- GRDDL Use Cases: Scenarios of extracting RDF data from XML documents, W3C Working Group Note 6 April 2007, [57].
- GRDDL Primer, W3C Working Group Note 28 June 2007.

And we participated in discussions on the specifications: Gleaning Resource Descriptions from Dialects of Languages (GRDDL) W3C Recommendation 11 September 2007

We also contributed to the RDFa working group providing:

- A GRDDL profile for RDFa available on the namespace server of INRIA ns.inria.fr;
- Tests reports on the latest specs;
- A GRDDL experiment for several microformats.

Finally we participated to two face-to-face meetings : Boston, Plenary from the 21st to the 27th of January ; Boston, Plenary from the 2nd to the 8th of November.

9. Dissemination

9.1. Animation of the Scientific Community

9.1.1. Program committees

Michel Buffa was member of the ACM Wikisym 2007 program committee conference.

Rose Dieng-Kuntz was member of the following program committees:

- 7ème édition de la conférence francophone Extraction et Gestion des Connaissances(EGC'2007), Namur, Belgique, January 23 - 26, 2007
- IC'2007 (18èmes Journées Francophones d'Ingénierie des Connaissances), July 2-6, 2007, Grenoble, AFIA Platform
- TIA'2007, Terminologie et Intelligence Artificielle, held at INRIA Sophia Antipolis, October 2007, [18].
- IJCAI'2007 Workshop on Knowledge Management and Organizational Memories held at Hyderabad, India, January 2007 [19].
- 14th ISPE Conference on Concurrent Engineering, São José dos Campos, Brazil, June 2007.
- KWEPSY2007, Knowledge Web PhD Symposium 2007, co-located with the 4th European Semantic Web Conference (ESWC'2007) in Innsbruck, Austria.
- 2nd International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice (TEL-CoPs'07), Crete, Greece, September 2007,
- Second International Conference on Knowledge Science, Engineering and Management, November 28-30, 2007, Melbourne, Australia.
- NETTAB'2007, A Semantic Web for Bioinformatics: Goals, Tools, Systems, Applications, June 12-15, 2007, University of Pisa, Italy.
- K-CAP'07 Workshop on Knowledge Management and Semantic Web for Engineering Design (KW4ED), Whistler, British Columbia, Canada, November 2007.
- SAAKM2007, Semantic Authoring, Annotation and Knowledge Markup Workshop, in conjunction with K-CAP 2007, Whistler, BC, Canada, October 28-31, 2007.

She is member of the steering committee of the EKAW, COOP and PAKM conferences.

Fabien Gandon was member of the program committees or reviewer for:

- European Semantic Web Conference (ESWC2007), IEEE/WIC/ACM Web Intelligence (WI2007), Ingénierie des Connaissances (IC2007), ACM Symposium on Applied Computing (SAC2007)
- Mobile Services and Ontologies (MOSO2007); K-CAP'07 Workshop on Knowledge Management and Semantic Web for Engineering Design (KW4ED); Knowledge Web PhD Symposium (KWEPSY2007); Ontologies and Information Systems for the Semantic Web (ONISW2007); Web Semantics 2007.

Alain Giboin was member of the following programme committees:

- DeViNT'2007, Cinquième journée "Déficients visuels et NTIC", Sophia Antipolis, France, 2007, 31 mai 2007, Website: http://devint.polytech.unice.fr/
- ÉPIQUE'2007, Quatrièmes journées d'étude en Psychologie ergonomique, September 2007, Nantes, France, Website : http://www.sfpsy.org/spip.php?article97.

Alain Giboin is member of the steering committee of COOP conferences.

9.1.2. Journals and Publishers

Rose Dieng-Kuntz is:

- Co-editor of the Series *Frontiers in Artifical Intelligence and applications*, at IOS Press,
- Member of Editorial Board of *Revue d'Intelligence Artificielle*.

and was reviewer for Journal of Universal Computer Science.

Fabien Gandon is:

- Editorial Board Member of *Electronic Commerce Research and Applications* journal, Elsevier.
- Reviewer for IEEE Internet Computing, IEEE Transactions on Knowledge and Data Engineering, IEEE Computer Magazine, IEEE Transaction on Multimedia, World Wide Web Journal, IJSWIS special issue for MOSO, Integrated Computer-Aided Engineering Journal, Journal of Universal Computer Science,

Alain Giboin is member of the expert committee of the review *Le Travail Humain*.

9.2. Organization of conferences and courses

- Rose Dieng-Kuntz was:
  - Organizer of TIA'2007, Terminologie et Intelligence Artificielle, held at INRIA Sophia Antipolis, October 2007, [18].
  - Co-organizer, with Nada Matta, of IJCAI'2007 Workshop on Knowledge Management and Organizational Memories held at Hyderabad, India, January 2007 [19].
- Hacène Cherfi and Rose Dieng-Kuntz organized *KW4ED: K-CAP'07 Workshop on Knowledge Management and Semantic Web for Engineering Design* [17], in conjunction with KCAP 2007 (*the Fourth International Conference on Knowledge Capture*), Whistler, British Columbia, Canada, November 2007. This workshop (see http://www-sop.inria.fr/acacia/WORKSHOPS/KW4ED-K-CAP2007), supported by the SevenPro IST project, enabled a liaison with the Acemedia IST project.
- Rose-Dieng Kuntz, Adil El Ghali, Alain Giboin and Amira Tifous organized the *Palette Training on Knowledge Management*, http://argentera.inria.fr/swikipalette/data/Main/KmTraining.jsp , on November 20-21, 2007, at INRIA Sophia Antipolis.
- Alain Giboin was:

- Co-organizer (with Pascal Salembier, IRIT Toulouse) of the Thematic Session "Intelligibilité mutuelle", during ÉPIQUE'2007, Quatrièmes journées d'étude en Psychologie ergonomique, September 2007, Nantes, France.
- Member of the organizing committee of DeViNT'2007, Cinquième journée "Déficients visuels et NTIC", Sophia Antipolis, France, 2007, Website: http://devint.polytech.unice.fr/
- Member of the organizing committee of the WUD Sophia 2007, the World-Usability-Day-labelled event organized in Sophia Antipolis (France) by the Use Age Association and INRIA, November 29, 2007, http://www.worldusabilityday.org/event/show/389. The WUD is an annual series of events founded by the Usability Professionals' Association (USA) to promote usability studies, and "to ensure that the services and products important to life are easier to access and simpler to use" (http://www.worldusabilityday.org/).

9.3. Others

9.3.1. Scientific Councils and Evaluation tasks

Michel Buffa is

- member of the scientific council of the CNRT Télius,
- member of the Commission des spécialistes for the 27th section, at UNSA.

Rose Dieng-Kuntz is member of:

- International Advisory Board of Cooperation Unit of EPFL,
- Fondation C.Genial for Scientific and Technical Culture.

Fabien Gandon was referee for ANR 2007 in MDCO program.

9.3.2. International Working Groups

Rose Dieng-Kuntz is chair of the IFIP Working Group on Knowledge Management.

9.3.3. Collective tasks

- Olivier Corby is member of: CDL (Commission for software development) at INRIA Sophia Antipolis.
- Fabien Gandon:
  - Secretary of INRIA Sophia Antipolis Project Committee until July 2007
  - Member of commissions: CCC, CSD.
  - Member of working groups GT Télécom and GT PDA Phones.
- Alain Giboin is
  - Member of the Cumir (Commission des Utilisateurs des Moyens Informatiques pour la Recherche). Coordinator, with Nicolas Tsingos, of a survey about the "Obsolescence du parc informatique de l'UR INRIA Sophia Antipolis". Writing of the final report of this survey. Coordinator, with Elisabeth Verplanken, of the Working Group "Hotline", aimed at specifying the Hotline aspects of the future "Contrat de service Semir".
  - Member of the Comorale group, aimed at analyzing the "Communication interpersonnelle à l'INRIA Sophia Antipolis", and at providing recommendations to improve this communication. Contribution to the data analysis and to the writing of the final report.
  - Member of the jury of four external competitive examinations to recruit four INRIA Financial Services' Agents.

9.3.4. Visits

The Edelweiss team welcomed:

- Rubens Pereira from Brazil,
- Emanuel Pietriga from In-Situ project-team (INRIA-Futurs),
- Stuart Madnick from MIT.

9.4. Teaching

9.4.1. University

The Edelweiss project is a welcoming team of the *École doctorale STIC of the Nice - Sophia Antipolis University (UNSA)*.

The members of the project gave the following courses:

- Olivier Corby, Catherine Faron-Zucker, Fabien Gandon and Alain Giboin are in charge of a course on Knowledge Engineering & Semantic Web. It is a one semester course during the last year of the curriculum at EPU (Ecole Polytechnique Universitaire de Nice - Sophia Antipolis, Master 5), 45 hours. They also supervised several student projects.
- Olivier Corby, Catherine Faron-Zucker and Fabien Gandon taught a course on Semantic Web at the Licence professionnelle, IUT, Nice Sophia Antipolis.
- Olivier Corby and Catherine Faron-Zucker taught a course on Semantic Web and Description Logics in a Research Master at UNSA.
- Rose Dieng-Kuntz is responsible for:
  - the course on *Knowledge Capitalization and Economic Intelligence* (20h) in the framework of the Masters “Audit Informationnel et Stratégique” at the Institut d’Administration d’Entreprises, UNSA,
  - the course on *Knowledge Management* at ENSI, Tunis.
- Fabien Gandon
  - Two weeks of courses and seminars at Gaston Berger University in St Louis (Senegal) from the 31st of January to the 14 of February on Web technologies with an emphasis on semantic Web frameworks.
  - Ingénierie des Connaissances et Web Sémantique, Master Pro Course, Polytech’Nice Master Pro.
  - Design and maintenance of online tutorials on Corese and Sewese.
- Alain Giboin gave the following courses:
  - EPU, ESSI 3rd year, Module "Interfaces graphiques homme-machine" (GUI), Université de Nice Sophia Antipolis (http://www.essi.fr/~pinna/MODULEIHM/): contribution to the organization of the module, lectures, participation to tutorials, and assessment of students’ GUI projects. Supervising the collaboration between the EPU-SI students and the ergonomics students who contribute also to the GUI projects.
  - Master "Ergonomie des Nouvelles Technologies de l’Information et de la Communication (ErgoNTIC)", Université de Nice Sophia Antipolis (http://www.unice.fr/master-ErgoNTIC/): in charge of the « EPU Project » with Anne-Marie Pinna-Déry (EPU-SI). The EPU project allows ergonomics students from ErgoNTIC and software engineering students from the EPU-SI to work together as early as the learning phase, so preparing future collaborations when working together in enterprise.
  - EPU (With Olivier Corby, Catherine Faron-Zucker, and Fabien Gandon) ESSI, Knowledge engineering applied to the Semantic Web course": lectures and tutoring of students projects.
- Noureddine Mokhtari : Practical courses on Programming and Web Security at University of Nice-Sophia Antipolis, Master 3, 24 hours.

9.4.2. Theses

Defended theses:

1. Sylvain Dehors: *Semantic Web and Knowledge Management for E-learning*, université de Nice - Sophia Antipolis,

Current theses:

1. Phuc-Hiep Luong: *Management of a Corporate Semantic Web Evolution*, Ecole Nationale Supérieure des Mines de Paris,
2. Noureddine Mokhtari: *Extraction and exploitation of contextual, evolving semantic annotations for a virtual community*, université de Nice - Sophia Antipolis,
3. Freddy Limpens: *Semantic Annotation of Usages and Persons*, university of Nice-Sophia Antipolis,
4. Guillaume Erétéo: *Detection of emerging communuties of practice and support to their life cycle through semantic annotation of activities, uses and persons*

Olivier Corby was reviewer of the PhD thesis of Stéphane Jean from University of Poitiers on *OntoQL, un langage d'exploitation des bases de données à base ontologique*, december 2007.

9.4.3. Training

We welcomed the following trainees:

- Ouassim Akabbal from Polytech'Nice-Sophia Antipolis:*Towards an automatic detection of spatiotemporal relations*,
- Carine Colombo from Polytech'Nice-Sophia Antipolis : *Evaluation of collaborative ontology editors*,
- Ibrahima Diop from University Saint-Louis, Senegal : *Approximate Matching in an Engineering Memory*,
- Cheikh Niang from University Saint-Louis, Senegal: *Distributed Memory Servers*,
- Bassem Makni from ENSI, Tunisie: *Semi-automatic Creation of an Ontology and of Semantic Annotations from a Discussion Forum of a Community of Practice*,
- Nizar Ghoula from ENSI, Tunisie: *Patent Mining for Generation of Semantic Annotations*,
- Yassine Mrabet from ENSI, Tunisie: *Personalization of a Biomedical Website Browsing through User Profile Learning*,
- Minh Tuan Nguyen from IFI, Vietnam: *Towards a generic platform for automatic semantic annotations of multimedia documents*.

9.5. Participation to conferences, seminars, invitations

Members of the team took part in conferences and *workshops* (see the bibliography). In addition to these conferences,

Olivier Corby was invited speaker at *Journées sur les mesures de similarité sémantique*, 21st May, ERIC laboratory, University Lyon 2.

Fabien Gandon was:

- Invited speaker of the W3C Track of WWW 2007 on "Advances in Semantic Web",
- Invited speaker at *Journées sur les mesures de similarité sémantique*, 21st May, ERIC laboratory, University Lyon 2,
- Invited speaker at C2EI CRCM 2007, «Capitalisation et Réutilisation des Connaissances Métier en conception de systèmes mécaniques», 24th of May 2007, Montbéliard (UTBM).

Alain Giboin gave a talk "Gestion de référentiels communs dans les activités coopératives : cas du projet Palette" on January 20, 2007, to the Laboratory of Usage, Sophia Antipolis.

10. Bibliography

Major publications by the team in recent years

[1] R. DIENG-KUNTZ, O. CORBY (editors). *Knowledge Engineering and Knowledge Management: Methods, Models and Tools, Proc. of the 12th International Conference, EKAW'2000*, Springer-Verlag, LNAI n.1937, Juan-les-Pins, October 2 -6 2000, http://www.inria.fr/acacia/ekaw2000.

[2] O. CORBY, R. DIENG-KUNTZ. *The WebCokace Knowledge Server*, in "IEEE Internet Computing", vol. 3, n^{o} 6, December 1999, p. 38-43.

[3] O. CORBY, R. DIENG-KUNTZ, C. FARON-ZUCKER. *Querying the Semantic Web with Corese Search Engine*, in "Proc. of the 16th European Conference on Artificial Intelligence (ECAI'2004), Prestigious Applications of Intelligent Systems, Valencia, Spain", R. LOPEZ DE MANTARAS, L. SAITTA (editors), August 22-27 2004, p. 705-709.

[4] O. CORBY, R. DIENG-KUNTZ, C. FARON-ZUCKER, F. GANDON. *Searching the Semantic Web: Approximate Query Processing based on Ontologies*, in "IEEE Intelligent Systems & their Applications", vol. 21, n^{o} 1, January-February 2006, p. 20-27.

[5] O. CORBY, R. DIENG-KUNTZ, C. HEBERT. *A Conceptual Graph Model for W3C Resource Description Framework*, in "Conceptual Structures: Theory, Tools and Applications, Proc. of the 8th Int. Conference on Conceptual Structures (ICCS'2000), Darmstadt, Allemagne", B. GANTER, G. W. MINEAU (editors), Springer-Verlag, LNAI n. 1867, August 13 -17 2000, p. 468-482.

[6] A. DELTEIL, C. FARON-ZUCKER. *A Graph-Based Knowledge Representation Language*, in "Proceedings of the 15th European Conference on Artificial Intelligence (ECAI 2002), Brighton, Lyon, France", F. VAN HARMELEN (editor), IOS Press, July 21- 26 2002, p. 297-301.

[7] R. DIENG-KUNTZ, O. CORBY, F. GANDON, A. GIBOIN, J. GOLEBIOWSKA, N. MATTA, M. RIBIÈRE. *Knowledge management: Méthodes et outils pour la gestion des connaissances, 3rd edition*, DUNOD, Octobre 2005.

[8] R. DIENG-KUNTZ, O. CORBY, A. GIBOIN, M. RIBIÈRE. *Methods and Tools for Corporate Knowledge Management*, in "International Journal of Human-Computer Studies, special issue on knowledge Management", vol. 51, December 1999, p. 567-598.

[9] R. DIENG-KUNTZ, A. GIBOIN, C. AMERGÉ, O. CORBY, S. DESPRÉS, L. ALPAY, S. LABIDI, S. LAPALUT. *Building of a Corporate Memory for Traffic-Accident Analysis*, in "AI Magazine", vol. 19, n^{o} 4, December 1998, p. 81-101.

[10] R. DIENG-KUNTZ, S. HUG. *Comparison of "Personal Ontologies" Represented through Conceptual Graphs*, in "Proc. of the 13th European Conference on Artifical Intelligence (ECAI'98), Brighton, UK", H. PRADE (editor), Wiley & Sons, August 1998, p. 341-345.

[11] F. GANDON, L. BERTHELOT, R. DIENG-KUNTZ. *A Multi-Agent Platform for a Corporate Semantic Web*, in "Proceedings of AAMAS'2002 (First International Joint Conference on Autonomous Agents and Multi-Agent Systems), Bringing People and Agents Together, Bologna, Italy", C. CASTELFRANCHI, W. JOHNSON (editors), July 15-19 2002, p. 1025-1032.

[12] F. GANDON. *Distributed Artificial Intelligence And Knowledge Management: Ontologies and Multi-Agents Systems for a Corporate Semantic Web*, Ph. D. Thesis, université de Nice - Sophia Antipolis, November 7th 2002.

[13] A. GIBOIN. *Conversational Remembering in Teams of Road Accident Analysts: Using a Model of Collective Memory for Designing an Organizational Memory System*, in "Le Travail Humain", vol. 63, n^{o} 3, 2000, p. 227-257.

[14] J. GOLEBIOWSKA, R. DIENG-KUNTZ, O. CORBY, D. MOUSSEAU. *Building and Exploiting Ontologies for an Automobile Project Memory*, in "Proc. of the First International Conference on Knowledge Capture (K-CAP 2001), Victoria, Canada", October 22 - 23 2001.

[15] P. MARTIN. *CGKAT: a Knowledge Acquisition and Retrieval Tool using Structured Documents and Ontologies*, in "Fulfilling Peirce's Dream, Proc. of the 5th International Conference on Conceptual Structures (ICCS'97), University of Washington, Seattle Washington USA", D. LUKOSE, H. DELUGACH, M. KEELER, L. SEARLE, J. SOWA (editors), Springer-Verlag, LNAI n. 1257, August 4-8 1997, p. 168-182.

[16] M. RIBIÈRE, R. DIENG-KUNTZ. *A Viewpoint Model for Cooperative Building of an Ontology*, in "Conceptual Structures : Integration and Interfaces, Proceedings of the 10th International Conference in Conceptual Structures (ICCS'2002), Darmstadt, Allemagne", U. PRISS, D. CORBETT, G. ANGELOVA (editors), Springer-Verlag, LNCS 2393, July 15-19 2002, p. 220-234.

Year Publications

Books and Monographs

[17] H. CHERFI, R. DIENG-KUNTZ (editors). *Proc. of KW4ED: the KCap'07 Workshop on Knowledge Management and Semantic Web for Engineering Design*, October 2007.

[18] R. DIENG-KUNTZ, C. ENGUEHARD (editors). *Proc. 7ème Conférence Terminologie et Intelligence Artificielle (TIA'2007)*, PUG, Sophia Antipolis, October 2007.

[19] R. DIENG-KUNTZ, N. MATTA (editors). *Proc. IJCAI'2007 Workshop on Knowledge Management and Organizational Memories*, January 2007.

Doctoral dissertations and Habilitation theses

[20] S. DEHORS. *Exploiting Semantic Web and Knowledge Management Technologies for E-learning*, Ph. D. Thesis, University of Nice-Sophia Antipolis, INRIA, February 2007.

[21] P.-H. LUONG. *Gestion de l'évolution d'un web sémantique d'entreprise*, Ph. D. Thesis, Ecole Nationale Supérieure des Mines de Paris, INRIA, December 2007.

Articles in refereed journals and book chapters

[22] F. Q. ALKHATEEB, J.-F. BAGET, J. EUZENAT. *Extending SPARQL with Regular Expression Patterns (for Querying RDF)*, in "Journal of Web Semantics", Accepted with revisions, 2007.

[23] M. BUFFA, G. ERÉTÉO, C. FARON-ZUCKER, F. GANDON, P. SANDER. *Semantic Wikis*, in "Journal of Web Semantics, special issue on Web 2.0 and the Semantic Web", accepted for publication, 2007.

[24] M. BUFFA, F. GANDON, G. ERÉTÉO. *A Wiki on the Semantic Web*, in "Emerging Technologies for Semantic Web Environments: Techniques, Methods and Applications, Germany", J. RECH, B. DECKER, E. RAS (editors), Fraunhofer Institute for Experimental Software Engineering (IESE), July 2007.

[25] K. KHELIF, R. DIENG-KUNTZ, P. BARBRY. *An Ontology-based Approach to Support Text Mining and Information Retrieval in the Biological Domain*, in "Journal of Universal Computer Science (JUCS), Special Issue on Ontologies and their Applications", accepted for publication, 2007.

[26] P.-H. LUONG, R. DIENG-KUNTZ. *A Rule-based Approach for Semantic Annotation Evolution*, in "The Computational Intelligence Journal", vol. 23, n^{o} 3, 2007, p. 320-338.

Publications in Conferences and Workshops

[27] T. L. BACH, R. DIENG-KUNTZ. *A Graph-Based Algorithm for Alignment of OWL Ontologies*, in "Proc. of IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, USA", November 2007.

[28] J.-F. BAGET. *A Datatype Extension for Simple Conceptual Graphs and Conceptual Graphs Rules*, in "Proceedings of the 15th International Conference on Conceptual Structures (ICCS 2007)", LNCS, Springer, 2007, p. 83-96.

[29] J.-F. BAGET, S. LABORIE. *Bi-Intervals for Backtracking on Temporal Constraint Networks*, in "Proceedings of the 14th International Symposium on Temporal Representation and Reasoning (TIME 2007)", IEEE Computer Society, 2007, p. 163-168.

[30] M. BUFFA, F. GANDON. *Wikis et Web Sémantique*, in "Proc. Ingénierie des Connaissances, IC'2007, Grenoble", July 2007, p. 49-60.

[31] H. CHERFI, O. CORBY, C. MASIA-TISSOT. *RDF(S) and SPARQL Epressiveness in Engineering Design Patterms*, in "Proc. of IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, USA", November 2007.

[32] O. CORBY, C. FARON-ZUCKER. *Implementation of SPARQL Query Language based on Graph Homomorphism*, in "Proc. of the 15th International Conference on Conceptual Structures (ICCS'2007), Sheffield, UK", July 2007, p. 472-475.

[33] O. CORBY, C. FARON-ZUCKER. *RDF/SPARQL Design Pattern for Contextual Metadata*, in "Proc. of IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, USA", November 2007.

[34] P. DURVILLE, F. GANDON. *Sewese : Semantic Web Server*, in "WWW'2007 Developers track, Banff, Canada", 2007.

[35] F. GANDON, V. BOTTOLLIER, O. CORBY, P. DURVILLE. *RDF/XML Source Declaration*, in "Proc. of IADIS International Conference WWW/Internet, Vila Real, Portugal", October 2007.

[36] A. E. GHALI, A. TIFOUS, M. BUFFA, A. GIBOIN, R. DIENG-KUNTZ. *Using a Semantic Wiki in Communities of Practice*, in "Proceedings of TEL-CoPs'07 workshop, EC-TEL'07, Crete, Greece", September 2007.

[37] N. GHOULA, K. KHELIF, R. DIENG-KUNTZ. *Supporting Patent Mining by using Ontology-based Semantic Annotations*, in "Proc. of IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, USA", November 2007.

[38] M. LO, F. GANDON. *Semantic Web Services in Corporate Memories*, in "ICIW 2007, International Conference on Internet and Web Applications and Services, Mauritius", May 2007.

[39] P.-H. LUONG, R. DIENG-KUNTZ, A. BOUCHER. *Evolution de l'ontologie et gestion des annotations sémantiques inconsistantes*, in "Proc. 7ème conférence francophone Extraction et Gestion des Connaissances (EGC'2007), Belgique", January 2007, p. 635-646.

[40] B. MAKNI, K. KHELIF, R. DIENG-KUNTZ, H. CHERFI. *Création semi-automatique d'une ontologie et des annotations sémantiques pour une liste de diffusion d'une communauté de pratique*, in "Atelier Ontologies et Textes associé à la 7ème conférence Terminologie et Intelligence Artificielle TIA'07, Sophia Antipolis", October 2007.

[41] I. MIRBEL. *Connecting Method Engineering Knowledge: a Community Based Approach*, in "Proc. IFIP WG8.1 Working Conference on Method Engineering, Geneva, Switzerland.", September 2007.

[42] Y. MRABET, K. KHELIF, R. DIENG-KUNTZ. *Recognising Professional-Activity Groups and Web Usage Mining for Web Browsing Personalisation*, in "Proc. of IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, USA", November 2007.

[43] R. THOMOPOULOS, J.-F. BAGET, O. HAEMMERLÉ. *Conceptual Graphs as Cooperative Formalism to Build and Validate a Domain Expertise*, in "Proceedings of the 15th International Conference on Conceptual Structures (ICCS 2007)", LNCS, Springer, 2007, p. 112-125.

[44] A. TIFOUS, A. E. GHALI, R. DIENG-KUNTZ, A. GIBOIN, C. EVANGELOU, G. VIDOU. *An Ontology for Supporting Communities of Practice*, in "Proceedings of KCAP'07, The Fourth International Conference on Knowledge Capture, Whistler, BC, Canada", October 2007, p. 28-31.

[45] A. TIFOUS, A. E. GHALI, A. GIBOIN, R. DIENG-KUNTZ. *O'CoP, an Ontology Dedicated to Communities of Practice*, in "Proceedings of I-KNOW'07, The 7th International Conference on Knowledge Management, Graz, Austria", September 2007.

Internal Reports

[46] Y. A. AMEUR, M. BENNIS, O. CORBY, R. DIENG-KUNTZ, P. DURVILLE, C. FANKAM, A. GIBOIN, B. GRILHERES, S. JEAN, P.-H. LUONG, G. PIERRA, E. SARDET. *Outils et Services de gestion des ontologies* , Deliverable e-WoK HUB, WP 2/Task 2, Technical report, May 2007.

[47] A. BURGER. *Specification of available services*, Deliverable of Sealife project, INRIA Contributors: Khaled Khelif, Rose Dieng-Kuntz, Technical report, n^{o} I2-D1, March 2007.

[48] H. CHERFI. *Extension of the ontology to support the text annotation process*, Deliverable of SevenPro project, v.1.1 INRIA Contributors: Hacène Cherfi, Technical report, n^{o} D.4.1.1, July 2007.

[49] H. CHERFI. *Semantic Annotation of Engineering Repositories Prototype*, Deliverable of SevenPro project, Prototype (report in December 2006), v.1.5 INRIA Contributors: Hacène Cherfi, Rose Dieng-Kuntz and Khaled Khélif, Technical report, n° D.3.3.1, February 2007.

[50] H. CHERFI. *Specification and implementation of annotation extractors for text documents*, Deliverable of SevenPro project, v.1.3, INRIA Contributors: Hacène Cherfi, Olivier Corby and Minh-Tuan Nguyen, Technical report, n° D4.1.2, October 2007.

[51] H. CHERFI. *Upper layers of the ontology*, Deliverable of SevenPro project, v.1.4 INRIA Contributors: Hacène Cherfi and Olivier Corby, Technical report, n° D.2.1.1, February 2007.

[52] C. COLOMBO. *Evaluation ergonomique de l'éditeur collaboratif d'ontologie ECCO*, Rapport final de stage pour le Master « Ergonomie des NTIC » , Technical report, November 2007.

[53] C. COLOMBO. *Intervention ergonomique sur ECCO: Éditeur collaboratif d'ontologie* , Rapport intermédiaire de stage pour le Master « Ergonomie des NTIC », Technical report, July 2007.

[54] A. DAELE. *Description of 6 scenarios and of the results of 6 validated trials*, Deliverable of Palette project, INRIA Contributors: Michel Buffa, Adil El Ghali, Alain Giboin and Amira Tifous, Technical report, n° D.PAR.03, September 2007.

[55] A. EL GHALI, R. DIENG-KUNTZ. *Basic CoP-oriented Knowledge Management Tool offering basic CoP-adapted KM Services*, Deliverable of Palette project, INRIA Contributors: Adil El Ghali, Alain Giboin and Amira Tifous, Technical report, n° D.KNO.04, November 2007.

[56] F. GANDON, V. BOTTOLLIER, O. CORBY, P. DURVILLE. *RDF/XML Source Declaration*, W3C Member Submission, Technical report, September 2007, http://www.w3.org/Submission/rdfsource.

[57] F. GANDON. *GRDDL Use Cases: Scenarios of extracting RDF data from XML documents*, Working Group Note, Technical report, W3C, April 2007, http://www.w3.org/TR/grddl-scenarios.

[58] F. GANDON. *Specification and implementation of Engineering Memory Tool*, Deliverable of SevenPro project, v.1.1 INRIA Contributors: Ibrahima Diop, Fabien Gandon and Cheikh Niang, Technical report, n° D.7.3.1, November 2007.

[59] N. GHOULA. *Fouille de brevets pour la génération d'annotations sémantiques*, Rapport de fin d'études PFE, Juillet 2007.

[60] J. HAKENBERG, M. SCHROEDER. *First demonstrator*, Deliverable of Sealife project, INRIA Contributors: Khaled Khelif, Rose Dieng-Kuntz, Technical report, n° I2-D1, March 2007.

[61] E. JAMIN. *Analysis, specification and implementation of a rule-based VR reasoning module*, Deliverable of SevenPro project, v.1.1 INRIA Contributors: Olivier Corby and Emmanuel Jamin, Technical report, n° D.6.3.1, November 2007.

[62] K. KHELIF, R. DIENG-KUNTZ. *Mining Biomedical Texts to Generate Semantic Annotations*, Technical report, n° RR-6102, INRIA, January 2007, http://hal.inria.fr/inria-00125266.

[63] K. KHELIF, R. DIENG-KUNTZ. *Analysis of existing term extraction methods*, Deliverable of Sealife project, INRIA Contributors: Khaled Khelif, Rose Dieng-Kuntz, Technical report, n^{o} I2-D1, March 2007.

[64] K. KHELIF, R. DIENG-KUNTZ. *Development and Implementation of Extraction Methods*, Deliverable of Sealife project, INRIA Contributors: Khaled Khelif, Rose Dieng-Kuntz, Nizar Ghoula, Yassine Mrabet, Olivier Corby, Technical report, n^{o} I2-D2, September 2007.

[65] Y. MRABET. *Personnalisation de la navigation dans un site Web biomédical par apprentissage de profils utilisateurs*, Rapport de fin d'études PFE, Juillet 2007.

[66] A. TIFOUS, R. DIENG-KUNTZ. *CoP-dependent ontologies*, Deliverable of Palette project, INRIA Contributors: Rose Dieng-Kuntz, Priscille Durville, Adil El Ghali, Alain Giboin and Amira Tifous, Technical report, n^{o} D.KNO.02, March 2007.

Miscellaneous

[67] B. MAKNI. *Construction semi-automatique d'une ontologie et des annotations sémantiques à partir d'une liste de diffusion d'une communauté de pratique*, Rapport de fin d'études extensible PFEE, ENSI, Tunisie, Septembre 2007.

References in notes

[68] I. ALEXANDER, N. MAIDEN (editors). *Scenarios, Stories, Use Cases. Through the Systems Development Life-Cycle*, Wiley, 2004.

[69] F. ALKHATEEB, J.-F. BAGET, J. EUZENAT. *RDF with Regular Expressions*, Technical report, n^{o} RR-6191, http://hal.inria.fr/inria-00144922/en.

[70] K. BAKER, S. GREENBERG, C. GUTWIN. *Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration*, in " Engineering for Human-Computer Interaction (8th IFIP International Conference, EHCI 2001, Toronto, Canada, May", M. LITTLE, L. NIGAY (editors), Lecture Notes in Computer Science Vol 2254, Springer-Verlag, 2001, p. 123-139.

[71] K. BAKER, S. GREENBERG, C. GUTWIN. *Empirical development of a heuristic evaluation methodology for shared workspace groupware*, in "Proceedings of the ACM Conference on Computer Supported Cooperative Work", Conference Chair-Elizabeth F. Churchill and Conference Chair-Joe McCarthy and Program Chair-Christine Neuwirth and Program Chair-Tom Rodden, ACM Press, 2002, p. 96-105.

[72] S. BECHHOFER, C. GOBLE, L. CARR, W. HALL, S. KAMPA, D. DE ROURE. *COHSE: Conceptual Open Hypermedia Service*, S. HANDSCHUH, S. STAAB (editors), IOS Press, Amsterdam, 2003.

[73] J. BRESLIN, A. HARTH, U. BOJARS, S. DECKER. *Towards Semantically-Interlinked Online Communities*, in "Second European Semantic Web Conference, ESWC 2005, Heraklion,", 2005, http://www.springerlink.com/content/xtyetvllxfnk/.

[74] J. M. CARROLL. *Making Use: Scenario-Based Design of Human-Computer Interactions*, The MIT Press, Cambridge, MA, 2000.

[75] P. CIMIANO, S. HANDSCHUH, S. STAAB. *Towards the self-annotating web*, in "Proc. 13th International World Wide Web Conference, (WWW 2004), New York, USA", May 2004.

[76] F. CIRAVEGNA, Y. WILKS. *Designing Adaptive Information Extraction for the Semantic Web in Amilcare*, in "Annotation for the Semantic Web in the Series Frontiers in Artificial Intelligence and Applications", S. HANDSCHUH, S. STAAB (editors), IOS Press, Amsterdam, 2003.

[77] O. CORBY, C. FARON-ZUCKER. *Corese: A Corporate Semantic Web Engine*, in "Proceedings of the International Workshop on Real World RDF and Semantic Web Applications, 11th International World Wide Web Conference, Hawai, USA", May 7 2002, http://www.cs.rutgers.edu/~shklar/www11/.

[78] R. DIENG-KUNTZ, O. CORBY, F. GANDON, A. GIBOIN, J. GOLEBIOWSKA, N. MATTA, M. RIBIÈRE. *Knowledge management: Méthodes et outils pour la gestion des connaissances, 3rd edition*, DUNOD, Octobre 2005.

[79] R. DIENG-KUNTZ. *Corporate Semantic Webs*, in "Encyclopaedia of Knowledge Management", D. SCHWARTZ (editor), Idea Publishing Group, September 2005, p. 67-80.

[80] L. DUBÉ, A. BOURHIS, R. JACOB. *Towards a Typology of Virtual Communities of Practice*, in "Interdisciplinary Journal of Information, Knowledge, and Management", vol. 1, 2006, p. 69-93.

[81] M. DZBOR, J. DOMINGUE, E. MOTTA. *Magpie: Towards a semantic web browser*, in "Proceedings of the Second International Semantic Web Conference, ISWC, Sanibel Island, Florida, USA", October 2003.

[82] F. GANDON, O. CORBY, A. GIBOIN, N. GRONNIER, C. GUIGARD. *Graph-based inferences in a Semantic Web Server for the Cartography of Competencies in a Telecom Valley*, in "Proceedings International Semantic Web Conference, Galway", Springer, Lecture Notes in Computer Science, November 6-10 2005.

[83] F. GANDON, R. DIENG-KUNTZ, O. CORBY, A. GIBOIN. *Semantic Web and Multi-Agents Approach to Corporate Memory Management*, in "Proceedings of the 17th IFIP World Computer Congress IIP Track - Intelligent Information Processing, Montréal, Canada", M. MUSEN, B. NEUMANN, R. STUDER (editors), August 2002, p. 103-115.

[84] F. GANDON. *Distributed Artificial Intelligence And Knowledge Management: Ontologies and Multi-Agents Systems for a Corporate Semantic Web*, Ph. D. Thesis, université de Nice Sophia Antipolis, 7 novembre 2002.

[85] D. GENEST, E. SALVAT. *A Platform Allowing Typed Nested Graphs: How CoGITo Became CoGITaNT*, in "Proceedings of the 6th International Conference on Conceptual Structures (ICCS'98), (session CG Tools)", LNCS, Springer, p. 154-161, http://cogitant.sourceforge.net.

[86] A. GIBOIN, F. GANDON, N. GRONNIER, C. GUIGARD, O. CORBY. *Comment ne pas perdre de vue les usage(r)s dans la construction d'une application à base d'ontologies ? Retour d'expérience sur le projet KmP*, in "Actes des 16e Journées francophones d'Ingénierie des connaissances (IC'2005), Grenoble. France", M. JAULENT (editor), PUG, 2005, p. 133-144.

[87] J. GOLEBIOWSKA, R. DIENG-KUNTZ, O. CORBY, D. MOUSSEAU. *Building and Exploiting Ontologies for an Automobile Project Memory*, in "Proc. of the First International Conference on Knowledge Capture (K-CAP 2001), Victoria, Canada", October 22 - 23 2001.

[88] J. GOLEBIOWSKA, R. DIENG-KUNTZ, O. CORBY, D. MOUSSEAU. *Samovar: Using Ontologies and Text-Mining for Building an Automobile Project Memory*, in "Knowledge Management and Organizational Memories", Kluwer Academic Publishers, July 2002, p. 89-102.

[89] J. GOLEBIOWSKA. *Exploitation des ontologies pour la mémoire d'un projet-véhicule: méthode et outil SAMOVAR*, Ph. D. Thesis, université de Nice Sophia Antipolis, 4 février 2002.

[90] T. GRUBER. *Ontology of Folksonomy: A Mash-up of Apples and Oranges*, in "Conference on Metadata and Semantics Research (MTSR)", 2005.

[91] H. HALPIN, V. ROBU, H. SHEPHERD. *The Complex Dynamics of Collaborative Tagging*, in "Proceedings of the 16th International World Wide Web Conference (WWW'07), New York, NY, USA", ACM Press, 2007, http://www2007.org/papers/paper635.pdf.

[92] A. JAMESON. *Usability and the Semantic Web*, in "The Semantic Web: Research and Applications: Proc. of the Third European Semantic Web Conference, ESWC ", Y. SURE, J. B. DOMINGUE (editors), Springer, 2006.

[93] K. J. KOCHUT, M. JANIK. *SPARQLeR: Extended SPARQL for Semantic Association Discovery*, in "Proc. European Semantic Web Conference, ESWC'2007, Innsbruck, Austria", 2007.

[94] M.-R. KOIVUNEN. *Annotea and Semantic Web Supported Collaboration,* , in "Invited talk at Workshop on User Aspects of the Semantic Web (UserSWeb), at European Semantic Web Conference (ESWC 2005), Heraklion, Greece", May 2005.

[95] C. MEDINA-RAMÍREZ, O. CORBY, R. DIENG-KUNTZ. *A Conceptual Graph and RDF(S) Approach for Representing and Querying Document Content*, in "Advances in Artificial Intelligence IBERAMIA 2002, Proceedings of the 8th International Ibero-American Conference on AI, Séville, Spain", F. J. GARIJO, J. C. RIQUELME, M. TORO (editors), Springer-Verlag, LNAI 2527, November 2002, p. 121-130.

[96] J. NIELSEN. *Heuristic Evaluation*, in "Usability Inspection Methods", John Wiley and Sons, New York, 1994, p. 25-62.

[97] T. O'REILLY. *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*, September 2005, http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

[98] A. PASSANT. *Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs*, in "International Conference on Weblogs and Social Media", 2007.

[99] R. SINHA. *Tagging from Personnal to social : observations and design principles*, in "WWW2006 Tagging Workshop", 2006.

[100] V. UREN, P. CIMIANO, J. IRIA, S. HANDSCHUH, M. VARGAS-VERA, E. MOTTA, F. CIRAVEGNA. *Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art*, in "Journal of Web Semantics", n^{o} 4, 2005, p. 14-28.

[101] G. VIDOU, R. DIENG-KUNTZ, A. EL GHALI, C. EVANGELOU, A. GIBOIN, A. TIFOUS, S. JACQUEMART. *Towards an Ontology for Knowledge Management in Communities of Practice*, in "Proceedings of 6th

International Conference on Practical Aspects of Knowledge Management, Vienna, Austria", Lecture Notes in Computer Science, Springer, Nov. 30th - Dec. 1st 2006.

[102] R. VOLZ, D. OBERLE, S. STAAB, B. MOTIK. *KAON SERVER - A Semantic Web Management System*, in "In Alternate Track, Proceedings of the 12th International World Wide Web Conference, WWW2003, Budapest, Hungary", ACM, May 2003.

[103] M. WEEBER, J. MORK, A. R. ARONSON. *Developing a test collection for biomedical word sense disambiguation*, in "Journal of the American Medical Informatics Association", 2001, p. 746-750.

[104] E. WENGER, R. MCDERMOTT, W. M. SNYDER. *Cultivating Communities of Practice - A guide to managing knowledge*, Harvard Business School Press, Boston, MA, 2002.

[105] E. WENGER. *Communities of Practice: Learning as a Social System*, in "Systems Thinker", vol. 9, n^{o} 5, 1998.

[106] M. ZACKLAD, M. LEWKOWICZ, J.-F. BOUJUT, F. DARSES, F. DÉTIENNE. *Forme et gestion des annotations numériques collectives en ingénierie collaborative*, in "14ème journées Francophones de l'ingénierie des connaissances, IC 2003, Laval, France", July 2003, p. 207-224.

[107] A. DE MOOR. *Patterns for the Pragmatic Web*, in "Proc. of the 13th International Conference on Conceptual Structures, ICCS 2005, Kassel, Germany", LNAI, Springer Verlag, July 2005, p. 1-18.