Project-Team PILGRIM

Graduality, Imprecision, and Mediation in Database Management Systems

Lannion

Activity Report

2011

1 Team

Head of the team

Olivier Pivert, Professor, Enssat

Administrative assistant

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

Université Rennes 1 personnel

Patrick Bosc, Professor, Enssat
Allel Hadjali, Associate Professor, Enssat
Hélène Jaudoin, Associate Professor, Enssat
Ludovic Liétard, Associate Professor, IUT Lannion
Daniel Rocacher, Associate Professor, HDR, Enssat
Grégory Smits, Associate Professor, IUT Lannion

PhD students

Amine Mokhtari, Région Bretagne grant and Conseil Général 22 grant, since October 2007
Nouredine Tamani, Région Bretagne grant and Conseil Général 22 grant, since January 2008
Katia Abbaci, Région Bretagne grant and ANR grant, since November 2009

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

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

natural language which can be hardly defined in terms of usual sets (e.g., high, young, small, 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 intentional 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 ($\perp$) tied by de Morgan's laws. Then:

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

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 envisaged, 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,
- Lukasiewicz 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]:

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

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

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 is:

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

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 (Π = N = 0 to Π = 1 and N = 0), then those which are completely possible and somewhat certain (from Π = 1 and N = 0 to Π = 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

[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.

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.

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:

- conditions calling on aggregate functions applying to fuzzy sets, for instance fuzzy quantified statements such as "most employees have a medium salary" which can be expressed in the context of an SQL-like language,
- the handling of fuzzy bags (fuzzy multisets) and their connection with fuzzy numbers.

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

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

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).

3.4 Query rewriting using views

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.

[Len02] M. Lenzerini, "Data Integration : A Theoretical Perspective", 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.

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 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.

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

[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.

[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.

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.

4 Application Domains

As to the aspect dealing with flexible queries, there are several potential application domains. Soft querying turns out to be relevant in many contexts, such as information retrieval, in particular on the Web (many commercial systems, e.g. Google or Yahoo use a technique to rank-order the answers), 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.

Even though most of the research works performed in Pilgrim assume relational data, many results can be transposed to other contexts such as information retrieval or multimedia database querying. We are currently working on the specification of a flexible route planning system involving fuzzy preferences (cf. Section 6.3), which should illustrate the utility of fuzzy queries in the context of intelligent transportation systems.

Databases involving imprecise data are not yet common in practice for two reasons: developing DBMSs supporting such data is a hard job and the demand is presently not so strong. However, many potential domains could take advantage of such advanced systems capable of storing and querying databases where some pieces of information are imprecise: 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.

5 Software

- FRIGA (Flexible RetrIeval and GRaded Answers) is a flexible querying prototype that aims at evaluating fuzzy queries addressed to regular databases. It takes the form of an additional software layer on top of PostgreSQL, whose function is to translate a fuzzy query into a procedural evaluation program including regular SQL queries in order to take advantage of the optimization mechanisms that exist in the DBMS. In its current version, the prototype is able to process "simple" fuzzy queries (i.e., fuzzy queries involving a single block) and we are now extending it so as to make it support nested queries and so-called contextual fuzzy queries (i.e. queries where some fuzzy terms do not have to be explicitly defined by the user but whose interpretation depends on a certain context that can be determined from the query itself).

- POSTGRESQLf is a prototype which implements a significant part of the SQLf language (i.e., the fuzzy query language developed by the team) on top of the open source database

management system PostgreSQL. Contrary to FRIGA which calls on external procedural programs, POSTGRESQLf evaluates fuzzy queries by means of stored functions written in PL/SQL or C.

- DISCORD (Database Interrogation System Computing OutRanking Degrees): this prototype, currently under development, implements an approach to database preference queries based on the notion of outranking (cf. Subsection 6.2.1), suited to the case where partial preferences are incommensurable. This model, which rests on a weighted majority rule, constitutes an alternative to the use of Pareto order — and also refines it.

- 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. 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.

- PROST (Preference-based ROute planning SysTem): As mentioned in Section 4, we decided to illustrate the interest of flexible querying techniques in the domain of Intelligent Transportation Systems (ITS). Along with two other IRISA teams, namely Cairn and Cordial, Pilgrim is involved in the development of a platform called MOB-ITS in the context of the CPER INVENT'IST 2007-2013. This platform aims at supporting mobile and interactive access to information for ITS applications. In this framework, Pilgrim has implemented a flexible route planner which will be integrated into that platform.

- 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).

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

6 New Results

6.1 Possibilistic database modeling and querying

Participants: Patrick Bosc, Olivier Pivert, Allel Hadjali.

- Query semantics in a possibilistic database framework. 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 [24], 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.

- Possibilistic skyline queries. In [19], we deal with Skyline queries in the context of possilistic databases, where uncertain attribute values are represented by possibility distributions. In this framework, Skyline queries aim at computing the extent to which any tuple from a given relation is possibly/certainly not dominated by any other tuple from that relation. Beside the interpretation of possibilistic Skyline queries, a basic algorithm suited to their evaluation is provided.

6.2 Flexible querying of classical databases

6.2.1 Preference modeling

Participants: Olivier Pivert, Patrick Bosc, Grégory Smits, Allel Hadjali.

- Possibilistic logic approach to preference queries. In [8], a new approach to database preferences queries is presented, where preferences are represented in a possibilistic logic manner, using symbolic weights. The symbolic weights may be processed without assessing their precise value, which leaves the freedom for the user to not specify any priority among the preferences. The user may also enforce a (partial) ordering between them, if necessary. The approach can be related to the processing of fuzzy queries whose components are conditionally weighted in terms of importance. Here, importance levels are symbolically processed, and refinements of both Pareto ordering and minimum ordering are used. The representational power of the proposed setting is stressed, while the approach is compared with database *Best* operator-like methods and with the CP-net approach developed in artificial intelligence. The paper also provides a structured and rather broad overview of the different lines of research in the literature dealing with the handling of preferences in database queries.

- Competitive conditional preferences. In [4], a new type of database queries involving preferences is introduced. The idea is to consider competitive conditional preference

clauses structured as a tree, of the type "preferably P_1 or ... or P_n; if P_1 then preferably $P_{1,1}$ or ...; if P_2 then preferably $P_{2,1}$ or ...", where the P_i's are not exclusive (thus the notion of competition). The paper defines two possible interpretations of such queries and outlines two evaluation techniques which follow from them.

- Study of three noncommutative fuzzy connectives. In most of query languages, conjunctive and disjunctive combinations of conditions remain the usual way for aggregation. Fuzzy query languages also offer trade-off operators, such as means in order to compensate between elementary conditions. In [23], we introduce a new type of condition basically founded on the interaction between two predicates, thus enriching the panoply of tools the user is provided with and the power of query languages. In [21], we investigate one of them in more detail, which makes it possible to express conditions of the form "P_A all the more as P_B".

- Preference query model based on the fusion of local orders. In [20], we define a new approach to database preference queries. The situation considered is that of queries involving incommensurable partial preferences, possibly associated with scoring functions. The basic principle is to rank the tuples according to each partial preference, then to merge the local orders obtained, using a linear function for aggregating the local scores attached to the tuples. Basically, a local score expresses the extent to which a tuple is strictly better than many others and not strictly worse than many others with respect to the partial preference attached to a given attribute.

- Inferred fuzzy predicates. In [31], we deal with database preference queries involving fuzzy conditions which do not explicitly refer to an attribute from the database, but whose meaning is rather inferred from a set of rules. The approach we propose, which is based on some concepts from the fuzzy control domain (aggregation and defuzzification, in particular), significantly increases the expressivity of fuzzy query languages inasmuch as it allows for new types of predicates. An implementation strategy involving a coupling between a DBMS and a fuzzy controller is outlined. In [28], an alternative approach based on the fuzzy inference pattern called generalized modus ponens is proposed.

- Contextual preferences. In [26, 1], we propose a fuzzy-rule-based model for the representation of contextual preferences in a database querying framework. We discuss the augmentation of a query with preferences deduced from information regarding the current context of the user. The approach proposed is based on a well-known inference scheme, namely the generalized modus ponens.

- Fuzzy skyline queries. In [27], we deal with database preference queries based on the skyline paragdigm, which aim at retrieving the tuples from a database which are not Pareto-dominated by any other tuple. We propose different ways to fuzzify such queries in order to make them more flexible, to increase their discrimination power, to make them more drastic or more tolerant. In particular, some of these extensions make it possible to reduce the risk of getting many incomparable tuples, even when the number of dimensions is high.

- Tolerant antidivision queries. In [6], we were interested in taking preferences into account for a family of queries inspired by the antidivision. An antidivision query aims at retrieving the elements associated with none of the elements of a specified set of values. We suggested the introduction of preferences inside such queries with the following specificities: i) the user gives his/her preferences in an ordinal way and ii) the preferences apply to the divisor which is defined as a hierarchy of sets. Different uses of the hierarchy are investigated, which leads to queries conveying different semantics and the property of the result delivered is characterized. Furthermore, the case where a conjunctive stratified antidivision query returns an empty set of answers was dealt with, and an approach aimed at relaxing such queries was proposed.

- Stratified division and a weakening/strengthening mechanism. In [5], we were interested in taking preferences into account for a family of queries inspired by the relational division. A division query aims at retrieving the elements associated with a specified set of values and usually the results remain not discriminated. So, we suggested the introduction of preferences inside such queries with the following two specificities: i) the user gives his/her preferences in an ordinal way and ii) the preferences apply to the divisor which is stratified, i.e., defined as a hierarchy of sets. Different uses of the hierarchy were investigated, which led to queries conveying different semantics and the property of the result in terms of a quotient was studied. A special attention was paid to the implementation of such extended division queries using a regular database management system along which some experiments to support the feasibility of the approach. Moreover, the issue of empty or overabundant answers was dealt with.

6.2.2 Bipolar fuzzy queries

Participants: Ludovic Liétard, Nouredine Tamani, Daniel Rocacher, Olivier Pivert, Patrick Bosc.

The concept of bipolar queries is a particular way to integrate preferences inside queries where mandatory preferences, called constraints, are distinguished from optional preferences, called wishes. Constraints and wishes are respectively defined by a set of acceptable values and a set of desired values. Tuples satisfying the constraints and the wishes are returned in priority to the user. If such answers do not exist, tuples satisfying only the constraints are delivered. Constraints are preferred to wishes since wishes are optional in the sense that they may be not fulfilled by the answers provided to the user. We consider the case of bipolar conditions where both the wishes and the constraints are defined by fuzzy sets thus defining bipolar fuzzy queries.

- Bipolar relational algebra. In [25], we presented an extension of relational algebra suited to 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 degrees in the unit interval.

- Negation of bipolar fuzzy conditions. In [22], we dealt with the negation operator in the context of a bipolar fuzzy relational algebra. Several possible definitions of the negation were studied and assessed with respect to some desirable properties. A negation operator which complies with all those desirable properties was proposed, which serves as a basis for the definition of the set difference operation in the extended relational algebraic framework considered.

- "Or else" connective. Previously studied fuzzy bipolar conditions of type "and if possible" are made of a mandatory condition c and an optional condition w. They make it possible to express complex preferences of a conjunctive nature. In [30], we study a new kind of fuzzy bipolar condition of the form "or else" which express complex preferences of a disjunctive nature. We show that the "or else" form can be used as a negation operator of the "and if possible" form and vice versa. We also show that both these forms are compatible and, therefore, fuzzy bipolar conditions of both types can be used together in the same bipolar query.

- Linguistic quantifiers and bipolarity. In [40], fuzzy bipolar conditions are combined with linguistic quantifiers and the notion of a bipolar quantified proposition is introduced and studied (in particular from an evaluation point of view).

- Bipolar fuzzy extension of SQL. In [35], we define the main elements of the Bipolar SQLf language, which is an SQL-like querying language based on a bipolar relational algebra. This language is an extension of the SQLf language defined earlier in the team. Basic statements (projection, selection, etc.) are first defined in terms of syntax, evaluation and calibration. Then, complex statements, such as bipolar queries based on nesting operators are studied in terms of expression, evaluation, query equivalence and backward compatibility with the SQLf language.

- Bipolar division. In [7], we investigated how bipolarity may impact the division operator in the context of relational databases. Various forms of bipolar divisions can indeed be devised, each of them conveying a specific semantics. Starting with a basic bipolar division with crisp relations where the divisor is made of two components (one representing values which are required, the other describing the values which are expected but not mandatory), we moved to more sophisticated forms of bipolar divisions: i) the dividend is a graded (fuzzy) relation where each tuple has a degree of membership to the fuzzy concept conveyed by the relation and ii) the universal quantifier is softened into possibly two different weaker forms (one related to the constraint, i.e., the mandatory values of the divisor, and the other for the wish part, i.e., the values of the divisor that are only desired). The result of all these bipolar divisions is characterized as a quotient, i.e., a maximal relation. In [36], this work about the bipolar division is extended to the case where the relations involved are defined by fuzzy bipolar conditions.

6.2.3 Cooperative answering to flexible database queries

Participants: Allel Hadjali, Grégory Smits, Olivier Pivert, Hélène Jaudoin, Patrick Bosc.

- Failing fuzzy queries. In [33], we deal with conjunctive fuzzy queries that yield an empty or unsatisfactory answer set. We propose a cooperative answering approach which efficiently retrieves the minimal failing subqueries of the initial query (which can then be used to explain the failure). The detection of the minimal failing subqueries relies on a prior step of fuzzy cardinalities computation. The main advantage of this strategy is to rely on a single scan of the database. Moreover, the fuzzy cardinalities which describe the data distributions easily fit in main memory.

- Diversity. In [32], we deal with fuzzy queries and describe an approach that aims at providing users with a set of answers which satisfies a diversity criterion on one or several attributes. Different cases are considered and two types of algorithms are described. The first one, which has a linear complexity in terms of the number of tuples in the result, is suited to the case where the notion of similarity underlying the definition of diversity is crisp. The second one, based on a trial and error strategy, makes it possible to deal with fuzzy similarity, but its high complexity means that it can be employed only when a relatively small sets of tuples is used to increase diversity in the result.

6.3 Fuzzy preferences in intelligent transportation systems

Participants: Ludovic Liétard, Nouredine Tamani, Allel Hadjali, Olivier Pivert, Amine Mokhtari, Daniel Rocacher.

Three years ago, we started investigating a new topic, namely the application of fuzzy set theory to the specification of a route planning system involving sophisticated user preferences. In 2011, the results obtained concern the following issues:

- Bipolarity in flexible querying of information systems dedicated to multimodal transport networks. In [37], a flexible approach aimed to help users organizing their trips by promoting public transportation networks (bus, subway, train, boat, plane, etc.) is presented. An integrated environment is defined, in which it is possible for the user to express queries with complex preferences so as to meet his/her expectations. Complex preferences are modeled by fuzzy bipolar conditions which associate negative and positive conditions. In the context considered, bipolar queries are addressed to multimodal transport information systems, which are often made of several distributed and heterogeneous databases. Therefore, semantic aspects have to be taken into consideration in the querying process so that only the most relevant data is targeted to evaluate queries. The flexible approach proposed in [37] combines a reasoning mechanism with a highly expressive fuzzy query language.

- Extension of a fuzzy ontology for flexible querying of an embedded database. In [38], we propose a personalized approach for flexible querying of information systems. This approach consists in the combination of the reasoning capabilities of the fuzzy DLR-Lite ontology and the expressivity of the SQLf language. The interpretation of the gradual inclusion (subsumption) axioms of the ontology is based on Gödel's fuzzy implication. Its generalization to a tree of inclusions is also proposed. This tree and its property of propagation of degrees are the basic theoretical elements of our application, which

consists in querying of a multimodal transport information system which is embedded in a mobile terminal characterized by limited storage and processing capabilities.

- Route planning. The Ph.D. dissertation [2] presents a set of contributions aimed at designing a new generation of route planners capable of dealing with complex user preferences. Fuzzy set theory is used as a formal basis. The first contribution is a typology of preferences that make sense in the context of a unimodal travel. The bipolar nature of user preferences in such a context is discussed and taken into account. The second contribution is a route query language, called RQL, based on a fuzzy extension of tuple relational calculus. It is also shown how to augment a query with new preferences deduced from the user's context. Finally, the architecture of a personalized route planning system is proposed, as well as an efficient query processing strategy. Experimental results that show the feasibility of the approach are presented and discussed.

6.4 Flexibility issues in data integration systems

Participants: Hélène Jaudoin, Olivier Pivert, Ludovic Liétard, Nouredine Tamani, Daniel Rocacher, Grégory Smits, Allel Hadjali.

- Processing fuzzy queries in a Peer Data Management System (PDMS) using fuzzy summaries. In [34], we considered the situation where a fuzzy query is submitted to distributed data sources. In order to save bandwith and processing cost, we proposed a technique whose aim is to forward the query to the most relevant sources only. It is assumed that a fuzzy-cardinality-based summary of every data source is available, and the approach we propose consists in estimating the relevance of a source with respect to a user query, based on its associated summary. The general case where the user does not necessarily employ the vocabulary (i.e., the labels from the fuzzy partitions) that was used for summarizing the source was considered.

- Rewriting a fuzzy query using imprecise views. In [29], we tackled the problem of answering queries using views when the queries and the views may involve fuzzy value constraints. These constraints allow for specifying the possible values of the attributes by associating them with a degree between 0 and 1. Such constraints represent user preferences in the queries, whereas in the views, they give a concise, flexible but informative description of data as done by summaries. The problem is formalized in the setting of the description logics $\mathcal{FL}_0$ extended to fuzzy value constraints. We proposed an algorithm of structural subsumption for this logic, that plays a key role in the query rewriting algorithm. Finally, we characterized the query rewriting forms.

- Expressing and evaluating bipolar fuzzy preferences in a context of distributed and heterogeneous information systems. In [39], it is assumed that complex preferences are modeled by fuzzy bipolar conditions which associate negative and positive conditions, and queries involving such bipolar conditions are addressed to information systems in which data from various distributed and heterogeneous databases are integrated. A new

approach is introduced, which combines a reasoning mechanism based on the fuzzy bipolar DLR-Lite (which is a fuzzy bipolar description logic) with Bipolar SQLf (which is a bipolar relational language).

6.5 Graph queries and web services

Participants: Ludovic Liétard, Allel Hadjali, Daniel Rocacher, Katia Abbaci.

- Similarity skyline of a graph query. One of the fundamental problems in graph databases is similarity search for graphs of interest. Existing approaches dealing with this problem rely on a single similarity measure between graph structures. In [10, 9], we suggest an alternative approach allowing for searching similar graphs to a graph query where similarity between graphs is rather modeled by a vector of scalars than a unique scalar. To this end, we introduce the notion of similarity skyline of a graph query defined by the subset of graphs of the target database that are the most similar to the query in a Pareto sense. The idea is to achieve a d-dimensional comparison between graphs in terms of d local distance (or similarity) measures and to retrieve those graphs that are maximally similar in the sense of the Pareto dominance relation. A diversity-based method for refining the retrieval result is proposed as well.

- Preferences in service retrieval. Current approaches for service discovery are based on semantic knowledge, such as ontologies and service behavior (described as a process model). However, these approaches still remain with a high selectivity rate, resulting in a large number of services offering similar functionalities and behavior. One way to improve the selectivity rate is to cope with user preferences defined on quality attributes. In [11, 12, 13], we we propose a novel approach for service retrieval that takes into account the service process model and relies both on preference satisfiability and structural similarity. User query and target process models are represented as annotated graphs, where user preferences on QoS attributes (such as response time, availability and throughput) are modelled by means of fuzzy sets. A flexible evaluation strategy based on fuzzy linguistic modifiers is introduced and different ranking methods are discussed. In [14], an extensive set of experiments based on real data sets is conducted, on one hand, to demonstrate the efficiency and the scalability of this approach, and on the other hand, to analyze the effectiveness and the accuracy of the proposed ranking methods compared to an expert evaluation.

- Data as a Service (DaaS) is a flexible way that allows enterprises to expose their data. Composition of DaaS services provides bridges to answer queries. User preferences are becoming increasingly important to personalizing the composition process. In [17, 3], we propose an approach to compose DaaS services in the context of preference queries where preferences are modeled by means of fuzzy sets that allow for a large variety of flexible terms such as “cheap”, “affordable” and “fairly expensive”. The proposed approach is based on RDF-based query rewritings to take into account the partial matching between individual DaaS services and parts of the user query. Matching degrees between DaaS

services and fuzzy preference constraints are computed by means of different constraints inclusion methods. Such degrees express to which extent a service is relevant to the resolution of the query. A fuzzification of Pareto dominance is also proposed to better rank composite services by computing the score of services. The resulting scores are then used to compute the top-k DaaS service compositions that cover the user query. In [18], a prototype named FuDoCS is presented. This framework implements a first algorithm that efficiently compute the top-k service compositions and a second algorithm — introduced in [15] — that diversifies the top-k service compositions by using a quality metric that combines both diversity and service accuracy.

- The widespread explosion of functional-similar Web services has led to a new challenge of selecting the most relevant services using quality of service (QoS) aspects. Traditionally, the relevance of a service is determined by computing an overall score that aggregates individual QoS values. Most of these approaches require users to assign weights over QoS attributes. This is a rather demanding task and an imprecise specification of the weights could miss user desired services. Moreover, recent approaches focus on computing service skyline over a set of QoS aspects. This can completely free users from assigning weights on QoS attributes. However, two main drawbacks characterize such approaches. First, the service skyline often privileges services with a bad compromise between different QoS attributes. Second, as the size of the service skyline may be quite large, users will be overwhelmed during the service selection process. In [16], a new concept is introduced, called alpha-dominant service skyline, to address the above issues and a suitable algorithm is developed for computing it efficiently. Experimental evaluation on synthetically generated datasets, demonstrates both the effectiveness of the proposed concept and the efficiency of the algorithm.

7 Other Grants and Activities

7.1 National actions

- Ludovic Liétard, Allel Hadjali, and Daniel Rocacher participate in the ANR project "AOC", which deals with the definition of matching methods for complex objects (graphs in particular). The other teams involved are from IRIT (Toulouse), PRISM (Versailles), LIRIS (Lyon), LIESP (Lyon).

- Hélène Jaudoin got a grant from the University of Rennes 1 in the framework of the Incitative Action "Projets scientifiques émergents 2010". An engineer has been hired in October 2011 and has started working about a fuzzy approach to flexible data integration.

- The team Pilgrim is involved in a PME project of the "pôle Images et Réseaux", named IntelSearch, in collaboration with Semsoft, Swid, and Ensai. This project is about flexible data integration system and the leader for Pilgrim is Hélène Jaudoin.

7.2 International actions

- Ms. Meryem Saidi, Ph.D student at the university Aboubakr Belkaid in Tlemcen (Algeria), spent one month in our team from December 26, 2010 to January 26, 2011.
- Ms. Samia Boulkrinat, Assistant Professor at the USTHB University in Algiers (Algeria), spent one month in our team (from September 29, 2011 to October 28, 2011) in order to start a research collaboration about fuzzy recommender systems.

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 speciality "Interaction Intelligente sur l'Information" of the Master's degree in computer science at University of Rennes 1, and at Enssat (third year level cursus).

A. Hadjali gave a Master's course entitled "Requêtes à préférences" at the University USTHB of Algiers (Algeria) in September 2011.

8.2 Scientific activities

8.2.1 Highlights of the year

- Allel Hadjali defended his "Habilitation à Diriger des Recherches" on December 7, 2011.
- Amine Mokhtari defended his Ph.D. thesis on March 25, 2011.
- The paper by P. Bosc and O. Pivert entitled "On the negation of bipolar fuzzy conditions" got an "outstanding paper" award at the 30$^{\text{th}}$ North American Fuzzy Information Processing Society (NAFIPS 2011) Conference.
- One demo paper accepted at VLDB'11.

8.2.2 Program committees

P. Bosc served as a member of the following program committees:

- 26$^{\text{th}}$ ACM Symposium on Applied Computing (SAC 2011), Taichung, Taiwan, March 21-25, 2011.
- 9$^{\text{th}}$ Conference on Flexible Query-Answering Systems (FQAS 2011), Ghent (Belgium), October 26-28, 2011.
- Joint 7$^{\text{th}}$ Conference of the European Society of Fuzzy Systems and Technology (EUSFLAT 2011) and Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2011), Aix-les-Bains, France, July 18-22, 2011.

- DEXA 6th International Workshop on Flexible Database and Information Systems Technology (FlexDBIST 2011), Toulouse, France, August 29-September 2, 2011.
- 3rd International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2011), St. Maarten, The Netherlands Antilles, January 23-28, 2011.
- 30th International Conference of the North American Fuzzy Information Processing Society (NAFIPS'11), El Paso, TX, USA, March 18-20, 2011.
- 22nd International Conference on Database and Expert Systems Applications (DEXA 2011), Toulouse, France, August 29-September 2, 2011.

A. Hadjali served as a member of the following program committees:

- 2e Atelier Graphes et Appariement d'Objects Complexes (GAOC 2011), in conjunction with EGC 2011, Brest, January 25, 2011.
- 9th Conference on Flexible Query-Answering Systems (FQAS 2011), Ghent (Belgium), October 26-28, 2011.
- 19th International Symposium on Methodologies for Intelligent Systems (ISMIS 2011), Warsaw, Poland, June 28-30, 2011.
- Joint 7th Conference of the European Society of Fuzzy Systems and Technology (EUSFLAT 2011) and Rencontres Francophones sur la Logique Floue et ses Applications (LFA'11), Aix-les-Bains, France, July 18-22, 2011.
- 1eres Rencontres sur la Recherche en Informatique (R2I 2011), Tizi-Ouzou, Algeria, June 12-14, 2011.
- Colloque sur l'Optimisation et les Systèmes d'Information, Guelma, Algeria, April 24-28, 2011.

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

- Colloque sur l'Optimisation et les Systèmes d'Information, Guelma, Algeria, April 24-28, 2011.

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

- 26th ACM Symposium on Applied Computing (SAC 2011), Taichung, Taiwan, March 21-25, 2011.
- 2e Atelier Graphes et Appariement d'Objects Complexes (GAOC 2011), in conjunction with EGC 2011, Brest, January 25, 2011.

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

- 26th ACM Symposium on Applied Computing (SAC 2011), Taichung, Taiwan, March 21-25, 2011.

- 9th Conference on Flexible Query-Answering Systems (FQAS 2011), Ghent (Belgium), October 26-28, 2011.
- 19th International Symposium on Methodologies for Intelligent Systems (ISMIS 2011), Warsaw, Poland, June 28-30, 2011.
- Joint 7th Conference of the European Society of Fuzzy Systems and Technology (EUSFLAT 2011) and Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2011), Aix-les-Bains, France, July 18-22, 2011.
- 29e Congrès INFORSID, Lille, France, May 24-26, 2011.
- 5th International Scalable Uncertainty Management Conference (SUM 2011), Dayton, OH, USA, October 10-12, 2011.
- VLDB 5th International Workshop on the Management of Uncertain Data (MUD 2011), Seattle, WA, USA, September 2, 2011.
- DEXA 6th International Workshop on Flexible Database and Information Systems Technology (FlexDBIST 2011), Toulouse, France, August 29-September 2, 2011.
- 27es Journées Bases de Données Avancées (BDA 2011), Rabat, Morocco, October 24-27, 2011.

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

- 2e Atelier Graphes et Appariement d'Objects Complexes (GAOC 2011), in conjunction with EGC 2011, Brest, January 25, 2011.
- Joint 7th Conference of the European Society of Fuzzy Systems and Technology (EUSFLAT 2011) and Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2011), Aix-les-Bains, France, July 18-22, 2011.

8.2.3 Editorial boards

Patrick Bosc is a member of the following editorial boards:

- IEEE Transactions of Fuzzy Systems,
- International Journal on Fuzziness, Uncertainty and Knowledge-Based Systems,
- Fuzzy Sets and Systems,
- Revue I3.

Allel Hadjali is a member of the following editorial board:

- Journal of Advanced Computing Technologies.

Olivier Pivert is a member of the following editorial boards:

- Journal of Intelligent Information Systems,
- Fuzzy Sets and Systems.

8.2.4 Edition of special issues

- Patrick Bosc, Allel Hadjali, and Olivier Pivert are involved as Guest Editors in the edition of a special issue of the international journal Fuzzy Sets and Systems, devoted to "Advances in Soft Computing Applied to Database and Information Systems", following a special session on this topic at the IFSA/EUSFLAT'09 international joint conference.

- Ludovic Liétard is involved in the edition of a special issue of the "Information-Interaction-Intelligence (I3)" Journal, devoted to "Complex Object Matching and Discover", with co-editor Daniela Grigori (PRiSM, University of Versailles St-Quentin en Yvelines, France).

8.2.5 Organization of special sessions

- Allel Hadjali and Olivier Pivert organized, along with Guy de Tré (from Univ. Ghent, Belgium) and Sławomir Zadrożny (from Systems Research Institute, Warsaw, Poland), a special session dedicated to "Fuzzy database and information retrieval systems" at the Joint 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT'11) and Rencontres Francophones sur la Logique Floue et ses Applications (LFA'11).

8.2.6 Invited talks

- Grégory Smits gave an invited talk about "Flexible queries, cooperative answering, and plethoric answer sets" at the LIP6 (Paris, Jussieu) seminary devoted to "Data and Machine Learning" (DAPA) on March 24, 2011.

9 Bibliography

Major publications by the team in recent years

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

[2] P.Bosc, A. Hadjali, O. Pivert, "Incremental controlled relaxation of failing flexible queries", *Journal of Intelligent Information Systems 33*, 3, 2009, p. 261–283.

[3] 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.

[4] 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.

Doctoral dissertations and "Habilitation" theses

[1] A. Hadjali, *Gradualité, incertitude et préférence*, PhD Thesis, Habilitation à Diriger les Recherches – University of Rennes 1, 7 décembre 2011.

[2] A. Mokhtari, *Système personnalisé de planification d'itinéraire unimodal : Une approche basée sur la théorie des ensembles flous*, PhD Thesis, University of Rennes 1 – Ecole doctorale MATISSE, 25 mars 2011, supervised by O. Pivert and A. Hadjali.

Articles in referred journals and book chapters

[3] K. Benouaret, D. Benslimane, A. Hadjali, "A Fuzzy Framework For Selecting Top-k Web Service Compositions", *Applied Computing Review 11*, 3, 2011, p. 32–40.

[4] P. Bosc, A. Hadjali, O. Pivert, "On Database Queries Involving Competitive Conditional Preferences", *International Journal of Intelligent Systems 26*, 3, 2011, p. 206–227.

[5] P. Bosc, O. Pivert, O. Soufflet, "On Three Classes of Division Queries Involving Ordinal Preferences", *Journal of Intelligent Information Systems 37*, 3, 2011, p. 315–331.

[6] P. Bosc, O. Pivert, O. Soufflet, "Strict and Tolerant Antidivision Queries with Ordinal Layered Preferences", *International Journal of Approximate Reasoning 52*, 1, 2011, p. 38–48.

[7] P. Bosc, O. Pivert, "On Diverse Approaches to Bipolar Division Operators", *International Journal of Intelligent Systems 26*, 10, 2011, p. 911–929.

[8] A. Hadjali, S. Kaci, H. Prade, "Database Preference Queries — A Possibilistic Logic Approach with Symbolic Priorities", *Ann. Math. Artif. Intell. 63*, 3-4, 2011, p. 357–383.

Publications in Conferences and Workshops

[9] K. Abbaci, A. Hadjali, L. Liétard, D. Rocacher, "A similarity skyline approach for handling graph queries – A preliminary report", *in : Proc. of the 2nd ICDE International Workshop on Graph Data Management: Techniques and Applications (GDM'11)*, p. 112–117, Hannover, Germany, 2011.

[10] K. Abbaci, A. Hadjali, L. Liétard, D. Rocacher, "Une approche skyline pour l'interrogation de bases de données de graphes", *in : Actes du 2ème Atelier Graphes et Appariement d'Objets Complexes (GAOC'11) conjoint à la 11ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC'11)*, p. 14–25, Brest, France, 2011.

[11] K. Abbaci, F. Lemos, A. Hadjali, D. Grigori, L. Liétard, D. Rocacher, M. Bouzeghoub, "An Approach Based on Fuzzy Sets to Handle Preferences in Service Retrieval", *in : Proc. of the 13th IEEE Conference on Commerce and Enterprise Computing (CEC'11)*, p. 213–218, Luxembourg, Luxembourg, 2011.

[12] K. Abbaci, F. Lemos, A. Hadjali, D. Grigori, L. Liétard, D. Rocacher, M. Bouzeghoub, "A Cooperative Answering Approach to Fuzzy Preferences Queries in Service Discovery", *in : Proc. of the 9th International Conference on Flexible Query Answering Systems (FQAS'11), LNAI vol. 7022*, Springer, p. 318–329, Ghent, Belgium, 2011.

[13] K. Abbaci, F. Lemos, A. Hadjali, D. Grigori, L. Liétard, D. Rocacher, M. Bouzeghoub, "A Fuzzy-Set-Based Approach to Handle Preferences in Service Retrieval", *in : Actes des 27èmes Journées Bases de Données Avancées*, Rabat, Maroc, 2011.

[14] K. Abbaci, F. Lemos, A. Hadjali, D. Grigori, L. Liétard, D. Rocacher, M. Bouzeghoub, "Selecting and Ranking Business Processes with Preferences: An Approach Based on Fuzzy Sets", *in : Proc. of the 19th International Conference on Cooperative Information Systems (CoopIS'11)*, p. 38–55, Crete, Grece, 2011.

[15] K. Benouaret, D. Benslimane, A. Hadjali, M. Barhamgi, "Top-k Web Service Compositions using Fuzzy Dominance Relationship", *in : Proc. of the 8th IEEE International Conference on Services Computing (SCC'11)*, p. 144–151, Washington, DC, USA, 2011.

[16] K. Benouaret, D. Benslimane, A. Hadjali, "On the use of fuzzy dominance for computing service skyline based on QoS", *in : Proc. of the 9th International Conference on Web Services (ICWS'11)*, p. 540–547, Washington, DC, USA, 2011.

[17] K. Benouaret, D. Benslimane, A. Hadjali, "Top-k service compositions: a fuzzy set-based approach", *in : Proc. of the 26th ACM Symposium on Applied Computing (SAC'11)*, p. 1033–1038, Taichung, Taiwan, 2011.

[18] D. Benslimane, K. Benouaret, A. Hadjali, M. Barhamgi, "FuDoCS: A web service composition system based on fuzzy dominance for preference query answering", *in : Proc. of the 37th International Conference on Very Large Data Bases (VLDB'11, Demo session)*, p. 1430–1433, Seattle, WA, USA, 2011.

[19] P. Bosc, A. Hadjali, O. Pivert, "On possibilistic skyline queries", *in : Proc. of the 9th International Conference on Flexible Query Answering Systems (FQAS'11), LNAI vol. 7022*, Springer, p. 412–423, Ghent, Belgium, 2011.

[20] P. Bosc, O. Pivert, G. Smits, "A preference query model based on a fusion of local orders", *in : Proc. of the 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'11), LNAI vol. 6717*, Springer, p. 725–736, Belfast, Northern Ireland, UK, 2011.

[21] P. Bosc, O. Pivert, "On a strengthening connective for flexible database querying", *in : Proc. of the 20th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'11)*, Taiwan, 2011.

[22] P. Bosc, O. Pivert, "On the negation of bipolar fuzzy conditions", *in : Proc. of the 30th International Conference of the North American Fuzzy Information Processing Society (NAFIPS'11)*, El Paso, TX, USA, 2011.

[23] P. Bosc, O. Pivert, "On three fuzzy connectives for flexible data retrieval and their axiomatization", *in : Proc. of the 26th ACM Symposium on Applied Computing (SAC'11)*, p. 1114–1118, Taichung, Taiwan, 2011.

[24] P. Bosc, O. Pivert, "On three semantics of queries in a possibilistic database framework", *in : Proc. of the World Conference on Soft Computing (WCSC'11)*, San Francisco, CA, USA, 2011.

[25] P. Bosc, O. Pivert, "Une algèbre relationnelle bipolaire floue", *in : Actes du 29ème Congrès INFORSID*, p. 79–94, Lille, France, 2011.

[26] A. Hadjali, A. Mokhtari, O. Pivert, "Une approche à base de règles floues pour les requêtes à préférences contextuelles", *in : Actes de la 11ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC 2011)*, p. 293–294, Brest, France, 2011.

[27] A. Hadjali, O. Pivert, H. Prade, "On different types of fuzzy skylines", *in : Proc. of the 19th International Symposium on Methodologies for Intelligent Systems (ISMIS'11), LNAI vol. 6804*, Springer, p. 581–591, Warsaw, Poland, 2011.

[28] A. Hadjali, O. Pivert, "A fuzzy-rule-based approach to the handling of inferred fuzzy predicates in database queries", *in : Proc. of the 9th International Conference on Flexible Query Answering Systems (FQAS'11), LNAI vol. 7022*, Springer, p. 448–459, Ghent, Belgium, 2011.

[29] H. Jaudoin, O. Pivert, "Answering fuzzy queries using imprecise views", *in : Proc. of the 15th East-European Conference on Advances in Database & Information Systems (ADBIS'11), LNCS vol. 6909*, Springer, p. 257–270, Vienna, Austria, 2011.

[30] L. Liétard, N. Tamani, D. Rocacher, "Fuzzy bipolar conditions of type "or else"", *in : Proc. of the 20th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'11)*, Taiwan, 2011.

[31] O. Pivert, A. Hadjali, G. Smits, "On database queries involving inferred fuzzy predicates", *in : Proc. of the 19th International Symposium on Methodologies for Intelligent Systems (ISMIS'11), LNAI vol. 6804*, Springer, p. 592–601, Warsaw, Poland, 2011.

[32] O. Pivert, A. Hadjali, G. Smits, "Searching for a compromise between satisfaction and diversity in database fuzzy querying", *in : Proc. of the Joint 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT'11) and Rencontres Francophones sur la Logique Floue et ses Applications (LFA'11)*, p. 402–408, Aix-Les-Bains, France, 2011.

[33] O. Pivert, G. Smits, A. Hadjali, H. Jaudoin, "Efficient detection of minimal failing subqueries in a fuzzy querying context", *in : Proc. of the 15th East-European Conference on Advances in Database & Information Systems (ADBIS'11), LNCS vol. 6909*, Springer, p. 243–256, Vienna, Austria, 2011.

[34] O. Pivert, G. Smits, A. Hadjali, "Processing fuzzy queries in a peer data management system using distributed fuzzy summaries", *in : Proc. of the 5th International Conference on Scalable Uncertainty Management (SUM'11), LNAI vol. 6929*, Springer, p. 359–372, Dayton, OH, USA, 2011.

[35] N. Tamani, L. Liétard, D. Rocacher, "Bipolar SQLf: A Flexible Querying Language for Relational Databases", *in : Proc. of the 9th International Conference on Flexible Query Answering Systems (FQAS'11), LNAI vol. 7022*, Springer, p. 472–484, Ghent, Belgium, 2011.

[36] N. Tamani, L. Liétard, D. Rocacher, "Bipolarity and the relational division", *in : Proc. of the Joint 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT'11) and Rencontres Francophones sur la Logique Floue et ses Applications (LFA'11)*, p. 424–430, Aix-Les-Bains, France, 2011.

[37] N. Tamani, L. Liétard, D. Rocacher, "Bipolarity in flexible querying of information systems dedicated to multimodal transport networks", *in : Proc. of the 10th International Symposium on Programming and Systems (ISPS'11)*, p. 108–115, Algiers, Algeria, 2011.

[38] N. Tamani, L. Liétard, D. Rocacher, "An extension of a fuzzy ontology for flexible querying", *in : Proc. of the 20th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'11)*, Taiwan, 2011.

[39] N. Tamani, L. Liétard, D. Rocacher, "Fuzzy bipolar DLR-Lite based approach for flexible querying of information systems", *in : Proc. of the World Conference on Soft Computing (WCSC'11)*, San Francisco, CA, USA, 2011.

[40] N. Tamani, L. Liétard, D. Rocacher, "Linguistic quantifiers and bipolarity", *in : Proc. of the Joint World Congress of International Fuzzy Systems Association (IFSA'11) and Asia Fuzzy Systems Society International Conference (AFSS'11)*, p. (FT–101–1)–(FT–101–7), Surabaya-Bali, Indonesia, 2011.