# Project-Team: BADINS

Bases de données et interrogation souple
Theme 3a
Rennes

November 25, 2005

# Contents

# 1 Team

**debug :** Module named "" for the project BADINS, section composition

**Head of the team**

Patrick Bosc [professor ENSSAT]

**Staff UMR, Located in Lannion**

Ludovic Liétard [assistant professor IUT-Lannion]

Olivier Pivert [assistant professor ENSSAT]

Daniel Rocacher [assistant professor ENSSAT]

Allel Hadjali [assistant professor ENSSAT]

Nadia Liétard [Ph.D. Student (regional grant) from October 1, 2004]

Salah-Eddine Tbahriti [Ph.D. Student (regional grant) from December 1, 2005]

# 2 Overall Objectives

**debug :** Module named "presentation" for the project BADINS, section presentation

**debug : body of the module**.

In a majority of works undertaken in the area of database management systems (DBMSs), it is assumed that data are perfectly known and that only Boolean queries can be expressed. The aim of the project is to soften this double hypothesis by introducing the notions of imprecision and graduality in information systems. Such concepts impact three aspects of database systems:

1. the expression of queries addressed to DBMSs, which become gradual or flexible. The idea is to allow the user to specify preferences instead of Boolean conditions, thus leading to a result whose elements are ordered accordingly, e.g., find the *young* employees who work in a *high* budget department,

2. the expression of properties calling on vague qualifications in the framework of data mining processes, e.g., employees with *similar* experiences have *close* salaries,

3. the description and the manipulation of imprecisely known data. Such data can be expressed in a linguistic way, e.g., a reservoir which is rather *big* or a person who is *young*.

The project reconsiders several aspects of DBMSs by taking imprecision and graduality into account. From a querying point of view, these two characteristics are indeed orthogonal, i.e., one may consider regular queries addressed to databases containing imprecise data, or flexible queries against regular databases. The originality of the project is to rely on a common scientific framework, namely fuzzy sets, to deal with these two sides and then to be able to consider the "joint" issue of flexible queries addressed to databases containing imprecise data.

The issue of flexible querying has received an increasing attention in the last years in the database community and the idea of including preferences inside queries gets more and more acceptance, even if there is a counterpart in terms of performances. The line followed in Badins 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 both relational and object-oriented databases.

Basically, a flexible 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, one 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 car" is meant for 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* 1500$. 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.

Since a couple of years, another important line of research concerns the querying of databases in the presence of imprecise data. We think that this type of database should receive more and more attention in the future, for instance in the context of automated recognition, data fusion, forecasts or incomplete archives, in which imprecision is intrinsically present. Some works in the 80s have established that null values and disjunctive data have a strong impact on the queries that can be processed in an efficient way. This situation remains true in the context considered in Badins, where the objective is to investigate various types of queries that can be processed efficiently in the relational setting. Although there is no hope for a general family of queries which would include all of the operators of the relational algebra, the identification of its largest subset which fits our requirements is undertaken. Beyond this target, specific queries that make sense and are tractable are sought, such as queries of the type: "to what extent is it possible that tuple t belongs to the result of query Q", where t is a given tuple and Q is an algebraic query. Such queries are called possibilistic queries (due to the presence of the word possible in its statement) and they stem from regular yes/no queries.

Querying imprecise data calls on a given theory of uncertainty. Possibility theory is used in Badins, essentially because it provides a framework which is coherent with the one serving for flexible queries. So doing, it becomes possible to study the issue of flexible queries against imprecise databases, even if this is not our first priority. It is worth noticing that the case where probabilistic data are used instead of possibilistic ones, is also a matter of interest.

Data mining is a third topic of research where the objective is to assess what fuzzy sets can bring in order to express properties. In other words, the idea is to discover properties inside which gradual linguistic terms are useful. The interest of the approach is to allow for a compact robust representation of properties relying on the fact that the terms used are not clear-cut. Different types of rules can be considered, among which extended association rules, gradual rules such as "the *lower* the salary, the *longer* the trip from home to work" and quantified propositions like "*most* of *low qualified* employees have a *fairly long* trip from home to work".

# 3 Scientific Foundations

**debug :** Module named "fondements" for the project BADINS, section fondements

**Abstract:** *The project investigates the issues of flexible queries against regular databases, regular queries addressed to databases involving imprecise data, as well as the discovery of properties calling on gradual predicates or terms. These various 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

**debug :** Module named "fondements2" for the project BADINS, section fondements
**debug : body of the module**.

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

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

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 $\alpha$ level-cut of a fuzzy set F is defined as the (regular) set of elements whose degree of membership is equal or over $\alpha$ 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 $\bar{A}$, is a fuzzy set such that: $\mu_{\bar{A}}(x) = neg(\mu_A(x))$, where neg is a strong negation operator and the complement to 1 is almost always 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, other 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 is more or less included in F. Different approaches can be envisaged, among which one is based on the notion of 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 $\Rightarrow_G$ 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, such as mean operators, which do not make sense for classical sets.

### 3.2 Possibility theory

**debug :** Module named “fondements3” for the project BADINS, section fondements
**debug : body of the module**.

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

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

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

[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.

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

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

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

### 3.3 Fuzzy sets, possibility theory and databases

**debug :** Module named "fondements4" for the project BADINS, section fondements
**debug : body of the module**.

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 Badins 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,

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

- the handling of fuzzy bags (fuzzy multi sets) and their connection with fuzzy numbers.

As to possibility distributions, they are used to represent imprecise (imperfect) data. So doing, a straightforward connection can be established between a possibilistic database [PT84] and regular ones. Indeed, a possibilistic database is nothing but a weighted set of regular databases (called worlds), obtained by choosing one candidate in every distribution appearing in any tuple of any 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 combinatorial 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).

## 4 Application Domains

**debug :** Module named "domaine" for the project BADINS, section domaine

**Abstract:** *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 Alta Vista use a hidden 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 tool for expressing information needs.*

*Databases involving imprecise data are not yet common in practice for two reasons: developing DBMSs supporting such data is probably a hard job and the demand is presently not 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

**debug :** Module named "logiciels" for the project BADINS, section logiciels

**Abstract:** *So far, software development has not been a high priority of the project. However, a certain number of applications have been developed and we*

[PT84] H. Prade, C. Testemale, "Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries", *Information Sciences 34*, 1984, p. 115–143.

*are currently implementing a querying system prototype, built above Oracle, that aims at evaluating fuzzy queries addressed to regular databases. More precisely, the objective is to add a software layer to Oracle, 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, called iSQLf, is able to process simple fuzzy queries (i.e., fuzzy queries involving a single block). It rests on a double client/server architecture. The iSQLf system is indeed a Web server but is also a client which accesses the DBMS via the local network. The implementation of this prototype is currently being pursued, the objective being to extend it in order to deal with fuzzy nested queries.*

*Querying possibilistic databases requires a first stage: the storage of imprecise data. This issue gave birth to a very first development using Oracle 8i. Two techniques have been compared: the use of user-defined types for the storage of possibility distributions along with specific methods for their manipulation and a more ad hoc approach where possibility distributions are stored in separate tables. The first results seem to indicate that the second way of doing is significantly more efficient than the former. Prototypes of kernel systems for dealing with different types of queries against such databases should also be tackled in the future.*

# 6 New Results

## 6.1 Querying databases containing imprecise attribute values

**debug :** Module named “resultats” for the project BADINS, section resultats

**Participants**: Patrick Bosc, Nadia Liétard, Olivier Pivert.

**Keywords**: Ill-known data, possibilistic databases, strong representation system, querying language.

**Abstract:** *Relational databases including ill-known attribute values represented by possibility distributions are considered. The main objective is to point out classes of queries that can be used to retrieve information from such databases, with the two following constraints: the queries must have a sound semantics, and the complexity of their evaluation must be reasonable.*

**debug : body of the module**.

Some well-known works conducted in the 80s have revealed that the presence of null values (in the sense of existing unknown values) in databases raises several serious problems. In particular, it is not feasible to extend the relational operators, defined for precise and complete relations, so as to make them work on relations containing imprecise information. Indeed, when null values are involved, the result of some operations (such as the join) cannot be represented by means of a "basic" relational table (a more sophisticated model is needed). Now, a null value only constitutes a special case of disjunction where no preferences are expressed over the candidates (which are exclusive). Therefore, the same problem obviously arises in the case of more elaborate ill-known values such as possibility distributions. It is thus important to identify classes of queries that can be used in a possibilistic database context (but it is worth noticing that this question also makes sense in a probabilistic database framework). Two main lines of research are investigated in the project:

- definition of a restricted querying algebra for possibilistic databases,
- study of user-oriented queries of the type "to which extent is it possible and certain that the answer to Q fulfills property P?" (where Q is an algebraic query and P is explicitly specified), called extended yes/no queries.

The first line aims at extending algebraic queries in the presence of ill-known data, i.e., at defining the algebraic operators that may be part of a querying language in a possibilistic database context. Let us recall that a possibilistic database corresponds to a set of interpretations (also called worlds) which are usual databases obtained by choosing a candidate value in each possibility distribution. The combinatorial nature of this mechanism leads to a considerable number of worlds even when the number of ill-known values in the possibilistic database is relatively small. A crucial objective is thus to define a "compact" processing method, i.e., a method that does not require to make the interpretations of the database explicit. Consequently, it is mandatory to design a data model that constitutes a strong representation system for the query language considered, i.e., a model that supports a closed set of operations whose results are consistent with a world-based interpretation. In other terms, if rep(D) denotes the set of worlds associated with the possibilistic database D, the following property must hold for any operation o of the language: o(rep(D)) = rep(o(D)). Such a framework guarantees a sound semantics for the operators of the language, and it makes it possible to envisage a tractable query evaluation process.

In the previous years, we showed that the "basic" possibilistic relational model (i.e., the regular relational model extended to the case where an attribute value may be a possibility distribution) had too limited an expression power (in terms of the operators it allows to define) to be useful in practice. Therefore, we adapted it and defined an extended possibilistic database model, which involves usual possibilistic relations enriched with:

- a degree N associated with every tuple, which expresses the certainty that the tuple has a representative in any world than can be derived from the relation,
- the introduction of nested relations used to represent dependencies between candidate values for different attributes and thus to model possibility distributions over several attributes.

We showed that this extended model constitutes a strong representation system for the operations of selection, projection and foreign key join (fk-join)[11]

In 2005, we mainly investigated the second line of research mentioned above and studied different ways a compact relation resulting from an algebraic query can be handled by a user. In the model we defined (see above), the result of an algebraic query is a "compact" table representing all the more or less possible worlds resulting from the query. One may think that this form of a result is not very practical for an end user and that some kind of interface allowing for its manipulation would be useful. This led us to study more user-oriented queries, called extended yes/no queries, whose general form is: "to what extent is it possible and certain that the answer to Q satisfies property P" where Q is an algebraic query. A "naïve" evaluation method would consist in processing Q in each world of the database and then to check whether property P is fulfilled by the result in each of these worlds, but this approach is clearly intractable due to the huge (possibly infinite) number of worlds. This observation led us to consider only specific queries which can be processed directly against the possibilistic database, while delivering a result equivalent to the one defined in terms of worlds. It has been shown [6] that these queries could be processed in a "compact" (then reasonable) way at the price of some restrictions as to the operations present in Q (the only "legal" operations are those for which the model constitutes a strong representation system).

Up to now, we have considered three possible forms for property P: P1) membership of a given tuple to the result: "to which extent is it possible and certain that tuple t belongs to the answer to Q?" [12], P2) vacuity of the result: "to which extent is it possible and certain that the answer to Q is non-empty?" [6], P3) condition on the cardinality of the result: "to which extent is it possible and certain that the answer to Q contains at least (at most, exactly) n elements ?" [18]. In each case, we defined an evaluation algorithm based on the "trial and error" pattern, that scans the set of possible worlds attached to the result of Q (the cardinality of this set is obviously much more limited than the number of worlds attached to the initial possibilistic database) in a non-exhaustive way thanks to the use of pruning conditions.

Interesting perspectives for further work include:

i) to perform some experimental studies in order to assess the performances of the processing method we proposed;

ii) to extend conditions of type P1 (membership) to the case of multiple tuples ("to what extent is it possible and certain that tuples t1, t2, ..., and tn belong jointly to the answer to Q");

iii) to extend conditions of type P3 so as to consider other types of aggregates than the cardinality (such as in: "to what extent is is possible and certain that the sum of the elements that constitute the answer to Q is over a given threshold");

iv) to extend conditions of type P3 in order to deal with quantified statements (such as in: "to what extent is is possible and certain that more than half of the elements that constitute the answer to Q satisfy a given condition C");

v) to study the impact of the presence of fuzzy conditions in query Q and/or property P on the processing method.

## 6.2 Fuzzy bags, fuzzy numbers and flexible querying

**debug :** Module named “resultats” for the project BADINS, section resultats

**Participants**: Patrick Bosc, Daniel Rocacher, Ludovic Liétard.

**Keywords**: fuzzy bag, gradual integer, gradual rational, quantified statement, flexible querying, object oriented databases, OQL.

**Abstract:** *This work takes place in a framework defining a query language which deals with quantification and preferences on data. In this context, fuzzy extensions of bags, integers and rational numbers are specifically studied.*

**debug : body of the module**.

Fuzzy set theory offers a general framework for dealing with flexible queries and priorities inside compound queries. The answers to such queries are then qualified and rank ordered. Besides, the bag type, which offers the ability to manage quantities (numbers of occurrences of data items), plays an important role in databases and data models (relational or object oriented) have been designed to support it. Systems taking into account both flexible queries and bags lead to deal with fuzzy bags. For example, a fuzzy bag can be obtained when some attributes are removed from a fuzzy set of tuples. This is illustrated by the query: *find the salaries of young employees* which calls on a projection (salary) of a fuzzy set of persons (the young employees) and delivers a fuzzy bag. As several employees may have the same salary, the collection of salaries returned may contain duplicates. Moreover, a given salary occurrence is associated with a more or less young employee and thus satisfies more or less the criterion "to be the salary of a young employee". Consequently, the different salaries returned by the query have to be managed both quantitatively and qualitatively thanks to a fuzzy bag which represents the distribution of the salaries of young employees. Our general objective is to devise new structures and queries able to deal with quantification and preferences over usual databases (relational or object-oriented). For examples, we may be interested in expressing requests about quantities, as in: *find the best five companies where most of the young employees are well-paid or find the best five companies in which the number of young employees is equal to or greater than the number of well-paid employees.* Even though we suggest here a role for fuzzy bags in the context of databases, it is worth mentioning that our investigations have a larger scope than the field of databases and many other potential application domains could also benefit from fuzzy bags, such as fuzzy data mining, summarization of data or fuzzy information retrieval.

A new approach to fuzzy bags based on the notion of fuzzy cardinalities of a fuzzy set has been proposed [Roc03]. In this proposition, an element x in a fuzzy bag A is characterized by the fuzzy cardinality of the fuzzy set of the occurrences of x in A, denoted by $\Omega_A(x)$. This kind of "fuzzy number", called a gradual natural integer (which belongs to the set $N_f$), is not interpreted as a possibility distribution describing the ill-known value of a variable, but is viewed as a conjunctive fuzzy set of integers providing an exact characterization of the number of elements belonging to A. This specificity has important consequences regarding the validity of group properties (in a mathematical meaning) of ($N_f$, $+$), which hold in this particular context. Fuzzy bags generalize sets, fuzzy sets and bags. Operations on crisp bags, based on operations on $N$, can be straightforwardly extended to fuzzy bags using operations on $N_f$. All of these collections (fuzzy or crisp) are homogeneous and compatible because they can be defined through a well-founded common framework. Consequently, a small number of generic operators can be applied to these different collections which are treated in a uniform way and can then be composed.

In this context, the difference between two fuzzy bags does not always exist and, in such cases, only approximations of the difference have been defined by several researchers. The problem comes from the fact that the fuzzy bag model considered so far is based on positive fuzzy integers. A first partial solution is to define approximations of the difference between fuzzy bags on $N_f$. A second solution is to enlarge our point of view. Because $N_f$ does not permit the existence of unrestricted subtraction, we have proposed the first ground for a new construction by introducing negative fuzzy integers. Thus, $N_f$ has been extended to $Z_f$ [14]. In this framework, the difference A - B of two fuzzy bags is always exactly defined in terms of one equivalence class of pairs of fuzzy natural integers. The method for defining such an extension is coherent with the one used for extending $N$ to $Z$. It has been shown that each equivalence class can be identified by a unique canonical representative and can easily be manipulated using $\alpha$-cuts. This approach has been followed up by extending $Z_f$ to $Q_f$ (the set of fuzzy rational numbers) where each number is defined in terms of one equivalence class of pairs of fuzzy relative integers also identified by a unique canonical representative [13]. These new frameworks provide an arithmetic basis where differences and ratios between fuzzy quantities can be exactly evaluated. For example, the query "find the average salary of well-paid employees", or the evaluation of inclusion grades and similarity measures between two fuzzy sets A and B, leads to perform an exact evaluation of a division on $Q_f$. The obtained results can then be used inside more complex calculations and composed. Furthermore, from an exact evaluation of an arithmetic expression on $Z_f$ or $Q_f$, it is possible to extract different approximations, on $N_f$ or $R$, for example, depending on users or applications needs.

Complementary studies have to be done to define fuzzy order relations between fuzzy numbers. Such comparisons between (fuzzy or crisp) quantities are essential, in particular for dealing with flexible queries using absolute or relative fuzzy quantifiers. A study, under submission, shows that these fuzzy order relations rely on difference or division operators of $Q_f$. As a first step, a gradual order relation, denoted by x $\geq_g$ y, strictly comparing two fuzzy

[Roc03] D. Rocacher, "On fuzzy bags and their application to flexible querying", *Fuzzy Sets and Systems 140*, 2003, p. 93–110.

quantities, has been proposed. Such a relation is based on a fuzzy measure $M$ of a fuzzy set S representing the global satisfaction of x $\geq_g$ y. S is an inventory of satisfactions by $\alpha$-cuts of x $\geq_g$ y and the measure $M$ summarizes the information described by S. This measure $M$ expresses the following semantics " the greater the number of high level $\alpha$-cuts satisfying $x_\alpha \geq y_\alpha$, the more globally satisfied the predicate x $\geq_g$ y". Such a behavior is dealt with using Choquet integrals. In a second step, a fuzzy tolerance notion has been introduced in a gradual order relation. The idea is to take into account gradual aspects in the comparison of x and y by allowing, to some extent, more or less satisfied cuts $x_\alpha \geq y_\alpha$. This fuzzy tolerant order relation extends the gradual one. It has been shown that gradual tolerant order relations can be viewed as generalizations of fuzzy R-implications. In this context, one point that could be developed in further investigation, is the use of aggregation operators in order to compose comparisons between fuzzy numbers. The idea is to combine global satisfactions of each comparison using the intersection and union of fuzzy sets and to apply a measure $M$ to the overall resulting global satisfaction.

As to quantified statements, two types of expression are generally distinguished: expressions denoted by "$Q$ X are $A$", meaning $Q$ elements belonging to X are $A$, and expressions written "$QB$ X are $A$" where a linguistic quantifier $Q$ is applied to a fuzzy set $B$ X. These two types of statement are illustrated by the following conditions: "*most of* employees are *well-paid*" and "*most of young* employees are *well-paid*". Many approaches have been proposed to evaluate such conditions. The most popular approaches are based on the OWA operator and the Sugeno fuzzy integral. Both deliver a degree of truth and are limited to statements involving increasing quantifiers (as "*most of*", "*at least* 3"). We are studying a new approach based on the arithmetic on gradual numbers ($N_f$, $Z_f$, $Q_f$). The advantages of our proposition is that no particular assumption is made on the monotonicity of the linguistic quantifier and that "$Q$ X are $A$" and "$Q$ $B$ X are $A$" expressions are dealing with the same method. In addition, our approach is a generalization of the OWA-based evaluation [24]. In the near future, we aim at comparing our gradual arithmetic-based approach to evaluate quantified statements with the Sugeno Fuzzy Integral-based interpretation.

### 6.3 Division of fuzzy relations and approximate inclusion

**debug :** Module named “resultats” for the project BADINS, section resultats

**Participants**: Patrick Bosc, Olivier Pivert, Daniel Rocacher.

**Keywords**: relational division, graded inclusion, quotient, softened universal quantifier.

**Abstract:** *The role and properties of the division are very well known in the context of queries addressed to regular relational databases. An extension of the division operator to fuzzy relations (i.e., relations made of weighted tuples) has been studied. The main objective of the work undertaken is to investigate which fuzzy implication can be used in order to get a result having the characteristics of a quotient. Moreover, the inclusion which is behind the division can be extended into*

*an approximate one where the universal quantifier is softened. As a consequence, an approximate division of both regular and fuzzy relations can be defined as a generalization of their usual division and the issue of assessing if its result is a quotient has been investigated.*

**debug : body of the module**.

In the context of the relational data model, the division operation is somewhat similar to the integer division. The division of r by s, whose respective schemas are (A, X) and (B) where A and B are attributes defined on the same domain, delivers a relation t whose schema is X. It turns out that t is a quotient, i.e., the largest relation such that:

$$t \times s \subseteq r$$

where $\times$ stands for the Cartesian product. The extension of this operation to fuzzy relations has been studied by various authors in the past decade, but the nature of the result obtained has not been characterized. In other words, it is not known whether the result of a given extended division is or not a quotient. We have investigated this question with the following point of departure. The regular division of r by s is defined as:

$$r \div s = \{x \mid s \subseteq K_r(x)\}$$

where $K_r(x) = \{a \mid < a, x > \in r\}$. When r and s are fuzzy relations, the previous definition becomes:

$$\mu_{r \div s}(x) = gr\text{-}inc(s, K_r(x))$$

where *gr-inc* denotes a gradual inclusion delivering a degree of inclusion and $K_r(x)$ is defined as:

$$K_r(x) = \{d/a \mid \mu_r(< a, x >) = d\}.$$

It turns out that there are two major types of graded inclusions. The first one is based on a logical vision of the inclusion calling on a fuzzy implication ($\Rightarrow_f$):

$$gr\text{-}inc(E, F) = inf_{x \in X} \mu_E(x) \Rightarrow_f \mu_F(x).$$

Fuzzy implications include notably R-implications and S-implications. Both of them generalize the regular implication (a $\Rightarrow$ b = (not a) or b) and they are generated using a given triangular norm ($\top$). The second approach to graded inclusions relies on the cardinalities of the two sets involved:

$$gr\text{-}inc(E, F) = card(E \cap F)/card(E)$$

where the intersection depends on a triangular norm. It has been shown [10, 9] that the only way for defining a sound division, i.e., delivering a quotient, is to use the first approach with any R-implication or S-implication associated with a continuous triangular norm. So doing, the result of the division is the largest fuzzy relation t such that:

$$t \times s \subseteq r$$

where $\mu_{t\times s}(a, x) = cnj(\mu_t(x), \mu_s(a))$ and $cnj$ is an operator generalizing the usual conjunction and strongly connected with the implication used for the division.

A complementary aspect of this work is concerned with the notion of an approximate division, which applies to both regular and fuzzy relations. The underlying idea is to soften the universal quantifier appearing in the logical view of a division. In other words, in the case of regular relations, an element may appear in the result of the division if it is connected (in the dividend) with *almost all* the elements of the divisor, instead of the strict requirement of being in relation with *all* the elements of the divisor. The approximate division of regular relations has been defined so that its result is a quotient. Similarly, the approximate division of fuzzy relations has been investigated. The key principle is once again to rely on a logical approach and to adapt what has been done for the crisp (non-approximate) division. In such a situation, it is necessary to combine the degrees coming from the fuzzy implication used in the approximate division on the one hand and the grades issued from the softened universal quantifier ("almost all") on the other hand. It has been shown [10, 9, 20] that with an appropriate definition, the result of the approximate division of fuzzy relations is a quotient, provided that the divisor is normalized (i.e., at least one of its elements has the degree 1) or the quantifier "almost all" is defined in a crisp way.

It turns out that, as the division and inclusion are pairwise connected, an approximate inclusion operator can be associated with the approximate division. The properties of fuzzy inclusion operators have been widely studied in the fuzzy community and we have examined those possessed by the approximate inclusion operator related to the approximate division [22].

### 6.4 Gradual properties of fuzzy sets in regular relational databases

**debug :** Module named “resultats5” for the project BADINS, section resultats

**Participants**: Patrick Bosc, Ludovic Liétard.

**Keywords**: relational databases, set-oriented functions, flexible queries, linguistic summaries.

**Abstract:** *Gradual set-oriented functions and their application to fuzzy sets can be found in flexible querying of regular relational databases, for example when the result of an aggregate is confronted to a fuzzy condition. We have proposed a general technique to achieve such a computation which can be interpreted in terms of properties of $\alpha$-cuts. Two applications of this technique in the context of regular*

*relational databases were also investigated. The first one concerns flexible querying and the definition of condition involving aggregates. The second one concerns the evaluation of linguistic summaries.*

**debug : body of the module.**

A gradual set-oriented function F is defined as a function applying to crisp subsets of a universe X (i.e. to sets belonging to $2^X$) which delivers a degree between 0 and 1. An example taken in the context of flexible querying of regular relational databases is provided by the satisfaction of the fuzzy condition "the cardinality is *high*" :

$$\begin{aligned} F: \quad 2^X &\rightarrow [0,1] \\ E &\rightarrow \mu_{high}(card(E)), \end{aligned}$$

where X is a given referential (a set of employees of a given firm for instance) and the fuzzy predicate *high* is defined by a fuzzy subset of the naturals. When E represents the employees from Paris, F(E) is the extent to which the number of employees from Paris is *high*. We propose a computation to estimate the value of function F when its arguments are fuzzy sets. If we refer to the given example, such a situation appears when we need to compute F on the fuzzy set of *young* employees (to determine the extent to which the number of *young* employees is *high*).

We consider the most general case where F has several arguments and applies to fuzzy sets: $A^1$, $A^2$, ..., $A^n$, n being the number of its arguments. According to our proposition, each fuzzy set $A^i$ is viewed as a collection of its $\alpha$-cuts which represent different interpretations (at different levels) of $A^i$. For a given level $\alpha$, the value of $F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha)$ can be computed, which is an interpretation (associated with level $\alpha$) of the result $F(A^1, A^2, ..., A^n)$.

The question now is about the integration of the results associated with these various interpretations. Intuitively, it seems reasonable to think that : *the more often* $F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha)$ *is high when* $\alpha$ *varies, the higher* $F(A^1, A^2, ..., A^n)$ and this interpretation is the meaning adopted for $F(A^1, A^2, ..., A^n)$.

A way is to look for the highest degree of satisfaction $\beta$ such that, for each level $\alpha$ $F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha)$ is at least equal to $\beta$. In other words:

$$F(A^1, A^2, ..., A^n) = max_{\beta \in [0,1]} min(\beta, each(\beta)),$$

where each($\beta$) means "for each level $\alpha$, $F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha) \geq \beta$". We propose to sum the lengths of intervals (of levels) where the threshold $\beta$ is reached :

$$\text{each}(\beta) = \sum_{]\alpha_i,\alpha_j] \text{ s.t. } \forall \alpha \in ]\alpha_i,\alpha_j],\ F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha) \text{ is defined and } F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha) \geq \beta} (\alpha_j - \alpha_i).$$

The higher each($\beta$), the more numerous the levels $\alpha$ for which $F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha) \geq \beta$. In particular, each($\beta$) equals 1 means that for each level $\alpha$, $F(A^1_\alpha, A^2_\alpha, ..., A^n_\alpha)$ is larger than (or equal to) $\beta$.

Two examples of use of our proposition have been provided [17]. The first one concerns the evaluation of conditions involving aggregate operators in flexible queries addressed to regular databases. When considering a condition of type "agg(A) is C", which means that the value of aggregate agg computed on fuzzy set A is C, for example "avg(*high* salaries) is *around 2000*", the given interpretation considers that the more $\alpha$-cuts of A highly satisfy the constraint "the aggregate is C", the more satisfied "agg(A) is C". Conditions of type "$agg_1(A)\ \theta\ agg_2(B)$" where $\theta$ is a fuzzy comparator (avg($A$) $\approx$ max($B$) where $\approx$ means *closed to*), can also be interpreted and, here again, the more $\alpha$-cuts of $A$ and $B$ highly satisfy $agg_1(A_\alpha)\ \theta\ agg_2(B_\alpha)$ the more satisfied $agg_1(A)\ \theta\ agg_2(B)$.

The second use of our approach is related to the evaluation of linguistic summaries of the type "$A^1$ is $C^1$ and $A^2$ is $C^2$ and ... and $A^n$ is $C^n$" where $A^i$s are attributes of a same relation and $C^i$s are different linguistic variables. The validity of such a summary over the entire database depends on two parameters : the quantity of tuples which satisfies the conjunct and their individual levels of satisfaction. As, in general, the higher the quantity, the smaller satisfaction, it has been shown [17] that our proposition determines a compromise between these two aspects. It delivers the highest p such that at least p% of the tuples satisfies "$A^1$ is $C^1$ and $A^2$ is $C^2$ and ... and $A^n$ is $C^n$" at a degree which is of at least p. As an example, if a validity of 0.8 is obtained for the linguistic summary "age is *young* and salary is *well-paid*" (on an employee relation), it means 80% of tuples in the relation satisfy *young* and *well-paid* at least at degree 0.8. As a consequence, its is possible to state that the linguistic summary is strongly satisfied (at degree 0.8) and it is also possible to provide a clear meaning for the obtained degree.

### 6.5 Query weakening

**debug :** Module named "resultats" for the project BADINS, section resultats

**Participants**: Patrick Bosc, Allel Hadjali, Olivier Pivert.

**Keywords**: Cooperative answering, flexible queries, query weakening, proximity relation, fuzzy relations.

**Abstract:** *The aim of this work is to propose an approach for dealing with the problem of query weakening in the framework of fuzzy querying of regular relational databases. The approach rests on a transformation mechanism that consists in applying a tolerance relation to fuzzy predicates contained in the query. A particular tolerance relation, which can be conveniently modeled in terms of a parameterized proximity relation, is discussed.*

**debug : body of the module**.

It is well known that the main problem approached in the field of cooperative answering systems is the "*empty answer problem*", that is, the problem of providing the user with some alternative data when there is no data fitting his query. *Relaxation* is one of the basic cooperative techniques that are used to deal with this problem. In the Boolean querying framework, relaxation consists in expanding the query by replacing some query conditions by more general conditions or by just eliminating some conditions. This allows the database to return answers related to the original query which are more convenient than an empty answer.

Weakening a *failing fuzzy query* consists in modifying the constraints involved in the query in order to obtain a less restrictive variant. Let, for instance, $Q$ be a query of the form $P_1$ *and* $P_2$ *and* ... *and* $P_k$ (where $P_i$ is a fuzzy predicate), and assume that the set of answers to $Q$ is empty. A simple way to weaken $Q$ is to apply a *basic transformation* to all the predicates $P_i$ or to some of them. This transformation process can be accomplished iteratively if necessary.

To be more efficient, a transformation $T$ should provide some *semantic limits*. Namely, what is the maximum number of weakening steps that is acceptable according to the user, i.e., such that the final modified query is *not too far*, semantically speaking, from the original one. Such limits can offer a rational tool for *controlling the relaxation process*.

Our method to transform a fuzzy predicate P into an enlarged one, with the objective to keep it semantically close to the initial one, relies on the notion of *proximity*. Let us note that the proximity used here originally stems from qualitative reasoning about orders of magnitude [HDP03,DHP01] and has been applied to define a fuzzy set of values that are close to some real-valued x. More generally, a *proximity* relation is intended for defining a set of predicates that are close to a given predicate P semantically speaking. This is why this notion appears appropriate in this topic.

In [BHP04], we have explored a proximity-based approach for relaxing fuzzy queries. The considered proximity is defined in a relative way and can be expressed by a *fuzzy relative closeness* relation (Cl) defined as:

$$\mu_{Cl[M]}(x, y) = \mu_M(x/y).$$

$M$ is called a tolerance parameter and should be chosen such that its support $S(M)$ lies in the validity interval V = $[(\sqrt{5} - 1)/2, (\sqrt{5} + 1)/2]$. The main feature of this approach to single-predicate query weakening is the following: it provides an efficient rational tool, which uses the only properties of the parameter M, for dealing with the stopping condition of the weakening process.

Since a *proximity relation* can be defined in a relative or in an absolute way, in [16] we have investigated the use of an *absolute proximity* for the purpose of query weakening. An *absolute*

[HDP03] A. Hadjali, D. Dubois, H. Prade, "Qualitative reasoning based on fuzzy relative orders of magnitude", *IEEE Transactions on Fuzzy Systems 11*, 2003, p. 9–23.

[DHP01] D. D. D., A. Hadjali, H. Prade, "Fuzzy qualitative reasoning with words", *in: Computing with Words (P.P. Wang, Ed.)*, p. 347–366, John Wiley and Son, 2001.

[BHP04] P. Bosc, A. Hadjali, O. Pivert, "Fuzzy closeness relation as a basis for weakening fuzzy relational queries", *in: Proceedings of the 6th Inter. Conference on Flexible Query Answering Systems (FQAS'04) - LNCS 3055 (Springer Verlag)*, p. 41–53, Lyon (France), 2004.

*proximity* is an approximate equality relation which can be modeled by a fuzzy relation of the form:

$$\mu_E(x, y) = \mu_Z(x - y),$$

which only depends on the value of the difference $x - y$, and where $Z$ is a fuzzy set centered in 0, such that: (i) $\mu_Z(u) = \mu_Z(-u)$; ii) $\mu_Z(0) = 1$; iii) its support $S(Z) = \{u | \mu_Z(u) > 0\}$ is bounded.

Assume that $P$ is a fuzzy predicate and $E[Z]$ an absolute proximity parameterized by a tolerance *indicator* $Z$. Relaxing $P$ into a fuzzy predicate $P'$ can be achieved in the following way:

$$\mu_{P'}(u) = sup_{v \in U} min(\mu_P(u'), \mu_{E[Z]}(u, v)),\ \forall u \in U.$$

We can easily check that $P' = P \oplus Z$. Here, the basic transformation $T$ is such that $T(P) = P \circ E[Z] = P \oplus Z$. In practise, if the set of answers to $Q$ ($Q = P$) is empty, then $Q$ is transformed into $Q_1 = P \oplus Z$. This progressive relaxation mechanism can be applied iteratively until the answer to the resulting query $Q_n = P \oplus n.Z$ is not empty. This strategy leads to a straight *symmetrical* weakening mechanism since the relaxation effect in the right and the left parts is the same. However, no information is provided about the *semantic limits*. To remedy this deficiency, one solution consists in asking the user to specify, along with his query, a fuzzy set Fp of non-authorized values in the related domain. Some additional properties of this technique are [15] :

- When applying this technique, the weakening is attribute domain-sensitive and predicate membership function-independent;,
- This technique is still valid for predicates expressed as traditional intervals.

Let us note that the problem of relaxing conjunctive queries has been investigated as well in [15]. Such queries are of the form $P_1$ **and** $P_2$ **and** ... **and** $P_k$, where $P_i$ is a fuzzy predicate. Two strategies can be envisaged for relaxing conjunctive queries:

- *Global Strategy.* It consists in applying the basic transformation to the entire query. Given a transformation $T$ and a conjunctive query $Q = P_1$ *and* $P_2$ *and* ... *and* $P_k$, the set of revised queries related to $Q$ resulting from applying $T$ is

  $$\{T^i(P_1)\ and\ T^i(P_2)\ and\ ...\ T^i(P_k)\},$$

  where $i \geq 0$ and $T^i$ means that the transformation $T$ is applied $i$ times.

- *Local Strategy.* In this case, the basic transformation applies only to subqueries. Given a transformation $T$ and a conjunctive query $Q = P_1$ *and* $P_2$ *and* ... *and* $P_k$, the set of modifications of $Q$ by $T$ is

$$\{T^{i_1}(P_1)\ and\ T^{i_2}(P_2)\ and\ ...\ T^{i_k}(P_k)\},$$

where $i_h \geq 0$ and $T^{i_h}$ means that the transformation $T$ is applied $i_h$ times.

In this later strategy, it is desirable that $T$ satisfies the property of *equal relaxation effect* on all terms. A possible way to define this property is to consider the ratio of the areas of the trapezes representing the membership functions associated to the original and the modified predicates. This ratio must be of the same magnitude when a certain transformation $T$ is applied.

Some directions of our future work concern:

- Define a measure of semantic closeness between the query Q and the modified query Q'.
- Investigate the dual problem of the empty answer problem, that of avalanche responses (which occurs when the database has a large amount of data that fully satisfy the user's query).

# 7 Contracts and Grants with Industry

**debug :** Module named "contrats" for the project BADINS, section contrats
**debug : body of the module**.

# 8 Other Grants and Activities

**debug :** Module named "international" for the project BADINS, section international

## 8.1 National actions

**debug :** Module named "national" for the project BADINS, section international
**debug : body of the module**.

Daniel Rocacher participates in a specific action called "Action Spécifique sur la Personnalisation de l'Information". The work of this group has given birth to a proposal for a project called APMD (Personalized Access to Massive Data) in response to the program ACI "masse de données". This project has been accepted and is in activity since october, 22.

## 8.2 International actions

**debug :** Module named "international2" for the project BADINS, section international
**debug : body of the module**.

# 9 Dissemination

**debug :** Module named "diffusion" for the project BADINS, section diffusion

### 9.1 Teaching

**debug :** Module named “ens” for the project BADINS, section diffusion
**debug : body of the module.**

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 DEA in computer science at University of Rennes 1, and at ENSSAT (third year level cursus).

### 9.2 Seminars and tutorials

**debug :** Module named “” for the project BADINS, section diffusion
**debug : body of the module.**

### 9.3 Panel discussion

**debug :** Module named “” for the project BADINS, section diffusion
**debug : body of the module.**

P. Bosc has organized a panel discussion at the BISCSE’05 conference (Berkeley, California, 2-5 November) about "Fuzzy Sets and Information Systems".

### 9.4 Scientific activities

**debug :** Module named “animation” for the project BADINS, section diffusion
**debug : body of the module.**

#### 9.4.1 Program committees

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

- EUSFLAT-LFA’05 (Barcelona, Spain, September, 7-9),
- FUZZ-IEEE’05(Reno, Nevada, May, 22-25),
- FLAIRS’05 (Clearwater Beach, Florida, May, 15-17),
- INFORSID’05 (President) (Grenoble, France, May, 24-25),
- KES’05 (Melbourne, Australia, September, 14-16),
- DEXA’05 Workshop on Integrating Data Mining, Databases and Information Retrieval (IDDI-05)(Copenhagen, Danemark, August, 22-26),
- BDA’05 (Saint Malo, France, October, 17-20),
- CORIA’05 (Grenoble, France, March, 9-11),
- EGC’05 (Amsterdam, The Netherlands, February, 14-16),

- BISCSE'05 (Berkeley, California, November, 2-5).

A. Hadjali has been a member of the program committee of the following conference:

- FUZZ-IEEE'05 (Reno, Nevada, May, 22-25).

D. Rocacher has been a member of the program committee of the following conferences:

- FLAIRS'05 (Reno, Nevada, May, 22-25),
- INFORSID'05 (Grenoble, France, May, 24-25).

### 9.4.2 Editorial boards

Patrick Bosc is a member of the following editorial boards:

- International Journal on Fuzziness, Uncertainty and Knowledge-Based Systems,
- Fuzzy Sets and Systems,
- Revue I3,
- Ingénierie des systèmes d'information (ISI).

## 10 Bibliography

### Major publications by the team in recent years

[1] P. Bosc, B. Buckles, F. Petry, O. Pivert, "Fuzzy databases (chapter 7)", *in: Fuzzy Sets in Approximate Reasoning and Information Systems*, J. Bezdek, D. Dubois, and H. Prade (editors), Kluwer Academic Publishers, 2000, p. 403 – 468.

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

[3] P. Bosc, O. Pivert, "SQLf: a relational database language for fuzzy querying", *IEEE Transaction on Fuzzy Systems 3*, 1, 1995.

[4] P. Bosc, H. Prade, "An introduction to the fuzzy set and possibility theory-based treatment of flexible queries and uncertain or imprecise databases (chapter 10)", *in: Uncertainty Management in Information Systems*, A. Motro and P. Smets (editors), Kluwer Academic Publishers, 1997, p. 285 – 324.

### Books and Monographs

[5] P. Bosc, L. Liétard, *Non monotonic aggregates applying to fuzzy sets in flexible querying, in Modern Information Processing: From Theory to Applications*, Elsevier, to appear.

[6] P. Bosc, O. Pivert, *Vacuity-oriented generalized yes/no queries addressed to possibilistic databases, in Flexible Databases Supporting Imprecision and Uncertainty*, Springer Verlag, to appear.

### Doctoral dissertations and "Habilitation" theses

[7] D. Rocacher, *Multi-ensembles flous et quantités graduelles - Application à l'interrogation flexible de bases de données*, PhD Thesis, Habilitation à diriger des recherches, IRISA - Université de Rennes I, December 2005.

### Articles in referred journals and book chapters

[8] P. Bosc, D. Kraft, F. Petry, "Fuzzy sets in database and information systems: status and opportunities", *Journal of Fuzzy Sets and Systems 156*, 3, 2005, p. 418–426.

[9] P. Bosc, O. Pivert, D. Rocacher, "About quotient and division of crisp and fuzzy relations", *Journal of Intelligent Information Systems*, to appear.

[10] P. Bosc, O. Pivert, D. Rocacher, "A propos de division usuelle et approchée de relations floues", *Technique et Science Informatiques*, to appear.

[11] P. Bosc, O. Pivert, "About selection - projection - join queries addressed to possibilistic relational databases", *IEEE Transactions on Fuzzy Systems 13*, 1, 2005, p. 124–139.

[12] P. Bosc, O. Pivert, "About yes/no queries against possibilistic databases", *International Journal of Intelligent Systems*, to appear.

[13] D. Rocacher, P. Bosc, "The set of fuzzy rational numbers and flexible querying", *Journal of Fuzzy Sets and Systems 155*, 3, 2005, p. 317–339.

[14] D. Rocacher, P. Bosc, "The set of fuzzy relative integers and fuzzy bags", *International Journal of Intelligent Systems*, to appear.

### Publications in Conferences and Workshops

[15] P. Bosc, A. Hadjali, O. Pivert, "Towards a tolerance-based technique for cooperative answering of fuzzy queries against regular databases", *in: Conference on Cooperative Information Systems (COOPIS'05) - LNCS 3760*, Larnaca (Cyprus), 2005.

[16] P. Bosc, A. Hadjali, O. Pivert, "Weakening of fuzzy relational queries: an absolute proximity relation-based approach", *in: Joint Conference of the European Society for Fuzzy Logic and Techniques and Rencontres Francophones sur la Logique Floue et ses Applications (EUSFLAT-LFA'05)*, p. 1275–1280, Barcelona (Spain), 2005.

[17] P. Bosc, L. Liétard, "A general technique to measure gradual properties of fuzzy sets", *in: 11th International Fuzzy Systems Association World Congress (IFSA'05)*, p. 485–490, Beijing (China), 2005.

[18] P. Bosc, N. Liétard, O. Pivert, "About cardinality-based possibilistic queries against possibilistic databases", *in: Joint Conference of the European Society for Fuzzy Logic and Techniques and Rencontres Francophones sur la Logique Floue et ses Applications (EUSFLAT-LFA'05)*, p. 1176–1181, Barcelona (Spain), 2005.

[19] P. Bosc, O. Pivert, D. Rocacher, "Characterizing the result of the division of fuzzy relations", *in: North American Fuzzy Information Processing Society Annual Conference (NAFIPS'05)*, p. 519–524, Ann Arbor (USA), 2005.

[20] P. Bosc, O. Pivert, D. Rocacher, "On the approximate division of fuzzy relations", *in : 15th International Symposium on the Management of Intelligent Systems (ISMIS'05) - LNAI 3488*, p. 314–322, Saratoga Springs (USA), 2005.

[21] P. Bosc, O. Pivert, "About the processing of possibilistic and probabilistic queries", *in : 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05) - LNAI 3571*, p. 812–823, Barcelona (Spain), 2005.

[22] P. Bosc, O. Pivert, "Approximate inclusion between fuzzy sets - application to the division of fuzzy relations", *in : IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'05)*, p. 561–566, Reno (USA), 2005.

[23] P. Ceravolo, D. Rocacher, "Detecting is-a and part-of relations in heterogeneous data flow", *in : 9th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES'05) - LCNS 3681*, p. 104–111, Melbourne (Australia), 2005.

[24] L. Liétard, D. Rocacher, "A generalization of the OWA operator to evaluate non monotonic quantifiers", *in : Joint Conference of the European Society for Fuzzy Logic and Techniques and Rencontres Francophones sur la Logique Floue et ses Applications (EUSFLAT-LFA'05)*, p. 183–188, Barcelona (Spain), 2005.