UMR IRISA

Project-Team PILGRIM

Graduality, Imprecision, and Mediation in Database Management Systems

Lannion

Activity Report

2013

1 Team

Head of the team

Olivier Pivert, Professor, Enssat

Administrative assistant

Joëlle Thépault, Enssat, 20%
Nelly Vaucelle, Enssat, 10%

Université Rennes 1 personnel

François Goasdoué, Professor, Enssat (since September 1, 2013)
Hélène Jaudoin, Associate Professor, Enssat
Ludovic Liétard, Associate Professor, HdR, IUT Lannion
Daniel Rocacher, Professor, Enssat
Grégory Smits, Associate Professor, IUT Lannion
Virginie Thion, Associate Professor, Enssat (since September 1, 2013)

PhD students

Katia Abbaci, Région Bretagne grant and ANR grant, since November 2009
William Correa Beltran, Région Bretagne grant and Conseil Général 22 grant, since October 2012

2 Overall Objectives

2.1 Introduction

In database research, the last two decades have witnessed a growing interest in preference queries on the one hand, and uncertain databases on the other hand.

Motivations for introducing preferences inside database queries are manifold. First, it has appeared to be desirable to offer more expressive query languages that can be more faithful to what a user intends to say. Second, the introduction of preferences in queries provides a basis for rank-ordering the retrieved items, which is especially valuable in case of large sets of items satisfying a query. Third, on the contrary, a classical query may also have an empty set of answers, while a relaxed (and thus less restrictive) version of the query might be matched by items in the database.

Approaches to database preference queries may be classified into two categories according to their qualitative or quantitative nature. In the qualitative approach, preferences are defined through binary preference relations. Among the representatives of this family of approaches, let us mention an approach based on CP-nets, and those relying on a dominance relation, e.g. Pareto order, in particular Skyline queries. In the quantitative approach, preferences are expressed quantitatively by a monotone scoring function (the overall score is positively correlated with partial scores). Since the scoring function associates each tuple with a numerical score, tuple t_1 is preferred to tuple t_2 if the score of t_1 is higher than the score of t_2. Well-known representatives of this family of approaches are top-k queries, and *fuzzy-set-based approaches*. The team Pilgrim particularly studies the latter, and the line followed is to focus on:

1. various types of flexible conditions, including non-trivial ones,
2. the semantics of such conditions from a user standpoint,
3. the design of query languages providing flexible capabilities in a relational setting.

Basically, a fuzzy query involves linguistic terms corresponding to gradual predicates, i.e., predicates which are more or less satisfied by a given (attribute) value. In addition, these various terms may have different degrees of importance, which means that they may be connected by operators beyond conjunction and disjunction. For instance, in the context of a search for used vehicles, a user might say that he/she wants a *compact* car *preferably French*, with a *medium* mileage, *around* 6 k$, whose color is *as close as possible* to light grey or blue. The terms appearing in this example must be specified, which requires a certain theoretical framework. For instance, one may think that "*preferably* French" corresponds to a complete satisfaction for French cars, a lower one for Italian and Spanish ones, a still smaller satisfaction for German cars and a total rejection for others. Similarly, "*medium* mileage" can be used to state that cars with less than 40000 km are totally acceptable while the satisfaction decreases as the mileage goes up to 75000 km which is an upper bound. Moreover, it is likely that some of the conditions are more important than others (e.g., the price with respect to the color). In such a context, answers are ordered according to their overall compliance with the query, which makes a major difference with respect to usual queries.

In the previous example, conditions are fairly simple, but it turns out that more complex ones can also intervene. A particular attention is paid to conditions calling on aggregate functions together with gradual predicates. For instance, one may look for departments where *most* employees are *close* to retirement, or where the average salary of *young* employees is *around* $2500. Such statements have their counterpart in regular query language, such as SQL, and the specification of their semantics, when gradual conditions come into play, is studied in the project.

Along this line, the ultimate goal of the project is to introduce gradual predicates inside database query languages, thus providing flexible querying capabilities. Algebraic languages as well as more user-oriented languages are under consideration in both the original and extended relational settings.

As to the second topic mentioned at the beginning of this introduction, i.e., uncertain databases, it already has a rather long history. Indeed, since the late 70s, many authors have made diverse proposals to model and handle databases involving uncertain or incomplete data. In particular, the last two decades have witnessed a profusion of research works on this topic. The notion of an uncertain database covers two aspects: i) attribute uncertainty: when some attribute values are ill-known; ii) existential uncertainty: when the existence of some tuples is itself uncertain. Even though most works about uncertain databases consider probability theory as the underlying uncertainty model, some approaches — in particular those proposed by Pilgrim — rather rely on possibility theory. The issue is not to demonstrate that the possibility-theory-based framework is "better" than the probabilistic one at modeling uncertain databases, but that it constitutes an interesting alternative inasmuch as it captures a different kind of uncertainty (of a subjective, nonfrequential, nature). A typical example is that of a person who witnesses a car accident and who does not remember for sure the model of the car involved. In such a case, it seems reasonable to model the uncertain value by means of a possibility distribution, e.g., {1/Mazda, 1/Toyota, 0.7/Honda} rather than with a probability distribution which would be artificially normalized. In contrast with probability theory, one expects the following advantages when using possibility theory:

- the qualitative nature of the model makes easier the elicitation of the degrees attached to the various candidate values;
- in probability theory, the fact that the sum of the degrees from a distribution must equal 1 makes it difficult to deal with incompletely known distributions;
- there does not exist any probabilistic logic which is complete and works locally as possibilistic logic does: this can be problematic in the case where the degrees attached to certain pieces of data must be automatically deduced from those attached to some other pieces of data (e.g., when data coming from different sources are merged into a single database).

A recent research topic in Pilgrim concerns flexible data integration systems. One considers a distributed database environment where several data sources are available. An extreme case is that of a totally decentralized P2P system. An intermediary situation corresponds to the case where several global schemas are available and where the sources can be accessed through views defined on one of these schemas (LAV approach). The problem consists in handling a user query (possibly involving preferences conveyed by fuzzy terms) so as to forward it (or part of it) to the relevant data sources, after rewriting it using the views. The overall objective is thus to define flexible query rewriting techniques which take into account both the approximate nature of the mappings and the graded nature of the initial query. A large scale environment is aimed, and the performance aspect is therefore crucial in such a context.

3 Scientific Foundations

The project investigates the issues of flexible queries against regular databases as well as regular queries addressed to databases involving imprecise data. These two aspects make use of two close theoretic settings: fuzzy sets for the support of flexibility and possibility theory for the representation and treatment of imprecise information.

3.1 Fuzzy sets

Fuzzy sets were introduced by L.A. Zadeh in 1965 [Zad65] in order to model sets or classes whose boundaries are not sharp. This is particularly the case for many adjectives of the natural language which can be hardly defined in terms of usual sets (e.g., high, young, small,

[Zad65] L. Zadeh, "Fuzzy sets", *Information and Control 8*, 1965, p. 338–353.

etc.), but are a matter of degree. A fuzzy (sub)set F of a universe X is defined thanks to a membership function denoted by μ_F which maps every element x of X into a degree $\mu_F(x)$ in the unit interval $[0, 1]$. When the degree equals 0, x does not belong at all to F, if it is 1, x is a full member of F and the closer $\mu_F(x)$ to 1 (resp. 0), the more (resp. less) x belongs to F. Clearly, a regular set is a special case of a fuzzy set where the values taken by the membership function are restricted to the pair $\{0,\ 1\}$. Beyond the intrinsic values of the degrees, the membership function offers a convenient way for ordering the elements of X and it defines a symbolic-numeric interface. The α level-cut of a fuzzy set F is defined as the (regular) set of elements whose degree of membership is greater than or equal to α and this concept bridges fuzzy sets and ordinary sets.

Similarly to a set A which is often seen as a predicate (namely, the one appearing in the intensional definition of A), a fuzzy set F is associated with a gradual (or fuzzy) predicate. For instance, if the membership function of the fuzzy set *young* is given by: $\mu_{young}(x) = 0$ for any $x \geq 30$, $\mu_{young}(x) = 1$ for any $x < 21$, $\mu_{young}(21) = 0.9$, $\mu_{young}(22) = 0.8$, ... , $\mu_{young}(29) = 0.1$, it is possible to use the predicate *young* to assess the extent to which Tom, who is 26 years old, is young ($\mu_{young}(26) = 0.4$).

The operations valid on sets (and their logical counterparts) have been extended to fuzzy sets. Their definition assumes the validity of the commensurability principle between the concerned fuzzy sets. It has been shown that it is impossible to maintain all of the properties of the Boolean algebra when fuzzy sets come into play. Fuzzy set theory starts with a strongly coupled definition of union and intersection which rely on triangular norms ($\top$) and co-norms ($\bot$) tied by de Morgan's laws. Then:

$$\mu_{A \cap B}(x) = \top(\mu_A(x),\ \mu_B(x)) \qquad \mu_{A \cup B}(x) = \bot(\mu_A(x),\ \mu_B(x))$$

The complement of a fuzzy set F, denoted by $\bar{F}$, is a fuzzy set such that: $\mu_{\bar{F}}(x) = neg(\mu_F(x))$, where *neg* is a strong negation operator and the complement to 1 is generally used. The conjunction and disjunction operators are the logical counterpart of intersection and union while the negation is the counterpart of the complement.

In practice, minimum and maximum are the most commonly used norm and co-norm because they have numerous properties among which:

- the satisfaction of all the properties of the usual intersection and union (including idempotency and double distributivity), except excluded-middle and non-contradiction laws,
- they still work with an ordinal scale, which is less demanding than numerical values over the unit interval,
- the simplicity of the underlying calculus.

Once these three operators given, others can be extended to fuzzy sets, such as the difference:

$$\mu_{E-F}(x) = \top(\mu_E(x),\ \mu_{\bar{F}}(x))$$

and the Cartesian product:

$$\mu_{E \times F}(x, y) = \top(\mu_E(x),\ \mu_F(y)).$$

The inclusion can be applied to fuzzy sets in a straightforward way: $E \subseteq F \Leftrightarrow \forall x, \mu_E(x) \leq \mu_F(x)$, but a gradual view of the inclusion can also be introduced. The idea is to consider that E may be more or less included in F. Different approaches can be considered, among which one is based on the notion of a fuzzy implication (the usual logical counterpart of the inclusion). The starting point is the following definition valid for sets:

$$E \subseteq F \Leftrightarrow \forall x, x \in E \Rightarrow x \in F$$

which becomes :

$$deg(E \subseteq F) = \top_x(\mu_E(x) \Rightarrow_f \mu_F(x))$$

where $\Rightarrow_f$ is a fuzzy implication whose arguments and result take their value in the unit interval. Different families of such implications have been identified (notably R-implications and S-implications) and the most common ones are:

- Kleene-Dienes implication : $a \Rightarrow_{K-D} b = \max(1 - a,\, b)$,
- Rescher-Gaines implication: $a \Rightarrow_{R-G} b = 1$ if $a \leq b$ and 0 otherwise,
- Gödel implication : $a \Rightarrow_{Go} b = 1$ is $a \leq b$ and b otherwise,
- Łukasiewicz implication : $a \Rightarrow_{Lu} b = \min(1,\, 1 - a + b)$.

Of course, fuzzy sets can also be combined in many other ways, for instance using mean operators, which do not make sense for classical sets.

3.2 Possibility theory

Possibility theory is a theory of uncertainty which aims at assessing the realization of events. The main difference with the probabilistic framework lies in the fact that it is mainly ordinal and it is not related with frequency of experiments. As in the probabilistic case, a measure (of possibility) is associated with an event. It obeys the following axioms [Zad78]:

- $\Pi(X) = 1$,
- $\Pi(\oslash) = 0$,
- $\Pi(A \cup B) = \max(\Pi(A),\, \Pi(B))$,

where X denotes the set of all events and A, B are two subsets of X. If $\Pi(A)$ equals 1, A is completely possible (but not certain), when it is 0, A is completely impossible and the closer to 1 $\Pi(A)$, the more possible A. From the last axiom, it appears that the possibility of $\bar{A}$, the opposite event of A, cannot be calculated from the possibility of A. The relationship between these two values (for Boolean events) is:

$$max(\Pi(A),\, \Pi(\bar{A})) = 1$$

[Zad78] L. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems 1*, 1978, p. 3–28.

which stems from the first and third axioms (where B is replaced by $\bar{A}$).

In other words, if A is completely possible, nothing can be deduced for $\Pi(\bar{A})$. This state of fact has led to introduce a complementary measure (N), called necessity, to assess the certainty of A. $N(A)$ is based on the fact that A is all the more certain as $\bar{A}$ is impossible [DP80]:

$$N(A) = 1 - \Pi(\bar{A})$$

and the closer to 1 $N(A)$, the more certain A. From the third axiom on possibility, one derives:

$$N(A \cap B) = \min(N(A),\, N(B)))$$

and, in general:

- $\Pi(A \cap B) \leq \min(\Pi(A),\, \Pi(B))$,
- $N(A \cup B) \geq \max(N(A),\, N(B))$.

In the possibilistic setting, a complete characterization of an event requires the computation of two measures: its possibility and its certainty. It is interesting to notice that the following property holds:

$$\Pi(A) < 1 \Rightarrow N(A) = 0.$$

It indicates that if an event is not completely possible, it is excluded that it is somewhat certain, which makes it possible to define a total order over events: first, the events which are somewhat possible but not at all certain (from ($\Pi = N = 0$ to $\Pi = 1$ and $N = 0$), then those which are completely possible and somewhat certain (from $\Pi = 1$ and $N = 0$ to $\Pi = N = 1$). This favorable situation (existence of a total order) is valid for usual events, but if fuzzy ones are taken into account, this is no longer true (because A $\cup\ \bar{A} = X$ is not true in general when A is a fuzzy set) and the only valid property is: $\forall$ A, $\Pi(A) \geq$ N(A).

The notion of a possibility distribution [Zad78], denoted by π, plays a role similar to that of a probability distribution. It is a function from the referential X into the unit interval and:

$$\forall A \subseteq X, \Pi(A) = \sup_{x \in A} \pi(x)$$

In order to comply with the second axiom above, a possibility distribution must be such that there exists (at least) an element x_0 of X for which $\pi(x_0) = 1$. Indeed, a possibility distribution can be seen as a normalized fuzzy set F which represents the knowledge about a given variable. The following formula:

$$\pi(x = a) = \mu_F(a)$$

which is often used, tells that the possibility that the actual value of the considered variable x is a, equals the degree of membership of a to the fuzzy set F. For example, Paul's age may be only imprecisely known as "close to 20", where a given fuzzy set is associated with this fuzzy linguistic expression.

[DP80] D. Dubois, H. Prade, *Fuzzy set and systems: theory and applications*, Academic Press, 1980.

[Zad78] L. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems 1*, 1978, p. 3–28.

3.3 Fuzzy sets, possibility theory and databases

The project is situated at the crossroads of databases and fuzzy sets. Its main objective is to broaden the capabilities offered by DBMSs according to two orthogonal lines in order to separate two distinct problems:

- flexible queries against regular databases so as to provide users with a qualitative result made of ordered elements,
- Boolean queries addressed to databases containing imprecise attribute values.

Once these two aspects solved separately, the joint issue of flexible queries against databases containing imprecise attribute values will also be considered. This can be envisaged because of the compatibility between the semantics of grades (preferences) in both fuzzy sets and possibility distributions.

It turns out that fuzzy sets offer a very convenient way for modeling gradual concepts and then flexible queries. It has been proven [BP92] that many *ad hoc* approaches (e.g., based on distances) were special cases of what is expressible using fuzzy set theory. This framework makes it possible to express sophisticated queries where the semantic choices of the user can take place (e.g., the meaning of the terms or the compensatory interaction desired between the various fuzzy conditions of a query). The works conducted in Pilgrim aim at extending algebraic as well as user-oriented query languages in both the relational and the object-oriented (extended relational in practice) settings. The relational algebra has already been revised in order to introduce flexible queries and a particular focus has been put on the division operation. Current works are oriented towards:

- bipolar fuzzy queries (including two parts: one viewed as a constraint, the other as a wish),
- the use of a predefined fuzzy vocabulary (which raises the question of its adequacy wrt to the actual content of the database),
- fuzzy extensions of Skyline queries (based on Pareto order),
- implementation and query optimization issues.

As to possibility distributions, they are used to represent imprecise (imperfect) data. By doing so, a straightforward connection can be established between a possibilistic database and regular ones. Indeed, a possibilistic database is nothing but a weighted set of regular databases (called worlds), obtained by choosing one candidate in every distribution appearing in any tuple of every possibilistic relation. According to this view, a query addressed to a possibilistic database has a natural semantics. However, it is not realistic to process it against all the worlds due to their huge number. Then, the question tied to the querying of a possibilistic database bears mainly on the efficiency, which imposes to obviate the combinatory explosion of the worlds. The objective of the project is to identify different families of queries which comply with this requirement in the context of the relational setting, even if the initial model must obviously be extended (in particular to support imprecise data).

[BP92] P. Bosc, O. Pivert, "Some approaches for relational databases flexible querying", *Journal of Intelligent Information Systems 1*, 1992, p. 323–354.

3.4 Query rewriting using views

3.4.1 Data integration

Information integration is the problem of combining information residing at disparate sources and providing the user with a unified view of that information. This problem has been a long standing challenge for the database community.

Two main approaches for information integration have been proposed. In the first approach, namely materialization or warehousing, data are periodically extracted from the sources and stored in a centralized repository, called a (data) warehouse. User queries are posed and executed at the warehouse with no need to access the remote information sources. Such an approach is useful in the context of intra-enterprise integration with few remote sources to integrate. It is, however, not feasible in open environments like the Web where the number of sources may be very large and dynamic.

In the second approach, called mediation or virtual integration, data stay at the sources and are collected dynamically in response to user queries [Len02,Hal03]. Mediation architectures are based on the mediator/wrapper paradigm where native information sources are *wrapped* into logical views through which the underlying sources may be accessed. The views are stored in the mediator component which additionally contains an integrated global schema that provides a single entry point to query the available information sources. The global schema acts as an interface between the user queries and the sources, freeing the users from the problem of source location and heterogeneity issues. In such an architecture, user queries posed on the global schema are rewritten in terms of logical views and then sent to the remote sources.

Briefly stated, two main approaches of mediation have been investigated [Hal01]: the GAV (Global As View) approach where the global schema is expressed as a set of views over the data sources, and the LAV (Local As View) approach where the data sources are defined as views over the global schema. Query processing is expected to be easier in the GAV approach as it can be achieved by a kind of unfolding of original queries. However, this approach suffers from a lack of extensibility as changing or adding new sources affects the global schema. On the contrary, the LAV approach is known to be highly extensible in the sense that source changes do not impact the global schema. However, in the context of the LAV approach, query processing is known to be more challenging.

A centralized mediation approach has several drawbacks including scalability, flexibility, and availability of information sources. To cope with such limitations, a new decentralized integration approach, based on a Peer-to-Peer (P2P) architecture, has been proposed. A P2P data management system [HIM+04] enables sharing heterogeneous data in a distributed and scalable way. Such a system is made of a set of peers each of which is an entire data source with its own distinct schema. Peers interested in sharing data can define pairwise

[Len02] M. Lenzerini, "Data Integration : A Theoretical Perspective", *in: PODS*, Madison, Wisconsin, 2002.

[Hal03] A. Halevy, "Data Integration : A status Report", *in: German Database Conference BTW-03*, Leipzig, Germany, 2003. Invited Talk.

[Hal01] A. Y. Halevy, "Answering queries using views: A survey", *VLDB Journal 10*, 4, 2001, p. 270–294.

[HIM+04] A. Y. Halevy, Z. G. Ives, J. Madhavan, P. Mork, D. Suciu, I. Tatarinov, "The Piazza Peer Data Management System.", *IEEE Trans. Knowl. Data Eng. 16*, 7, 2004, p. 787–798.

mappings between their schemas. Users formulate queries over a given peer schema then a query answering system exploits relevant mappings to reformulate the original query into set of queries that enable to retrieve data from other peers.

3.4.2 Query answering in information integration systems

The problem of answering queries in mediation systems has been intensively investigated during the last decade. In particular, the investigation of this problem in the context of a LAV approach led to a great piece of fundamental theory. Recent works show that query processing in data integration is related to the general problem of answering queries using views [Hal01, Len02]. In such a setting, the semantics of queries can be formalized in terms of certain answers [AD98]. Intuitively, a certain answer to a query Q over a global (mediated) schema with respect to a set of source instances is an answer to Q in any database over the global schema that is consistent with the source instances. Therefore, the problem of answering queries in LAV-based mediation systems can be formalized as the problem of computing all the certain answers to the queries. As shown recently, this problem has a strong connection with the problem of query answering in database with incomplete information under constraints.

One of the common approaches to effectively computing query answers in mediation systems is to reduce this problem into a query rewriting problem, usually called *query rewriting using views* [Hal01,Len02,TH04]. Given a user query expressed over the global (or a peer) schema, the data sources that are relevant to answer the query are selected by means of a rewriting algorithm that allows to reformulate the user query into an equivalent or maximally subsumed (contained) query whose definition refers only to source descriptions.

The problem of rewriting queries in terms of views has been intensively investigated in the last decade (see [Hal01,Len02] for a survey). Existing research works differ w.r.t. the languages used to express a global schema, views and queries as well as w.r.t. the type of rewriting considered (i.e., maximally contained or equivalent rewriting). In a nutshell, this problem has been studied for different classes of languages ranging from various sub-languages of datalog, hybrid languages combining Horn rules and description logics to semistructured data models. Recently, the problem of rewriting queries in terms of views has been investigated in the context of P2P DBMSs [HIM+04,TH04] in order to ensure scalability in terms of the number of data sources. A few recent papers also contributed to the development of data integration systems capable of taking into account imprecision or uncertainty. Most of the works along that line use probability theory in order to capture the form of uncertainty that stems from the schema definition process, or that associated with the mere existence of data, or aim at modelling the approximate nature of the semantic links between the data sources and the mediated schema.

[AD98] S. Abiteboul, O. Duschka, "Complexity of Answering Queries Using Materialized Views.", *in: PODS*, p. 254–263, 1998.

[TH04] I. Tatarinov, A. Halevy, "Efficient query reformulation in peer data management systems", *in: SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, ACM Press, p. 539–550, New York, NY, USA, 2004.

4 Application Domains

Flexible queries have many potential application domains. Indeed, soft querying turns out to be relevant in a great variety of contexts, such as web search engines, yellow pages, classified advertisements, image or multimedia retrieval. One may guess that the richer the semantics of stored information (for instance images or video), the more difficult it is for the user to characterize his search criterion in a crisp way, i.e., using Boolean conditions. In this kind of situation, flexible queries which involve imprecise descriptions (or goals) and vague terms, may provide a convenient means for expressing information needs.

As for uncertain data management, many potential domains could take advantage of advanced systems capable of storing and querying databases where some pieces of information are imprecise/uncertain: military information systems, automated recognition of objects in images, data warehouses where information coming from more or less reliable sources must be fused and stored, etc.

In the near future, we intend to focus on two application domains:

- Open data management. One of the challenges in web data management today is to define adequate tools allowing users to extract the data that are the most likely to fulfill all or part of their information needs, then to understand and automatically correlate these data in order to elaborate relevant answers or analyses. Open data may be of various levels of quality: they may be imprecise, incomplete, inconsistent and/or their reliability/freshness may be somewhat questionable. An appropriate data model and suitable querying tools must then be defined for dealing with the imperfection that may pervade data in this context. On the other hand, it is of prime importance to provide end-users with simple and flexible means to better understand and analyze open data. The standards of W3C offer popular languages for representing both open and structured data. Another objective is to propose analytical tools suited to these languages through the construction of RDF data warehouses, whereas fuzzy-set-based data summarization approaches should constitute an important step towards making open data more intelligible to non-expert users. This work about open data management is expected to be carried out in collaboration with DGA and Semsoft. Another potential partner is Rennes Métropole, with which contacts have already been established.

- Environmental information systems. This work will be performed in collaboration with the Biological Station based in Roscoff (Finistère). The general objective is to define an information system architecture (along with an associated "toolbox") suited to the context of marine biodiversity monitoring and environmental protection. We intend to study three main aspects:

  – definition of a data warehouse model suited to this context, capable of dealing with missing values, imprecise information (a situation which often occurs due to the way data is collected and described, through sampling campaigns and human-performed labeling, in particular), uncertain data (uncertainty is unavoidable when data are obtained by means of predictive models, for instance).

– identification of new needs in terms of query expression: new OLAP operators suitable for the model, making it possible to handle dimensions described by fuzzy concept trees, to manage fuzzy cardinalities, possibility distributions and so on.
– knowledge discovery: we are notably interested in exploiting a concept that comes from artificial intelligence but has not been applied in the domain of data management yet: that of an analogical proportion, which underlies propositions of the type "A is to B as C is to D". We believe that discovering such "regularities" in a dataset could prove very useful for many purposes connected to environmental monitoring issues, in particular when it comes to predict the evolution of an ecosystem or the population of a species, etc.

5 Software

- PostgreSQLF is a flexible querying prototype that aims at evaluating fuzzy queries addressed to regular databases. It is an extension of PostgreSQL which implements the fuzzy query language SQLf defined in the team. This prototype is coupled with a graphical interface names ReqFlex [7] that makes it easy for an end user to specify his/her fuzzy queries.

- CORTEX (CORrelaTion-based Query EXpansion): Retrieving data from large-scale databases sometimes leads to plethoric answers especially when queries are underspecified. To overcome this problem, we proposed an approach which strengthens the initial query by adding new predicates (cf. Subsection 6.2.4). These predicates are selected among predefined ones principally according to their degree of semantic correlation with the initial query. This way, we avoid an excessive modification of its initial scope. Considering the size of the initial answer set and the number of expected results specified by the user, fuzzy cardinalities are used to assess the reduction capability of these correlated predefined predicates. This approach has been implemented as a research prototype, named CORTEX, to query a database containing 10,000 ads about second hand cars [BHPS10].

- LUCIFER (Leveraging Unveiled Conflicts In FlexiblE Requests): This prototype deals with conjunctive fuzzy queries that yield an empty or poorly satisfactory answer set. It implements a cooperative answering approach which efficiently retrieves the minimal failing subqueries of the initial query (which can then be used to explain the failure and revise the query) [PSHJ12].

- FALSTAFF (FAceted search engine Leveraging Summaries of daTA with Fuzzy Features): Faced with the difficulty of formulating precise queries to retrieve items from

[BHPS10] P. Bosc, A. Hadjali, O. Pivert, G. Smits, "CORTEX — CORrelaTion-based query EXpansion", *in: Actes des 26e Journées Bases de Données Avancées (BDA'10), session démonstration*, 2010.

[PSHJ12] O. Pivert, G. Smits, A. Hadjali, H. Jaudoin, "LUCIFER : Un système de détection de conflits dans les requêtes flexibles", *in: Actes de la 12e Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC'12)*, p. 617–620, 2012.

large scale databases, interactive interfaces implementing a faceted search strategy help the users navigate through the data by successively selecting facet-value pairs. This prototype uses a faceted search strategy to construct fuzzy queries. The interactive query construction process relies on precomputed metadata that informs about the data distribution over a predefined vocabulary [SP12].

- COKE (COnnected KEywords): Keyword queries have emerged as the most convenient way to query data sources especially for unexperienced users. Introduced initially for document retrieval on the web, such queries are defined as an enumeration of keywords corresponding to a rough description of what users are looking for. The interpretation process of keyword queries has then been adapted to handle structured data like relational databases or XML documents. Instead of considering queries as an unstructured enumeration of keywords, the approach underlying the COKE system lets users structure their keyword queries using simple but meaningful grammatical connectors. Using the data structure intensively, a COKE query is translated into SQL to retrieve exact answers. An autocompletion strategy is also proposed to help users take advantage of connectors in their keyword queries [SPJP13]. An experimentation shows that the COKE system efficiently retrieves more relevant and precise answers than classical queries made of keywords enumerations and offers a good coverage of possible query patterns.

6 New Results

6.1 Possibilistic database modeling and querying

Participants: Olivier Pivert.

Many works have been undertaken in the area of “fuzzy databases” in the last twenty years. This term is sometimes misused or misleading since it covers both fuzzy querying against regular databases and the handling of databases that are pervaded with imprecision or uncertainty in the data (as opposed to queries). In [4], we review different semantics that queries may have when they are addressed to a database where imperfect data is involved. We limit the scope to the case where imprecise data is represented in the framework of the possibilistic model, although the categorization proposed would apply to a large extent to probabilistic databases as well.

[SP12] G. Smits, O. Pivert, “A Fuzzy-Summary-Based Approach to Faceted Search in Relational Databases”, *in: Proc. of the 16th East-European Conference on Advances in Databases and Information Systems (ADBIS'12)*, T. Morzy, T. Haerder, R. Wrembel (editors), *LNCS*, *7503*, Springer, p. 357–370, 2012.

[SPJP13] G. Smits, O. Pivert, H. Jaudoin, F. Paulus, “An Autocompletion Mechanism for Enriched Keyword Queries to RDF Data Sources”, *in: Proc. of the 10th International Conference on Flexible Query Answering Systems (FQAS'13)*, 2013.

6.2 Flexible querying of classical databases

6.2.1 Preference queries

Participants: Olivier Pivert, Grégory Smits, Ludovic Liétard, Daniel Rocacher , Katia Abbaci.

The works presented hereafter deal with different aspects of preference queries (fuzzy and others) in a database context.

- *Handling bipolar fuzzy queries and relations.* In [3], we present an extension of relational algebra suitable for the handling of bipolar concepts. The type of queries considered involves two parts: a first one which expresses a (possibly flexible) constraint, and a second one that corresponds to a (possibly flexible) wish. The framework considered is that of bipolar fuzzy relations where each tuple is associated with a pair of satisfaction degrees. It is shown that the query equivalences that exist in the framework of classical relational algebra remain valid in the extended context considered. In [18], we pay a special attention to queries where an inclusion-based condition comes into play. This leads us to defining generalized inclusion indicators in the context of bipolar fuzzy relations. The paper proposes an interpretation of such indicators and investigates their properties as well as their practical use in a database querying context (for expressing division queries, in particular). An alternative interpretation of the division of bipolar fuzzy relations, based on a generalized R-implication, is presented in [24].

- *Skyline refinement.* Skyline queries are a popular and powerful paradigm for extracting interesting objects from a d-dimensional dataset. They rely on Pareto dominance principle to identify the skyline objects, i.e., the set of incomparable objects which are not dominated by any other object from the dataset. Two main problems may be faced when using such queries: (i) the skyline contains a small number of objects, which may be insufficient to meet the users' needs; and (ii) the skyline includes a huge number of objects, which is difficult for the users to exploit. In [10], the latter problem is tackled and an approach is proposed, whose principle consists in refining the skyline in order to discriminate its elements and select the best ones. A definition of a dominance relationship based on the fuzzy quantifier "almost all" is introduced.

- *Implementation.* In [21, 7, 20], we present an implementation strategy for a fuzzy querying system embedded in a regular DBMS. This system relies on the language SQLf that makes it possible to express a great variety of fuzzy queries. Experiments show that this implementation strategy induces performance gains with respect to existing strategies based on a loose (or milder) coupling between a fuzzy querying layer and a DBMS, that necessitate an external postprocessing so as to compute the result in the form of a fuzzy relation. We also describe a user-friendly interface aimed at helping nonexpert users express their fuzzy queries in an intuitive manner (using graphical tools).

- *Adequacy of a user-defined fuzzy vocabulary.* Clustering methods are of a particular interest to discover and to summarize the structure of a data set. However, interpreting clusters may be abstruse for unexperienced users who most of the time possess their own vocabulary to describe data and properties. In [16, 17], an approach is proposed to determine and quantify how appropriate a user-defined vocabulary is regarding the structure captured on the data distribution using a clustering method. Two measures of vocabulary appropriateness based on clustering are proposed and tested on artificial data.

6.2.2 Cooperative answering to flexible database queries

Participants: Grégory Smits, Olivier Pivert, Hélène Jaudoin.

The practical need for endowing information systems with the ability to exhibit cooperative behavior (thus making them more "intelligent") has been recognized at least since the early 90s. The main intent of cooperative systems is to provide correct, non-misleading and useful answers, rather than literal answers to user queries. Different aspects of this problem are tackled in the works presented hereafter.

- *Empty and plethoric sets of answers.* Cooperative approaches to relational database querying help users retrieve the tuples that are the most relevant with respect to their information needs. In [8], we propose a unified framework that relies on a fuzzy cardinality-based summary of the database. We show how this summary can be efficiently used to explain failing queries or to revise queries returning a plethoric answer set.

- *Association-based retrieval of similar objects.* [19] deals with the issue of extending the scope of a user query in order to retrieve objects which are similar to its "strict answers". The approach proposed exploits associations between database items, corresponding, e.g., to the presence of foreign keys in the database schema. Fuzzy concepts such as typicality, similarity and linguistic quantifiers are at the heart of the approach and make it possible to obtain a ranked list of similar answers.

6.3 Flexibility issues in data integration systems

Participants: François Goasdoué, Hélène Jaudoin, Olivier Pivert, Ludovic Liétard, Daniel Rocacher, Grégory Smits.

- *Fuzzy ontology for database querying with bipolar preferences.* [9] deals with the expression and the evaluation of complex user preferences in the context of distributed and heterogeneous information systems. Complex preferences are modeled by fuzzy bipolar conditions, which associate negative and positive conditions. Bipolar queries are addressed to information systems such as web applications, built on several distributed and heterogeneous databases. This querying can lead to process huge volumes of data and can deliver massive responses, in which it may be difficult for the user to distinguish the relevant answers from irrelevant ones. Semantic aspects make it possible to address this problem by providing a personalized data access method, so that only the most relevant data are targeted to evaluate queries. [9] presents a new approach for flexible querying of complex information systems that combines a reasoning mechanism (an ontology based on a fuzzy bipolar extension of the description logic DLR-Lite) with a bipolar relational language of a high expressivity (Bipolar SQLf language). The reasoning mechanism can also answer queries in an approximate way, based on degrees expressing at which extent it is possible to substitute a concept in the query with other concepts, while still meaningful to the user. In [23], this approach is extended to bipolar conditions of type "or else".

- *Enriched keyword queries.* [22] introduces a novel keyword-based query solution for end users in order to retrieve precise answers from semantic data sources. Contrary to existing approaches, connectives corresponding to linking words or verbal structures from natural languages are used inside queries to specify the meaning of each keyword, thus leading to a complete and explicit definition of the intent of the search. In order to help users formulate such connected keywords queries and to translate them into SPARQL, an interactive mechanism based on autocompletion has been developed, which is presented in [22].

- *Ontology-based Web data integration.* In [5], we define an XML-RDF hybrid framework, called XR, for managing XML documents with fine-grained interlinked RDF annotations. In particular, we allow putting *any RDF description* on *any piece (aka node) of an XML document*. Hence, we can have RDF annotations interlinking tree databases, RSS feeds, XHTML pages, ODF documents, etc. To query uniformly our annotated documents, we also define a query language, called XRQ, integrating the Tree-Pattern Queries (TPQs) of the standard XML query language XQuery, and the BGP queries of SPARQL for RDF. We thus allow querying annotated documents wrt both their structure (with XML tree-shape constraints) and their semantics (with RDF ontological constraints). Finally, we investigate how such a uniform XML-RDF setting can be used for optimizing query answering against annotated documents, by pushing some information from XML to RDF, or RDF to XML, during query processing. An XR data management platform has been developed and demonstrated on a fact-checking scenario [13, 14]. See also [2].

  In [6], we propose the new notion of *robust* module-based data management; we devise it for DL-lite, the description logic family that underlying OWL2 QL, the W3C standard for scalable data management through rather expressive ontologies. Module-based data management amounts to handling data using an ontology – a *module* – that derives from that of a preexisting ontology-based data management application. Adding *robustness* properties to a module allows using it to enhance data integrity and to complement the answers to queries in ontology-based data integration applications.

  In [12], we present preliminary results for redesigning core data analytics concepts and tools in the context of the RDF data model, to provide the first complete formal framework for warehouse-style RDF analytics. A prototype implementing our results has been demonstrated in [12].

  In [15], we have devised techniques for efficiently querying RDF databases. [11] is a preliminary extension of this work considering huge RDF datasets (possibly integrated from many sources) that require massively parallel techniques to be processed.

7 Other Grants and Activities

7.1 National actions

- In 2013, Pilgrim signed an External Research Contract (15 k€) with a start-up based in Rennes, named Semsoft, that works in the domain of web data management. This contract concerns a technology watch activity about cooperative approaches to data access. An internal research report consisting of a survey of approaches from the literature that deal with the empty answer set or the plethoric answer set problems has been produced.

- Hélène Jaudoin got a grant (12 k€) from the University of Rennes 1 in the framework of the Incitative Action "Projets scientifiques émergents 2010" on the topic of "argumented answer agreggation". This project allowed us to investigate a new axis related to the computation of cooperative answers by leveraging different types of metadata (cluster-based summaries in particular [16]) and information extracted from the database and/or its schema [19].

- Pilgrim also submitted a DGA project (RAPID) proposal about flexible RDF data management, with two other partners, namely Semsoft (Rennes) and LRI (Orsay). It has been accepted and will start in 2014. Two Ph.D. students and a post-doctorate will be recruited.

François Goasdoué is involved in the following projects:

- Datalyse (Investissements d'Avenir, *Big Data / Cloud computing*, 2013–2016). This project deals with Big Data management in a cloud architecture. The consortium is made of industrial partners (Eolas – Business & Decision and Les Mousquetaires), academic partners (Inria, LIFL of Univ. Lille, LIG of Univ. Grenoble, LIRMM of Univ. Montpellier), as well as the city of Grenoble as an open data provider.

- ANR JCJC Pagoda (2013–2017). PAGODA (Practical algorithms for ontology-based data access) is a basic research project whose objective is to improve the efficiency and robustness of ontology-based data access by developing scalable algorithms for query answering in the presence of ontologies as well as pragmatic approaches to handling inconsistent data. Partners are from LIG of Univ. Grenoble, LIRMM of Univ. Montpellier, and LRI of Univ. Paris-Sud.

8 Dissemination

8.1 Teaching

Project members give lectures in different faculties of engineering, in the third cycle University curriculum: "Bases de données avancées" in the track "Intelligent Interaction with Information" of the Master's degree in computer science at University of Rennes 1 (École Doctorale Matisse), and at Enssat (third year level cursus).

8.2 Scientific activities

8.2.1 Highlights of the year

- The paper by P. Bosc, O. Pivert, and G. Smits, entitled "An Approach to Database Preference Queries Based on an Outranking Relation" received the International Fuzzy Systems Association 2013 Best Paper Award (L.A. Zadeh Prize).
- Katia Abbaci defended her Ph.D. thesis [1] on December 12, 2013.
- Demo paper accepted at VLDB'13 [7].

8.2.2 Program committees

François Goasdoué served as a member of the following program committee:

- 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013), Beijing, China, August 3-9, 2013 (Senior PC member).
- IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2013), Washington DC, USA, November 4-6, 2013.
- 2nd International Workshop on Open Data (WOD'13), Paris, France, June 3, 2013.

H. Jaudoin served as a member of the following program committee:

- 29^{e} Journées Bases de Données Avancées, Nantes, 23-25 octobre 2013.

L. Liétard served as a member of the following program committees:

- 28th ACM Symposium on Applied Computing (SAC 2013), Coimbra, Portugal, March 18-22, 2013.
- 8th European Society for Fuzzy Logic and Technology Conference (EUSFLAT 2013), Milano, Italy, September 11-13, 2013.

O. Pivert served as a member of the following program committees:

- 28th ACM Symposium on Applied Computing (SAC 2013), Coimbra, Portugal, March 18-22, 2013.
- IEEE Symposium Series on Computational Intelligence (SSCI 2013), Singapore, April 15-19, 2013.
- 10th Conference on Flexible Query-Answering Systems (FQAS 2013), Granada (Spain), September 18-20, 2013.
- IEEE International Conference on Fuzzy Systems (Fuzz-IEEE 2013), Hyderabad, India, July 7-10, 2013.
- 24th International Conference on Database and Expert Systems Applications (DEXA 2013), Prague, Czech Republic, August 26-30, 2013.

- 5th International Conference on Intelligent Networking and Collaborative Systems (InCoS 2013), Xi'an, China, 9-11 September, 2013.
- 8th European Society for Fuzzy Logic and Technology Conference (EUSFLAT 2013), Milano, Italy, September 11-13, 2013.
- 7th International Conference on Scalable Uncertainty Management (SUM 2013), Washington DC, USA, September 16-18, 2013.
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2013), Reims, France, October 10-11, 2013.

D. Rocacher served as a member of the following program committees:

- 10^{e} Conférence en Recherche d'Information et Applications (CORIA 2013), Neuchâtel, Suisse, 3-5 avril 2013.
- 8th European Society for Fuzzy Logic and Technology Conference (EUSFLAT 2013), Milano, Italy, September 11-13, 2013.
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2013), Reims, France, October 10-11, 2013.

8.2.3 Editorial boards

Olivier Pivert is a member of the following editorial boards:

- Journal of Intelligent Information Systems,
- Fuzzy Sets and Systems,
- International Journal of Fuzziness, Uncertainty and Knowledge-Based Systems,

8.2.4 Steering committees

O. Pivert is as a member of the steering committee of the French-speaking conference "Rencontres Francophones sur la Logique Floue et ses Applications" (LFA).

8.2.5 International advisory boards

O. Pivert is as a member of the international advisory boards of

- the International Conference on Flexible Query-Answering Systems (FQAS).
- a special issue of IJSC (International Journal of Semantic Computing) dedicated to Prof. Lotfi Zadeh.

8.2.6 Organization of special sessions

- Ludovic Liétard co-organized, with Allel Hadjali (from ENSMA, Poitiers) and Henri Prade (from IRIT, Toulouse), a special session dedicated to "Advances in Bipolarity Applied to Databases and Information Systems" at the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT'13).

9 Bibliography

Major publications by the team in recent years

[1] P. Bosc, A. Hadjali, O. Pivert, "Incremental Controlled Relaxation of Failing Flexible Queries", *Journal of Intelligent Information Systems 33*, 3, 2009, p. 261–283.

[2] P. Bosc, L. Liétard, O. Pivert, D. Rocacher, *Gradualité et imprécision dans les bases de données*, Ellipses, 2004.

[3] P. Bosc, O. Pivert, "Modeling and Querying Uncertain Relational Databases: A Survey of Approaches Based on the Possible Worlds Semantics", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 18*, 5, 2010, p. 565–603.

[4] P.Bosc, O. Pivert, D.Rocacher, "About Quotient and Division of Crisp and Fuzzy Relations", *Journal of Intelligent Information Systems 29*, 2, 2007, p. 185–210.

[5] P.Bosc, O. Pivert, "About Projection-Selection-Join Queries Addressed to Possibilistic Relational Databases", *IEEE Transactions on Fuzzy Systems 13*, 1, 2005, p. 124–139.

[6] P.Bosc, O. Pivert, "About Possibilistic Queries and their Evaluation", *IEEE Transactions on Fuzzy Systems 15*, 1, 2007, p. 439–452.

[7] O. Pivert, P. Bosc, *Fuzzy Preference Queries to Relational Databases*, Imperial College Press, London, UK, 2012.

Doctoral dissertations and "Habilitation" theses

[1] K. Abbaci, *Contribution à l'interrogation flexible et personnalisée d'objets complexes modélisés par des graphes*, PhD Thesis, University of Rennes 1 – Ecole doctorale MATISSE, December 12, 2013, supervised by A. Hadjali, D. Rocacher, and L. Liétard.

[2] J. Leblay, *Techniques d'optimisation pour des données semi-structurées du web sémantique*, PhD Thesis, University Paris-Sud – Ecole doctorale EDIPS, September 27, 2013, supervised by F. Goasdoué and I. Manolescu.

Articles in referred journals and book chapters

[3] P. Bosc, O. Pivert, "On a Fuzzy Bipolar Relational Algebra", *Information Sciences 219*, 2013, p. 1–16.

[4] P. Bosc, O. Pivert, "Querying Possibilistic Databases: Three Interpretations", *in: Soft Computing: State of the Art Theory and Novel Applications*, R. Yager, A. Abbasov, M. Reformat, and S. Shahbazova (editors), *Studies in Fuzziness and Soft Computing*, *291*, Springer, Heidelberg, Germany, 2013, p. 161–176.

[5] F. Goasdoué, K. Karanasos, Y. Katsis, J. Leblay, I. Manolescu, S. Zampetakis, "Growing Triples on Trees: an XML-RDF Hybrid Model for Annotated Documents", *VLDB Journal*, 2013, http://hal.inria.fr/hal-00828906.

[6] F. Goasdoué, M.-C. Rousset, "Robust Module-Based Data Management", *IEEE Trans. Knowl. Data Eng. 25*, 3, 2013, p. 648–661.

[7] G. Smits, O. Pivert, T. Girault, "ReqFlex: Fuzzy Queries for Everyone", *PVLDB 6*, 12, 2013, p. 1206–1209.

[8] G. Smits, O. Pivert, A. Hadjali, "Fuzzy Cardinalities as a Basis to Cooperative Answering", *in: Flexible Approaches in Data, Information and Knowledge Management*, O. Pivert and S. Zadrożny (editors), *Studies in Computational Intelligence*, *497*, Springer, Heidelberg, Germany, 2013, p. 261–289.

[9] N. Tamani, L. Liétard, D. Rocacher, "A Fuzzy Ontology for Database Querying with Bipolar Preferences", *International Journal of Intelligent Systems 28*, 2013, p. 4–36.

Publications in Conferences and Workshops

[10] K. Abbaci, A. Hadjali, L. Liétard, D. Rocacher, "A Linguistic-Quantifier-Based Approach for Skyline Refinement", *in: Proc. of the IFSA-NAFIPS Joint Congress*, Edmonton, Canada, 2013.

[11] D. Colazzo, T. Ghosh, F. Goasdoué, I. Manolescu, A. Roatis, "WaRG: Warehousing RDF Graphs", *in: Journées de Bases de Données Avancées (BDA)*, Nantes, France, October 2013. Demonstration, http://hal.inria.fr/hal-00868670.

[12] D. Colazzo, F. Goasdoué, I. Manolescu, A. Roatis, "Warehousing RDF Graphs", *in: Journées de Bases de Données Avancées (BDA)*, Nantes, France, October 2013, http://hal.inria.fr/hal-00868616.

[13] F. Goasdoué, K. Karanasos, Y. Katsis, J. Leblay, I. Manolescu, S. Zampetakis, "Fact Checking and Analyzing the Web", *in: SIGMOD – ACM International Conference on Management of Data*, 2013, http://hal.inria.fr/hal-00814285.

[14] F. Goasdoué, K. Karanasos, Y. Katsis, J. Leblay, I. Manolescu, S. Zampetakis, "Fact checking and Analyzing the web", *in: Journées de Bases de Données Avancées (BDA)*, Nantes, France, October 2013. Demonstration, http://hal.inria.fr/hal-00867749.

[15] F. Goasdoué, I. Manolescu, A. Roatis, "Efficient query answering against dynamic RDF databases", *in: Proc. of EDBT'13*, p. 299–310, 2013.

[16] M.-J. Lesot, G. Smits, O. Pivert, "Adequacy of a User-Defined Vocabulary to the Data Structure", *in: Proc. of the 22nd IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'13)*, Hyderabad, India, 2013.

[17] M.-J. Lesot, G. Smits, O. Pivert, "Mesures d'adéquation entre vocabulaire expert et structure de données", *in: Actes des Rencontres Francophones sur la Logique Floue et ses Applications (LFA'13)*, p. 243–250, Reims, France, 2013.

[18] L. Liétard, O. Pivert, D. Rocacher, "On a Gradual Inclusion of Bipolar Fuzzy Relations", *in: Proc. of the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT'13)*, Milano, Italy, 2013.

[19] O. Pivert, G. Smits, H. Jaudoin, "Finding Similar Objects in Relational Databases — An Association-Based Fuzzy Approach", *in: Proc. of the 10th International Conference on Flexible Query Answering Systems (FQAS'13)*, p. 425–436, Granada, Spain, 2013.

[20] G. Smits, O. Pivert, T. Girault, "Requêtes floues et SGBD relationnels : vers un couplage renforcé", *in: Actes des Rencontres Francophones sur la Logique Floue et ses Applications (LFA'13)*, p. 77–84, Reims, France, 2013.

[21] G. Smits, O. Pivert, T. Girault, "Towards Reconciling Expressivity, Efficiency and User-Friendliness in Database Flexible Querying", *in: Proc. of the 22nd IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'13)*, Hyderabad, India, 2013.

[22] G. Smits, O. Pivert, H. Jaudoin, F. Paulus, "An Autocompletion Mechanism for Enriched Keyword Queries to RDF Data Sources", *in: Proc. of the 10th International Conference on Flexible Query Answering Systems (FQAS'13)*, p. 601–612, Granada, Spain, 2013.

[23] N. Tamani, L. Liétard, D. Rocacher, "Bipolar Conjunctive Query Evaluation for Ontology-Based Database Querying", *in: Proc. of the 10th International Conference on Flexible Query Answering Systems (FQAS'13)*, p. 389–400, Granada, Spain, 2013.

[24] N. Tamani, L. Liétard, D. Rocacher, "A Relational Division Based on a Fuzzy Bipolar R-implication Operator", *in: Proc. of the 22nd IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'13)*, Hyderabad, India, 2013.