



**HAL**  
open science

## Evaluating data reliability : an evidential answer with application to a web-enabled data warehouse

Sébastien Destercke, Patrice Buche, Brigitte Charnomordic

### ► To cite this version:

Sébastien Destercke, Patrice Buche, Brigitte Charnomordic. Evaluating data reliability : an evidential answer with application to a web-enabled data warehouse. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25 (1), pp.92-105. 10.1109/tkde.2011.179 . hal-01267940

**HAL Id: hal-01267940**

**<https://hal.science/hal-01267940v1>**

Submitted on 29 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Evaluating data reliability: an evidential answer with application to a Web-enabled data warehouse

Sebastien Destercke, Patrice Buche and Brigitte Charnomordic

**Abstract**—There are many available methods to integrate information source reliability in an uncertainty representation, but there are only a few works focusing on the problem of evaluating this reliability. However, data reliability and confidence are essential components of a data warehousing system, as they influence subsequent retrieval and analysis. In this paper, we propose a generic method to assess data reliability from a set of criteria using the theory of belief functions. Customizable criteria and insightful decisions are provided. The chosen illustrative example comes from real-world data issued from the *Sym'Previus* predictive microbiology oriented data warehouse.

**Index Terms**—Belief functions, evidence, information fusion, confidence, maximal coherent subsets, trust, data quality, relevance.

## I. INTRODUCTION

Estimating data reliability is a major issue for many scientists, as these data are used in further inferences. During collection, data reliability is mostly ensured by measurement device calibration, by adapted experimental design and by statistical repetition. However, full traceability is no longer ensured when data are reused at a later time by other scientists. If a validated physical model exists and data values fall within the range of the model validated domain, then data reliability can be assessed by comparing data to the model predictions. However, such models are not always available and data reliability must then be estimated by other means. This estimation is especially important in areas where data are scarce and difficult to obtain (e.g., for economical or technical reasons), as it is the case, for example, in Life Sciences.

The growth of the web and the emergence of dedicated data warehouses offer great opportunities to collect additional data, be it to build models or to make decisions. The reliability of these data depends on many different aspects and meta-information: data source, experimental protocol, ... Developing generic tools to evaluate this reliability represents a true challenge for the proper use of distributed data.

In classical statistical procedures, a preprocessing step is generally done to remove outliers. In procedures using web facilities and data warehouses, this step is often omitted,

implicit or simplistic. There are also very few works that propose a solution to evaluate data reliability. It is nevertheless close to other notions that have received more attention, such as trust [1] (see Section II for details).

We propose a method to evaluate data reliability from meta-information. Several criteria are used, each one providing a piece of information about data reliability. These pieces are then aggregated into a global assessment that is sent back, after proper post-treatment, to the end-user. In our opinion, such a method should

- deal with conflicting information, as different criteria may provide conflicting information about the reliability. For example, data may come from a reputed journal, but have been collected with rather unreliable instruments;
- be traceable, as it is important to be able to detect conflict and to provide insights about their origins, or in the absence of such conflicts, to know why such data has been declared poorly (or highly) reliable;
- be readable, both in its different input parameters and results, as the method and the system it is implemented in will be used mainly by non-computer scientists.

The method presented here answers these needs, by addressing two issues: first we propose a generic approach to evaluate global reliability from a set of criteria, second we consider the problem of ordering the reliability assessments so that they are presented in a useful manner to the end-users. Indeed, the goal of the present work is to propose a partly automatic decision-support system to help in a data selection process.

As evaluating data reliability is subject to some uncertainties, we propose to model information by the means of evidence theory, for its capacity to model uncertainty and for its richness in fusion operators. Each criterion value is related with a reliability assessment by the means of fuzzy sets latter transformed in basic belief assignments, for the use of fuzzy sets facilitates expert elicitation. Fusion is achieved by a compromise rule that both copes with conflicting information and provides insights about conflict origins. Finally, interval-valued evaluations based on lower and upper expectation notions are used to numerically summarise the results, for their capacity to reflect the imprecision (through interval width) in the final knowledge. As an application area, we focus on Life Science and on reliability evaluation of experimental data issued from arrays in electronic documents.

Section II explains what we understand by reliability and discusses related notions and works. Section III is dedicated

Sebastien Destercke is with CIRAD, UMR IATE, Campus SupAgro, Montpellier, France, email sebastien.destercke@cirad.fr

Patrice Buche is with INRA, UMR IATE, Campus SupAgro, Montpellier, France, email patrice.buche@supagro.inra.fr

Brigitte Charnomordic is with INRA, UMR MISTEA 2, place Pierre Viala 34060 Montpellier Cedex 1, FRANCE email: bch@supagro.inra.fr

to an analysis of the information available to infer data reliability (with a focus on experimental data). Section IV describes the method used to model this information and to merge the different criteria using evidence theory. Section V addresses the question of data ordering by groups of decreasing reliability and subsequently the presentation of informative results to end-users. Section VI is devoted to the practical implementation of the approach to the case of the @Web data warehouse [2], [3]. It also presents a use case in the field of predictive microbiology.

## II. RELATED NOTIONS AND WORKS

In this paper, a source (e.g., expert, sensor, ...) is considered as reliable if its information can be used safely (i.e., be trusted), while the information of an unreliable source has to be used with caution (note that information coming from an unreliable source may be true, but nothing guarantees it).

This section makes a short review of topics covered in this paper and of related notions, i.e. how to evaluate reliability, what are the notions related to reliability, and how reliability evaluations should be presented to the end-user.

### A. Reliability evaluation

In practice, an information source is seldom always right or wrong, and evaluating/modelling the reliability of a source can be complex, especially if source information cannot be compared to a reference value.

In evidence theory, methods to evaluate reliability consist in choosing reliability scores that minimise an error function [4]. In spirit, the approach is similar to compare source assessments with reference values (as done to evaluate experts in probabilistic [5] or possibilistic [6] methods). It requires the definition of an objective error function and a fair amount of data with known reference value. This is hardly applicable in our case, as data are sparse and can be collected and stored for later use, i.e. not having a specific purpose in mind during collection. Other approaches rely on the analysis of conflict between source information [7], assuming that a source is more reliable when it agrees with the others. This comes down to make the assumption that the majority opinion is more reliable. If one accepts this assumption, then the results of such methods could possibly complement our approach.

### B. Related notions

Reliability has strong connections with other notions such as relevance, truthfulness, trust, data quality ... All these related concepts are, however, either different from or less specific than the notion of reliability.

First, there is a difference between data *reliability*, i.e., the trust we can have in data values, and data *relevance*, i.e. the degree to which data can answer our need. Indeed, some data can be totally reliable and totally irrelevant for a given situation. For instance, a perfect measurement of temperature is reliable but would be irrelevant if one needs to have a pressure measurement. Also, while the notion of reliability

is rather objective (data are reliable or not, irrespectively of the situation), the notion of relevance only makes sense w.r.t. a particular question of interest. Note that in some occasion a same criterion can be used to evaluate the reliability and relevance, but the two notions can behave differently w.r.t. this criterion. For example, if the criterion is the scientific journal where the data have been published, data relevance will be mainly impacted by the scope of the journal, while data reliability will be mainly impacted by the journal overall reputation. Another example is the data imprecision, to which the notions react in opposite ways: when data are more imprecise, data reliability increases (i.e., they are more likely to contain the true value) and data relevance decreases (i.e., they are less likely to answer a specific question). So, both notions of data reliability and relevance need to be considered in real-world cases, but they should not be confused and they require separate treatment and evaluation.

Similarly, the notion of source sincerity or *truthfulness* [8] is different and should be handled individually. This notion only makes sense if the source can be suspected of lying in order to gain some advantage, and is distinct from reliability. For instance, one may infer the truth from a lying source if the lying mechanism is known, while nothing may be inferred from an unreliable source. In our particular application area, sources are seldom untruthful, contrary to what may happen in other domains (e.g., multi-agent systems).

Finally, the problem of evaluating data reliability can be paralleled with the one of trust evaluation that has recently received much attention, as it is an important aspect of multi-agent systems [1] and of the semantic web [9]. Trust is an integral part of the Semantic Web architecture, and many research works [10] focus on issues such as source authentication, reputation or cryptography. Indeed, data reliability can be seen as the amount of "trust" we have in the data validity. However, the notion of trust is often considered in a large sense and can include considerations related to reliability as well as to relevance or truthfulness.

In multi-agent systems, the notion of trust as well as the information used to evaluate it can take many forms [1]. One can differentiate between individual-level and system-level trusts, the former concerning the trust one has in a particular agent, while the latter concerns the overall system and how it ensures that no one will be able to take advantage of the system. The collected information about the trustworthiness of an agent may be direct, i.e. coming from past transactions one has done with this agent. It can also be indirect, i.e. provided by third-party agents. In that case, it can be based on an evaluation of the agent reputation or on information concerning some of its characteristics. This latter type of evaluation is the closest to the current work.

However the information itself may have an impact on trust assignment, either due to its nature or to its content. As the Semantic Web provides tools to formalize and exchange the content of resources, a new potential arises for further innovative techniques that examine the information content when determining trust. A recent paper [11] describes at

length several factors that affect how users determine trust in content provided by a Web resource. Naturally these include source authority and direct experience. Among the remaining factors, one can find items like topic and criticality, which are somehow related to data relevance. The limitation of resources may play a role, as well as the incentive to provide *good* information or on the contrary to be biased or deceptive (elements that are related to the notion of truthfulness). Source agreement and user expertise also have an impact. Some factors that we considered as particularly important in the present work are highlighted, such as citations (through related resources or recommendations), or age/freshness, this last point being very domain dependent.

Another paper [12] advocates a multi-faceted approach to trust models in internet environments. The authors point out the great number of terms and intertwined meanings of trust, and the difficulty to capture the wide range of subjective views of trust in single-faceted approaches. They propose an OWL based ontology of trust related concepts, such as credibility, honesty, reliability, reputation or competency, as well as a meta-model of relationships between concepts. Through domain specific models of trust, they can propose personalised models suited to different needs. The idea is to provide internal trust management systems, i.e. the trust assessment being made inside the system, while using the annotation power of a user community to collect trust data.

Among methods proposing solutions to evaluate trust or data quality in web applications, the method presented in [13] for recommendation systems is close to our proposal, but uses possibility theory as a basis for evaluations rather than belief functions. Another difference between this approach and ours is that global information is not obtained by a fusion of multiple uncertainty models, but by the propagation of uncertain criteria through an aggregation function (e.g., a weighted mean). Each method has its pros and cons: it is easier to integrate criteria interactions in aggregation functions, while it is easier to retrieve explanations of the final result in our approach.

The problem with using available trust evaluation systems on the web is that they are often dedicated to a particular application. They propose relatively simple representation and aggregation tools, notably due to the fact that Semantic Web applications are confronted to scalability issues. Our situation is somehow different, since we aim for a general method applicable to situations where the number of items will seldom exceed tens of thousands, and will in fact be often limited to some dozens.

### C. Output post-treatment

In [14], the impact of data quality on decision making is explored, and an experimental study about the consequences of providing various kinds of information (none, two-point ordinal and interval scale) regarding the quality of data is performed. They point out that the availability of the information is not enough, and that an important consideration is how data quality information is recorded and presented.

Decision tasks (apartment or restaurant selection) were experimented using two groups of subjects, one group performing the tasks without data quality information and the other one with a given quality format. Users were asked to explain their decision process, and issues like complacency, consensus and consistency were examined. Complacency is a measure of the degree to which data-quality information is ignored. The following interesting result was observed.

For simple tasks, the smallest level of complacency, corresponding to the greatest impact of data quality information, was observed when comparing groups with interval scaled data quality information with groups with no data quality information. For complex tasks, there seems to be an information *overload* effect, and no statistically significant conclusions appear. This is an important point in favor of giving a lot of attention in presenting readable results to end-users. Interval scaled quality is used in the present paper, together with group ordering by decreasing reliability.

### III. FROM WHAT INFORMATION SHOULD WE INFER RELIABILITY?

In this section, we present the type of information we have considered to evaluate the reliability of experimental data in Life Science. These criteria are elements that are usually found in publications (reports, papers, ...) reporting experimental results. Note that most of these criteria are not specific to Life Science, and can be used for any experimental data. The list of criteria is, of course, not exhaustive.

For other popular cases such as touristic data or other applications of the semantic web, some criteria used here are universal enough to be valid, but they must be completed by other proper criteria. The approach itself remains generic. Table I summarizes the various criteria that can be considered in our applicative context:

- a first group concerns the data source itself. It contains features such as the source type (e.g., scientific publication, governmental report, web page, ...), the source reputation (e.g., is the laboratory that has produced data known for its reliability), the number of times the source has been cited or the publication date (data freshness being important in Life Science, due to rapid evolution of measurement devices and experimental protocol);
- a second group is related to the means used to collect data. Information related to these criteria is typically included in a section called *material and method* in papers based on experiments in Life Science, which thoroughly describes the experimental protocol and material. Some methods may be known to be less accurate than others, but still be chosen for practical considerations;
- a third group is related to statistical procedures: presence of repetitions, uncertainty quantification, elaboration of an experimental design.

These criteria can be reduced or enriched, according to the available information about the data and the relevant features to evaluate reliability.



Source	Production	Statistics
<b>Type</b>	<b>Protocol</b>	<b>Repetitions</b>
<b>Reputation</b>	<b>Material</b>	<b>Uncertainty quantification</b>
<b>Citation count</b>		(variance, confidence interval)
<b>Publication date</b>		<b>Experimental design</b>

TABLE I  
RELIABILITY CRITERIA

#### IV. METHOD

This section describes the method we propose to evaluate and use reliability. We first describe how information is collected and modelled. Then we briefly recall the basics of evidence theory needed in this paper.

##### A. Notations and information collection

We assume that reliability takes its value on a finite ordered space  $\Theta = \theta_1, \dots, \theta_N$  having an odd number of elements and such that  $\theta_i < \theta_j$  iff  $i < j$ .  $\theta_1$  corresponds to total unreliability, while  $\theta_N$  corresponds to total reliability, neutral value standing in the middle (i.e.,  $\theta_{(N+1)/2}$ ). We denote by  $I_{a,b} = \{\theta_a, \dots, \theta_b\}$  a set such that  $a \leq b$  and  $\forall c$  s.t.  $a \leq c \leq b$ ,  $\theta_c \in I_{a,b}$ . Such sets include all values between their minimum value  $\theta_a$  and maximum value  $\theta_b$ , and using a slight stretch of language we call them *intervals*.

The evaluation is based on the values taken by  $S$  groups  $A_1, \dots, A_S$  of criteria. A group may be composed of multiple criteria from Table I, considered as relevant to evaluate the reliability (e.g., source type, number of citation  $\times$  publication date, measurement device). Groups are built so that the impact of each group  $A_i$  on the reliability evaluation can be judged (almost) independent of the impact of any other group  $A_j$ . Each group  $A_i$ ,  $i = 1, \dots, S$  can assume  $C_i$  distinct values on a finite space  $\mathcal{A}_i = \{a_{i1}, \dots, a_{iC_i}\}$ .

For each possible value of each criteria group  $A_1, \dots, A_S$ , a domain expert is asked to give its opinion about the corresponding data reliability. To facilitate expert elicitation, a linguistic scale with a reasonable number of terms is used, for instance the five modalities *very unreliable*, *slightly unreliable*, *neutral*, *slightly reliable* and *very reliable*. To respect the imprecision inherent to linguistic concepts, we model the expert opinions<sup>1</sup> as fuzzy sets [15]. A fuzzy set  $\mu$  defined on a space  $\mathcal{A}$  is a function  $\mu : \mathcal{A} \rightarrow [0, 1]$  with  $\mu(x)$  the membership degree of  $x$ . Recall that the support  $S(\mu)$  and kernel  $K(\mu)$  of a fuzzy sets are the sets  $S(\mu) = \{x \in \mathcal{A} | \mu(x) > 0\}$  and  $K(\mu) = \{x \in \mathcal{A} | \mu(x) = 1\}$ . Such fuzzy sets describe some ill-known value of reliability, and one can define two uncertainty measures from it, namely a possibility and a necessity measure [16], such that

$$\Pi(E) = \sup_{x \in E} \mu(x); \quad N(E) = 1 - \Pi(E^c) = \inf_{x \in E^c} (1 - \mu(x)).$$

Possibility and necessity functions respectively measure the plausibility and the certainty of an event  $E \subseteq \mathcal{A}$ .

Denote by  $\mathcal{F}(\Theta)$  the set of all fuzzy sets defined over space  $\Theta$ . For each group  $A_i$ , we define a mapping  $\Gamma_{A_i} : \mathcal{A}_i \rightarrow \mathcal{F}(\Theta)$  built from the expert opinions,  $\Gamma_{A_i}(a)$  being

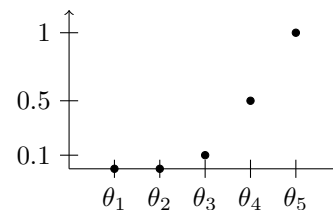


Fig. 1. Fuzzy set corresponding to the term *very reliable* defined on  $\Theta$  with  $N = 5$ .

the interpretation on  $\Theta$  of the expert information about the reliability when  $A_i = a$ . We denote by  $\mu_a$  the fuzzy set  $\Gamma_{A_i}(a)$ . The expert can select from a limited number of linguistic terms or combinations of them, using "or" disjunctions<sup>2</sup>. An additional term allows to express total ignorance. By doing so, expressiveness is respected while limiting the number of linguistic concepts. Fig. 1 illustrates a fuzzy set corresponding to the term *very reliable*.

The number of linguistic terms and of elements in the space  $\Theta$  should be low enough to ensure readability and ease-of-use, and high enough to ensure a reasonable flexibility in the modelling. While 3 seems a bit low (merely allowing to distinguish between bad/neutral/good), 5 or 7 will be adequate for most situations [17]. We propose the following guidelines to build the membership functions (MF) of fuzzy sets to model expert opinions:

- The Kernels are disjoint and ordered accordingly to linguistic terms (e.g.,  $K(\mu) = \theta_4$  for  $\mu = \textit{slightly reliable}$ , and  $K(\mu) = \theta_5$  for  $\mu = \textit{very reliable}$ ).
- The support of positive (resp. negative) linguistic terms should be  $S(\mu) = \{\theta_{(N+1)/2}, \dots, \theta_N\}$  (resp.  $S(\mu) = \{\theta_1, \dots, \theta_{(N+1)/2}\}$ ).
- The linguistic term *neutral* should have a kernel  $K(\mu) = \theta_{(N+1)/2}$  and be symmetric around this kernel.

MF parameters can be set according to these guidelines and with sufficient differences (at least 0.1) between two consecutive membership values. Note that the merging method used in Section IV-B is not sensitive to small changes in the MF values (see [18, Sect. 5, Prop 11.]), therefore the above guidelines appear both sufficient and not too restrictive.

**Example 1.** Consider the two groups  $A_1 = \textit{source type}$  and  $A_2 = \textit{experience repetition}$  such that

$$A_1 = \{a_{11} = \textit{journal paper}, a_{12} = \textit{governmental report}, a_{13} = \textit{project report}, a_{14} = \textit{other}\}$$

$$A_2 = \{a_{21} = \textit{repetitions}, a_{22} = \textit{no repetition}\}$$

The expert then provides his opinion about the reliability value for the different values of these two criteria. These opinions are summarised below

$$\begin{aligned} \mu_{a_{11}} &= \textit{very reliable}, \mu_{a_{12}} = \textit{slightly reliable}, \\ \mu_{a_{13}} &= \textit{neutral}, \mu_{a_{14}} = \textit{slightly unreliable}; \\ \mu_{a_{21}} &= \textit{very reliable}, \mu_{a_{22}} = \textit{slightly unreliable}. \end{aligned}$$

<sup>1</sup>The opinion can actually be any belief function.  
<sup>2</sup>In this case, fuzzy sets are combined by the classical t-conorm max answer with application to a web-enabled data warehouse. IEEE Transactions on Knowledge and Data Engineering, 25 (1), 92-105. DOI : 10.1109/tkde.2011.179

### B. Global reliability information through merging

For a data item, each group  $A_i$  takes a particular value, providing  $S$  different fuzzy sets as pieces of information. We propose to use evidence theory [19] to merge these information into a global representation. This choice is motivated by the richness of the merging rules [20] it provides and by the good compromise it represents in terms of expressiveness and tractability. Indeed it encompasses fuzzy sets and probability distributions as particular representations.

1) *Evidence theory*: We recall the basics of the theory and its links with fuzzy sets. A basic belief assignment (*bba*)  $m$  on a space  $\Theta$  is a mapping  $m : 2^{|\Theta|} \rightarrow [0, 1]$  from the power set of  $\Theta$  to the unit interval, such that  $\sum_{E \subseteq \Theta} m(E) = 1$ ,  $m(E) \geq 0$  and  $m(\emptyset) = 0$ . Sets  $E$  receiving a strictly positive mass are called **focal elements**. We denote by  $\mathcal{F}_m$  the set of focal elements of  $m$ . The mass  $m(E)$  can be seen as the probability that the most precise description of what is known about a particular situation is of the form " $x \in E$ ". From this mass assignment, Shafer [19] defines two set functions, called *belief and plausibility functions*, for any event  $A \subseteq \Theta$ :

$$Bel(A) = \sum_{E, E \subseteq A} m(E); \quad Pl(A) = 1 - Bel(A^c) = \sum_{E, E \cap A \neq \emptyset} m(E),$$

where the belief function measures the certainty of  $A$  (i.e., sums all masses that cannot be distributed outside  $A$ ) and the plausibility function measures the plausibility of  $A$  (i.e., sums all masses that it is possible to distribute inside  $A$ ). In this view, sets  $E$  are composed of mutually exclusive elements and represent incomplete information inducing uncertainty<sup>3</sup>. As the two functions  $Bel, Pl$  are conjugate, specifying one of them for all events is enough to characterize the other one.

Several special kinds of *bbas* are commonly recognized:

- a *bba* is *vacuous* if  $m(\Theta) = 1$ . It models total ignorance, making all conclusions equally possible (i.e.,  $Pl(A) = 1$  and  $Bel(A) = 0$  for any  $A \neq \{\emptyset, \Theta\}$ ).
- A *bba* is *empty* if  $m(\emptyset) = 1$ , and models a completely conflicting information.
- A *bba* is *categorical* if there is a set  $E \subset \Theta$  such that  $m(E) = 1$
- A *bba* reduces to a probability distribution if all focal elements are singletons.

It can be shown [19] that any necessity (resp. possibility) function is a special kind of belief (resp. plausibility) function, whose associated *bba* has nested focal elements. In this case, the *bba* is called *consonant*. Consequently *bbas* include fuzzy sets as a special case. A fuzzy set  $\mu$  with  $M$  distinct membership degrees  $1 = \alpha_1 > \dots > \alpha_M > \alpha_{M+1} = 0$  defines a *bba*  $m$  having, for  $i = 1, \dots, M$ , the following focal elements  $E_i$  with masses  $m(E_i)$  [21]:

$$\begin{cases} E_i = \{\theta \in \Theta | \mu(\theta) \geq \alpha_i\} = A_{\alpha_i}, \\ m(E_i) = \alpha_i - \alpha_{i+1}. \end{cases} \quad (1)$$

Therefore, each fuzzy set provided by experts (see Section IV-A) can easily be mapped into an equivalent *bba*.

**Example 2.** Consider the fuzzy set depicted in Fig. 1. Its equivalent *bba*  $m$  is such that

$$\begin{aligned} m(E_1 = \{\theta_5\}) &= 0.5, \quad m(E_2 = \{\theta_4, \theta_5\}) = 0.4, \\ m(E_3 = \{\theta_3, \theta_4, \theta_5\}) &= 0.1. \end{aligned}$$

2) *Merging multiple pieces of information*: When  $S$  different sources provide pieces of information modelled as *bbas*  $m_1, \dots, m_S$  over a same space  $\Theta$ , it is necessary to merge them to obtain a global model. Two main issues related to this merging are (1) how to merge non-independent sources [22] and (2) how to deal with sources providing conflicting information [20].

In this paper, sources (criteria groups) were selected to be independent, thus we do not have to deal with the first issue. However, it would be unrealistic to assume that all sources agree. Therefore we have to handle the case of conflicting information. Given the fact that sources are independent, the general merging of *bbas*  $m_1, \dots, m_S$  can be written,

$$\forall E \subseteq \Theta \quad m(E) = \sum_{\substack{E_i \in \mathcal{F}_i \\ \bigoplus_{i=1}^S (E_i) = E}} \prod_{i=1}^S m_i(E_i), \quad (2)$$

with  $\mathcal{F}_i$  the focal elements of  $m_i$ , and  $\bigoplus_{i=1}^S (E_i) = E$  an aggregation operator on sets. The well-known unnormalized Dempster's rule of combination [23] and the disjunctive combination correspond  $\oplus = \cap$  and  $\oplus = \cup$ , respectively. However, the former is not well adapted to conflicting information (the result is then close to an empty *bba*), while the latter often results in a very imprecise model (yielding a *bba* close to the vacuous one).

To deal with the problem of conflicting information, we propose a merging strategy based on maximal coherent subsets (MCS). This notion has been introduced by Rescher and Manor [24] as a means to infer from inconsistent logic bases, and can be easily extended to the case of quantitative uncertainty representations [18]. Given a set of conflicting sources, MCS consists in applying a conjunctive operator within each non-conflicting (maximal) subset of sources, and then using a disjunctive operator between the partial results [25]. With such a method, as much precision as possible is gained while not neglecting any source, an attractive feature in information fusion. In general, detecting maximal coherent subsets has a NP-hard complexity, however in some particular cases this complexity may significantly be reduced.

Consider  $N = \{I_{a_1, b_1}, \dots, I_{a_k, b_k}\}$  a set of  $k$  intervals. Using the MCS method on such intervals consists in taking the intersection over subsets  $\overline{K}_j \subset N$  s.t.  $\bigcap_{i \in \overline{K}_j} I_{a_i, b_i} \neq \emptyset$  that are maximal with this property, and then in considering the union of these intersections as the final result (i.e.  $\bigcup_j \bigcap_{i \in \overline{K}_j} I_{a_i, b_i}$ ). We denote by  $\oplus_{MCS}$  the MCS aggregation operator on sets. In the case of ordered spaces (such as  $\Theta$ ), Algorithm 1 detailed in [26] gives an easy means to detect maximal coherent subsets. Its result is shown in Fig. 2, where four (real valued) intervals  $I_1, I_2, I_3, I_4$  are merged by MCS aggregation. The two MCS are  $(I_1, I_2)$  and  $(I_2, I_3, I_4)$  and the final result is  $(I_1 \cap I_2) \cup (I_2 \cap I_3 \cap I_4)$ . If all

<sup>3</sup>This is in contrast with other uses of random sets. Destercke, S., Buche, P., and Denoeux, B. (2013). Evaluating data reliability: an evidential answer with application to a web-enabled data warehouse. IEEE Transactions on Knowledge and Data Engineering, 25 (1), 92-105. DOI : 10.1109/tkde.2011.179

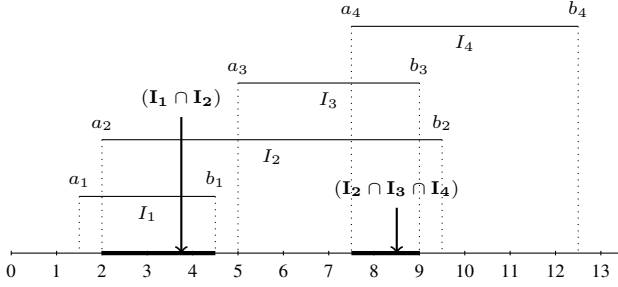


Fig. 2. Illustration of maximal coherent subsets merging.

intervals are consistent, conjunctive merging is retrieved, while disjunction is retrieved when every pair of intervals conflicts. An interesting feature is that the groups of intervals forming maximal coherent subsets may be used as elements explaining the result, as we shall see in the sequel.

---

**Algorithm 1:** Maximal coherent subset detection

---

**Input:**  $k$  intervals  $I_{a,b}$   
**Output:** List of maximal coherent subsets  $\overline{K}_j$

- 1 List =  $\emptyset$  ;  $j=1$  ;  $\overline{K} = \emptyset$  ;
- 2 Order by increasing value  
 $\{a_i, i = 1, \dots, k\} \cup \{b_i, i = 1, \dots, k\}$  (in case of ties,  
 $a_i$  is before  $b_i$ ) ;
- 3 Rename them  $\{c_i, i = 1, \dots, 2k\}$  with  $type(i) = a$  if  
 $c_i = a_m$  and  $type(i) = b$  if  $c_i = b_m$  ;
- 4 **for**  $i = 1, \dots, 2k - 1$  **do**
- 5     **if**  $type(i) = a$  **then**
- 6         Add source  $m$  to  $\overline{K}$  s.t.  $c_i = a_m$  ;
- 7         **if**  $type(i + 1) = b$  **then**
- 8              $\overline{K}_j = \overline{K}$  ;
- 9             Add  $\overline{K}_j$  to List ;
- 10              $j = j + 1$  ;
- 11     **else**
- 12         Remove source  $m$  from  $\overline{K}$  s.t.  $c_i = b_m$  ;

---

3) *Application of MCS to reliability evaluation:* As different groups of criteria can provide conflicting information about the data reliability, MCS appears as a natural solution to solve these conflicts while preserving a maximal amount of information. Consider a group  $A_i$ . According to Eq. 1 the fuzzy set  $\mu_{a_{ij}}$  given for each value  $a_{ij} \in A_i$  can be mapped into an equivalent *bba*  $m_{a_{ij}} : 2^{|\Theta|} \rightarrow [0, 1]$  representing the expert opinion about reliability when  $A_i = \{a_{ij}\}$ .

**Example 3.** Consider the two groups of Example 1. Now, assume that the retrieved data come from a journal paper ( $A_1 = a_{11}$ ) but that the experiment has not been repeated ( $A_2 = a_{22}$ ). Value  $a_{11}$  corresponds to "very reliable", while  $a_{22}$  corresponds to "slightly unreliable". Each group  $A_i$  thus provides an individual *bba* corresponding to the criterion

value. These *bbas* are given in the following tables:

$E_{11} = \{\theta_3, \theta_4, \theta_5\}$	$E_{12} = \{\theta_4, \theta_5\}$	$E_{13} = \{\theta_5\}$	
$m_{a_{11}}$	0.1	0.4	0.5
$E_{21} = \{\theta_1, \theta_2, \theta_3\}$	$E_{22} = \{\theta_2, \theta_3\}$	$E_{23} = \{\theta_2\}$	
$m_{a_{22}}$	0.1	0.4	0.5

The *bbas* must then be merged into a global representation. Denote by  $j_i$  the index of the criterion value for a given data item, and  $m_g$  the *bba* obtained by merging  $m_{a_{1j_1}}, \dots, m_{a_{sj_s}}$  through Eq. (2) with  $\oplus = \oplus_{MCS}$ . Merging the *bbas* of Example 3 result in a *bba* given by Table II.

Criteria forming MCS	MCS focal sets	Focal set	Mass of focal set
$\{A_1, A_2\}$	$E_{11} \cap E_{21}; E_{11} \cap E_{22}$	$\{\theta_3\}$	0.05
$\{A_1\}; \{A_2\}$	$E_{11} \cup E_{23}; E_{12} \cup E_{22}$	$\{\theta_2, \dots, \theta_5\}$	0.21
$\{A_1\}; \{A_2\}$	$E_{12} \cup E_{21}$	$\{\theta_1, \dots, \theta_5\}$	0.04
$\{A_1\}; \{A_2\}$	$E_{12} \cup E_{23}$	$\{\theta_2, \theta_4, \theta_5\}$	0.20
$\{A_1\}; \{A_2\}$	$E_{13} \cup E_{21}$	$\{\theta_1, \theta_2, \theta_3, \theta_5\}$	0.05
$\{A_1\}; \{A_2\}$	$E_{13} \cup E_{23}$	$\{\theta_2, \theta_5\}$	0.25
$\{A_1\}; \{A_2\}$	$E_{13} \cup E_{22}$	$\{\theta_2, \theta_3, \theta_5\}$	0.2

TABLE II  
EXAMPLE OF MERGING INDEPENDENT INFORMATION USING MCS

This simple example using two conflicting sources is sufficient to show that the MCS procedure solves conflicts by using a disjunctive approach, that is by considering that at least one source is right, but we do not know which one. In this merging procedure, all sources are treated equally, i.e. no source is judged more important than others. In practice, however, there will be cases where some criteria may be considered as more important than others. Integrating such information is the matter of future works.

## V. PRESENTATION OF RESULTS

Table II, obtained by the merging of two *bbas*, is difficult to read and is of little practical use. Hence, it is necessary to provide tools to present the final results in a readable way.

Given a set  $D = \{e_1, \dots, e_d\}$  of  $d$  data, we propose three complementary means to summarise their reliability evaluations: by ordering them, by providing a summarising interval and by giving insights into the origin of results.

In this section, we use the notion of lower and upper expectations of a function  $f : \Theta \rightarrow \mathbb{R}$  induced by a *bba*  $m_g$ . The lower expectation is defined as

$$\underline{\mathbb{E}}_g(f) = \sum_{A \subseteq \Theta} m(A) \min_{\theta \in A} f(\theta), \quad (3)$$

and the upper expectation  $\overline{\mathbb{E}}_g(f)$  is obtained by replacing  $\min$  with  $\max$  in Eq. (3). They are the infimum and supremum values of all expectations of  $f$  w.r.t. probability measures dominating the belief function induced by  $m_g$ .

### A. Comparing, evaluating and ordering data

Let  $m_{g_1}, \dots, m_{g_d}$  be the global *bbas* representing our knowledge about the reliability of  $e_1, \dots, e_d$ . To induce an order among the data, we need to compare *bbas*  $m_{g_1}, \dots, m_{g_d}$ . Such comparisons usually induce a partial order [27]. Here, we propose to achieve this comparison in two different ways: the first uses a numerical and interval-valued estimation of data reliability, using a particular function in Eq. (3); the second uses a generalisation of classical stochastic dominance to ordered spaces.

1) *Function based ordering*: We first propose to use lower and upper expectations of the function  $f_\Theta : \Theta \rightarrow \mathbb{R}$  such that  $f_\Theta(\theta_i) = i$  (each element  $\theta_i$  receives its rank as value). To the data item  $e_i$  is associated a reliability interval in the form

$$[\underline{\mathbb{E}}_{g_i}(f_\Theta), \overline{\mathbb{E}}_{g_i}(f_\Theta)],$$

where  $\underline{\mathbb{E}}_{g_i}(f_\Theta)$  (resp.  $\overline{\mathbb{E}}_{g_i}(f_\Theta)$ ) is the lower (resp. upper) expectation of  $f_\Theta$  w.r.t. the *bba*  $m_{g_i}$ .

**Example 4.** Consider the three *bbas*  $m_{g_1}, m_{g_2}, m_{g_3}$  respectively representing the reliability of  $e_1, e_2, e_3$  (e.g. resulting from the merging process illustrated in Example 3), defined over  $\Theta = \{\theta_1, \dots, \theta_5\}$  such that

$$m_{g_1}(\{\theta_1, \theta_2, \theta_3\}) = 0.3, \quad m_{g_1}(\{\theta_2, \theta_3\}) = 0.7;$$

$$m_{g_2}(\{\theta_3, \theta_4\}) = 0.5, \quad m_{g_2}(\{\theta_4, \theta_5\}) = 0.5;$$

$$m_{g_3}(\{\theta_1\}) = 0.4, \quad m_{g_3}(\{\theta_5\}) = 0.4, \quad m_{g_3}(\{\Theta\}) = 0.2.$$

Corresponding reliability intervals are:

$$[\underline{\mathbb{E}}_{g_1}(f_\Theta), \overline{\mathbb{E}}_{g_1}(f_\Theta)] = [1.7, 3]; \quad [\underline{\mathbb{E}}_{g_2}(f_\Theta), \overline{\mathbb{E}}_{g_2}(f_\Theta)] = [3.5, 4.5];$$

$$[\underline{\mathbb{E}}_{g_3}(f_\Theta), \overline{\mathbb{E}}_{g_3}(f_\Theta)] = [2.6, 3.4];$$

**Partial order** We propose to order the *bbas* according to the (partial) order  $\leq_{\mathbb{E}}$  such that

$$m_g \leq_{\mathbb{E}} m_{g'} \text{ iff } \underline{\mathbb{E}}_g(f_\Theta) \leq \underline{\mathbb{E}}_{g'}(f_\Theta) \text{ and } \overline{\mathbb{E}}_g(f_\Theta) \leq \overline{\mathbb{E}}_{g'}(f_\Theta)$$

In Example 4, we have  $e_1 <_{\mathbb{E}} e_3 <_{\mathbb{E}} e_2$  (in the sequel, when ordering data, we will make no difference between a datum  $e_i$  and its *bba*  $m_{g_i}$ ), obtaining in this case a complete order. However, as  $\leq_{\mathbb{E}}$  will be, in general, a partial order, we propose an algorithm allowing to build from it a complete (pre-)order, so that users are given an ordered list, easier to understand and interpret.

**Building groups** The next step is to order data by groups of decreasing reliability according to  $\leq_{\mathbb{E}}$ , i.e., to build an ordered partition  $\{D_1, \dots, D_O\}$  of  $D$ , where  $D_1$  corresponds to the most reliable data. Given a subset  $F \subseteq \{e_1, \dots, e_d\}$ , denote by  $opt(\mathbb{E}, F)$  the set of optimal data in the sense of reliability, i.e. not dominated w.r.t.  $\leq_{\mathbb{E}}$

$$opt(\mathbb{E}, F) = \{e_i \in F \mid \nexists e_j \in F, \text{ such as } e_i \leq_{\mathbb{E}} e_j\}.$$

The partition  $\{D_1, \dots, D_O\}$  can now be defined recursively as follows

$$D_i = opt(\mathbb{E}, (\{e_1, \dots, e_d\} \setminus \bigcup_{j=0}^{i-1} D_j)), \quad (4)$$

---

### Algorithm 2: Ordering data by decreasing reliability (function expectation based)

---

**Input:**  $\{e_1, \dots, e_d\}, m_{g_1}, \dots, m_{g_d}$  on  $\Theta$

**Output:** Partition  $\{D_1, \dots, D_O\}$

```

1 F = D = {e_1, ..., e_d}; j=1;
2 for i = 1, ..., d do
3   Compute [E_{g_i}(f_\Theta), \overline{E}_{g_i}(f_\Theta)]
4 while F \neq \emptyset do
5   foreach e_i \in F do
6     if \nexists e_j \in F s.t. e_i \leq_{\mathbb{E}} e_j then
7       Include e_i in D_j
8   F = F \setminus D_j;
9   j = j + 1

```

---

with  $D_0 = \emptyset$ . Algorithm 2 summarises the procedure. This comes down to retain, at each step, the non-dominated data in the sense of  $\leq_{\mathbb{E}}$  (lines 6-7), and then to remove them from the list of elements to order (line 8).

**Optimistic vs pessimistic** The partitioning could be done the other way around, starting by putting in a class  $D_i$  all data that do not dominate another one in the sense of  $\leq_{\mathbb{E}}$  (i.e., the worst data), before removing them from  $D$  and iterating. The main difference would be the behaviour of the partitioning algorithm towards data with imprecise reliability (i.e., *bbas* close to the vacuous one). In the case of Algorithm 2, such *bbas* are considered as reliable and put in a subset close to  $D_1$ , while they are considered as poorly reliable when partitioning is done in the other direction. Choosing between the two partitioning methods then depends of the attitude (either "optimistic" or "pessimistic") one wants to adopt towards data whose reliability is ill-known.

The advantage of using  $f_\Theta$  to order data is that it provides a numerical evaluation, easier to explain and to use, as scientists are used to numerical evaluations on bounded scales. However, one may object that the choice of  $f_\Theta$  is to some extent arbitrary, and that any choice of an increasing function on  $\Theta$  would be as justified as  $f_\Theta$  (e.g.,  $f(\theta_i) = i^2$ ). This is why we propose another ordering method that does not depend on the choice of a numerical function.

2) *Stochastic dominance based ordering*: Since the space  $\Theta$  is completely ordered, we can use a notion similar to stochastic ordering to compare and order different *bbas* defined on  $\Theta$ . Let  $\Omega_i = \{\theta_1, \dots, \theta_i\}$  be the set containing the  $i$  first elements of  $\Theta$ . Then, we propose to order *bbas* according to the (partial) order  $\ll$  such that

$$m_g \ll m_{g'} \text{ iff } Bel_g(\Omega_i) \geq Bel_{g'}(\Omega_i) \text{ and } Pl_g(\Omega_i) \geq Pl_{g'}(\Omega_i) \quad \forall i = 1, \dots, N$$

When  $m_g$  and  $m_{g'}$  are probability distributions, the order  $\ll$  reduces to the stochastic ordering on ordered spaces. When  $m_g$  and  $m_{g'}$  are two categorical *bbas* such that  $m_g(I_{a,b}) = 1$  and  $m_{g'}(I_{a',b'}) = 1$ , then  $m_g \ll m_{g'}$  iff  $\max(I_{a,b}, I_{a',b'}) = I_{a,b}$ , therefore  $\ll$  is also coherent with



the ordering of  $\Theta$  elements. The latter remark also holds when using expectation bounds, while the first does not necessarily hold (two probability distributions can be incomparable w.r.t. stochastic ordering, but will always be comparable w.r.t. classical expectations). Another interesting property [28] is that for any real-valued increasing function  $f : \Theta \rightarrow \mathbb{R}$  on  $\Theta$  (i.e.,  $f(\theta_i) \leq f(\theta_j)$  if  $i \leq j$ ),  $m_g \ll m_{g'}$  implies that  $\underline{\mathbb{E}}_g(f) \leq \underline{\mathbb{E}}_{g'}(f)$  and  $\overline{\mathbb{E}}_g(f) \leq \overline{\mathbb{E}}_{g'}(f)$ . This shows that  $\ll$  is compatible with any increasing function  $f$  on  $\Theta$  and avoids the choice of one, at least to compare and order the *bbas*  $m_{g_1}, \dots, m_{g_d}$ . It also means that using a function  $f$  and  $\leq_{\mathbb{E}}$  will provide a refinement of  $\ll$ .

**Example 5.** Let us again consider the three *bbas* of Example 4. Belief and plausibility measures on events  $\Omega_i = \{\theta_1, \dots, \theta_i\}$  are summarised in the following table

	$\Omega_1$	$\Omega_2$	$\Omega_3$	$\Omega_4$	$\Omega_5$
$Bel_{g_1}$	0	0	1	1	1
$Bel_{g_2}$	0	0	0	0.5	1
$Bel_{g_3}$	0.4	0.4	0.4	0.4	1
$Pl_{g_1}$	0.3	1	1	1	1
$Pl_{g_2}$	0	0	0.5	1	1
$Pl_{g_3}$	0.6	0.6	0.6	0.6	1

According to  $\ll$ , we have  $m_{g_1} \ll m_{g_2}$ , but  $m_{g_3}$  is not comparable with  $m_{g_1}$  or  $m_{g_2}$ . The ordering obtained in Example 4 is a refinement of the ordering obtained with  $\ll$ .

Algorithm 2 can be adapted to  $\ll$ , simply replacing  $\leq_{\mathbb{E}}$  with  $\ll$  (Line 6). The previous remarks concerning the user attitude towards ill-known reliabilities hold for both cases.

**Example 6.** Consider the three *bbas* of Example 4. Applying Algorithm 2 with  $\ll$ , we get  $D_1 = \{e_2, e_3\}$  and  $D_2 = \{e_1\}$ . Partitioning by starting from non-dominating data would have given  $D_1 = \{e_2\}$  and  $D_2 = \{e_1, e_3\}$ . This example illustrates the fact that both ordering procedures allow to differentiate the attitude towards imprecision and incomparable elements.

### B. Explaining the results

Beyond the ordering of data and the evaluation of their reliability, it is natural to be interested in the reasons explaining why some particular data were assessed as (un)reliable. We now show how maximal coherent subsets of criteria, i.e. groups of agreeing criteria, may provide some insight as to which reasons have led to a particular assessment.

According to our method, the more often  $F \subseteq \{A_1, \dots, A_S\}$  is an MCS appearing in  $\bigoplus_{i=1}^S (E_i) = E$  (see Eq. 2), the more important the impact of the maximal coherent subset  $F \subseteq \{A_1, \dots, A_S\}$  is on the global reliability score  $m_g$ . Therefore, we propose to measure the importance  $w(F)$  that a maximal coherent subset  $F$  of criteria has in  $m_g$  by summing all the masses of  $m_g$  for which it has been a maximal coherent subset, that is

$$w(F) = \left\{ \sum_{i=1}^S \prod_{E_i \in F} m_i(E_i) \mid F \text{ is an MCS of } \bigoplus_{i=1}^S (E_i) \right\}. \quad (5)$$

The value  $w(F)$  lies in the unit interval  $[0, 1]$ . It is maximal if  $F$  is an MCS for every combination of focal sets, and minimal if  $F$  is never an MCS. The higher the value  $w(F)$  of a given subset of criteria, the more these criteria are agreeing.

**Example 7.** Consider the results of Table II. The impact  $w$  of the different encountered maximal coherent subsets is evaluated as follows:

$$w(\{A_1\}) = 0.95, \quad w(\{A_2\}) = 0.95, \quad w(\{A_1, A_2\}) = 0.05,$$

from which it can be inferred that the two criteria  $\{A_1\}$  and  $\{A_2\}$ , appearing often alone, do not agree with each other. This means that the imprecision in the reliability interval can be explained by the conflict between criteria  $A_1$  and  $A_2$ .

In this example, this information may seem unnecessary, as the analysis is straightforward. However, when dealing with thousands of data and over five criteria groups, such tools may help users to do a quick analysis and retain the data that best serve their purposes. Also, in the case where the user is not an expert (but has vague knowledge of the considered scientific domain), even such slight insight may be useful.

## VI. APPLICATION TO THE DESIGN OF A WEB-ENABLED DATA WAREHOUSE

In this section, we present an application of the method to @Web, a web-enabled data warehouse. Indeed, the framework developed in this paper was originally motivated by the need to estimate the reliability of scientific experimental results collected in open data warehouses. To lighten the burden laid upon domain experts when selecting data for a particular application, it is necessary to give them indicative reliability estimations. Formalizing reliability criteria will hopefully be a better asset for them to justify their choices and to capitalize knowledge than the use of an *ad-hoc* estimation.

For this application, numerical evaluations (Section V-A1) were chosen, the reason being that the initial *ad-hoc* evaluation system proposed numerical evaluations, hence users were more at ease with them. Tools development was carefully done semantic web recommended languages, so that created tools would be generic and reusable in other data warehouses. This required an advanced design step, which is important to ensure modularity and to foresee future evolutions.

The current version of @Web has been implemented using the W3C recommended languages (<http://www.w3.org/TR/>): OWL to represent the domain ontology, RDF to annotate Web tables and SPARQL to query annotated Web tables. The readers not familiar with the semantic web tools may refer to [29] for details, as this section is centered on the extended version of the OWL ontology, RDF annotations and SPARQL querying necessary for Web data reliability management. Nevertheless, to show the potential of these tools, we will illustrate the concepts with examples easy to understand. All belief functions-related computations have been done thanks to the R [30] package belief [31].

The section is organized as follows. We first describe the purpose and architecture of the data warehouse. We then

focus on the extension developed to implement the reliability estimation. Finally we provide a real-world use case in Food predictive microbiology.

It is worthwhile to note that this application provides another good example of the difference between the notions of relevance and reliability: while relevant data are the answers returned by a query, some of these answers may be unreliable. Indeed, data base queries return relevant, but possibly unreliable, answers.

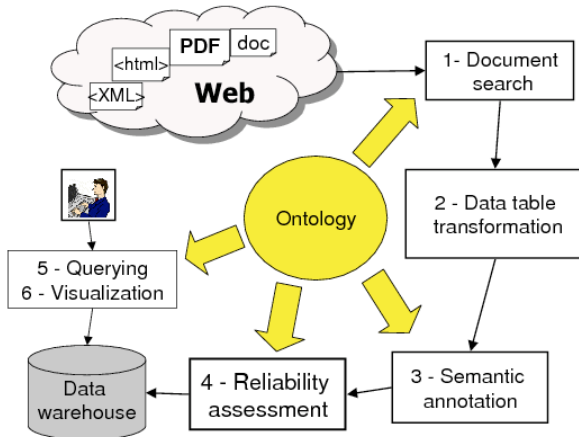


Fig. 3. Main steps of the document workflow in @Web

#### A. @Web presentation

@Web is a data warehouse opened on the Web [2], [3]. Its current version is centered on the integration of heterogeneous data tables extracted from Web documents. The focus has been put on Web tables for two reasons: (i) experimental data are often summarized in tables and (ii) data are already structured and easier to integrate in a data warehouse than, e.g., text or graphics.

The main steps of Web table integration are given in Fig. 3. The key point of data integration in @Web is the central role played by the domain ontology. This ontology describes the concepts, their terminology and the relationships between concepts proper to a given application domain. Thanks to this feature, @Web can be instantiated for any application domain by defining the corresponding ontology including the domain knowledge. For instance, @Web has already been instantiated and tested in various domains such as food predictive microbiology, chemical risk in food and aeronautics [3].

Once the ontology is built, data integration in the warehouse is done according to the steps of Fig. 3. Concepts found in a data table and semantic relations linking these concepts are automatically recognised and annotated, which allows interrogation and querying in an homogeneous way.

The @Web instance used here is implemented in the Sym'Previous [32] decision support system which simulates the growth of a pathogenic microorganism in a food product. Semantic relations in this system include for example the *GrowthRate* linking a microorganism and a food product to

the corresponding growth rate and its associated parameters. After semantic annotation, data retrieved from tables can be used for various tasks (e.g., estimate a model parameter).

1) *@Web generic ontology*: The current OWL ontology used in the @Web system is composed of two main parts: a generic part, the *core ontology*, which contains the structuring concepts of the Web table integration task, and a specific part, the *domain ontology*, which contains the concepts specific to the domain of interest. The *core ontology* is composed of symbolic concepts, numeric concepts and relations between them. It is therefore separated from the definition of the concepts and relations specific to a given domain, the *domain ontology*. All the ontology concepts are materialized by OWL classes. For example, in the microbiological ontology, the symbolic concept *Microorganism* and the numeric concept *pH* are represented by OWL classes that are subclasses of the generic classes *SymbolicConcept* and *NumericConcept*, respectively. Fig. 4 gives an excerpt of an OWL class organization for symbolic concepts.

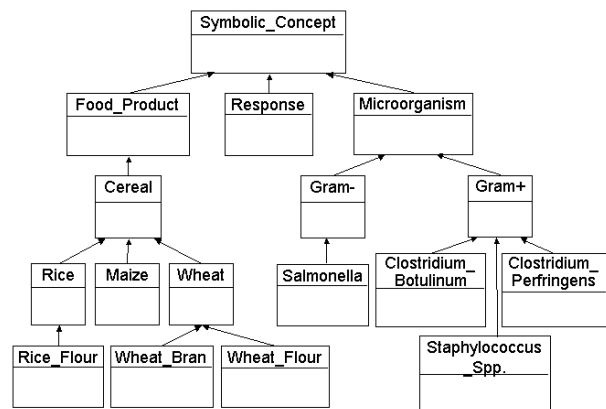


Fig. 4. Excerpt of OWL class hierarchy for symbolic concepts in the microbial domain

2) *@Web workflow*: The first three steps of @Web workflow (see Fig. 3) are the following: the first task consists in retrieving relevant Web documents (in html or pdf) for the application domain, using key-words extracted from the domain ontology. It does so by defining queries executed by different crawlers; in the second task, data tables are extracted from the retrieved documents and are semi-automatically translated into a generic XML format. The Web tables are then represented in a classical and generic way – a table is a set of lines, each line being a set of cells; in the third task, the Web tables are semantically annotated according to the domain ontology. This annotation consists in identifying what semantic relations of the domain ontology can be recognized in each row of the Web table (see [3] for details). This process generates RDF descriptions.

**Example 8.** Table III is an example of a Web table in which the semantic relation *GrowthParameterAw* has been identified. The domain of this relation is a *Microorganism* and its range is food product water activity ( $a_w$ ), a dimensionless value. For instance, the first row indicates that *Clostridium*

Organism	$a_w$ min.	$a_w$ optimum	$a_w$ max.
Clostridium	0.943	0.95-0.96	0.97
Staphylococcus	0.88	0.98	0.99
Salmonella	0.94	0.99	0.991

TABLE III  
EXAMPLE OF A WEB TABLE

water activity ( $a_w$ ) ranges from 0.943 to 0.97, and is known to be optimal in the range [0.95, 0.96].

Some of the RDF descriptions associated with Web tables by our semantic annotation process include values expressed as fuzzy sets (e.g.,  $a_w$  values). Let us now introduce the use of fuzzy sets in @Web before illustrating it.

3) *Use of fuzzy sets in @Web:* We distinguish two kinds of fuzzy sets: (i) discrete and (ii) continuous. Each kind will be used in @Web for specific purposes.

**Definition 1.** A **discrete fuzzy set**  $\mu$ , denoted by DFS in the RDF description, is a fuzzy set associated with a relation or a symbolic concept of the ontology. Its definition domain is the set of relations or the set of subclasses of the symbolic concept. We denote by  $\{(x, y), \dots\}$  the fact that element  $x$  has membership degree  $y$ .

**Definition 2.** A **continuous fuzzy set**  $\mu$ , denoted by CFS in the RDF description, is a trapezoidal fuzzy set associated with a numeric concept of the ontology. A trapezoidal fuzzy set  $\mu$  is defined by its four characteristic points  $[a, b, c, d]$  which correspond to its support  $[a, d]$  and its kernel  $[b, c]$  (with a linear interpolation between  $a, b$  and  $c, d$ ). Its definition domain is the interval of possible values for the concept.

The fuzzy values used to annotate Web tables may express two of the three classical semantics of fuzzy sets (see [33]): similarity or imprecision. In the @Web system, similarity interpretation is used to recognize symbolic concepts and relations inside the table, while imprecision interpretation is used when modelling some ill-known values about some particular instances of numerical concepts.

**Example 9.** Fig. 5 gives the main part of the RDF description corresponding to the recognition of the relation *GrowthParameterAw* in the first row of the Web Table III, denoted by *uriRow1* in Figure 5.

Starting from the left part of the figure, the row is annotated by a discrete fuzzy set *DFSR1*. A list of closest relations is extracted from within the ontology, in the present case a single element corresponding to the relation *GrowthParameterAw* (*GPaw1*). The membership degree (1 here) is a certainty score, denoted *ps*, expressing the degree of certainty associated with the relation recognition. In the top right part of the figure, the domain of the relation *GrowthParameterAw*, an instance of the symbolic concept *Microorganism*, is annotated by a discrete fuzzy set. This fuzzy set, typed by the OWL class *DFS*, has a semantic of similarity and gives the list of closest ontology concepts compared to *Clostridium* (First row of Table III). Starting from the end, we see two instances of symbolic concepts, *CPerfring* for *Clostridium*

*Perfringens*, and *CBotulinum* for *Clostridium Botulinum*, with membership degree equal to 0.5 for each of them.

Finally one can see the use of continuous fuzzy sets, like *CFS1* on the bottom right, to express numerical values associated with the range of the relation *GrowthParameterAw*.

4) *SPARQL querying of RDF graphs:* In the XML/RDF data warehouse, the querying is done through *MIEL++* queries. We briefly recall how *MIEL++* queries are executed in the current version of @Web (see [2] for details). A *MIEL++* query is asked in a view which corresponds to a given relation of the ontology (e.g., the relation *GrowthParameterAw* of example 8). A *MIEL++* query is an instantiation of a given view by the end-user, specifying among the set of queryable attributes of the view, which are the selection attributes (i.e., the one used to select relevant answers) and their corresponding searched values, and which are the projection attributes (i.e., the one displayed in the answers).

In such *MIEL++* queries, fuzzy sets allow representing end-users preferences (the third semantic [33] of fuzzy sets) and are used to retrieve not only exact answers but also answers which are semantically close (kernel matching versus support matching). Since the XML/RDF data warehouse contains fuzzy values generated by the annotation process, the query processing has (1) to consider the certainty score associated with the semantic relations identified in Web tables and (2) to compare a fuzzy set expressing querying preferences to a fuzzy set, generated by the annotation process, having a semantic of similarity or imprecision.

**Example 10.** Let us define a *MIEL++* query  $Q$  expressed in the view *GrowthParameterAw* as follows:

$$Q = \{ \text{Microorganism}, aw \mid \\ (\text{GrowthParameterAw}(\text{Microorganism}, aw) \wedge \\ (\text{Microorganism} \approx \text{MicroPreferences}) \wedge \\ (aw \approx \text{awPreferences}) \wedge (ps \geq 0.5)) \}.$$

In  $Q$ , projection attributes are *Microorganism* and *aw*, while the second part describe selection attributes. The discrete fuzzy set *MicroPreferences*, which is equal to  $\{(Gram+, 1), (Gram-, 0.5)\}$ , means that the end-user is firstly interested in microorganisms which are *Gram+* and secondly *Gram-* (see Fig. 4). The continuous fuzzy set *awPreferences* with characteristic points [0.9, 0.94, 0.97, 0.99] means that the end-user is first interested in  $a_w$  values in the interval [0.94, 0.97] (the kernel of the fuzzy set) but that he/she accepts to enlarge the querying till the interval [0.9, 0.99] (the support of the fuzzy set). *GrowthParameterAw* relations having a certainty score (denoted *ps*) inferior to 0.5 are discarded.

Since fuzzy sets are not supported in a standard SPARQL query, some technical difficulties arise. A complete solution to translate a *MIEL++* query into a standard SPARQL query is presented in [2]. In this paper, we only recall how is measured the satisfaction of a *MIEL++* selection criterion. The two semantics -imprecision and similarity- associated with fuzzy values of the XML/RDF data warehouse are con-



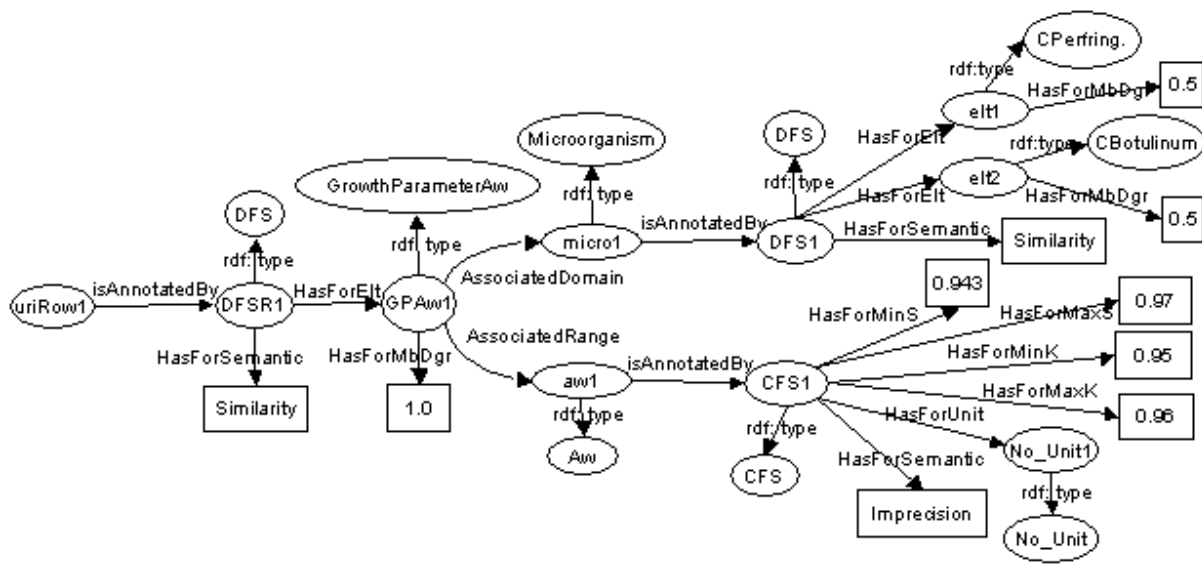


Fig. 5. Example of RDF annotations generated from the Web Table III

sidered. On one hand, two classical measures ([34]) have been proposed to compare a fuzzy set representing preferences to a fuzzy set having a semantic of imprecision: a possibility degree of matching denoted  $\Pi$  and a necessity degree of matching denoted  $N$ . On the other hand, an adequation degree [35], denoted  $ad$ , to compare a fuzzy set representing preferences to a fuzzy set having a semantic of similarity is computed. Based on those scores, we propose to define a lexicographic order on the answers which gives greater importance to the most pertinent answers compared with the domain ontology. The answers are ordered first according to  $ps$  (the certainty score associated with the relation), second to  $ad$ , third to  $N$  and finally to  $\Pi$ ,  $N$  being considered more important than  $\Pi$ .

**Example 11.** The answers to the SPARQL query associated with the MIEL++ query of Example 10 compared with the Web Table III are summarised in Table IV.

### B. Extending @Web for data reliability management

This section presents @Web extension integrating a reliability estimation to each table, in order to display the results of a user query ordered by decreasing reliability values. Even if a data table can include several items, the table level has been retained, as data from a given table are usually issued from the same experimental set-up and therefore share the same reliability criteria.

#### 1) Extending the ontology to include reliability criteria:

Some criteria retained for reliability estimation are part of the domain knowledge. For example, methods to count microorganisms all have equivalent precision, while the accuracy of methods to appreciate wheat grain size greatly varies. Therefore, it is natural to include criteria in the domain ontology. This solution allows designers to adapt the choice of the criteria associated with a given application domain, preserving the @Web generic approach at the same time.

In the extended ontology, the *core ontology* is enriched with corresponding symbolic and numeric criteria. The *domain ontology* is completed by the definition of the criteria selected to evaluate the reliability in the application domain, together with their possible values. For example, the symbolic criterion *SourceType* and the numeric criterion *CitationNumber* are represented by OWL classes and are in the *domain ontology*. They are subclasses of the generic classes *SymbolicCriterion* and *NumericCriterion*, respectively, which belong to the *core ontology*. As for symbolic concepts, the values associated with a symbolic criterion are represented by OWL classes that are subclasses of the OWL class representing the criterion.

2) *Storing data reliability criteria in RDF graphs:* In the extended version of @Web, an additional fourth task is introduced (see Fig. 3) to implement the reliability management. Users (currently) manually enter the values associated with the reliability criteria for each Web table. This information is stored in a RDF graph associated with the Web table, since this was the retained level.

**Example 12.** Fig. 6 presents the RDF descriptions representing the reliability criteria and values associated with the Web Table III. They express that the table (having the *uriTable1* identifier within the XML document) has for associated criteria the same values as in Example 3: journal paper and no repetitions of experiments.

The fourth task output of the extended @Web system is an XML/RDF data warehouse composed of a set of XML documents which represent Web tables, together with the (fuzzy) RDF annotations corresponding to the recognized semantic relations and the reliability criteria values.

3) *SPARQL queries and data reliability:* To evaluate the reliability of the answers associated with a MIEL++ query, the following post-processing is executed. Each answer being



$ps$	$ad$	$N$	$\Pi$	$Microorg$	$aw$
1	0.5	1	1	{(Clost. Perf., 0.5), (Clost. Botu., 0.5)}	[0.943, 0.95, 0.96, 0.97]
1	0.5	0.5	0.68	{(Staphy. spp., 0.5), (Staphy. Aureus., 0.5)}	[0.88, 0.98, 0.98, 0.99]
1	0.5	0	0.965	{(Salmonella, 1.0)}	[0.94, 0.99, 0.99, 0.991]

TABLE IV  
ANSWERS TO EXAMPLE 10 MIEL++ QUERY

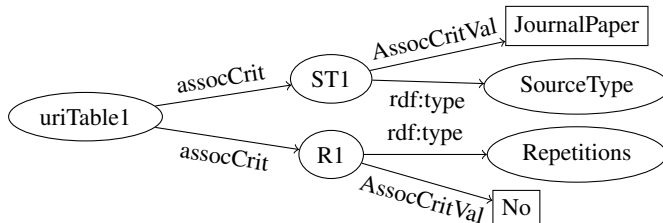


Fig. 6. Example of RDF annotations representing the criteria and values associated with the Web Table III

associated with a given row of a Web table, the reliability criteria are retrieved thanks to SPARQL queries, automatically generated using the ontology and table URIs, determined from URIs associated with rows. Answers are then compared and ordered using the framework of Section IV-B and related reliability intervals are calculated according to Section V .

### C. Use case in the field of predictive food microbiology

This part is dedicated to a use case in the field of predictive food microbiology, namely the selection of reliable parameters for simulation models. We first give the criteria suited to this field, as well as the corresponding expert opinions and fuzzy sets. We then detail the use case query and results.

1) *Customized criteria*: We define criteria corresponding to the three kinds discussed in Section III and recalled in Table I: source, production and statistics. They have been determined with the expert, and their values are given in Table V, together with the labels. The  $\mathcal{A}_4$  criterion: number of strains used in experiments, is domain specific.

	$\mu(\theta_1)$	$\mu(\theta_2)$	$\mu(\theta_3)$	$\mu(\theta_4)$	$\mu(\theta_5)$
Unknown	1	1	1	1	1
Very unrel.	1	0.5	0.1	0	0
Slightly unrel.	0.9	1	0.5	0	0
Neutral	0	0.5	1	0.5	0
Reliable	0	0	0.5	1	0.9
Very reliable	0.	0	0.1	0.5	1

TABLE VI  
FUZZY SETS ON  $\Theta$  CORRESPONDING TO LINGUISTIC MODALITIES

2) *Fuzzy sets definition*: Associated fuzzy sets defined on  $\Theta$  are shown in Table VI. Let us point out the difference between the *Unknown* modality, meaning a lack of information, and the *Neutral* one, associated to a neutral opinion.

3) *Expert opinions*: Experts provided opinions about reliability values for the different criteria labels, according to the linguistic terms defined in Table VI. They are summarised in Table VII for  $\mathcal{A}_1$ ,  $\mathcal{A}_2$  and  $\mathcal{A}_4$  and in Table VIII for the  $\mathcal{A}_3$  group (*citation number and publication age*).

$\mu_{a11}$	very reliable
$\mu_{a12}$	very reliable or reliable
$\mu_{a13}$	neutral
$\mu_{a14}$	slightly unreliable
$\mu_{a21}$	reliable
$\mu_{a22}$	slightly unreliable
$\mu_{a41}$	very unreliable
$\mu_{a42}$	reliable
$\mu_{a43}$	very reliable

TABLE VII  
EXPERT OPINIONS ABOUT RELIABILITY FOR  $\mathcal{A}_1$ ,  $\mathcal{A}_2$  AND  $\mathcal{A}_4$

#cit. \ age	0-2	3-8	+8
0-5	unknown	slightly unrel.	very unrel.
6-10	unknown	neutral	slightly unrel.
11-20	slightly rel.	slightly rel.	slightly unrel.
21-40	very rel.	very rel.	slightly rel.
+40	very rel.	very rel.	very rel.

TABLE VIII  
EXPERT OPINIONS ABOUT RELIABILITY FOR  $\mathcal{A}_3$  - CITATION NUMBER AND PUBLICATION AGE INTERDEPENDENT CRITERIA

4) *Illustration*: Consider a search for the most reliable input parameter to simulate the growth of *Listeria monocytogenes* or *Clostridium Perfringens*, in a given food product. To do that, the data warehouse is queried using the *GrowthParameterTemp* semantic relation of the ontology, i.e., the relevant relation to retrieve bacterial growth parameters. This allows users to retrieve the minimum temperature for the growth of *Listeria monocytogenes* or *Clostridium Perfringens*. It corresponds to the following MIEL++ query  $Q$ , expressed in the *GrowthParameterTemp* view:

$$Q = \{ \text{Microorganism, Temperature} \mid \\ (\text{GrowthParameterTemp}(\text{Microorganism, Temperature}) \\ \wedge (\text{Microorganism} \approx \text{MicroPreferences})) \wedge (ps \geq 0.5) \}$$

with the discrete fuzzy set (here reduced to two crisp values) *MicroPreferences* equals to  $\{(Listeria\ monocytogenes, 1.0), (Clostridium\ Perfringens, 1.0)\}$ . In that case, five particular instances of Web tables issued from @Web match the query. The citation number has been determined using Google Scholar, and reliability features are given in Table IX for all five instances.

The answers to the user query including reliability assessments are given below. The first column contains the reliability summary as a numerical interval (see Section V-A1). The  $ps$  and  $ad$  columns are MIEL++ answer scores, already presented in Example 11. Note that the value of  $nb$  strains has no influence on the reliability summary of Web tables  $e_2$ ,  $e_3$  and  $e_4$ , as it is not available for them.

Source				Statistics		Production	
$\mathcal{A}_1$ source type				$\mathcal{A}_2$ experiment repetition		$\mathcal{A}_3$ citation number $\times$ publication age	$\mathcal{A}_4$ nb strains
journal paper $a_{11}$	international report $a_{12}$	project $a_{13}$	other $a_{14}$	Yes $a_{21}$	No $a_{22}$	{[0,5],[6,10],[11,20],[21,40],[40+]} (number) $\times$ {[0,2],[3,8],[8+]} (age)	{[1,5],[6,20],[20+]} $\{a_{41}, a_{42}, a_{43}\}$

TABLE V  
RELIABILITY CRITERIA

Table Id	$\mathcal{A}_\infty$ (#cit vs age)	$\mathcal{A}_\in$ repetitions	$\mathcal{A}_\supset$ source type	$\mathcal{A}_\Delta$ nb strains
$e_1$	(21,5)	yes	journal paper	21
$e_2$	(42,6)	no	internat. report	-
$e_3$	(61,23)	no	journal paper	-
$e_4$	(0,6)	no	internat. report	-
$e_5$	(21,5)	yes	journal paper	5

TABLE IX  
RELIABILITY CRITERIA VALUES ASSOCIATED WITH THE FIVE INSTANCES OF WEB TABLES CORRESPONDING TO THE QUERY

$e_i$	$\mathbb{E}_{g_i}(f_\Theta), \bar{\mathbb{E}}_{g_i}(f_\Theta)$	$ps$	$ad$	<i>Microorg.</i>	$T_{\min}(\text{°C})$
$e_1$	[4.60, 4.95]	1	1.0	{(L. M., 1.0)}	[-1.7, -1.7]
$e_2$	[1.65, 4.95]	1	1.0	{(L. M., 1.0)}	[0, 2]
$e_3$	[1.29, 4.99]	1	0.5	{(L. M., 0.5)}	[-1.4, -1.0]
$e_4$	[1.44, 4.25]	1	1.0	{(L. M., 1.0)}	[-0.4, -0.4]
$e_5$	[1.21, 4.90]	1	1.0	{(C. P., 1.0)}	[12.0, 12.0]

Using Algorithm 2, we get the ordering  $\{e_1, e_3\} \leq_{\mathbb{E}} \{e_2\} \leq_{\mathbb{E}} \{e_3, e_5\}$ .  $e_1$  precise reliability assessment comes from  $e_1$  having good scores for the four reliability criteria without discordance between them. This makes  $e_1$  the best candidate. The other ones have conflicting criteria, as shown by the large difference between  $\mathbb{E}_{g_i}(f_\Theta)$  and  $\bar{\mathbb{E}}_{g_i}(f_\Theta)$ ,  $\forall i \in \{2, 3, 4, 5\}$ . For instance, if we consider the Web table  $e_4$ , the impact  $w$  of the different encountered maximal coherent subsets is evaluated as follows (Eq. (5)):

$$w(\{A_2, A_3, A_4\}) = 0.875, \quad w(\{A_1, A_4\}) = 0.625,$$

$$w(\{A_1, A_2, A_4\}) = 0.125, \quad w(\{A_1, A_3, A_4\}) = 0.125,$$

Examination of these results shows that group  $\{A_1, A_4\}$  and group  $\{A_2, A_3, A_4\}$  of criteria do not agree with each other. The imprecision in the reliability score can be explained by the conflict between these groups. The *source type* is reliable but, from the other criteria (*repetition* and *nb citations vs age*) point of view, the Web table  $e_4$  is not very reliable ( $A_4$  being unknown). Therefore, in term of reliability, the user is clearly encouraged to choose  $e_1$ . This case study also shows that domain or expert knowledge is essential in reliability evaluation, since both  $e_1$  and  $e_5$  tables come from the same paper (hence have the same source) but have very different reliability, due to the differences in experimental setup.

To study the method sensitivity to changes in  $\Theta$  and in fuzzy set definition, we have changed  $N$  from 5 to 7 and have applied random changes between  $-0.1$  and  $0.1$  to membership degrees. Except for numerical scale bounds, none of these changes modified the evaluation results.

#### D. Discussion

Implementation allowed us to demonstrate the approach feasibility and to have returns from non-computer scientists, more specifically from experts involved in the case study.

On the pros side, first returns were enthusiastic, the expert found the reliability information useful and the method (once explained) more meaningful and readable than the current ad hoc technique. Criteria are available for display from within the interface and the expert can customise the system. Another good point is that interface complexity is only slightly increased from the user point of view, even if underlying mechanisms are more complex. Finally the presentation of the results by ordered reliability groups seems to be satisfactory, as it allows users to easily select the most reliable data or comparable ones in case of ties.

On the cons side, some limits were detected. Considering criteria importance, e.g., using weights, appears as an important request. Also, some additional efforts to explain or summarise the results seems necessary, in particular in presence of several (e.g., 8 to 10) criteria and answers (over a hundred). Some criteria, such as *source type*, also appear to be the subjects of debates, and they could necessitate some dynamical construction process. Finally, users pointed out the need to integrate reliability assessments in further processing, such as in the aggregation of multiple data.

#### VII. CONCLUSION AND PERSPECTIVES

We proposed a generic method to evaluate the reliability of data automatically retrieved from the Web or from electronic documents. Even if the method is generic, we were more specifically interested in scientific experimental data.

The method evaluates data reliability from a set of common sense (and general) criteria. It relies on the use of basic probabilistic assignments and of induced belief functions, since they offer a good compromise between flexibility and computational tractability. To handle conflicting information while keeping a maximal amount of it, the information merging follows a maximal coherent subset approach. Finally, reliability evaluations and ordering of data tables are achieved by using lower/upper expectations, allowing us to reflect uncertainty in the evaluation. The results displayed to end-users is an ordered list of tables, from the most to the least reliable ones, together with an interval-valued evaluation.

We have demonstrated the applicability of the method by its integration in the @Web system, and its use on the *Sym'Previus* data warehouse. As future works, we see two main possible evolutions:

- complementing the current method with useful additional features: the possibility to cope with multiple

experts, with criteria of non-equal importance and with uncertainly known criteria;

- combining the current approach with other notions or sources of information: relevance, in particular, appear to be equally important to characterise experimental data. Also, we may consider adding users feedback as an additional (and parallel) source of information about reliability or relevance, as it is done in Web applications.

## REFERENCES

- [1] S. Ramchurn, D. Huynh, and N. Jennings, "Trust in multi-agent systems," *The Knowledge Engineering Review*, vol. 19, pp. 1–25, 2004.
- [2] P. Buche, J. Dibie-Barthélemy, and H. Chebil, "Flexible sparql querying of web data tables driven by an ontology," in *FQAS*, ser. Lecture Notes in Computer Science, vol. 5822, 2009, pp. 345–357.
- [3] G. Hignette, P. Buche, J. Dibie-Barthélemy, and O. Haemmerlé, "Fuzzy annotation of web data tables driven by a domain ontology," in *ESWC*, ser. Lecture Notes in Computer Science, vol. 5554, 2009, pp. 638–653.
- [4] D. Mercier, B. Quost, and T. Denoeux, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Information Fusion*, vol. 9, pp. 246–258, 2008.
- [5] R. Cooke, *Experts in uncertainty*. Oxford, UK: Oxford University Press, 1991.
- [6] S. Sandri, D. Dubois, and H. Kalfsbeek, "Elicitation, assessment and pooling of expert judgments using possibility theory," *IEEE Trans. on Fuzzy Systems*, vol. 3, no. 3, pp. 313–335, August 1995.
- [7] F. Delmotte and P. Borne, "Modeling of reliability with possibility theory," *IEEE Trans. on Syst., Man, and Cybern. A*, vol. 28, no. 1, pp. 78–88, 1998.
- [8] F. Pichon, D. Dubois, and T. Denoeux, "Relevance and truthfulness in information correction and fusion." in *International Journal of Approximate Reasoning*, vol. Accepted for Publication, 2011.
- [9] J. Sabater and S. Sierra, "Review on computational trust and reputation models," *Artificial Intelligence Review*, vol. 24, no. 33–60, 2005.
- [10] J. Golbeck and J. Hendler, "Inferring reputation on the semantic web," in *Proceedings of the 13th international World Wide Web conference*, New York, NY, USA, 2004.
- [11] Y. Gil and D. Artz, "Towards content trust of web resources," in *WWW'06 : Proceedings of the 15th international conference on World Wide Web*, New York, NY, USA, 2006, pp. 565–574.
- [12] K. Quinn, D. Lewis, D. O'Sullivan, and V. Wade, "An analysis of accuracy experiments carried out over a multi-faceted model of trust," *International Journal of Information Security*, vol. 8, pp. 103–119, 2009.
- [13] A. Denguir-Rekik, J. Montmain, and G. Mauris, "A possibilistic-valued multi-criteria decision-making support for marketing activities in e-commerce: Feedback based diagnosis system," *European Journal of Operational Research*, vol. 195, no. 3, pp. 876–888, 2009.
- [14] I. N. Chengalur-Smith, D. P. Ballou, and H. L. Pazer, "The impact of data quality information on decision making: An exploratory analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, pp. 853–864, 1999.
- [15] L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-i," *Information Sciences*, vol. 8, pp. 199–249, 1975.
- [16] —, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, pp. 3–28, 1978.
- [17] G. Miller, "The magical number seven, plus or minus two," *The Psychological Review*, vol. 63, pp. 81–97, 1956.
- [18] S. Destercke, D. Dubois, and E. Chojnacki, "Possibilistic information fusion using maximal coherent subsets," *IEEE Trans. on Fuzzy Systems*, vol. 17, pp. 79–92, 2009.
- [19] G. Shafer, *A mathematical Theory of Evidence*. New Jersey: Princeton University Press, 1976.
- [20] P. Smets, "Analyzing the combination of conflicting belief functions," *Information Fusion*, vol. 8, pp. 387–412, 2006.
- [21] D. Dubois and H. Prade, "On several representations of an uncertain body of evidence," in *Fuzzy Information and Decision Processes*, M. Gupta and E. Sanchez, Eds. North-Holland, 1982, pp. 167–181.
- [22] T. Denoeux, "Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence," *Artificial Intelligence*, vol. 172, pp. 234–264, 2008.
- [23] A. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.
- [24] N. Rescher and R. Manor, "On inference from inconsistent premises," *Theory and Decision*, vol. 1, pp. 179–219, 1970.
- [25] S. Benferhat, D. Dubois, and H. Prade, "Reasoning in inconsistent stratified knowledge bases," in *Proc. of the 26th Inter. Symp. on Multi-Valued Logic (ISMVL)*, 1996, pp. 184–189.
- [26] D. Dubois, H. Fargier, and H. Prade, "Multi-source information fusion: a way to cope with incoherences," in *Proc. of French Days on Fuzzy Logic and Applications (LFA)*, Cepadues, Ed. La Rochelle: Cepadues, 2000, pp. 123–130.
- [27] D. Dubois and H. Prade, "A set-theoretic view on belief functions: logical operations and approximations by fuzzy sets," *I. J. of General Systems*, vol. 12, pp. 193–226, 1986.
- [28] T. Denoeux, "Extending stochastic ordering to belief functions on the real line," *Information Sciences*, vol. 179, pp. 1362–1376, 2009.
- [29] G. Antoniou and F. van Harmelen, *A semantic Web primer*. Cambridge, Massachusetts: The MIT Press, 2008.
- [30] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2005, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org>
- [31] N. Maillat, B. Charnomordic, and S. Destercke, *belief: Contains basic functions to manipulate belief functions and associated mass assignments*, 2010, r package version 1.0. [Online]. Available: <http://CRAN.R-project.org/package=belief>
- [32] P. Buche, O. Couvert, J. Dibie-Barthélemy, G. Hignette, E. Mettler, and L. Soler, "Flexible querying of web data to simulate bacterial growth in food," *Food Microbiology*, vol. 28, pp. 685–693, 2010.
- [33] D. Dubois and H. Prade, "The three semantics of fuzzy sets," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 141–150, 1997.
- [34] —, *Possibility theory- An approach to computerized processing of uncertainty*. Plenum Press, New York, 1988.
- [35] M. Baziz, M. Boughanem, H. Prade, and G. Pasi, *A fuzzy logic approach to information retrieval using a ontology-based representation of documents*. in *Fuzzy logic and the Semantic Web*, Elsevier, 2006, pp. 363–377.



**Sebastien Destercke** Sebastien received the Ph.D. degree in computer science from Universit Paul Sabatier, in Toulouse (France) in 2008. He is currently a researcher with the French agricultural research centre working for international development. His main research interests are in the field of decision making and uncertainty modeling and treatment (learning, propagation, fusion, independence modeling) with imprecise probabilistic models.



**Patrice Buche** Patrice Buche received the PhD degree in computer science from the University of Rennes, France, in 1990. He has been a research engineer with INRA, Agricultural Research Institute since 2002 and assistant professor with AgroParisTech, Paris since 1992. His research works focus on data integration from heterogeneous sources and fuzzy querying in structured and weakly structured databases with applied projects in food safety and biodegradable packaging design.



**Brigitte Charnomordic** Brigitte Charnomordic received the PhD degree in Physics (Quantum Mechanics) from the University of Lyon in 1976. She is currently a research engineer with INRA, Agricultural Research Institute, Department of Applied Mathematics and Informatics. Her research interests are focused on fuzzy modelling, knowledge induction, system interpretability and learning from data, with applied projects in agriculture and environment. Her research work is accompanied by an important software development, in particular

FisPro, open source software for fuzzy inference system design and optimization.