Weakening of Fuzzy Relational Queries: an Absolute Proximity Relation-Based Approach

Patrick BOSC **Allel HADJALI** **Olivier PIVERT**

IRISA/ENSSAT
6, rue de kerampont – BP 80518
22305 Lannion Cedex, France
{bosc, hadjali, pivert}@enssat.fr

Abstract

In a previous work, we have explored a proximity-based approach for relaxing queries involving fuzzy predicates. This proximity was defined in a relative way. In this paper we consider an absolute proximity and we show how it can be used for the purpose of query weakening. The main features of the weakening mechanism resulting from the use of this kind of proximity are investigated as well. Lastly, a comparative study of some methods proposed to deal with the issue of fuzzy query weakening is provided.

Keywords: Cooperative answering, fuzzy query weakening, proximity relation, fuzzy interval.

1 Introduction

Since the 90's, cooperative query answering has attracted the attention of many researchers, especially in the database area [9]. The most well-known problem approached in this field 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. Several approaches have been proposed to deal with this issue. Some of them are based on a relaxation paradigm that expands the scope of the query [3][8]. This allows the database to return answers related to the original query which are more convenient than an empty answer.

On the other hand, relying on fuzzy queries has the main advantage of diminishing the risk of empty answers. Indeed, fuzzy queries are based on preferences and retrieve elements that are more or less satisfactory rather than ideal. However, it still may happen that the database does not have any element that satisfies, even partially, the criterion formulated by the user. Then, an additional relaxation level must be performed on the fuzzy query to avoid such empty answers. This can be accomplished by replacing fuzzy predicates involved in the query with weakened ones. The resulting query is then less restrictive.

In the fuzzy framework, query weakening consists in modifying the constraints contained in the query in order to obtain a less restrictive variant. Such a modification can be achieved by applying a basic transformation to each predicate of the query. Note that research on fuzzy query weakening has not received much attention in the literature. Very few works are concerned with this problem. The study done by Andreasen and Pivert [1] is considered as pioneering in this area. This approach is based on a transformation that uses a particular *linguistic modifier*. More recently in [4], another solution has been proposed. It is based on a particular tolerance relation modeled by a *relative proximity* parameterized by a tolerance indicator. This notion of proximity, which originates from qualitative reasoning about fuzzy orders of magnitude [10], is intended for defining a set of predicates that are close, semantically speaking, to a given predicate *P*. This approach is significantly different from the previous one. It provides the semantic basis for defining a stopping criterion of the iterative weakening process. Let us also mention the work done in [11] where the authors consider queries addressed to data summaries and propose a method based on a specified distance to repair failing queries. Repairing query appears as relaxing the constraints of the retrieval since the resulting query, called substituting query, is more permissive than the initial one. See also the platform PRETI [2] which includes a flexible querying module that is endowed with an empirical method to avoid empty answers to user's request expressing a search for a house to let.

It is well known that there are two points of view which can be considered to compare numbers and thus orders of magnitude x and y on the real line [6]. We can evaluate to what extent the difference $x - y$ is *large*, *small* or *close* to *0*; this is the absolute comparative approach. Or, we may use relative orders of magnitude, i.e., evaluate to what extent the ratio x/y is *close* to *1* or not. A relative closeness based approach for fuzzy query weakening has already been proposed in [4]. Our objective in this paper is to investigate the behavior of the weakening mechanism when using an absolute proximity as a basis for the transformation of fuzzy predicates involved in the query.

The paper is structured as follows. The next section recalls the problem of fuzzy query weakening on one hand, and summarizes some methods that are proposed to solve this problem on the other hand. Section 3 shows how an absolute proximity can be used for generating more fuzzy permissive predicates and for achieving query relaxation. In section 4, we provide a comparative study of three basic techniques to relax fuzzy queries. Last, we attempt to outline some future works.

2 Fuzzy query weakening

Weakening a "failing" fuzzy query consists in modifying the constraints involved in the query in order to obtain a less restrictive variant. Let Q be a fuzzy 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 natural way to relax Q, in order to obtain a non empty set of answers, is to apply a basic uniform transformation to each predicate P_i. This transformation process can be accomplished iteratively if necessary. Two desirable properties are required for any transformation T when applied to a predicate P: ($\mathcal{C}_1$) it does not decrease the membership degree for any element of the domain, i.e. $\forall u \in$ domain(A), $\mu_{T(P)}(u) \geq \mu_P(u)$ where A denotes the attribute concerned by P; ($\mathcal{C}_2$) it extends the support of the fuzzy term, i.e. $S(P) = \{u / \mu_P(u) > 0\} \subset S(T(P)) = \{u / \mu_{T(P)}(u) > 0\}$.
Then, if P is a fuzzy predicate represented by the trapezoidal membership function (A, B, a, b), the desired transformation T is such that $P' = T(P) = (A, B, T(A, a), T(B, b))$. See Figure 1.

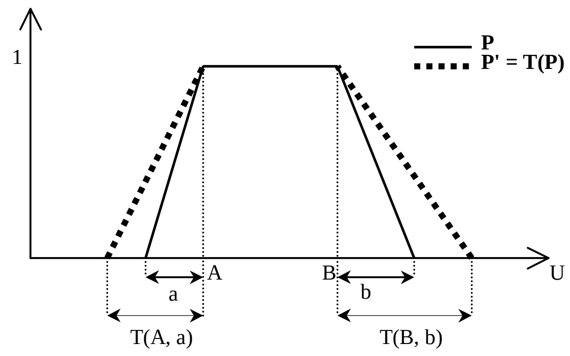

Figure 1: The basic transformation

As mentioned in the introduction, very few studies exist to deal with the issue of query weakening in the fuzzy setting. In each of them, a specific basic transformation is used. In the approach by Andreasen and Pivert [1], the transformation is based on a particular linguistic modifier. On the other hand, in the more recent one [4], the relaxation strategy makes use of a particular relative proximity expressed by a convenient fuzzy closeness relation.
The rest of this section is devoted to outline the two above approaches.

2.1 Linguistic modifier-based approach

To illustrate this approach, let us consider a query that involves only one fuzzy predicate. As pointed out in [1], one way to weaken such a query is to apply a linguistic modifier to the fuzzy term. For instance, the query "find the employees who are *young*" can be transformed into "find the employees who are *more-or-less young*". In [5] Bouchon-Meunier has proposed a family of linguistic modifiers which have interesting properties. Especially, one modifier of this family, called *ν-rather*, is of particular interest for the purpose of query weakening.

Let P be a fuzzy predicate represented *by* (A, B, a, b). The linguistic modifier *ν-rather* is defined such that $\nu\text{-}rather(P) = (A, B, a/\nu, b/\nu)$, with $\nu \in [1/2, 1[$. Then, the transformation T based on this modifier is such that $T(A, a) = a/\nu$ and $T(B, b) = b/\nu$, see Figure 1. Denoting by a' = a/ν and b' = b/ν, we can write $a' = a + \theta \cdot a$ and $b' = b + \theta \cdot b$, with $\theta = (1 - \nu)/\nu \in$

]0, 1]. Then, we can easily see that the resulting weakening effects in the left and right sides are obtained using the same parameter θ. This is why the approach is said to be *quasi-symmetric* (if $a = b$ holds, it is *symmetric*). As can be seen, this transformation preserves the specificity of the attribute as well (which means that μ_P and $\mu_{T(P)}$ are equal to 1 for the same elements of U) and satisfies the two desired properties ($\mathcal{C}_1$) and ($\mathcal{C}_2$).

Let us now explain how this modifier can be used in order to weaken a query involving one predicate *P*. If the set of answers is empty, the query *Q* is transformed into Q_1 = *rather (P)* and the process can be repeated n times until the answer to the question Q_n = *rather(rather(…rather (P)…))* is not empty. In practice, the difficulty when applying this technique concerns its *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 Q_n is *not too far,* semantically speaking, from the original one. Indeed, no intrinsic criterion is provided for stopping the iterative process. To overcome this problem, one solution consists in asking the user to specify, along with his query, a fuzzy set $\mathbb{F}_p$ of non-authorized values in the related domain. Then, the satisfaction degree of an element *u* becomes $min(\mu_{Q_i}(u),\ 1\ -\ \mu_{\mathbb{F}_P}(u))$ with respect to the modified query Q_i resulting from *i* weakening steps. The weakening process will now stop when the answer is non-empty or when the core of the complementary of the support of Q_i is included in the core of $\mathbb{F}_P$ (i.e., $min(\mu_{Q_i}(u),\ 1\ -\ \mu_{\mathbb{F}_P}(u)) = 0$).

2.2 Fuzzy relative closeness-based approach

In the framework of a study about relative orders of magnitude in qualitative reasoning, a fuzzy semantics has been proposed to closeness, negligibility and comparability relations [10]. It was shown that such relations can be represented by means of fuzzy relations controlled by tolerance parameters. The idea of relative closeness which expresses an approximate equality between two real numbers *x* and *y*, can be captured by the following definition.

Definition. The *closeness relation* (Cl) is a reflexive and symmetric fuzzy relation such that:

$\mu_{Cl}(x, y) = \mu_M(x/y)$,

where the characteristic function μ_M is that of a fuzzy number "*close to 1*", such that: i) $\mu_M(1) = 1$ (since *x* is close to *x*); ii) $\mu_M(t) = 0$ if $t \leq 0$ (assuming that two numbers which are close should have the same sign); iii) $\mu_M(t) = \mu_M(1/t)$ (since closeness is naturally symmetric).

M is called a *tolerance parameter*. Strict equality is recovered for $M = \mathbf{1}$ defined as $\mu_{\mathbf{1}}(x/y) = 1$ if $x = y$ and $\mu_{\mathbf{1}}(x/y) = 0$ otherwise. According to this point of view, we evaluate the extent to which the ratio *x/y* is close to *1*. The closer *x* and *y* are, the closer to *1 x/y* must be according to *M*.

Semantic properties of M. It has been demonstrated in [10] that the fuzzy number *M* which parameterizes closeness and negligibility should be chosen such that its support *S(M)* lies in the *validity interval* $\mathbb{V} = [(\sqrt{5} - 1)/2, (\sqrt{5} + 1)/2]$ in order to ensure that the closeness relation is more restrictive than the relation "*not negligible*". This means that if the support of a tolerance parameter associated with a closeness relation *Cl* is not included in $\mathbb{V}$, then the relation *Cl* is not in agreement with the intuitive semantics underlying this notion. It is worth noticing that the *validity interval* $\mathbb{V}$ plays a key role in the query weakening process. As it will be shown later, it constitutes the basis for defining a stopping criterion of an iterative weakening process.

Principle of the approach. As pointed out in [4], a way to perform query weakening is to apply a tolerance relation to the fuzzy terms involved in the query. A particular tolerance relation which is of interest in the context of query weakening can be conveniently modeled by the proposed parameterized closeness relation. Let us consider a query which only involves one fuzzy predicate *P*, and a closeness relation parameterized by a tolerance indicator *M*, *Cl[M]*. Now, to relax this query we replace the predicate *P* by an enlarged fuzzy predicate *P'* defined as follows:

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

Using the extension principle, it is easy to check that $P' = P \otimes M$, where $\otimes$ is the product operation extended to fuzzy numbers, see [7]. Clearly, the closeness-based transformation leads to a modified predicate *P'* which gathers the elements of *P* and the elements outside *P* which are somewhat close to an element in *P*.

Then the two desirable properties ($\mathcal{C}_1$) and ($\mathcal{C}_2$) are satisfied by the predicate *P'*. Namely, we have: i)

$\forall\, u,\ \mu_{P'}(u) \geq \mu_P(u)$; ii) $S(P) \subset S(P')$. In terms of trapezoidal membership functions, if $P = (A, B, a, b)$ and $M = (1, 1, \varepsilon, \varepsilon/(1 - \varepsilon))$, with $\varepsilon \in [0, (3 - \sqrt{5})/2]$, then, $P' = (A, B, a + A \cdot \varepsilon, b + B \cdot \varepsilon/(1 - \varepsilon))$ by applying the above arithmetic formula. Then, the transformation T is such that $T(P) = P \circ Cl[M] = P \otimes M$, where $\circ$ stands for the fuzzy composition operation. This leads to the following equalities $T(A, a) = a + A \cdot \varepsilon$ and $T(B, b) = b + B \cdot \varepsilon/(1 - \varepsilon)$ (see Figure 1), which expresses that the weakening effect is of a *non-symmetrical* nature.

In practice, if Q is a query containing one predicate P (i.e., $Q = P$) and if the set of answers to Q is empty, then Q is relaxed by transforming it into $Q_1 = P \otimes M$. This transformation is repeated n times until the answer to the question $Q_n = P \otimes M^n$ is not empty. Now, in order to ensure that the query Q_n is semantically close enough to the original one, the support of M^n should be included in the validity interval $\mathbb{V}$. Then, the above iterative procedure will stop either when the answer is non-empty or when $S(M^n) \not\subset \mathbb{V}$. As we can see, the strength of this approach lies in the fact that it provides *semantic limits* for the relaxation mechanism.

3 Absolute proximity-based approach to fuzzy query weakening

The purpose of this section is twofold. First, the notion of *absolute proximity* relation is introduced. Then, the method based on this proximity to address the problem of query weakening is discussed.

3.1 Absolute proximity relation

An *absolute proximity* is an approximate equality relation which can be modeled by a fuzzy relation of the form [6]:

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

which only depends on the value of the difference $x - y$, and where Z is a fuzzy set centered in 0, such that: i) $\mu_Z(r) = \mu_Z(-r)$; ii) $\mu_Z(0) = 1$; iii) its support $S(Z) = \{r \,/\, \mu_Z(r) > 0\}$ is bounded and is denoted by $[-\delta, \delta]$ where δ is a real number.

Property (i) ensures the symmetry of the approximate equality relation ($\mu_E(x, y) = \mu_E(y, x)$); (ii) expresses that x is approximately equal to itself with a degree 1. Here we evaluate to what extent the amount $x - y$ is close to 0. The closer x is to y, the closer $x - y$ and 0 are. Classical equality is recovered for $Z = \mathbf{0}$ defined as $\mu_{\mathbf{0}}(x - y) = 1$ if $x = y$ and $\mu_{\mathbf{0}}(x - y) = 0$ otherwise. See [6] for other interesting properties of the relation E.

3.2 Principle of the approach

As emphasized in section 2, a basic approach to fuzzy query weakening consists in the transformation of each fuzzy predicate P_i involved in the query into a weakened one, with the aim to keep it semantically close to the original one. The transformation explored in section 2.2 aims at finding a set of predicates that are close to a given predicate P. This is achieved by composing the predicate P with an appropriate relative proximity expressed by the fuzzy closeness relation $Cl[M]$. To do so, it is possible to use an absolute proximity.

Let us sketch how it can be done. 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:

$$\begin{aligned}\mu_{P'}(u) &= \sup_{v \in U} \min(\mu_P(v), \mu_{E[Z]}(u, v)),\ \ \forall\, u \in U \\ &= \sup_{v \in U} \min(\mu_P(v), \mu_Z(u - v)), \\ &= \mu_{P \oplus Z}(u),\ \text{ observing that } v + (u - v) = u.\end{aligned}$$

This means that $P' = P \oplus Z$, where $\oplus$ is the addition operation extended to fuzzy numbers [7]. As we can see P' contains P and the elements outside P which are in the neighborhood of an element of P. Then, this transformation is in agreement with the two requirements ($\mathcal{C}_1$) and ($\mathcal{C}_2$). In terms of trapezoidal membership functions, if $P = (A, B, a, b)$ and $Z = (0, 0, \delta, \delta)$, then $P' = (A, B, a + \delta, b + \delta)$ using the above arithmetic formula. This means that the transformation T is such that $T(P) = P \circ E[Z] = P \oplus Z$, which is illustrated in Figure 2 (where $T(A, a) = a + \delta$ and $T(B, b) = b + \delta$).

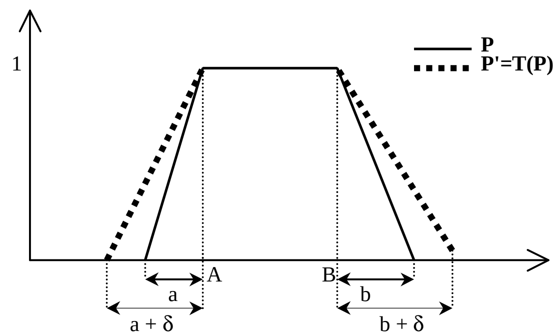

Figure 2: Absolute proximity-based weakening

Let us now show how this kind of proximity can be used to relax the query *Q* containing one predicate *P* ($Q = P$). If the set of answers to *Q* 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 $Q_n = P \oplus n{\cdot}Z$ is not empty. This strategy to single-predicate queries provides an implicit measure of *nearness* such that: Q_k is nearer to Q than Q_l if $k < l$. Moreover, this method leads to a straight *symmetrical* weakening mechanism since the relaxation effect in the right and the left parts is the same.

From a practical point of view, this mechanism is very simple to implement. However, no information is provided about the *semantic limits*. This means that the user does not know the maximum number of weakening steps that he (she) is authorized to make such that the final relaxed query Q_n remains semantically close to the initial one. To remedy this deficiency, once again we can use the fuzzy set $\mathbb{F}_p$ (introduced in section 2.1). Thus, the weakening process will stop if ($\Sigma_{Q_i} \neq \emptyset$ or $min(\mu_{Q_i}(u), 1 - \mu_{\mathbb{F}_p}(u)) = 0$), where Σ_{Q_i} stands for the set of answers to Q_i. This stopping condition is the same as in the modifier-based approach. This weakening technique can be sketched by the following algorithm (where $\overline{S(Q_i)}$ denotes the complementary of the support of Q_i):

```
let Q := P
let δ be a tolerance value        /* Z = (0, 0, δ, δ) */
i := 0
Qi := Q
compute ΣQi
while (ΣQi = ∅ and (C( S(Qi) ) ⊄ C(𝔽p)) 1) do
      begin
            i := i+1
            Qi := P ⊕ i·Z
            compute ΣQi
      end
if ΣQi ≠ ∅ then return ΣQi  endif.
```

4 A Comparative Study

Let us first investigate the main features of the absolute proximity-based approach. For the two other approaches, a complete study is available in [4].

[1] C(A) denotes the core of A, i.e., $\{u \in U \,/\, \mu_A(u) = 1\}$.

As it is illustrated in Figure 3, the slopes and the relative position of the membership function have no impact on the weakening effect when using the absolute proximity-based approach. However, the attribute domain is identified as a major factor affecting the weakening because δ is an absolute value which is added and subtracted (δ will be different for the attributes "age" and "salary").

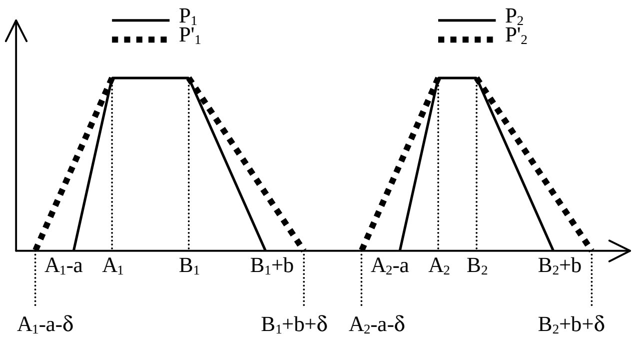

Figure 3: Impact of the slopes and the relative position of the membership functions ($a_1=a_2=a < b_1=b_2=b$)

Now from a practical point of view, it is interesting to compare the behaviors of the three weakening methods presented above in order to design a guide enabling the user to choose which method is the most suitable. To do this, we have listed five criteria that seem of major importance for the user:

(i) *Preservation/modification of the specificity of the attribute*.

(ii) *Symmetric/non-symmetric weakening*: it consists in checking whether the weakening effect in the right and left parts is the same or not.

(iii) *Semantic control of the relaxation*.

(iv) *Factors related to the domain and the predicate*: it consists in checking whether the attribute domain and the shape (or relative position) of the predicate membership function can have some impact on the weakening effect.

(v) *Applicability in the crisp case*: is the transformation still valid for predicates expressed as traditional intervals?

In table 1, we summarize the behavior of each query weakening technique with respect to the above five criteria. As it is shown in the table, the two advantages of the absolute proximity-based approach with respect to the ν-rather modifier-based one are the symmetrical nature of its weakening effect and its applicability when relaxing conventional queries. However, the most interesting

Table 1: Comparative study

Criteria	ν-rather modifier-based approach	Relative closeness-based approach	Absolute proximity-based approach
(i)	Attribute specificity preserved	Attribute specificity preserved	Attribute specificity preserved
(ii)	Symmetrical weakening under certain conditions	Non symmetrical weakening	Symmetrical weakening by nature
(iii)	No intrinsic semantic limits	semantic limits provided	No intrinsic semantic limits
(iv)	Attribute domain-independent and predicate membership function-dependent	Attribute domain-independent and predicate membership function-dependent	Attribute domain-dependent and predicate membership function-independent
(v)	Inappropriate in the crisp case	Still effective in the crisp case	Still effective in the crisp case

features of the relative closeness-based approach remains the rigorous semantic limits for controlling the query relaxation level that it provides.

5 Conclusion

An alternative fuzzy set-based approach for handling query failure is proposed. It contributes to enrich cooperative answering techniques in the context of usual database fuzzy querying. The proposed method is based on the notion of absolute proximity to define a predicate transformation. This transformation aims at finding a set of the closest predicates in the sense of the considered proximity.

In our future work, we plan to: i) define a measure of weakening intensity in order to be able to achieve uniform transformation in the case of fuzzy multi-predicates queries; ii) investigate the dual problem, i.e., that of plethoric answers (which occurs when the database has a large amount of data that fully satisfy the user's query).

References

[1] T. Andreasen and O. Pivert, "On the weakening of fuzzy relational queries", in Proc. 8th *Int. Symp. on Meth. for Intell. Syst.*, Charlotte, USA, 1994, pp. 144-151.

[2] M. de Calmès, D. Dubois, E. Hullermeier, H. Prade and F. Sedes, "Flexibility and fuzzy case-based evaluation in querying: An illustration in an experimental setting", *Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems*, 11(1), 2003, pp. 43-66.

[3] F. Benamara and P. Saint Dizier, "Advanced relaxation for cooperative question answering", in *New Directions in Question Answering*, Chapter 21, M. T. Maybury Ed., AAAI/MIT Press, 2004.

[4] P. Bosc, A. HadjAli and O. Pivert, "Fuzzy closeness relation as a basis for weakening fuzzy relational queries," in Proc. of 6th *Inter. Conf. FQAS*, Lyon, France, 2004, pp. 41-53.

[5] B. Bouchon-Meunier, "Stability of linguistic modifiers compatible with a fuzzy logic", in *Uncertainty in Intelligent Systems*, LNCS, 313, 1988, pp. 63-70.

[6] D. Dubois, A. HadjAli and H. Prade, "Fuzzy qualitative reasoning with words", in: *Computing with Words* (P.P. Wang, Ed.), Volume 3 in a Series of Books on Intelligent Systems, John wiley & Son, 2001, pp. 347-366.

[7] D. Dubois and H. Prade, "*Fundamentals of Fuzzy Sets*", The Handbooks of Fuzzy Sets Series (D. Dubois and H. Prade, Eds), Vol. 3, Kluwer Academic Publishers, Dordrecht, Netherlands, 2000.

[8] T. Gaasterland, "Cooperative answering through controlled query relaxation", *IEEE Expert*, 12(5), Sep/Oct 1997, pp. 48-59.

[9] T. Gaasterland, P. Godfrey and J. Minker, "An overview of cooperative answering", *Journal of Intelligent Information Systems*, 1(2), 1992, pp. 123-157.

[10] A. HadjAli, D. Dubois and H. Prade, "Qualitative reasoning based on fuzzy relative orders of magnitude", *IEEE Transactions on Fuzzy Systems*, Vol. 11, No 1, 2003, pp. 9-23.

[11] W.A. Voglozin, G Rashia, L. Ughetto and N. Mouaddib, "Interrogation de résumés de données et réparation de requêtes", in Proc. of 10th *Logique Floue & Application Conf.* (LFA), Nantes, France, 2004, pp. 239-246.