UMR IRISA

Project-Team PILGRIM

Gradedness, Imprecision, and Mediation in Database Management Systems

Lannion

Activity Report

2010

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]:

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

[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 Postgres, 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).

- 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 (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.

- 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 is currently implementing a flexible route planner (cf. Subsection 6.3) which will be integrated into that platform.

6 New Results

6.1 Possibilistic database modeling and querying

Participants: Patrick Bosc, Olivier Pivert.

- Modeling and querying uncertain relational databases. In [2], we gave an overview of the most representative approaches aimed at querying databases containing ill-known data, starting from the pioneering works by Codd and Lipski and up to very recent proposals. This study focuses on approaches with a clear and sound semantics, based on the notion of possible worlds. Three types of queries are considered: i) those about attribute values (in an algebraic or SQL-like framework), ii) those about the properties satisfied by a given set of worlds (i.e., a set of instances of an imprecise database), and iii) those about the representation of uncertain data. For the first two types, it is emphasized that a trade-off has to be found between expressivity (of the model) and tractability (of the queries in the context of a given model).

- Certainty-based model. In [12, 11], we dealt with modeling and querying databases containing uncertain attribute values, in the situation where some knowledge is available about the more or less certain value (or disjunction of values) that a given attribute in a given tuple can take. This is represented in the setting of possibility theory. A relational database model suited to this context was introduced and the operators of relational algebra were extended so as to handle such relations. It was shown that i) the proposed model is a strong representation system for the whole relational algebra, and that ii) the data complexity associated with the extended operators in this context is the same as in the classical database case, which makes the approach highly scalable.

- Preference queries to an uncertain database. In [10], it is shown how three different types of hierarchical queries involving qualitative preference and leading to totally ordered results, which have been recently considered in the setting of division operation, can be conveniently encoded in possibilistic logic. This enables such a handling of preference queries to be interfaced with a recently proposed approach for dealing with pieces of data associated with certainty levels represented in the framework of possibilistic logic.

6.2 Flexible querying of classical databases

6.2.1 Preference modeling

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

- Outranking. In [14], we describe an approach to database preference queries based on the notion of outranking, 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 based on the concept of dominance. Even though outranking does not define an order in the strict sense of the term, we describe a technique which yields a complete pre-order, based on a global aggregation of the outranking degrees computed for each pair of tuples. The model proposed in [13] constitutes an alternative to the approach introduced in [14] where i) indifferent criteria were agregated with concordant ones in the calculus of the outranking degree ("broad" preference model), whereas the model described in [13] founds the comparison of tuples on discriminating preferences only ("strict" preference model), ii) the transition between the notions of concordance and indifference (resp. indifference and discordance) was crisp in [14] whereas it is fuzzy in this revisited model.

- Contextual preferences. In [21], 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. To this end, we present an approach based on generalized modus ponens.

- Contextual skyline queries. In [22], we propose a possibility theory-based approach to the treatment of missing user preferences in skyline queries. To compensate this lack of knowledge, we show how a set of plausible preferences suitable for the current context can be derived either in a case-based reasoning manner, or using an extended possibilistic logic setting. Uncertain dominance relationships are defined in a possibilistic way and the notion of possibilistic contextual skyline is introduced. [22] also includes a structured overview of the different types of "fuzzy" skylines.

- Fuzzy group-by [16, 15]. Group-by is a core database operation that is used extensively in data analysis and decision support systems. In many application scenarios, it appears useful to group values according to their compliance with a certain concept instead of founding the grouping on value equality. In [16], we propose a new SQLf construct that supports fuzzy-partition-based group-by (FGB). We show in [15] that FGB can be used to generate fuzzy summaries as well as to mine fuzzy association rules in a practical and efficient way.

6.2.2 Extended division operators

Participants: Patrick Bosc, Olivier Pivert.

- Stratified division and a weakening/strengthening mechanism. In [19], 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.

- Queries mixing division and antidivision and their relaxed/strenghtened versions. In [20, 3], we dealt with queries involving two components: one describing desired associations, another specifying forbidden associations. In addition, those preferences are cited in the context of a hierarchy expressing some strength about what is wanted and rejected. So doing an ordinal ordering over the answers is made available in order to distinguish among the elements of the answer. The situation where no answer is returned is also tackled and it is proposed to soften the initial query in order to get a non-empty answer. It is shown that this can be achieved still using an ordinal framework which is an extension of the initial one.

- Application to IR. Some IR models make use of an implication to match a document d and a query q, computing either "q implies d" (e.g. in fuzzy inclusion models) or, the other way, "d implies q" (e.g. in logical IR models). In [29], the IR models using both approaches are analyzed from a theoretical point of view. Even if the above notations seem to be opposite, it is shown that they sometimes come from different formulations of the same paradigm, which led to mistakes in the literature. Then the paper comes back to fuzzy models based on "q implies d" (q included in d) and shows their efficiency, and compares them to models based on "d implies q" (d included in q). The latter is attractive from a theoretical point of view, but turns out to be less efficient in practice, and is rarely adopted in the literature. At last, attempts to use "d implies q" in a fuzzy model are discussed, and we try to explain their inefficiency.

6.2.3 Bipolar fuzzy queries

Participants: Patrick Bosc, Olivier Pivert, Ludovic Liétard, Daniel Rocacher, Amine Mokhtari.

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 [9], 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.

- Division of bipolar relations. In [17], we dealt with the relational division operation and extended it so that it can handle bipolar fuzzy relations, i.e., relations where each tuple t is attached a pair of grades in the unit interval expressing the extent to which t satisfies a flexible constraint and a flexible wish tied by a consistency condition. The framework considered is that of an extended relational algebra. It is shown that the result of a division of bipolar relations can be characterized as a twofold quotient. The question of the (non-) primitivity of the operator is also tackled and we have pointed out an equivalent (even though complex) formulation for the division of bipolar fuzzy relations.

- Negation of bipolar fuzzy conditions. In [18], 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.

6.2.4 Cooperative answering to flexible database queries

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

- Plethoric answers. Retrieving data from large-scale databases often leads to plethoric answers. To overcome this problem, we propose in [8] an approach which selects a set of predicates that the user may use for augmenting his/her query. These predicates are selected among predefined ones according principally to their degree of semantic correlation with the initial query. This makes it possible to avoid a deep modification of the semantic scope of the user query. The extension of this approach to the *fuzzy* query case is presented in [7]. A further refinement of the approach is given in [6], where it is suggested to use fuzzy cardinalities to assess the reduction capability of the correlated predefined predicates added to the query. A prototype that implements this approach, named CORTEX, is described in [5].

- Failing fuzzy queries. In [24], we investigate the problem of handling of failing queries involving fuzzy predicates. We propose an approach that leverages data distribution of the target database. It involves two steps: i) query translation that aims at translating the failing fuzzy query into a crisp query by means of a particular semantic distance between sets; ii) query relaxation which consists in expanding the translated query criteria with similar values. A method is also proposed to rank-order the approximate query results.
- Empty answers and predicate substitution. In [26], we proposed an approach aimed at obviating empty answers for a family of conjunctive queries involving value constraints. Contrary to the approaches based on a relaxation of the predicates involved in the query, the principle suggested in [26] consists in replacing the query by a similar one which has been processed previously and whose answer is known to be non-empty. This case-based reasoning technique thus avoids the combinatory explosion induced by classical relaxation-based approaches.

6.3 Fuzzy preferences in intelligent transportation systems

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

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 2010, the results obtained concern the following issues:

- Extension of a fuzzy ontology for flexible querying of an embedded database. In [28] a personalized approach for flexible information systems querying is proposed. It is based on the association of the reasoning capacities of the fuzzy DLRLite and the expressiveness of the SQLf language. The interpretation of the inclusion is based on Gödel's fuzzy implication and its generalization with a tree of inclusions is introduced. This tree and its property of degrees propagation are the basic elements of our application context. The illustrative example consists in querying an information system dedicated to multimodal transportation networks which is embedded in user terminal characterized by a limited storage and processing capabilities.

6.4 Flexibility issues in data integration systems

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

- The ANR project FORUM, which Pilgrim was part of and which ended in June 2009, aimed at extending existing data integration techniques in order to facilitate the development of mediation systems in large and dynamic environments. It is well known from the literature that a crucial point that hampers the development and wide adoption of mediation systems lies in the high entry and maintenance costs of such systems. To overcome these barriers, the FORUM project investigated three main research issues: (i) automatic discovery of schema matchings, (ii) consistency maintenance of mappings, and (iii) tolerant rewriting of queries in the presence of approximate mappings. The contributions of the project are presented in [1].

- Processing fuzzy queries in a Peer Data Management System (PDMS) using fuzzy summaries. In [25], 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. An alternative approach is presented in [4], which uses a different type of fuzzy summary, and exploits a distributed routing index in order to find the most relevant peers.

- Rewriting a fuzzy query using imprecise views. In [23], 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.

- Handling dirty integrated databases: from user warning to data cleaning. One can conceive many reasonable ways of characterizing how dirty a database is with respect to a set of integrity constraints (e.g., functional dependencies). However, dirtiness measures, as good as they can be, are difficult to interpret for an end-user and do not give the database administrator much hint about how to clean the base. In [27], these aspects are discussed and some methods are proposed which aim at both the user and the administrator for overcoming the limitations of dirtiness measures when it comes to handling dirty databases.

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).

- Incitative Action "Projets scientifiques émergents 2010" (Univ. Rennes 1). Topic: Flexible data integration system. Leader: Hélène Jaudoin. The main objective of this project is to define formal tools for building and querying data integration systems which are "uncertainty aware". We focus on the following topics: i) definition of possible global schemas from those attached with data sources, ii) specification of the semantics of queries on possible schemas and iii) conception of efficient algorithms to answer such queries.

- PME project of the "Pôle Images et Réseaux", named IntelSearch, in collaboration with Semsoft, Swid, and Ensai. Topic: Flexible data integration system. Leader for Pilgrim: Hélène Jaudoin. The purpose of this project is to provide an innovative software to query audiovisual contents. This software lies on a mediation architecture in which our contribution mainly concerns the definition of a flexible querying approach. A first objective is to rank query rewritings according to the probability that they correctly answer a given query. Secondly, we aim at defining a query rewriting technique that authorizes a form of *exceptions*.

7.2 International actions

Lakhdar Amrani, Associate Professor at the University Ferhat Abbas in Sétif (Algeria), spent one week in our team, from January 3 to January 8, 2010.

8 Dissemination

8.1 Teaching

Project members give lectures in different faculties of engineering, in the third cycle University curriculum: ”Bases de données, gradualité et imprécision“ in the speciality ”Intelligence Artificielle et Images“ 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 2010.

8.2 Scientific activities

8.2.1 Highlights of the year

- Organization of LFA 2010, Lannion, November 18–19.

8.2.2 Program committees

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

- 25^{th} ACM Symposium on Applied Computing (SAC'10), Special Track on Information Access and Retrieval, Sierre, Switzerland, March 22–26, 2010.
- 2^{nd} International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2010), Menuires, France, April 11–16, 2010.
- 13^{th} International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2010), Dortmund, Germany, June 28–July 2, 2010.
- 10^{emes} Journées Francophones Extraction et Gestion des Connaissances (EGC 2010), Hammamet, Tunisia, Jan. 27–29, 2010.
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2010), Lannion, France, Nov. 18–19, 2010.
- IEEE/WIC/ACM International Conference on Web Intelligence (WI'10), Toronto, Canada, August 31–September 3, 2010.
- 19^{th} IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'10), Barcelona, Spain, July 18–23, 2010.
- VLDB 2010 4^{th} International Workshop on the Management of Uncertain Data (MUD 2010), Singapore, September 13, 2010.
- DEXA 2010 5^{th} International Workshop on Flexible Database and Information Systems Technology (FlexDBIST 2010), Bilbao, Spain, August 30-September 3, 2010.
- 29^{th} International Conference of the North American Fuzzy Information Processing Society (NAFIPS'10), Toronto, Canada, July 12–14, 2010.
- 21^{st} Int. Conference on Database and Expert Systems Applications (DEXA'10), Bilbao, Spain, August 30-September 3, 2010.
- 3^{rd} North American Simulation Technology Conference (NASTEC 2010), Redmond-Seattle, USA, August 25–27, 2010.

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

- Workshop GAOC (Graphes et Appariement d'Objets Complexes), held in conjunction with EGC'10, Hammamet, Tunisia, Jan. 26, 2010.
- Colloque sur l'Optimisation et les Systèmes d'Information, Ouargla, Algeria, 18–20 avril, 2010.
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2010), Lannion (France), 18–19 novembre, 2010.

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

- Colloque sur l'Optimisation et les Systèmes d'Information, Ouargla, Algeria, 18–20 avril, 2010.

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

- Workshop GAOC (Graphes et Appariement d'Objets Complexes), held in conjunction with EGC'10, Jan. 26, 2010.

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

- 25^{th} ACM Symposium on Applied Computing (SAC'10), Special Track on Information Access and Retrieval, Sierre, Switzerland, March 22–26, 2010.
- 13^{th} International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2010), Dortmund, Germany, June 28–July 2, 2010.
- 5^{th} IEEE International Conference on Intelligent Systems (IS 2010), London, UK, July 7–9, 2010.
- WI/IAT'10 Workshop on Soft approaches to Information Access on the Web, Toronto, Canada, August 31–September 3, 2010.
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2010), Lannion (France), 18–19 novembre, 2010.
- VLDB 2010 4^{th} International Workshop on the Management of Uncertain Data (MUD 2010), Singapore, September 13, 2010.
- DEXA 2010 5^{th} International Workshop on Flexible Database and Information Systems Technology (FlexDBIST 2010), Bilbao, Spain, August 30-September 3, 2010.

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

- Workshop GAOC (Graphes et Appariement d'Objets Complexes), held in conjunction with EGC'10, Jan. 26, 2010.
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2010), Lannion (France), 18–19 novembre 2010.

8.2.3 Organizing committees

Olivier Pivert chaired the following organizing committee:

- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2010), Lannion (France), 18–19 novembre 2010

whose members also included Katia Abbaci, Allel Hadjali, Hélène Jaudoin, Ludovic Liétard, Amine Mokhtari, Daniel Rocacher, Grégory Smits, and Nouredine Tamani.

8.2.4 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.

Olivier Pivert is a member of the following editorial board:

- Journal of Intelligent Information Systems.

8.2.5 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 IFA/EUSFLAT'09 international joint conference.

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.

Articles in referred journals and book chapters

[1] Z. Bellahsene, S. Benbernou, H. Jaudoin, F. Pinet, O. Pivert, F. Toumani, S. Bernard, P. Colomb, R. Coletta, E. Coquery, F. de Marchi, F. Duchâteau, M. Hacid, A. Hadjali, M. Roche, "FORUM: A Flexible Data Integration System Based on Data Semantics", *SIGMOD Record 39*, 2, 2010, p. 11–18.

[2] 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.

[3] P. Bosc, O. Pivert, "On some uses of a stratified divisor in an ordinal framework", *in: Uncertainty Approaches for Spatial Data Modeling and Processing: A Decision Support Perspective*, F. Petry, J. Kacprzyk, and A. Yazici (editors), *Studies in Computational Intelligence, vol. 271*, Springer, 2010, p. 133–154.

Publications in Conferences and Workshops

[4] A. Alem, A. Hadjali, "Evaluation de requêtes flexibles dans un contexte non-centralisé : Une approche basée sur les résumés distribués", *in: Actes du Colloque sur l'Optimisation et les Systèmes d'Information (COSI'10)*, Ouargla, Algeria, 2010.

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

[6] P. Bosc, A. Hadjali, O. Pivert, G. Smits, "On the use of fuzzy cardinalities for reducing plethoric answers to fuzzy queries", *in: Proc. of the 4th International Conference on Scalable Uncertainty Management (SUM'10), LNAI vol. 6379*, Springer, p. 98–111, Toulouse, France, 2010.

[7] P. Bosc, A. Hadjali, O. Pivert, G. Smits, "Trimming plethoric answers to fuzzy queries: An approach based on predicate correlation", *in: Computational Intelligence for Knowledge-Based Systems Design (Proc. of IPMU'10)*, E. Hüllermeier, R. Kruse, F. Hoffmann (editors), *Lecture Notes in Computer Science, 6178*, Springer, p. 595–604, 2010.

[8] P. Bosc, A. Hadjali, O. Pivert, G. Smits, "Une approche fondée sur la corrélation entre prédicats pour le traitement des réponses pléthoriques", *in: Actes de la 10ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (EGC'10)*, p. 273–284, Hammamet, Tunisie, 2010.

[9] P. Bosc, O. Pivert, L. Liétard, A. Mokhtari, "Extending relational algebra to handle bipolarity", *in: Proc. of the 25th ACM Symposium on Applied Computing (SAC'10)*, p. 1718–1722, Sierre, Switzerland, 2010.

[10] P. Bosc, O. Pivert, H. Prade, "A possibilistic logic view of preference queries to an uncertain database", *in: Proc. of the 19th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'10)*, p. 379–384, Barcelona, Spain, 2010.

[11] P. Bosc, O. Pivert, H. Prade, "Un modèle fondé sur la certitude pour les bases de données incertaines", *in: Actes des 26èmes Journées Bases de Données Avancées (BDA'10)*, 2010.

[12] P. Bosc, O. Pivert, H. Prade, "An uncertain database model and a query algebra based on possibilistic certainty", *in: Proc. of the 2nd International Conference on Soft Computing and Pattern Recognition (SoCPaR'10)*, p. 63–68, Cergy-Pontoise, France, 2010.

[13] P. Bosc, O. Pivert, G. Smits, "A database preference query model based on a fuzzy outranking relation", *in: Proc. of the 19th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'10)*, p. 38–43, Barcelona, Spain, 2010.

[14] P. Bosc, O. Pivert, G. Smits, "A model based on outranking for database preference queries", *in: Information Processing and Management of Uncertainty in Knowledge-Based Systems (Proc. of IPMU'10)*, E. Hüllermeier, R. Kruse, F. Hoffmann (editors), *Communications in Computer and Information Science, 81*, Springer, p. 95–104, 2010.

[15] P. Bosc, O. Pivert, G. Smits, "On a fuzzy group-by and its use for fuzzy association rule mining", *in: Proc. of the 14th East-European Conference on Advances in Databases and Information Systems (ADBIS'10), LNCS vol. 6295*, Springer, p. 88–102, Novi Sad, Serbia, 2010.

[16] P. Bosc, O. Pivert, "On a fuzzy group-by clause in SQLf", *in: Proc. of the 19th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'10)*, p. 2409–2414, Barcelona, Spain, 2010.

[17] P. Bosc, O. Pivert, "On the division of bipolar fuzzy relations", *in: Proc. of the 29th International Conference of the North American Fuzzy Information Processing Society (NAFIPS'10)*, Toronto, Canada, 2010.

[18] P. Bosc, O. Pivert, "A propos de la négation de conditions floues bipolaires", *in: Actes des Rencontres Francophones sur la Logique Floue et ses Applications (LFA'10)*, p. 21–28, Lannion, France, 2010.

[19] P. Bosc, O. Pivert, "A propos de requêtes à préférences et diviseur stratifié", *in: Actes du 28ème Congrès INFORSID*, p. 311–326, Marseille, France, 2010.

[20] P. Bosc, O. Pivert, "Queries mixing positive and negative associations and their weakening", *in: Proc. of the 29th International Conference of the North American Fuzzy Information Processing Society (NAFIPS'10)*, Toronto, Canada, 2010.

[21] A. Hadjali, A. Mokhtari, O. Pivert, "A fuzzy-rule-based approach to contextual preference queries", *in: Computational Intelligence for Knowledge-Based Systems Design (Proc. of IPMU'10)*, E. Hüllermeier, R. Kruse, F. Hoffmann (editors), *Lecture Notes in Computer Science, 6178*, Springer, p. 532–541, 2010.

[22] A. Hadjali, O. Pivert, H. Prade, "Possibilistic contextual skylines with incomplete preferences", *in: Proc. of the 2nd International Conference on Soft Computing and Pattern Recognition (SoCPaR'10)*, p. 57–62, Cergy-Pontoise, France, 2010.

[23] H. Jaudoin, O. Pivert, "Répondre à une requête floue en termes de vues imprécises", *in: Actes des Rencontres Francophones sur la Logique Floue et ses Applications (LFA'10)*, p. 37–44, Lannion, France, 2010.

[24] L. Oudjoudi, A. Hadjali, "Supporting failing database queries in a flexible context: A novel approach", *in: Actes du Colloque sur l'Optimisation et les Systèmes d'Information (COSI'10)*, Ouargla, Algeria, 2010.

[25] O. Pivert, A. Hadjali, G. Smits, "Estimating the relevance of a data source using a fuzzy-cardinality-based summary", *in: Proc. of the 5th IEEE International Conference on Intelligent Systems (IEEE IS'10)*, p. 96–101, London, Great-Britain, 2010.

[26] O. Pivert, H. Jaudoin, C. Brando, A. Hadjali, "A method based on query caching and predicate substitution for the treatment of failing database queries", *in: Proc. of the 18th International Conference on Case-Based Reasoning (ICCBR'10), LNAI 6176*, p. 436–450, Alessandria, Italy, 2010.

[27] O. Pivert, H. Prade, "Handling dirty databases: From user warning to data cleaning — Towards an interactive approach", *in: Proc. of the 4th International Conference on Scalable Uncertainty Management (SUM'10), LNAI vol. 6379*, Springer, p. 292–305, Toulouse, France, 2010.

[28] N. Tamani, L. Liétard, D. Rocacher, "Extension d'une ontologie floue pour l'interrogation flexible d'une base de données embarquée", *in : Actes des Rencontres Francophones sur la Logique Floue et ses Applications (LFA'10)*, p. 29–36, Lannion, France, 2010.

[29] L. Ughetto, G. Pasi, V. Claveau, O. Pivert, P. Bosc, "Implication in information retrieval systems", *in : Proc. of the 9th International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information (RIAO'10)*, Paris, France, 2010.