Activity Report 2017

Team SHAMAN

A Symbolic and Human-Centric View of Data Management

D7 – Data and Knowledge Management

1 Team

Head of the team
Olivier Pivert, Professor, Enssat

Administrative assistant
Joëlle Thépault, Enssat, 20%
Angélique Le Pennec, Enssat, 10%

University of Rennes faculty members
Laurent D'Orazio, Professor, IUT Lannion
François Goasdoué, Professor, Enssat
Hélène Jaudoin, Associate Professor, Enssat
Ludovic Liétard, Associate Professor, HDR, IUT Lannion
Pierre Nerzic, Associate Professor, IUT Lannion
Daniel Rocacher, Professor, Enssat
Grégory Smits, Associate Professor, IUT Lannion
Virginie Thion, Associate Professor, Enssat

Post-doctorate researcher
Ahmed Abid, DGA contract, since October 2017.

PhD students
Le Trung Dung, Vietnam government grant MOET 911, since September 2015; in Shaman since September 2016
Ngoc Toan Duong, CIFRE grant with Semsoft, since October 2016
Ludivine Duroyon, ANR grant, since October 2017
Sara El Hassad, Région Bretagne grant and Lannion Trégor Communauté grant, since October 2014
Cheikh Brahim El Vaigh, INRIA grant since October 2017
Aurélien Moreau, DGA contract, since November 2014
Thi To Quyen, grant of the French Ministry of Foreign Affairs, since October 2017
Olfa Slama, DGA contract, since November 2014
Van Hoang Tran, PEC grant since December 2017.

2 Overall Objectives

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

Motivations for introducing preferences inside database queries are manifold. First, it has appeared to be desirable to offer more expressive query languages that can be more faithful to

what a user intends to say. Second, the introduction of preferences in queries provides a basis for rank-ordering the retrieved items, which is especially valuable in case of large sets of items satisfying a query. Third, on the contrary, a classical query may also have an empty set of answers, while a more flexible version of the query might be matched by items in the database.

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

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

Basically, a fuzzy query involves linguistic terms corresponding to gradual predicates, i.e., predicates which are more or less satisfied by a given (attribute) value. In addition, these various terms may have different degrees of importance, which means that they may be connected by operators beyond conjunction and disjunction. For instance, in the context of a search for used vehicles, a user might say that he/she wants a *compact* car *preferably French*, with a *medium* mileage, *around* 6 k$, whose color is *as close as possible* to light grey or blue. The terms appearing in this example must be specified, which requires a certain theoretical framework. For instance, one may think that "*preferably* French" corresponds to a complete satisfaction for French cars, a lower one for Italian and Spanish ones, a still smaller satisfaction for German cars and a total rejection for others. Similarly, "*medium* mileage" can be used to state that cars with less than 40 000 km are totally acceptable while the satisfaction decreases as the mileage goes up to 75 000 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 be handled. A particular attention is paid to conditions calling on aggregate functions together with gradual predicates. For instance, one may look for departments where *most* employees are *close* to retirement, or where the average salary of *young* employees is *around* $2500. Such statements have their counterpart in regular query language, such as SQL, and the specification of their semantics, when gradual conditions come into play, is studied in the project.

Along this line, the ultimate goal of the project is to introduce gradual predicates inside database query languages, thus providing flexible querying capabilities. Algebraic languages as

well as more user-oriented languages are under consideration in both the original and extended relational settings.

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

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

A research axis in Shaman concerns the definition of models and languages for dealing with uncertain databases where uncertainty is represented in a qualitative manner, with the crucial objective of finding a good compromise between the expressivity of the model and the efficiency of query processing.

3 Scientific Foundations

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

3.1 Fuzzy sets

Fuzzy sets were introduced by L.A. Zadeh in 1965 [Zad65] in order to model sets or classes whose boundaries are not sharp. This is particularly the case for many adjectives of the natural language which can be hardly defined in terms of usual sets (e.g., high, young, small, etc.), but are a matter of degree. A fuzzy (sub)set F of a universe X is defined thanks to a membership function denoted by μ_F which maps every element x of X into a degree $\mu_F(x)$ in the unit interval $[0, 1]$. When the degree equals 0, x does not belong at all to F, if it is 1, x is a full member of F and the closer $\mu_F(x)$ to 1 (resp. 0), the more (resp. less) x belongs to F. Clearly, a regular set is a special case of a fuzzy set where the values taken by the membership function are restricted to the pair $\{0,\ 1\}$. Beyond the intrinsic values of the degrees, the membership function offers a convenient way for ordering the elements of X and it defines a symbolic-numeric interface. The α level-cut of a fuzzy set F is defined as the (regular) set of elements whose degree of membership is greater than or equal to α and this concept bridges fuzzy sets and ordinary sets.

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

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

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

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

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

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

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

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

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

and the Cartesian product:

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

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

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

which becomes :

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

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

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

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

3.2 Possibility theory

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

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

[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 (for Boolean events) is:

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

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

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

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

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

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

and, in general:

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

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

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

It indicates that if an event is not completely possible, it is excluded that it is somewhat certain, which makes it possible to define a total order over events: first, the events which are somewhat possible but not at all certain (from ($\Pi = N = 0$ 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 $\mathrm{A} \cup \bar{A} = X$ is not true in general when A is a fuzzy set) and the only valid property is: $\forall\, A$, $\Pi(A) \geq \mathrm{N}(A)$.

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

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

In order to comply with the second axiom above, a possibility distribution must be such that there exists (at least) an element x_0 of X for which $\pi(x_0) = 1$. Indeed, a possibility

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

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

distribution can be seen as a normalized fuzzy set F which represents the knowledge about a given variable. The following formula:

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

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

3.3 Fuzzy sets, possibility theory and databases

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

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

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

It turns out that fuzzy sets offer a very convenient way for modeling gradual concepts and then flexible queries. It has been proven [BP92] that many *ad hoc* approaches (e.g., based on distances) were special cases of what is expressible using fuzzy set theory. This framework makes it possible to express sophisticated queries where the semantic choices of the user can take place (e.g., the meaning of the terms or the compensatory interaction desired between the various fuzzy conditions of a query). Current works conducted in Shaman are oriented towards:

- the use of a predefined fuzzy vocabulary (which raises the question of its adequacy wrt to the actual content of the database),
- implementation and query optimization issues, in particular in a context of *Big Data*.

As to possibility distributions, they are used to represent uncertain data. By doing so, a straightforward connection can be established between a possibilistic database and regular ones. Indeed, a possibilistic database is nothing but a weighted set of regular databases (called worlds), obtained by choosing one candidate in every distribution appearing in any tuple of every possibilistic relation. According to this view, a query addressed to a possibilistic database has a natural semantics. However, it is not realistic to process it against all the worlds due to

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

their huge number. Then, the question tied to the querying of a possibilistic database bears mainly on the efficiency, which imposes to obviate the combinatory explosion of the worlds. The objective of the project is to identify different families of queries which comply with this requirement in the context of different possibility-theory-based extensions of the relational data model.

3.4 Ontology-based data management

Data management is a longstanding research topic in *Knowledge Representation* (KR), a prominent discipline of *Artificial Intelligence* (AI), and — of course — in *Databases* (DB).

Till the end of the 20^{th} century, there have been few interactions between these two research fields concerning data management, essentially because they were addressing it from different perspectives. KR was investigating data management according to human cognitive schemes for the sake of intelligibility, e.g. using *Conceptual Graphs* [CM08] or *Description Logics* [BCM+03], while DB was focusing on data management according to simple mathematical structures for the sake of efficiency, e.g. using the *relational model* [AHV95] or the *eXtensible Markup Language* [AMR+12].

In the beginning of the 21^{st} century, these ideological stances have changed with the new era of *ontology-based data management* [Len11]. Roughly speaking, ontology-based data management brings data management one step closer to end-users, especially to those that are not computer scientists or engineers. It basically revisits the traditional architecture of database management systems by decoupling the models with which data is exposed to end-users from the models with which data is stored. Notably, ontology-based data management advocates the use of conceptual models from KR as human intelligible front-ends called *ontologies* [Gru09], relegating DB models to back-end storage.

The *World Wide Web Consortium* (W3C) has greatly contributed to ontology-based data management by providing *standards* for handling data through ontologies, the two *Semantic Web* data models. The first standard, the *Resource Description Framework* (RDF) [W3Ca], was introduced in 1998. It is a graph data model coming with a very simple ontology language, *RDF Schema*, strongly related to description logics. The second standard, the *Web Ontology Language* (OWL) [W3Cb], was introduced in 2004. It is actually a family of well-established description logics with varying expressivity/complexity tradeoffs.

[CM08] M. Chein, M.-L. Mugnier, *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs*, Springer Publishing Company, Incorporated, 2008.

[BCM+03] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, P. F. Patel-Schneider (editors), *The Description Logic Handbook: Theory, Implementation, and Applications*, Cambridge University Press, 2003.

[AHV95] S. Abiteboul, R. Hull, V. Vianu, *Foundations of Databases*, Addison-Wesley, 1995.

[AMR+12] S. Abiteboul, I. Manolescu, P. Rigaux, M.-C. Rousset, P. Senellart, *Web Data Management*, Cambridge University Press, 2012.

[Len11] M. Lenzerini, "Ontology-based data management", 2011.

[Gru09] T. Gruber, "Ontology", *in: Encyclopedia of Database Systems*, Springer US, 2009, p. 1963–1965.

[W3Ca] W3C, "Resource Description Framework", *research report.*

[W3Cb] W3C, "Web Ontology Language", *research report.*

The advent of RDF and OWL has rapidly focused the attention of academia and industry on *practical* ontology-based data management. The research community has undertaken this challenge at the highest level, leading to pioneering and compelling contributions in top venues on Artificial Intelligence (e.g. AAAI, ECAI, IJCAI, and KR), on Databases e.g. ICDT/EDBT, ICDE, SIGMOD/PODS, and VLDB), and on the Web (e.g. ESWC, ISWC, and WWW). Also, open-source and commercial software providers are releasing an ever-growing number of tools allowing effective RDF and OWL data management (e.g. Jena, ORACLE 10/11g, OWLIM, Protégé, RDF-3X, and Sesame).

Last but not least, large societies have promptly adhered to RDF and OWL data management (e.g. library and information science, life science, and medicine), sustaining and begetting further efforts towards always more convenient, efficient, and scalable ontology-based data management techniques.

3.5 Big Data management

While large volumes of data have always been the interest of many research efforts, recent results in both the distributed systems and data base communities have lead to an increasing attention.

Large scale distributed file system such as Google File System [GGL03], parallel processing paradigm/environment like MapReduce [DG08] have been the foundation of a new ecosystem with data management contributions in major conferences and journals on databases, such as VLDB, VLDBJ, SIGMOD, TODS, ICDE, IEEE DEB, ICDE and EDBT. Different (often open-source) systems have been provided such as Pig [ORS+08], Hive [TSJ+10] or more recently Spark [ZCD+12] and Flink [CKE+15], making it easier to use data centers resources for managing big data.

[GGL03] S. Ghemawat, H. Gobioff, S.-T. Leung, "The Google file system", *in: Proceedings of the Symposium on Operating Systems Principles (SOSP)*, p. 29–43, Bolton Landing, NY, USA, 2003.

[DG08] J. Dean, S. Ghemawat, "MapReduce: simplified data processing on large clusters", *Communications of the ACM 51*, 1, 2008, p. 107–113.

[ORS+08] C. Olston, B. Reed, U. Srivastava, R. Kumar, A. Tomkins, "Pig latin: a not-so-foreign language for data processing", *in: Proceedings of the SIGMOD International Conference on Management of Data*, p. 1099–1110, Vancouver, BC, Canada, 2008.

[TSJ+10] A. Thusoo, J. S. Sarma, N. Jain, Z. Shao, P. Chakka, N. Zhang, S. Anthony, H. Liu, R. Murthy, "Hive - a petabyte scale data warehouse using Hadoop", *in: Proceedings of the International Conference on Data Engineering ({ICDE})*, p. 996–1005, Long Beach, California, {USA}, 2010.

[ZCD+12] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauly, M. J. Franklin, S. Shenker, I. Stoica, "Resilient Distributed Datasets: {A} Fault-Tolerant Abstraction for In-Memory Cluster Computing", *in: Proceedings of the {USENIX} Symposium on Networked Systems Design and Implementation (NSDI)*, p. 15–28, San Jose, CA, USA, 2012.

[CKE+15] P. Carbone, A. Katsifodimos, S. Ewen, V. Markl, S. Haridi, K. Tzoumas, "Apache Flink{\texttrademark}: Stream and Batch Processing in a Single Engine", *{IEEE} Data Engineering Bulletin 38*, 4, 2015, p. 28–38.

4 Application Domains

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

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

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

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

- Cyber Security. One subdomain of Cybersecurity aims to guarantee the safety of systems, continuously monitoring *unusual events* (that may be characterized in a fuzzy way). The development of technologies (mobility, embedded systems, Internet Of the Things) leads to a huge amount of heterogeneous information to be managed efficiently. Cloud Computing provide software and hardware resources for Big Data management in several contexts (telecom, social networks, healthcare, etc.). However, these solutions do not address specific properties of cyber security. In addition, they do not rely on opportunities of modern hardware. This work, in collaboration with Nokia and DGA, aims to address this problem by proposing optimization techniques based on new infrastructure for a Big Data platform dedicated to cybersecurity.

5 Software

Only the most recent prototypes developed by the team are described hereafter. Some more can be found here: `http://www-shaman.irisa.fr/shaman-software/`.

- PostgreSQLF is a flexible querying prototype that aims at evaluating fuzzy queries addressed to regular databases. It is an extension of PostgreSQL which implements the fuzzy query language SQLf defined in the team. This prototype is coupled with a graphical interface names ReqFlex [SPG13] that makes it easy for an end user to specify his/her fuzzy queries.
- IKEYS [DSP16] is an interactive and cooperative querying systems dedicated to corporate data, that allows users define unambiguous queries in an intuitive way. Users first express their information needs through coarse keyword queries (e.g. "track Jim Morrison 1971") that may then be refined with explicit projection and selection statements involving comparison operators and aggregation functions (e.g., "titles of tracks composed by Jim Morrison before 1971").
- FUDGE/SUGAR: FUDGE [PST15] is a query language allowing to query graph databases — fuzzy or not — in a flexible way. It makes it possible to express preferences queries where preference criteria may concern i) the content of the vertices of the graph and ii) the structure of the graph (which may include weighted vertices and edges when the graph is fuzzy). SUGAR is a prototype, based on Neo4j, implementing the FUDGE language. More information can be found here: `https://www-shaman.irisa.fr/fudge-prototype/`.
- TAMARI (Quality Alerts Management in Graph Databases using Rabbithole) is a prototype, based on the Neo4j graph databases management system, that makes it possible to introduce some functionalities for quality management of graph databases. Based on quality annotation (tags) attached to subgraphs of the data, a quality vocabulary, and user quality profiles, TAMARI implements an extension of the Neo4j Cypher language in order to introduce quality-awareness in queries. See `https://www-shaman.irisa.fr/tamari/`.
- MRFrequentSubGraphMining [AdMN15] is a prototype that makes it possible to mine frequent sub-graphs in a large graph data base using the main algorithms of the literature

[SPG13] G. Smits, O. Pivert, T. Girault, "ReqFlex: Fuzzy Queries for Everyone", *PVLDB 6*, 12, 2013, p. 1206–1209.

[DSP16] K. Dramé, G. Smits, O. Pivert, "IKEYS: Interactive KEYword Search Dedicated to Corporate Data", *in: Knowledge Engineering and Knowledge Management - EKAW 2016 Satellite Events, EKM and Drift-an-LOD, Bologna, Italy, November 19-23, 2016, Revised Selected Papers*, P. Ciancarini, F. Poggi, M. Horridge, J. Zhao, T. Groza, M. C. Suárez-Figueroa, M. d'Aquin, V. Presutti (editors), *Lecture Notes in Computer Science*, *10180*, Springer, p. 105–108, 2016.

[PST15] O. Pivert, G. Smits, V. Thion, "Expression and Efficient Processing of Fuzzy Queries in a Graph Database Context", *in: Proc. of the 24th IEEE International Conference on Fuzzy Systems (Fuzz-IEEE'15)*, Istanbul, Turkey, 2015.

[AdMN15] S. Aridhi, L. d'Orazio, M. Maddouri, E. M. Nguifo, "Density-based data partitioning strategy to approximate large-scale subgraph mining", *Inf. Syst. 48*, 2015, p. 213–223.

(gSpan, Gaston, FSG, etc.) on top of MapReduce. The prototype consists in a framework, enabling to test new algorithms or using different stores. In addition, it integrates our proposed partitioning strategy based on graphs' density.

- HYTORMO [NdTH16] is a HYTORMO, a new model to store and query medical (more precisely DICOM) data. HYTORMO uses a hybrid data storage strategy that is aimed not only to leverage the advantage of both row and column stores, but also to attempt to keep a trade-off among reducing disk I/O cost, reducing tuple construction cost and reducing storage space. In addition, Bloom filters are applied to reduce network I/O cost during query processing. The prototype is built on top of Spark.

6 New Results

6.1 Flexible database querying

6.1.1 Preference queries

Participants: Olivier Pivert, Hélène Jaudoin, Grégory Smits, Virginie Thion, Ludovic Liétard, Daniel Rocacher.

The works presented hereafter deal with fuzzy preference queries in the context of RDF databases on the one hand, and of uncertain relation databases on the other hand.

- *Fuzzy Quantified Queries to Fuzzy RDF Databases.* In a relational database context, fuzzy quantified queries have been long recognized for their ability to express different types of imprecise and flexible information needs. In [14], we introduce the notion of fuzzy quantified statements in a (fuzzy) RDF database context. We show how these statements can be defined and implemented in FURQL, which is a fuzzy extension of the SPARQL query language that we previously proposed. Then, we present some experimental results that show the feasibility of this approach.

- *Soft querying of sensorial data.* Data of a sensory profile represent human's evaluations and feelings about intensities of several criterias to describe and compare different products (as the criteria *odor of red fruits* for a red wine). [6] concerns the representation and querying of such data. It is shown that sensorial data are intrinsically imprecise due to this human evaluation (possibilistic data), and especially for untrained people. The data treatment can take advantages of a querying with user's preferences (flexible querying with fuzzy predicates). The classical approach to evaluate a fuzzy predicate on a possibility distribution is based on a possibility and a necessity measures of a fuzzy event and it is shown that this approach may be not convenient. A new expression for the evaluation of a fuzzy predicate on a possibility distribution is then introduced. More

[NdTH16] D. Nguyen-Cong, L. d'Orazio, N. Tran, M. Hacid, "Storing and Querying DICOM Data with HYTORMO", *in: Data Management and Analytics for Medicine and Healthcare - Second International Workshop, DMAH 2016, Held at VLDB 2016, New Delhi, India, September 9, 2016, Revised Selected Papers*, F. Wang, L. Yao, G. Luo (editors), *Lecture Notes in Computer Science, 10186*, p. 43–61, 2016.

complex flexible queries on possibilistic data are defined and methods to rank the answers are also proposed.

6.1.2 Cooperative answering, data summarization

Participants: Grégory Smits, Olivier Pivert, Pierre Nerzic, Aurélien Moreau.

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

- *Interactive Data Exploration on Top of Linguistic Summaries.* Extracting useful and interpretable knowledge from raw data is a crucial issue that has been largely addressed by the data mining community in particular. In [18], we propose an interactive data exploration approach that involves two steps. First, a personalized linguistic summary of the dataset concerned is built and displayed as a tag cloud. Then, exploration functionalities are provided on top of the summary to help the user discover interesting properties in the data, such as frequent/atypical/diversified associations between properties. In [17], different strategies are studied and compared for storing and using the linguistic summary efficiently.

- *Typicality-based recommendation.* In [13], we introduce a new recommendation approach leveraging demographic data. Items are associated with the audience who liked them, and we consider similarity based on audiences. More precisely, recommendations are computed on the basis of the (fuzzy) typical demographic properties (age, sex, occupation, etc) of the audience associated with every item. Experiments on the MovieLens dataset show that our approach can find predictions that other tested state-of-the-art systems cannot.

- *Fuzzy vocabulary elicitation.* Linguistic descriptions of numerical data using a vocabulary defined as linguistic variables are particularly useful to help a user understand the content of a dataset. When dealing with data structured with classes, the relevance of the linguistic descriptions strongly relies on the adequacy between the vocabulary and this data structure. In [16], we proposes a criterion to quantify this relevance, understood as informativeness and measured in terms of specificity. It then proposes various strategies to elicit appropriate fuzzy partitions to define the modalities of relevant linguistic variables and experimentally examines their performance on artificial data sets.

6.2 Big data management

Participants: Laurent D'Orazio, François Goasdoué, Virginie Thion.

- *Algebraic approach.* In [4], we present the algebraic formalism underlying the management of (possibly very large) relational databases.

- *User- and Super-user- based interaction for Multi-objective Query Processing.* In [19] we introduce a new user interaction model in multi-objective query processing. This model introduces the administrators, or super users, to the user interaction process, allowing them to preset Weight Profiles and their logical descriptions. Weight Profiles contain objective preferences for the users before the query is executed. By using this model, the users can select a Weight Profile that will obtain their optimal query execution plan, and the process of choosing will be more accurate and efficient.

- *Graph Constraints in Urban Computing: dealing with conditions in processing urban data.* In [7] we present our views towards developing techniques for querying and evolving graph-modeled datasets based on user-defined constraints. Our focus is to show how these techniques can be applied to effectively retrieve urban data and have automated mechanisms that guarantee data consistency.

- *Quality management in graph databases.* Much work has been done about data quality management in *relational* databases. However, even though relational databases are still widely used, the need to handle *complex* data has led to the emergence of other types of data models. In the last few years, *graph databases* have started to attract a lot of attention in the database world. Their basic purpose is to manage networks of entities, the underlying data model of many open data applications like e.g. social networks, biological or bibliographic databases. This context raises new challenges in terms of data quality management. In [15], we introduce the notion of quality aware (graph pattern) query, which allows to take quality information into account, at evaluation time, in a query addressed to a graph database. The quality notion is usage-dependent, defined according to user quality profiles. We also implement such a quality awareness and study the complexity of providing the quality awareness extension. An associated open-source prototype was developed (the TAMARI prototype, see Section 5).

6.3 Ontology-based data management

Participants: Sara El Hassad, François Goasdoué, Hélène Jaudoin.

- *Summarization of RDF graphs.* RDF is the data model of choice for Semantic Web applications. RDF graphs are often large and have heterogeneous, complex structure. Graph summaries are compact structures computed from the input graph; they are typically used to simplify users' experience and to speed up graph processing. In [5], we introduce a formal RDF summarization framework, based on graph quotients and RDF node equivalence; our framework can be instantiated with many such equivalence relations. We show that our summaries represent the structure and semantics of the input graph, and establish a sufficient condition on the RDF equivalence relation which ensures that a graph can be summarized more efficiently, without materializing its implicit triples.

- *Reasoning using ontologies.* Finding commonalities between descriptions of data or knowledge is a fundamental task in Machine Learning. The formal notion characterizing precisely such commonalities is known as least general generalization of descriptions and

was introduced by G. Plotkin in the early 70's, in First Order Logic. Identifying least general generalizations has a large scope of database applications ranging from query optimization (e.g., to share commonalities between queries in view selection or multi-query optimization) to recommendation in social networks (e.g., to establish connections between users based on their commonalities between profiles or searches). We revisited the notion of least general generalization in the entire Resource Description Framework (RDF) [8, 9] and popular conjunctive fragment of SPARQL [12, 10, 11], a.k.a. Basic Graph Pattern (BGP) queries. Our contributions include the definition and the computation of least general generalizations in these two settings, which amounts to finding the largest set of commonalities between incomplete databases and conjunctive queries, under deductive constraints.

7 Other Grants and Activities

7.1 National actions

François Goasdoué is involved in the following projects:

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

- ANR ContentCheck (2016–2019), whose other partners are INRIA Saclay, LIMSI (Orsay), LIRIS (Lyon) and the team in charge of the blog "Les Décodeurs" associated with the newspaper Le Monde (http://www.lemonde.fr/les-decodeurs/).

- INRIA Project Lab iCoda – Knowledge-mediated Content and Data Analytics (2017–2020), whose partners are INRIA Montpellier (GRAPHIK), INRIA Saclay (CEDAR & ILDA), INRIA/IRISA Rennes (LINKMEDIA & SHAMAN), as well as AFP, Ouest France and Le Monde. The goal of this project is the design of algorithms that allow analysts to efficiently infer useful information and knowledge by collaboratively inspecting heterogeneous information sources, from structured data to unstructured content, taking data journalism as an emblematic use-case.

François Goasdoué, Hélène Jaudoin, Olivier Pivert, Grégory Smits, and Virginie Thion are involved in the DGA project ODIN (Open Data INtelligence) which started in November 2014. The other partners involved are Semsoft and INRIA Saclay. The ODIN project aims to propose a data management and business intelligence solution for big data, i.e., large-scale heterogeneous and imperfect data distributed over several sources. For doing so, we intend to conceive a data processing and multidimensional analysis chain suitable for RDF data, taking into account the data quality aspect.

Virginie Thion coordinates the project GioQoso (défi CNRS mastodons 2016) about quality management of open musical scores (see `https://gioqoso.irisa.fr/` for more details).

Apart from IRISA/Shaman, the other participants are the teams CNAM/CEDRIC (Paris), CNRS/IREMUS (Paris) and CESR (Tours).

Laurent d'Orazio is involved of the following projects:

- Action incitative Université de Rennes 1 Data ANalysis for Cyber sEcurity (DANCE). This project aims to provide a system for storing and querying big data for cyber security thanks to preliminary collaborations with Nokia and DGA.
- PEPS CNRS MULTImodal Platform fOr e-socIal sciences aNalyTics: application to political sciences (as the coordinator) about big data management of tweets in political sciences. Partners are from ILCEA4, IRISA, LIFO and LIG.

7.2 International actions

- Grégory Smits gave a Master's course about Fuzzy Preferences Queries at the Hanoi University of Science and Technology (HUST) in January 2017.
- Laurent d'Orazio is involved in the NSF MOCCAD project (`www.cs.ou.edu/~database/MOCCAD/`) about multi-objective query processing in a mobile cloud environment.

8 Dissemination

8.1 Teaching

Project members give lectures in different faculties of engineering, in the third cycle University curriculum: "Bases de données avancées" and "Web data management" in the Master's degree in computer science (M2 SIF) at University of Rennes 1, and at Enssat (third year level cursus).

8.2 Scientific activities

8.2.1 Highlights of the year

- Olfa Slama defended her Ph.D. thesis on November 22, 2017 [3].
- Hélène Jaudoin was Program Committee Co-Chair of the international conference FQAS'17 [1] that took place in London, UK from June 21 to June 23, 2017.
- Olivier Pivert was Program Committee Co-Chair of the international conference SUM'17 [2] that took place in Granada, Spain from October 4 to October 6, 2017.

8.2.2 Program committees

Laurent D'Orazio served as a member of the following program committees:

- International Conference on Big Data Analytics and Knowledge Discovery (DAWAK@DEXA 2017);

- International workshop on Uncertainty in Cloud Computing (UCC@DEXA 2017);
- International Conference on Cyber-Security in Aviation, Computer Science, and Electrical Engineering (UND-SCIEI ICCS 2017);
- Journées Francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2017);
- Colloque sur l'Optimisation et les Systmes d'Information (COSI 2017);
- International Symposium on Information and Communication Technology (SoICT 2017).

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

- International Conference on Extending Database Technology (EDBT 2017);
- ACM Special Interest Group On Data Management (SIGMOD 2017);
- International World Wide Web Conference (WWW 2017).

H. Jaudoin served as the PC Co-Chair of the following program committee:

- 12$^{\text{th}}$ International Conference on Flexible Query Answering Systems (FQAS 2017), London, UK, June 21-23, 2017.

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

- 32$^{\text{nd}}$ ACM Symposium on Applied Computing (SAC 2017), Marrakech, Morocco, April 3-6, 2017;
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2017), Amiens, France, October 18-19, 2017.

O. Pivert served as the PC Co-Chair of the following program committee:

- 11$^{\text{th}}$ International Conference on Scalable Uncertainty Management (SUM 2017), Granada, Spain, October 4-6, 2017.

and as a member of the following program committees:

- 32$^{\text{nd}}$ ACM Symposium on Applied Computing (SAC 2017), Marrakech, Morocco, April 3-6, 2017;
- 23$^{\text{rd}}$ International Symposium on Methodologies for Intelligent Systems (ISMIS 2017), Warsaw, Poland, June 26-29, 2017;
- 12$^{\text{th}}$ International Conference on Flexible Query Answering Systems (FQAS 2017), London, UK, June 21-23, 2017.
- Joint 17$^{\text{th}}$ World Congress of International Fuzzy Systems Association and 9$^{\text{th}}$ International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS 2017), Otsu, Japan, June 27-30, 2017;

- 17^{e} Conférence sur l'Extraction et la Gestion des Connaissances (EGC 2017), Grenoble, France, January 23-27, 2017;
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2017), Amiens, France, October 18-19, 2017.

G. Smits served as a member of the following program committees:

- 23rd International Symposium on Methodologies for Intelligent Systems (ISMIS 2017), Warsaw, Poland, June 26-29, 2017;
- 12th International Conference on Flexible Query Answering Systems (FQAS 2017), London, UK, June 21-23, 2017.
- Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS 2017), Otsu, Japan, June 27-30, 2017;
- Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2017), Amiens, France, October 18-19, 2017.

V. Thion served as a member of the following program committees:

- the Asian Conference on Intelligent Information and Database Systems (ACIIDS), Kanazawa, Japan, April 3-5, 2017;
- the Congrès INFormatique des ORganisations et Systèmes d'Information et de Décision (INFORSID), Toulouse, France May 30- June 2, 2017;
- the Workshop Qualité des Données du Web (Quality of Linked Open Data) (QLOD), en conjonction avec EGC, Grenoble, France, January 23-27.

8.2.3 Editorial boards

Olivier Pivert is a member of the following editorial boards:

- Journal of Intelligent Information Systems,
- Fuzzy Sets and Systems,
- International Journal of Fuzziness, Uncertainty and Knowledge-Based Systems,
- Ingénierie des Systèmes d'Information.

8.2.4 Steering committees

Olivier Pivert is as a member of the steering committee of

- the French-speaking conference "Rencontres Francophones sur la Logique Floue et ses Applications" (LFA);
- the International Symposium on Methodologies for Intelligent Systems (ISMIS).

8.2.5 International advisory boards

Olivier Pivert is as a member of the international advisory board of the International Conference on Flexible Query-Answering Systems (FQAS).

9 Bibliography

Major publications by the team in recent years

[1] M. Bienvenu, C. Bourgaux, F. Goasdoué, "Explaining Inconsistency-Tolerant Query Answering over Description Logic Knowledge Bases", *in : Proc. of the 30th AAAI Conference on Artificial Intelligence (AAAI'16)*, Phoenix, Arizona, USA, 2016.

[2] M. Bienvenu, C. Bourgaux, F. Goasdoué, "Query-driven Repairing of Inconsistent DL-Lite Knowledge Bases", *in : Proc. of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16)*, New York, NY, USA, 2016.

[3] P. Bosc, O. Pivert, "On a fuzzy bipolar relational algebra", *Inf. Sci. 219*, 2013, p. 1–16.

[4] D. Bursztyn, F. Goasdoué, I. Manolescu, "Teaching an RDBMS about Ontological Constraints", *in : Proc. of the 42nd International Conference on Very Large Data Bases (PVLDB'16)*, New Delhi, India, 2016.

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

[6] O. Pivert, H. Prade, "A Certainty-Based Model for Uncertain Databases", *IEEE Trans. Fuzzy Systems 23*, 4, 2015, p. 1181–1196.

[7] O. Pivert, G. Smits, V. Thion, "Expression and Efficient Processing of Fuzzy Queries in a Graph Database Context", *in : Proc. of the 24th IEEE International Conference on Fuzzy Systems (Fuzz-IEEE'15)*, Istanbul, Turkey, 2015.

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

Books and Monographs

[1] H. Christiansen, H. Jaudoin, P. Chountas, T. Andreasen, H. Larsen (editors), *Flexible Query Answering Systems, 12th International Conference, FQAS 2017, London, UK, June 21-22, 2017, Proceedings*, *Lecture Notes in Computer Science*, *10333*, Springer, 2017.

[2] S. Moral, O. Pivert, D. Sanchez, N. Marin (editors), *Scalable Uncertainty management, 11th International Conference, SUM 2017, Granada, Spain, October 4-6, 2017. Proceedings*, *Lecture Notes on Artificial Intelligence Intelligence*, *10564*, Heidelberg, Germany, Springer, 2017.

Doctoral dissertations and "Habilitation" theses

[3] O. Slama, *Flexible Querying of RDF Databases: A Contribution Based on Fuzzy Logic*, PhD Thesis, University of Rennes 1 – École doctorale MathSTIC, November 22, 2017, supervised by O. Pivert and V. Thion.

Articles in referred journals and book chapters

[4] F. Goasdoué, V. Thion, "Approche algébrique de la manipulation de données", *in : Les Big Data à Découvert*, M. Bouzeghoub and R. Mosseri (editors), CNRS Éditions, 2017.

Publications in Conferences and Workshops

[5] S. Cebiric, F. Goasdoué, I. Manolescu, "A Framework for Efficient Representative Summarization of RDF Graphs", *in : Proc. of the 16th International Semantic Web Conference (ISWC'17), Poster & demo track*, Vienna, Austria, 2017.

[6] C. Coulon-Leroy, L. Liétard, "Soft Querying of Sensorial Data", *in : Proc. of the 26th IEEE International Conference on Fuzzy Systems (Fuzz-IEEE'17)*, Naples, Italy, 2017.

[7] L. D'Orazio, M. Halfeld-Ferrari, C. Satie Hara, N. Kozievitch, M. Musicante, "Graph Constraints in Urban Computing: Dealing with conditions in processing urban data", *in : Proceedings of the International workshop on Data Analytics solutions for Real-LIfe APplications (DARLI-AP)*, Exeter, United Kingdom, UK, 2017.

[8] S. El Hassad, F. Goasdoué, H. Jaudoin, "Learning Commonalities in RDF", *in : Proc. of the 14th Extended Semantic Web Conference (ESWC'17)*, Portoroz, Slovenia, 2017.

[9] S. El Hassad, F. Goasdoué, H. Jaudoin, "Learning Commonalities in RDF", *in : Proc. of the 27th International Conference on Inductive Logic Programming (ILP'17), "Recently published papers" track*, Orléans, France, 2017.

[10] S. El Hassad, F. Goasdoué, H. Jaudoin, "Learning Commonalities in SPARQL", *in : Proc. of the 16th International Semantic Web Conference (ISWC'17)*, Vienna, Austria, 2017.

[11] S. El Hassad, F. Goasdoué, H. Jaudoin, "Learning Commonalities in SPARQL", *in : Actes de la 33^e Conférence sur la Gestion de Données — Principes, Technologies et Applications (BDA'17)*, Nancy, France, 2017.

[12] S. El Hassad, F. Goasdoué, H. Jaudoin, "Towards Learning Commonalities in SPARQL", *in : Proc. of the 14th Extended Semantic Web Conference (ESWC'17), Poster session*, Portoroz, Slovenia, 2017.

[13] A. Moreau, O. Pivert, G. Smits, "A Typicality-Based Recommendation Approach Leveraging Demographic Data", *in : Proc. of the 12th International Conference on Flexible Query Answering Systems (FQAS'17)*, p. 71–83, London, UK, 2017.

[14] O. Pivert, O. Slama, V. Thion, "Fuzzy Quantified Queries to Fuzzy RDF Databases", *in : Proc. of the 26th IEEE International Conference on Fuzzy Systems (Fuzz-IEEE'17)*, Naples, Italy, 2017.

[15] P. Rigaux, V. Thion, "Quality Awareness over Graph Pattern Queries", *in : Proceedings of the International Database Engineering & Applications Symposium (IDEAS)*, Bristol, United Kingdom, 2017.

[16] G. Smits, M.-J. Lesot, O. Pivert, "Vocabulary Elicitation for Informative Descriptions of Classes", *in : Proc. of the Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS'17)*, Otsu, Japan, 2017.

[17] G. Smits, O. Pivert, P. Nerzic, "Calcul, stockage et utilisation efficaces de résumés linguistiques de données massives", *in : Actes des Rencontres Francophones sur la Logique Floue et ses Applications (LFA'17)*, Amiens, France, 2017.

[18] G. Smits, R. Yager, O. Pivert, "Interactive Data Exploration on Top of Linguistic Summaries", *in : Proc. of the 26th IEEE International Conference on Fuzzy Systems (Fuzz-IEEE'17)*, Naples, Italy, 2017.

[19] C. Wang, J. Arenson, F. Helff, L. Gruenwald, L. d'Orazio, "Weighted Sum Model for Multi-Objective Query Optimization for Mobile-Cloud Database Environments", *in : Proceedings of the International Workshop on Scalable Cloud Data Management (SCDM@IEEE Big Data)*, Boston, MA, USA, 2017.