INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team EXMO

Computer-mediated communication of structured knowledge

Rhône-Alpes

THEME SYM

Rapport d'Activité

2005

Table of contents

1. Team

Head of project-team
Jérôme Euzenat [Research scientist (DR INRIA)]

Administrative assistant
Stéphanie Berger [25% (IA INRIA)]

Research scientist
Jean-François Baget [CR INRIA]

Post doctoral fellow
Jason Jung [INRIA, since September 1st, 2005]

Ph. D. students
Faisal Al-Khateeb [Jordan government, since October 1st, 2005]
Sébastien Laborie [MESR grant]
Jérôme Pierson [CIFRE (France Telecom R&D)]
Antoine Zimmermann [INRIA]

2. Overall Objectives

Keywords: *Alignment API*, *OLA*, *OWL*, *RDF*, *RDF Path*, *Transmorpher*, *XSLT*, *content representation*, *context*, *knowledge representation*, *knowledge transformation*, *multimedia document adaptation*, *ontologies*, *ontology alignment*, *property preservation*, *semantic web*, *semantics of knowledge representation*, *semiotics*.

We consider that, in future information systems, formalized knowledge will be massively exchanged. In this communication process, the computer can complement its medium and memory roles by formatting, filtering, classifying, consistency checking or generalizing knowledge. These manipulations can be thought of as transformations. This approach is developing with the generalized use of standardized exchange languages (XML) in network communication as well as more precisely defined languages (RDF and OWL) on the semantic web. Consequently, the users must ask for more guarantees from the performed transformations. *Exmo*'s goal is the development of theoretical and software tools for enabling interoperability in formalized knowledge exchange.

Our main approach is the investigation of the properties that transformations must satisfy. Among these properties are content or structure preservation, traceability, or, conversely, confidentiality. We want to contribute to a "general theory of transformations" grounded on transformation properties rather than on transformations themselves.

On the one hand, we aim at showing that this approach can be applied to various contexts depending on languages, properties and transformations. We study information preservation modelled by lattices or scenario preservation in adapted multimedia documents. On the other hand, we investigate more precisely semantic properties when transforming from one knowledge representation language to another. The main question is: are the consequences of the initial representation still consequences of the transformed one? For helping answering it, we establish relationships between several kinds of semantic properties, known or new. We also investigate more restricted cases (family of languages, patterns) that can use these relationships for establishing properties more quickly.

Our medium term objectives are (1) the opportunity to analyse transformations in terms of a set of composition operators, (2) the isolation of useful properties and the study of their behavior regarding these operators and (3) the capability to prove these properties on this kind of transformations.

On a longer term, we want to explore properties that we call semiotic, i.e., which concern the interpretation of the communicated representation by a human user. This goal should require an analysis of the extra-semantic rules that govern the choice of subsets of models.

Anticipated applications are in transformation system engineering (in which the information system is seen as a transformation flow) and semantic web infrastructure.

3. Scientific Foundations

3.1. Knowledge representation semantics

We usually work with semantically defined knowledge representation languages (like description logics [26], conceptual graphs and object-based languages). Their semantics is usually defined within model theory initially developed for logics. The languages dedicated to the semantic web (RDF and OWL) follow that approach.

We consider a language L as a set of syntactically defined expressions (often inductively defined by applying constructors over other expressions). A representation (r) is a set of such expressions. An interpretation function (I) is inductively defined over the structure of the language to a structure called interpretation domain (D). This expresses the construction of the "meaning" of an expression in function of its components. A formula is satisfied by an interpretation if it fulfills a condition (in general being interpreted over a particular subset of the domain). A model of a set of expressions is an interpretation satisfying all these interpretations. An expression (δ) is then a consequence of a set of expressions (r) if it is satisfied by all of its models (noted $r \models \delta$).

A computer must determine if a particular expression (taken as a query, for instance) is the consequence of a set of axioms (a knowledge base). For that purpose, it uses programs, called provers, that can be based on the processing of a set of inference rules, on the construction of models or on more classical programming. These programs are able to deduce theorems (noted $r \vdash \delta$). They are said to be sound if they only find theorems which are indeed consequences and to be complete if they find all the consequences as theorems. However, depending on the language and its semantics, the decidability (i.e., the ability to create sound and complete provers) is not warranted. Even for decidable languages, the algorithmic complexity of provers can prohibit their exploitation.

To solve this problem a trade-off between the expressivity of the language and the complexity of its provers has to be found. These considerations have led to the definition of languages with limited complexity - like conceptual graphs and object-based representations - or of modular families of languages with associated modular prover algorithms - like description logics.

Exmo mainly considers languages with well-defined semantics (such as RDF and OWL that we contributed to define), and defines the semantics of some languages such as multimedia specification languages, in order to establish the properties of computer manipulations of the representations.

3.2. Transformations and properties

A transformation (τ) is an algorithmic manner to generate a representation ($\tau(r)$) from another one (r), not necessarily in the same language. We will focus on transformations made by composition of more elementary transformations for which we only know the input, output and assumed properties.

A transformation system is characterized by a set of elementary transformations and a set of composition operators. A transformation flow is the composition of elementary transformation instances whose input/output are connected by channels. A transformation flow is itself a transformation. More precisely, our work concerns syntactic transformations of XML ("eXtensible Markup Language") structures encoding knowledge representation languages. We take advantage of the XSLT transformation language ("XML Style Language Transformations" [27]) recommended by W3C, for which we put forward a compound transformation language (see §5.1).

The design of information systems as transformation flows requires the ability to express such flows and to determine their properties. A property is a boolean predicate about the transformation (e.g., "preserving information" is such a predicate - it is true or false of a transformation - which is satisfied if there exists an algorithmic means to recover r from $\tau(r)$).

We consider more closely preservation properties that can allow the preservation (or anti-preservation) of an order relation between the source representations and the target representations. For instance, one can identify:

- Syntactic properties: based on the organization of syntactic elements, like the completion ($\tau(r) \leq r$, in which $\leq$ denotes structural subsumption between representations);
- Semantic properties: based on model and consequence concepts, like consequence preservation ($\forall r \subseteq L, \forall \delta \in L, r \models_L \delta \Rightarrow \tau(r) \models_{L'} \tau(\delta)$);
- Semiotic properties: based on the interpretation as signs of the manipulated objects, like interpretation preservation (let σ be the interpretation rules and $\models_{L,i}$ be the interpretation of individual i, $\forall r \subseteq L, \forall \delta \in L, \forall i, j, r, \sigma \models_{L,i} \delta \Rightarrow \tau(r), \tau(\sigma) \models_{L',j} \tau(\delta)$).

Our goal is to study transformations based on transformation properties rather than on representations or transformation structures. This does not deal only with semantics but considers various properties (e.g., content or structure preservation, traceability, and confidentiality). However, we more specifically address semantic properties. We also consider properties of transformation systems (given a transformation system, is information preservation decidable and at what cost?). We try to characterize, given a particular type of property, which transformations leave them invariant and what is the action of composition operators.

3.3. Ontology alignment

When different conceptual models (called ontologies) for representations are used, it is necessary to identify their correspondences before designing a transformation (or another use). This task is called ontology matching and its result is an alignment. It can be described as follows: given two ontologies, each describing a set of discrete entities (which can be classes, properties, rules, predicates, etc.), find the relationships (e.g., equivalence or subsumption) that hold between these entities.

An alignment between two ontologies *O* and *O'* is a set of correspondences <*e, e', r, n*> in which

- *e*∈*O* and *e'*∈*O'* are the entities between which a relation is asserted by the correspondence (e.g., formulas, terms, classes, individuals);
- *r* is the relation asserted to hold between *e* and *e'*. Nothing is said about the relation but that it must apply to the pair of entities. For instance, this relation can be a simple set-theoretic relation (applied to entities seen as sets or their interpretation seen as sets), a relationship defined in representation languages (such as subsumption), a fuzzy relation, a probabilistic distribution over a complete set of relations, a similarity measure, etc.
- *n* is a degree of confidence in this particular correspondence. The degree of confidence belongs to a structure <*D*, $\leq$, $\top$, $\bot$> such that *D* is the set of degrees, $\leq$ is an order on *D*×*D* such that whatever *d*∈*D*, $\bot \leq d \leq \top$. This structure is applicable to a wide number of measures (e.g., boolean lattice, fuzzy degrees, probabilities, any lattice). The confidence degree can be computed in many ways, including users' feedback or log analysis.

A variety of methods from the literature may be used for this task, basically they perform pair-wise comparison of entities from each of the ontologies and select the most similar pairs. Most matching algorithms provide correspondences between named entities, more rarely between compond terms. The relationships are generally the equivalence between these entities. Some systems are able to provide subsumption relations as well as other relations in the support language (like incompatibility or instanciation). Confidence measures are usually given a value between 0 and 1 and are used for expressing preferences between two correspondences.

4. Application Domains

Two application contexts motivate and spur our work: transformation system engineering (§4.1) and the "semantic web" infrastructure (§4.2)

4.1. Transformation system engineering

Computerization and networking lead organizations to exchange information in machine-readable form. E-commerce generates a continuous flow of such documents. As the transmitted information is neither addressed nor adapted to all the members of the organization, it is necessary to transform document structure and content. Similarly, web sites are generated from databases or primary funds and e-commerce documents are applied various transformations before goods are shipped. Additionally, the Object Management Group Model-Driven Architecture (MDA) considers the future of software development as a composition of transformations from (platform independent) domain models to other (platform dependent) models in function of platform description models. This is considering any implementation as adaptation.

Interoperability requirements have led to the definition of the structured document representation language XML which helps handling the syntax of documents straightforwardly. Other languages such as XSLT or Omnimark, enable the implementation of standalone transformations.

However, this view of transformations is only partial and local. It seems unavoidable that, in the future, we will have to deal with complex transformation flows automating the combination of transformations, some of which coming from external sources. This will require the global understanding of the behavior of the flow of transformations. This calls for real "transformation system engineering" which should address the following issues:

- the lack of global consideration of transformations: they are processed in relation with other transformations;
- the need to consider the properties of transformations and especially their semantic properties: this will require the semantic analysis of the transformations;
- the design of transformation flows from external resources (as it is in software engineering): this will require the ability to consider the properties of imported transformations.

Transformation system engineering will require tools, methodologies and formal methods. As a matter of fact, it will be necessary to check that a particular transformation system does not export sensitive information or that the transformation process terminates. For that purpose, the transformation flow must be expressed in a parsable way and the expected properties of the flow must be expressed (see §3.2). *Exmo* is concerned by tools and formal methods and aims at combining them in solutions for transformation flow design environments (see §5.1).

4.2. Semantic web technologies

Internet technologies support access and sharing of knowledge, often difficult to access in a documentary form, in the organizations. However, the technologies quickly reach their limits: web site organization is expensive and full-text search inefficient. Content-based information search is becoming a necessity. Content representation will enable computers to manipulate knowledge on a more formal ground and to carry out similarity or generality search... Knowledge representation formalisms are good candidates for expressing the content.

The vision of a "semantic web" [28] supplies the web, as we know it (informal) with annotations expressed in a machine-processible form and linked together. In the context where web documents are formally annotated, it becomes necessary to import and manipulate the annotations according to their semantics and their use. Taking advantage of this semantic web will require the manipulation of various knowledge representation formats. *Exmo* concerns are thus central to the semantic web implementation.

Our work aims at enhancing content understanding, including the intelligibility of communicated knowledge and formal knowledge transformations. The semantic web idea is essentially based on the notion of ontology (that can be quickly described as conceptual schemes or knowledge bases). Even if a standard knowledge representation language emerges, it will still be necessary to import and exchange ontologies in such a way that the semantics of their representation language is taken care of. We work on finding the correspondences between various knowledge representation languages and ontologies (see §6.3) in order to take advantage of them in ontology merging and bridging or message translation. Bringing solutions to this problem is part of the ambition of *Exmo*.

5. Software

Exmo's work can be implemented in software: we have proposed an API for expressing ontology alignment (§5.1) and we have designed and developed a system for expressing and processing relatively complex transformation flows (§5.2).

5.1. Alignment API: manipulating ontology alignments

Participants: Jérôme Euzenat [Contact], Jérôme Pierson, Antoine Zimmermann.

We have designed a format for expressing alignments in a uniform way [2]. The goal of this format is to be able to share on the web the available alignments. It should help systems using alignments (e.g., mergers, translators) to take advantage of any alignment algorithm and it will help alignment algorithms to be used in many different tasks. This format is expressed in RDF, so it is freely extensible, and has been defined by a DTD (for RDF/XML), an OWL ontology and an RDF Schema. Aligned entities are identified by their URIs.

The API itself is a *Java* description of tools for accessing the common format. It defines four main interfaces (Alignment, Cell, Relation and Evaluator) and proposes the following services:

- Storing, finding, and sharing alignments;
- Piping alignment algorithms (improving an existing alignment);
- Manipulating (thresholding and hardening);
- Generating processing output (transformations, axioms, rules);
- Comparing alignments.

It will be extended with web service and agent communication languages in order to be used as an alignment service on the semantic web.

The API can be used for producing transformations, rules or bridge axioms independently from the algorithm which produced the alignment. The proposed API implementation is based on the OWL-API and offers floating point measures between 0 and 1. It features:

- a base implementation of the interfaces with all useful facilities;
- a library of sample aligners;
- a library of renderers (XSLT, SWRL, OWL, C-OWL, SEKT mapping language);
- a library of evaluators (precision/recall, generalized precision/recall, precision/recall graphs and weighted Hamming distance);
- a parser for the format.

To instanciate the API, it is sufficient to refine the base implementation by implementing the `align()` method. Doing so, the new implementation will benefit from all the services already implemented in the base implementation.

The *Alignment API* has been used for the processing of the EON 2004 Ontology Alignment Contest and Ontology Alignment Evaluation Initiative 2005. It is used in the people's portal alignment tool at DERI Innsbruck and used or output by a number of alignment tools (among which OLA that we develop jointly with the University of Montréal or CMS from University of Southampton).

This year we have specified extensions of this API for dealing with more expressive languages [20], and using the API as an alignment server [19][15].

The *Alignment API* is freely available since december 2003 under the LGPL licence at http://co4.inrialpes.fr/align

5.2. Transmorpher : expression and processing of XML transformation flows

Participant: Jérôme Euzenat [Contact].

In order to prove or check the properties of transformations, it is necessary to have a representation of these transformations. The XSLT language enables the expression of a transformation in XML but is relatively difficult to analyse. In order to overcome this problem, we have designed and developed in collaboration with the *Fluxmedia* company, the *Transmorpher* environment.

Transmorpher enables the definition and processing of generic transformations of XML documents. It aims at providing XSLT extensions in order to:

- Describe straightforwardly simple transformations (e.g., removing elements, replacing attribute names, merging documents, applying regular expression substitution);
- Composing transformations by connecting their (multiple) input and output;
- Applying transformations until closure;
- Calling external transformation engines.

Transmorpher describes the transformation flows in XML. Input/output channels carry the information, mainly XML, from one transformation to another. Transformations can be other transformation flows or elementary transformations. *Transmorpher* provides a set of abstract elementary transformations (including their execution model) and one default instantiation of each. Such elementary transformations include external call (e.g. XSLT), dispatcher, serializer, query engine, iterator, merger, generator and rule sets.

Rule sets describe transformations in a language simpler than XSLT. It is currently processed by transformation to XSLT (extended by regular expression substitution).

Transmorpher provides a set of documented *Java* classes (which can be refined or integrated into another software) and a transformation flow processing engine. A transformation flow can be described by programming in *Java* or providing an XML description. *Transmorpher* has been registered by the "Agence de Protection des Programmes" (APP) under the IDDN.FR.001.430009.000.R.P.2003.000.30620. It is freely available, under the GPL licence, since June 2001 at http://transmorpher.inrialpes.fr.

An extension of *Transmorpher* towards safer transformations consists of attaching assertions to the transformations in a transformation flow in order to tell if a property is assumed, proved or to be checked. This will allow experiments on proving properties of compound transformations.

6. New Results

The results in 2005 are mainly related to semantic adaptation of multimedia documents (§6.1), graph-based knowledge representation languages (§6.2), and ontology alignment (§6.3). We also started investigating the use of alignments related to context in ambient computing applications (§6.4).

6.1. Semantic adaptation of multimedia documents

Participants: Sébastien Laborie [Contact], Jérôme Euzenat, Nabil Layaïda [WAM team].

When a multimedia document is played on platforms with limited resources (e.g., a mobile phone that can only display one image at once or an interactive display without keyboard), it is necessary to adapt the document to the target device. In order to assess the meaning of adaptation, we have defined a semantic approach [4], which considers a model of a multimedia document as one of its potential execution (an execution satisfying its specification). In a first approximation, adaptation reduces the set of models of a specification by selecting those satisfying the adaptation constraints.

Adapting amounts to finding this subset of models or, when it is empty, finding a compatible execution as close as possible to the initial execution. For that purpose, we proposed to express the set of possible interpretation by a resolved relation graph. Each relation of this graph could be a temporal, spatial, spatio-temporal relation... This approach has been applied to the temporal and spatial dimensions based on Allen and RCC algebras respectively.

This year, we have followed on our work on semantic adaptation for the SMIL 2.0 language [18].

We have extended this framework to the spatial dimension. Our goal was to find a qualitative spatial representation that computes in a reasonable time a set of adaptations close to the initial document satisfying the adaptation constraints. We have studied different qualitative spatial representations. Some are very expressive and precise, like the directional representation, but the computational cost of adapting is high. Others, like topological representation, can be adapted quickly, but offer poor expressiveness. Thus, we have introduced a new spatial representation which trades off between expressiveness and speed. We have compared experimentally each spatial representation on SMIL examples.

Moreover, we have proposed a method to adapt SMIL documents in the spatio-temporal dimension. We have first focussed our framework on multimedia documents without animations (i.e., every multimedia object appears and disappears at the same place). We showed that chaining temporal and spatial adaptation independently would yield non minimal solutions. Thus, we proposed to use a spatio-temporal representation and show how to calculate proximity between spatio-temporal relations. Then, we have considered multimedia documents with animation. This consists of adding multiple spatial relations to the spatio-temporal representation and taking into account the property of continuity when the adaptation is applied.

6.2. Algorithms for graph-based knowledge representation languages

We are interested in various knowledge representation languages whose formulas can be represented by graphs and where the truth value of a formula in a model-theoretic interpretation can be expressed by some kind of graph homomorphism. Such languages naturally encompass conceptual graphs or RDF, but also propositional logics, constraint networks or positive Datalog.

6.2.1. Graph homomorphisms and entailment

Participants: Jean-François Baget [Contact], Faisal Al-Khateeb.

Proofs of soundness and completeness results for entailment in both the conceptual graphs and RDF formalisms have been unified into a larger framework [11]. We have identified similarities, not only between the algorithms used to compute entailment (different kinds of graph homomorphisms), but also in the way their semantics are expressed. In order to identify the logics that can benefit from this property, we obtained the following results:

- For any logic L, the consequence relation is a preorder, and is equivalent to a lattice if L is conjunctive.
- For any preorder R, we can build a logic L whose logical consequence relation is R, and where the interpretations are formulas of the language (we call such a logic a self-interpreted logic).
- Any preorder on a countable set of objects is a suborder of graph homomorphism.

Any conjunctive logic having a countable set of formulas is thus equivalent to a self-interpreted logic where both truth values and entailment are expressed via a graph homomorphism. However, the third item above being non constructive, such an equivalent logic cannot be automatically generated. We have instead adopted another approach, where the semantics of a language is described via graph homomorphisms.

Consider a logic, i.e., a set of formulas, a set of interpretations, and a relation expressing which interpretations are models for a logic.

- Both formulas and model-theoretic interpretations can be encoded as graphs (respectively formula graphs and interpretation graphs)
- Assume a graph homomorphism, called a projection (a mapping of nodes fulfilling particular properties) such that:

– The interpretation *I* is a model of the formula *f* if and only if the graph of *f* projects into the graph of *I*;
– For every satisfiable formula *f*, there exists a model *I* of *f* such that the graph of *I* projects into the graph of *f*.

In that case, the projection considered is a sound and complete entailment mechanism for the logic involved.

Through this framework, we have characterized a class of logics, including RDF or conceptual graphs, whose semantics can be expressed via a graph homomorphism. The main interest of this approach is to automatically generate a decision procedure for the entailment relation from the description of the semantics (as graph homomorphism computation).

6.2.2. Path RDF as a query language for RDF

Participants: Faisal Al-Khateeb [Contact], Jean-François Baget.

Querying RDF graphs can be reduced to computing entailment and a popular way to compute entailment is to find a projection from an RDF graph to another. Another way to query RDF graphs is to query for path expressed by regular expressions holding between nodes. The two kinds of queries do not identify the same set of queries (the former allows for full graphs branching and cycling as queries, the latter allows for indetermined lengths of paths).

We have introduced Path RDF as a simple extension of RDF in which relations can be replaced by regular path expressions over relations. This allows to express that there is a path combining relations in a particular way between two resources. For instance, one can query that there is transportation means by any combination of bus or train between two cities. We have provided the syntax and semantics of Path RDF and provided an extension of the projection operations between graphs that can be used between Path RDF graphs.

By using the above-mentioned framework, we have shown that the new projection is sound and complete for computing entailment (and hence answering queries) when a query in Path RDF is evaluated against an RDF graph. This projection is also sound when used between two Path RDF graphs (hence computing query containment). In general, this projection operation is not complete. However, we have shown that if the query is grounded (there is no blank node), then the projection is complete as well.

We are working towards extending Path RDF projection to wider classes and to take into account RDFS and OWL fragments.

6.3. Ontology alignment

In the context of the Knowledge web network of excellence, we are developing activities for structuring the research on ontology alignment at the European level. This has led to numerous activities and results.

6.3.1. Benchmarking

Participants: Jérôme Euzenat [Contact], Marc Ehrig [Karlruhe Universität].

Like last year [13], we have organized an ontology alignment evaluation for assessing the degree of achievement of actual ontology alignment algorithms [5][14][22]. This year's event has been held in Banff, Canada. Seven different teams entered the evaluation which consisted in three sets of tests (ontology sampling benchmark, directory alignment and large expressive ontology matching in the domain of anatomy).

In order to evaluate the performance of ontology matching algorithms it is necessary to confront them with test ontologies and to compare the results. The most prominent criteria are precision and recall originating from information retrieval. However, it can happen that an alignment be very close to the expected result and another quite remote from it and they both share the same precision and recall. This is due to the inability of precision and recall to measure the proximity of the results. To overcome this problem, we have proposed a framework for generalizing precision and recall [12].

This framework replaces the intersection of retrieved and expected sets by some expression of a similarity between these sets. This similarity must satisfy some constraints (preserving the results obtained by precision and recall, being maximized by the size of the largest set) that guarantee that precision/recall is one of its instances.

We have instantiated the framework with three different measures and have shown in a motivating example that the proposed measures are prone to solve the problem of rigidity of classical precision and recall. We are now working on providing measures that are semantically justified.

6.3.2. Towards an alignment algebra

Participants: Antoine Zimmermann [Contact], Jérôme Euzenat.

We have designed a theoretical framework to formally define ontology alignments and operations that can be done with them. This framework, based on category theory, considers ontologies and ontology alignments as first order objects independently of the representation language, as opposed to entity based definitions. Indeed, if ontologies are objects in a category, and morphisms correspond to ontology refinements, then an ontology alignment is a categorical relation between two ontologies, i.e. a pair of morphisms with the same domain. So, if $f: A \rightarrow O$ and $f': A \rightarrow O'$ are both ontology refinements, then $\langle A, f, f' \rangle$ is an alignment of O and O', A approximates ontologies O and O' and it describes a part of the knowledge that is common to them. We call this structure a V-alignment. With such a characterization, an algebra has been designed which describes what is ontology merging, alignment comparison, composition, union and intersection using well known categorical constructions.

Concrete categories of ontologies exist in the literature, but fail to express complex alignments, e.g. alignments expressing subsumption relation between concepts of two ontologies. To solve this problem, we investigated two approaches: defining more complex categories or improving the structure of our categorical alignments. On the one hand, more complex categories enhance the expressivity of the alignments, but the categories lose some interesting properties such as the existence of the merge for any V-alignment and pair of ontologies. On the other hand, simpler categories, such as the category of theories and theory morphisms in institution theory, can describe complex alignments if the alignment structure is more elaborate. In collaboration with the University of Karlsruhe, we introduced W-alignments: a structure having an additional ontology containing bridge axioms relating to O and O' by two V-alignments [21]. This raises the expressivity of alignments while keeping desirable properties of the category. However, the resulting algebra is less natural (composition is not associative).

6.3.3. Similarity for ontology alignment

Participants: Jérôme Euzenat [Contact], Petko Valtchev [Université de Montréal].

In order to be able to align ontologies written in OWL-Lite, we developed an algorithm (OLA), adapted from a method for measuring object-based similarity. OLA relies on a universal measure for comparing the entities of two ontologies that combines in a homogeneous way all the knowledge used in entity descriptions: it deals successfully with external data types, internal structure of classes as given by their properties and constraints, external structure of classes as given by their relationships to other classes and the availability of individuals. This is an improvement over other methods that usually take advantage of only a subpart of the language features.

The proposed method does not only compose linearly individual methods for assessing the similarity between entities, it uses an integrated similarity definition that makes them interact. OLA is based on the definition of a distance between entities of two ontologies as a system of equations that has to be solved in order to extract an alignment.

These equations are parameterized by a number of weigths corresponding to the respective importance of different components of ontologies. This year we introduced a preprocessing step which considers the ontologies to align and evaluate the availability of the corresponding features in order to choose the corresponding weights [15]. This automated weight adaptation has been instrumental in providing good results for OLA at the OAEI-2005 campaign.

This work is developed in collaboration with the University of Montréal.

6.4. Dynamic context management in pervasive computing

Participants: Jérôme Pierson [Contact], Jérôme Euzenat.

In a pervasive computing environment, the environment itself is the interface between services and user. Using context information coming from sensors, location technologies and aggregation services, applications adapt their run-time behaviour to the context in which users evolve (e.g., physical location, social or hierarchical position, current tasks as well as related information). These applications have to deal with the dynamic integration in the environment of new elements (users or devices), and the environment has to provide context information to newly designed applications. We study and develop a dynamic context management system for pervasive application. It must be flexible enough to be used by heterogeneous applications and to run dynamically with new incoming devices.

We have designed an architecture in which context information is distributed in the environment. Each device or service implements a context manager component in charge of maintaining its local context. It can communicate with other context manager components: some of them are context information producers, some of them are context information consumers and some of them are both. We have defined a simple protocol to allow a consumer to identify and determine the producer for the information it needs. Context manager components express their context information using a OWL ontology, and exchange RDF triples with each other. The openness of ontology description languages makes possible the extension of context descriptions and ontology matching helps dealing with independently developed ontologies. Thus, this architecture allows the introduction of new components and new applications without interrupting what is working.

We contributed the development of a demonstrator application for pervasive computing called ambient instant messaging. An appartment is divided in three zones. In each zone, there are different user interface devices. An instant messaging application (such as Jabber, MSN, or ICQ) is accessible to users in each room with different modalities and interfaces corresponding to their locations and activities. We worked on a centralized management system for context information used in this demonstrator. It takes information from distributed sensors in the environment and from a location system, and it provides user "chatting" activity and communication status to applications.

The next step will be to use an alignment service in order to be able to match application needs in terms of context information with that provided by the other devices.

This work is developed in collaboration with France Telecom R&D.

7. Contracts and Grants with Industry

7.1. France Telecom R&D

Participants: Jérôme Pierson [Contact], Jérôme Euzenat.

The work of Jérôme Pierson is developed in collaboration with the Centre Norbert Segard of France Telecom R&D. Jérôme Pierson's thesis is jointly supervised by Fano Ramparany and Jérôme Euzenat under a CIFRE contract. The topic of this thesis is to investigate the notion of context in ambient computing and the ability to match ontologies when shifting context (see §6.4).

8. Other Grants and Activities

8.1. European initiatives

8.1.1. Knowledge web network of excellence: realising the semantic web

Participants: Jérôme Euzenat [Contact], Faisal Al-Khateeb, Jean-François Baget, Antoine Zimmermann.

Exmo, as part of INRIA is a founding and active member of the Knowledge web Network of Excellence. The INRIA node comprises the *Acacia*, *Exmo* and *Orpailleur* teams. *Exmo* is in charge of the work package 2.2 on

"Heterogeneity" (see §6.3). Jérôme Euzenat is vice-scientific director of Knowledge web and represents INRIA in the executive management board.

More information on *Knowledge web* can be found at http://knowledgeweb.semanticweb.org.

8.1.2. Collaboration with University of Karlsruhe

Participants: Jérôme Euzenat [Contact], Antoine Zimmermann.

In the context of the T-Rex programme of Knowledge web, Marc Ehrig from university of Karlsruhe visited *Exmo* during the month of May. With Marc we have organized the Ontology Alignment Evaluation Intiative 2005 held in Banff in October and we have worked on introducing generalization of precision and recall measures for evaluating ontology alignments (see §6.3.1).

Markus Kröszch came visiting *Exmo* in April and Antoine Zimmermann visited Karlsruhe in June.

8.1.3. Collaboration with University of Trento

Participant: Jérôme Euzenat [Contact].

Following last year visit of Pavel Shvaiko, we carry on collaborating on foundations of ontology matching. In particular, we developed a tutorial that has been produced at the European Semantic Web Conference 2004 and a review paper [10].

8.2. International Initiatives

8.2.1. W3C Data Access Working Group

Participants: Jean-François Baget [Contact], Faisal Al-Khateeb.

The Data Access Working Group is one of the two working groups (along with the Best Practice Group) that form the "Semantic Web Activity" of the World Wide Web Consortium (W3C). It is chartered to develop a query language (named SPARQL) and protocol for querying RDF and OWL data. Jean-François Baget represents INRIA in this working group. The work on Path RDF is a continuation of the SPARQL development.

More information on RDFDAG can be found at http://www.w3.org/2001/sw/DataAccess/.

8.2.2. Collaboration with University of Montréal

Participant: Jérôme Euzenat [Contact].

Exmo collaborates with the team of Petko Valtchev at the university of Montréal (DIRO) on the topic of "Ontology alignment".

Petko Valtchev visited *Exmo* in May 2005 and Jérôme Euzenat visited University of Montréal in September 2005. This resulted in good results at the Ontology Alignment Evaluation Initiative 2005 with the OLA algorithm (see §6.3.3).

9. Dissemination

9.1. Leadership within scientific community

- Jérôme Euzenat is elected member of the board of the French artificial intelligence society (AFIA).
- Jérôme Euzenat is founding member of the "Semantic web science association" (steering committee for the ISWC conference series), member of the steering committee of the LMO conference series and member of the steering committee for the RFIA 2006 conference.
- Jérôme Euzenat has been invited to the brainstorming workshop on efforts worth supporting in FP7 "Knowledge and Content Technologies" in Galway.
- Jérôme Euzenat is vice-scientific director and coordinator of "Heterogeneity" work package of the Knowledge web network of excellence (see §8.1) (2004-2007).

9.2. Editorial boards, conference and workshop committees

- Jérôme Euzenat has been the general chair of the "European Semantic Web Conference (ESWC)" (Heraklion, GR), May 28th-June 1st, 2005 [7].
- Editorial board of the journal "L'objet", "Journal électronique d'intelligence artificielle (JEDAI)", "Journal of Web Semantics" and "Journal of Data Semantics" (Jérôme Euzenat).
- Program committee for the 2005 issues of the conferences "International Semantic Web Conference (ISWC)", "World-wide web conference (WWW)", "Symposium on Abstraction, Reformulation and Approximation" (SARA), "Langages et modèles à objets" (LMO), "Foundational aspects of ontologies" workshop at the 28th Künstliche Intelligenz conference.

9.3. Conferences, meetings and tutorial organization

- Organizer (with Mikalai Yatsewitch and Heiner Stuckenschmidt) of the Ontology Alignment Evaluation Initiative 2005 at the Integrating ontologies" workshop at K-Cap conference 2005, Banff (CA), October 2005 [14].
- Organizer (with Pavel Shvaiko, Alain Léger, Deborah McGuinness and Holger Wache) for the AAAI-2005 Context and ontologies workshop, Pittsburgh (PA US), 2005 [6]
- Organizer (with Marc Ehrig, Todd Hugues and Heiner Stuckenschmidt) of the "Ontology integration" workshop in the 3rd conference on Knowledge Capture (K-Cap), Banff (CA), 2005 [5].
- Organizer (Antoine Zimmermann) of the SDK Ontology Working Group meeting, 7-8 April 2005 in Grenoble (SDK is the cluster of three FP6 European projects: SEKT, DIP and Knowledge web).

9.4. Invited conferences and other talks

- Ontology reconciliation for the semantic web and agents, invited talk, AgentLink Technical forum, Ljubljana (SL), 1st March 2005
- De la sémantique dans le web sémantique, Séminaire Talana, Paris (FR), 23 May 2005
- Participer à et faire émerger un web sémantique, Séminaire Thalès-communications, Colombes (FR), 24 May 2005
- Opportunities and Challenges ahead for the Semantic Web, Semantic web days, München (DE), 7 October 2005

9.5. Teaching

- Co-ordination of the Artificial intelligence profile of the second year of mathematics and informatics master, Intelligence, Interaction, Information track -Joseph Fourier university and INPG - Grenoble - resp. Yves Demazeau
- Schema and ontology matching tutorial: European Semantic Web Conference - Hersounisous (GR) - 3h- resp. Jos De Bruyn, September 28-29th (Jérôme Euzenat and Pavel Shvaiko)
- Web, Connaissance, Sémantique: Lecture (18h), Master "Mathématiques et informatique" research option, "Intelligence, interaction, information" track, Université Joseph-Fourier-INP Grenoble (Jean-François Baget and Jérôme Euzenat)

9.6. Miscellaneous

- Development and maintainance of the web site (http://jedai.afia-france.org) for the Journal électronique d'intelligence artificielle (JEDAI).
- Jérôme Euzenat is redactor-in-chief (with Philippe Morignot) of the Bulletin de l'AFIA.

10. Bibliography

Major publications by the team in recent years

[1] ROLAND DUCOURNAU, JÉRÔME EUZENAT, GÉRALD MASINI, AMEDEO NAPOLI (editors). *Langages et modèles à objets: état et perspectives de la recherche*, collection "Didactique", vol. 19, INRIA, Rocquencourt (FR), 1998, http://www.inria.fr/rrrt/d-019.html.

[2] JÉRÔME EUZENAT. *An API for ontology alignment*, in "Proc. 3rd conference on international semantic web conference (ISWC), Hiroshima (JP)", FRANK VAN HARMELEN, SHEILA MCILRAITH, DIMITRIS PLEXOUSAKIS (editors). , 2004, in "Lecture notes in computer science", Springer Verlag, Heildelberg (DE), vol. 3298, p. 698-712, 2004, http://www.springerlink.com/index/DY8Y9F31A9GT9762.

[3] JÉRÔME EUZENAT. *Représentations de connaissance : de l'approximation à la confrontation*, Habilitation à diriger des recherches, Université Joseph Fourier, Grenoble (FR), janvier 1999, ftp://ftp.inrialpes.fr/pub/sherpa/theses/hdr-euzenat.ps.gz.

[4] JÉRÔME EUZENAT, NABIL LAYAÏDA, VICTOR DIAS. *A semantic framework for multimedia document adaptation*, in "Proc. 18th International Joint Conference on Artificial Intelligence (IJCAI), Acapulco (MX)", p. 31-36, Morgan Kauffman, San-Mateo (CA US), 2003.

Books and Monographs

[5] BEN ASHPOLE, MARC EHRIG, JÉRÔME EUZENAT, HEINER STUCKENSCHMIDT (editors). *Proceedings K-Cap 2005 integrating ontology workshopProc. K-Cap 2005 workshop on Integrating ontologies, Banff (CA)*, 2005, 2005, http://ceur-ws.org/Vol-156/.

[6] PAVEL SHVAIKO, JÉRÔME EUZENAT, ALAIN LÉGER, DEBORAH MCGUINNESS, HOLGER WACHE (editors). *Context and ontologies: theory and practiceProc. workshop on Context and ontologies: theory and practice, Pittsburg (PA US)*, 2005, 2005, http://www.cando.org.

[7] ASUNCIÓN GÓMEZ PÉREZ, JÉRÔME EUZENAT (editors). *The semantic web: research and applications*, 2005, in "Lecture notes in computer science", Springer verlag, Berlin (DE), vol. 3532, 2005, http://www.springeronline.com/3-540-26124-9.

Articles in refereed journals and book chapters

[8] JÉRÔME EUZENAT, ANGELO MONTANARI. *Time granularity*, MICHAEL FISHER, DOV GABBAY, LLUIS VILA (editors). , 2005, Elsevier, Amsterdam (NL), p. 59-118, 2005, ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2005a.pdf.

[9] JÉRÔME EUZENAT. *L'annotation formelle de documents en (8) questions*, RÉGINE TEULIER, JEAN CHARLET, PIERRE TCHOUNIKINE (editors). , 2005, L'Harmattan, Paris (FR), p. 251-271, 2005, ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2005c.pdf.

[10] PAVEL SHVAIKO, JÉRÔME EUZENAT. *A survey of schema-based matching approaches*, in "Journal on data semantics", Springer Verlag, Heildelberg (DE), vol. 4, p. 146-171, 2005, http://dx.doi.org/10.1007/11603412_5.

Publications in Conferences and Workshops

[11] JEAN-FRANÇOIS BAGET. *RDF Entailment as a Graph Homomorphism*, in "Proc. 4th conference on international semantic web conference (ISWC), Galway (EI)", YOLANDA GIL, ENRICO MOTTA, RICHARD BENJAMINS, MARK MUSEN (editors). , 2005, in "Lecture notes in computer science", Springer Verlag, Heildelberg (DE), vol. 3729, 82=96, 2005, ftp://ftp.inrialpes.fr/pub/exmo/publications/baget2005a.pdf.

[12] MARC EHRIG, JÉRÔME EUZENAT. *Relaxed Precision and Recall for Ontology Matching*, in "Proc. K-Cap 2005 workshop on Integrating ontology, Banff (CA)", p. 25-32, 2005, http://ceur-ws.org/Vol-156/paper5.pdf.

[13] JÉRÔME EUZENAT. *Evaluating ontology alignment methods*, in "Proc. Dagstuhl seminar on Semantic interoperability and integration, Wadern (DE)", n° 04391, 2005, http://drops.dagstuhl.de/opus/volltexte/2005/36.

[14] JÉRÔME EUZENAT, HEINER STUCKENSCHMIDT, MIKALAÏ YATSKEVICH. *Introduction to the Ontology Alignment Evaluation 2005*, in "Proc. K-Cap 2005 workshop on Integrating ontology, Banff (ALB CA)", p. 61-71, 2005, http://ceur-ws.org/Vol-156/paper10.pdf.

[15] JÉRÔME EUZENAT, PHILIPPE GUÉRIN, PETKO VALTCHEV. *OLA in the OAEI 2005 alignment contest*, in "Proc. K-Cap 2005 workshop on Integrating ontology, Banff (CA)", p. 97-102, 2005, http://ceur-ws.org/Vol-156/paper15.pdf.

[16] JÉRÔME EUZENAT. *Alignment infrastructure for ontology mediation and other applications*, in "Proc. 1st international workshop on Mediation in semantic web services, Amsterdam (NL)", p. 81-95, 2005, http://ceur-ws.org/Vol-168/MEDIATE2005-paper6.pdf.

[17] JASON JUNG, INAY HA, GEUN-SIK JO. *BlogGrid: towards an efficient information pushing service on blogspaces*, in "Proc. 4th conference on International Conference on Grid and Cooperative Computing (GCC), Beijing (CN)", HAI ZHUGE, GEOFFREY FOX (editors). , 2005, in "Lecture notes in computer science", Springer Verlag, Heildelberg (DE), vol. 3795, p. 178-183, 2005, ftp://ftp.inrialpes.fr/pub/exmo/publications/jung2005a.pdf.

[18] SÉBASTIEN LABORIE, JÉRÔME EUZENAT, NABIL LAYAÏDA. *Adapter temporellement un document SMIL*, in "Proc. atelier on Connaissance et document temporel, Nice (FR)", p. 47-58, 2005.

Internal Reports

[19] JÉRÔME EUZENAT, LOREDANA LAERA, VALENTINA TAMMA, ALEXANDRE VIOLLET. *Negociation/argumentation techniques among agents complying to different ontologies*, deliverable, 2005, n° 2.3.7, Knowledge web NoE, , IST Knowledge web NoE, 2005.

[20] JÉRÔME EUZENAT, FRANÇOIS SCHARFFE, LUCIANO SERAFINI. *Specification of the delivery alignment format*, deliverable, 2005, n° 2.2.6, Knowledge web NoE, , IST Knowledge web NoE, 2005.

[21] PASCAL HITZLER, JÉRÔME EUZENAT, MARKUS KRÖTZSCH, LUCIANO SERAFINI, HEINER STUCKENSCHMIDT, HOLGER WACHE, ANTOINE ZIMMERMANN. *Integrated view and comparison of alignment semantics*, deliverable, 2005, n° 2.2.5, Knowledge web NoE, , IST Knowledge web NoE, 2005, ftp://ftp.inrialpes.fr/pub/exmo/reports/kweb-225.pdf.

[22] HEINER STUCKENSCHMIDT, MARC EHRIG, JÉRÔME EUZENAT, ANDREAS HESS, ROBERT VAN HAGE, WEI HU, NINGSHENG JIAN, GONG CHEN, YUZHONG QU, GEORGE STOILOS, GIORGO STAMOU, UMBERTO STRACCIA, VOJTECH SVATEK, RAPHAËL TRONCY, PETKO VALTCHEV, MIKALAÏ YATSKEWICH. *Description of alignment implementation and benchmarking results*, deliverable, 2005, n° 2.2.4, Knowledge web NoE, , IST Knowledge web NoE, 2005.

Miscellaneous

[23] FAISAL AL-KHATEEB, JEAN-FRANÇOIS BAGET, JÉRÔME EUZENAT. *Complex Path Queries for RDF Graphs*, in "Proc. 6th poster session on , Galway (IE)", p. PID-52, 2005, ftp://ftp.inrialpes.fr/pub/exmo/publications/alkhateeb2005a.pdf.

[24] MARC EHRIG, JÉRÔME EUZENAT. *Generalizing precision and recall for evaluating ontology matching*, in "Proc. 6th poster session on , Galway (IE)", p. PID-54, 2005, ftp://ftp.inrialpes.fr/pub/exmo/publications/ehrig2005b.pdf.

[25] PAVEL SHVAIKO, JÉRÔME EUZENAT. *Ontology Matching*, 2005, in "DLib magazine", CNRI, Preston (VA US), vol. 12, n° 11, 2005, http://www.dlib.org/dlib/december05/12inbrief.html#PAVEL.

Bibliography in notes

[26] FRANZ BAADER, DIEGO CALVANESE, DEBORAH MCGUINNESS, DANIELE NARDI, PETER PATEL-SCHNEIDER (editors). *The description logics handbook*, Cambridge university pressCambridge (UK), 2003.

[27] JAMES CLARK (editor). *XSL Transformations (XSLT)*, Recommendation, 1999, W3C, http://www.w3.org/TR/xslt.

[28] TIM BERNERS-LEE, JAMES HENDLER, ORA LASSILA. *The semantic web*, in "Scientific american", vol. 279, n° 5, p. 35-43, 2001.