Aide à la décision sémantique pour la diffusion d'informations

Amandine Bellenger

Institut National des Sciences Appliquées de Rouen
Laboratoire d'Informatique de traitement de l'Information et des Systèmes

PHD THESIS

Speciality : Computer Science

Defended by
Amandine Bellenger

to obtain the title of
PhD of Science of INSA de ROUEN

Semantic Decision Support for Information Fusion Applications

June 3rd, 2013

Jury :

Laurence Cholvy	-	*Research Director, Reviewer*	-	ONERA
Arnaud Martin	-	*Professor, Reviewer*	-	Université de Rennes
Sebastian Rudolph	-	*Professor, Reviewer*	-	University of Dresden
Paulo Costa	-	*Associate Professor, Examiner*	-	George Mason University
Habib Abdulrab	-	*Professor, PhD Director*	-	INSA de Rouen
Jean-Philippe Kotowicz	-	*Associate Professor, PhD Advisor*	-	INSA de Rouen
Xavier Lerouvreur	-	*Fusion Expert, Invited*	-	CASSIDIAN
Stephan Brunessaux	-	*Head of IPCC Department, Invited*	-	CASSIDIAN

"Knowing ignorance is strength.
Ignoring knowledge is sickness."

Ancient Chinese philosopher Lao Tsu,
in book Tao Te Ching (ca. 600 b.c.)

Remerciements

C'est avec une sincère reconnaissance que je tiens à remercier toutes les personnes ayant pu contribuer au bon déroulement de cette thèse.

Je souhaiterais commencer par remercier Sylvain et Stéphan qui m'ont permis d'effectuer un stage d'Ingénieur au sein de Cassidian, qui a débouché ensuite sur ce projet de thèse Cifre. Ainsi, Sylvain a pu m'entraîner dans des thématiques de recherches novatrices qui ont été à l'origine des idées développées au cours de la thèse. Bien évidemment, je tiens à remercier Habib et Jean-Philippe pour leur encadrement, leur soutien et leurs encouragements durant ces trois années. Merci à Xavier pour son support technique et ses relectures rigoureuses spécialement durant cette dernière année de thèse.

J'ai une pensée toute particulière à l'intention des doctorants (ou pour certains maintenant docteurs!) de l'équipe IPCC qui ont toujours su partager leurs conseils, expériences et bonne humeur. Plus généralement, je remercie chaleureusement tous les membres de l'équipe ainsi que ceux du laboratoire LITIS à l'INSA de Rouen.

Merci à Sebastian Rudolph pour s'être intéressé à mes travaux et pour m'avoir prodigué de judicieuses remarques lors de mon séjour à Karlsruhe. *Vielen Dank!*
Je remercie également Arnaud Martin, Laurence Cholvy et Paulo Costa d'avoir accepté d'être membres du jury de thèse.

Enfin, plus récemment dans la chronologie, je suis très reconnaissante à Stéphan et à l'ensemble de la hiérarchie pour m'avoir témoigné leur confiance en me maintenant au sein de l'équipe depuis le 1^{er} Février dernier.

Enfin, d'un point de vue plus personnel, je souhaiterais également remercier mes proches, pour m'avoir témoigné leur soutien.

Contents

Part I

Preamble

Chapter 1
Introduction

Contents

Fusing semantic beliefs is part of everyday life for any human being. It all starts from the human process to observe our environment (through our sight, smell, hearing, etc.). Then, we - as human - use a cognitive process, taking into account our *a priori* knowledge and memory, to aggregate all these information in a continuous manner so as to understand what is going on around us. Understanding our environment permits us to take decisions and thus act.

A simple example such as a car driver illustrates this point. One adapts his driving according to the traffic, to its dashboard or even to possible weird sounds of the motor. In the same area, a pedestrian in the street has to pay a lot of attention to his environment before crossing the street. A myriad of such situations can be found when one plays sport, music, cooking and so on.

In all the above use cases, the situation assessment can be affected by various factors: external factors (nightfall, fog, rain, etc.) or factors specific to the individual (the training of the person, his attention, his physical condition, etc.). All these imperfect observations lead to uncertainties. These uncertainties are taken into account by our human brain which is able to perform a natural fusion of all these diverse information so as to identify events and act accordingly.

This natural fusion system performed by the human brain is remarkably effective. Therefore, a large amount of techniques developed for the information fusion attempts to reproduce this processing.

However, it is not so straightforward to mimic this human reasoning. Even if information technology has made great advances both on hardware and software, this brings also an increased complexity. Pieces of information are richer and more complex, of different nature and credibility.

This thesis tries to address these issues and is as such part of the artificial intelligence scientific area. Our approach directly relies on the human mimicking and is applied for dynamic system where information are uncertain and heterogeneous. Therefore, systems have not only to take into account numerical degrees of uncertainty but also the semantics of the information.

This chapter aims to lay out the deep motivations, the context of the dissertation and the open problems the thesis intends to address. In the following, we give a brief informal overview of information fusion, relevant aspects of the Defense and Security domain and more particularly of the maritime surveillance context. The open problems in the management of high-level information and uncertainty handling - that raise from the described context - are underlined. This chapter contains then a brief description of the chosen approach. Finally, the contributions of the dissertation have been listed here. At last, the reader can find at the end of the chapter, a rough guide on how is organized the whole dissertation document.

1.1 Motivations and Purposes of Information Fusion

Many scientific communities are interested in the so-called *fusion of information*. These communities find their roots in applications from the medical domain, the environment, air traffic, security and military domain, to quote only a few. Clearly, their commonality resides in the fact that they have to manage real-time dynamic systems with a multitude of data that need to be synthesized into a single operational picture so as to enable a better understanding of what is going on, i.e. awareness of the situation.

In order to reach this goal, dimensionality of data has to be reduced and information quality has to be improved by combining data. The first point (reduction of data dimensionality) refers to information overload. This is a common recurrent issue in nowadays society: today, a huge quantity of messages is produced at a high rate and need to be processed. This trend is quite true for our applications, where this large amount of messages comes from either sensor data (cameras, radars, chemical sensors, etc.), operational information, human intelligence reports, or even information obtained from open sources (internet, papers, radio, TV, etc.). It produces a large amount of multimedia information that need to be automatically processed in order to estimate objects and gain knowledge of the entire domain of interest. As this information is heterogeneous and geographically distributed, getting the information induces gathering information from various sources and then fusing them accordingly to their underlying semantics in order to get a consistent and more informative set of information. Obviously, classic manual processing is here too time consuming and would simply be impossible. The second point concerning the improvement of information quality is reached through an efficient fusion scheme, which should improve confidence and reduce uncertainty in the synthetized information, thanks to the use of complementary sources.

The definition of information fusion within the Defense and Security domain is totally consistent with the above introductory description of information fusion. "Defense and Security" refers both to civil and military contexts. More specifically, they are geared towards the protection of critical infrastructures (military bases, government headquarters, harbours, airports, train stations, etc.), the prevention of piracy and terrorism acts, and persistent surveillance of immense territory such as borders, maritime zones, etc. We point out that in this type of application, there is - as previously mentioned - a real need to prevent cognitive overload and free man from tedious tasks of information and data processing and to provide assistance to military commanders, hence letting them concentrate on the final decision process. The improvement of information quality is also very important, due to countermeasures or non-reliability of the sources in adverse environmental conditions. Moreover, decision-makers at all levels and all types of organisations need timely and accurate awareness of the situation in their respective area of responsibility as well as prediction of likely intentions of the participants. As a result, automatic processes have to support human users by affording them possible current and predicted future states of the environment, including global situations and possible threats. This type of applications are being developed by Cassidian - an EADS company, which has funded this PhD thesis through a CIFRE[1] contract.

Recently, Cassidian particularly focuses on maritime situation awareness applications. Indeed, the maritime domain presents a growing interest for information fusion applications, since 90% of all international trade is carried out by sea and it represents an important theatre for suspicious activities, such as terrorism, smuggling activities, and illegal immigration. Therefore, there is an increasing need to combat those illegal activities. For instance, detection of unusual vessel activities is considered as an important objective. Here again the sources of information are multiple. At least, we can mention:

- the watchmen in semaphores/watchtowers located along the coasts;
- physical sensor systems such as punctual radars, satellites, drones and cameras which collect information about ships without their cooperation; on the opposite, we have collaborative reporting systems, which rely on ships cooperation to provide information, such as VMS, AIS and many non-automatic reporting systems (such as radio);
- databases gathering historical ships information (characteristics, maintenance data, voyages);
- and finally other sources of information that come from the collaboration with other data fusion systems and that return their own enrichment of the maritime situation picture.

The observed information concerns ship kinematics, ship identifications (obtained by any means), ship characteristics, maintenance and voyages information, con-

[1] "Convention Industrielle de Formation par la REcherche" in French

textual information (ongoing events, environmental information, maritime related activities, infrastructures, water space management). The actors for the information fusion are the operational centres which gather the observed information and have at disposal means to react on the situation. To enable threat recognition, appropriate situational awareness is needed which implies recognition and identification of the objects present in the scene, their interaction with the environment and their intention on basis of threat hypotheses in order to foresee the situation in the near future.

1.2 Open Problems

Since the beginning of data fusion applications, the scientific literature tends to be numerically based. This one deals with physical and kinematic properties of physical observed entities. Questions such as: "how a radar imagery can be used to determine the velocity of an object?" have been handled over the past years. Much progress have thus been made on these topics. However, less research has been allocated to deal with symbolic and semantic information. Yet, this should provide the ability of a fusion system to capture and reason about global situations, rather than about single distinct objects depicted through numerical vector states. Thus, representation of complex relations between objects, representation of threats that can arise from these behaviours, or representation of intents and goals are made possible through qualitative information. Moreover, the explicit semantic of this symbolic information makes possible to use enhanced reasoning tools. Here, questions such as "how an open source report on geopolitical context can be used to infer and predict the behaviour of a group of activists?" wish to be handled. We talk here more about information and knowledge rather than simple raw data.

This new-found interest in information fusion may reflect changes within the Defense and Security context. As a matter of fact, there is a transition from a focus on traditional military problems to asymmetrical threat problems. These are more difficult to reveal, since they take place during the course of normal life. For example, during maritime missions in littoral environments, acts of piracy, drug trafficking and other threatening events become obscured in the crowd of everyday fisheries, cargo traders, ferries and pleasure cruises. The hostile intent of objects is therefore not always easy to determine because of its ability to cloak and hide among the regular vessel traffic.

Information fusion systems lack capabilities that enable a deep, semantic modeling of the domain and drawing conclusions on top of it. This research area is known as high-level information fusion. After numerical based research, the fusion community is for a decade, and in the future, turned towards semantic fusion.

In order to handle semantic symbolic information, the Semantic Web technologies are standing out. They come from the new generation of the Web - the one which intends to make it possible for any software agents to use the content of the Web. These technologies even promising are still rarely used in information fusion applications. Thus, a still open research challenge is to build global systems that support high-level information fusion based on Semantic Web technologies.

An additional problem is that uncertainty is of major concern in information fusion applications. Observations received from sensors are sullied by uncertainty, due to operational conditions of observations (meteorological, luminosity, etc.). Observations may also be uncertain due to doubtful reliability of the sources. An example from the maritime domain can be related to the AIS messages. It is underlined that AIS (Automatic Identification System) data, which is available in principle for all large vessels, can be manipulated and its reliability depends on the willingness of the crew to collaborate, since AIS is a self-reporting communication system. Finally, even by combining certain information from different sources, some inconsistencies may also arise. Uncertainty can also naturally occur in the fusion system itself, when processes are not deterministic. Although there are some indications which suggest that the Semantic Web technologies have been designed to capture a minimum of the uncertainty inherently presents in the knowledge, these do not allow quantifying that uncertainty. As such, **one open major challenge is to adapt these Semantic Web technologies to the context of uncertainty representation and reasoning for the stake of information fusion**.

As a result, automatic information fusion processes have to support human users by affording them as well as possible current and predicted future states of the environment (situations and threat), with likelihood or plausibility tags assigned to them.

1.3 Overview of our Approach

Globally, an information fusion system takes in input all available information and provides mathematical and logical tools so as to obtain an inferred and more complete representation of the situation in order to present it to the final decision-maker. The core of our approach resides in the "mathematical and logical tools" layer of the information fusion workflow, as depicted on figure 1.1.

More specifically, our approach is twofold. On one hand, the approach consists in proposing adequate solutions for the management of the semantics for information fusion applications. On the other hand, the approach aims to provide a model and reasoning tools for the management of uncertainty in such an environment.

According to the state-of-the-art, we believe that semantic web technologies and more particularly OWL ontologies are well suited for representing information within information fusion applications. We have thus opted for the OWL2 language to describe the domain of knowledge related to a situation observed by the sources of an information fusion system. More particularly, we represent terminology of our domain through classes, properties and logical axioms, whereas the observations from different sources are asserted through instances of classes and properties.

However, we have noticed in the meantime that these technologies do not inherently support the management of uncertainty. We have thus studied different math-

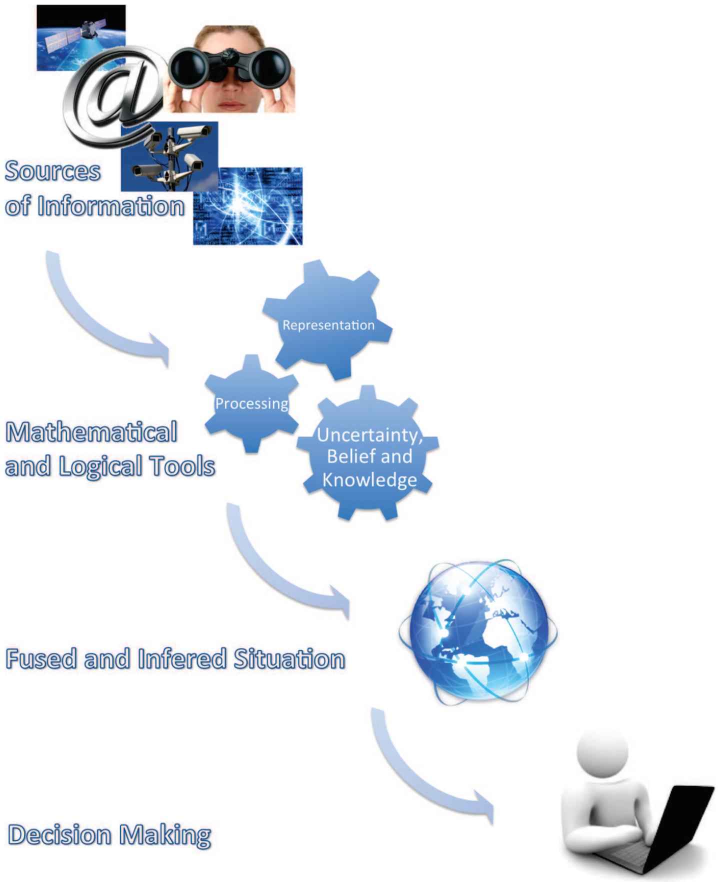

Figure 1.1: General semantic decision-support workflow.

ematical theories tackling uncertainty. This survey has outlined that the Dempster-Shafer theory is particularly adequate for the context and issues of information fusion applications. It is a generalization of the probabilities and is well suited to information fusion and decision-support when information in addition to being unsure, may also be incomplete and imprecise. This theory permits the combination of information coming from different sources and to decide which hypothesis is the best one.

We need in our applications to represent and reason on uncertain observations and more generally on uncertain statements about a given situation. Thus, we believe that the representation of uncertainty has to be made at the instance level of the ontology. We consider that the terminology of the domain ontology is deterministic (no uncertainty is involved). This proposition dissociates us from other existing approaches. Uncertain instances can be embodied by either uncertain instances of classes (i.e. individuals) or instances of properties of an ontology. Moreover, instances can themselves be related to other properties or individuals. In fact, an uncertain instance is a description of the phenomenon that can be perceived / understood differently depending on the source. For each uncertain instance, we decide to associate a degree of belief and its reporting source. Moreover, the sum of the degrees of belief assigned by one source for one given phenomenon should be equal to one - as it is proposed by the theory of Dempster-Shafer. Our approach is based on the expression "*semantic belief*" that we introduced. It refers to one or several uncertain candidate instances - described semantically through a domain ontology -, with its degree of belief and reporting source.

We have chosen to fix the semantics of this representation of uncertainty through an ontology, that we called the *DS-Ontology*. It can thus be seen as a meta-ontology, since it specifies the knowledge-structuring construct for uncertain instances regarding any domain of knowledge (e.g. the maritime domain, the medical domain, etc.).

This knowledge representation step is really important, since it permits to gather and share the information in a consistent and sound way, however the representation alone does not provide us with any added-value on the information itself. Therefore, we decide to propose an innovative way to fuse all these semantic beliefs and thus be able to decide which instance best holds. We decide naturally to rely on the fusion process and the decision-support proposed by the Dempster-Shafer theory.

However, the important distinction with classical examples of the Dempster-Shafer theory is that we deal here with semantic beliefs that may have semantic implicit dependencies between them, in comparison to "raw beliefs" that are already pre-processed and where dependencies are explicit in the set theory.

For example, the belief hypotheses: `car`, `truck` and `land vehicle` are semantically related, since we can intuitively say - as humans - that the belief in a land vehicle has no contradictory statements with the fact that it can be a car or a truck. Taking a second example, we can say that the belief hypotheses: `car` and `truck` are semantically more related than those beliefs with the belief: `aircraft`. Therefore, we define - what we called - *semantic set operators*. Namely: *semantic inclusion* and *semantic intersection*. They enable to automatically discover the semantic relations between our beliefs. These relations are made explicit in the set theory.

We have then described a specific process to refine a frame of semantic beliefs into a frame compliant with the Dempster-Shafer theory, called the frame of discernment. This process relies on the above definitions of the semantic set operators.

In order to understand concretely the need of the semantic set operators in this process, let's take the same example beliefs as previously. One source states its beliefs on `land vehicle` and `aircraft` whereas a second source says that it may rather be a `car` or a `truck`. With our newly introduced operators, one can say that the two sources are not in total conflict, and in consequence that a decision process is possible. Generally speaking, taking into account only the labels of the hypotheses leads to false results, by increasing the amount of conflict and under-estimating credibility and plausibility of the beliefs of interest.

In our process, the dynamic created semantic beliefs are automatically computed into an adequate mathematical model for the Belief Functions Theory. This specific process is embodied by the "mathematical and logical tools" level in the summary diagram of figure 1.2. It finally permits to apply the fusion process and the decision-support proposed by the Dempster-Shafer theory.

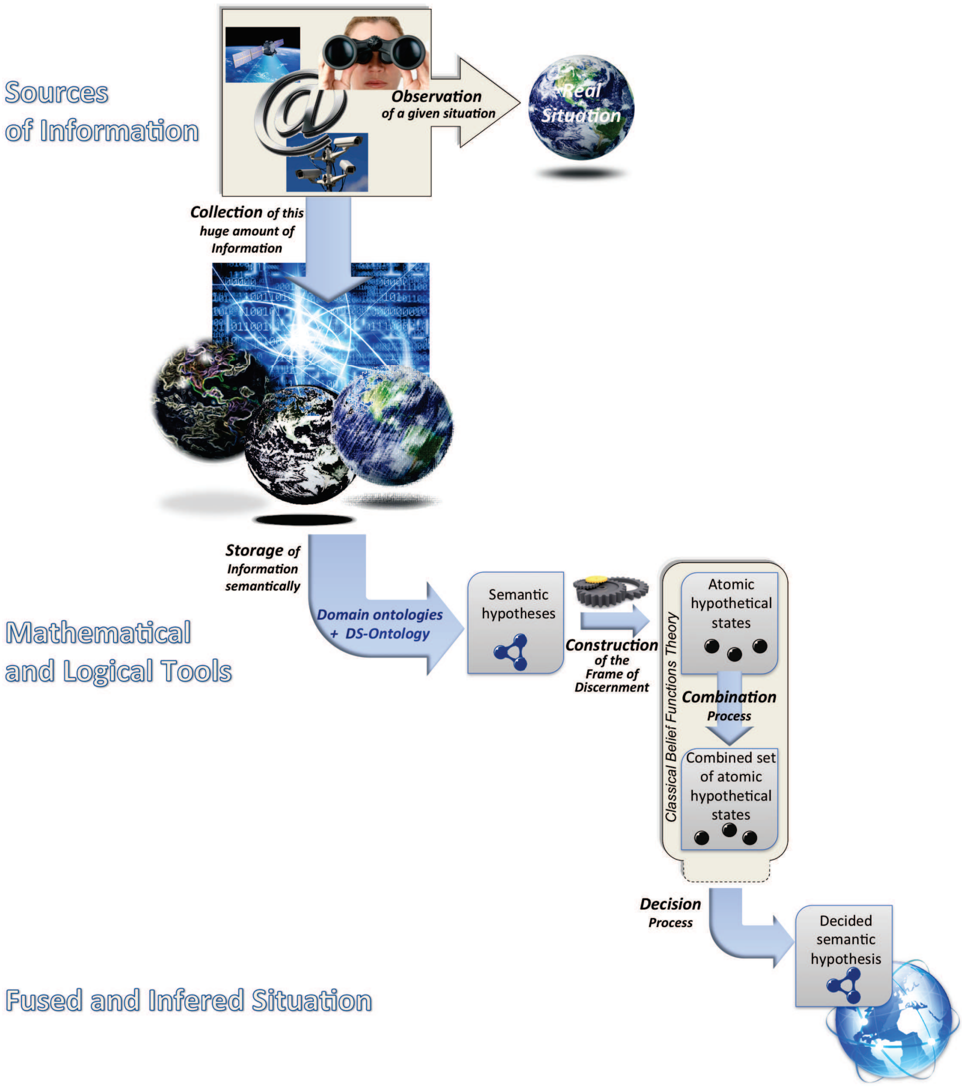

Figure 1.2: Detailled semantic decision-support workflow.

1.4 Specificities and Innovative Aspects of our Approach

The dissertation efforts contribute to advances of computer science in the areas of knowledge representation and uncertain information modeling, and advances of modern software engineering in the area of combining uncertain semantic information. Principally, our contribution consists of the following points:

- A theoretical and consistent framework for representing and reasoning on semantic beliefs.
  - The *DS-Ontology*.

    An innovative modeling contribution is the proposal of the *DS-Ontology*, an ontology dedicated to the representation of uncertainty in any domain

ontology. It is a meta ontology that describes the available annotations to instantiate uncertain statements.

– Definition of Semantic Set Operators.

  This dissertation offers a formal definition of *semantic set operators* we have introduced. It consists namely of: *semantic inclusion* and *semantic intersection*. They enable us to discover the semantic relations between instances of an ontology. These relations are made explicit in the set theory. Algorithms to compute these operators are proposed in this document. The definitions of these operators have been validated by experimentations performed on a pool of test-persons.

– Automatic construction of a frame of discernment in the Evidential theory.

  Defining the frame of discernment is not a straightforward task. There might be no clear boundary between the different hypotheses. By taking into account the intrinsic meaning of candidate instances, we permit a decision-support system to reason not only on predefined labels of exclusives hypotheses but on any hypotheses that are modelled semantically through an ontology. Our semantic decision-support system is indeed able to directly formulate a frame of discernment by taking into account the semantic of hypotheses and enables an adequate combination and decision process. Thus, sources do not have to care about the different levels of granularity of the hypotheses on which they assigned a degree of belief. This process can be done intuitively by a human brain but the interest is here that it is automatically computed. This is actually appreciable when having high dynamical environments with systems that may be not aware of the domain of interest modelled by a given ontology.

• The semantic layer of the *FusionLab* platform.

  A technical contribution has been the creation of a set of OWL ontologies - called the *FusionLabOntologies* - that supports a high-level information fusion system in its semantic description of a given situation. This set of ontologies is well organized around a *FusionLabUpper*, *Core* and application domain ontologies in order to provide high maintainability, interoperability and adaptability to a service oriented information fusion platform - called the *FusionLab* platform. They are expressed in OWL2, a World Wide Web Consortium (W3C) standard[2]. Within the platform, ontologies can be compared as semantic glue between the services that enable them to exchange/share and process the information in the same and consistent way. The *FusionLabUpper* ontology can be seen as a generic language for services interface and is based on the well-established JC3IEDM conceptual military and security NATO model. The *FusionLabCore* deepens the high-level concepts of the *FusionLabUpper*, such

[2] http://www.w3.org/TR/owl2-overview/

as time, space, object taxonomy, reporting data, etc. into a detailed description. This partition avoids too much cross-dependency in the model and permits to maximize the usage of already existing standards to promote interoperability. This information is understood by all services in order to perform their processing. The last level is composed of application domain ontologies, as for example the *FusionLabMaritime* ontology. They rely on the *FusionLabUpper* and *Core* ontologies but are relative to a particular instantiation of the framework concerning a particular application.

As such, an important innovative characteristic of the *FusionLab* is to enable semantic knowledge processing on top of only kinematic data. It uses "artificial intelligence" correlation functionalities in addition to standard statistical based functionalities. The *FusionLab* is intended to be instantiated for different operational needs: ground, maritime, air or joint levels but also security domains such as border surveillance, critical infrastructure security or camp protection. The idea is to propose generic reasoning tools and to adapt or combine them for specific domain contexts. The *FusionLab* promotes also interoperability by proposing a flexible knowledge model with associated mapping tools.

- A reusable *FusionLab* service.

  The implementation of the fusion and decision algorithms have been encapsulated as a capitalized web service of the *FusionLab* platform. As such, projects based on the *FusionLab* platform can rapidly integrate this module in their whole architecture. The *DS-Ontology* is part of the *FusionLab* ontologies.

- A Protégé plugin for a graphical visualisation.

  A graphical interface has been also developed within the famous *Protégé* ontology editor. It permits to create and edit easily semantic beliefs of a given ontology and to visualize the process and results of the several combination and decision process proposed in the Dempster-Shafer theory.

- Definition and Implementation of a maritime use case.

  The dissertation offers the definition of a use case for the semantic beliefs framework. This use case concerns the maritime surveillance context. It has been funded through the *Seabilla* European research project[3]. In the global maritime surveillance system, we have proposed a service for vessel identification based on our framework. This latter underlines, through expert-systems producing uncertain results, the need to have such a decision-support tool being able to perform in high-dynamical environments.

- Universal multi-domain perspective.

[3] http://www.seabilla.eu

As mentioned above, the semantic beliefs framework can be applied in the *FusionLab* platform, where the domain ontologies are the *FusionLab* ontologies. However, this dissertation offers a framework that is generic and not only restricted to the information fusion of the Defense and Security domain. It can be applied whenever several sources of information are stating their beliefs concerning a particular phenomenon of interest (i.e. they are making hypotheses about this phenomenon) and that at one moment one needs to gather all these beliefs - so as to get a global amount of belief for each given hypothesis and be able to decide which hypothesis represents the best the particular phenomenon we were initially interested in. This phenomenon can be either a physical/abstract object or an event, etc.; in fact, it can be everything that can be represented through instances of an ontology.

As such, many current hot research topics could be interested by this framework. For instance, there could be many applications coming from the Web itself. Fusion of information from web sites is an ubiquitous task on the Web that is still mainly performed by humans. Indeed, Web users (human or software agents) often need to aggregate information from multiple sources on the Web. However, that set of information acquired from multiple sources about the same statement may be inconsistent. Thus, it may be difficult to decide in favor of a single alternative. By weighting sources trust, for example, and by using our framework, an automatic process can combine and propose a decision-support to the user. Other situations in which knowledge on the Semantic Web needs to represent uncertainty range from recommendation (trust and provenance in Web) and extraction/annotation to belief fusion/opinion pooling and healthcare/life sciences.

- Participation to the ETUR Working Group.

  One of our contribution lays on our participation to the ETUR WG[4]. ETUR stands for Evaluation of Techniques for Uncertainty Representation. This group has been created in 2011 as part of the International Society for Information Fusion (ISIF). The goal of the ETUR WG is to bring together experts, researchers, and practitioners from the Fusion community to leverage the advances and developments in the area of uncertainty representation in order to address the problem of evaluating uncertainty representation and reasoning approaches for High-Level Information Fusion systems. For that reason, we have participated in developing a set of use cases involving information exchange and fusion requiring sophisticated reasoning and inference under uncertainty; and in defining evaluation criteria.

- Publications.

  Major results of the dissertation were presented in proceedings of International conferences.

[4] http://eturwg.c4i.gmu.edu/

[1] Bellenger Amandine, Gatepaille Sylvain. *Uncertainty in Ontologies: Dempster-Shafer Theory for Data Fusion Applications.* In Workshop on Theory of Belief Functions, Brest, France, April 2010.

[2] Bellenger Amandine, Lerouvreur Xavier, Gatepaille Sylvain, Abdulrab Habib and Kotowicz Jean-Philippe. *An Information Fusion Semantic and Service Enablement Platform: the FusionLab Approach.* In the 14th International Conference on Information Fusion, Chicago, USA, July 2011.

[3] Bellenger Amandine, Gatepaille Sylvain, Abdulrab Habib and Kotowicz Jean-Philippe. *An Evidential Approach for Modeling and Reasoning on Uncertainty in Semantic Fusion Applications.* In the 7th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW), Bonn, Germany, October 2011.

[4] Bellenger Amandine, Lerouvreur Xavier, Abdulrab Habib and Kotowicz Jean-Philippe. *Semantic Beliefs Fusion.* In the 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), Catania, Italy, July 2012.

Finally, a patent has also been submitted:

A. Bellenger, S. Gatepaille and H. Abdulrab. *Procédé permettant la fusion de croyances sémantiques* French, PCT/FR12/000140, April 2012.

1.5 Outline of the Document

We now give a brief guide to this thesis.

Following this introduction, Part II of this document reviews the literature, so that the novice reader has the key concepts in hand to understand the challenges and the refined goals we will progressively set.

First, we highlight the characteristics of high-level information fusion regarding data fusion in general and by opposition to low levels. For that purpose, we overview, in Chapter II, existing reference models that have been set up over the years in order to facilitate discussions and exchanges between engineers, researchers, by establishing common and clear definitions, concepts and terms, within the global range of fusion functions and levels. This will lead us, in section 2.3, to the discussion of the remaining key issues of information fusion, that we have particularly focused on. Namely, it is largely agreed in the community that the knowledge representation of a given situation and the management of uncertainty are very important features on which the community has to work on in order to improve information fusion applications.

Having stated that, in the two following chapters: Chapter 3 and 4, we naturally focus on uncertain mathematical theories and knowledge representation means, respectively. In the meantime, we present our key candidate technologies: the Evidential theory and ontologies, respectively. Finally, Chapter 5 will be dedicated to some reviews on existing works mixing ontologies and uncertain mathematical theories. Even if many works are here recapped, we see that only a very few are specific to our context of semantic fusion applications.

Once having exploring the background material of semantic technologies and management of uncertainty for information fusion, we present our theoretical framework in Part III of this document. It presents the core theoretical contributions of this thesis. It is divided into three chapters: the representation of semantic uncertain beliefs in ontologies (Chapter 6), the introduction of semantic set operators (Chapter 7) and the projection to a consistent modeling enabling the evidential reasoning process (Chapter 8).

The fourth part of this document is devoted to the practical implementation of this theoretical frame. The *FusionLab* platform as well as the graphical interface developed through the Protégé editor are described in Chapter 9. The experimentations, concerning namely the human feedback on the semantic set operators and a particular industrial project are presented in Chapter 10. This chapter also presents an analysis of the ETUR evaluation set of criteria regarding our theoretical framework, implementation and particular maritime use case.

Part IV of this document concludes the thesis with a review of contributions, summary of outstanding issues, and suggestions for future research in the area.

Part II

Review of the Literature

Chapter 2

High-Level Information Fusion

Contents

Many scientific communities are interested in information fusion. These communities range from the medical domain, the environment, air traffic, to security and military domains of applications. Each has different objectives but clearly, their commonality resides in the fact that a multitude of data needs to be synthesized into a single operating picture so as to enable a better understanding of what is going on.

Since this research work takes place in an industrial context of defense and security applications, we emphasize first of all the operational point of view by introducing an intuitive definition of information fusion.

Then, a state-of-the-art concerning formal existing definitions of information fusion is presented.

The two last sections will finally lead us to conclude on the major and remaining issues of information fusion. These issues will actually be the central focus of this whole dissertation.

2.1 Defense and Security Operational Motivation

From a pessimistic view, this 21st century will bring many challenges to overcome. According to [Alberts 2010], it is to be expected: a variety of humanitarian disasters (earthquakes, floods, tsunamis), failed states, instability, global terrorism, intractable conflicts, pandemics, economic crises, and poverty. This implies that Defense & Security departments are geared towards the protection of critical infrastructures (military bases, Government Headquarters, Harbours and airports or train stations and the like), the prevention of piracy and terrorism acts, and persistent surveillance of immense territories such as borders, maritime zones, etc.

As such, it is important to collect observations from the domain of interest. These observations are made by geographically distributed sources. Sources of information can be either sensor data (cameras, radars, chemical sensors, etc.), operational information, human intelligence reports, or even information obtained from open sources (internet, papers, radio, TV, etc). The following figure illustrates the entanglement of sources used in Defense & Security applications.

Figure 2.1: Context of Defense & Security applications.

These observations are made so that users can gain knowledge on the domain of interest. Among these people, some are called the "decision-makers". At all levels of the hierarchy, as recalled in [Jousselme 2011], they need to rapidly develop situational awareness (e.g. understand how a situation has occurred and is expected to evolve) and share understandings of the operational environment. They also need to plan operations and monitor the situation to check the execution of the plans.

In order to help the decision-makers in their task, computerized systems that are intended to interact with and complement a human decision-maker have been developed. They are so-called: Decision Support Systems (DSS). An ideal DSS from the point of view of [Jousselme 2011] would be the one that provides the information needed by the human decision-maker, as opposed to raw data; it would be able to be controlled effortlessly by the human (transparent to the user); it would efficiently complement the cognitive power of the human mind; and support a wide variety of problem solving strategies (from instinctive reactions to knowledge based reasoning).

In background, information fusion algorithms support DSS. An information fusion system is thus often used as a synonym of a DSS. Among the many reasons for interest in this technology, data and information fusion:

- reduces data dimensionality,
- enables a better understanding of the situation,
- enables the decision-maker to cope with the complexity and tempo of operations in moderns, dynamic operational theatres.

The first point refers to information overload. Today, sensors produce a huge quantity of messages at a high rate that need to be processed. Obviously, classic manual processing is much too time consuming and would simply be impossible. This is a major problem in the military context. Lieutenant General Deptula, USAF deputy chief of staff for intelligence, surveillance, and reconnaissance said [DefenceWeb 2011]: "We are going to find ourselves in the not too distant future swimming in sensors and drowning in data". Information fusion systems are thus deployed so as to reduce substantively the amount of data that need to be presented to operators within the command and control rooms.

The second point is achieved through an efficient fusion scheme, which would improve confidence in decisions due to the use of complementary information (e.g. silhouette of objects from visible image, active/non-active status from Infra-Red image, speed and range from radar, etc.). It would also improve performance against countermeasures (it is very hard to hide an object in all possible wave-bands) and performance in adverse environmental conditions. Typically smoke or fog cause bad visible contrast and some weather conditions (rain) cause low thermal contrast (Infra Red imaging), combining both types of sensors should give better overall performance.

The third point refers to the fact that commanders at all levels and all types of organisations need timely and accurate awareness of the situation in their respective area of responsibility as well as prediction of likely intentions of the participants. The underlying issues are the distributed aspects of information: the required data, information and knowledge and the services, applications, tools and products originate from various systems. The level of responsibility of a decision-maker refers to the information release constraints due to different security levels (the "need to know" protocol).

In the Defense & Security domain, the welcome operational output of information fusion is a so-called COP: Common Operational Picture. It is a single identical display of relevant (operational) information (e.g. position of own troops and enemy troops, position and status of important infrastructure such as bridges, roads, etc.) shared by more than one Command. It facilitates collaborative planning and assists all echelons to achieve situational awareness. Previously, headquarters prepared maps with various symbols to show the locations of friendly and enemy troops and other relevant information. From now, the COP is the automatic output of an information fusion system.

2.2 Attempts for a Formal Description of Information Fusion

The above intuitive and operational definition of information fusion has been reviewed by the scientific community to propose a formalism which would enable a precise characterisation of the information fusion processes. As information fusion is still in its infancy stage, there are not yet a rigorous conceptual model but rather vaguely perceived concepts and principles. That is the reason why, since the 1980s, many models have been developed. They all underline their own vision of the main principles of information fusion. Each of them creates its own additional value as it provides particular insights in this field.

Actually, we present in this section models that have gained significant interest within the community. Our purpose is to give the reader a comprehensive view of the research context. As a consequence, it should provide a basis to understand the current fundamental challenges of information fusion which will be the subject of the next section.

2.2.1 JDL Data Fusion Model

This model is widely used for categorizing data fusion related functions. It is certainly the oldest one, since its origins are found in the early 1986 [Kessler 1992]. The Data Fusion model was developed by the Joint Directors of Laboratories (JDL) Data Fusion Group, a US DoD government committee overseeing US defense technology.

The goal of the JDL model is to facilitate understanding, communication and coordination among acquisition managers, theoreticians, designers, evaluators, and users of data fusion technology to permit cost-effect system design, development, and operation.

The more precise stated purpose for that model, according to [Steinberg 2004] and its subsequent revisions (1998, 2004), has been:

- to categorize different types of fusion processes;

- to provide a common frame of reference for fusion discussions ;

- to facilitate understanding of the types of problems which data fusion is applicable for;

- to codify the commonality among problems;

- to ease the extension of previous solutions;

- to provide a framework for investment in automation.

One has to note that this model is above all a functional model and not a process model or an architectural paradigm. It should be underlined that the following levels are not necessarily orderly processed. Hereafter is the current recommended

JDL model which has been lastly revised in 2004 by Steinberg and Bowman in [Steinberg 2004].

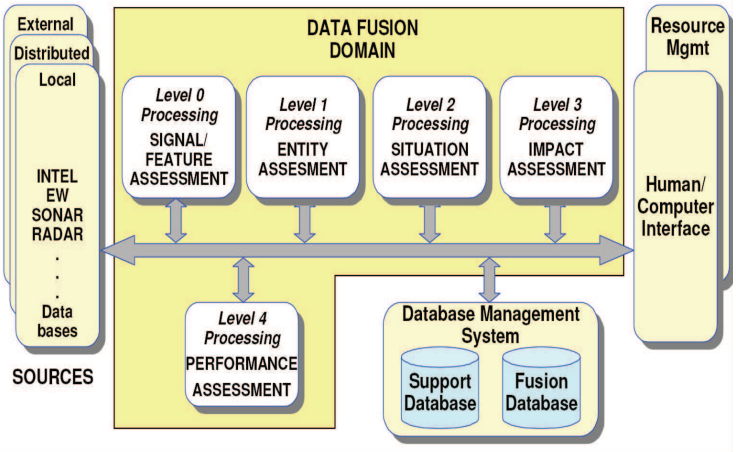

Figure 2.2: Recommended revised data fusion model of 2004 (from [Steinberg 2004]).

The model is comprised of five levels, which are defined as follows:

- **Level 0: Signal/feature assessment.** At this level signal- or object-observable states are estimated and predicted on the basis of pixel/signal-level data association and characterization. Signals and features may be defined as patterns that are inferred from observations or measurements. These may be static or dynamic and may have locatable or causal origins (e.g. an emitter, a weather front, etc.).

- **Level 1: Entity assessment.** This level performs the estimation of entity parametric and attributive states. It is an iterative process of fusing data to determine the identity and other attributes (e.g. position, velocity, type) of entities and also to build tracks to represent their chronological evolution. The term entity refers here to a distinct object. For example, an entity can be a single military truck. A track is usually directly based on detections of an entity.

- **Level 2: Situation assessment.** This level estimates the structure relationships among entities and their implications for the states of the related entities. It is an iterative process of fusing the spatial and temporal relationships between entities to group them together and form an abstracted interpretation of the patterns involved in a specific situation. A group of entities is for example a trucks aggregate where each single track has already been assessed in JDL level 1. Level 2 differs from level 1 in its emphasis on relations among objects rather than on the characteristics of single objects. While the number of such characteristics grows linearly with the number of objects considered by an information fusion system, this cannot be said about the number of possible relations, which can grow exponentially.

- **Level 3: Impact assessment.** This one is in charge of the estimation of the utility/cost of signal, entity, or situation states, including predicted utility/cost given a system's alternative courses of action. In other words, it is also an iterative process of estimating and predicting the utility or cost of an estimated world state (i.e. situation) to a user objective (e.g. mission). An example of a level 3 type of estimation problem is the prediction of the impact of the estimated current situational state on the mission utility. Another example could be to estimate the threat for a current situation (this level was indeed previously called threat assessment).

- **Level 4: Performance assessment.** This level deals with a system self-estimation of its performance as compared to desired states and measures of effectiveness. It could be noted that this level was previously referred to as "Process refinement". Level 4 interacts with all the other levels.

2.2.2 Endsley's Model for Situation Awareness

Endsley's model stands out as the reference for most works done in Situation Awareness (SAW). As defined by Mica Endsley in [Endsley 1995, Endsley 2000], SAW is "the perception of the elements of the environment, the comprehension of their meaning (understanding), and the projection (prediction) of their status in order to enable decision superiority". Endsley considers Situation Awareness as a state of knowledge that results from a process. This process, which may vary widely among individuals and contexts, is referred to as Situation Assessment, or as the process of achieving, acquiring, or maintaining Situation Awareness.

Her model, depicted in figure 2.3, has two main parts: the core Situation Awareness portion and the various factors affecting Situation Awareness. The core portion follows Endsley's proposition that Situation Awareness has three levels of mental representation:

- **Perception of the elements in the environment.** It provides information about the presence, characteristics, and activities of relevant elements in the environment. These perceived elements are actually only a subset of elements present in the environment. The set is structured into meaningful events situated in time and space, creating an active working memory, which is the basis for situation awareness.

- **Comprehension of the current situation.** It encompasses the combination, interpretation, storage, and retention of information, yielding an organized representation of the current situation by determining the significance of objects and events. This aspect is a synthesis of the previous one disjointed elements. It thus provides an organized picture of the elements with a comprehension of the significance of objects and events. This comprehension requires that the problem of meaning be tackled head on. Meaning must be considered both in the sense of subjective interpretation and in the sense of objective significance or importance. Therefore, mental models (complex schemata representing a given system) are the basis of this SAW aspect.

- **Projection of future status.** It involves forecasting future events which encompasses the highest level of SAW. This ability characterizes the decision-makers who have the highest level of understanding of the situation. Through knowledge of the current status and the dynamics of the situations elements and the comprehension of the situation, one can achieve predictions about future states in a near future.

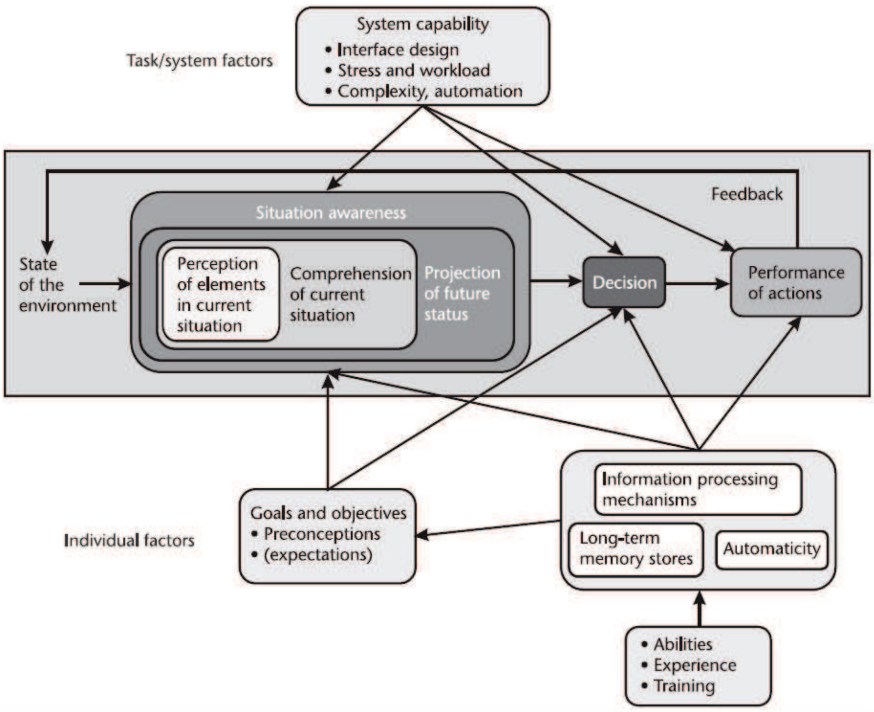

Figure 2.3: Endsley's situation awareness model (from [Endsley 1995]).

McGuiness and Foy in [McGuinness 2000] state that Perception is the attempt to answer the question "What are the current facts?"; Comprehension asks "What is actually going on?" and Projection asks "What is most likely to happen if...?".

Looking at the two above models, it clearly seems that, these two models are dual. The notion of *assessment* in the JDL model refers to a process, whereas *awareness* is a result of it. Specifically, *awareness* is a mental state (of a human) that can be obtained by an *assessment* process. Thus, the JDL model is more about functions, whereas the Endsley's model is more about results. In some manner, the JDL seems more adequate to numerical data and information, while the Endsley's model is nearer from the human reasoning.

Other models have been proposed to formalize the definition of information fusion. However, they are usually relying on the JDL, on the Endsley's model or a combination of both. Among them, the λJDL [Price 2004] or the Salerno's model [Salerno 2007] can be mentioned.

2.2.3 Information Fusion versus Sensor Fusion

The distinction between sensor fusion and information fusion is equivalent to the one that the data fusion community tends to draw between low and high-level fusion processing. As a matter of fact, the term sensor fusion typically applies to levels 0 and 1 of the JDL model, whereas the term information fusion is often used to refer to levels 2, 3 and related parts of level 4. The main conceptual differences have been summarized by Waltz and Llinas in [Waltz 1990], and have been graphically represented by [Blasch 2010](see figure 2.4).

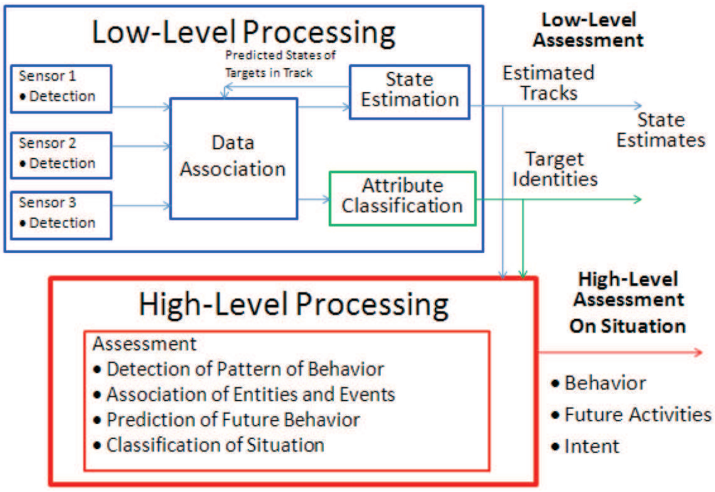

Figure 2.4: Low Levels versus High-Level Processing (from [Blasch 2010]).

On one hand, sensor fusion involves objects having properties, which are usually measurable. Indeed, its goal is largely to infer and predict physical, kinematic properties of physical entities. For instance, radar can be used to infer existence and velocity of an object. As such, mathematical models and algorithms are here closely tied to physics and are constrained by the laws of physics. Therefore, literature dealing with low-levels data fusion tends to be numerically based. Much progress has been made on these levels during the last years, and are still in course, with major implication of the scientific community.

On the other hand, while low-level processes support target classification, identification, and tracking, high-level processes support situation, impact, and fusion process refinement. As a consequence, high-level fusion processes are complex processes involving many elements and interactions among a wide variety of situation components. According to [Boury-Brisset 2004], high-level processes have the following properties:

- they emphasize on symbolic reasoning rather than numeric reasoning;
- they manipulate both concrete and abstract entities;
- there exist numerous constituency-dependency relationships among objects as well as events and activities of interest.

Comparing to the world of objects handled by low-levels data fusion, information fusion takes the world to be a world of facts, where facts involve the application of relations between objects. Information fusion represents the world by using symbols to make claims about the world, which purport to express facts. More generally, higher-levels rely on the realm of cognitive, social and organizational phenomena. For example, information fusion may be interested in knowing how the platoons are organized into companies and battalions and what their missions are. Here, physical laws seem to be insufficient. This distinction between low and high-level information fusion is dual to the notion of semantic growth of data, information and knowledge, which will be recalled in Chapter 4, section 4.1.2.

Finally, one can note that high-level processing is still nowadays mainly performed by humans. This is largely due to remaining problems that need to be tackled and solved. The following section therefore addresses the list of issues that information fusion still encounters.

2.3 Information Fusion Issues

Yesterday's data fusion problems were largely involved in sensor fusion. Indeed, up until the break of the 21st century, much activity was being spent in sensor fusion especially concerning target recognition and tracking algorithms and in developing algorithms to perform model assessment.

Since the late 1990s, there have been few cumulative updates in the progress of information fusion. Therefore, there are still the same remaining issues and challenges. However, these latter have been through the years refined and more explicitly stated. For instance, during the Fusion'2005 conference [Blasch 2006], a panel discussion was organized with leading experts to elicit and summarize current issues and challenges in information fusion that need to be searched in the next decade. Another more recent initiative was the paper [Toth 2008] which particularly focuses on the challenges of the academic community.

We can, however, wonder why the community has shown so lately interest in these higher levels of fusion? According to [Toth 2008], the implicit underlying assumption of the community is that in order to make significant progress in higher-level information fusion, basic research was needed.

This section will, therefore, point out how these information fusion factors, which are characteristic of most of today's national security concerns, contrast with the classical sensor fusion problems that dominated military concerns of the twentieth century.

Issues such as: *how to represent the knowledge issued from the sensors?* and *how* *to handle and reason about uncertainty?* are crucial questions that are often raised in the community. They especially focus on challenges, which concern namely the knowledge representation and the management of uncertainty.

2.3.1 Knowledge Representation

One key research topic for higher-level information fusion is knowledge representation. As a matter of fact, one needs a sufficiently expressive knowledge representation in order to capture complex behavior characterizing situations. As mentioned in previous section, data is insufficient to enable those complex situations. By handling more complex "objects" (groups, activities, or situations) one could increase overall understanding and permit to deal with situation, impact and threat assessment.

In [Lambert 2003], Lambert identifies a number of important challenges for information fusion. The majority of these challenges are related to the knowledge representation issue. Among them, we can quote the semantic, epistemic and interface challenges. There are here further explained:

- Semantic Challenge: What symbols should be used and how do these symbols acquire meaning? The semantic challenge is said to transcend philosophical, mathematical and computational dimensions. As a matter of fact, a philosophical theory is required to conceptualise the domain of interest; a mathematical theory is required to impose structure on that conceptualisation; and a computational theory is required to bring that conceptualised structure to life. The conceptualisation suggests symbols that might be used. The (formal) mathematical theory prescribes the meaning of those symbols.
- Epistemic Challenge: What information should be represented, and how should it be represented and processed within the machine? While the semantic challenge for information fusion relates to the choice of symbols and what those symbols mean, the epistemic challenge involves the choice of knowledge content and the choice of a symbolic knowledge representation scheme. This requires a modeling framework, a means of capturing the domain knowledge within that framework, and a means of automating that captured knowledge within a machine.
- Interface Challenge: How should people be interfaced to complex symbolic information stored within machines? To be of practical use, knowledge models should be fairly straightforward to implement by engineers. Indeed, as symbolic information fusion matures within the machine, a means is required of interfacing complex information encapsulated in symbolic machine representations to people interacting with those machines.

2.3.2 Managing Uncertainty

Uncertainty is finally one of the very first most important characteristics of the data and information handled by data fusion applications. Indeed, the principal aim of information fusion is to overcome the lack of knowledge we have about a situation. To some extent, if perceptions of the information were including no uncertainty, there would be no need for information fusion applications.

As underlined in [Smets 1997], even if there may or may not be uncertainty in the real world, still the picture we capture from it, which corresponds to the only information we can cope with, never reaches perfection. Thus, information is always imperfect in those systems. As a matter of fact, observations received from sensors are sullied with uncertainty, due to operational conditions of observations (concealment, jamming, dissemination, and the like). Uncertainty can be due to measuring errors made by physical sensors (human or electronic). For example, a radar can observe the speed of a vehicle with an error of + or - 3 km/h, an image interpretation can also lead to the conclusion: "Possibility of existence of a regiment position (x,y,t)". As such, neither devices nor people are error-free observers or reporters. Observations may contain differences in detail which introduce uncertainty into the mix; sometimes they may even be contradictory. Some sources may be unreliable. Inevitably, there will be some potentially duplicate or missing information [Kruger 2008].

As we have seen, uncertainty in low-levels is brought through geospatial, temporal or existential characteristics. But, uncertainty is not the sole preserve of low-level observations.

Uncertainty is also ubiquitous and still remains a major issue in higher level fusion. Here, uncertainty can also be involved in social (relationships, interactions, collaborations), behavioural (intentions, threats, defensive postures), or motivational (needs, desires, wants) domains. But, these variables are harder to define and quantify. This can be particularly the case when addressing uncertainty regarding rare situations that may have not even occurred before (potential threats for instance).

As a conclusion, uncertainty is inherent to data, numerical or symbolic. However, uncertainty also naturally occurs in the fusion process itself. Issues related to uncertainty arise in case the set of information acquired from multiple sources about the same fact is inconsistent, or - more generally - in case that multiple information sources attribute different grades of belief to the same statement. If the user/application is not able to decide in favour of a single alternative (due to insufficient trust in the respective information sources), the aggregated statement resulting from the fusion of multiple statements is typically uncertain. The result needs to reflect and weight the different information inputs appropriately, which typically leads to uncertainty.

Therefore, research is needed on how best to represent and propagate uncertainty in information fusion.

2.4 Conclusion

This chapter has presented the background context of information fusion applications and its current issues, most of which is necessary for understanding later chapters.

Concerning data fusion in general, more than thirty fusion models have attempted over the years to define formal models in great detail. However, no model has become as influential as the JDL one. Actually, the JDL model is surely the most widely used model for categorizing data fusion related functions. Relying on these models, the emphasis has been given on the comparison of information fusion with sensor fusion.

This analysis has underlined major challenges that arise from information fusion. Indeed, as information fusion handles more information and knowledge, rather than raw numerical data, this raises the issue of representing that knowledge: which deals not just with objects, but also with parts and wholes, roles and relationships. Chapter 4 is therefore devoted to propose solutions to handle that first issue thanks to ontologies. An additional, but fundamental issue, concerns the management of uncertainty, since information fusion as sensor fusion is fraught with uncertainty. As a consequence, the next chapter is dedicated to the study of the different mathematical theories tackling uncertainty. However, there are several theories since uncertainty is a large area ranging from empirical belief, ambiguity, inconsistency, vagueness to imprecision. Therefore, next chapter is intended to give first an outlook of the numerous flavours of uncertainty and then recall the different mathematical theories to address uncertainty.

Chapter 3

Uncertainty through the Dempster-Shafer Theory of Evidence

Contents

Uncertainty is an important characteristic of data and information handled by real-world applications. As Chapter 2 has underlined it, this is quite true for information fusion applications. Indeed the observations, our understanding of them, their encoding or the fusion process itself are always sullied with uncertainty.

In this chapter, we introduce what the different origins of uncertainty are at a mathematical level. It ranges from aleatory uncertainty, epistemic uncertainty to inconsistency.

Then, a brief introduction to major existing mathematical uncertainty theories is given. According to an analysis of the specific issues of information fusion applications, we draw the advantages of the theory of evidence. This theory is either called the Evidential theory, the Dempster-Shafer theory or the belief functions theory.

Finally, a more in-depth section is devoted to this theory, and especially focuses on its mathematical formalism, its combination and decision formulas.

3.1 Several Facets of Uncertainty

As point out in [Rogova 2010], there are several criteria to assess information quality. Among them, one can distinguish the availability of the information, its accessibility, relevance, timeliness and finally its uncertainty.

The term "uncertainty" refers to a variety of forms of imperfect knowledge, such as incompleteness, vagueness, randomness, inconsistency and ambiguity. Uncertainty is thus a synonym of imperfection of the data.

In [Dubois 2011], the author distinguishes three different types of uncertainty arising from different origins. He makes the distinction between aleatory, epistemic and inconsistent uncertainty as represented in figure 3.1.

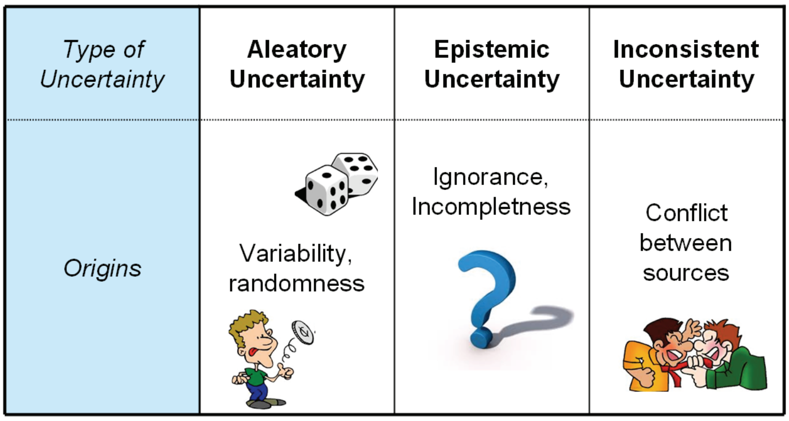

Figure 3.1: Origins of Aleatory, Epistemic, Inconsistent Uncertainty (adapted from [Dubois 2011]).

Aleatory uncertainty is induced by the outcomes of random (repeatable) experiments. It is often referred to as the randomness aspect of an observed phenomenon, since this uncertainty is inherent in the phenomenon. A typical example of randomness is the drawing of a colored ball from an urn. Another classical example is the outcome of the throw of a coin or a dice. Aleatory uncertainty is also induced by the variability of entities in populations. Weather forecast is a good example to show the variability of a phenomenon. It may change every day. The common aspect of aleatory uncertainty is that it can be estimated through statistical data. These estimations are objective, i.e. independent from the observer.

The second kind of uncertainty is the epistemic one. It is due to lack of knowledge. Contrary to the variability characterizing the aleatory uncertainty, the epistemic uncertainty is characterized by the ignorance, or incompleteness. As a matter of fact, because information is often lacking, knowledge about issues of interest is generally not perfect. Examples given in [Dubois 2011] are the inability to distinguish the color of a ball because of color blindness, or the birth date of Brazilian President. Contrary to aleatory uncertainty, here statistics on birth dates of other presidents do not help much. Ignorance can lead also to imprecision, which thus can be classified as epistemic uncertainty. This imprecision can be either sub typed in disjunctive, negative, interval - or fuzzy- valued [Das 2011]. A disjunctive imprecision is embodied by multiple singleton hypotheses where one of them should represent the real phenomenon, but we do not know which one. For example: "the unit type is tank or armored personnel carrier". A negative imprecision is when we express our certainty through a negation, for example: "the unit type is not a tank". Then, an interval-valued imprecision occurs when we deal with numerical data (e.g. "the unit is travelling between 25 and 40km/h"). Finally, fuzzy-valued imprecision appears with natural language expression (e.g. "the unit is moving fast"). The two last subtypes are also often referred to as vague information.

Finally, the third kind of uncertainty is the inconsistency. It is due to conflicting testimonies or reports. Indeed, the more sources, the more likely the inconsistency.

Many research have tried to classify formally these different aspects of uncertainty in order to get a clear and shared specification of the intended meaning of the uncertainty regarding particular scenario. This has been realized through taxonomies (see for example [Rogova 2010]) or through ontologies in [Laskey 2008]. In [Laskey 2008], the URW3-XG (see also section 4.7, or 5.1) proposes a first ontology whose aim is to focus discussion on their use cases. It is not intended to be used for annotating uncertainty in software applications, which is our goal and will be addressed in Chapter 6. The top level of the URW3-XG ontology is shown in Figure 3.2. According to the ontology, uncertainty is associated with sentences that make

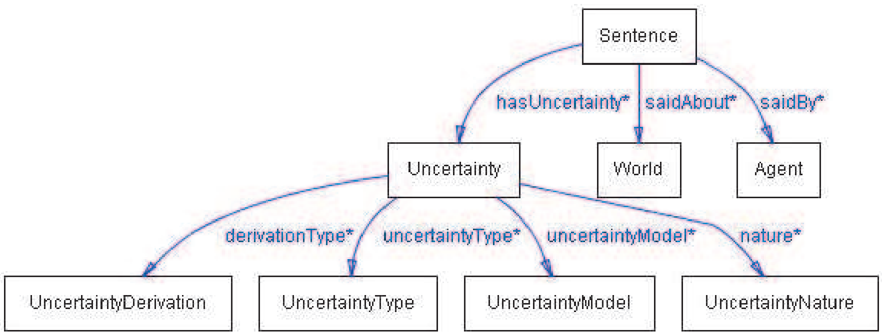

Figure 3.2: Top level of URW3-XG Uncertainty Ontology (from [Laskey 2008]).

assertions about the world, and are asserted by agents (human or computer). The different aspects previously discussed in this section are rephrased through the *UncertaintyDerivation*, *UncertaintyType* and *UncertaintyNature* concepts. The uncertainty derivation may be objective (via a formal, repeatable process) or subjective (judgment or guess). Uncertainty type includes ambiguity, empirical uncertainty (i.e. the fact that an event is satisfied or not), randomness, vagueness, inconsistency and incompleteness. Uncertainty nature includes aleatory or epistemic. Uncertainty models include probability, fuzzy logic, belief functions, rough sets, and other mathematical models for reasoning under uncertainty. It is the purpose of the next section to review these different mathematical models and theories.

Now, let's analyze which particular aspects of uncertainty are needed by information fusion applications. From the analysis of uncertainty management needs of section 2.3.2, the categorization of uncertainty aspects information earlier in this chapter and our application needs, it turns out that, fusion techniques try to manage mostly epistemic uncertainty and inconsistency, which are inherent to the pieces of information provided by the sources. In other words, the aleatory aspect of the information is not a specific challenge on which information fusion needs for the moment to focus.

3.2 Different Mathematical Theories

Before, the probability theory was used to model almost all aspects of uncertainty underlined in the previous section (aleatory, epistemic and inconsistency). However, over the last three decades, new theories have been developed. Indeed, the mathematical community has realized that uncertainty can encompass a large variety of aspects and that the probabilities can not cope with all of them.

Currently, different frameworks exist for reasoning with uncertainty. In this section, we list the most well-known ones: the Probability theory, the Fuzzy-Set theory, the Possibility theory and the Evidential theory. Other theories dealing with uncertainty can be mentioned such as the Imprecise probability theory [Walley 1991] or Rough set theory [Pawlak 1995]. We will not present here the mathematical formalism attached to each theory but rather focus on their purpose and range of action.

As a result, we would like to select one mathematical theory which considers only the epistemic uncertainty, due to lack of knowledge and the inconsistency, due to conflicting testimonies or reports. As we can guess from the title of next section, this analysis will be in favour of the theory of evidence.

3.2.1 Probabilities

Probability domain is surely the most known mathematical theory dealing with uncertainty. In brief, it provides a sound mathematical representation language and formal calculus for rational degrees of belief. This theory relies on solid bases and well-known axioms. Probabilistic reasoning gives us a way of finding out how likely something is the case. More precisely, the roles of probability are twofold [Dubois 2011]. On one hand, probability theory is used for representing aleatory phenomena. The probabilities are thus capturing variability and randomness through repeated observations. Probabilities are considered as objective quantities and interpreted as frequencies or limits of frequencies. On the other hand, Probability theory can be used to represent epistemic uncertainty. Probabilities are here subjective and have to be interpreted as degrees of belief. A belief describes a person's opinion on the occurrence of a singular event. As opposed to frequentist probability, subjective probability that models unreliable evidence is not necessarily related to statistics. Let's recall that often degrees of beliefs in information fusion are provided by experts (see section 5.1.1.3) and thus may be scarce and subjective. However, one major drawback of the Probability theory resides in the requirement of a perfect knowledge of the probabilities, and especially of the *a priori* probabilities. Unfortunately, when knowledge on the problem is imperfect, probabilities cannot be estimated. The representation of information absence is somehow badly taken into account by the theory of probability. Indeed, prior and conditional probabilities need to be specified into probabilistic methods. This requirement often leads in using a symmetry (minimax error) argument to assign prior probabilities to random variables (e.g. assigning 0.5 to binary values for which no information is available about which is more likely). As such, a single probability distribution cannot distinguish between uncertainty due to randomness and uncertainty due to lack of knowledge. It causes some ambiguity.

In our context, we are more interested in subjective probability, which is also often referred as the Bayesian approach. Indeed, in these methods, update of information (modelled through probability distributions) is carried out thanks to the Bayes theorem.

A major advantage of the probability theory is that it provides a powerful graphical model called the Bayesian networks. These are directed acyclic graphs, used to represent a set of random variables and their conditional dependencies.

As the probability theory is the oldest theory to deal with uncertainty and is still the most known, data fusion methods for uncertainty management have been first proposed under the Bayesian approach. Now, data fusion relies generally on confidence measure which includes either probabilistic, possibility or evidential measures.

3.2.2 Fuzzy Set Theory

Fuzzy formalisms allow the representation and the gradual assessment of truth about vague information. The fuzzy set theory can be used in a wide range of domains when information is incomplete or imprecise. The role of fuzzy sets is to model vagueness phenomenon. Fuzzy sets are sets whose elements have degrees of membership. Fuzzy sets were introduced in [Zadeh 1965] as an extension of the classical notion of set. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition - an element either belongs or does not belong to the set. By extension, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described through a membership function valued in the real unit interval [0,1]. As such, the degree of truth is a value between 0 and 1, as in the Probability theory. However, the degree of truth in Fuzzy sets represents a membership in vaguely defined set, whereas a probability represents a likelihood of the membership itself. Hence, fuzzy logic sees the world as continuous instead of binary, while probabilistic logics make a claim about the randomness of the world or the observer's state of certainty. For example, the vague proposition "John is tall" may be more or less true, and it is thus associated with a truth value in [0, 1], depending on the body height of John.

Regarding our aspects of uncertainty of interest, the fuzzy set theory supports indeed epistemic uncertainty, but refers more to vagueness rather than about the veracity of a belief itself. Moreover, it does not manage inconsistency.

3.2.3 Possibility Theory

The theory of possibilities has been introduced in [Zadeh 1978] as a generalization of the Fuzzy sets. The aim was to enable the management of uncertainty, which is not probabilistic in nature. Thus, in the possibility theory, imprecise knowledge and uncertain knowledge can coexist and been handled together. The theory considers situations are more or less possible than others. It does not model a degree of belief or verity, but rather the preference we have for a hypothesis.

Contrary to the probability theory, which associates with every event a unique probability, we now have possibility distributions, each of which associates with every event a unique possibility and a unique necessity.

3.2.4 The Dempster-Shafer Theory

The Dempster-Shafer theory has been proposed by Shafer in 1975 (see [Shafer 1976]) taking support on the work made by Arthur P. Dempster. It is also known as the Evidential theory or the theory of belief functions.

The Dempster-Shafer theory allows the combination of distinct evidence from different sources in order to calculate a global amount of belief for a given hypothesis. Modeling of information is done using belief functions. Each event has a degree of belief (credibility) and a degree of plausibility, instead of a single degree of probability. One of the major advantages of the Dempster-Shafer theory over probability theory is thus to allow one to specify a degree of ignorance in a situation instead of being forced to supply prior probabilities. This ability to explicitly model the degree of ignorance makes the theory very appealing. Moreover, probabilistic approaches reason only on singletons. On the contrary, Dempster-Shafer theory enables us not only to affect belief on elementary hypotheses but also on composite ones. This last point illustrates the fact that Dempster-Shafer theory manages also imprecision, vagueness and inaccuracies. In addition, there is in the theory of Probabilities, a strong relation between an event and its negation, since its sum equals to 1. The theory of evidence implies no relations between the existence or not of an event. Thus, it models only the belief associated to a class, without influencing the belief allots to others classes.

Once the belief functions are obtained, the merger is realized through the combination rule of Dempster. This evidence combination rule provides an interesting operator to integrate multiple pieces of information from different sources. Thus, it is very helpful when working on pieces of information that come from various sources, as in information fusion applications. Finally, decision on the optimal hypothesis choice can be made in a flexible and rational manner.

In regards to previous sections, two points of view can be distinguished. On one hand, the view defended by the probabilities relies on the fact that all the previous mentioned theories lead to results that could have been obtained with a probabilistic model provided that the model has been sufficiently adapted to the problem. On the other hand, followers of the possibility theory and the evidential theory prefer modeling the available information as closely as possible. Nevertheless, we believe that belief functions theory proposes major advantages compared to the probabilities.

Actually, all these advantages find their roots in the fact that the Evidential theory is an extension of both the Probability, the Possibility and the Sets theory [Dubois 2011]. As a matter of fact, the theory includes extensions of probabilistic notions (conditioning, marginalization), of the possibility theory and includes also set-theoretic notions (intersection, union, inclusion, etc.). Indeed, Probability distributions are good for expressing variability (randomness), but they are information demanding, whereas Sets are good for representing incomplete information, but often crude representation of uncertainty. To that point of view, the theory of evidence is much more flexible than the probability theory or the possibility theory. It permits to manage as well uncertainties as the inaccuracies and the ignorance.

Finally, it is worth noting that there is a general trend to use increasingly belief functions in the fusion community as the two special sessions dedicated to that topic have shown it during the 14th International Conference on Information Fusion in Chicago, USA, in 2011.

3.3 Formalism and Reasoning Process of the Dempster-Shafer Theory of Evidence

Considering the advantages presented in the previous section, we have chosen to rely on the Dempster-Shafer theory of evidence. This section is devoted to give more in depth details about its mathematical formalism and use.

3.3.1 Fundamental Concepts

3.3.1.1 Frame of Discernment

Let Ω be the universal set: the set of all the N states (hypothesis) under consideration. This set is also called the frame of discernment.

$$\Omega = \{H_1, H_2, ..H_N\}. \tag{3.1}$$

We suppose that the universal set is exhaustive and that all hypotheses are exclusives. This assumption is also called the closed-world assumption. From this universal set, we can define a set, noted 2^{Ω}. It is called the power set and is the set of all possible subsets of Ω, including the empty set. It is defined as follows:

$$2^{\Omega} = \{A \mid A \subseteq \Omega\} = \{\emptyset, \{H_1\}, ..., \{H_N\}, \{H_1, H_2\}, ..., \Omega\}. \tag{3.2}$$

For example, if $\Omega = \{a, b\}$, then $2^\Omega = \{\emptyset, \{a\}, \{b\}, \Omega\}$. In the following, we will note H_n a singleton hypothesis (also sometimes called atomic state) and A a proposition referring indifferently to an hypothesis or a disjunction of hypotheses. The elements of the power set can be taken to represent propositions that one might be interested in, by containing all and only the states in which this proposition is true.

3.3.1.2 Basic Mass Assignment

An observer/source who believes that one or more states in the power set of Ω might be true can assign belief mass to these states. The theory of evidence assigns a belief mass to each element of the power set. This mass induces an opinion on the state of a system for instance. This piece of information can be issued from a sensor, an agent, an expert, etc. The latter will be referred to as the source of the information. Formally, a mass function is defined by:

$$m : 2^\Omega \to [0, 1] \tag{3.3}$$

It is called a basic mass assignment and has two properties. First, the mass of the empty set is zero:

$$m(\emptyset) = 0 \; . \tag{3.4}$$

Second, the masses of the members of the power set add up to a total of 1:

$$\sum_{A \in 2^\Omega} m(A) = 1 \; . \tag{3.5}$$

The mass $m(A)$ is interpreted as the part of belief placed strictly on A. It expresses the proportion of all relevant and available evidence that supports the claim that the actual state belongs to A but to no particular subset of A. The value of $m(A)$ pertains only to the set A and makes no additional claim about any subsets of A, each of which has, by definition, their own mass. This quantity differs from a probability since the total mass can be given either to singleton hypotheses H_n or to composite ones A.

Belief mass on an atomic state is interpreted as the belief that the state in question is true. Belief mass on a non-atomic state is interpreted as the belief that one of the atomic states it contains is true, but the source is uncertain about which of them is true.

Elements of Ω that have a non null mass are called focal elements. A focal set is a set A where $m(A) \neq 0$. Moreover, the union of focal elements is called a kernel.

As with most other representations, the major difficulty relies on assigning belief masses to each hypothesis. The objective is to model expert (source) opinions and statistical information using belief functions. We can also take into account the source reliability by discounting possibly the belief functions. Every method to define a mass function is potentially acceptable. Most of existing modeling depends on the considered application [Denoeux 2011].

3.3.1.3 Other Belief Measures

From the basic belief assignment, presented above, we can calculate other belief measures. We present here the credibility and plausibility belief measures.

Belief The belief - also called the credibility - for a set A is defined as the sum of all the masses of subsets of the set of interest:

$$bel(A) = \sum_{B|B \subseteq A} m(B) \quad \forall A \subseteq \Omega \; . \tag{3.6}$$

$bel(A)$ represents the set of belief mass given to the elements of this disjunction of hypotheses. It is the degree of evidence that directly supports the given hypothesis A at least in part, forming a lower bound. The special case, where $bel(A) = 1$ can be understood as A is certain.

This function verifies the following properties:

- $bel(\emptyset) = 0$,
- $bel(\Omega) = 1$,
- $bel(A_1 \cup A_2 \cup ...A_n) \geqslant \sum_{I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} bel(\cap_{i \in I} A_i)$.

The following picture illustrates graphically the above definition applied for the calculation of the belief of the set $A5$.

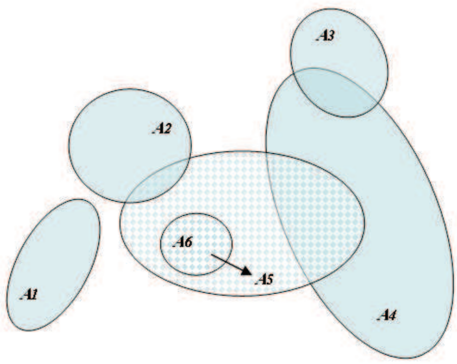

Figure 3.3: Graphical Illustration of Belief function.

The mass distribution m and the belief measure bel are two equivalent representations of the same information. As a matter of fact, the Möbius transformation permits to calculate the mass distribution from the belief one thanks to the following relation:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} bel(B) \; . \tag{3.7}$$

Where $| A \setminus B |$ is the cardinal of the set of elements of A, which do not belong to B.

Plausibility The plausibility $pl(A)$ is the sum of all the masses of the sets B that intersect the set of interest A:

$$pl(A) = \sum_{B|B\cap A\neq\emptyset} m(B) \quad \forall A \subseteq \Omega\ . \tag{3.8}$$

$pl(A)$ can be interpreted as the part of belief which could be potentially allocated to A, taking into account the elements that do not contradict this hypothesis. This is depicted through an example on figure 3.4.

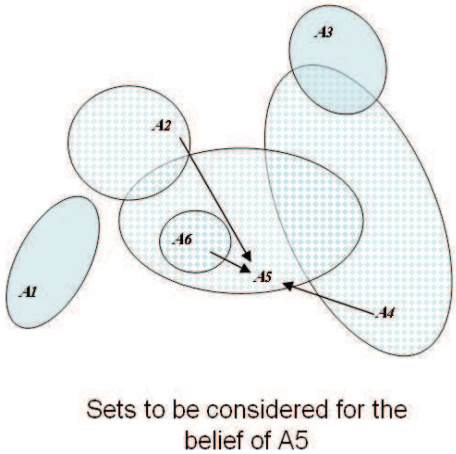

Figure 3.4: Graphical Illustration of Plausibility function.

Plausibility is 1 minus the sum of the masses of all sets whose intersection with the hypothesis is empty. It is an upper bound on the possibility that the hypothesis could possibly happen, i.e. it "could possibly happen" up to that value, because there is only so much evidence that contradicts that hypothesis. $pl(A)$ is also said to be the upper bound on the degree of support that could be assigned to A if more specific information became available. The special case, where $pl(A) = 0$ can be understood as A is impossible. This plausibility function verifies the following properties:

- $pl(\emptyset) = 0$,
- $pl(\Omega) = 1$,
- $pl(A_1 \cap A_2 \cap ...A_n) \leqslant \sum_{I\subseteq\{1,...,n\}}(-1)^{|I|+1}pl(\cup_{i\in I}A_i)$.

Relation between Belief and Plausibility

"Certain implies plausible"

The following equation holds for each hypothesis A within the frame of discernment. This can be understood as "certain implies plausible".

$$bel(A) \leq pl(A) \quad \forall A \subset \Omega\ . \tag{3.9}$$

From the ordering of belief and plausibility measure, the upper and lower bounds of a probability interval can be defined. This interval contains the precise probability of a set of interest (in the classical sense) as shown in the following equation:

$$bel(A) \leq P(A) \leq pl(A)\ . \tag{3.10}$$

Moreover, if focal sets are only singletons (i.e. we assign only masses to singleton hypothesis), then the mass distributions, credibility measures, plausibility ones and commonalities are merged and coincide with a probability distribution. A probabilistic distribution is thus a special case of the mass assignments.

Duality between belief and plausibility

This duality is depicted through the following equation and graphically illustrated in figure 3.5.

$$pl(A) = 1 - bel(\bar{A}) \quad \forall A \subset \Omega\ . \tag{3.11}$$

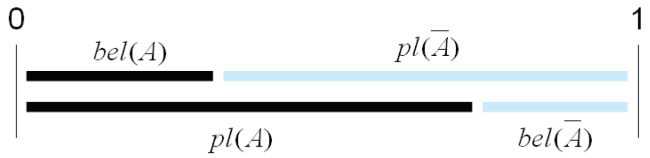

Figure 3.5: Relations of Belief and Plausibility and their Negation.

It follows from the above and from the Möbius transformation that one needs to know only one of the three (mass, belief, or plausibility) to deduce the other two, though one may need to know the values for many sets in order to calculate one of the other values for a particular set. m, bel and pl are thus three equivalent representations of a piece of evidence or, equivalently, a state of belief induced by this evidence.

Ignorance Interpretation

The difference $pl(A) - bel(A)$ quantifies ignorance about a specific hypothesis A. For example, having $pl(A) = 1$ and $bel(A) = 0$ for A brings no additional information. This can be interpreted as the total ignorance of the source regarding A. When the source cannot allocate its knowledge to a smaller set than Ω, it assigns: $m(\Omega) = 1$, which implies $m(A) = 0$, $A \neq \Omega$ and $bel(A) = 0$, $pl(A) = 1$. In other words, the source knows that the solution hypothesis is within the universal set, but cannot say anything more. This refers also to the total ignorance of the source. On the opposite $m(A) = 1$ indicates the total certainty of the source on singleton hypothesis A.

3.3.2 Combination Process

Defining masses within the Dempster-Shafer theory would be useless without an adequate combination process enabling the fusion of beliefs from different sources. We propose here to browse the major rules developed and used in the fusion community working with belief functions through the last past thirty years (as recalled in [Martin 2008]).

3.3.2.1 Unnormalized Dempster's rule

Let m_1 and m_2 be two mass functions on Ω induced by two independent items of evidence. We note $m = m_1 \wedge m_2$, the fused mass distribution issued from the fusion of the two distributions m_1 and m_2 and defined with the following formula.

$$(m_1 \wedge m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \ . \tag{3.12}$$

3.3.2.2 Normalized Dempster's rule

The Dempster's combination is calculated from two sets of masses m_1 and m_2 in the following manner:

$$(m_1 \oplus m_2)(A) = \begin{cases} \frac{(m_1 \wedge m_2)(A)}{1 - K_{12}} & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases} \ . \tag{3.13}$$

where $K_{12} = (m_1 \wedge m_2)(\emptyset)$ is the degree of conflict. K is a measure of the amount of conflict between two mass sets. It is ranging from 0 to 1. It has for value 1, when all the focal subsets assigned by one source have a null intersection with each focal subsets assigned by the other source. It has for value 0, when each previously mentioned intersection is not null. Consequently, if K is close to 0, the two sets of masses are almost not in conflict, while if K is close to 1, they are almost in total conflict. It has the effect of completely ignoring conflict and attributing any mass associated with conflict to the null set.

Let's take an example to illustrate this combination. We consider the frame of discernment $\Omega = \{H_1, H_2\}$ and we suppose two sources: S_1 and S_2.
S_1 provides the following mass assignment:

$$\begin{cases} m_1(H_1) = \frac{1}{3} \\ m_1(H_2) = \frac{1}{6} \\ m_1(H_1, H_2) = \frac{1}{2} \end{cases}$$

where $m_1(H_1, H_2)$ represents the imprecision of the source S_1.
S_2 provides the following information:

$$\begin{cases} m_2(H_1) = \frac{3}{4} \\ m_2(H_2) = \frac{1}{4} \end{cases}$$

The product of mass intersections are represented in the following table 3.6.

$S_1 \backslash S_2$	$m_2(H_1) = \frac{3}{4}$	$m_2(H_2) = \frac{1}{4}$
$m_1(H_1) = \frac{1}{3}$	$\frac{1}{4}$	$\emptyset$ (conflict)
$m_1(H_2) = \frac{1}{6}$	$\emptyset$ (conflict)	$\frac{1}{24}$
$m_1(H_1, H_2) = \frac{1}{2}$	$\frac{3}{8}$	$\frac{1}{8}$

Figure 3.6: Preliminary step of the Dempster's rule: intersection product.

The calculi of the conflict gives: $K_{12} = \sum_{B \cap C = \emptyset} m_1(B)m_2(C) = \frac{1}{6} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{3} = \frac{5}{24} \simeq 0.2$.

Therefore, after normalizing, we deduce:

$$\begin{cases} m(H_1) = \frac{\frac{1}{4}}{1-\frac{5}{24}} = \frac{6}{19} \simeq 0.3 \\ m(H_2) = \frac{\frac{1}{24}}{1-\frac{5}{24}} = \frac{1}{19} \simeq 0.05 \\ m(H_1, H_2) = \frac{\frac{3}{8}+\frac{1}{8}}{1-\frac{5}{24}} = \frac{12}{19} \simeq 0.63 \end{cases}$$

Many other rules of combination exist in the literature and are still proposed, they mainly try to handle differently the conflict repartition. Among them, we can mentioned the Yager's rule [Yager 1987], which was proposed to overcome counter-intuitive results in case of high conflict within the Dempster's rule. Other noteworthy combination rules are the disjunctive rule [Smets 1993], the Dubois and Prade mix rule [Dubois 2008] or the cautious rule [Denoeux 2006].

3.3.3 Decision Making

Belief modeling and combination enable the elaboration of decision processes.

We can distinguish two principal types of decision process: decision on singletons and decisions on composite ones. In the latter case, it is still possible for example to limit the cardinality of composite hypotheses to 2 or 3 singletons for instance.

Main decision processes on sets (which are not necessary restricted to singletons) are through the maximum of credibility or the maximum of plausibility. In the first case, we choose the hypothesis or the group of hypotheses whose credibility is the most important. That amounts choosing the subset whose implications for this subset are maximal. In the second case, we choose the hypothesis or the group of hypotheses whose plausibility is the most important. That amounts choosing the subset which contradicts the less the whole set of available information. It is also possible to fix a user specific criterion for decision. For example, one decision criteria could be to choose a subset whose difference between its credibility and plausibility is the weaker and whose credibility or plausibility is the higher.

We can also apply the maximum of pignistic probability proposed by [Smets 1990]. Let m be any mass distributions, the pignistic probability distribution is defined such as:

$$\forall A \subseteq \Omega,\ P_m(A) = \sum_{\emptyset \neq B \subseteq \Omega} m(B) \frac{\mid A \cap B \mid}{\mid B \mid} . \tag{3.14}$$

Then, we choose the hypothesis whose pignistic probability is the maximal. In other words, we consider that, during the decision process, masses should be distributed such as probability coefficients.

3.4 Conclusion and Remarks

Information fusion techniques try to manage epistemic uncertainty - due to lack of knowledge of the sources or of the processes - and inconsistent pieces of information. The Dempster-Shafer theory seems to be well adequate for this purpose. Indeed, this mathematical theory can model epistemic uncertainty: both empirical uncertainty (a sentence about a world is either satisfied or not) and imprecision. Empirical uncertainty is modelled through the value of masses or other belief functions assigned to the hypotheses. Imprecision is modelled through a set encompassing more or less hypotheses. The Dempster-Shafer theory can also deal with inconsistent pieces of information, thanks to the conflict management in the different combination rules.

In all this Chapter, hypotheses have been considered and restricted to simple labels - as it is often presented in classical mathematical formalisms (e.g. H_1, H_2). However, as we will see in next chapter (Chapter 4), the knowledge acquired by different sources will be modelled in our applications through ontologies (and more particularly through instantiations of an ontology). As such, uncertain hypotheses are also modelled through ontologies and have thus an explicit semantic attached to them. For example, hypotheses might have different levels of semantic granularity. Whereas a first hypothesis can refer to a "vehicle", a second hypothesis could be a "red car". These hypotheses are not contradictory and are semantically related. As a matter of fact, semantic descriptions of hypotheses may depend on the level of expertise or the capabilities of the sources.

Then, Chapter 5 will review the very recent approaches in that research area that try to handle uncertainty within ontologies or taking the problem the other way round : that try to handle semantics within rigid mathematical theories that we presented in this chapter.

Chapter 4

Ontology and Semantic Reasoning

Contents

As previously presented (see section 2.3.1), knowledge representation is a major challenge in information fusion applications. It is the field of study within Artificial Intelligence concerned with using symbols to represent a collection of propositions believed by some agents. It is considered as the fundamental support for further processing carried out on knowledge.

This chapter introduces the notion of OWL ontology, which is relatively new in the landscape of knowledge representation. It provides explicit semantics to the modelled information. As such, a clear benefit of ontologies with respect to simpler representation formalisms consists in the support of reasoning tasks based upon the expressed semantics.

An ontology is seen as a powerful tool and a key technology for information fusion, which permits to answer the knowledge representation challenge among others.

4.1 Knowledge Representation

One of the best options to understand what knowledge representation is, is to simply mention what it is intended for. Its mission is to make knowledge as explicit as possible. This is necessary since knowledge is stored in implicit form [Garcia González 2006]. Implicit knowledge is what an agent obtains when it observes its environment and makes its own internal representations of its beliefs.

However, to facilitate knowledge sharing, it is necessary to make it explicit.

Therefore, the role of reasoning is to bridge the gap between what is represented explicitly and the implicit knowledge. This is the reason why, the fundamental goal of knowledge representation is to represent knowledge in a manner that facilitates inference, that is to say draw conclusions from knowledge (see section 4.4).

In the following, conceptual notions composing the landscape of knowledge representation are introduced.

4.1.1 Semiotic Triangle

One practical way to understand more deeply the difference between implicit knowledge and its explicit representation can be carried out through the Semiotic Triangle vision, depicted in figure 4.1.

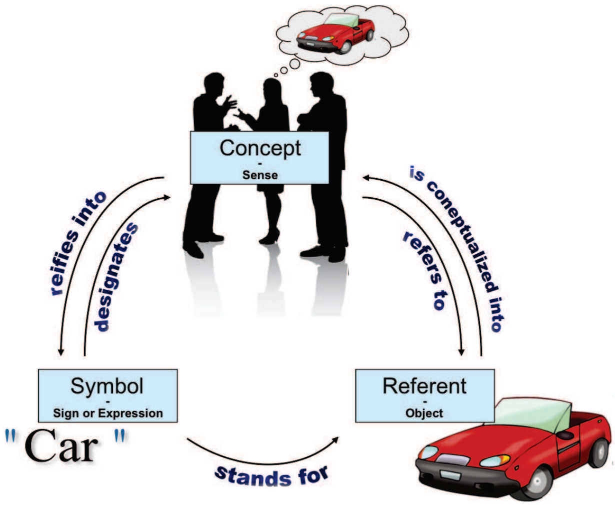

Figure 4.1: Semiotic triangle (adapted from [Sølvberg 1999]).

It all starts from an object. The latter is also called a referent and is something that exists in itself and that is to be the focus of a particular interest. For instance, a referent can be a specific car I saw this morning in the street. Through our limited senses and intellect, we create a certain conceptualization of this referent (depicted by the top corner of the triangle), biased by a specific interest or point of view. However, this mental representation is only useful for the bearer of the conceptualization, unless it is shared with others. In order to do anything that transcends the boundaries of one's intellect, a set of symbols (represented by the

"Car" symbol in the figure) that denotes the concept must be explicitly defined. Symbols are indeed the only means by which concepts can be shared among people.

As such, one can say that semantics deals with how knowledge representations are related by agents to the things they stand for.

4.1.2 Data, Information and Knowledge

The concept of data versus information, and knowledge is not new and much literature can be found on the subject (see for instance [Aamodt 1995, Bellinger 2004, Boisot 2004]). This literature leads generally to the following schema 4.2, which is often referred to as the "Knowledge/Information Hierarchy", or the "Knowledge Pyramid". In this graph, the higher a piece of data is situated, the more meanings

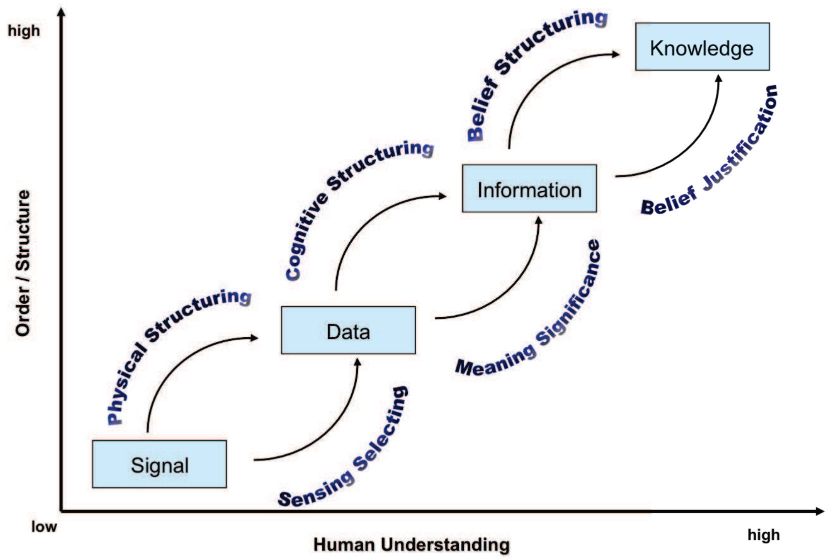

Figure 4.2: Signal, Data, Information and Knowledge.

is embedded. There is a semantic growth, gathering more and more reasoning and understanding. In the following, we provide explanations concerning these concepts.

In [Dretske 1981], the author offers a useful definition. It says that, "Information is that commodity capable of yielding knowledge, and what information a signal carries is what we can learn from it. Knowledge is identified with information-produced (or sustained) belief, but the information a person receives is relative to what he or she already knows about the possibilities at the source."

In other words, data is the raw material of information and is represented through raw symbols. It simply exists and has no significance beyond its existence (in and of itself). Information is a flow of messages. Information is intended to be data with meaning that are processed to be useful. The patterns and relationships in the data are pointed out and discussed. It is the output from data interpretation as well as the input to, and output from, the knowledge-based process of decision-making. Knowledge can be seen as information incorporated in an agent's reasoning resources and context, and made ready for active use within a decision process.

4.1.3 Conceptual Levels of Knowledge Sharing

This section introduces now some useful remarks about the different levels of semiotics, which is another view of distinguishing three fundamental branches of communication. This distinction will be valuable to understand the different positions of languages presented in the following (namely sections 4.3.2 and 4.3.3). These levels ranges from the most basic one to the most expressive one, where each level puts an additional layer to the previous one:

- syntactic level (forms of language). It is concerned with the rules for building up sentences. It solves a technical problem: how accurately the symbols used in communications can be transmitted?
- semantic level (meanings of language). It examines the meaning of signs in relation to the represented objects or actions. It solves a representation problem: how intelligibly the transmitted signs represent the intended message, and, how precisely the transmitted symbols convey the desired meaning?
- pragmatic level (use or function of the language). It features how the senders and receivers evaluate and understand the meaning including psychological impact, action consequences, etc. At the pragmatic level, we solve an efficiency problem: how efficiently the received message influences the behaviour of the receiver, or more precisely, how effectively the received meaning affects the conduct in the desired way?

4.1.4 Existing Representation Techniques

Many knowledge representation methods were tried in the 1970s and early 1980s, such as heuristic question-answering, neural networks, theorem proving, and expert systems, with varying success.

In computer science, knowledge is contained in knowledge bases, which can be of different types, more or less expressive. Dictionary is the less expressive knowledge base. It consists of a words and terms collection having no relations between them. More expressive knowledge representations - based also on non-logic approaches - are the well known production rules, semantic networks [Quillian 1968] or conceptual graphs [Sowa 1976].

Description Logics form another family of models, structured and formal, to represent knowledge and reason on the represented knowledge. In recent years, they have inspired the development of ontology languages. This is the topic of following sections (see section 4.3.3). Description Logics are considered as the logical basics and inference mechanisms of ontologies.

4.2 The Concept of Ontology

There does not exist one single, universal, and commonly accepted definition but rather several definitions and attempts to explain the concept of ontology. This may be due to its origin and its recent transition from an abstract A.I. - Artificial Intelligence - concept to its current use in computer science.

Originally, the term of ontology comes from the domain of philosophy. Indeed, ontology has arisen out of the branch of philosophy known as metaphysics, which deals with the nature of reality, that is to say of what exists. Ontology tends to explain concepts that exist in the world and how they are organized and related to each others.

Relying on the semiotic triangle, we introduce in figure 4.3 the distinct meaning of "ontology" within the branch of philosophy and within computer science. As a matter of fact, ontology is a word that the computer science has borrowed to philosophy, at the beginning of the 1990s.

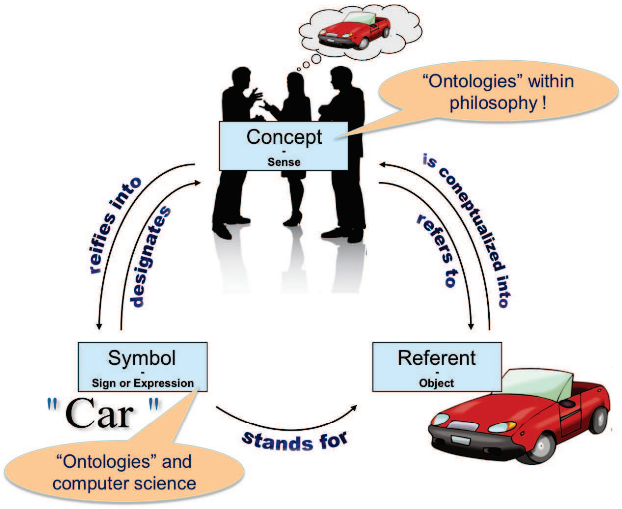

Figure 4.3: Semiotic triangle revisited with the concept of Ontology.

In Artificial Intelligence, one of the most accepted definitions has been stated in [Gruber 1993]. Grubber specifies: "an ontology is an explicit specification of a conceptualization". This definition was then refined in [Borst 1997]: an ontology is "a formal, explicit specification of a shared conceptualization".

The two previous definitions are rich of sophisticated words. We try to explain them in the following. First of all, the two definitions use the word *conceptualization* to refer to what an ontology is trying to represent. A conceptualization has to be understood as a simplified view of the world we wish to represent for some purpose. This view is composed of identified concepts, objects, and other entities that are presumed to exist in this view. An *explicit specification* refers to an unambiguous definition of the different concepts, the relationships between them and the constraints on their use related to the abstract model. This implies machine-readable and understandable. Finally, adding the two adjectives *formal* and *shared* restrict a little more the definition of ontology. *Formal* adds the idea of a well-established procedure of how to specify the representation of knowledge, while *shared* implies that the knowledge view is based on a consensus. It reflects the notion that an ontology contains knowledge used and reused across different applications. In other words, ontologies propose a shared and common understanding of an area of knowledge that can be communicated between people or heterogeneous application systems. This area of knowledge underlines the fact that an ontology is domain specific and is not intended to describe or represent the whole world.

Specific points, commonly accepted by the community of knowledge engineering and inspired from the description logics, lead to a more concrete explanation of what an ontology is. The knowledge of a domain is divided into two levels that represent the knowledge of an application domain (the "world").

First, it defines the relevant classes and relations of the domain. This level is called the terminology, or the *T-Box* and refers to the structural knowledge. The notion of classes, also equivalently called concepts, is ubiquitous. They are organized through relations of subsumption, resulting in taxonomy. Superclasses represent higher-level concepts and subclasses represent finer concepts, and the finer concepts have all the attributes and features that the higher concepts have. There is also the notion of relation between concepts. Each class is typically associated with various properties describing its features and attributes as well as various restrictions on them.

Secondly, it uses these concepts to specify properties of objects and individuals occurring in the domain (the world description). This is the assertion level, called *A-Box*, which constitutes factual knowledge or a set of assertions, also called instantiations.

One call a knowledge base the couple (*T-Box, A-Box*).

Ontologies satisfy several purposes: most notably, they prevent misunderstandings in human communication and they ensure that software behaves in a uniform, predictable way and works well with other software. However, in order to complete and implement this vision, we need to realize and provide the *explicit specification* part of the definition. As such, an ontology language is needed, and this is the topic of next section.

4.3 OWL - Web Ontology Language

An ontology language is a formal language used to encode the ontology. There has been a number of such languages for ontologies, both proprietary and standard-based. Among them, we can list: Common logic [Group 2003], DOGMA [Jarrar 2002], KIF [Genesereth 1992], F-Logic [Kifer 1989], OIL [Fensel 2001] and last but not least OWL.

OWL - Web Ontology Language - is nowadays without any doubt the most popular language for creating ontologies. As its acronym underlines it, most important efforts in developing this language have been made in the field of the Semantic Web.

Therefore, a special section is devoted to explain the Semantic Web vision, which will permit to present the main objectives of OWL. Due to its origins, OWL has its roots in its own web language predecessors (i.e. XML, RDF and RDFS). However, OWL is naturally not restricted to the area of the Web. It has already attracted both academic and commercial interests in several domain applications.

4.3.1 Originated from the Semantic Web

The Semantic Web is the new vision of the Web, whose main goal is to make Web contents not only human readable but also machine readable and processable. Indeed, the current World Wide Web (or WWW, or simply the Web) is more a syntactic web, where pages are designed to be read by humans and not by computer programs.

Tim Berners-Lee *et al* presented in May 2001 [Berners-Lee 2001], the idea of the Semantic Web. It is "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation". Therefore the Semantic Web is not a separate Web but an extension of the current one.

The Web has had an enormous success. Since its creation in early 1990s, it has grown exponentially, and it has reached currently enormous proportions. This is good news for its users: there is an immense amount of information and opportunities to exploit. However, simple accumulation is not the answer. In order to efficiently exploit it and extract its full potential, more elaborated mechanisms should be layered over the basic HTML pages that the Web was previously providing. Therefore, the main intent of the Semantic Web is to give machines much better access to information resources so that they can be information intermediaries in support of humans. The potential for increasing knowledge availability and the ability of machines to effectively work with it is enormous.

During this last decade, the W3C - World Wide Web Consortium - has relied on a set of languages, depicted in figure 4.4. This set of languages is called by the W3C the Semantic Web Stack, also referred as the semantic layer cake.

Technologies from the bottom of the stack up to OWL are currently standardized. They are widely accepted to build Semantic Web applications. First layers (URI/IRI and XML) are inherited from the previous Web. The others try to build the Semantic Web (see section 4.3.2 for a presentation of the four layers up to OWL). However, it is still not clear how the top of the stack is going to be implemented.

4.3.1.1 Semantic Web Assumptions

The essential property of the Web is its universality. However, or maybe due to that, the current Web is quite unrestricted; it sacrifices link integrity for scalability. This great lack of restrictions in the Web design makes it fundamentally different from traditional hypertext systems. This design principle continues leading the Semantic

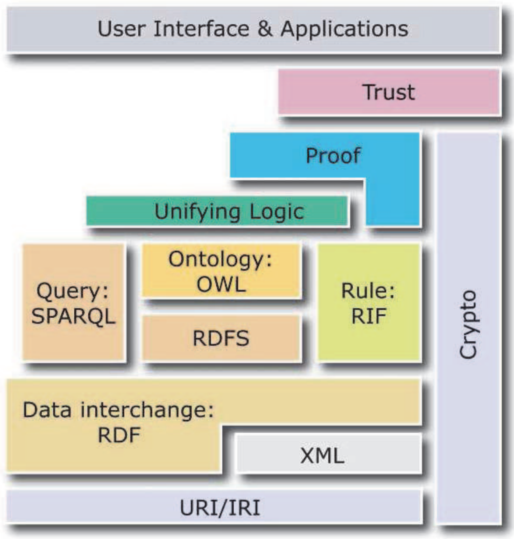

Figure 4.4: Semantic layer cake.

Web view. This lack of constraints paradoxically leads to three major ideas, as explained in [Allemang 2008]:

- **AAA**. The first assumption is "Anyone can say Anything about Any topic". It is often referred to the AAA principle or partial information principle and is relative to one of the basic tenets of the Web in general and the Semantic Web in particular. Indeed, as mentioned just before, current Web is quite unrestricted. Therefore, there should be no fundamental constraint relating what is said, what is said about, and where it is said. Consequently, it is not expected to have global consistency of all data.

- **Open World Assumption**. The second assumption is a consequence of the AAA slogan and states that there could always be something new that someone will say; this means that we assume that there is always more information that could be known. It is called the open world assumption. In open world assumption, everything, which was not specified explicitly, is unknown. This is the opposite assumption of databases, which are typically developed under a closed world assumption. In databases, the only possible instances of a given relation are those implied by the objects existing in the database. In other words, if something is not represented there, then it does not exist. This closed world assumption is also adopted by almost all logic programming languages. The Semantic Web, on the contrary, does not necessarily carry the assumption that not being represented entails non-existence.

- **Non-Unique Name Assumption**. Finally, there is the Non-unique naming principle. Indeed, since the speakers on the Web do not necessarily coordinate their naming efforts, the same entity could be known by more than one name.

These requirements have set the foundation for the design of the Semantic Web technologies.

4.3.2 Incremental Technology Path to OWL

OWL is considered as a sophisticated language since it was designed based on several communities and on several existing languages. In fact, previous mentioned knowledge representations such as Description Logics and early Semantic Web technologies have largely influenced the development of OWL. More particularly, to create such a language, important technologies for developing the Semantic Web have been used: URI/IRI, XML, RDF, and RDFS, as illustrated by the semantic layer cake.

4.3.2.1 Uniform/Internationalized Resource Identifier - URI/IRI

This is the bottom layer of the Semantic Web, which provides its global perspective, already presents in the WWW.

URIs/IRIs[1] are a basic mechanism of the Web where all hyperlinks are expressed in URI/IRI format. This latter is a chain of characters that allows identifying a resource in a unique manner. A resource, in the Semantic Web domain is a basic entity used to represent some knowledge.

4.3.2.2 eXtensible Markup Language - XML

A common characteristic of the W3C standardized languages is to have the faculty of identifying resources through URIs and to be expressible according to an XML syntax. XML[2] is as such considered as the basic language for the Semantic Web.

XML is a meta-language (language that is used to define another language) easing the elaboration of specialized tag-based language. XML provides a definition and structural frame of notions that constitute a document format. A notion is characterized by a tag or an XML attribute. These notions are then instantiated in documents that respect this format.

4.3.2.3 Resource Description Framework - RDF

The RDF[3] language defines the building blocks to realize the Semantic Web. This is the first layer that was specially developed for it.

RDF structures information more homogeneously than XML. Indeed, RDF is a formalism for knowledge representation based around the fundamental notion of a knowledge *triple*. A triple is a representation of a relation between a subject, a predicate, and an object. A triple can be interpreted as "the subject has for predicate the object". Each component of this triple is referred to as the generic term "*resource*".

[1] http://www.w3.org/Addressing/
[2] http://www.w3.org/XML/
[3] http://www.w3.org/RDF/

4.3.2.4 RDF Schema - RDFS

RDFS[4] aims to extend RDF by providing a structuring for RDF resources. In order to do this, it enables the building of classes and properties. A class embodies a type for a set of subject or object; a property embodies a type for a set of predicates. These resources are called assertions. We can note that an assertion whose type is a class is also called an instance of that class; this term is often preferred in this case. Properties have a direction of application, from a domain to a range. This indicates that the subject of a property must belong to the set of instances of the indicated class (domain), and respectively the object of a property must belong to the set of instances of the indicated class (range).

RDFS proposes also a way to organize the set of classes and properties in taxonomy, i.e. in hierarchy.

The joint use of RDF and RDFS is often noted RDF(S). It offers the possibility to represent knowledge of a domain on two conceptual levels. The elements modelled in RDFS represent structural knowledge of the represented domain; the elements modelled in RDF represent the assertional knowledge.

4.3.3 General Description of OWL

In order to offer deeper modeling, the W3C recommends the OWL language, which enables the creation of expressive ontologies. OWL[5] is the pseudo acronym of Web Ontology Language. OWL came out in 2001 and has only been a W3C recommendation in 2004. It was then revisited in OWL2 and was subject to a second W3C recommendation in October 2009 ([Hitzler 2009]). The OWL language extends RDF(S), by providing a richer vocabulary. The most common semantics for ontology languages is description logic, a decidable fragment of first-order logic. OWL is a quasi-rewriting of some description logics. The syntactic features of description logic have been recapped in figure 4.5, where C and D are concepts, a and b are individuals, and R is a role. Ontologically speaking, C and D are classes, a and b instances, and R a property.

OWL adds more vocabulary than RDFS for describing properties and classes. For example, more restrictions can be used to describe relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry) and enumerated classes. An exhaustive description of this language is given in [Hitzler 2009].

OWL defines three sublanguages that offer different advantages in particular application scenarios, as for OWL Lite, DL and Full. As a matter of fact, OWL Lite is a restricted form of OWL that is intended to be easy to understand, and easier to implement in applications. OWL-DL contains all the OWL language primitives, but imposes restrictions on their use. OWL-DL is designed to maximize expressiveness

[4] http://www.w3.org/TR/rdf-schema/
[5] http://www.w3.org/TR/owl2-overview/

Symbol	Description	Example	Read
$\top$	All concept names	$\top$	top
$\bot$	Empty concept	$\bot$	bottom
$\sqsubseteq$	Concept inclusion	$C \sqsubseteq D$	all C are D
$\forall$	Universal restriction	$\forall R.C$	all R-successors are in C
$\exists$	Existential restriction	$\exists R.C$	an R-successor exists in C
$\sqcap$	Intersection or conjunction of concepts	$C \sqcap D$	C and D
$\sqcup$	Union or disjunction of concepts	$C \sqcup D$	C or D
$\equiv$	Concept equivalence	$C \equiv D$	C is equivalent to D
$\neg$	Negation or complement of concepts	$\neg C$	not C
:	Concept assertion	$a : C$	a is a C
:	Role assertion	$(a, b) : R$	a is R-related to b
U	All role names	U	top role
$= n$	Cardinality of role	$C \equiv = nR.\top$	C has n R-successors

Figure 4.5: Conventional Notation of Description Logic.

while retaining decidability and completeness of reasoning. OWL Full is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees.

Moreover, three additional sublanguages (or profiles) have been also defined in OWL 2: EL, QL, and RL. Briefly, OWL 2 EL is a fragment that has polynomial time reasoning complexity. OWL 2 QL is designed to enable easier access and query to data stored in databases, and OWL 2 RL is a rule subset of OWL 2.

Each of the sublanguages/profiles trades off different aspects of OWL's expressive power in return for different computational and/or implementational benefits.

4.4 Automatic Implicit Inference

OWL is describing a state of knowledge in a logical way. However, in order to use this intrinsic logic and get something more than a notation, appropriate tools are needed: the so-called reasoners. As underlined by [Allemang 2008], semantics and inference are thus strongly connected.

Reasoning tasks enhance systems and improve queries possibilities by inferring implicit knowledge. It infers new information related both to the terminology (T-Box) and the instantiation (A-Box) part of the ontology. It is carried out by rules that determine how patterns are generated from others. Appropriate inference rules allow reasoning mechanisms automation and, thus, the generation of new knowledge from previous one.

We focus here on reasoners that could deduce automatically logical consequences of the encoded domain knowledge in an ontology. Among the many automated reasoning tasks, one can mention:

- Ontology consistency. In [Haase 2005], the authors distinguish different forms of ontology consistency, among them:

  - Structural consistency: This notion of consistency ensures that the ontology conforms to the ontology language constraints imposed by this language. Structural consistency can be enforced by verifying a set of structural conditions related to the ontology language in use. As an example of structural conditions we can state "The complement of a class must be a class".

  - Logical consistency: An ontology is logically consistent if it does not contain contradictory information, it conforms to the underlying formal semantics of the ontology language.

- Concept satisfiability. Its aim is to verify whether a concept does not necessarily imply the empty concept (i.e. does this concept can be instantiated?).

- Concept subsumption. It checks whether a concept is considered more general than another one.

- Concept equivalence. Two concepts are equivalent if they subsume each other.

- Concept disjointness. Two concepts are disjoint if they do not have a common instance.

An analogy to an iceberg can be useful to understand the scope of this reasoning (see Figure 4.6). The original knowledge, which has been stated explicitly in the ontology, represents only the tip of an iceberg. However, nine-tenths below the surface correspond to the implicit knowledge that will be inferred and stated explicitly thanks to an automatic reasoner.

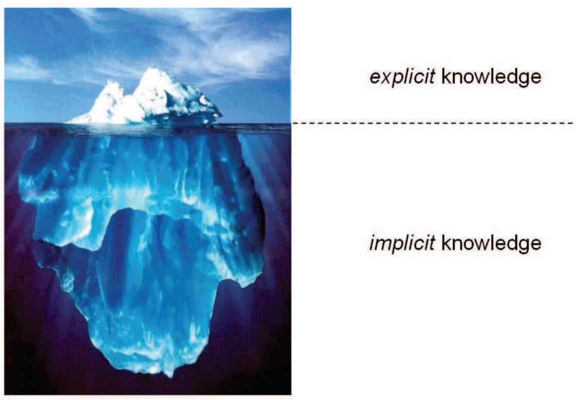

Figure 4.6: Iceberg Analogy.

Still, even if the inferred knowledge comes from universal rules, the way this resulting knowledge has been reached is quite complex and may be difficult for people to understand. Indeed, the way ontological axioms interact to infer knowledge can be very subtle [Hitzler 2009]. As such, reasoners can discover information that a person would not have spotted.

To support automatically the tasks listed above, a variety of reasoners have been proposed. Among them, we can list: Hermit[6], Fact++[7], RacerPro[8], Pellet[9].

4.5 A Domain-Ontology Example

This section illustrates with a concrete toy example the notions that have been developed throughout this Chapter 4. An example of a domain ontology is here provided. It is a quite rather simple example with no specific application purpose. However, it recalls concretely the notions of terminology: classes, properties, logical axioms, the notions of assertional facts through instances, and the notion of inferred axioms. Moreover, this example is introduced here, but it is important to note that it is a thread example for the whole dissertation (it will be particularly used in sections: 6.3, 7.3, 8.2.4).

This domain ontology is depicted here through the Description Logic syntax introduced in section 4.3.3. Indeed, this syntax is rather compact and quite universal when expressing logical axioms. For seek of completeness, this ontology example is also given in computable OWL syntax (here, the OWL functional syntax) in appendix A. Finally, this section also introduces the Protégé editor[10], proposed by Stanford University, which permits to manage graphically and easily ontologies. We will refer again to this tool in this dissertation - namely in section 9.1 of Chapter 9.

First, the taxonomy of classes is provided through the following axioms. This taxonomy is also graphically shown on the left hand side of figure 4.7.

1- $\texttt{Book, Color, Direction, Vehicle} \sqsubseteq \top$

2- $\texttt{Aircraft, LandVehicle, WaterCraft} \sqsubseteq \texttt{Vehicle}$

3- $\texttt{Bicycle, Car, Truck} \sqsubseteq \texttt{LandVehicle}$

4- $\texttt{FireTruck} \sqsubseteq \texttt{Truck}$

5- $\texttt{SubsurfaceVessel, SurfaceVessel} \sqsubseteq \texttt{WaterCraft}$

The above classes have relations between them, which are defined through the following axioms. The two first lines represent the taxonomy of these object properties (represented in the center figure 4.7). Then, the domain and range definitions are given. Axioms of the form: $\exists R.\top \sqsubseteq D$ indicates that a role: R has for domain the D class and axiom $\top \sqsubseteq \forall R.C$ indicates that R role has for range the C class.

1- $\texttt{hasMainColor, isStoppedNear, movesTowards} \sqsubseteq \mathrm{U}$

2- $\texttt{movesFastTowards, movesSlowlyTowards} \sqsubseteq \texttt{movesTowards}$

3- $\exists\texttt{isStoppedNear}.\top \sqsubseteq \texttt{Vehicle}$

[6] http://hermit-reasoner.com/
[7] http://code.google.com/p/factplusplus/
[8] http://www.racer-systems.com/
[9] http://clarkparsia.com/pellet
[10] http://protege.stanford.edu

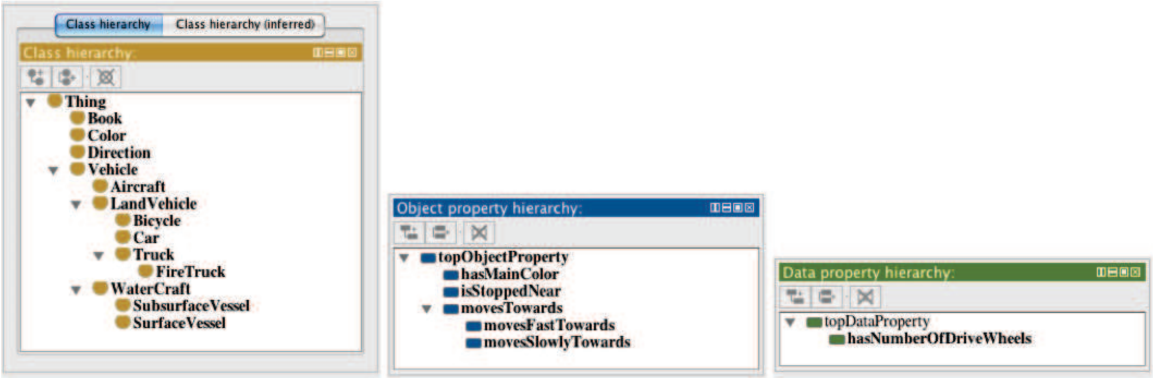

Figure 4.7: Protégé snapshots of the hierachical structure of ontological concepts.

4- $\top \sqsubseteq \forall$`isStoppedNear.Vehicle`
5- $\exists$`movesTowards`.$\top \sqsubseteq$ `Vehicle`
6- $\top \sqsubseteq \forall$`movesTowards`.(`Vehicle` $\sqcup$ `Direction`)
7- $\exists$`hasMainColor`.$\top \sqsubseteq$ `Vehicle`
8- $\top \sqsubseteq \forall$`hasMainColor.Color`

One datatype property is also defined:
1- `hasNumberOfDriveWheels` $\sqsubseteq$ U

Some instances of classes (i.e. individuals) have been asserted within the ontology. They are here noted with an additional "#" prefix so as to easily distinguish them from their respective classes in the ontology. This list of individuals is graphically represented in figure 4.8.
1- `#aircraft` : `Aircraft`
2- `#redCar` : `Car`
3- (`#redCar`, `#red`) : `hasMainColor`
4- `#red, #green, #blue` : `Color`
5- `#south, #east, #west, #north` : `Direction`
6- `#fireTruck` : `FireTruck`
7- `#landVehicle` : `LandVehicle`

Regarding this list of individuals, an additional axiom is added to further define the `FireTruck` class within the terminology. It states that all instances of `FireTruck` class are instances of `Truck` class and have the `hasMainColor` object property linked to the `#red` instance.
1- `FireTruck` $\sqsubseteq$ `Truck` $\sqcap$ $\exists$`hasMainColor`.{`#red`}

Thanks to this additional axiom, when a reasoner will be run over this ontology, it will state explicitly that the fire truck object is also red:
1- (`#fireTruck`, `#red`) : `hasMainColor`

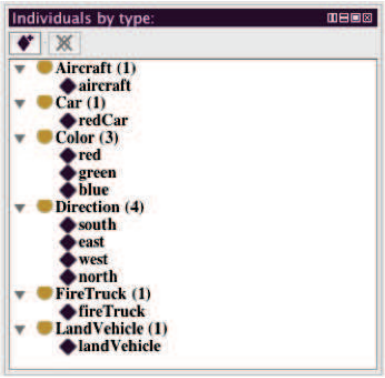

Figure 4.8: Protégé snapshot of the list of instances.

4.6 Interest in Information Fusion

Since the development of semantic technologies, many domain ontologies have been created. The scientific community in many domains - other than the Semantic Web (Biomedical informatics, Software Engineering, Information architecture, Artificial Intelligence) - is following this area of research with great interest. This section underlines commonalities between information fusion applications and the Semantic Web. The advantages of such technologies for information fusion are then presented.

4.6.1 Semantic Web and Fusion Domain Applications: Same Assumptions

Some connections between information fusion and the Semantic Web environment can be pointed out according to their fundamental principles and respective goals.

As presented in section 4.3.1.1, there are three essential assumptions that lead the development of Semantic Web technologies: the AAA statement, the Open World assumption and the Non-Unique Name postulation.

It is clear that these three principles can be also applied to the domain of information fusion. As a matter of fact, fusion systems are used in environments, where available data do not come from a single source, but from many heterogeneous sources. In other words, different sources deliver observations from the same object of interest. This enables the system to deliver a more accurate description of the object, which is one typical goal of information fusion. This heterogeneity of sources and of observations illustrates particularly the AAA principle: "Anyone can say Anything about Any topic". Moreover, as our sensors are not ubiquitous or omniscient, there could always be more information that could be known as in the Open World Assumption holding for the Semantic Web. Finally, the same object could be labelled with different names by different sensors since it is the fusion objective to determine if different observations refer really to the same entity (known as the "Association" step of a fusion application). Therefore, the last principle of the Non-Unique Name holds here also.

Fusion and Semantic Web applications are sharing the same basic and fundamental underlying principles. They are actually evolving in the same type of environment, with the same goal of representing knowledge and fusing it in order to form a consistent and global view of the knowledge. As a consequence, it sounds logical trying to use the same technologies for both domains. This leads to the deduction that semantic technologies: XML, RDF(S) and OWL, which have been largely developed within the World Wide Web, can also be useful for information fusion applications.

4.6.2 Benefits of Ontology-based Fusion

As seen in Chapter 2, information fusion is a domain highly related to artificial intelligence and more especially associated to knowledge representation. Ontologies are seen as a way out to answer some problems that need to be tackled and solved within information fusion applications.

First of all, one key research topic for information fusion is knowledge representation (see section 2.3.1). As a matter of fact, one needs a sufficiently expressive knowledge representation in order to capture complex behavior characterizing situations. Vintage information fusion systems have usually relied on application-specific database schemas and other rudimentary representation formats. These were adequate for lower levels of information fusion in standalone systems, but as the trend towards higher levels of the information fusion spectrum became clearer their limitations have become apparent. This first issue can be tackled through the use of ontologies. Indeed, ontologies represent a shared understanding of some domain through classes, instances and relations and enjoy of a logical expressivity greater than taxonomies. More generally, it is true that organizing knowledge in terms of classes, properties, restrictions and individuals has been proven to be well accepted by domain experts and software developers, since this paradigm is very similar to object-oriented modeling or UML. As such, it has promoted ontology design in the fusion community as well. Ontologies are thus valuable to represent the situation observed from different sources and to complete it as our comprehension on the situation increases, that is to say as the process of information fusion is operating. As such, while revisiting the JDL model (see section 2.2.1), James Llinas, et al wrote in 2004 [Llinas 2004]: "we see ontology as aiding the fusion community in moving ahead with Level 2 and 3 capability development because it will provide adequate specificity in defining the L2, L3 states and the relationships within and among those states". In that paper, they propose to extend the JDL model to include among others: remarks about the need for, and exploitation of, an ontology-based approach for data fusion process design.

Ontologies also improve the power of reasoning thanks to inference mechanisms. Knowledge expressed using OWL can be logically processed, i.e., inferences can be made upon it. A major concern for making our data more useful in information fusion (as it is the case for data living in the Web) is to have them behave in a consistent way when they are combined with data from multiple sources. The strategy of basing the meaning of our constraint terms on inferences provides a robust solution to understand the meaning of novel combinations of terms. A concrete example of the inference mechanism is the easy way to access knowledge through queries. The relationships linking concepts (i.e. abstraction and composition) can be used to extend or refine queries. For example, a search for information about helicopter will derive a search for information for CH-146 given that CH-146 is a type of helicopter. So, information about CH-146 will be retrieved even if they do not contain the word helicopter. More complex relations and their properties could be exploited to process queries and draw inferences leading to more relevant results. For instance, a query that would search for information about an aircraft that has horizontal rotors should lead us to a derived research about helicopter, once again! Finally, in 2007, in [Bossé 2007], the author wrote "ontologies as formal theories will become a significant and fundamental element of the mathematical foundation of information fusion".

Moreover, ontologies present a growing interest to deal automatically with new kinds of information that begin to exceed military sensor systems, such as web sites, public media, blogs, anonymous tips, direct human sources, which are expressed in natural language. Especially in military operations, critical information comes also from the commander's intent, orders, doctrines and directives, which flow as free text elements. These new kinds of information can be of crucial importance in understanding a situation and predict/evaluate possible threat from a political/geographical/etc. context. However, many terms used in natural language have several distinct meanings. Therefore, to access and analyze automatically this type of data, these terms need to be tagged with metadata, which may tell information fusion systems how to access, process and understand them. This issue is completely solved when dealing with Semantic resources. We recall here that an ontology constrains the semantic interpretation of the terms employed in natural languages by providing formal definitions. This is called ontological commitment and means mapping between ontology terms and their intended meanings. In such a way, ontologies support content extraction.

Benefits of using ontologies can also be found for instance on systems interoperability (by providing a knowledge level description of a domain that can be mapped to heterogeneous data or information sources). As written in [Bossé 2007], "One main benefit typically cited for basing the development of an information process on an ontological footing is: interoperability with other local and also external processes, which leads also to shared understanding". Another quote taken from a NATO report on Semantic Interoperability [RTG-044 2011] says that ontology support "flexible mediation of data and information between heterogeneous systems, such as C4IS and decision support systems".

For all these reasons, there has been since 2003 a growing interest of ontology-based technologies within information fusion. This is particularly shown by the increasing number of papers in international conferences. For instance during the 6th International Conference on Information Fusion, held in Australia in 2003, was organized for the first time a special session about ontology within data fusion. This special session was entitled "Ontology Needs and Issues for Higher-Level Fusion". Since 2003, many approaches have started to create ontologies (e.g. [Boury-Brisset 2003] which reviews military ontological models, or [Little 2006] for a formal ontological analysis of threat), to analyse potential use cases (see for example [Kokar 2006]) and to use them in their fusion process (e.g. [Smart 2005] in the domain of humanitarian operations, or [Laudy 2009] for enhancing observations association).

4.7 Ontologies as a Silver Bullet - What about Uncertainty?

Through this chapter, ontologies tend to be viewed as a silver bullet to address the issues and needs of information fusion. Indeed, ontology engineering is becoming a major aspect of research in this area.

However, recalling Chapter 2 and Chapter 3, one major characteristic of the information handled is uncertainty. In the case of fusion systems, we have found that these uncertainties are caused, for instance, by lack of knowledge about the environment, unreliable sources of information, shortcomings in data produced by sources, unknown or unexpected results of actions, or unknown intentions of other communication partners. In [Costa 2010], the author identifies three fundamental requirements for a representational framework in support of effective information fusion. The first requirement is a rigorous mathematical foundation. Then, this framework should be an efficient and scalable support for automated reasoning. Finally, it comes to the ability to represent intricate pattern of uncertainty. Current ontology formalisms deliver a partial answer to the two first items. We examine here the third requirement and determine if OWL ontologies in nature are able to deal with this fundamental need of fusion applications.

We have seen previously in this chapter that OWL has its roots in its own web language predecessors (i.e. XML, RDF(S)), and in traditional knowledge representation formalisms (Description Logics) that have historically not considered uncertainty. As such, XML, RDF and thus OWL are declarative languages that do not support uncertainty. Semantic Web standards assume indeed a bivalent logic. Bivalent means that statements are either true or false; no third possibility, such as unknown, exists, nor anything in between true and false. In other words, currently designed ontologies contain only concepts and relations that describe asserted facts about the world. According to [Little 2008], this means that the ontology itself is not uncertain in nature, but rather presents an *a priori* model of the world, which has to be taken as true by its users.

We agree with [Hois 2009], that specifications of a domain can be strict and well-defined in ontologies and thus are not primarily supposed to represent uncertainties. However, even though the ontology itself may not be affected by uncertainties, a system instantiation of it is or may be. As a matter of fact, as soon as an ontology is instantiated in a system, different types of uncertainties arise. Facts of a domain as specified by an ontology can either be true or false.

However, there are some indications (Open World and Not Unique Name assumptions) which suggest that these technologies have been designed to capture a minimum of the uncertainty inherently present in knowledge. Indeed, there are important pieces of evidence that uncertainty is an essential factor of ontology engineering [Costa 2005]. First, OWL ontologies are relying on the Open World Assumption (see also section 4.3.1.1) and thus, they do not include the principle that not being represented entails non-existence. On the contrary, it means that queries about which there is insufficient information in an ontology to be proved cannot be assumed as being false. Secondly, the Unique Name assumption means that every resource name of a knowledge base denotes a distinct resource. According to [Klinov 2011], its rejection helps to capture uncertainty about identity of objects, i.e., two constants can denote the same object. This is a clear sign that uncertainty is an intrinsic component of ontology engineering. Moreover, thanks to the hierarchical structure of ontologies, concepts may be modelled with more or less general definitions, thus representing the level of knowledge of a user. This feature is related to the uncertain aspect capturing vague and imprecise information. Thus, OWL ontologies do provide some limited support for capturing uncertainty. But, these do not allow to quantify or measure that uncertainty.

This review leads us to conclude that OWL language has no built-in support for quantifying uncertainty. As a consequence, one of the major limitations of traditional ontology formalisms is the lack of consistent mathematical support for uncertainty. Let's recall that in many cases, it is preferable to store a piece of information even uncertain, rather than to interpret its contents in a restrictive manner, which may lead to store erroneous pieces of information.

This lack in ontologies has spurred the development of approaches that try to represent and reason over uncertainty. This research area is however still in its infancy stage and we propose a state-of-the-art of these approaches in next Chapter.

Chapter 5

Semantic Fusion and Uncertainty

Contents

During the last decade, many approaches considering uncertainty have been proposed by the Semantic Web community. Indeed, there has been recently a plethora of demands for uncertainty reasoning technologies among Semantic Web researchers and developers. In the meantime, attempts to fill this gap can be also found within the mathematical belief functions community. Without talking already of Semantic Web technologies, they propose approaches to take into account semantic hypotheses or semantic degrees of beliefs. However, rare are the approaches in the fusion community that combine both semantic technologies and uncertainty reasoning techniques.

In this chapter, we outline some of these approaches in order to position our work. A focus is particularly made on the design choices for representing elements of uncertainty that are needed to support reasoning for the Web. Considering each design choice, the underlying motivations regarding their respective use cases are presented. The existing reasoning processes are also presented and organized through their underlying mathematical theory.

We will see that some approaches aim to quantify the degree of overlap or inclusion between two concepts, aim to assess the gradual truth about the type of an instance or aim to handle inconsistency in OWL ontologies and ontology mapping. The mathematical underlying theory is often the probability or fuzzy sets theory, but is seldom the evidential one.

However, the reader has to consider that this is not an exhaustive state-of-the-art, since new approaches dealing with both uncertainty and the Semantic Web come up every day. Moreover, we focus here on concrete approaches that have been implemented on top of RDF or OWL languages. In other words, theoretical approaches on extensions of Description Logics or on Logic Programming formalism are not dealt in this thesis. However, the reader could easily find some of them on the Web or in the following articles: [Lukasiewicz 2008, Straccia 2008, Predoiu 2009].

5.1 Recent Efforts in the Semantic Web Community

As previously mentioned, many researchers are currently trying to enhance ontological capabilities to fill the gap of uncertainty representation. To illustrate this, let's mention that the World Wide Web Consortium (W3C) has set up the Uncertainty Reasoning for the World Wide Web Incubator Group - URW3 XG[1] - in 2007, in order to explore and better define the challenges of reasoning with and representing uncertain information in the context of the World Wide Web. Another illustration of this new-found interest is the ISWC's URSW (Uncertainty Reasoning for the Semantic Web) workshop series[2] held as part of the annual International Semantic Web Conference since 2005.

The important interest in the matter for this community is largely motivated by the broad range of possible use cases found on the Web. For instance, the fusion of information on web sites is a large issue on the web. As a matter of fact, like the current Web, the Semantic Web is going to contain controversial pieces of information coming from different web sources. This can be handled by associating every web source a value (belief or probability) describing its degree of reliability. As a result, the combined knowledge has to encode this uncertainty and allows further inference on it. This issue is also related to the recurrent notions of trust and provenance in Web. Other reasoning challenges specific to the Web encompasses the discovery of Web Services or order processing via Web Services and would benefit of uncertainty management progress in the field. Considering more specifically the ontology technical domain, one can found many use cases in the domain of ontology learning and ontology engineering ([Keet 2010]), automatic inference ([Nikolov 2008]) or ontology mapping ([Ding 2006, Mitra 2005]). Ontology mapping provides also many challenges for uncertainty reasoning. It indeed assumes that ontologies are not themselves uncertain, but uncertainty is produced when a set of ontologies are mapped together. Finally, the Web is inherently related to natural language processing (with the user queries or with web pages content) and as such vague expressions need to be reflected in ontologies.

5.1.1 Representation Choices for Uncertainty in Ontologies

In addition to being logically related, the concepts of an ontology are generally also uncertainly related. As such, in this subsection, technical aspects of representing uncertainties in ontologies are investigated in more details. Existing approaches define either uncertainties within an ontology or across different ontologies. Within one ontology, the uncertainty definition can be made in various manners: either in annotations, or in reified statements, and either using semantic labels, or numerical values or intervals of values. In this section, we review the different design choices that can be encountered for that goal.

A preliminary remark is that uncertainty is often perceived as something beside

[1] http://www.w3.org/2005/Incubator/urw3/
[2] http://c4i.gmu.edu/ursw/2011/

"normal" knowledge. It is often characterized as meta-knowledge (as well as the origin of the fact, the time it has been asserted, etc.) and generally the representation of meta-knowledge in ontologies has often been neglected so far.

Secondly, the representation of uncertainty may be required to represent uncertainty inherent in the data, but also uncertainty related to the processing of data and the delivery of processing results, as underlined in [Laskey 2008]. On one hand, the objective of uncertainty representation is to provide as much as possible a standardized common interchange syntax so that people can identify and process this information quickly. On the other hand, the representation should determine how to represent the meta-information on processing and decide how detailed the meta-information would need to be and where it should reside.

5.1.1.1 Ontological Elements affected by Uncertainty

In [Hois 2009], the author reviews all the ontological constructions (see section 4.3.3) and how or in which cases they can be affected by uncertainty. Studied ontological constructions are class constructions, property constructions, class restrictions and instance constructions. The author especially separates the cases where only one ontology is taken into account or the case where uncertainty come across different ontologies which is typically the case of ontology mapping.

In the first case, the author considers that classes are meant to be well defined; the definitions of subclasses should therefore be strict. For example, the class "Man" is subclass of "Human" and that's it. By the way, the introduction of union and intersection constructions to propose new classes allows flexible relationships in class definitions and the latter should therefore not be influenced by uncertainty. Even if the author in [Hois 2009] considers that classes should not encompass epistemic uncertainty, uncertainty can still arise from classes when handling fuzziness notions. This is the focus of numerous researches (e.g. [Bobillo 2008, Ding 2006]).

Property constructions simply define the domain and range of classes that are related with each others[3]. Here again, the general definition of possible relations between instances is strictly specified and should not be affected by uncertainties.

Class restrictions (*allValuesFrom*, *someValuesFrom* and cardinality constraints) introduce flexible definitions in ontologies and additional uncertainties are here not indicated or not necessary.

Finally, the instance constructions are the main constructions that suffer from uncertain information. These constructions include both instantiations of classes and of properties. An agent instantiates indeed an ontology to represent its environment, and as we have previously underlined it (see section 2.3.2): this is subject to various kinds of uncertainties. This environmental information is subject to inaccuracies, incompleteness, ambiguity and incorrectness, due to noise, unreliable sources, or source limitations. As such, environmental entities are uncertainly instantiated as a specific class or properties [Hois 2009, Klinov 2008, Laskey 2011].

[3] we are not here referring to relations between instances but only to their definitions, which are different things

For example, we may have an uncertainty considering "Sam" may be an instance of the class "Man" or "Woman".

While considering the case of ontology mapping, relations - enabling the link between two ontologies - appear more likely to be affected by uncertainty. As such, construction artifacts such as: *equivalentClass* and *equivalentProperty* or more complex user builtin relations [Mitra 2005] between different ontologies will have to embody a certain degree of uncertainty.

5.1.1.2 Structural Design Choices

The previous section has focused on which ontological elements are affected by uncertainty, we now are going to present how this can be practically formulated and structured in ontological constructions. Existing works have made various design choices that we review here.

First a quite intuitive and convenient way to assign meta-information to RDF statements is to use the notion of reified statements or its extension through named-graphs. The RDF reification vocabulary is designed to talk about statements-individuals that are instances of `rdf:Statement`. A statement is an subject-predicate-object triple (see section 4.3.2.3). Reification in RDF is used to put additional information about this triple, such as the source of the information in the triple, for example. This design choice has been adopted by namely the following approaches: [Carroll 2005, Dividino 2009].

Similar approaches integrate meta-knowledge as OWL annotations. One of the differences with the previous method is that these additional statements are explicitly stated in the OWL ontology with the type *owl:AnnotationProperty*. There are some annotations predefined by OWL, such as `owl:versionInfo`, `rdfs:label`, etc. This method was adopted by [Klinov 2008] with the use of OWL version 1.1 axiom annotations to associate probability intervals with uncertain OWL axioms.

However, through these two visions, the meta-knowledge become implicit notation, with no semantic consequences; the meta-knowledge is just syntactic part of the ontology, which is totally different from a formally extended model of RDF or OWL which provides explicit notation of meta-knowledge. As such, in our opinion, one of the drawbacks is the poor expressiveness of uncertainty. The other disadvantage, which derives from the first one, is that as the uncertainty representation is not "constrained"; meta-knowledge can be expressed on every statement: either assertions, terminology or logical axioms, etc. One can consider this as an advantage, while others immediately think about the complexity it induces on the dedicated reasoning process to apply. On the other hand, it retains upward compatibility with existing usage of the language and corresponding tools and methods.

Finally, most of the existing approaches do not use simple RDF or OWL annotations. As such, they formally propose extended models of RDF or OWL. For example, [Costa 2006], [Ding 2004], [Bobillo 2008] or [Essaid 2009] propose some sort of upper ontologies defining artifacts of uncertainty and that are independent from the domain ontology.

5.1.1.3 Degrees of Uncertainty Assignment Choices

Regardless the structural design choice, at one moment the user (a software agent or a human) has still to assign a degree of certainty to the uncertain proposition. This is sometimes done very simply through linguistic metadata, as for example in the BFO upper ontology [Arp 2008] (only for relations). Otherwise, the degree of uncertainty takes most often the form of a single numerical value. This numerical value can have different semantics: percentage, probability, belief, etc. For example, the foundational ontology SUMO [Niles 2001] provides a relation *ProbabilityRelation* that assigns a percentage to the probability of an event. Intervals are also used to assign uncertainty values, as for example in [Klinov 2008] where probability intervals are assigned to uncertain OWL axioms. Finally, functions can also be described in ontologies to represent the possible uncertainty values. It is the case, for example, of fuzzy functions in [Bobillo 2008] (for example, the concept *LessThan18* is represented by a crisp function with arguments (0, 100, 0, 18)).

One can also wonder how and by who these degrees of uncertainty are assigned. For example, considering agent-based uncertain instantiations of classes and relations, degrees of uncertainty are logically defined by belief values. This belief may then, for instance, be defined by the Dempster-Shafer theory of evidence (see section 3.3), but also by probability theories (see section 3.2.1). The numerical values of beliefs can then be provided by "experts", statistics, or averaged probabilities over class restrictions, etc. Considering ontology mapping, numerical values reflect the similarity of related classes or relations. These values can be defined manually by developers that provide mapping relations themselves or by the system that automatically detects mapping relations.

5.1.2 Existing Approaches for Reasoning on Uncertainty in Ontologies

Uncertainty is encoded in ontologies (in various ways as we have seen in the previous section), so that users (software agents or humans) can exchange this information. But above all, the uncertainty is encoded so that software can process the data appropriately towards a specific goal. In order to process these uncertain data, a software agent has to rely on a mathematical uncertain theory according to the way uncertainty values have been represented depending on their final objective. The reasoning through uncertainty combination or inference calculations is obviously always provided outside ontological structures. This means that this reasoning process is completely disjoint from the automatic logical inference presented in section 4.4. The literature contains several probabilistic generalizations of ontology languages and as such many reasoning techniques are based on probabilities or Bayesian networks. Obviously, other approaches are using fuzzy logics, rough set theory and the Evidential theory (see section 3.2). This section goes briefly over the existing approaches and provides references to the interested reader.

5.1.2.1 Reasoning through the Probabilities

The Probabilities is surely the most known mathematical theory dealing with uncertainty. Not surprisingly, many current researches trying to fill the uncertainty gap in ontologies are based on this theory and more specifically on the Bayesian networks, which are a quite convenient way to represent and reason graphically on a set of variables and their conditional dependencies.

In particular, Costa in [Costa 2006] suggests a probabilistic generalization of OWL, called PR-OWL, whose probabilistic semantics is based on an extension of Bayesian networks: multi-entity Bayesian networks (MEBNs). MEBN is based on a Bayesian logic that combines first-order logic with Bayesian networks. The probabilistic upper ontology, PR-OWL, provides OWL constructs for representing these MEBNs. In order to use it, one has to import the PR-OWL ontology into an ontology editor (e.g. Protégé). Then, one can start constructing domain-specific concepts using the PR-OWL definitions to represent uncertainty about their attributes and relationships according to the MEBN model. PR-OWL is one of the most quoted works in the literature on Uncertainty within OWL ontologies, however, the initial version fell short in several important aspects of the compatibility with OWL domain ontologies. As a matter of fact, PR-OWL was offering OWL constructs for representing extensions of Bayesian networks, but the variables involved in these networks were not referring to a semantic description in a previous domain ontology. As such, PR-OWL was more a semantic description of MEBN rather than a probabilistic extension of standard OWL domain ontologies. These shortcomings were especially the focus of the latest version, PR-OWL 2 [Carvalho 2010], proposed in 2010. It indeed proposes new features and describes the process of constructing a PR-OWL 2 ontology using an existing OWL ontology as a starting point.

BayesOWL [Ding 2004, Ding 2006], proposed by Ding et al., is another well known and early work in the domain, to model uncertainty in OWL ontologies through Bayesian networks. It provides a set of rules and procedures to express OWL ontologies as Bayesian networks by adding a second ontology to this translation which declares the probabilistic relationships. The generated Bayesian network, which preserves the semantics of the original ontology and which is consistent with all the given probability constraints, supports ontology reasoning, both within and across ontologies, as Bayesian inferences. It is used to quantify the degree of the overlap or inclusion between two classes. In [Pan 2005], Pan et al. also describe an application of the BayesOWL approach in ontology mapping.

Considering still Bayesian networks for mapping purposes, we can quote the work of Mitra et al. in [Mitra 2005], which describes an implemented technique, called OMEN (Ontology Mapping ENhancer), to enhance existing ontology mappings by using a Bayesian network representing the influences between potential concept mappings across ontologies. More concretely, OMEN is based on a simple ontology model similar to RDF Schema. It uses a set of meta-rules that capture the influence of the ontology structure and the semantics of ontology relations, and matches nodes that are neighbors of already matched nodes in the two ontologies.

Yang and Calmet [Yang 2005] present another integration of the web ontology language OWL with Bayesian networks, called OntoBayes. The distinction with BayesOWL is that the random variable are multivalued (they were only true/false in BayesOWL networks) and that the authors make use of dependency-annotated OWL to represent uncertain information in Bayesian networks, which is a more generic dependency modeling than the set-theoretic approach found in BayesOWL. The work also describes an application in risk analysis for insurance and natural disaster management.

There are also the so-called probabilistic extensions to Description Logics that want to be seen as an alternative to more traditional Bayesian approaches. For instance, Pronto [Klinov 2008] is a probabilistic DL reasoner prototype. Pronto is able to represent and reason about uncertainty in both, generic background knowledge and individual facts (respectively probabilistic relationships between OWL classes and relationships between an OWL class and an individual). For example, considering that a certain number of menopausal women have a risk of breast cancer, questions such as "how likely Helen has such a risk?" can be asked.

5.1.2.2 Reasoning through the Possibility Theory

Regarding the theory of Possibility, we can mention the works of [Lesot 2008] and [Coucharière 2010], which provide some possibilistic extension of the tableau algorithm. The tableau algorithm aims to decide consistency of ALC ontologies and all other standard DL reasoning problems. This extension is dedicated to the computation of the inconsistency degree of the ontology. For that purpose, the inconsistency degree of the knowledge base is computed based on the logic expressed in the base and the degree of certainty of the statements. These works are particularly interesting since they are addressed from an operational information system point of view. In that context, it should permit to determine if some elementary information can coexist (if they are referring to the same object) and to some extent they permit to interpret the situation awareness (i.e. infer additional information concerning the situation). However, the imprecision of the modeling choice of the agent can not be taken into account by this inconsistency. In other words, similarity of elementary information is not taken into account, which may be a limitation to the fusion process.

5.1.2.3 Reasoning through the Fuzzy Logic and Rough Set Theory

Fuzzy formalisms allow the representation and the gradual assessment of truth about vague information. Regarding ontologies, fuzzy approaches ([Bobillo 2008, Simou 2007]) consider classes to have unsharp definitions. Umberto Straccia has especially conducted extensive work on Fuzzy Description Logics. His joint work with Fernando Bobillo has lead to FuzzyDL [Bobillo 2008], which is one of the most succeeded tools proposing a fuzzy description language associated to a reasoning engine. It aims at representing and reasoning about a membership function specifying the degree to which an instance belongs to a class.

Approaches in [Schlobach 2007, Keet 2010] are relying on rough set theory - which considers the indiscernibility between objects. In that case, classes are not restricted to a crisp representation; they may be coarsely described with approximations. In [Keet 2010], the author is using rough classes to generate new subclasses or relations by mining an important set of instances already existing. This can be part of the ontology engineering process.

5.1.2.4 Reasoning through the Evidential Theory

Existing approaches using the Evidential theory have been most applied in ontology mapping. Among them, [Nagy 2006] incorporates the Dempster-Shafer theory into the mapping process, by combining the similarities which were originally created by both syntactic and semantic similarity algorithms, in order to improve the correctness of the mapping. [Yaghlane 2007] deals with uncertainty inherent to the mapping process, especially when interpreting and combining the results returned by different matchers.

Another recent published approach [Nikolov 2008] is concentrating on uncertain reasoning on instances of an ontology using the evidential theory and some similarity measures. Their main objective is to propose an alternative ABox inductive reasoning - by classifying individuals (determining their class- or role- memberships or value for datatype properties) through a prediction based on an evidential nearest neighbor procedure. Their reasoning addresses here another way to tackle automatic inference from a classical ontology. This automatic inference aims to derive new or implicit knowledge about the current representation of the world, on the basis of the asserted knowledge.

[Essaid 2009] transforms uncertain statements in belief networks. It focuses on translating an OWL taxonomy into a directed evidential network based on the evidence theory. This work has quite the same purpose as [Yang 2005] and [Ding 2004] by transforming the terminology of an ontology into a graphical probabilistic/evidential model. The added-value of using the Evidential theory is for Amira Essaid to be able to deal with incomplete information. However, the semantics of the variables is still here not taken into account, which would have yet enabled to consider the most conflicting hypotheses or on the inverse the implied hypotheses.

5.2 Semantic Extensions of the Evidential Theory

Within uncertain mathematical theories, some approaches tend to introduce notions of semantics behind hypotheses or behind the degrees of belief assigned to these hypotheses.

In the first case, degrees of belief are expressed through qualitative linguistic labels (which induce the use of qualitative operators) [Martin 2008]. Human experts can indeed provide more easily these labels rather than numerical degrees.

In the second case, approaches try to structure a proper frame of discernment by taking into account the semantic of hypotheses.

In fact, first of all, there are the well known operations: refining/coarsening of the Evidential Theory [Shafer 1976] that permit to take into account the different levels of granularity within hypotheses, by either reducing the size of the frame of discernment in merging some elements together, or the other way round. This mechanism is illustrated on the following figure 5.1, where Ω is a refinement of Θ and inversely Θ is a coarsening of Ω. We consider here two frames of discernment

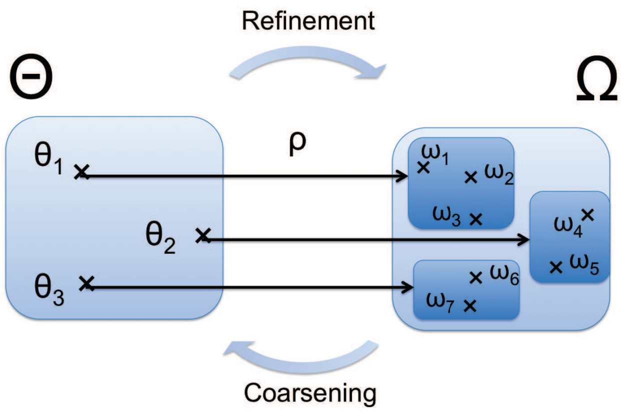

Figure 5.1: Refinement and Coarsening operations.

Θ and Ω. The idea that Ω is obtained from Θ by analysing or splitting some or all the elements of Θ is characterized by the refining operation. The refining in the schema is represented by the function ρ, which specify for each hypothesis of Θ a subset of Ω. The mass function of Ω is obtained by the following expression: $m^{\Omega}(\rho(A)) = m^{\Theta}(A), \forall A \subseteq \Theta$. The coarsening operation, which is the inverse of the refining operation, provides a way to aggregate some hypotheses. The mass function on Ω is transformed to a mass function on Θ according to the following expression: $m^{\Theta}(A) = \sum_{\substack{B \subseteq \Omega \\ A=\{\omega_i | \rho(\omega_i) \cap B \neq \emptyset\}}} m^{\Omega}(B)$

This mechanism is quite interesting because it shows that the intrinsic meaning of hypotheses - through their granularity with the other hypotheses of the frame of discernment - is of major concern. However, here the granularity should be explicited manually through inclusion relations. Another remark is that through this mechanism, one underlying assumption is that Θ has still disjoint elements (due to assumptions of the frame of discernment in the Evidential theory) and that also $\rho(\theta_i)$ has to constitute a partition of Ω. In other words, all hypotheses, in either a coarsened or a refined frame of discernment, have to be disjoint, which can be hard to obtain in practice.

Dezert and Smarandache have underlined [Dezert 2004] this issue concerning the disjointness of hypotheses within a frame of discernment. They point out that this exclusivity assumption ($H_i \cap H_j \neq \emptyset$, for further details see section 3.3.1.1) is too restrictive for many applications and is just impossible to obtain in reality. Indeed, hypotheses can have different intrinsic nature and appear vague and imprecise, as described in natural language so that exclusive hypothesis elements cannot be properly defined. This is typically the example of color palette hypotheses, as shown on figure 5.2 where a frame of discernment would be $\Omega = \{H_1 = (R)\}, \{H_2 = (G)\}, \{H_3 = (B)\}$. The trouble is that the boundaries of these hypotheses may be blurred (the lines between H_1, H_2 and H_3 can not be precisely defined), due to different color perception by any individuals.

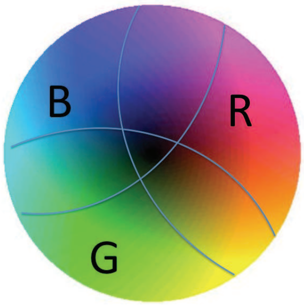

Figure 5.2: Illustrative example for motivating the use of the DSmT framework.

Therefore, they propose in 2002 an extension (called DSmT [Dezert 2004] - Dezert-Smarandache Theory -) of the Belief Functions Theory, where a frame of discernment has potentially overlapped elements. In DSmT framework, one considers $\Omega = \{H_1, H_2, ..H_N\}$ be a finite set of N exhaustive elements only (i.e. elements of Ω have not to be exclusive). The hyper-power set of Ω denoted D^Ω is defined as:

- $\emptyset, H_1, H_2, ..H_N \in D^\Omega$,
- If $A, B \in D^\Omega$, then $A \cap B, A \cup B \in D^\Omega$,
- No other elements belong to D^Ω, except those obtained by the first two rules.

As a consequence, if $|\Omega| = N$, then $|D^\Omega| \leq 2^{2^N}$. For example, if $\Omega = \{H_1, H_2\}$, then $D^\Omega = \{\emptyset, H_1 \cap H_2, H_1, H_2, H_1 \cup H_2\}$. DSmT can thus be seen as a generalization of the Dempster Shafer theory. However, as a consequence, the complexity in terms of calculi is higher than the Dempster-Shafer theory and is often a severe critic of this approach. Moreover, taking into account that all elements are overlapped is to our mind a too much relaxed assumption. Indeed, the previous example of color palette is totally adequate for the DSmT, however, we often face cases where only some elements are overlapped and the others are disjoint. Therefore, a finer modeling of the frame of discernment should be often made possible relying on the semantic description of its elements.

This last observation has been especially the trigger for the introduction of the reduced model of DSmT - D_r^Ω [Martin 2009]. It is indeed possible to constrain and alleviate the hyper-power set D^Ω from certain elements which would represent an empty intersection. This space is therefore much larger than 2^{Ω} but much smaller than $2^{2^{\Omega}}$. However, the constraints applied to this space should be already known by the user and entered manually before any processing. It would indeed be interesting to automatically infer and apply these semantic constraints on the frame of discernment.

5.3 Conclusion, Opinion and Remarks

In this chapter, we have reviewed the different methods and approaches that relate to both uncertainty and ontologies (extended to semantics). We have seen that some aspects of these techniques are shared by many approaches even for different goals. However, some technical choices are turned towards specific goals on which we do not run.

Let's remind that we are focused on information fusion applications that deal with real world environment and that the uncertainty comes from some agents that observe, analyze or make inferences on the knowledge. As such, our immediate objective is not to propose enhanced ontology engineering tools such as ontology mapping or inference.

That is why, according to this chapter, instantiations are the constructions that suffer from uncertainty. Indeed, we have previously shown that ontology constructions were only partially affected by uncertainties and considering our purpose, uncertainties are not supposed to be modelled by some class or properties definitions nor to be indicated by logical formalisms. As a matter of fact, while a relation may hold or not, its definition itself is strictly specified. As such, an instance may have uncertain classes or properties, but the general definition of classes or possible relations between instances is not affected by uncertainties. Moreover, we do not consider that uncertainty has to be represented across different ontologies since we are not handling ontology mapping techniques.

We have also introduced some methods that can represent uncertainty on any triple without constraining this representation with particular ontological constructions. Even if this is very flexible, it may introduce a high complexity to reason on all kinds of constructions. Moreover, through this method, uncertainty is not represented semantically, and thus places uncertainty besides the "normal" knowledge, whereas we prefer to consider uncertainty be included within the characterization of the fact itself, rather than additional information. Therefore, we promote semantically-driven approaches for representing uncertainties in ontologies. We think indeed that the ontology should then provide definitions of uncertainties by specifying syntax and semantics for modeling them. Thus, uncertainty is not kept implicit.

A major part of this chapter has also focused on the presentation of the different researches for handling uncertainty reasoning.

Not surprisingly, there are many probabilistic approaches. However, only PR-OWL [Costa 2006] or Pronto [Klinov 2008] are relevant to our objective of handling uncertainty at the instance level (other mentioned approaches were more related to terminological uncertainty). Yet, as underlined in the previous chapter, probabilities suffer from the lack of ignorance and imprecision management in comparison to the Evidential theory. In addition, the MEBN community, on which PR-OWL relies, is not wide enough - to our knowledge - to be considered as an emerging standard. Thus it represents a major difficulty to manipulate this tool. Regarding, the Evidential theory, the existing approaches were either related to ontology mapping or ontology engineering, but none were handling uncertain instances. Finally, even if it could be interesting to take into account fuzzy aspects of hypotheses especially those formulated by human sources, this is not related to our initial priority purpose.

Last but not least, we have underlined that taking into account the semantics of hypotheses in the Evidential theory implies to construct a proper frame of discernment still consistent with the exclusivity assumption while still being consistent with the global granularity of the hypotheses. Indeed, introducing semantic hypotheses can lead to hypotheses that overlap each other, have the same meaning or on the contrary have different intrinsic semantic natures. However, contrary to the presented approaches, we believe that it should be possible to still rely on the Evidential theory. As such, we believe that each hypothesis is clearly defined in the ontology so that exclusivity can be roughly expressed in the Evidential theory. The semantics of the hypotheses should be thoroughly taken into account during the construction of the frame of discernment.

Part III

A Framework for the Fusion of Semantic Beliefs

Chapter 6

Modeling Semantic Beliefs

Contents

Within information fusion applications, uncertainty arises from the observations, analyses and inferences that the agents are making over the knowledge of a situation. In the previous chapter, we have come to the conclusion that uncertainties are caused by agent-based instantiations of the ontology. In other words, we consider that instantiations are the constructions in an ontology that suffer from uncertainty: an instance may be of an uncertain type and may be related to uncertain things. An instance has here to be understood as either an instance of a class (sometimes called an individual) or as an instance of a property. The general definition of concepts (classes and properties) is not affected by uncertainties. From the previous chapter, we also decide to consider uncertainty as being included within the characterization of the fact itself. This means that an ontology should provide definitions of uncertainty by specifying syntax and semantics for modeling it. In such a manner, uncertainty is not kept implicit.

Since currently no standard exists for the representation of uncertainty in ontologies and even less considering our above needs, we propose, in this chapter, a new semantically-driven approach for representing uncertainties in the instantiation part of ontologies.

As previously specified (see section 3.1), it is worth recalling that we consider and restrict uncertainty to epistemic uncertainty - due to lack of knowledge of the agent - and inconsistent pieces of information.

In this chapter, we first specify where the uncertainty is represented in ontologies and what is considered to be deterministic. Then, we formally provide the definitions and semantics for the different artifacts used to represent that uncertainty in ontologies. These artifacts are gathered in an upper ontology called the *DS-Ontology*. These artifacts permit to represent *semantic beliefs*. This notion refers to one or several uncertain candidate instances - described semantically through a domain ontology -, as well as their associated degree of belief and reporting source. A schematic example of one *semantic belief* is given on figure 6.1.

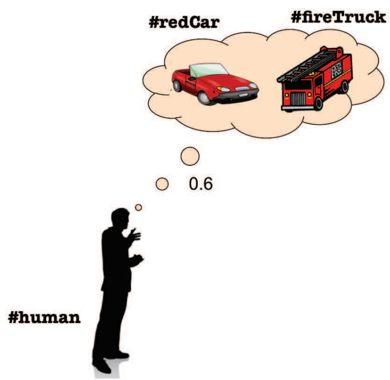

Figure 6.1: An example of one semantic belief. The uncertain candidate instances are here the `#redCar` and `#fireTruck` instances. The degree of belief is 0.6 and the reporting source is the `#human` instance.

No doubt that these *semantic beliefs* are inspired by belief functions from the Dempster-Shafer theory. We will see, however, that the former has weaker assumptions concerning their semantic set relations.

Finally, examples are provided to illustrate the use of the *DS-Ontology* for a particular domain ontology.

6.1 Uncertain Instantiation Part of a Domain-Ontology

As seen in section 4.2, it is common to draw a distinction between the *T-Box* and the *A-Box* of an ontology. The *T-Box* refers to the terminology of an ontology. It is a finite set of axioms describing the classes, properties and datatypes in a domain of discourse. Whereas, the *A-Box* refers to the instances used to populate the *T-Box*. It is a finite set of assertions of classes and properties. Considering OWL2-DL, instances are strictly separated from the classes and properties defined by the ontology.

However, within the instances of the *A-Box* itself, it is also important to distinguish instances that are yet part of the terminology from those that are mere "facts". The first ones can be called *terminological instances* and are those instances needed as conceptual entities for defining the "vocabulary" of the domain of interest. Others are *factual instances*. They represent "real world" entities, which are specific examples of the structure defined by classes and properties. For example, in an ontology about colors, the palette of colors (blue, red, green, etc.) that commonly exists are related to *terminology instances*, whereas the specific color that an artist has mixed and applied on his painting is not part of the ontological terminology about color, but is a more a *factual instance*. The exact border between these two types of instances is sometimes blurred and kept implicit; however, it still exists and is a design decision in any particular case. Generally and by extension, we usually use the term *instance* to refer to these *factual instances*.

Within our framework, as mentioned before, it appears natural that it should be the instances (and more particularly factual instances) that embody uncertainty. This is illustrated in figure 6.2.

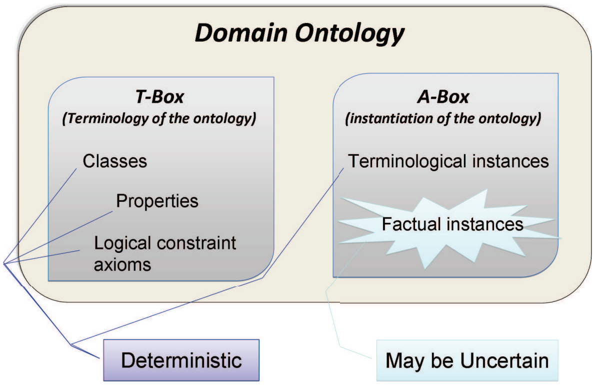

Figure 6.2: Where does uncertainty rely?

In our framework, the ontological terminology is fixed (i.e. it is modelled *a priori*), then agents populate it as the system runs according to their beliefs.

Therefore, on one hand, classes and properties, defined in the terminology of the ontology, are deterministic.

On the other hand, instantiation part of the ontology may be uncertain, but is not restricted to be uncertain: it is a pair composed of a deterministic and an uncertain part. The deterministic instantiation part contains instances that are known to be true. The uncertain instantiation part contains instances for which we only have a degree of evidence regarding a phenomenon of interest. We may have several uncertain instances referring to the same phenomenon that may be in conflict. However, we know that these instances are associated to the same phenomenon (an object, an action, a property, etc.). Each instance may either be part of the deterministic or the uncertain *ABox* but not both. The uncertain instantiation part of the ontology refers to uncertain instances of classes (i.e. individuals) and to uncertain instances of properties.

In the following, we will refer to the ontology of interest/discourse as the domain-ontology in opposition to the further described *DS-Ontology*. Descriptions of uncertain phenomena - through hypothetical instance - are relevant to that domain-ontology.

6.2 The *DS-Ontology*, an Upper Ontology

For the stake of uncertainty representation within ontological instances, a specific ontology has been created [Bellenger 2010, Bellenger 2011a, Bellenger 2012a], called the *DS-Ontology*. This name comes from the acronym of Dempster-Shafer. It is a wink to the Dempster-Shafer theory on which our representation borrows the main concepts. However, one has not to expect a perfect match between the concepts of the Dempster-Shafer theory and the one of the *DS-Ontology*, since the latter relaxes some assumptions concerning the frame of discernment (which is in the following referred to as the universal set of candidate instances). The link between the two formalisms will be the subject of Chapter 8.

The *DS-Ontology* is an upper ontology since one can use it to represent uncertainty in every area of knowledge. It is non domain specific. It can be also assimilated to a meta ontology, since it specifies the knowledge-structuring construct for uncertain instances. However, it should not be confused with top-level ontologies, which model the most basic fundaments of the world.

The *DS-Ontology* and domain-ontology are used in combination to instantiate the domain-ontology in an uncertain manner. This process is depicted on figure 6.3.

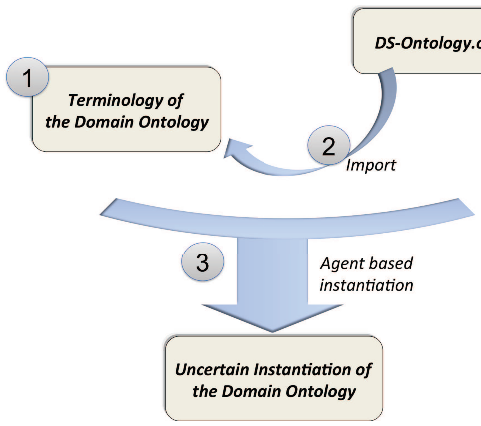

Figure 6.3: Process for using the *DS-Ontology* in combination with any domain-ontology.

The preliminary step is obviously to define the terminology of the domain-ontology. Then, one has to import the *DS-Ontology* in the domain-ontology. This step refers to the physical OWL import of the *DS-Ontology* file: "DS-Ontology.owl". For that, the `owl:imports` statement is used. Finally, an agent (either software or human) can start instantiate it while using the artifacts defined in the *DS-Ontology* to encapsulate uncertain beliefs about the domain of interest.

An important point for the design of the *DS-Ontology* is the definition of relevant meta knowledge vocabulary (candidate instance, sources, beliefs, etc.) and their suitable interpretation. Implementation of the *DS-Ontology* has been performed with the OWL2-DL language [Hitzler 2009] and is conveyed in appendix B.

However, for ease of explanation and transposition to other languages, we expose in the following sections its formal vocabulary and semantics.

6.2.1 Vocabulary

We expose here the formal definitions of the *DS-Ontology* vocabulary.

Definition (Candidate Instance).

A *candidate instance* is an instance of the domain ontology that is a proposed explanation for a phenomenon. It is an hypothetical ontological instance. It may either be an instance of a class (i.e. an individual) or an instance of a property. A *candidate instance* is denoted by I_j.

The ***universal set of candidate instances*** is the finite set of all possible *candidate instances* for a given phenomenon. It is denoted by $\Psi = \{I_j\}_{j=1..N}$ when there are N *candidate instances*. In case of a set of candidate individuals, there is no restriction on the uncertainty of the individual: it can be any individual of the domain ontology. In case of a set of candidate instance of properties, the subject and object of the uncertain property are fixed: only the predicate is uncertain. The difference of uncertainty restriction is illustrated of figure 6.4.

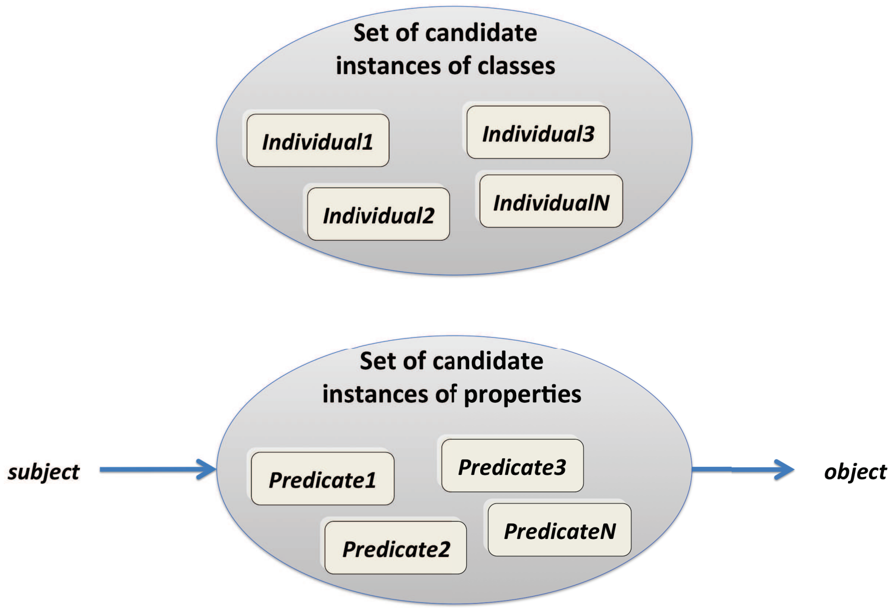

Figure 6.4: Set of candidate instances of classes versus of properties.

The ***power set of candidate instances***, 2^Ψ, is the set of all subset of Ψ including the empty set. For example, if $\Psi = \{\texttt{\#car, \#vehicle}\}$ then $2^\Psi = \{\emptyset, \{\texttt{\#car}\}, \{\texttt{\#vehicle}\}, \Psi\}$. A ***set of candidate instances***, A, is an element of 2^Ψ.

Definition (Reporting Source).

A *reporting source* S is anything that can indicate its state of belief. A device, a human or a computational process or combination could play the role of a reporting source.

Definition (Mass Value).

A *mass value* is a specific datatype whose value is a real between 0 and 1. It is symbolized by m_{value}.

Definition (DS Concept).

A *DS concept* is a 3-tuple $\{S, m_{value}, A\}$ which specifies that a *reporting source* S believes in a *set of candidate instances* A to a certain degree m_{value}. It is subcategorized in *DS class* and *DS property* when A is a set of candidate instances of classes, and instances of properties, respectively. For example, a *DS class* equals to {`#human`, 0.4, `{#car, #vehicle}`} means that a source identified by the URI local name: `#human`, assigns 0.4 as mass value to the fact that it can be either the instance `#car` or `#vehicle`.

Definition (Uncertain Concept).

For a given phenomenon, an *uncertain concept* gathers all *DS concept* items which are related to that phenomenon. It is denoted by $U = \cup_{A \in 2^{\psi}} \{S, m_{value}, A\}$. It is subcategorized in *uncertain class* and *uncertain property* when A is a set of candidate instances of classes, and instances of properties, respectively. For a given reporting source, an uncertain concept verifies that the sum of its mass values is equal to 1. More formally, we have $\forall S$, $\sum_{m_{value}} m_{value} = 1$ such that $\{S, m_{value}, A\} \in U, A \in 2^{\psi}$.

6.2.2 Modeling in OWL

The vocabulary of concepts - handled in the *DS-Ontology* - has been introduced formally in the previous section. Now, let's see how this vocabulary is concretely encoded in the *DS-Ontology*. We present in the following this encoding for each of the concepts listed above.

- *Candidate Instance*: A *candidate individual* is encoded as any individual of the domain ontology. A *candidate instance of property* is encoded as any instance of property defined in the domain ontology. Thus, they are not associated to a particular class or property in the *DS-Ontology*.

- *Reporting Source*: It is represented as an instance of the class `Reporting_Source` in the *DS-Ontology*.

- *Mass Value*: It is translated as a user-defined OWL datatype `SpecificUncertaintyDatatype`, whose value is restricted to an xsd:double ranging from 0 to 1.
- *DS Concept*: It is encoded as an instance of the class `DS_concept` in the *DS-Ontology*.
- *Uncertain Concept*: It is an instance of the class `Uncertain_concept` in the *DS-Ontology*

The OWL2-DL language is a syntactic variant of description logic (see section 4.2). We will use this notation to describe the semantics attached to the *DS-Ontology*. Moreover, the reader can refer in the meantime to the figure 6.6 on page 88 for ease of understanding, since it proposes a schematic view of the basic relations between classes and properties of the *DS-Ontology*. For recall, the syntactic features of description logic have been recapped in figure 6.5, where C and D are concepts, a and b are individuals, and R is a role. Ontologically speaking, C and D are classes, a and b instances, and R a property.

Symbol	Description	Example	Read
$\top$	All concept names	$\top$	top
$\bot$	Empty concept	$\bot$	bottom
$\sqsubseteq$	Concept inclusion	$C \sqsubseteq D$	all C are D
$\forall$	Universal restriction	$\forall R.C$	all R-successors are in C
$\exists$	Existential restriction	$\exists R.C$	an R-successor exists in C
$\sqcap$	Intersection or conjunction of concepts	$C \sqcap D$	C and D
$\sqcup$	Union or disjunction of concepts	$C \sqcup D$	C or D
$\equiv$	Concept equivalence	$C \equiv D$	C is equivalent to D
$\neg$	Negation or complement of concepts	$\neg C$	not C
:	Concept assertion	$a : C$	a is a C
:	Role assertion	$(a, b) : R$	a is R-related to b
U	All role names	U	top role
$= n$	Cardinality of role	$C \equiv = nR.\top$	C has n R-successors

Figure 6.5: Conventional Notation of Description Logic.

Hereafter concept inclusions of the *DS-Ontology* are described. It corresponds to the *subclassOf* axiom in OWL language. Namespace are here deliberately forgotten for ease of readability. All these concepts are disjoint from each other.

1- `Uncertain_concept` $\sqsubseteq \top$
2- `DS_concept` $\sqsubseteq \top$
3- `Reporting_Source` $\sqsubseteq \top$
4- `Uncertain_class` $\sqsubseteq$ `Uncertain_concept`
5- `Uncertain_property` $\sqsubseteq$ `Uncertain_concept`

6- DS_class $\sqsubseteq$ DS_concept, DS_property $\sqsubseteq$ DS_concept

In the following, axioms of the form: $\exists R.\top \sqsubseteq D$ indicates that a role: R has for domain the D class and axiom $\top \sqsubseteq \forall R.C$ indicates that R role has for range the C class. The following axioms introduce roles in the *DS-Ontology* and their associated domain and range. Moreover, some axioms using roles define more deeply the semantics of concepts of the *DS-Ontology*. There are also presented here.

The five following axioms are concerned about the hasDS_concept role. The two first ones state that an instance of the Uncertain_concept class is linked to instances of the DS_concept class through the hasDS_concept property. The third one expresses that every instance of Uncertain_concept should be linked to at least one DS_concept instance. The fourth one states that an instance of Uncertain_concept which is related to a DS_class via the hasDS_concept is in fact also an instance of Uncertain_class. Respectively, an instance of Uncertain_concept which is related to a DS_property via the hasDS_concept is in fact also an instance of Uncertain_property.

1- $\exists$hasDS_concept.$\top \sqsubseteq$ Uncertain_concept

2- $\top \sqsubseteq \forall$hasDS_concept.DS_concept

3- Uncertain_concept $\equiv \exists$hasDS_concept.DS_concept

4- Uncertain_class $\equiv$ Uncertain_concept $\sqcap \forall$hasDS_concept.DS_class

5- Uncertain_property $\equiv$ Uncertain_concept $\sqcap$ $\forall$hasDS_concept.DS_property

The fact that an instance of the DS_concept class is related to an instance of the Reporting_Source thanks to the hasDS_source property is depicted by the two first axioms below. This hasDS_source property permits also to add another logical axiom (the third one) on the DS_concept class definition. This latter says that a DS_concept should have one and only one hasDS_source property.

1- $\exists$hasDS_source.$\top \sqsubseteq$ DS_concept

2- $\top \sqsubseteq \forall$hasDS_source.Reporting_Source

3- DS_concept $\equiv\, =1$hasDS_source.$\top$

An instance of the DS_concept is also related by a datatype property DS_mass to a specific user datatype massValue restricted through the fourth axiom.

1- $\exists$DS_mass.$\top \sqsubseteq$ DS_concept

2- $\top \sqsubseteq \forall$DS_mass.specificUncertaintyDatatype

3- specificUncertaintyDatatype $\equiv$ $double[\geqslant 0.0, \leqslant 1.0]$

4- DS_concept $\equiv\, =1$DS_mass.$\top$

Last but not least, an instance of the DS_concept class is also related to its candidate instance(s). Two different cases can be distinguished depending if the candidate instances are instances of class (individuals) or instances of properties. In the first case, the hasDS_hypothesis object property relates an instance of DS_class to a candidate individual (see first axiom below). The candidate individual is thus an instance of any class of the domain-ontology. Therefore, the range of the `hasDS_hypothesis` property is here not specified. In the second case, concerning candidate properties, things have been done differently. Indeed, properties are not "first-class citizens" (since OWL ontology are based on first order-logic), contrary to individuals; in other words, properties cannot be related to each others: a property cannot be the subject or object of another property. To get around this, an object property `hasUncertain_property` has been introduced. The original subject of the candidate property is the subject of `hasUncertain_property`. The domain of `hasUncertain_property` is - as its name indicates it - the `class Uncertain_property` (denoted by the second axiom below). Then, a `DS_property` instance is directly the subject of the candidate property(ies) while their objects remain unchanged.

1- $\exists \texttt{hasDS\_hypothesis}.\top \sqsubseteq \texttt{DS\_class}$

2- $\top \sqsubseteq \forall \texttt{hasUncertain\_property}.\texttt{Uncertain\_property}$

3- $\texttt{DS\_class} \equiv \exists \texttt{hasDS\_hypothesis}.\top$

4- $\texttt{Uncertain\_property} \equiv \exists \texttt{hasUncertain\_property}^{-1}.\top$

In our model, `Uncertain_concept` and `DS_concept` are classes that let grouping together collected pieces of information about an uncertain instance we want to model and reason about. It can be viewed as a reification process, where an addressable object is created as a proxy for non-addressable objects. Informally, reification is often referred to as "making something a first-class citizen" within the scope of a particular system. Reification is one of the most frequently used techniques of conceptual analysis and knowledge representation.

Figure 6.6 recaps graphically the different classes, hierarchical relations (*subClassOf*) and object/datatype properties (solid arrows) of the *DS-Ontology*.

This section has explicited formally the semantics of the *DS-Ontology* terminology introduced in section 6.2.1. This semantic expressed in OWL2-DL language has been quite successful to represent the relations between all the concepts of the terminology. However, for the completeness of this document, one can remark that there is a deficiency to represent one constraint axiom mentioned in section 6.2.1, which consequently does not appear in this section. We recall here this constraint: "For a given reporting source, an uncertain concept verifies that the sum of its mass values is equal to 1". The trouble is that with the OWL language, one cannot add this type of numerical constraint axiom: a numerical sum is not possible to represent in OWL. We can only mention that in the implementation of our systems that reason over this uncertain knowledge (see Chapter 9), we add a function in programming language to verify this constraint.

An illustration of the use of the *DS-Ontology* is given in the next section.

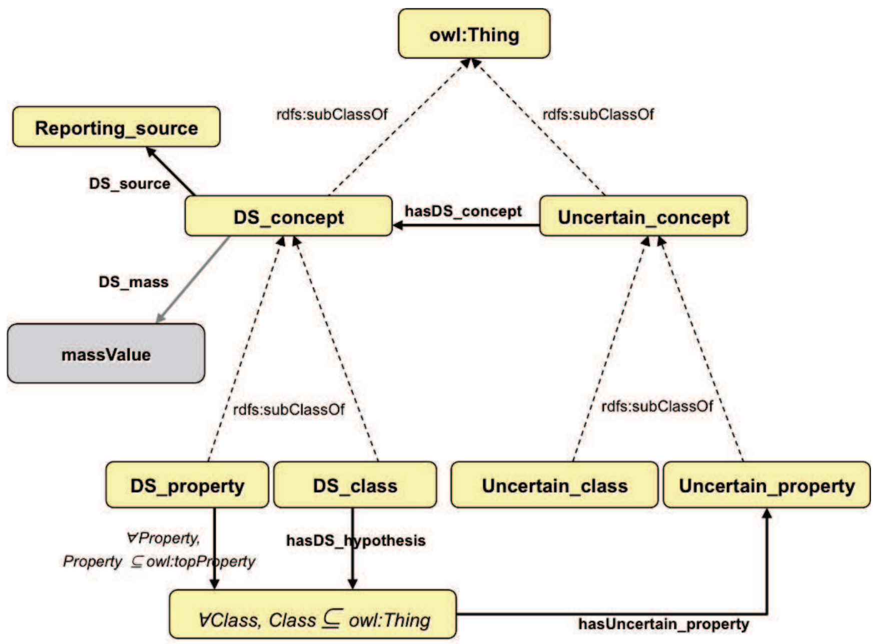

Figure 6.6: Informal schema of the DS-Ontology structure. Yellow boxes represent OWL classes. Grey ones refer to datatypes (XML ones and user defined datatype). Arrows symbolize properties. Elements in italic refer to elements of the domain-ontology.

6.3 Examples of Semantic Beliefs Representation

This section presents two examples of representing uncertain instances in ontologies. The first one is about uncertain instances of classes. The second one considers uncertain instances of properties. They are both based on the domain ontology already defined in section 4.5 of Chapter 4.

6.3.1 Uncertain candidate individuals

This example involves two distinct sources. One is a human while the other is an automatic sensor, such as radar. They both want to express that something is moving towards a specific direction; the "something" entity is the same object for both sources; however, they are not sure about what is this object.

Here, the radar source can only distinguish a land vehicle from an aircraft; it assigns here a more important degree of belief to the fact that it is a land vehicle. The second source, who is a human, is quite far away from the situation. He is assigning different beliefs to the fact that it is a red car, or a fire truck or more imprecisely that it is a vehicle on the ground.

According to the sources and to the domain ontology (see section 4.5), the universal set of candidate instances is $\Psi = \{\texttt{\#landVehicle}, \texttt{\#aircraft}, \texttt{\#fireTruck}, \texttt{\#redCar}\}$.

The sources are represented by two instances of the `#Reporting_Source` class of the *DS-Ontology*. These are `#human` and `#radar` individuals.

The sources assign mass values to their beliefs. These values are represented on the following figure.

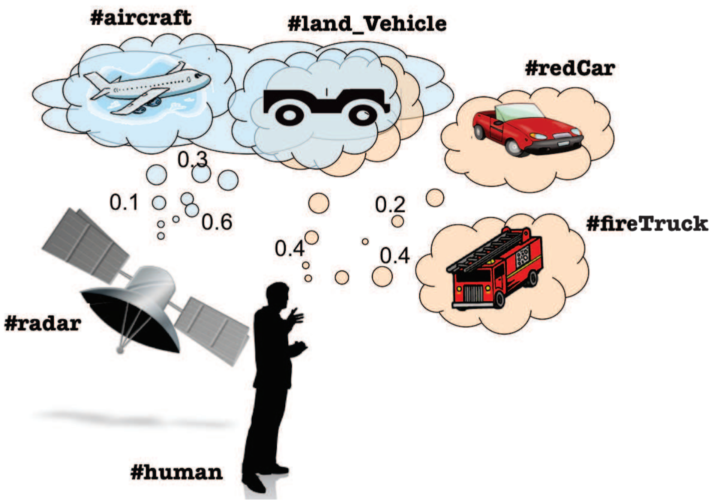

Figure 6.7: Schematic view of the example.

Finally, the *DS concepts* are the following tuples:

1- $\{$#human$, 0.2, \{$#redCar$\}\}$
2- $\{$#human$, 0.4, \{$#fireTruck$\}\}$
3- $\{$#human$, 0.4, \{$#landVehicle$\}\}$
4- $\{$#radar$, 0.6, \{$#landVehicle$\}\}$
5- $\{$#radar$, 0.3, \{$#landVehicle$,$ #aircraft$\}\}$
6- $\{$#radar$, 0.1, \{$#aircraft$\}\}$

The whole example is represented by the following axioms in description logic (see section 4.3.3 for notations):

1- #uncertain_class : Uncertain_class
2- (#uncertain_class, #ds_class_1) : hasDS_concept
3- (#ds_class_1, #human) : hasDS_source
4- (#ds_class_1, 0.2) : DS_mass
5- (#ds_class_1, #redCar) : hasDS_hypothesis
6- (#uncertain_class, #ds_class_2) : hasDS_concept
7- (#ds_class_2, #human) : hasDS_source
8- (#ds_class_2, 0.4) : DS_mass
9- (#ds_class_2, #fireTruck) : hasDS_hypothesis
10- (#uncertain_class, #ds_class_3) : hasDS_concept
11- (#ds_class_3, #human) : hasDS_source
12- (#ds_class_3, 0.4) : DS_mass
13- (#ds_class_3, #landVehicle) : hasDS_hypothesis

14- (#uncertain_class, #ds_class_4) : hasDS_concept
15- (#ds_class_4, #radar) : hasDS_source
16- (#ds_class_4, 0.6) : DS_mass
17- (#ds_class_4, #landVehicle) : hasDS_hypothesis
18- (#uncertain_class, #ds_class_5) : hasDS_concept
19- (#ds_class_5, #radar) : hasDS_source
20- (#ds_class_5, 0.3) : DS_mass
21- (#ds_class_5, #landVehicle) : hasDS_hypothesis
22- (#ds_class_5, #aircraft) : hasDS_hypothesis
23- (#uncertain_class, #ds_class_6) : hasDS_concept
24- (#ds_class_6, #radar) : hasDS_source
25- (#ds_class_6, 0.1) : DS_mass
26- (#ds_class_6, #aircraft) : hasDS_hypothesis
27- (#uncertain_class, #south) : movesTowards

The following figure illustrates graphically - through a non-formal ontological schema - the above axioms. Moreover, for completeness purposes, the OWL functional syntax is provided in appendix C.

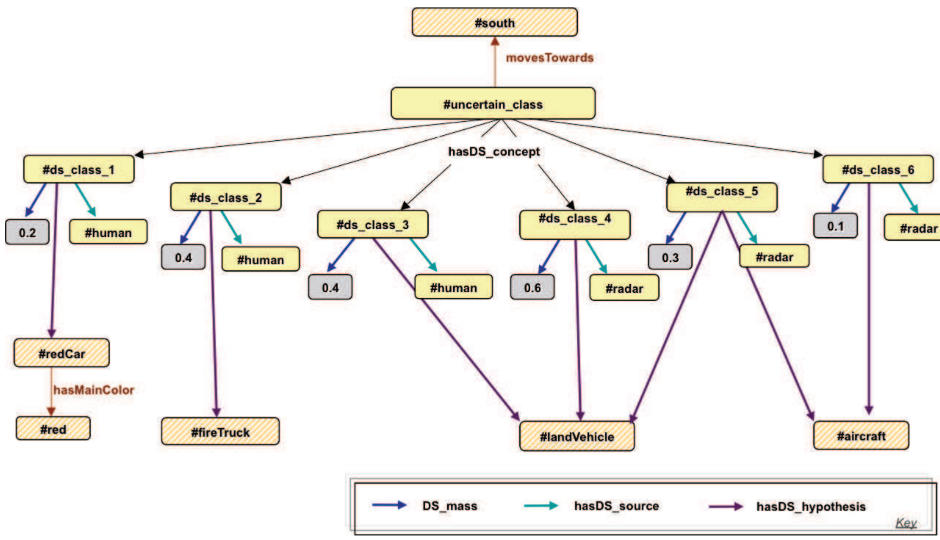

Figure 6.8: Schematic representations of the axioms.

6.3.2 Uncertain candidate instances of properties

The following example is related to uncertain instances of properties. This example is even simpler than the previous one, since only one source is involved. It focuses on the relation between two instances: a #policeCar and a #suspectCar in a chase scenario. The uncertainty resides in the instances isStoppedNear, movesTowards and movesSlowlyTowards, which define Ψ.

The *DS Concepts* are provided in the following tuples:
1- $\{$`#source`$, 0.3, \{$`isStoppedNear`$,$ `movesTowards`$\}\}$
2- $\{$`#source`$, 0.7, \{$`movesSlowlyTowards`$\}\}$

These *DS Concepts* are encoded by the following DL statements:
1- `#uncertain_property` : `Uncertain_property`
2- (`#policeCar`, `#uncertain_property`) : `hasUncertain_property`
3- (`#uncertain_property`, `#ds_property_1`) : `hasDS_concept`
4- (`#ds_property_1`, `#source`) : `hasDS_source`
5- (`#ds_property_1`, 0.3) : `DS_mass`
6- (`#ds_property_1`, `#suspectCar`) : `isStoppedNear`
7- (`#ds_property_1`, `#suspectCar`) : `movesTowards`
8- (`#uncertain_property`, `#ds_property_2`) : `hasDS_concept`
9- (`#ds_property_2`, `#source`) : `hasDS_source`
10- (`#ds_property_2`, 0.7) : `DS_mass`
11- (`#ds_property_2`, `#suspectCar`) : `movesSlowlyTowards`

Figure 6.9 represents graphically the encoding of this example. Moreover, as previously, the OWL functional syntax of this encoding is also included in appendix C.

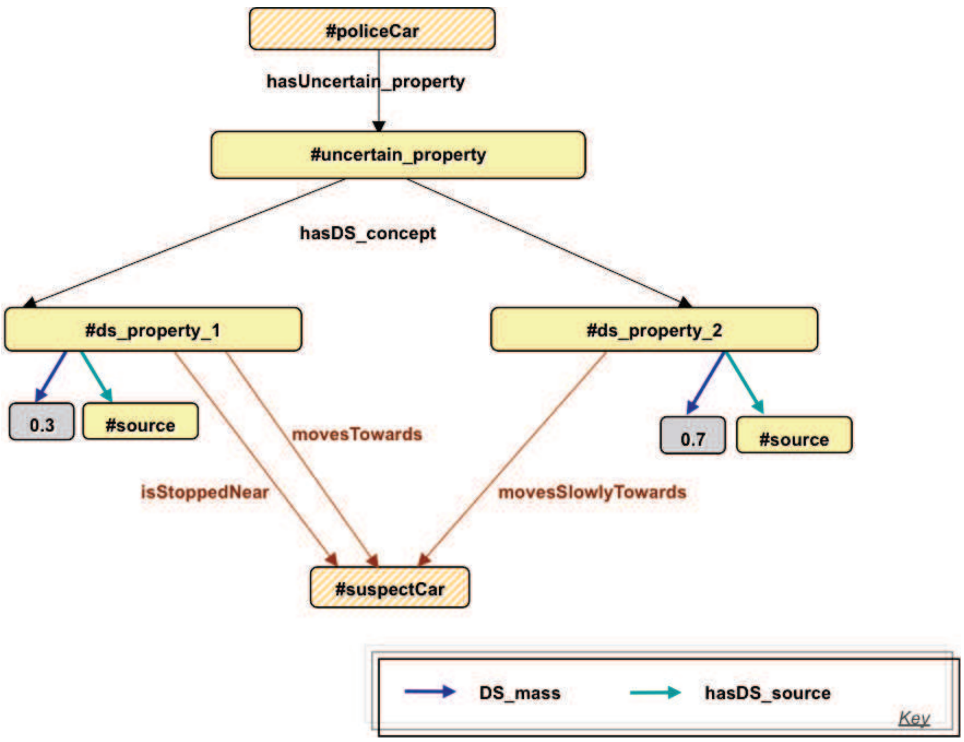

Figure 6.9: Uncertain property scenario.

6.4 Conclusion

In this chapter, we have presented our *DS-Ontology*. This is an upper ontology which defines the terminology to use when asserting uncertainty on instances of a domain ontology. Formal meaning has been given to the pieces of this ontology. This approach is compliant with the basic principle of OWL ontologies to structure knowledge on two levels: structural and assertional.

We have used the expression *semantic belief*, which refers to one or several uncertain candidate instances - described semantically through a domain ontology -, with a mass value and a reporting source.

Our representation of *semantic belief* answers the needs of information fusion systems in the field of uncertainty representation (see Chapters 2 and 3). First, through the *DS-Ontology*, one can represent empirical uncertainty. We enable the fact that an instance about the world is either satisfied or not, or to a certain degree of belief. This is made by representing it deterministically or by representing it as a candidate instance with a certain mass value in the *DS-Ontology*. Secondly, imprecision can also be represented in our framework. This is modeled by the fact that a semantic belief can encompass one or several uncertain candidate instances (the semantic belief associated to the candidate instances {`#landVehicle`, `#aircraft`} is less precise than the semantic belief with only `#landVehicle`), but also thanks to the granularity of candidate instances (the candidate instance `#landVehicle` is less precise than the candidate instance `#redCar`). Finally, inconsistent candidate instances can also be inserted by sources. For all these reasons, we can conclude that our semantic beliefs modeling permits to represent epistemic uncertainty (due to lack of knowledge) and inconsistent pieces of information, which are both fundamental in information fusion.

In this chapter, the use of the *DS-Ontology* to represent *semantic beliefs* has been illustrated through an example. When being aware of the *DS-Ontology* structure, its use presents no major difficulty. Still, we would like to draw the attention of the reader to the tool described in section 9.1 of Part IV. It permits to create and edit graphically these *semantic beliefs* with the famous Protégé editor while hiding the internal structure of the *DS-Ontology*. The same example presented in this chapter 6 will be recurrent in the two following Chapters 7 and 8.

In the following chapters, we will provide the formal relation between our uncertain instantiation obtained by using the *DS-Ontology* and the classical Dempster-Shafer formalism. To give a foretaste and recall what has already been outlined in Chapter 5, one can say that the major difference relies in the management of our uncertain candidate instances which have an explicit semantic attached to them. As a matter of fact, it would be unadvised to consider them as just simple labels and ignore their associated semantic. Indeed, hypotheses may have different levels of semantic granularity and may be *semantically related*. As we have seen in our example, one candidate instance is `#landVehicle`, a second one is `#redCar`. These candidate instances are not contradictory and are semantically related. This is exactly the purpose of the next chapter to formalize the definition and the discovery of these semantic relations.

Chapter 7

Semantic Set Operators

Contents

The previous chapter was dedicated to the representation of uncertain instances in ontologies. With the perspective to analyze which hypothesis should hold the best, it is necessary to take into account both degrees of belief and semantic relations between uncertain instances. We therefore introduce this chapter at this stage, since it permits us to present definitions of notions that will be extremely useful to handle the exclusivity assumption on the hypotheses of the Evidential frame of discernment and thus understand the reasoning process on semantic beliefs described in next chapter (Chapter 8).

As the reader is probably already aware, classical set theory is the branch of mathematics that studies sets. It all begins with a fundamental membership relation between an object I_j and a set I, denoted by $I_j \in I$. Since sets are objects, this relation can relate sets as well. Although any type of object can be collected into a set, set theory is applied most often to objects that are relevant to mathematics. For example $\{1, 2\}$ is a set where its objects are mathematical integers.

In the framework of semantic beliefs, we consider an ontological instance as a set. Its elementary objects are however not explicitly stated. We define in this chapter the semantic inclusion and disjointness operators, which can be understood as part of a semantic extension of the set theory. We will see for example that an instance may be included in another one. These operators correspond to the intuitive relation we -as human- can determine between two instances. In our case, this intuition is automatically computed by software. This software is aware of the terminology of the ontology and of the instances of the ontology that are related to our instances of interest.

This chapter first defines formally semantic inclusion and semantic intersection. It uses the following notation: I_j is an ontological instance (either an instance of a class or of a property). In addition, let $ABox_{Prop}$ be the set of instances of properties and $ABox_{Class}$ be the set of instances of classes.

Secondly, this chapter introduces a complete example in continuity to the previous ones from Chapter 4 and 6, so as to illustrate the involved calculi.

7.1 Semantic Inclusion

As stated in the introduction, this chapter begins by defining semantic inclusion.

Definition

We note I_j is semantically included in I_k by $I_j \subseteq_{sem} I_k$. In case I_j and I_k are instances of properties, we say that I_j is semantically included in I_k if I_j is a subproperty of I_k (formula (7.1)).

$$\forall I_j, I_k \in ABox_{Prop},\ (I_j \sqsubseteq I_k) \Rightarrow (I_j \subseteq_{sem} I_k)\ . \tag{7.1}$$

In case I_j and I_k are instances of classes, we say that I_j is semantically included in I_k if I_j is an instance of every classes C of I_k and all relations (datatype-properties and object-properties with their value and object, respectively) of I_k are also relations R of I_j (formula (7.2)).

$$\begin{gathered} I_j, I_k, I_q \in ABox_{Class}\ , \\ ((\forall C,\ I_k : C \Rightarrow I_j : C)\ \wedge\ (\forall R,\ (I_k, I_q) : R \Rightarrow (I_j, I_q) : R)) \\ \Rightarrow (I_j \subseteq_{sem} I_k)\ . \end{gathered} \tag{7.2}$$

In other words, the above definition says that $I_j \subseteq_{sem} I_k$ if I_j has no contradictory statements with I_k. Indeed, a `#redCar` instance would be included in a `#car` instance but not vice versa. This definition of semantic inclusion also refers to the notion of *hyponymy* (versus *hypernymy*).

7.2 Semantic Intersection

First, we introduce our aggregated semantic similarity measure, so that notions of semantic intersection or non-disjointness relying on this similarity measure can be defined.

7.2.1 Semantic Similarity

Generally speaking, semantic similarity between concepts is used in applications that need to exploit the knowledge that have been modelled in the ontology. It has received a lot of attention from the research field of ontology alignment, where semantic similarity is calculated between classes from distinct ontological terminologies. Regarding this literature, only few methods to assess similarities among instances have been proposed ([Albertoni 2006, Laudy 2009, d'Amato 2007]).

However, our purpose here is to assess the semantic similarity between our candidate instances and therefore we focus more on instances rather than on classes. Moreover, we do not want to use statistical means or external dictionaries, as it is sometimes proposed in the literature ([Lord 2003, Rodriguez 2003, Couto 2007]). Our semantic similarity measure should only rely on the domain ontology through which our instances of interest have been defined.

Therefore, we consider here semantic similarity as assessing the closeness between two instances of a same ontology. It is defined as a symmetric function, that is to say the similarity between the instances `#redCar` and `#aircraft` and the similarity between `#aircraft` and `#redCar` are the same. It returns a value between 0 and 1. The closest to 1 it is, the more similar the concepts are. In the literature, many works use the term of *distance*. In fact, similarity is just the contrary. In other words: $similarity = 1 - distance$.

As in [Albertoni 2006] or [Laudy 2009], we propose here our own aggregated function, which combines and extends different similarities measure already defined in the literature. Our function is denoted by $sim(I_j, I_k)$, where I_j and I_k are two instances. It is applicable both for instances of classes, and instances of properties. Our function takes into account the structural comparison of instances regarding the *a priori* terminology of the ontology, but also - in case of instances of classes - a comparison in term of their datatype- and object-properties.

7.2.1.1 Semantic Similarity regarding the Terminology of the Ontology

Our function takes into account the structural comparison of instances in terms of the concepts (class/property) the instances belong to. Considering instances of classes, we provide a similarity measure of their classes. And considering instances of properties, we respectively provide a similarity measure of their property definition. These measures rely on the hierarchical structure of the terminology. However, in case of instances of classes, we also consider the common object properties definition in which the classes of interest are involved.

Taxonomic Measure for Concepts Similarity - $sim_{Taxonomy}$

The concept (class or property) hierarchy is seen as a graph where the distance between two nodes can be calculated.

Therefore, one of the very first distance measure (and the simpler one) in the literature was just determining the shortest path between two concepts within the taxonomy (number of edges between the nodes). Thereby, distances between two brother nodes is equal to 2 (for example, in the figure 7.1 below, the nodes `Car` and `Truck`), whereas distance between cousin nodes is 4 (for example, in the figure 7.1, `Car` and `SurfaceVessel`).

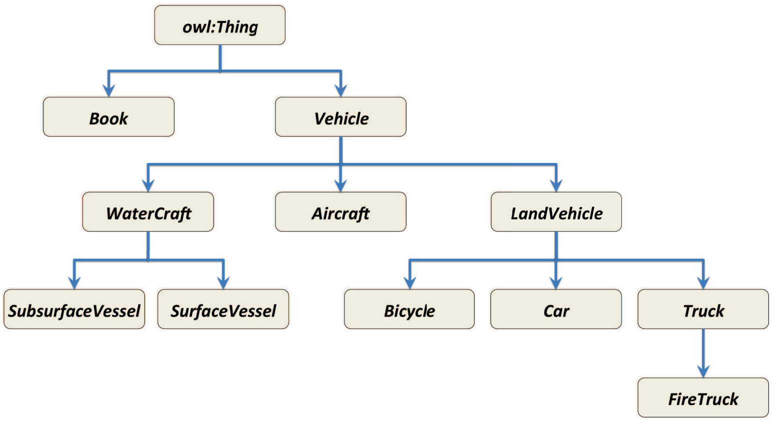

Figure 7.1: Example of concepts taxonomy.

However, one important drawback of this simple above method is that the similarity between two brother nodes is the same regardless of their depth in the hierarchy. For example, `Car` and `Truck` would have the same similarity measure as `Aircraft` and `LandVehicle`! This is actually not the intended result: we would prefer that the similarity measure increase with the depth in the hierarchy. Indeed, the intuitive idea behind this intended result is that, the deeper we go in the hierarchy (i.e. the farer from the root type), the less important the difference is between different levels (a node and its nodes). In other words, `Car` and `Truck` shall have a greater similarity measure than the pair `Aircraft` and `LandVehicle`. This depth criterion is often represented by the least (most specific) common subsumer of the two concepts.

For example, the Wu and Palmer [Wu 1994] similarity measure has been proposed to fulfill these above criteria. It measures indeed the distance that separates two types in the hierarchy and their position with the root. Equation (7.3) depicts this formula. C_j is a terminological concept. C_j is a class in case of instances of classes, and C_j is the property in case of instances of properties. Concept C is the least common subsumer of C_j and C_k. $depth(C)$ function is the number of edges separating C from the root. $depth_C(C_j)$ is the number of edges which separate C_j from the root while passing by C.

$$\forall I_j, I_k \in ABox,\ sim_{Taxonomy}(I_j, I_k) = \frac{2.depth(C)}{depth_C(C_j) + depth_C(C_k)} \tag{7.3}$$
$$where\ I_j : C_j,\ I_k : C_k\ if\ I_j, I_k \in ABox_{Class}$$
$$and\ C_j := I_j,\ C_k := I_k\ if\ I_j, I_k \in ABox_{Prop}\ .$$

For ease of explanation, we have not considered in the above formulas, the case of class multi-inheritance (an instance of class may have multiple classes and one class can have multiple superclasses). However, this is taken into account by modeling C as the class being the most specific common subsumer of all possible C obtained with the possible pairwise. Moreover, C_j is the class with whom the value $depth_C(C_j)$ is the minimum.

Measure for Classes Similarity according to the Domain and Range Definition of Properties - $sim_{PropDef}$

The terminology of the ontology is not limited to the hierarchical structure of the concepts but also to the domain and range definitions within properties. So similarity between classes should also entail the comparison of properties (datatype and object properties) that are possibly linked to instances of these classes. For example, the classes `Car` and `Truck` are both part of the domain of the object property `hasNumberOfDriveWheels`, whereas `Bicycle` class is not related to this property. Therefore, `Car` and `Truck` classes shall have a higher semantic similarity.

The first equation in (7.4) calculates the similarity for the domain definitions of object properties, in which classes of I_j and I_k are involved. The second one is considering the range definition of object properties. The last one considers the datatype properties definition (here classes can only belong to the domain definition, since the range is always a datatype). The notation $ObPropDomain(C_j)$ refers to the set of object properties that have the C_j class in their domain definition. Respectively, $DaPropDomain(C_j)$ refers to the range definition of the datatype properties. The notation $\mid ObPropDomain(C_j) \mid$ entails the cardinality of this set. Last of all, $\mid ObPropDomain(C_j) \cap ObPropDomain(C_k) \mid$ is the number of properties for which both C_j and C_k classes lay in the domain definition.

$$
\begin{aligned}
&\forall I_j, I_k \in ABox_{Class}, \\
sim_{ObjectPropDomain}(I_j, I_k) &= \frac{2. \mid ObPropDomain(C_j) \cap ObPropDomain(C_k) \mid}{\mid ObPropDomain(C_j) \mid + \mid ObPropDomain(C_k) \mid} \\
sim_{ObjectPropRange}(I_j, I_k) &= \frac{2. \mid ObPropRange(C_j) \cap ObPropRange(C_k) \mid}{\mid ObPropRange(C_j) \mid + \mid ObPropRange(C_k) \mid} \\
sim_{DataPropDomain}(I_j, I_k) &= \frac{2. \mid DaPropDomain(C_j) \cap DaPropDomain(C_k) \mid}{\mid DaPropDomain(C_j) \mid + \mid DaPropDomain(C_k) \mid} \\
&where\ I_j : C_j,\ I_k : C_k\ if\ I_j, I_k \in ABox_{Class}\ .
\end{aligned}
\tag{7.4}
$$

Moreover, we state that in case the denominator of one of the above equations is equal to 0, then the associated similarity (the right hand side of the concerned equation) is equal to 1.

Finally, the above equations are combined in the following amalgamated equation.

$$\forall I_j, I_k \in ABox_{Class},$$
$$sim_{PropDef}(I_j, I_k) = \frac{\frac{sim_{ObjectPropDomain}(I_j,I_k)+sim_{ObjectPropRange}(I_j,I_k)}{2} + sim_{DataPropDomain}(I_j, I_k)}{2}$$
$$where\ I_j : C_j,\ I_k : C_k\ if\ I_j, I_k \in ABox_{Class}\ . \tag{7.5}$$

7.2.1.2 Semantic Similarity regarding the Population of the Ontology

While looking at the similarity of two instances, it seems natural to take also into account their context of assertion in the ontology, in addition to their definition in the terminology. This context of assertion refers to the asserted properties related to the instances of interest (in case of instances of classes). For example, a `#redCar` instance is intuitively more similar to a `#fireTruck`, which have both in common the property `hasMainColor` related to the `#red` instance, whereas to a `#blueTruck`.

This similarity measure counts the number of identical properties versus the number of properties related to these individuals. It is calculated both for object properties and datatype properties (equations (7.6)). For object properties, $\mid ObProp(I_j) \mid$ is the number of object properties of individual I_j (respectively, for datatype properties: $\mid DaProp(I_j) \mid$). $\mid ObProp(I_j) \cap ObProp(I_k) \mid$ is the number of common properties - identical predicate and related individual or value - for the two individuals I_j and I_k.

$$\forall I_j, I_k \in ABox_{Class},$$
$$sim_{ObjectProp}(I_j, I_k) = \frac{2. \mid ObProp(I_j) \cap ObProp(I_k) \mid}{\mid ObProp(I_j) \mid + \mid ObProp(I_k) \mid} \tag{7.6}$$
$$sim_{DataProp}(I_j, I_k) = \frac{2. \mid DaProp(I_j) \cap DaProp(I_k) \mid}{\mid DaProp(I_j) \mid + \mid DaProp(I_k) \mid}\ .$$

Moreover, as previously mentioned, we state here also that in case the denominator of one of the above equations is equal to 0, then the associated similarity (the right hand side of the concerned equation) is equal to 1.

All these previous measures are quite intuitive and objective in regards to the Open World assumption (see section 4.3.1.1). For example, we could have also chosen $sim_{ObjectProp}(I_j, I_k) = \frac{|ObProp(I_j) \cap ObProp(I_k)|}{min\{|ObProp(I_j)|,|ObProp(I_k)|\}}$ or $= \frac{|ObProp(I_j) \cap ObProp(I_k)|}{max\{|ObProp(I_j)|,|ObProp(I_k)|\}}$. However, the first version is always greater or equal to our measure (upper-bound). It is an optimistic view of the Open World assumption (the relations that are not stated by an individual are the same as the other individual). On the opposite, the other mentioned version is always lower or equal to our measure (lower-bound). It can be seen as a pessimistic view of the Open World assumption (the relations that are not stated by an individual have different values than the other individual).

We could have also customized the calculation of $sim_{DataProp}$ by taking into account the similarity between the values of the object (e.g. numerical/date/string similarity). This has not been done here, but could be part of future works.

7.2.1.3 Our Final Aggregated Semantic Similarity Function

We now define our final similarity measure as an aggregation function, that combines the similarity measures seen above.

It is defined by $sim : ABox \times ABox \rightarrow]0,1]$ with:

$$sim(I_j, I_k) = \begin{cases} w_1.sim_{Taxonomy}(I_j, I_k) & \\ + w_2.sim_{PropDef}(I_j, I_k) & \\ + w_3.sim_{ObjectProp}(I_j, I_k) & \\ + w_4.sim_{DataProp}(I_j, I_k) & \text{if } I_j, I_k \in ABox_{Class} \\ & \\ sim_{Taxonomy}(I_j, I_k) & \text{if } I_j, I_k \in ABox_{Prop} \ . \end{cases} \tag{7.7}$$

where $\sum_i w_i = 1$.

w_i are weights that allow to give more importance to some elements with regards to the others and that enable to normalize the final similarity measure (in order to be consistent with similarity measure definition).

7.2.1.4 Generalization of the Semantic Similarity Measures for Multiple Elements

We generalize here our semantic similarity measure to more than two elements. This generalization should necessarily satisfy the following constraint:

$$sim(A, B) \geqslant sim(A, B, C) \ .$$

Considering the Wu and Palmer measure (see formula 7.3), we can generalize its equation by:

$$sim_{Taxonomy}(I_1, .., I_n) = \frac{n.depth(C)}{\sum_{j=1}^{n} depth_C(C_j)} \tag{7.8}$$

$$where\ I_j : C_j\ if\ \forall j \in \{1, .., n\},\ I_j \in ABox_{Class}$$
$$and\ C_j := I_j\ if\ \forall j \in \{1, .., n\},\ I_j \in ABox_{Prop}\ .$$

The other similarity measures ($sim_{PropDef}$, $sim_{ObjectProp}$ and $sim_{DataProp}$) can be easily generalized by following this template:

$$sim_{XXX}(I_1, .., I_n) = \frac{n. \mid Prop(I_1) \cap .. \cap Prop(I_n) \mid}{\sum_{j=1}^{n} \mid Prop(I_j) \mid} \tag{7.9}$$

7.2.2 Definition of Semantic Intersection, Semantic Non-Disjointness

The semantic intersection of I_j and I_k is denoted by $I_j \cap_{sem} I_k$.

Formally, two instances I_j and I_k are semantically disjoint if their semantic intersection is the empty set, denoted by $I_j \cap_{sem} I_k = \emptyset$.

Instances I_j and I_k are semantically non-disjoint (or commonly speaking: have a semantic intersection) if the similarity measure between I_j and I_k exceeds a certain threshold. This threshold has a value between 0 and 1.

$$\begin{gathered} \forall I_j, I_k \in A_{Box}, \\ If\ (sim(I_j, I_k) > threshold) \\ \Rightarrow (I_j \cap_{sem} I_k \neq \emptyset)\ . \end{gathered} \tag{7.10}$$

This definition can also be generalized to the computation of multiple intersections, such as:

$$\begin{gathered} \forall I_j \in A_{Box},\ where\ j \in \{1, .., n\}, \\ If\ (sim(I_1, .., I_n) > threshold) \\ \Rightarrow (I_1 \cap_{sem} .. \cap_{sem} I_n \neq \emptyset)\ . \end{gathered}$$

The threshold can be either manually fixed or automatically computed. In the next section, we focus on an automatic calculation of the threshold considering a particular finite set of instances.

7.2.3 Semantic Intersection within a Set of Instances

We focus here on a special case where we are not interested only by the semantic intersection between two instances, but by all the semantic intersections within a finite set of instances. Considering this set of instances, we want to determine the sets of instances for which a semantic intersection holds. This section is twofold. It permits us to recap the process for determining a semantic intersection between instances while applying this for a special case of a set of instances. Moreover, this section introduces a method to calculate automatically a semantic intersection threshold.

First, we propose a procedure (presented in figure 7.2) for semantic intersection between all combinations of instances of a set Υ of instances. From line 6 to 11, the procedure is dedicated to the computation of pairwise-intersection within all instances of the set. Similarities are first computed according to section 7.2.1, then, on line 8, a threshold is calculated through the call of `calculateThreshold()`. Afterwards, from line 12 to 26, we search for multiple (more than pairwise) intersections. The similarity value is calculated only for sets of instances which fulfill the following

Require: Υ: Finite Set of Instances
Ensure: $hasIntersection$: All sets which have an intersection

```
1: hasIntersection ← null
2: similarity{} ← null
3: powerSet ← getPowerSet(Υ)
4: setsOfCardinalityTwo ← getSetsOfSizeTwo(powerSet)
5: setsOfCardinalityMoreThanTwo ← getSetsOfSizeMoreThanTwo(powerSet)

6: for all I in setsOfCardinalityTwo do
7:    similarity{I} ← calculateSimilarity(I)
8: threshold ← calculateThreshold(similarity)
9: for all I in setsOfCardinalityTwo do
10:    if similarity{I} > threshold then
11:       hasIntersection ← hasIntersection ∪ I
12: for all I in setsOfCardinalityMoreThanTwo do
13:    findIntersection ← false
14:    for all I_j in I do
15:       if findIntersection = false then
16:          I' ← I \ {I_j}
17:          if contains(hasIntersection, I') then
18:             subIntersectionSatisfied ← true
19:             for all I'_j in I' do
20:                if notContains(hasIntersection, {I_j, I'_j}) then
21:                   subIntersectionSatisfied ← false
22:             if subIntersectionSatisfied = true then
23:                similarity{I} ← calculateSimilarity(I)
24:                if similarity{I} > threshold then
25:                   hasIntersection ← hasIntersection ∪ I
26:                   findIntersection ← true
27: return hasIntersection
```

Figure 7.2: Multiple semantic intersections algorithm.

schema:

$$\begin{gathered} If\ I_j \in A_{Box},\ where\ j \in \{1,..,n\},\ I_1 \cap_{sem} .. \cap_{sem} I_n \neq \emptyset \\ and\ \forall j \in \{1,..,n\},\ I_j \cap_{sem} I_{n+1} \neq \emptyset \\ \textit{Then, we calculate } sim(I_1 \cap_{sem} .. \cap_{sem} I_n \cap_{sem} I_{n+1}) \ . \end{gathered}$$

The first line of this criterion is represented by line 17 of the algorithm, the second line corresponds to lines 19 and 20 of the algorithm 7.2.

One can note that the threshold calculation is here performed only with semantic similarity values between pairwise instances. Indeed, if the threshold was calculated according to every similarity measures between all combination of instances, the threshold would necessarily be lowered, and nearly most of the pairwise instances would have an intersection. Thus, we decide to calculate the threshold according to similarity values between the same number of instances and here between pairwise instances.

The threshold is in fact automatically computed from the list of pairwise semantic similarity values. This threshold is computed by a partitioning method, where the partitions are here lists of similarity values and their number is fixed to maximum two: the list of similarities values that lead to an intersection (the list of value closest to 1) and the others. The chosen algorithm `calculateThreshold` is depicted in figure 7.3.

This algorithm implements the Fisher's criterion (see formuli 7.11). It performs classification in a one-dimensional space. It maximizes the distance between the means of the two sublists while minimizing the variance within each sublist. In other words, it gets larger with the distance between the sublist barycenters and gets larger when the sublists become more "compact". The goal being to identify the limit value (i.e. threshold) on which the sublists will be "as separated as possible".

$$criterion_{Fisher} = \frac{\mid m_{sup} - m_{inf} \mid^2}{s_{sup}^2 + s_{inf}^2} \ . \qquad (7.11)$$

where m represents the mean, s^2 represents the variance and the subscripts denote the two sublists $list_{sup}$ and $list_{inf}$.

The threshold is varying according to the list of all semantic similarity values. This method allows adapting the threshold to the granularity of the set of instances. It translates our general impression that the concept of a compact car is closer to the concept of a minivan than of a plane's; however the concept of a compact car is closer to the concept of plane than of a book's. In the first case, the intersection should be brought by the pair (compact car, minivan), whereas in the latter, it should be brought by the pair of (compact car, plane). It should be noted that, in both cases, the concepts of compact car and plane have the same semantic similarity value.

7.3 Examples of Semantic Set Operators Computation

This section illustrates the notions of semantic set operators presented in this chapter. It shows indeed the reasoning and calculi involved for a specific example of a set of instances. We actually take the set of instances introduced in section 4.5, which has been also taken for example in the previous chapter 6.3. Let's remind that this set of individual was {`#aircraft`, `#landVehicle`, `#fireTruck`, `#redCar`}.

In the following, we present which instances semantically include other ones and the reason why. We display also the results of semantic similarities in this set, and focus on the calculations of some pairs of instances. The threshold determination as well as the final results for semantic intersection are explained.

Require: $similarity\{\}$: Dictionary of similarity measures
Ensure: $threshold$: Real between 0 and 1

$list[] \leftarrow getAllValues(similarity\{\})$
$k \leftarrow getSize(list[])$
$list[] \leftarrow sort(list[])$
$listinf[] \leftarrow null,\ listinf[1] \leftarrow 0$
$listsup[] \leftarrow list[]$
$maxCriterion \leftarrow 0$
for $i : 0..k$ **do**
  if $i > 0$ **then**
    if $i = 1$ **then**
      $listinf.remove(\text{"0"})$
    $listinf.add(list[i])$
    $listsup.remove(list[i])$
    if $getSize(listsup) = 0$ **then**
      $listsup.add(\text{"1"})$
  $criterion \leftarrow \frac{(mean(listsup)-mean(listinf))^2}{variance(listsup)+variance(listinf)}$
  if $maxCriterion < criterion$ **then**
    $maxCriterion \leftarrow criterion$
    $threshold \leftarrow listinf[getSize(listinf)]$
return $threshold$

Figure 7.3: Threshold determination algorithm for a finite set of instances.

7.3.1 Semantic Inclusion Determination

For each instance in {`#aircraft`, `#landVehicle`, `#fireTruck`, `#redCar`}, we list and explain below their semantic inclusions if any, according to the semantic inclusion definition of section 7.1. More particularly, as it is a set of instances of classes, we refer to the formula 7.2 of section 7.1.

- `#aircraft`: its class `Aircraft` is not a superclass of any classes of other instances. As a matter of fact, `Aircraft` is neither a superclass of `LandVehicle`, `Car`, `Truck`, nor `FireTruck`. Therefore, `#aircraft` has no semantically included instances.

- `#landVehicle`: its class `LandVehicle` is not a superclass of `Aircraft`, but is a superclass of `Car` and `FireTruck`. Moreover, the `#landVehicle` instance has no contradictory property with `#redCar` nor `#fireTruck`. In fact, it has no property at all here. Therefore, `#landVehicle` semantically includes `#redCar` and `#fireTruck`. One can remark, as a teaching example, that if the `#landVehicle` would have had a property of blue color or a property concerning a specific speed, then `#landVehicle` would have had here no included instances.

- Concerning the instances: `#fireTruck` and `#redCar`, they have no semantic included instances since their respective classes `FireTruck` and `Car` are final leaves of the class hierarchy and thus have no subclasses.

7.3.2 Semantic Similarity Determination

We display in the following the subresults of the semantic similarity computation according to the criteria defined in section 7.2.1. The subresults are presented as symmetric tables since similarities measures are symetric.

Table 7.4 comes from the results obtained by equation 7.3 on page 98. It takes into account the class hierarchy of our set of instances. You can see the previous graphical schema 7.1 which explicitly represents this hierarchy.

	#aircraft	#landVehicle	#fireTruck	#redCar
#aircraft	1	-	-	-
#landVehicle	0.5	1	-	-
#fireTruck	0.33	0.67	1	-
#redCar	0.4	0.8	0.57	1

Figure 7.4: Calculated similarity values for class hierarchy.

For example, this table states that $sim_{Taxonomy}(\texttt{\#redCar}, \texttt{\#aircraft}) = 0.4$. Here, $I_j = \texttt{\#redCar}$ and $I_k = \texttt{\#aircraft}$ in the equation 7.3. Consequently, their respective direct classes are $C_j = \texttt{Car}$ and $C_k = \texttt{Aircraft}$. The least common subsumer of these classes is $C = \texttt{Vehicle}$, so $depth(C) = 1$. We have $depth(C_j) = 3$ and $depth(C_k) = 2$. Therefore, $sim_{Taxonomy}(\texttt{\#redCar}, \texttt{\#aircraft}) = \frac{2}{3+2} = 0.4$.

Another cell of this table states that $sim_{Taxonomy}(\texttt{\#redCar}, \texttt{\#fireTruck}) = 0.57$. Here, $I_j = \texttt{\#redCar}$ and $I_k = \texttt{\#fireTruck}$. Similarly, their respective classes are $C_j = \texttt{Car}$ and $C_k = \texttt{FireTruck}$. The least common subsumer of these classes is $C = \texttt{LandVehicle}$, so $depth(C) = 2$. We have $depth(C_j) = 3$ and $depth(C_k) = 4$. Therefore, $sim_{Taxonomy}(\texttt{\#redCar}, \texttt{\#fireTruck}) = \frac{2\times2}{3+4} \simeq 0.57$.

The second subresult is depicted in table 7.5. It represents the similarities of the instances regarding the common properties definition in which their classes are involved.

	#aircraft	#landVehicle	#fireTruck	#redCar
#aircraft	1	-	-	-
#landVehicle	0.5	1	-	-
#fireTruck	0.5	1	1	-
#redCar	0.5	1	1	1

Figure 7.5: Calculated similarity values for property definitions.

For example, the table shows that $sim_{PropDef}(\texttt{\#redCar}, \texttt{\#aircraft}) =$

0.5. This can be explained by the fact that none object properties distinguish `Aircraft` or `Car` classes in their domain of range definitions (so $sim_{ObjectPropDomain}(\texttt{\#redCar}, \texttt{\#aircraft}) = sim_{ObjectPropRange}(\texttt{\#redCar}, \texttt{\#aircraft}) = 1$).

And $sim_{DataPropDomain}(\texttt{\#redCar}, \texttt{\#aircraft}) = 0$ since the datatype property `#hasNumberOfDriveWheels` defines for domain the `LandVehicle` class whose subtype is `Car`. Thereby, $sim_{PropDef}(\texttt{\#redCar}, \texttt{\#aircraft}) = \frac{\frac{1+1}{2}+0}{2} = 0.5$

Concerning `#redCar` and `#fireTruck`, their similarity regarding domain and range of object properties is also equal to 1 due to the same reasons as above. Moreover, their classes are also involved in the same domain definition of datatype property (`#hasNumberOfDriveWheels`). Therefore $sim_{DataPropDomain}(\texttt{\#redCar}, \texttt{\#fireTruck}) = sim_{PropDef}(\texttt{\#redCar}, \texttt{\#aircraft}) = 1$. This third table in figure 7.6 presents the similarity results for the instances regarding their asserted object properties.

	#aircraft	#landVehicle	#fireTruck	#redCar
#aircraft	1	-	-	-
#landVehicle	1	1	-	-
#fireTruck	0	0	1	-
#redCar	0	0	1	1

Figure 7.6: Calculated similarity values regarding object properties.

For example, this table states that $sim_{ObjectProp}(\texttt{\#redCar}, \texttt{\#aircraft}) = 0$. The justification is rather simple, since these two instances have no object properties in common.

A second example of result in the table is $sim_{ObjectProp}(\texttt{\#redCar}, \texttt{\#fireTruck}) = 1$. These two instances have both the object property `hasMainColor` associated to the object `#red` instance. They have no other object properties. Thus, $sim_{ObjectProp}(\texttt{\#redCar}, \texttt{\#fireTruck}) = \frac{2\times1}{1+1} = 1$.

The last subtable below (7.7) shows the similarity results of the instances regarding their asserted datatype properties. The results are here quite straightforward to understand, since there are no datatype properties associated to any of these instances. Thus, there are no differences between each others.

	#aircraft	#landVehicle	#fireTruck	#redCar
#aircraft	1	-	-	-
#landVehicle	1	1	-	-
#fireTruck	1	1	1	-
#redCar	1	1	1	1

Figure 7.7: Calculated similarity values regarding datatypes properties.

The following table 7.8 fused all the subresults obtained so far. It has been calculated with the equation (7.7) and with the following parameters: $w_1 = 0.6$, $w_2 = 0.2$, $w_3 = 0.1$ and $w_4 = 0.1$. An important weight w_1 is given to the structure of the terminology, which is always very informative. The other criteria w_2, w_3 and w_4 are taken into account, but not magnified.

	`#aircraft`	`#landVehicle`	`#fireTruck`	`#redCar`
`#aircraft`	1	-	-	-
`#landVehicle`	0.6	1	-	-
`#fireTruck`	0.4	0.7	1	-
`#redCar`	0.44	0.78	0.74	1

Figure 7.8: Calculated global semantic similarities values.

For example, we have actually $sim(\texttt{\#redCar}, \texttt{\#aircraft}) = 0.6\times0.4+0.2\times0.5+0.1 \times 0 + 0.1 \times 1 = 0.44$.

As well, $sim(\texttt{\#redCar}, \texttt{\#fireTruck}) = 0.6\times0.57+0.2\times1+0.1\times1+0.1\times1 \simeq 0.74$.

7.3.3 Threshold Determination

In order to determine the threshold of this set of instances, we apply the algorithm 7.3. Line 3 of this algorithm gives the sorted list of all semantic similarities obtained from the above table: {0.4, 0.44, 0.6, 0.7, 0.74, 0.78}.

From this list, there are seven configurations of (ordered) sublist pairs possibles. For each sublist pairs, we have calculated the Fisher's criterion, and we have found that the maximal distance is 15.72 and is obtained for the following sublist pair: *listinf* = {0.4, 0.44} and *listsup* = {0.6, 0.7, 0.74, 0.78}. The last member of *listinf* = 0.44 gives the value of the threshold.

7.3.4 Semantic Intersection Determination

The semantic intersections are inferred by comparing the previously obtained similarity values with the above calculated threshold. Therefore, a semantic intersection is found when a similarity value is strictly greater than 0.44.

Therefore, according to table 7.8:

- `#redCar` and `#fireTruck` have a semantic intersection,
- `#redCar` and `#landVehicle` have a semantic intersection,
- `#fireTruck` and `#landVehicle` have a semantic intersection and
- `#aircraft` and `#landVehicle` have a semantic intersection.

7.3.5 Conclusion Example

The following schema represents graphically the semantic set relations we discovered in this example.

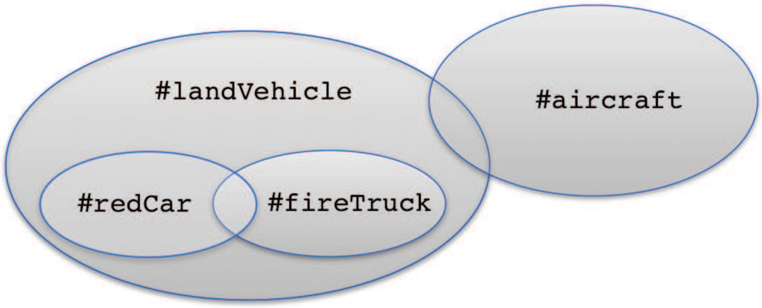

Figure 7.9: Graphical representation of the semantic set relations in this example.

7.4 Conclusion

In this chapter, we have proposed a semantic extension of the set theory through the semantic intersection and semantic inclusion operators. These operators have been defined for both instances of classes and instances of properties.

Semantic inclusion is quite straightforward to determine since it mostly relies on the hierarchy of the concepts. Semantic intersection needs however more computation since first a semantic similarity has to be calculated. This semantic similarity has many ways to be defined. We propose one in this chapter based on an aggregated function taking into account the terminology of the ontology (both its hierarchy and definitions of properties) and the populated ontology (properties associated to individuals).

Chapter 8

Reasoning on Semantic Beliefs

Contents

Once uncertainty contained in the information has been represented (see Chapter 6), reasoning processes have to be conducted to fuse the different beliefs and eventually choose the "best" instance.

In the previous chapter, we have underlined that the introduction of semantic hypotheses modelled with instances lead to hypotheses that may overlap each others, have the same meaning or on the contrary have totally disjoint intrinsic semantic natures. All this can be discovered with our semantic set operators.

This chapter emphasizes the need of semantic set relations, namely semantic inclusion and semantic intersection discovered between candidate instances so as to construct a proper frame of discernment still consistent with the exclusivity assumption of the Evidential theory. Indeed, contrary to the existing approaches presented in Chapter 5, we believe that it is possible to rely on the Evidential theory. We believe indeed that each hypothesis is clearly defined in the ontology so that the exclusivity assumption remains in the Evidence theory. For that purpose, the semantics of the hypotheses should be thoroughly taken into account during the construction of the frame of discernment.

Sections of this chapter can be viewed as the chronological steps realized by the system in order to reason on uncertain instances represented through the *DS-Ontology*. These steps are illustrated by the following figure 8.1.

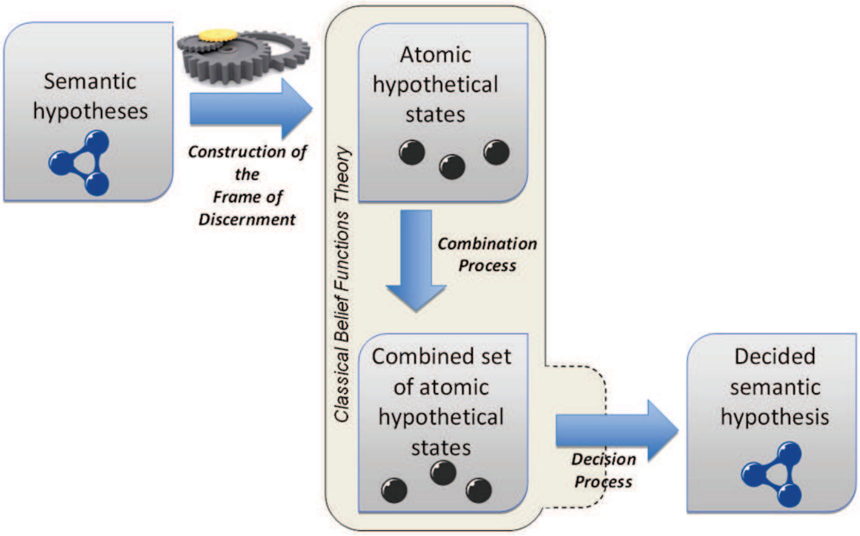

Figure 8.1: Workflow of reasoning on semantic hypotheses.

8.1 Running an Automatic Reasoner

As we mentioned in section 4.4, a benefit of ontologies with respect to simpler representation formalisms consists in the support of automatic reasoning tasks. Indeed, on the basis of the asserted knowledge it is possible to automatically derive new or implicit knowledge about the current situation. These inferences are based on the defined classes, properties or other constraint axioms and on the instances asserted while the information fusion system is operating. This reasoning is an essential preliminary step to any software agent that wants to access to the information gathered in an ontology.

This type of reasoning is called implicit or even automatic reasoning, since it is based on the ontology definition itself (versus external rules that would explicitly encode inference process). As such, it has some well known limits, which are its lack of mathematical calculi handling, among others.

Therefore, our approach does not consist in extending an automatic reasoner to apply our combination and decision process on semantic beliefs but rather encapsulates processes in an external reasoning system that still launch an automatic reasoner as its preliminary step.

8.2 Mapping Semantic Beliefs to Atomic Beliefs

8.2.1 Motivation

In order to apply the classical evidential combination and decision processes, a consistent frame of discernment is required (see Chapter 3).

Actually, in our semantic beliefs framework, the *universal set of candidate instances* Ψ is what is closest to the frame of discernment Ω in the Evidential theory.

However, elements of Ψ may not satisfy the underlying assumption of exclusivity. Indeed, *candidate instances* are not necessary disjoint from each other: ontological instances are not all on the same level of granularity and some instances may be semantically included or have a semantic intersection with other instances (see Chapter 7).

One can recall that set relations are quite important in the Evidential theory, since they influence the calculi of belief functions and of the Dempster's combination. As a matter of fact, if inclusion was not taken into account, we would for example under-estimate the belief ($bel(A) = \sum_{B|B\subseteq A} m(B)$ $, \forall A \subseteq \Omega$ would be limited to $bel(A) = m(A)$). If intersection was not taken into account, we would also under-estimate the plausibility ($pl(A) = \sum_{B|B\cap A\neq\emptyset} m(B)$ $, \forall A \subseteq \Omega$ would be limited to $pl(A) = m(A)$). Finally if different sources are involved, we would over-estimate the amount of conflict K ($K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$).

Therefore, Ω can not be considered as equal to Ψ.

The purpose of this section is thus to reformulate Ψ to obtain a frame of discernment Ω consistent with the assumptions of the Evidential theory, by relying on the semantic set operators, as exposed on schema 8.2. In other words, it is the issue of how to make the semantic of the hypotheses explicit in the set theory.

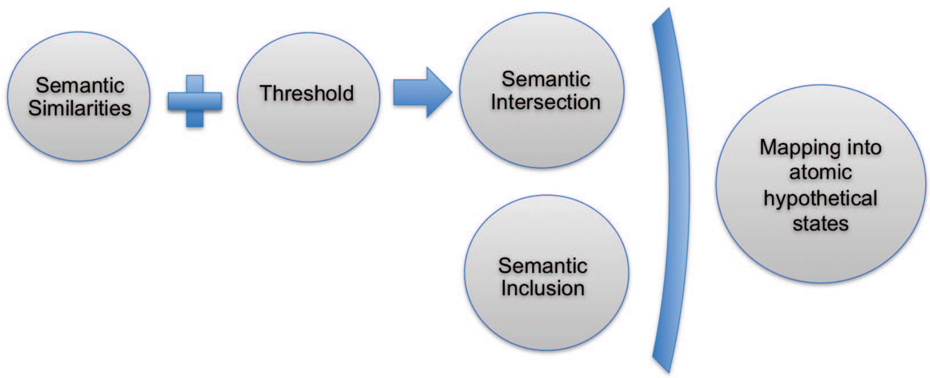

Figure 8.2: Use of semantic set operators for the construction of the frame of discernment.

8.2.2 Translation of the Frame of Discernment

Once semantic intersections and inclusions within a set of candidate instances have been identified (Chapter 7), we set up a consistent frame of discernment.

Let's consider the set of candidate instances $\Psi = \{I_{j_1}, ..., I_{j_n}\}$, then we propose a consistent frame of discernment Ω defined as:

$$\Omega = \{H_M \mid M \subseteq \{j_1, ..., j_n\},\ M \neq \emptyset,\ \bigcap_{j\in M} {}_{sem} I_j \neq \emptyset,$$
$$if\ I_j \subseteq_{sem} I_{j'}\ and\ j \in M\ then\ j' \in M\}\ . \tag{8.1}$$

Visually, it corresponds to all delimited zones in a graphical representation of the semantic set operators results (as the one in figure 8.6 of page 118, where one can count six different zones).

Consequently, we can express a mapping function f_{map} that enables to map one candidate instance to one or several virtual atomic states H_i of the evidential frame of discernment.

$$f_{map} : \Psi \to 2^{\Omega}$$
$$f_{map}(I_j) = \{H_M \mid j \in M\} \ . \tag{8.2}$$

For ease of understanding of the two last mathematical formulas, the reader could also consider the following expression of f_{map}, which is however restricted to only double intersection (and not multiple intersection):

$$f_{map} : \Psi \to 2^{\Omega}$$
$$f_{map}(I_j) = \{H_j\} \cup \{H_{j,k} \mid I_j \cap_{sem} I_k \neq \emptyset,\ I_j \not\subseteq_{sem} I_k,\ I_k \not\subseteq_{sem} I_j\} \cup \bigcup_{I_q \subsetneq_{sem} I_j} f_{map}(I_q) \ . \tag{8.3}$$

This last expression is recursive - it calls its own definition in case of included instances. By taking into account the transitivity of the semantic inclusion, we can rewrite this function through the chain of two mapping functions, such that:

$$f_{map} = f_{mapIncl} \circ f_{mapInter} \tag{8.4}$$

where

$$f_{mapIncl}(I_j) = \bigcup_{I_q \subseteq_{sem} I_j} f_{mapInter}(I_q)$$

and

$$f_{mapInter}(I_j) = \{H_j\} \cup \{H_{j,k} \mid I_j \cap_{sem} I_k \neq \emptyset,\ I_j \not\subseteq_{sem} I_k,\ I_k \not\subseteq_{sem} I_j\} \ .$$

For example, in case of a discovered semantic intersection between two candidate instances I_1 and I_2, I_1 is reformulated as the union of two singletons $\{H_1, H_{1,2}\}$ and I_2 as $\{H_2, H_{1,2}\}$. In case of a discovered semantic inclusion between two candidate instances I_1 and I_2, where I_1 is semantically included in I_2, I_1 is represented by a single hypothesis $\{H_1\}$ and I_2 by the union of hypotheses $\{H_1, H_2\}$.

Considering a set of candidate instance Ψ of cardinality $\mid \Psi \mid= N$, the size of the discernement frame Ω is in the worst case equals to $2^N - 1$.

$$If \ \mid \Psi \mid= N \ then \ \mid \Omega \mid \leq 2^N - 1 \ . \tag{8.5}$$

Let's take the example of $\Psi = \{I_1, I_2, I_3\}$, then $\mid \Psi \mid= 3$. We assume that all instances have a semantic intersection with one another, which is the worst case for the size of the created frame of discernment (as a matter of fact, discovered semantic inclusions do not lead to the creation of new atomic hypotheses in Ω). This would indeed lead to the creation of $\Omega = \{H_1, H_2, H_3, H_{1,2}, H_{1,3}, H_{2,3}, H_{1,2,3}\}$, thus having the worst case $\mid \Omega \mid= 2^3 - 1 = 7$.

Consequently, the final complexity of the frame of discernment is lower than the one which would have been obtained by the DSmT - Dezert and Smarandache extension of the Evidential theory (see section 5.2). Indeed, here a more finer analysis is performed on the semantic of the candidate instances instead of producing a set which would contained all possible intersections even when they are semantically null. We thus find ourselves in the particular case of the reduced model D_r^Ω of DSmT (see again section 5.2), except that here not only intersections have been taken into account but also included hypotheses have been considered.

The following algorithm illustrates the implementation of the above formula 8.1.

Require: Ψ: Universal Set of Candidate Instances
Ensure: Ω: Frame of Discernment

$hasInclusion[][] \leftarrow run\ SemanticInclusion$
$hasIntersection \leftarrow run\ SemanticIntersection$
$\Omega \leftarrow getPowerSet(\Psi)$
for all H **in** Ω **do**
  if $hasIntersection(H) =$ **false then**
    $\Omega \leftarrow \Omega \setminus \{H\}$
  else
    for all I_j **in** H **do**
      $I \leftarrow getInstancesThatInclude(I_j)$
      for all I_k **in** I **do**
        if $contains(H, I_k) =$ **false then**
          $\Omega \leftarrow \Omega \setminus \{H\}$
return Ω

Figure 8.3: Mapping algorithm to construct an Evidential frame of discernment

`SemanticInclusion` and `SemanticIntersection` are two procedures. Their definitions have both been provided in Chapter 7. More particularly, the `SemanticIntersection` procedure is given in figure 7.2 where the set of instance Υ is here equal to the set of candidate instance Ψ. Considering the same set of candidate instances of cardinality N, the semantic inclusion determination is performed $N(N-1)$ times. The semantic intersection calculi is computed $N(N-1)/2$ times, since the similarity between two instances is symmetric.

8.2.3 Mass Re-Assignment

Finally, every *DS concept* expressed by $\{S, m_{value}, A\}$ in the semantic beliefs formalism is mapped to a mass function of the form: $m_S^\Omega(f_{map}(A)) = m_{value}$ in the classical Evidential formalism.

8.2.4 Example

We take one more time our main thread example. The figure 8.4 below recalls the semantic set operators results obtained in the previous chapter.

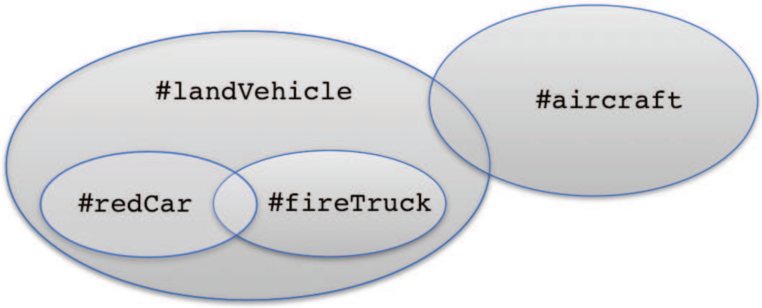

Figure 8.4: Graphical recall of the results obtained by the semantic set operators.

We first determine our frame of discernment by using the formula 8.1. This process is depicted on table 8.5. Each line of this table corresponds to one possible H_M of formula 8.1 coming from the combination of all candidate instances. Two columns ("Removal reasons" label) indicate if an H_M satisfies all constraints so as to be considered as an atomic state of the frame of discernment. The final column of this table gives a simplified (shorter) label for those final atomic states. Be careful, these shorter labels are not equal to the set of instances matched in the first columns: these shorter labels are the resulting atomic states that have been created by taking into account this set of instances.

Finally, from the above table, we get 6 elements for the frame of discernment, where $\Omega = \{a, b, c, d, e, f\}$. The graphical interpretation of these atomic states is given on figure 8.6 where each atomic state is delimited by a single zone in the schema.

Following formula 8.3, candidate instances of Ψ are now translated as such:

- `#aircraft` $= \{e, f\}$
- `#redCar` $= \{c, d\}$
- `#fireTruck` $= \{b, d\}$
- `#landVehicle` $=\{a, b, c, d, f\}$

Once the frame of discernment has been obtained and according to the initial mass values (see section 6.3), we can reformulate the basic mass assignment of the scenario within the mathematical Dempster-Shafer formalism:

- The mass function for the $\#radar$ source:

  $m_{\#radar}(\{a, b, c, d, f\}) = 0.6$

  $m_{\#radar}(\{e, f\}) = 0.1$

  $m_{\#radar}(\{a, b, c, d, e, f\}) = m_{\#radar}(\Omega) = 0.3$

- The mass function for the $\#human$ source:

Combination of candidate instances				Removal reasons		
#aircraft	#redCar	#fireTruck	#landVehicle	Inclusion constraint	Intersection constraint	
			×	✓	✓	a
		×		✗ (#fireTruck $\subseteq_{sem}$ #landVehicle)	✓	-
		×	×	✓	✓	b
	×			✗ (#redCar $\subseteq_{sem}$ #landVehicle)	✓	-
	×		×	✓	✓	c
	×	×		✗ (#redCar $\subseteq_{sem}$ #landVehicle and #fireTruck $\subseteq_{sem}$ #landVehicle)	✓	-
	×	×	×	✓	✓	d
×				✓	✓	e
×			×	✓	✓	f
×		×		✗ (#fireTruck $\subseteq_{sem}$ #landVehicle)	✗ (#fireTruck $\cap_{sem}$ #aircraft $= \emptyset$)	-
×		×	×	✓	✗ (#fireTruck $\cap_{sem}$ #aircraft $= \emptyset$)	-
×	×			✗ (#redCar $\subseteq_{sem}$ #landVehicle)	✗ (#redCar $\cap_{sem}$ #aircraft $= \emptyset$)	-
×	×		×	✓	✗ (#redCar $\cap_{sem}$ #aircraft $= \emptyset$)	-
×	×	×		✗ (#redCar $\subseteq_{sem}$ #landVehicle and #fireTruck $\subseteq_{sem}$ #landVehicle)	✗ (#redCar $\cap_{sem}$ #aircraft $= \emptyset$ and #fireTruck $\cap_{sem}$ #aircraft $= \emptyset$)	-
×	×	×	×	✓	✗ (#redCar $\cap_{sem}$ #aircraft $= \emptyset$ and #fireTruck $\cap_{sem}$ #aircraft $= \emptyset$)	-

Figure 8.5: Frame of discernment creation.

$$m_{\#human}(\{c, d\}) = 0.2$$

$$m_{\#human}(\{b, d\}) = 0.4$$

$$m_{\#human}(\{a, b, c, d, f\}) = 0.4$$

This new formalism makes it possible to apply directly classical combination rules of the Dempster-Shafer theory, and then go through some decision processes.

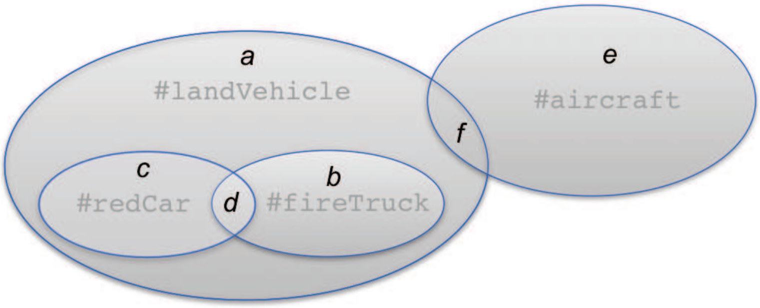

Figure 8.6: Graphical visualisation of the atomic states interpretation.

8.3 Classical Combination Process

This step is performed if more than one reporting source have stated their beliefs. In our framework, this condition is realised if more than one instance of `Reporting_source` class is associated (via the `DS_Concept` instances) to the same instance of `Uncertain_Concept` class.

In that case, the classical combination processes (seen in section 3.3.2) can be directly applied on the mass functions defined in the frame of discernment Ω.

A particular attention is paid to the Open World assumption of ontologies. Indeed, ontologies assume that we can always know more information in the future. However, the original Evidential theory assumes a closed world and that is why the measure of the amount of conflict K exists. Now, Evidential method supports both closed-world and open-world cases. In a closed-world scenario it is possible to indicate conflict of the input sources, while in an open-world scenario a high value at the empty set might indicate that the answer is outside of the current frame of discernment.

We recall here our main example, and apply the conjunctive combination first (see table 8.7).

	$m_{\#human}(\{c,d\}) = 0.2$	$m_{\#human}(\{b,d\}) = 0.4$	$m_{\#human}(\{a,b,c,d,f\}) = 0.4$
$m_{\#radar}(\{a,b,c,d,f\}) = 0.6$	0.12	0.24	0.24
$m_{\#radar}(\{e,f\}) = 0.1$	0 (conflict)	0 (conflict)	0.04
$m_{\#radar}(\Omega) = 0.3$	0.06	0.12	0.12

Figure 8.7: Preliminary step of combination: intersection product.

This table leads us to conclude:
$m(\{b,d\}) = 0.24 + 0.12 = 0.36$
$m(\{c,d\}) = 0.12 + 0.06 = 0.18$

$m(\{a, b, c, d, f\}) = 0.24 + 0.12 = 0.36$
$m(\{f\}) = 0.24 + 0.12 = 0.04$

Here $m(\emptyset) = K = 1 - (0.12 + 0.24 + 0.24 + 0.04 + 0.06 + 0.12 + 0.12) = 0.06$. The conflict, which is here interpretated as the fact that it is "something else" is equal to 0.06. This value being quite insignificant, we can consider being in a closed world and apply the Dempster's rule:
$m(\{b, d\}) = \frac{0.36}{1-0.06} \simeq 0.38$
$m(\{c, d\}) = \frac{0.18}{1-0.06} \simeq 0.19$
$m(\{a, b, c, d, f\}) = \frac{0.36}{1-0.06} \simeq 0.38$
$m(\{f\}) = \frac{0.04}{1-0.06} \simeq 0.04$

8.4 Decision Process

We propose in this section to review classical decision processes (introduced in section 3.3.3) through their possible use in our framework.

Within our framework, the objective of the decision process is to put forward one candidate instance among the set of candidate instances Ψ already proposed by the sources. It has to be noted, that decision is limited to one single candidate instance (not composite candidate instances). Moreover, the decision process is here not performed directly on hypotheses of the frame of discernment Ω. As a matter of fact, decisions criteria are applied to the $f_{map}(I_j)$, $\forall I_j \in \Psi$ which are the sets of atomic states corresponding to candidate instances.

For example, considering the maximum of credibility criteria, we calculate the belief (see section 3.3.1.3) of each translated candidate instance that is to say $bel(f_{map}(I_j))$, $\forall I_j \in \Psi$ and find I_j such that its belief is the maximum one.

Considering the maximum of pignistic probability, we first calculate the pignistic probability distribution over the candidate instances, thanks to the following equation adapted from the classical one.

$$\forall I_j \in \Psi,\ P_m(f_{map}(I_j)) = \sum_{\emptyset \neq B \subseteq \Omega} m(B) \frac{\mid f_{map}(I_j) \cap B \mid}{\mid B \mid} . \tag{8.6}$$

Then, we choose the candidate instance whose pignistic value is the maximal.

Other criteria could be further listed according to the specialized literature, but they would be implemented the same way according to the rough method presented here.

Finally, one could also think to extend this decision criteria to instances of ontology specially created during the decision process by applying first the classical decision process on composite hypotheses of the frame of discernment and then "translate" or make an inverse mapping of f_{map} to get instances that were not necessarily part of the set of initial candidate instances Ψ. Further thoughts are leaded in section 12.2 as future perspectives of this work.

8.5 Summary of the Semantic Beliefs Representation and Reasoning Process

This last section of Part III intends to recap the major steps detailled by Chapter 6, 7 and 8. All these chapters were turned towards the set up of a semantic decision-support system, that we use to call as our *fusion of semantic beliefs* system.

To sum up, we first recall back figure 8.8 which has been previously shown in the introduction of this document.

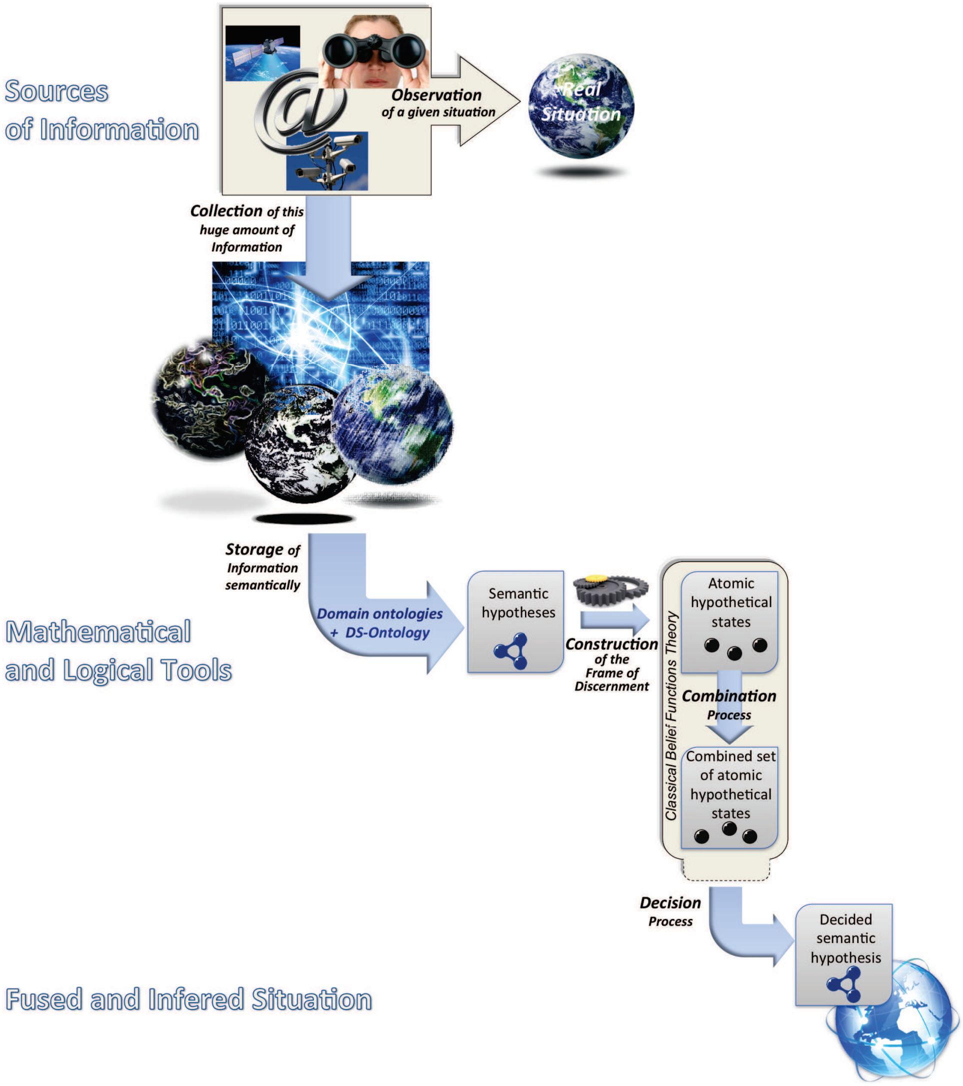

Figure 8.8: Detailled semantic decision-support workflow.

This figure includes the following steps (as mentioned in [Bellenger 2012b]):

1. First, we collect information coming from different sources for a same phenomenon;

2. We store the received information in a knowledge base through instances of a given ontology;

3. According to the initial information sources or to some information processes, degrees of belief are associated to the instances describing the phenomenon, thus defining candidate instances for the description of this phenomenon;

4. Then, we determine semantic inclusions between each candidate instances of a given phenomenon;

5. We calculate also the semantic similarities between each candidate instances of a given phenomenon;

6. From the previous semantic similarities, we calculate a threshold, above which we considered that semantic similarities represent semantic intersections;

7. According to semantic inclusions and intersections, we map the set of all candidate instances of a given phenomenon into an evidential frame of discernment, i.e., a set where all elements of this frame are exclusives;

8. We reassign the belief degrees into this frame of discernment in such a way that the assigned degree of belief of a candidate instance is equal to the sum of belief degrees of the union of its mapped elements in the frame of discernment;

9. In case several sources of information have stated their beliefs, we combine them at the level of the frame of discernment;

10. We applied some decision criteria existing in the Evidential theory, in order to determine which set of elements of the frame of discernment corresponds to the best hypothesis, among the sets of elements mapped from a candidate instance;

11. Finally, we return the initial candidate instances with their associated degree of beliefs as well as a decision support indicating the best candidate instance.

Steps 1, 2 and 3 are the initialization steps that gather the input pieces of information for the decision-support tool in an ontological format described in Chapter 6.

Steps 4, 5 and 6 have been the topic of Chapter 7 entitled Semantic Set operators.

Steps 7 and 8 are related to the projections of semantic beliefs in the Belief Functions Theory (described in this Chapter), in order to apply 9 and 10 stages. Step 11 gives the output of the decision-support tool.

Part IV of this document focuses on the implementation of the framework we formally presented. It also concentrates on experimentations that have been led in order to validate the approach.

Part IV

Implementations and Experimentations: The FusionLab Platform

CHAPTER 9

Implemented Systems and Tools

Contents

This chapter overviews software implementations which support practical application of theoretical concepts and reasoning process presented in Part III of this document. A Java library relying on the OWL API[1] has been encoded. As depicted on figure 9.1, this library has been used for the development of a Protégé plugin as well as for its encapsulation in a *FusionLab* service.

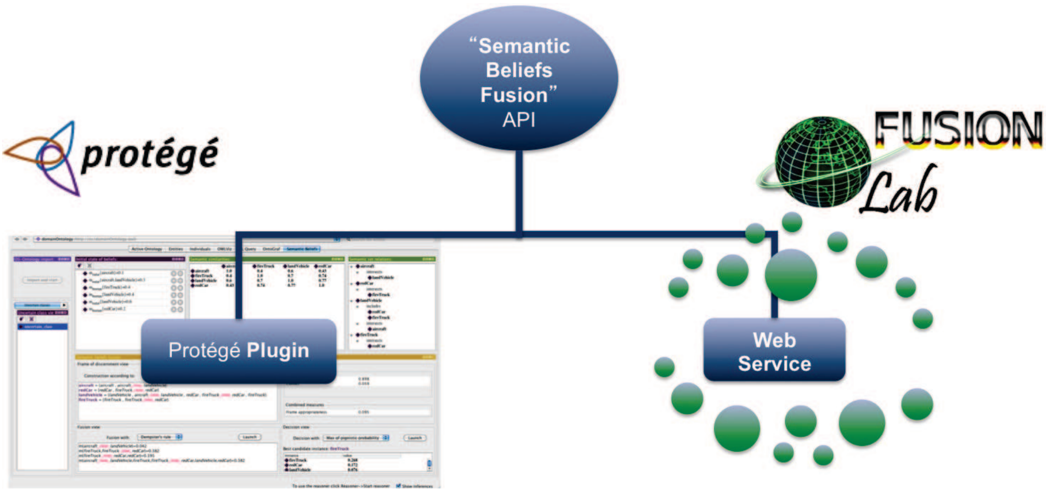

Figure 9.1: Implemented tools following the "Fusion of Semantic Beliefs" Application Programming Interface.

[1]http://owlapi.sourceforge.net/

9.1 A standalone Graphical User Interface

In order to test, promote and validate our theoretical approach, a graphical interface was needed. It should give the user the opportunity to get an outlook of what the system handles and how it is processing. In other words, this graphical interface provides the transparency and the traceability needed by the user so that he could better accept it. In addition to this passive behaviour, the user may want to be more active and have the possibility to parameterize himself some processes in order to understand the effects of each different parameters and possibly configure parameters to its type of objectives and use cases. Moreover, starting from scratch, this interface should permit to create and edit graphically and easily "semantic beliefs". The user being not aware of the *DS-Ontology* terminology should still be able to instantiate its domain ontology in an uncertain manner, by partially graphically hiding the way of structural storage of these information. Finally, it would be ideal if the graphical style of the interface was consistent with common tools adopted by ontology engineering community.

The Protégé system[2] (also called Protégé-OWL) has been introduced at the beginning of this document in section 4.5 and actually permits to manage graphically and easily ontologies. Clearly, it is the ontology editor which benefits from the largest interest in the community. Protégé is indeed supported by a strong community of developers and academic, government and corporate users, who are using Protégé for knowledge solutions in diverse areas. In our context, one important remark is that Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for application development[3].

Our graphical interface has thus been built upon the open source Protégé system. More precisely, it has been developed as a set of Protégé plugins, encapsulated as a new tab in the Protégé user interface. A screenshot of the tab entitled "Semantic Beliefs" is included in figure 9.2. It is instantiated with our thread example that has been continuously handled in Part II and Part III. The slots of this tab are organized according to the chronological steps of a fusion process starting from the creation of semantic beliefs till the decision process. There are six distinct slots composing the tab, which are named:

- *DS-Ontology import.* This slot permits to add the DS-Ontology.owl ontology file into the ontology already opened in Protégé.
- *Uncertain class view / Uncertain property view.* This slot permits to edit uncertain concepts: either uncertain classes as shown on figure 9.2 or uncertain properties. Once having creating at least one uncertain concept the following slot is enabled.
- *Initial state of beliefs.* It concerns the edition of initial semantic beliefs specific to a selected uncertain concept.

[2] http://protege.stanford.edu
[3] http://protege.stanford.edu/doc/dev.html

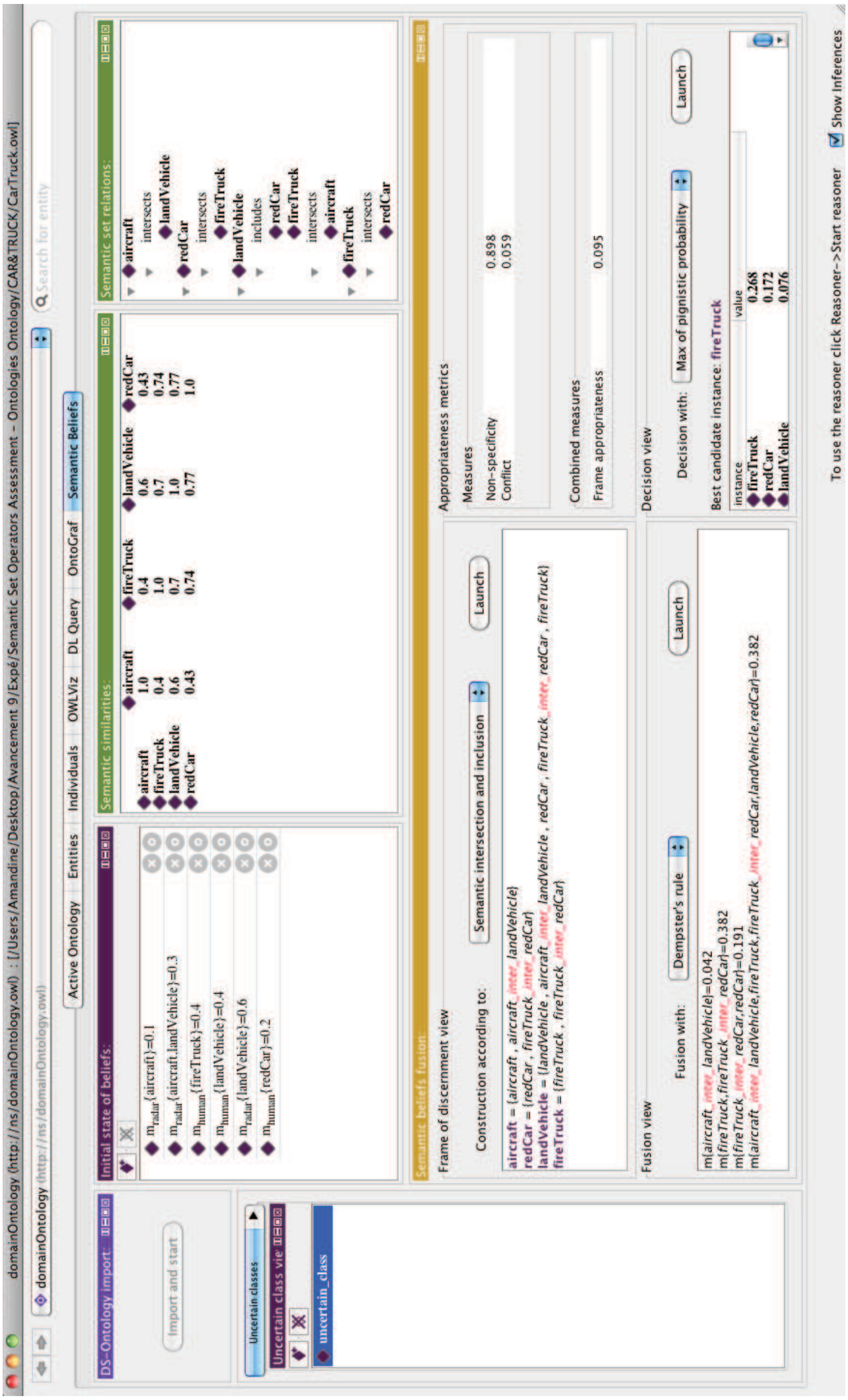

Figure 9.2: Global semantic decision-support process example.

- *Semantic similarities.* It is a passive (only readable) slot that shows the semantic similarities calculated between all semantic hypotheses mentioned in previous slot (i.e. the set of candidate instances). The obtained results for this example have been explained in section 7.3.

- *Semantic set relations.* This slot is also a passive one, which shows the semantic set relations (see sections 7.1 and 7.2) discovered between the semantic hypotheses. Only computed inclusions and intersections are represented here, in other words, intersections consecutive to inclusions are not shown here for ease of readability.

- *Semantic beliefs fusion.* This slot consists of multiple elements:

  - the frame of discernment view (results for this example have been explained in section 8.2.4) which can be created by taking all semantic set relations into account (the one by default), or only semantic inclusions or only semantic intersections or none.
  - some appropriateness metrics concerning the frame of discernment newly created. It includes the amount of conflict related to the new modeling. This amount decreases while taking into account the semantics of candidate instances (it reduces modeling errors). It presents also the non-specificity of the new frame of discernment. This amount logically increases by taking into account the semantics of candidate instances (it adds new atomic states). These metrics have not been further exploited in this work.
  - the fusion view (see section 8.3) which presents results obtained with the Dempster's rule, the conjunctive or the disjunctive rule (and others could be easily added).
  - the decision view (see section 8.4) recaps the "best" candidate instance according to the chosen criterion: maximum of credibility, of plausibility or of pignistic value.

9.2 An Information Fusion System

This thesis having been realized within the Cassidian company, one important requirement regarding the implementation was to realize a capitalized prototype within the current platform for high-level information fusion processing.

Since 2010, Cassidian is developing an integration platform dedicated to this type of processing and is called the *FusionLab*. We have actively contributed to the born and rise of this platform. Our contribution is twofold here. First, we have proposed technical and modeling solutions for the semantic exploitation of the information. Secondly, we have integrated the capability to have a fusion mechanism handling uncertain semantic statements of the world.

This section is thus going to sketch the software architecture of the platform. Then, we will focus on the semantic layer of the FusionLab. Afterwards, we will briefly present how our method for fusing semantic beliefs has been integrated to this platform.

9.2.1 FusionLab

The *FusionLab* platform has been set up in order to provide an environment to easily and rapidly integrate, demonstrate, and evaluate innovative capabilities of information fusion.

The aim of this platform is to be flexible enough to support various operational needs and in particular multiple scopes of application (ground, maritime, air or joint picture but also security domains such as border surveillance and critical infrastructure security, camp protection). The *FusionLab* comes from the combination of needs of various research projects in which the *Information Processing, Control and Cognition* department of Cassidian is involved. The idea is to propose tools for generic reasoning and to use them differently according to the context of use but also of being able to aggregate them.

It is based on the paradigm of service oriented architectures. Its architecture is largely inspired from the one of the *WebLab*™[Giroux 2008] developed also within Cassidian[4], which aims at providing intelligence systems and any other applications that need to process multimedia data (text, image, audio and video).

Within the *FusionLab* platform, each processing is identified as a service which can be then used in a whole particular application. By composing various services, it becomes possible to create information fusion processings integrating multiple heterogeneous components. This platform is born from the Cassidian expertise in this domain since it often positions itself as an integrator for its clients.

The *FusionLab* is thus based on a generic data model, describing different sensed objects of a situation. In order to describe each situation element, it is possible to annotate them. For example, a service of abnormal behaviour detection, able to discover anomalies considering a particular object and / or its relations with others, will add annotations on the detected anomaly. These annotations are based on semantic statements formalized following the terminology of the *FusionLab* set of ontologies. Further details on these ontologies are provided in the following section 9.2.2.

9.2.1.1 Logical Architecture of FusionLab

Figure 9.3 presents a global schema of a logical architecture of an application based on the *FusionLab*. The architecture is composed of five layers that we are going to introduce.

The first layer represents the access to data. There is an access point for data input (for example, in case of simulation: choose scenarios, speed of scenario replay),

[4] `http://weblab.ow2.org`

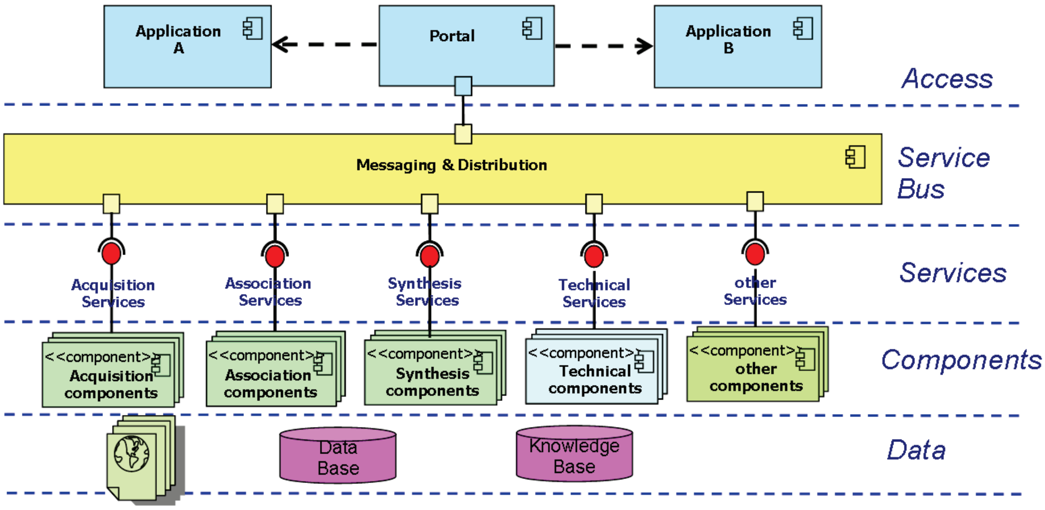

Figure 9.3: Logical architecture of an application based on the *FusionLab*.

and an access point for presentation of the COP - Common Operational Picture (see section 2.1).

The second layer is embodied by the service bus. It is a key element in service oriented architecture. It enables the distribution of messages to services.

The third layer deals with services. It is composed mainly of generic interfaces of services (contract) proposed by the *FusionLab*. This layer is actually in charge of integrating components of the lower layer.

The fourth layer, as previously mentioned, is composed of various components integrated in the platform. They are also-called business processes. It can be any type of software components used in the composition of a platform of information fusion processing (such as data alignment, association, tracking, identification, classification, visualisation, performance measurements, etc.).

The last layer gathers all persistent information in the system. It can be external *a priori* databases (contextual, intelligence, etc.) or information in the *FusionLab* OWL format that we are going to detail in the following section. The latter corresponds to the information produced and needed by the services as the application is running.

9.2.2 Semantic Glue of the FusionLab

The *FusionLab* platform addresses the problems of high-level information fusion. In that context, the use of information at a semantic level is considered. We have chosen ontologies (see section 4.6) as the underlying knowledge exchange mean.

Within this framework, ontologies can be compared as semantic glue between the services that enables them to deal with the information in the same and consistent way and to exchange this information. This information is shared as ontological instances according to the underlying terminology of the *FusionLab* ontologies. As a consequence, flexibility of the platform comes also from this suitable underlying knowledge base.

9.2.2.1 Structure of FusionLab Set of Ontologies

We have taken into account the service oriented architecture paradigm when proposing a global structure for our terminology knowledge modeling.

As the figure 9.4 depicts it, the model is split into three ontological layers: *FusionLabUpper*, *FusionLabCore* and application domain ontologies. The first level of ontologies serves as a generic language for services interface and is presented in section 9.2.2.2). The second level of ontologies (see section 9.2.2.3) is more semantically detailed and concerns the definition of concepts that are recurrent in any fusion applications. Finally, the third level consists of diverse ontologies that are specific to particular domains of applications (see section 9.2.2.4). The second and third levels are used by services so as to perform their processing.

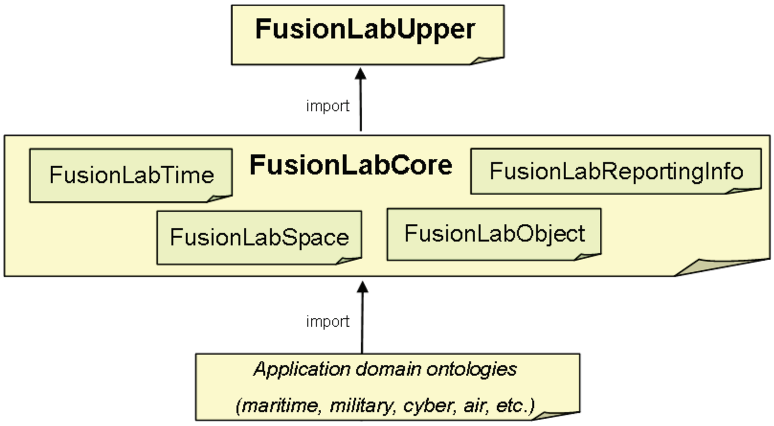

Figure 9.4: FusionLab Ontologies structure.

9.2.2.2 FusionLabUpper Ontology

The *FusionLabUpper* ontology constitutes a high-level overview of the global knowledge model of the *FusionLab*. It serves as a broad guide to the contents of the knowledge model specification. As a matter of fact, its goal is to indicate the scope of the model in covering information categories of interest to the operational user/services in the *FusionLab* platform. It describes the basic and abstract notions that are common to each information fusion application domain. It is generic enough to accommodate joint, land, air, and sea situations. It comprises the following concepts: object, action, space and time locations, observing source, etc.

The JC3IEDM[5] (Joint Consultation Command Control Information Exchange

[5] https://mipsite.lsec.dnd.ca/Public%20Document%20Library/Forms/AllItems.aspx?RootFolder=%2FPublic%20Document%20Library%2F04-Baseline_3.1%2FInterface-Specification%2FJC3IEDM

Data Model) is precisely defining this set of minimum data. It is an established data model in the NATO community and is managed by the Multi-lateral Interoperability Programme[6] (MIP). It specifies the minimum set of data that needs to be exchanged in coalition or multi-national operations. In comparison to other military models, the JC3IEDM is seen as a relevant defense and security standard, well structured through its three levels of management: conceptual, logical and physical. The JC3IEDM conceptual model has the specific objective of being a top-level data model of generalized concepts. In the JC3IEDM conceptual model, there are indeed only 15 independent entities composing the model.

The aforementioned statement leads us to the following report: the JC3IEDM model and especially its conceptual model would serve as a sound base for our *FusionLabUpper* ontology. It permits us to rely and be compatible with standards promoting interoperability. This choice has also been made in [Dorion 2005, Matheus 2006, Wartik 2009] for example, where the authors have decided to rely on the JC3IEDM to create their own ontologies. However, the JC3IEDM being specified as a relational data model, we adapted manually (contrary to [Dorion 2005, Matheus 2006, Wartik 2009] which have adopted an automatic transformation of the whole Jc3IEDM) the translation into an OWL ontology - partially depicted on figure 9.5.

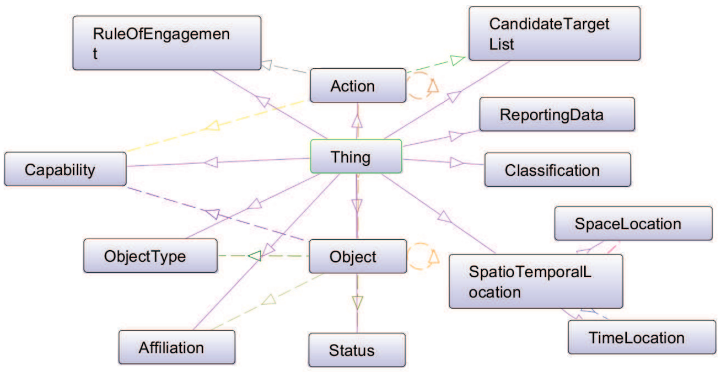

Figure 9.5: FusionLabUpper ontology overview.

The aim of this first layer of the *FusionLab* set of ontologies is twofold. It should serves as the mother classes of annotations that will be further specified by the other layers of the *FusionLab* set of ontologies. It is thus a solid base, which should be not modified even when FusionLabCore ontologies may be superseded by new ones (see section 9.2.2.3). Its second aim is to be used as the description of service interfaces. Indeed, the concepts used to described the information needed in input of services and that are expected in output of services comes from the *FusionLabUpper* layer. However, one can remark that the *FusionLab* is not meant to be a platform supporting semantic services using ontology to describe themselves, with for instance

[6] https://mipsite.lsec.dnd.ca/

OWL-S[7] or WSMO[8], which are still 2004 / 2005 W3C submissions. Instead, the *FusionLab* is meant to provide a service oriented architecture where services are exchanging and reasoning over knowledge (a.k.a. ontologies). Therefore, *FusionLab* is based upon the standard way to provide service-to-service interaction, i.e. through WSDL[9]. However, an XML schema (see section 4.3.2.2) is needed to support this interaction. It is limited to generic concepts that comes from the adaptation of the *FusionLabUpper* ontology and have in addition a specific slot providing the concrete message content (the "annotations") as an ontology.

9.2.2.3 FusionLabCore Ontology

The *FusionLabCore* deepens the high-level concepts of the *FusionLabUpper*, such as time, space, object taxonomy, reporting data, etc. into a detailed description. This partition avoids too much cross-dependency in the model and permits to use as much as possible already existing standards to promote interoperability. This type of information is understood within services so as to perform their processing.

The knowledge content includes time (relying on the *OWL-Time*[10] ontology [Hobbs 2006]) and space representations (including notions from OGC's standards[11]) which are common to any Information Fusion system.

The core ontology also consists of an object type taxonomy and affiliation (declined in Organization, Person, Material, Facility and Feature, etc. inspired from the JC3IEDM), and some reporting information descriptions (associating to each report, its source, its effective and reporting date/time, the degree of reliability of information, etc.).

The description of each *FusionLabCore* ontologies has been provided with much detail in [Bellenger 2011b].

As said before, almost all of the above mentioned notions have already been included in the *FusionLabUpper*, the role of the core ontology here is to deepen/detail and describe each branch of these notions in order to be effectively useful by services. In other words, it breaks down the high-level concepts into specific information which are regularly used.

9.2.2.4 FusionLab Application Domain Ontologies

The last level is composed of application domain ontologies, as for example the *FusionLabMaritime* ontology. They are relying on the *FusionLabUpper* and *Core* ontologies but are relative to a particular instantiation of the platform concerning a particular need. For the *FusionLabMaritime* ontology, a more detailed taxonomy of vessel (*n.b.*: the vessel concept coming from the *FusionLabObject* of the core set of ontologies) including for example Pleasure Boat, Ferry Boat and the specific

[7] http://www.w3.org/Submission/OWL-S/
[8] http://www.w3.org/Submission/WSMO/
[9] http://www.w3.org/TR/wsdl
[10] http://www.w3.org/TR/owl-time/
[11] http://www.opengeospatial.org/

measures to the maritime domain (e.g. nautical miles) as well as some specific kinds of maritime events (e.g. transshipping) are provided in that layer.

9.2.2.5 Related Components/Services

As seen previously, the *FusionLab* platform is based on the *FusionLab* ontologies. Native *FusionLab* services are using this set of ontologies in order to understand and reason over information. Yet, semantic services are not always compliant with *FusionLab* ontologies (services that come from other projects, or that are anterior to the creation date of this platform, etc.). As a result, in order to achieve semantic interoperability between services, we need to find some mappings between ontologies. Thus, services for mapping ontologies can be specified in the *FusionLab* framework, whose goals are to find correspondences between semantically related entities of different ontologies. A set of correspondences is also called an alignment. Mapping is the oriented, or directed, version of an alignment: it maps the entities of one ontology to one entity of another ontology. More concretely, it transforms instances from a specific ontology into instances of the *FusionLab* ontologies. The *FusionLab* framework offers the user the possibility to graphically create the rules through drag-and-drop actions, or/and to add some specific rules written by hand (for more details, see [Bellenger 2011b]).

So as to realize the fifth layer of the *FusionLab* architecture, providing persistency of the information, a semantic repository is required[12]. We have thus some technical services dedicated to the management of the information contained within this repository namely for the modification (delete, modify, write) or for request only.

9.2.3 The DS-Ontology and the Semantic Beliefs Fusion Service

Let's see now how our specific contributions are physically enclosed in the *FusionLab* platform so as to propose innovative solutions to tackle with the inherent uncertain nature of the information manipulated.

First of all, the *DS-Ontology* has been included in the *FusionLab* set of ontologies. It has been logically put under the *FusionLabReportingInfo* branch of the ontology. Let's remind that the latter gathers information about the reporting context of the newly created statements (reporting source, reporting time, reporting space, reporting uncertainty) and thus can be seen as meta information concerning facts of the situation.

As introduced on figure 9.1, the "Semantic Beliefs Fusion" API has been included as a capitalized service within the *FusionLab* platform. Regarding figure 9.3 on page 130, this service belongs to a *synthesis* service type. Consequently, it follows a generic *FusionLab synthesis* interface. In that configuration, other services (*association*, *assessment* services for example) are in charge of providing possible hypotheses to the input of the *synthesis* service which itself returns

[12] for example Sesame (`http://www.openrdf.org/`

a synthesized view of the situation by processing the combination and decision steps.

We can remark that for the moment no implementation link relates our Protégé plugin with the *FusionLab* system. However, it is easy to foresee the added value of integrating both in order to provide the domain expert user a graphical interface to visualize and have traceability of the decision process and / or to permit him to integrate its own beliefs to the system. This is part of future works.

In the next chapter, section 10.2 provides a description of the implementation context in which this service is used for a maritime surveillance project.

Chapter 10

Experimentations and Evaluations

Contents

The evaluation of our framework for the fusion of semantic beliefs is a complex task since we address several research axes. Moreover, as an innovative approach, which enables processes that were not automatically possible until now, it makes no sense to compare it directly to other approaches. No benchmark is for example related to our works.

However, this chapter proposes to take a look at experimentations or thoughts we have conducted to perform this evaluation task on different levels. As previously mentioned, our work is in charge of mimicking the human reasoning on the semantic level to construct a proper frame of discernment, thus the first section of this chapter focuses on the adequacy of the semantic set operators we have introduced with the human reasoning. We then focus more on the integration of such a system into a global information fusion platform such as the *FusionLab* through the ETUR working group and a *FusionLab* based application of maritime surveillance.

10.1 Semantic Set Operators Evaluation

This first experimentation has been designed in order to validate the notions of semantic inclusion and semantic intersection introduced in Chapter 7. These operators have been defined so as to be consistent with the human reasoning while facing semantic concepts that need to be characterized through classical set operators. In this section, a protocol to assess these operators is proposed. It relies on the comparison of results provided by our system and the ones issued from human reasoning.

10.1.1 Performed Experimentation

10.1.1.1 General Protocol

18 persons have participated to this experiment. These persons will in the following be referred to as test persons. Indeed, they are not users of a particular application we have implemented, but they serve us more as validators of our theoretical approach. They have been selected as having already basic knowledge on ontology engineering. Another requirement for the experiment was that the user should have an ontology editor installed on his computer.

First an email has been sent to potential users. This email described briefly the need of such experimentation, its context and the requirements to participate to the experiment. It also explains how to participate to the experiment and the approximate time it should take (Ca. 30 minutes). Finally, the email provided a web link to access an on-line questionnaire.

This questionnaire has been implemented as a Google Doc form in French (since recipients of this email - and thus potential test persons - were French people). The Google Doc form[1] is a tool that allows collecting information via a personalized quiz (see our initial questionnaire page shown on figure 10.1). The information is collected and automatically connected to a spreadsheet. The spreadsheet is populated with the questionnaire responses of the test persons.

The raw questionnaire is depicted in appendix D. It is composed of three parts, each focusing on a particular ontology. We detailed the three ontologies in the following section. For each part, first a brief description of the ontology was proposed. Then, the test person was invited to download the ontology and open it through an ontology editor (for example, the Protégé editor). Finally, he had to evaluate a set of propositions of the questionnaire by saying if they seem true or false with some degrees of certainty ("true", "rather true", "rather false" or "false"). Each part of the questionnaire had different types of questions:

- Some preliminary questions were asked to assess the level of the user in terms of ontology modeling and reasoning. It can be seen as a filter process to weight the following results of this user.
- The other questions, which deal directly with semantic set operators, are the real subject of our experimentation.

In total, there were 30 questions to answers. Among them, 10 were preliminary questions which were spread over the three parts. Moreover, for each ontology questionnaire, we have tried to keep the same level of "difficulty".

[1] `http://support.google.com/drive/bin/answer.py?hl=en&answer=87809`

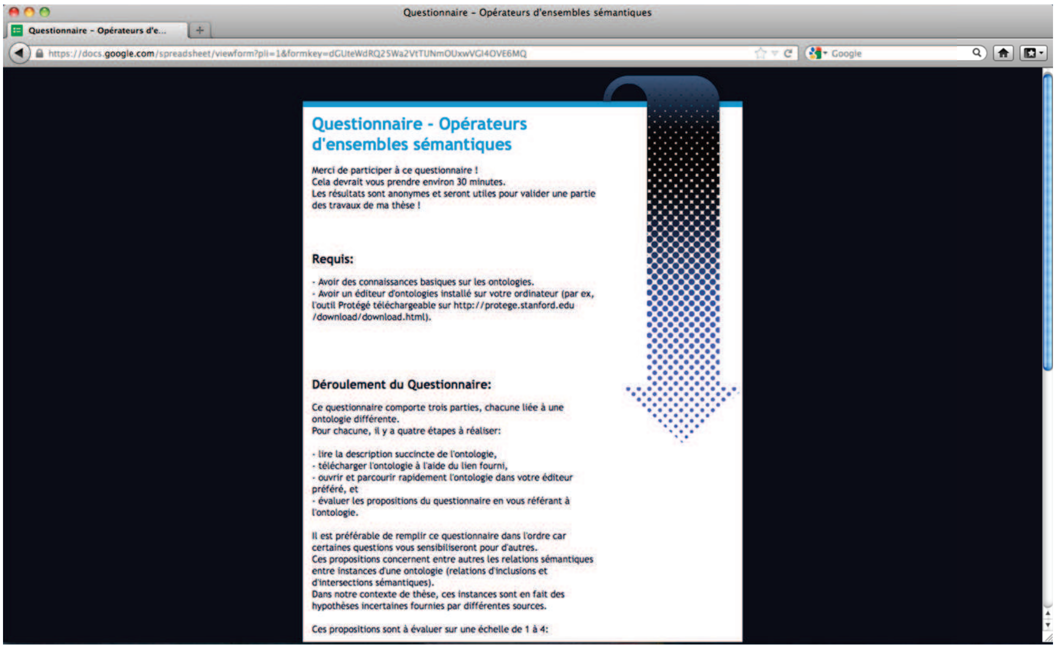

Figure 10.1: Start web page of our Google Doc questionnaire (in French).

10.1.1.2 Ontologies used in the Experimentation Process

Three ontologies have been used to support the experimentation. These ontologies have been selected so as to be not specific to a particular expert domain, to be enough expressive and to come from the public domain. In such a manner, three well-known ontologies that are generally used as a teaching purpose in the community of ontology engineering have been chosen: the Pizza ontology, the People and Pets ontology and the Travel ontology. Their domain of interest are slightly described in the following. Snapshots of the Protégé editor reproducing what the user could have look at when browsing the ontologies are given in Appendix E. Moreover, for each of them, metrics on their number of classes, logical axioms, etc. are provided in table 10.2.

These ontologies have been slightly adapted to reflect the experiments: only some instances have been added so that semantic set operators relate "interesting" instances of the ontologies. Moreover, we have paid a particular attention to the naming of the instances of interest. For each ontology, instances are either named through descriptive labels (e.g.: "hotPizza" for the Pizza ontology), or nominative labels (e.g.: "Minnie" for the People and Pets ontology), or arbitrary variable names (e.g.: "InstanceA" for the Travel ontology).

The adapted Pizza Ontology: This ontology describes the pizza domain. It goes from the pizza base (crust, thickness) till the pizza toppings (mozzarella, pepper, tomato, etc.). Relying on these notions, the ontology defines some pizza cat-

	Pizza	People & Pets	Travel
Class count	100	23	34
Object property count	8	10	6
Data property count	0	0	4
Individual count	17	16	22
Logical axiom count	731	55	104
DL expressivity	SHOIN	ALCHIN	SOIN(D)

Figure 10.2: Ontologies Metrics.

egories (vegetarian, cheesey, etc.) as well as pizza names (Margherita, la Reine, etc.).

This ontology is adapted from the one found on `http://www.co-ode.org/ontologies/pizza/2007/02/12/` and provided by the Manchester University. The Pizza ontology is nearly the most "famous" ontology, which is mainly due to the fact that it is often provided as teaching purpose and for semantic tools testing[2].

Snapshots of the Protégé editor when browsing this ontology are given in figure E.1 of appendix E (page 189). Elements with a yellow background are ontological axioms that have been inferred from an OWL reasoner proposed in the Protégé editor.

The adapted People and Pets Ontology: This ontology describes different persons categories (male, female, parent, child, etc.) and logical links defining categories from others (old lady, woman, etc.). Animals are also represented (dog, cat giraffe, etc.) as well as their relations with humans (is pet of, etc.).

The People and Pets Ontology comes from `http://protege.cim3.net/file/pub/ontologies/people.pets/people+pets.owl` and has been designed for the purpose of a tutorial on OWL by Sean Bechhofer, Ian Horrocks, and Peter Patel-Schneider at the ISWC 2003 conference.

Here, however, only a subset of this ontology has been taken (without a loose of expressivity of this subset) in order to decrease the size of the initial ontology. Nearly half of the concepts have been removed in the ontology we proposed for the experimentation (from 60 classes to 23 classes).

Snapshots of the Protégé editor when browsing this ontology are given in figure E.2 of appendix E (page 190).

The adapted Travel Ontology: This ontology is dedicated to the tourism domain. Concepts related to the type of accommodations (hotel, bed and breakfast, etc.), activities (adventure, sport, relaxation, etc.) and destinations (beach, rural area, city, etc) are described.

[2]see for example the tutorial of Protégé editor `http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/`

This ontology has been retrieved "as is" from `http://protege.cim3.net/file/pub/ontologies/travel/travel.owl` and has been originally designed by Holger Knublauch from the Stanford University for tutorial purposes.

Snapshots of the Protégé editor when browsing this ontology are given in figure E.3 of appendix E (page 191).

10.1.2 Results

Since the goal of the experiments is to compare the human results on the questionnaire with the results provided by the system implementing the semantic set operators, this section first presents the results given by our Protégé plugin and then analyses and compares them to the test persons results.

10.1.2.1 Outlook on System Results

As a seek of completeness, we provide here the visual output of the Protégé plugin which implements our definitions of semantic set operators for the three above set of instances. The threshold used for semantic intersections has been configured according to the Fisher's criterion as proposed in algorithm 7.3 of Chapter 7.

Screenshots from the Protégé plugin (described in section 9.1) are displayed in figure 10.3. For each screenshot, the semantic similarity view and the semantic set relations view are provided. We recall here that the Protégé plugin view of semantic set relations displays the included instances and all the intersections (except the intersections between instances, between which there is already an inclusion relation - for ease of readability). The first snapshot on top of the figure concerns the set of instances of the Pizza ontology. The computed threshold for semantic intersection is equal to 0.66. The second snapshot deals with the People and Pets ontology. The threshold is 0.5. Finally, the bottom snapshot is about the set of instances of the Travel ontology. The semantic intersection was determined with a threshold equals to 0.93.

This section gives us also the opportunity to show that our implemented framework is able to deal with very expressive ontologies. Indeed, we have illustrated this document with mainly our thread example (with `#redCar`, `#fireTruck`, etc.) which is a toy example with a limited number of classes, properties, instances and logical axioms. Here, these three new examples demonstrate that the system is still consistent with larger and more expressive ontologies.

Finally, at this stage, a first analysis of the system answers has been performed on each question of the questionnaire related to semantic set operators. Three categories of answers system reasons can be listed: "taxonomy" category, "property" category and "inference" category. Answers may be only motivated by taking into account the taxonomy of ontology concepts. In addition to taking into account the taxonomy, one can also consider properties, when they are relating instances of interest. When these relations influence system answers, the question belongs to the second category. "Inference" category may either be related to the first or to the

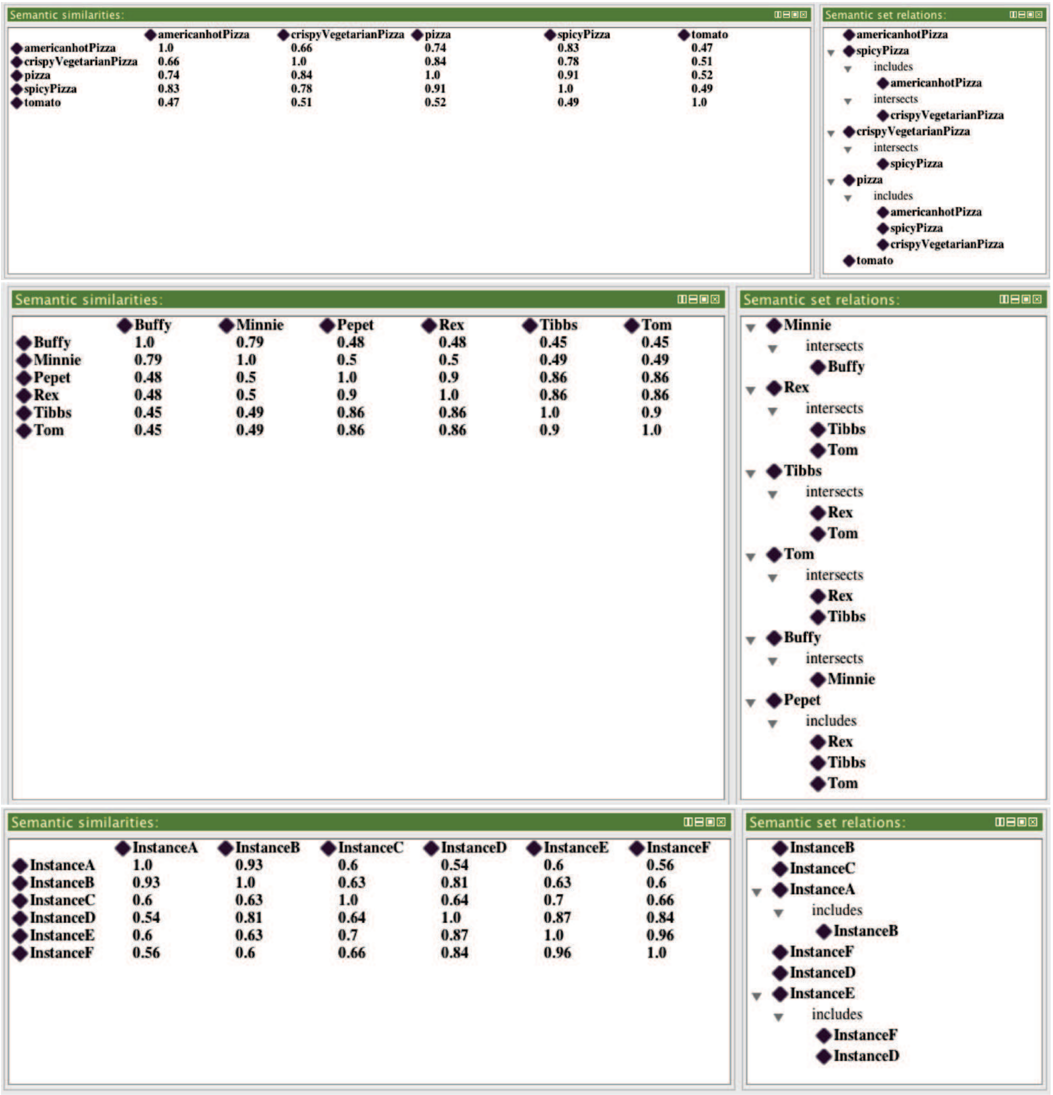

Figure 10.3: Semantic set operators for the three set of instances computed by the system.

second category, but first of all, the system has had to perform an inference process to discover either "hidden" information about the taxonomy or about properties. This remark will be useful to analyze the behaviour of test persons in next section.

10.1.2.2 Comparison with Human Reasoning

In the following, test persons answers are analyzed by comparing them to the above system results.

Different Test Person Levels

The experiment has been conducted on 18 voluntary people. Following the requirement to participate to the experiment, they all have knowledge background on ontologies. Test persons are essentially PhD students, research engineers, and academic researchers. These test persons have been grouped into different categories regarding their answers to the preliminary questions (concerning ontology modeling and reasoning). Three categories have been thus highlighted. A "high-level" test person corresponds to a person that has correctly answered more than 80% preliminary questions. The second level is called "good-level" and is assigned to exactly a 80% correct answers. Finally, the "average-level" corresponds to test persons with less than 80% correct answers. In fact, in this category, and thus in all our test persons, no one has less than 50% of correct answers, which demonstrates the good initial selection of our test persons. Consequently, no data has been rejected from our test panel. Correct answers encompass two degrees of answers: either when the user was sure or less sure but still correct answers. These levels also reflect the ability of the user to handle its ontology editor. The following pie chart (figure 10.4) shows the proportion of test persons regarding their "levels". More than a half of test persons have correctly answered to exactly 80% the preliminary questions (i.e. "good-level"). This critical percentage can be explained by the fact that there were around eight preliminary questions out of ten which were not dealing with complex inferences.

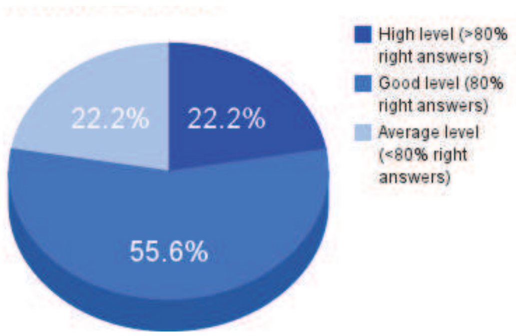

Figure 10.4: Test persons level assessment on basic notions of ontology modeling.

Semantic Set Operators Answers regarding Test Person Levels

Overall, the human intuition matches quite well our proposed definition and implementation of semantic set operators. Indeed, considering all test persons combined, more than three quarters (75,83%) of answers concerning questions about semantic set operators correspond to the answers given by the system. We use in the following the term of "matched answers" rather than "correct answers" as we did in the previous paragraph since there is no *a priori* right answers: test persons are not aware of the definitions we proposed about semantic set operators. Therefore, their answers are guided by their intuition and their knowledge in ontology modeling.

Taking into account the levels of test persons leads us to some important remarks.

There are namely 68,06% matched answers for "average-level" test persons. Then, there are 73,19% matched answers for "good-level" test persons and 92,01% for the the remaining "high-level". Diagram on top of figure 10.5 shows visually this trend. On this type of graphics, each vertical bar is split into maximum four horizontal layers. On the bottom, the green layer depicts the "fully" matched answers (the answers for which the person was sure and that are compliant with the system), the above light green layer represents still matched answers but for which the user has ticked that he was not exactly sure about it. On the top, the red layer respectively depicts the non "fully" matched answers and the light red: the non matched answers with doubt. This graphic represents this trend with an increasing amount of all green surfaces (and a decreasing amount of red surfaces) from the right to left bars.

This trend is verified for each ontology of the questionnaire as underlined in the tree diagrams on the bottom of figure 10.5. This increasing trend while going in the higher levels is a really demonstrative argument that the deeper and better ontology modeling knowledge background the person has, the more he agrees on our semantic set operators definitions.

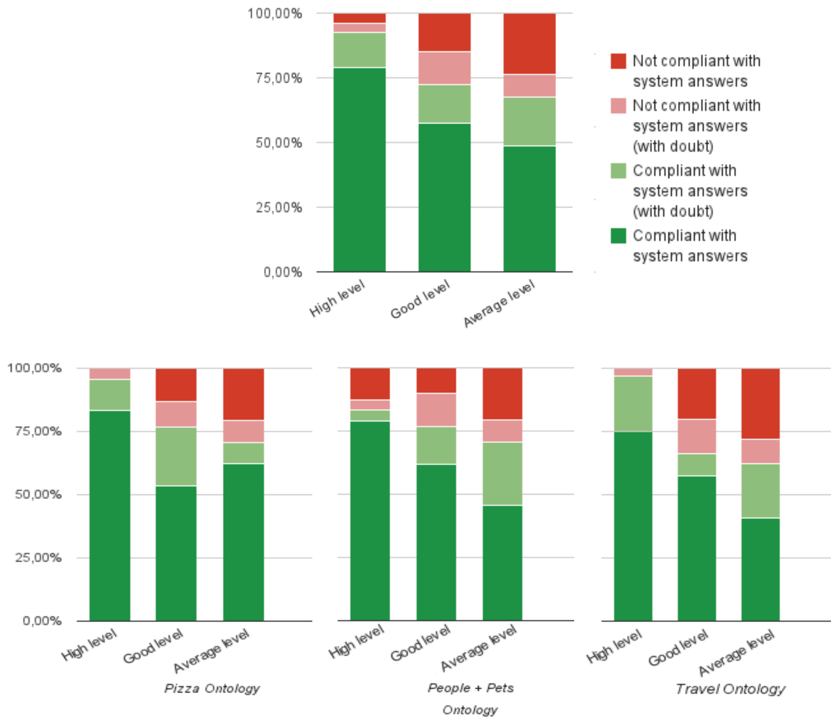

Figure 10.5: Influence of test person levels on their answers of semantic set operators relative questions for each of the three ontologies.

Influence of Ontologies and Instance Labels

Looking back at the above figure 10.5, one can remark that there is another observable trend among the three test-support ontologies. For each test person level, the percentage of matched answers is generally better for the Pizza ontology than for the People and Pets ontology, itself better than the Travel ontology. This observation is more striking on figure 10.6, where all matched answers are represented for each of the three ontologies (without considering test person levels). First, one can recall, that we have paid a particular attention when proposing the experimentation protocol regarding the labels of instances of the three ontologies (see section 10.1.1.2). Instances of the Pizza ontology have been named with descriptive labels, whereas instances of the People and Pets ontology with nominative labels and for the Travel ontology with arbitrary variable labels. One can intuitively think that descriptive labels make things easier for the human reasoning to discover our semantic set relations. For example, the fact that `#pizza` includes semantically `#spicyPizza` may seems simpler than while reasoning if `#instanceA` includes semantically `#instanceB`. As a matter of fact, in the first case, one can to some extents only refer on the label of instances and not look at the ontology while in the other case one is obliged to have a look at the ontology. One can also recall, that we have tried to create as much as possible equivalent questionnaire for each ontology in terms of "difficulty". However, this trend may also be explained by other factors such as the fact that the questionnaire was ordered and that the test person might be more concentrated with the first part of the questionnaire (which deals with the Pizza ontology) than at the end (with the Travel ontology). The counterpart of this argument would be that, on the contrary, the test person should do better with the questionnaire over time after being used to that type of questions and of handling its editor.

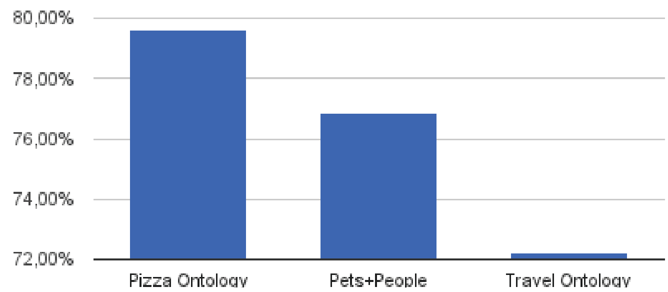

Figure 10.6: Influence of ontologies and instance labels on semantic set operators answers.

Analysis of Answers regarding Semantic Set Operators Type: Semantic Inclusion versus Semantic Intersection

Splitting semantic set operators related questions in questions regarding either semantic inclusion or semantic intersection, one obtain the two graphics at the bottom of figure 10.7. For ease of visual comparison, one can rely on the two horizontal bars on top of the same figure 10.7, which mixes up all users levels. From their analysis, one can say that the results are quite similar: no major significant difference results from their comparison. This leads us to conclude that there is no specific semantic set operator showing a particular reject from human reasoning. One can still remark that light color zones are larger in case of the semantic intersection than in case of semantic inclusion. This observation matches our semantic set operators formal definition. Indeed, it can be interpreted by the fact that semantic inclusion is a set operator where no fuzzy aspect comes into play. On the contrary, semantic intersection takes into account a threshold that can be manually fixed and thus is more subject to variability among users.

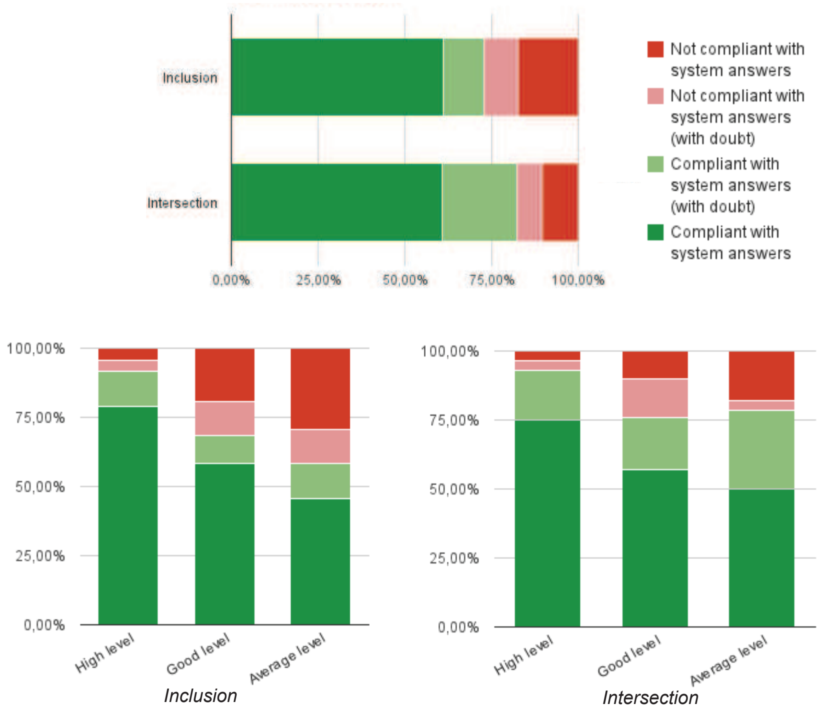

Figure 10.7: Matched answers regarding semantic inclusion versus semantic intersection.

Within semantic intersection related questions, there was a specific sequence of three questions we would like to analyze. These three questions are the last three questions concerning the Pizza ontology. The first question asks about the closeness (similarity) between two individuals "a" and "b" (`#spicyPizza` and `#crispyVegetarianPizza`) comparing to another individual "c" (`#tomato`): Is "a" nearer for "b" than "c"? The two following questions were dealing with the semantic intersection between "a" and "c" and between "a" and "b". Our system answers that "a" is indeed nearer from "b" than "c" and that there is no semantic intersection between "a" and "c" but there is one between "a" and "b". The threshold here is indeed greater than the similarity of "a" and "c" but lower than between "a" and "b". We are here interested in analyzing the implication of the first question (closeness of individuals) on the answers to the following questions (on semantic intersection). First of all, we observe that 100% of our test persons have answered correctly to the first question on closeness. Then, all test person levels considered, 61% of people have answered the same way as our system (illustrated by the red line on figure 10.8). As above, this percentage is influenced by the test person level as shown by the blue decreasing bars on figure 10.8. However, it is interesting to note that 22% of test persons have intuitively considered a higher threshold (i.e.: no intersection for "a" with "b" and neither for "a" and "c") and 11% of test persons have considered a lower threshold (i.e.: intersection for "a" with "b" and for "a" and "c" too). In fact only one marginal test person has produced a total counter result from the system (intersection between "a" and "c" but not between "a" and "b"). This analysis leads us to conclude that semantic intersection is consistent with the notion of closeness of individual and that the computed threshold matches the majority of test person's intuition but this semantic intersection includes still some fuzziness aspects according to the threshold we apply.

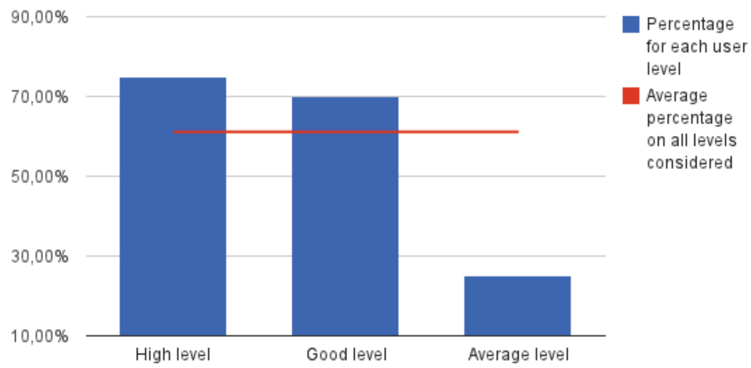

Figure 10.8: Relation between similarity / closeness and semantic intersection.

Analysis of Answers regarding Ontological Reason Types: Taxonomy, Properties versus Inference

We now give a look at the reasons which may have motivated the human reasoning results on semantic set operators. Let's recall, that, in the previous section, we have categorized questions according to their necessary reasons of system answer: taxonomy, property, inference. We present on top of figure 10.9, the test person results (all levels together) regarding the reason types of answers. The taxonomy is the parameter of semantic set operators definition which seems the most intuitive for all human testers. Then, properties are also taken into account in the human reasoning but to a lower scale. Finally, the initial step of inference is often forgotten. These conclusions seem quite understandable once again. Indeed, taxonomy is one of the simpler modeling notions that is straightforward to notice on any ontology editor. On the contrary, inferred axioms either inferred types of instances (classes) or inferred related properties are sometimes more difficult to discover all the most when an automatic reasoner has not been previously launched. However, once again if we look at the three diagrams (see bottom of figure 10.9) taking into account the levels of test persons, it appears that high-level persons have the highest rate of matched answers both in taxonomy, properties and still in inference and that these rates are greater than the one of good-level test persons, themselves greater than the ones of low-level test persons.

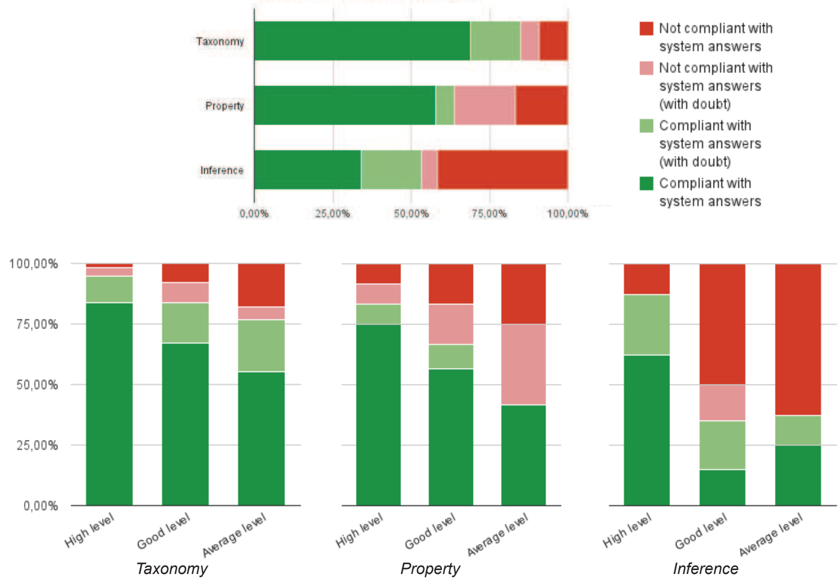

Figure 10.9: Influence of ontological reason types (taxonomy, property, inference) on semantic set operators answers.

Conclusion and Remarks

This experiment has shown that our definition of semantic set operators were quite intuitive for our test persons and that these definitions were compliant with ontology modeling and reasoning philosophy (underlined by test person levels). Of course, descriptive labels of instances have guided the test person answers but we have also shown that test persons have been doing the questionnaire seriously by relying concretely on the ontology and still obtain a significant rate of matched answers. Both semantic inclusion and semantic intersection definitions seems to be adequate with the human reasoning. Moreover, even if the semantic intersection definition includes the numerical computation of a threshold, human reasoning seems to be in range with our proposed threshold computation (Fisher's criterion) and still demonstrates that this is a sort of fuzzy notion since the degree of certainty of answers was a little bit lowered. Each component of semantic set operators (roughly speaking: first perform an inference process, look at concept taxonomy and then at related properties) is done (or wanted to be done!) by the human reasoning but this experiment also shows that the system performs better - since it is harder for a human brain to discover all logical intrications by hand.

However, one can still comment on some choices we made to conduct this experiment. An initial comment would be that this experiment has only focused on semantic set operators about instances of classes, whereas our proposed definitions applied also to instances of properties. Instances of class have been here promoted since semantic set operators are more complex to assess on individuals (it implies looking at more ontological aspects). In addition, the three selected ontologies were not suitable to propose a set of distinct interesting properties due to the limited amount of properties.

Secondly, the scale of our experiment is limited to only 18 test persons and to three support ontologies encompassing 30 questions. Number of test persons may seem at first sight quite low, however, it appears that for each level of test persons the distribution of answers are very similar and that the addition of new test persons does not change the distribution in significant way.

Moreover, we had some informal feedbacks from the test persons that they were sometimes a little bit confused talking about semantic set operators between instances. They explain that they were disturbed by semantic intersection between two instances, whereas for them, instances were representing distinct entities of real world. Even if their remarks were justified, we can remind that ontologies rely on the non unique name assumption (for example, through the `owl:sameAs` axioms) and thus instances are not always disjoint. Moreover in our context, as we have underlined it in previous chapters, the idea is that our set of candidate instances embodies the same real world entity but which might be described differently. Our goal is to find the best suitable semantic description.

10.2 Experimentation conducted on the SeaBILLA project

This section presents first our contributions within the SeaBILLA project, which are based on the *FusionLab* framework augmented with "semantic beliefs fusion" capabilities. Then, it focuses on the use case implementation: where do the original beliefs come from? where do the *DS-Ontology* and the "semantic beliefs fusion" service intervene in this particular *Fusionlab* instance architecture? Through this section, we would like to demonstrate the genericity and adaptability of our approach and implementation. However, as underlined in [Costa 2012], we would like to stress the difficulty to closely evaluate the performance of our added-value in that context. As a matter of fact, it is hard to distinguish the performance of an overall information fusion system from the performance of the uncertainty handling. Indeed, the first one is more encompassing in scope than the latter.

10.2.1 SeaBILLA

Started in June 2010, the project named SeaBILLA[3] (acronym of: Sea Border surveILLAnce system) will end up on February 2014. It is a European FP7 project. This latter is co-financed by the European Union within the Seventh Framework

[3] http://www.seabilla.eu

Programme (FP7) calls for proposals[4].

As already mentioned, the maritime domain presents a growing interest for information fusion applications. There is, indeed, a significant need to combat and prevent illegal activities on sea, such as terrorism, smuggling activities, and illegal immigration. One way of contributing to that combat is to alleviate operator workload in his numerous and tedious tasks. Vessel identification is one of these tasks. As a matter of fact, in real-life scenarios, there are a large number of objects to be identified simultaneously, and consequently a strong need for automated assistance.

One of our specific goals within SeaBILLA is thus to improve the task of vessel identification, which consists in determining the nature of a vessel (e.g. fishing boat, wooden boat, cargo...). Identification techniques for maritime domain awareness cannot be based upon tracking and statistical techniques only, *a priori* and contextual knowledge is of utmost importance and for that semantic technologies are used. Our innovation in that matter is to support the reasoning processes that may be applied with an adequate framework: the *FusionLab* (see section 9.2), and means to tackle with the inherently uncertain nature of the information manipulated.

10.2.2 A Vessel Identification Decision-Support System

We have implemented an identification system that handles uncertain multi-hypotheses due to the necessity to consider uncertain collected pieces of information and non-deterministic identification conclusions.

Relying on the *FusionLab* architecture, this vessel identification decision-support system is in fact a composite service. It is composed of different identification hypotheses *assessment* services that are run in parallel and of an identification *synthesis* service. Figure 10.10 represents this composition.

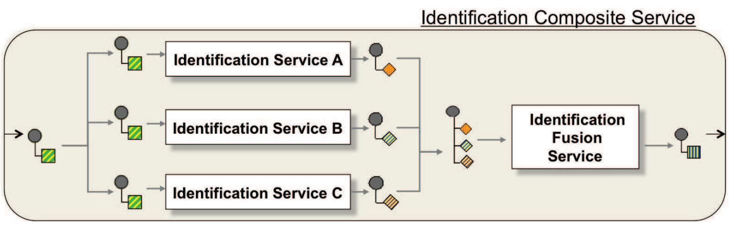

Figure 10.10: Focus on the identification composite service.

This composite service gets in input an instance of a specific vessel and returns back the identification of this vessel which should best hold regarding the available information on this vessel. The information related to this vessel are found in the semantic repository that have been continuously fed by incoming observations and output of previous processing. Finally, we assume here that an association step on raw observations has already been realized (by other services). We thus consider

[4] http://cordis.europa.eu/fp7/home_en.html

being able to retrieve "well-defined tracks", representing specific vessels with all their related static data as well as their trajectories.

10.2.2.1 Identification Hypotheses Assessment Services

Our identification assessment services are based on rule-based expert systems. In these kind of systems [Jackson 1990], a rule takes the form of "If <*premise*> Then <*conclusion*>". Premises contain a collection of conditions that must be satisfied before the rule may be used. The *conclusion* part of a rule contains a set of actions to be performed when the rule is applied. In our services, premises are directly relying on instances of ontologies that are related to the vessel of interest. If some premises of a given rule hold, then conclusions on the possible vessel identification are fired. Conclusions of rules are instances of the `VesselType` class or subclass in the *FusionLabMaritime* ontology (see figure 10.11).

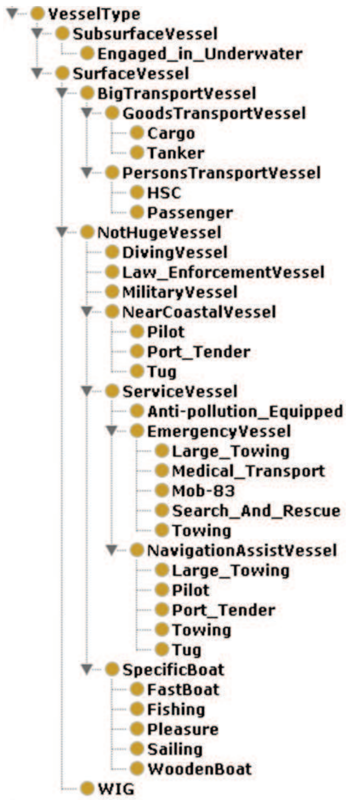

Figure 10.11: Vessel identification taxonomy.

Contrary to automatic inferences on ontologies (see section 4.4), these rules are manually defined outside of the ontology structure. They are complex business rules that cannot be defined within the internal structure of the ontology due to known limitation of the OWL language. As a matter of fact, our business rules need to handle numerical calculi, to manipulate some complex chain of properties and need to assign degrees of beliefs to their conclusions.

Rules premises are taking into account information about:

- the source of the observations. For example, the fact that certain radar has been able to detect the track can give clues on the minimal dimension of the vessel and leads to possible vessel identification. The fact that observations are coming from AIS sensor also means that it is some international voyaging ship with gross tonnage of 300 or more tons, or a passenger ship.
- the content of the observations themselves. Obviously, taking into account information such as the speed of the vessel, its dimension, etc. brings clues on the possible types of vessel.
- the context / environment of the vessel. For example, commercial lanes or fishing zones increase the probability that the observed vessel is commercial or respectively a fishing boat. Moreover, knowing the distance of the vessel from the coast gives information on the autonomy of the vessel and thus also on the type of the vessel.
- intelligence *a priori* data bases.
- etc.

Since, there can be a wide number of rules, the identification expert rules are distributed among different identification hypotheses assessment services. Within a service, rules may be intricated, but there are no relations/interactions between rules belonging to different services. Input information are of different types/natures for each service. For example (as illustrated on figure 10.13 page 153), one service could rely on the intrinsic attributes of the vessel (its dimension, speed, etc.); a service could rely on the contextual information of the vessel (its geographical zone, its shipment cargo, etc.); another service could rely on *a priori* intelligence database; others could rely on the form of the trajectory of the vessel; etc.

All these business rules coming from the expert knowledge in the maritime surveillance area are imprecise and have many exceptions (which we cannot afford to enumerate). Instead of ignoring or enumerating exceptions, our approach has been to summarize them by providing explicit weights on each conclusion to indicates the expert beliefs on it. However, it is very difficult to have human experts providing precise weights to the conclusion of their rules. They rather prefer to assign imprecise linguistic belief degrees: "High", "Medium" or "Low".

We make use of the Jena rule-based reasoner[5] to implement these business rules. It provides a rather simple textual syntax for the rules. For instance, on figure 10.12, the first rule states a quite straightforward inference: "Any vessel that is claiming through its AIS message to be a Fishing vessel leads to a high belief in the conclusion that it is effectively a Fishing vessel". These rules use some built-in functions such as `getAllPositions` or `getTheFarestPositionFromTheCoast`, etc. These built-in functions are specific function pattern that are linked for example here to a Geographic Information System (GIS) database.

[5] `http://jena.apache.org/documentation/inference/index.html#rules`

```
@prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
@prefix fusionLabMar:
            http://www.eads.com/fusionlab/models/fusionlabDomainOntologies/fusionlabMaritime.owl#
[Fishing:
        (?a rdf:type fusionLabMar:Vessel)
        (?a fusionLabMar:hasAISDesignType "Fishing"^^xsd:string)
        ->
        createInterResult(AISRule, ?a, fusionLabMar:Fishing, High)
        print ( "Vessel: " ?a " possible identification " fusionLabMar:Fishing)
]
[FarFromCoast:
        (?a rdf:type fusionLabMar:Vessel)
        getAllPositions(?a, ?positions)
        getTheFarestPositionFromTheCoast(?positions, ?theFarest)
        greaterThan(?theFarest, autonomyLimitForDistinguishingBigTransportVessel)
       ->
       createInterResult(FarFromCoast, ?a, fusionLabMar:BigTransportVessel, Medium)
       print ( "Vessel: " ?a " possible identification " fusionLabMar:BigTransportVessel)
]
```

Figure 10.12: Rule syntax example.

At the end of each identification hypotheses assessment service, a post process of the rules results is performed. Imprecise beliefs are transformed into well defined mass values (as shown on figure 10.13). For that, imprecise beliefs are first mapped to their corresponding values between 0 and 1. Afterwards, a normalisation step is performed: considering a particular vessel identification hypothesis, its mass value is given by the normalized sum of the weights that have been assigned by different rules to that identification hypothesis. Finally, the sum of the mass values assigned by a given service should be equal to 1. At last, the output is embedded in ontological instances by using the *DS-Ontology*.

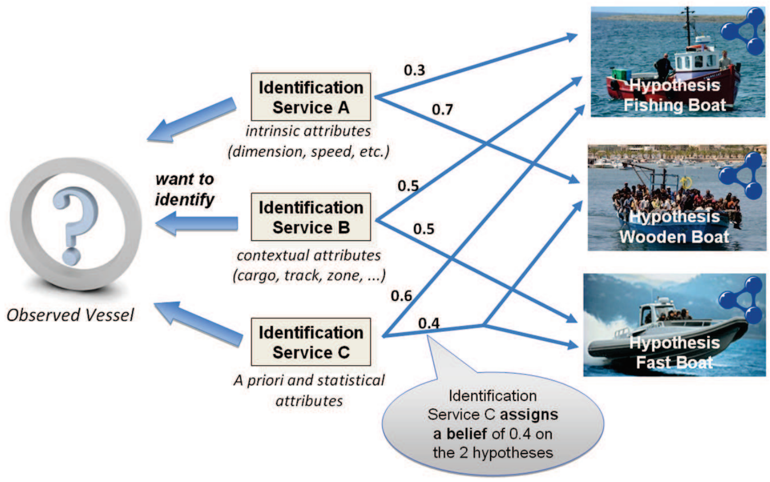

Figure 10.13: Identification hypotheses assessment services outputs.

Each identification hypotheses assessment service can be compared to an independent source of information that gives its own state of beliefs on the possible identification of the vessel.

10.2.2.2 Identification Synthesis Service

The identification synthesis service performs the combination and decision processes of the identification hypotheses coming from the above assessment services. It uses in fact the semantic beliefs fusion service of the *FusionLab* (described in section 9.2.3). As shown on figure 10.14, it takes into account the semantics of the candidate instances as well as their degrees of belief when combining them and decides which candidate instance - describing the vessel identification - should best hold.

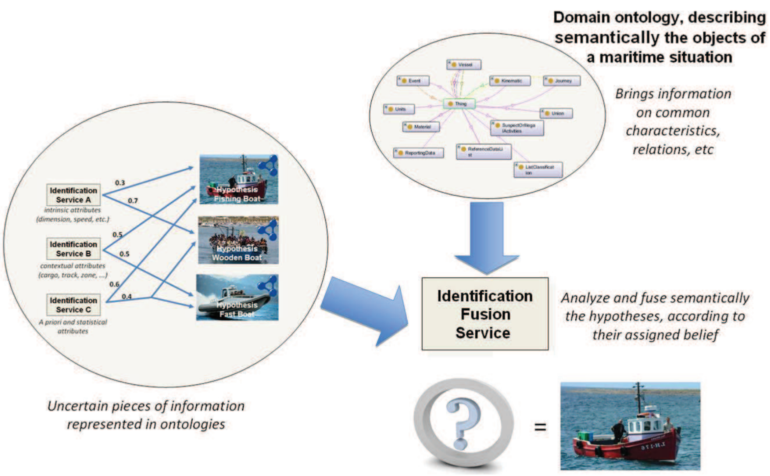

Figure 10.14: Semantic-based multi-hypotheses combination and decision.

The concrete use of our approach and implementation in a European project demonstrates its usefulness and adaptability. The whole system has been run over synthetic data and has proven in the specific vignette scenarios to be computationally tractable and to correctly identify the particular vessels of interest. However, no ground truth was available for the identification of vessels: only the identification of main scenario "actors" could have been compared. Yet, the SeaBILLA project is still an on-going process and future demonstrations should permit to get some feedback of operators.

Nevertheless, the whole approach and this particular maritime use case implementation provide us elements to be able to assess criteria of the ETUR framework, introduced in next section.

10.3 An Evaluation of Techniques for Uncertainty Representation Framework

From a practical standpoint, it is not possible to perform a complete evaluation of our approach for fusing semantic beliefs. On one hand, it is difficult to have a ground truth within SeaBILLA use cases and, then the SeaBILLA implementation is encompassing not only the fusion of semantic beliefs but also the beliefs creation. Instead of carrying out a deep numerical evaluation process, we perform an analysis of the uncertainty level associated to various elements of our model. Therefore, in this section, we discuss criteria proposed by the ETUR evaluation framework for our approach. This section can also be seen as a sum up of added-value or drawbacks brought by our approach and systems.

This section thus begins by presenting the ETUR working group in which we have actively participated, the proposed evaluation framework and then how it can be used to assess our approach and implementations.

10.3.1 The ETUR Working Group

The goal of the ETUR[6] working group is to bring together experts, researchers, and practitioners from the fusion community to leverage the advances and developments in the area of evaluation of uncertainty representation to address the problem of evaluating uncertainty representation and reasoning approaches for high-level information fusion systems. This group has been set up in 2011 and has been created as part of the International Society for Information Fusion (ISIF)[7]. The acronym ETUR stands for Evaluation of Techniques for Uncertainty Representation.

First, within this group, a set of use cases has been proposed. These use cases involve information exchange and fusion requiring sophisticated reasoning and inference under uncertainty. Among the five proposed use cases, one was very closed to our SeaBILLA project. As a matter of fact, a use case is called "Ship Locating and Tracking scenario" [ETURWG 2012] and is based upon illegal immigration detection in a maritime environment.

In parallel, some evaluation criteria were proposed and defined. Basically, the idea of this framework is to take into account issues of representing and reasoning with uncertainty, but also pragmatic topics that germane to knowledge exchange problems such as expressive power, solid mathematical foundation, and performance and scalability.

10.3.2 The URREF Ontology

This set of criteria has been organized in an ontology. This ontology is called the URREF ontology and stands for "Uncertainty Representation and Reasoning Evaluation Framework". As figure 10.15 illustrates it, it groups criteria under four categories. The first one groups together criteria on the input of the system to evaluate. The second and third one deal with uncertainty representation characteristics and respectively uncertainty reasoning characteristics. Finally, the last group of criteria takes into account how the system manages and presents the output.

[6] http://eturwg.c4i.gmu.edu
[7] http://www.isif.org/

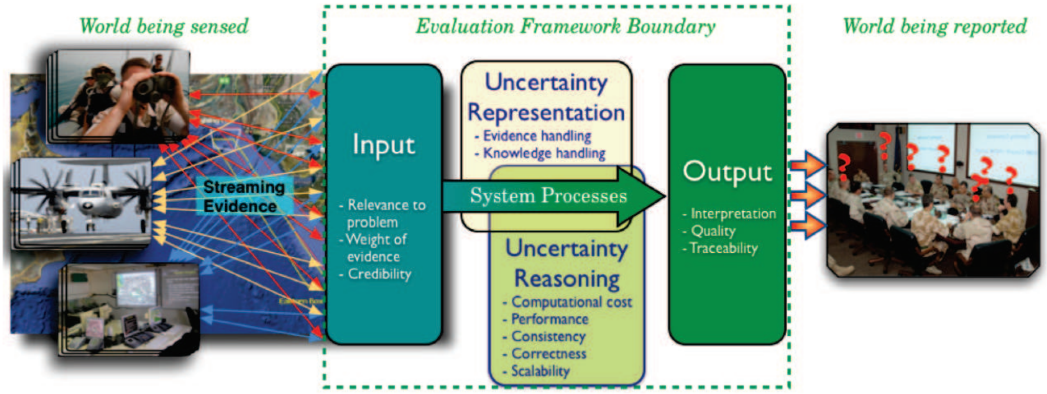

Figure 10.15: Boundaries of the Uncertainty Representation and Reasoning Evaluation Framework (from [Costa 2012]).

Until now, the ontology part dealing with these criteria is limited to a taxonomy. No properties are yet relating these concepts and no instances are populating this ontology. Accordingly, the URREF ontology permits to have a structured set of evaluation criteria with their associated definitions which is useful to share a common understanding between humans, but it is not yet an ontology that can be used by a software agent to concretely takes part of an automatic evaluation process.

The part of the ontology dedicated to the description of the Criteria class and its hierarchy is graphically presented on figure 10.16. Definitions of each criterion (i.e. each node of the ontology) are presented in appendix F. However, concrete numerical measures have not been yet proposed and should be the subject of future topics to get onto by the ETUR working group. The lay out of these criteria is still a work in progress, as the working group has regular teleconference meetings in which each criterion is discussed and its position in the ontology may be revised. At the moment of writing, this version of the ontology is the only stable one which was also used to present preliminary works on the subject at the annual 2012 Fusion conference[8].

This set of criteria has been organized according to a certain logic. This logic follows here the flow of information in the system (input, representation, reasoning, output sets of criteria). It can be however noted that as always in an ontology this is a subjective way of modeling this set of criteria. Another one could have been for example to split criteria according to a business view, a user view and a technical view. The first one could have included some quality, correctness, consistency, etc. criteria. The "user view" branch of the ontology could have grouped for example: traceability, interpretation, simplicity, assessment, etc. criteria. The technical view could have considered scalability, computational cost, or performance criteria.

[8]http://fusion2012.org/public.asp?page=special_session.htm#9

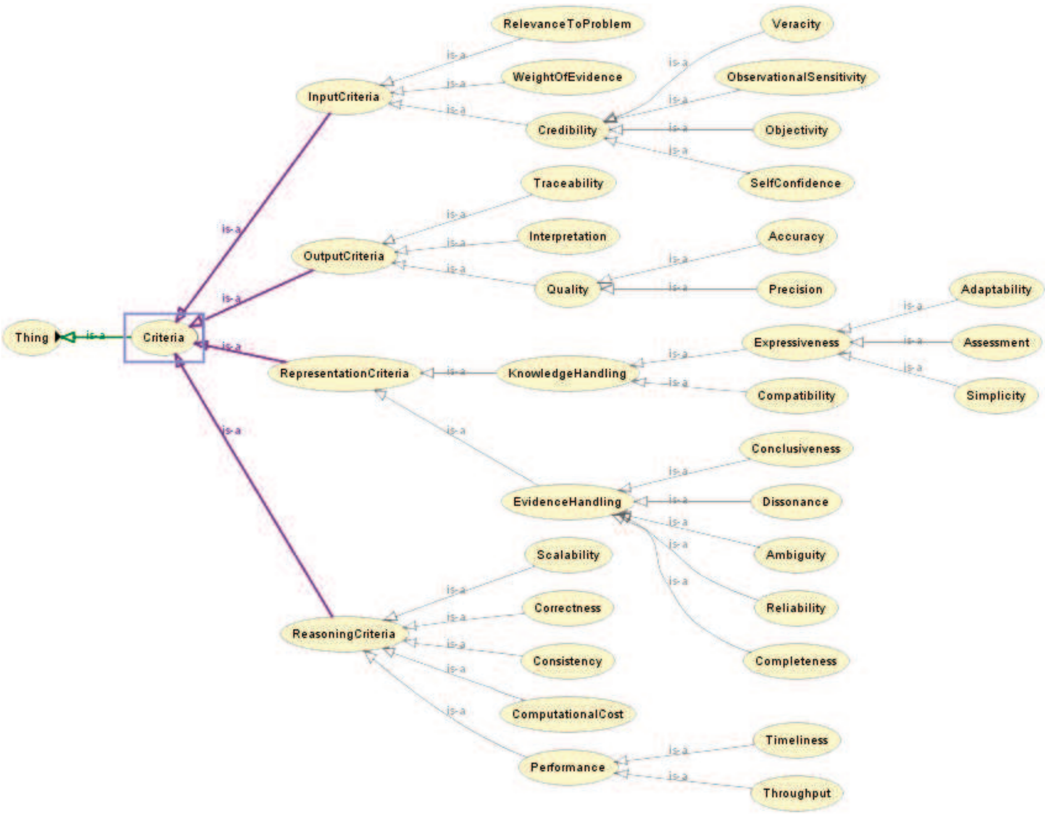

Figure 10.16: The Criteria class of the URREF ontology.

10.3.3 Discussions and Analysis

Relying on these definitions, we present table 10.17 where each criterion (lines of the table) of the URREF ontology is discussed according to its domains of relevance in our work. Criteria can be discussed while considering the uncertainty calculus in general (the theoretical approach) with either characteristics that pertain to the theory of Dempster-Shafer or enhancements that are added thanks to the semantic beliefs framework; criteria could be also discussed considering assets provided by the developed tools (either the Protégé plugin or the *FusionLab* platform); finally, criteria could also take into account the use of theoretical frameworks and developments for concrete use cases (beliefs assignment, the particular fusion algorithm, the decision process, etc.), such as SeaBILLA for instance. For each criterion, a smiley symbol is included in the columns of table 10.17 for which it is relevant (a dash is put on the others columns). A written description and/or justifications are provided in the following subsections, regarding criteria definitions of appendix F.

				Domain of Relevance				
				Theoretical Approach		Implementation		Use Case
				Dempster-Shafer Theory	Semantic Added Value	Protégé Plugin	*FusionLab* Framework	SeaBILLA
Input Criteria	Relevance to Problem			-	-	-	-	☺
	Weight or Force of Evidence			-	-	-	-	☺
	Credibility	Veracity		☺	-	-	-	😐
		Objectivity		☺	-	-	-	😐
		Observational sensitivity		☺	-	-	-	😐
		Self-Confidence		-	-	-	-	☺
Representation Criteria	Evidence Handling	Conclusiveness		☺	-	-	-	-
		Ambiguity		-	☺	-	-	-
		Completeness		☺	☺	-	-	-
		Reliability		-	-	-	-	😐
		Dissonance		☺	-	-	-	-
	Knowledge Handling	Expressiveness	Assessment	-	☺	☺	-	☺
			Adaptability	-	☺	☺	☺	☺
			Simplicity	-	-	☺	-	☺
		Compatibility		☺	☺	-	☺	☺
Reasoning Criteria	Correctness			☺	-	-	-	😐
	Consistency			☺	☺	☺	☺	☺
	Scalability			-	-	-		☺
	Computational Cost			-	☺	-	-	-
	Performance			-	-	-	-	😐
Output Criteria	Quality	Accuracy		☺	-	-	-	-
		Precision		☺	-	-	-	-
	Interpretation			-	☺	☺	-	-
	Traceability			-	-	☺	-	-

Figure 10.17: Overview of URREF criteria according to their domains of relevance.

10.3.3.1 Input Criteria

Relevance to Problem: The uncertainty representation captures well the SeaBILLA input of our identification fusion problem. Data input comes indeed from the inferred conclusions of expert rules which are thus relevant for the identification fusion problem.

Weight or Force of Evidence: Weights assigned to beliefs are called mass values in our theoretical framework. In SeaBILLA, they are assigned by experts through qualitative linguistic variables, which are then transformed automatically into numerical degrees.

Credibility - Veracity, Objectivity and Observational sensitivity: A discounting coefficient can weight the initial set of mass values, according to the veracity, objectivity and observational sensitivity of the source. Concerning the SeaBILLA project, the initial degrees of beliefs associated by the expert rules are taken "as is", since the sources (i.e. the experts are expected to be neutral and accurate).

Credibility - Self-Confidence: Self-confidence refers to the credibility as evaluated by the sensor itself. In SeaBILLA, self-confidence is embodied by the linguistic variables (initial set of mass values) assigned by experts on their rules.

10.3.3.2 Representation Criteria

Evidence Handling - Conclusiveness: The conclusiveness depends firstly on the amount of conflict between sources. As a matter of fact, the lower the conflict is, the more the sources will agree on the final conclusion. The conclusiveness relies as such on the combination rule adopted and also on the decision criterion chosen.

Evidence Handling - Ambiguity: The initial set of data does not encompass semantic ambiguity since hypotheses are described explicitly in ontologies. However, sources can introduce explicit ambiguity by playing on the granularity of those hypotheses or on multiple hypotheses to be able to support different conclusions.

Evidence Handling - Completeness: Ignorance and imprecision are well managed by the Dempster-Shafer Theory. Moreover, management of imprecision is reinforced by the adaptable granularity of semantic hypotheses.

Evidence Handling - Reliability: Reliability is understood here as the overall truthfulness of the input evidence related to a particular scenario. To measure it within the SeaBILLA use case, we should have the ground truth of the data and compare it to the output of identification hypotheses assessment services.

Evidence Handling - Dissonance: The amount of conflict of the Dempster-Shafer theory permits to measure the dissonance of evidence.

Knowledge Handling - Expressiveness - Assessment: Assessment criterion deals with the facility to provide beliefs to the system. We believe that the creation of beliefs is eased by our theoretical framework, since the user does not have to care about the different levels of granularity of the hypotheses on which they assigned a degree of belief. The Protégé plugin gives some guidance for the user on how to make the assessments (create its beliefs). In SeaBILLA, these assessments are eased by qualitative linguistic variables that have to be given by experts to their rules.

Knowledge Handling - Expressiveness - Adaptability: First, the theoretical approach provides adaptability through the import of the *DS-Ontology* in any domain ontology. Then, adaptability is a key goal for the *FusionLab* platform. For example, the semantic beliefs fusion service can be used in any applications, and configuration of combination and decision rules provides adaptability to the problem of interest. Finally, in SeaBILLA, adaptability is provided by the ease to add new expert rules providing new semantic beliefs.

Knowledge Handling - Expressiveness - Simplicity: The Protégé set of plugins permits to graphically edit semantic beliefs and manage the combination and decision process, which may be easier for the user who does not have a deep knowledge about the inner details of the technique. In SeaBILLA, simplicity is provided to the experts through ease of implementing new rules.

Knowledge Handling - Compatibility: The approach is compatible with the OWL W3C standard language to represent knowledge. Concerning the representation of uncertainty itself, our modelling is consistent with but extends the Dempster-Shafer formalism. The representation of the situation itself, which can also be seen as the representation of hypotheses is also of utmost importance of the compatibility criterion. A specific effort has been made for that criterion during the development stage of the *FusionLab* framework. The knowledge representation is indeed based on the JC3IEDM, OGC, and OWL-Time standards.

10.3.3.3 Reasoning Criteria

Correctness: Correctness is an inner property of the Dempster-Shafer Theory as it performs correct reasoning. In SeaBILLA, there is no ground truth, but the results match what is expected by experts (on the vessel of interest). However, these SeaBILLA results do not depend only on the management of uncertainty but also on the hypotheses produced (i.e. the expert rules).

Consistency: The system produces the same results when input with the same data within the same conditions.

Scalability: Our SeaBILLA implementation is scalable on the amount of data provided by the simulation and on the average numbers of expert rules.

Computational Cost: The computational cost is less than DSmT (as shown in section 8.2.2), however computation of semantic intersections and inclusions are additionally required. Logically, the bigger the set of initial candidate instances is, the more it becomes computationally costly.

Performance: The timeliness aspect of this criterion has not been measured (seems suitable for a maritime surveillance context).

10.3.3.4 Output Criteria

Quality - Accuracy and Precision: Accuracy and precision depend on the decision criterion chosen (for example, the plausibility criteria on would lead to a better accuracy, but a lower precision and inversely for credibility decision process). For SeaBILLA, there is no ground truth to evaluate the output.

Interpretation: There is a clear meaning of the output as it is described explicitly in ontologies. Within the Protégé set of plugins, the interpretation is even simpler for the human user as it is part of a dedicated graphical view.

Traceability: The output is easily explained when using the Protégé plugin.

10.3.3.5 Conclusion of Discussions

As a conclusion of table 10.17 and of our discussions, we can recall that evaluating a general uncertainty theory is quite a complex task since the evaluation clearly depends on the modeling of the input information and on the way fusion algorithms will be fine tuned. Therefore, this table has presented different specific criteria that help evaluating not only the general approach, but also the implemented tools and their use within a particular application. For each criterion, we have tried to raise the main assets or specific points of our approach for which the criterion was particularly relevant. Generally, assets of our approach, both in terms of implementation and of theoretical added-value or in the adaptation to the SeaBILLA use case, are mostly found on the representation of uncertainty and more especially in knowledge handling aspects. These criteria are dealing particularly with the interface between the input information and the system. This is actually the major strength of our approach to ease and enable a better information transmission within the system.

Our evaluation has been here limited to qualitative assessments. Thus, these assessments may be sometimes subjective. In future works of the ETUR WG, members will try to propose numerical calculi that could be used as quantitative assessments of the aforementioned criteria and should provide an impartial assessment. These quantitative assessments would be performed on specific running of the system over data. They could be applied for example to a particular maritime scenario. The underlying idea is to be able to compare and share our results with other approaches on same scenarios within the ETUR WG and to a greater extent once the ETUR framework will be in a stable version. Finally, a set of well defined scenarios could be understood as a sort of benchmark on which we will compare our systems.

Currently, the URREF ontology is still in discussion and is subject to reorganisation. For example, many criteria seems to be intricate - inter-dependant (the WG has for instance underlined that "weight of evidence" should rely on the "reliability" and "credibility" criteria). This kind of relations should be explicitly mentioned in the ontology (through properties definitions for instance, or through multi-inheritance), before proposing numerical calculi for them. Definitions should also be adapted to the internal discussions of the WG.

Part V

Conclusion

Chapter 11

Synthesis

Contents

Our approach takes place in the artificial intelligence landscape and more particularly in the uncertain knowledge management field.

We have proposed in this PhD thesis, a mechanism to both represent semantic uncertain pieces of information coming from different sources, and reason over this information in a consistent way by relying on sound and solid mathematical foundations. This work finds particularly its illustration and motivation from information fusion applications in the defense and security domain.

As explained all along this document, we have tackled the following problems of information fusion:

- the modeling of the situation of interest,
- the uncertainty representation, and
- the information synthesis phase, where the combination of uncertain pieces of information and a decision process is performed.

In this chapter, we summarize our contributions within the three above aspects and recall the validation issues. Contributions of this thesis are not only found in information fusion, but also in the Semantic Web, and uncertainty mathematical theories.

In next chapter, we will then broaden out with ideas of prospects and future studies that would complement this work.

11.1 Situation Modeling through Semantic Web Technologies

We have strengthened the Fusion community to use technologies from the Semantic Web to represent and reason over high-level information. We have underlined indeed that more powerful knowledge representation and reasoning tools are needed to manage semantic information. These tools are based on ontologies expressed in

OWL2 language. This formalism is domain independent. Thus, it allows representing information and knowledge from various application domains. This feature is of utmost importance regarding the genericity and modularity of our approach. Within the development of the innovative *FusionLab* platform, aiming to be an integration platform for high-level information fusion processing, we have proposed and provided a consistent semantic layer. This layer consists namely of a set of OWL2 ontologies. Only the terminology is defined, the instances are provided as the *FusionLab* ran over new observations of a particular use case. This layer also consists in providing concrete tools in the Semantic Web area for the *FusionLab* system: mapping techniques, knowledge repository management tools, automatic reasoners and explicit rules inference tools.

11.2 Uncertain Situation Modeling

We have underlined that the concept of information and situation modeling is closely connected to the concept of uncertainty. The information may be, for example, incomplete, imprecise, fragmentary, unreliable, vague, or contradictory.

The expression "semantic beliefs" was introduced to refer to uncertain instances of an ontology reported by different sources with associated numerical degrees of uncertainty.

A meta ontology has been created (the *DS-Ontology*) in order to represent these "semantic beliefs". It is based on the Dempster-Shafer formalism excepting that candidate instances do not have to be exclusive. This freedom is a must while handling dynamic sources that are totally independent from each others and while having a direct transcription of the raw information into ontologically encoded information. Recalling the epigraph of this document "Knowing ignorance is strength. Ignoring knowledge is sickness", our modeling approach permits to deal with that statement honestly. It permits indeed to represent all available information (both certain and uncertain), but enables also to represent the degree of (or full) ignorance of the source by assigning degrees of belief to high-level abstract ontological concepts.

An important aspect of this approach is that it promotes genericity. The *DS-Ontology* can be associated to any domain ontology so as to permit its instantiation in an uncertain manner.

This genericity is illustrated through the developed add-on to the famous Protégé ontology editor. A set of plugins has been indeed implemented to support and help editing semantic beliefs in a graphical way for any domain-ontology. Genericity is also exemplified by the inclusion of the *DS-Ontology* into the set of the *FusionLab* ontologies.

11.3 Uncertain Situation Reasoning

Our reasoning over uncertainty is based on the Dempster-Shafer theory which is known to capture well the issues found in information fusion. It provides thus a solid mathematical foundation to our reasoning.

However, the bridge between our "semantic beliefs" and the evidential Dempster-Shafer formalism was not straightforward for an automatic process. The novelty of this proposed technique is that it takes into account the semantics - the implicit sense of the hypotheses - which adds a semantic dimension to classical uncertainty management applications.

To make this possible, we have introduced the notions of semantic set operators: semantic inclusion and intersection. We have extended the classical Set theory to handle objects that are ontological instances. To our knowledge, there is no such approach in the current state of the art that has made the bridge between these two scientific areas. These operators correspond to the intuitive relation we - as human - can determine by two instances. These operators have been validated by experimentations on human reasoning.

Relying on these semantic set operators, a projection to a proper formalism of the evidential theory has been mathematically defined.

This reasoning process permits also to implicitly address the well-known problem of combining different frames of discernment. This problem refers to heterogeneous sources which provide their hypotheses on different definition spaces (i.e. different frames of discernment) [Schubert 2010, Rombaut 2001]. Provided that the domain-ontology is expressive enough, our creation of the frame of discernment inherently deals with that constraint. In our thread example, this has been illustrated by the hypotheses on the color of the vehicle.

Finally, the developed Protégé plugin helps also to configure and visualize the construction of the formalism into the evidential theory, as well as the combination and decision processes. Last, but not least, we performed a post-analysis on the whole framework we proposed regarding the set of criteria defined in the active ETUR Working Group. This latter has underlined clearly the assets of our framework in terms of knowledge handling.

11.4 Possible Range of Applications

Contributions have also led to the implementation of a Java library. This one has been encapsulated as a modular service within the *FusionLab* platform. As a consequence, it can be easily used for any information fusion projects based on the *FusionLab*. Actually, this has been the case for the European SeaBILLA project.

However, our approach can be used within any types of application that need to fuse both semantic and uncertain information. For example, the URW3-XG [Laskey 2008] provides a large group of situations in which knowledge on the Semantic Web needs to represent uncertainty and reason over uncertainty. This includes the need to fuse pieces of information from different web sites which are ambiguous and controversial. This raises issues of trust and provenance in Web. The idea is here the same as in information fusion applications recalling the "Anyone can say Anything about Any topic" Web assumption. The trust and uncertainty domains, although having different meanings, can be represented in a similar manner. Still relying on the Web, one can find many applications of our semantic set operators. For example, the information retrieval domain could benefit of these operators to increase its recall. These operators can be applied between query concepts and possible answers concepts of the information retrieval tasks. It thus permits to indicate how well the two match (concept overlap through semantic intersection) and give a measure of the output relevance. Semantic intersection operators may also be used in personalization and recommender systems.

Chapter 12

Perspectives

Contents

This last chapter sketches some possible ways on how the results and methods achieved in this work can be generalized and extended by future research. These ideas are organized on the basis of three time buckets: short, mid and long term. It should specify what would be possible to achieve in such a period.

12.1 Short-term perspectives

Using the Fusion of Semantic Beliefs for Hypotheses Testing

All along this document, we have been assuming that a set of semantic beliefs (i.e. candidate instances) was referring to a same phenomenon of interest. There can be several phenomena of interest, but regarding a particular one, our framework considers for granted that all its related hypotheses can be retrieved. The *DS-Ontology* formalism in section 6.2.1 has been created in that sense with the `Uncertain_Concept` class. To ensure this assumption, a preliminary "association" step (also known as hypotheses testing phase) should always be performed before calling our service.

However, in an environment where several different events occur, we may find, as underlined in [Blasch 2012], not only uncertain pieces of evidence but also pieces of evidence that are so weakly specified that it may not be certain to which event they refer. We must then make sure that we do not by mistake combine beliefs that are not referring to the same event. In the case the beliefs would not be referring to a same event, the conflict of the Dempster's rule calculating from the sources beliefs would increase.

Our framework for the fusion of semantic beliefs can be also used for that purpose. In our framework, the conflict of the Dempster's rule is calculated. However, in practice in our applications, we consider our conflict as always negligible (as the hypotheses testing step was assumed). Yet, if the hypotheses testing step is not performed, the conflict of the Dempster's rule when the beliefs are combined can be used so as to give an indication on whether they are referring and describing the same event.

Based on that criteria, a clustering method can be applied, where the conflict can be seen as the distance measure between different clusters. A cluster would refer to a group of belief functions relative to the same event. A high conflict between two belief functions would be an indication that they do not belong to the same cluster. This process is in fact also called the subproblem management and has been handled in [Bengtsson 2001, Schubert 2008]. Our framework, relying on the semantics of the pieces of evidence, would provide an added-value to this clustering method.

Semantic Set Operators through OWL Reasoners Only

A future prospect would be to investigate a slightly different transcription of the hypotheses into the domain ontology. The candidate instances would be restricted to be only linked to a particular type of class or property with no other relations linked to these instances. We can note that this is already the case for candidate instances of properties (which are not linked to any properties due to first-order OWL logic). But, it would change things for instances of classes. In order not to loose any expressivity concerning these instances, new classes should be created. For instance, if we recall our candidate instance `#redCar`, then instead of being an instance of the `Car` class and being linked to the property `hasMainColor #red`, the candidate instance would be directly linked to a `RedCar` class. This class would be defined as: $\texttt{RedCar} \equiv \texttt{Car} \sqcap \exists \texttt{hasMainColor}.\{\texttt{\#red}\}$.

The underlying motivation would be to permit direct determination of the semantic set operators through automatic OWL reasoning (see section 4.4). As a consequence, semantic inclusion could be limited to subsumption of classes. Semantic intersection could rely only on the disjoint OWL axioms that can be defined between classes. This straightforward criterion would avoid computation of semantic similarity between candidate instances. In the previous example, the semantic inclusion between `RedCar` and `LandVehicle` would be straightforward to determine by only asking an OWL reasoner if `RedCar` is a subclass of `LandVehicle`.

Even if this method is quite attractive and should be analyzed more in depth, it has to be carefully applied according to the domain ontologies at hand. As a matter of fact, this particular semantic beliefs representation assumes that the domain ontology is semantically enough sophisticated for that kind of processing. For instance, disjoint axioms are often left out. This could be detrimental to the computation of semantic intersections which would finally lead to an increased size of the frame of discernment.

12.2 Mid-term perspectives

Why limiting the decision process to candidate instances?

Future works should rely on the semantic interpretation of atomic states of the frame of discernment. For the moment, we restrict the decision process to candidate instances that have been proposed by the sources. Decision criteria are indeed applied directly to sets of atomic states corresponding to candidate instances. However, it should be interesting to perform the decision process on the entire frame of discernment and then - from the decided set of elements - create new instances if this set does not correspond to any existing candidate instance.

For example, recalling our main thread example, let's imagine that the decision process would have come up with the atomic state result: $\{d\}$ which corresponds to the intersection of `#redCar` and `#fireTruck` instances (see section 8.2.4). One could then imagine to automatically create a `#redLandVehicle` instance which should entail all the commonalities between the two latter instances. This process could be called inverse mapping in reference to our proposed mapping function that maps a candidate instance to a set of atomic states. However, it still raises some pending issues such as how to map the atomic state: $\{a\}$, where $\{a\}$ corresponds formally to the characteristics of the `#landVehicle` instance except the ones from the `#redCar`, `#fireTruck` and `#aircraft` instances. This inverse mapping seems promising and should be further thought to represent results in new instances reflecting the evidential reasoning as best as possible.

The Human in the Loop

Future works could be brought on a better integration of the human in the information fusion cycle.

This integration includes naturally a better traceability of reasoning over uncertainty within information fusion applications. This could be first done by integrating the Protégé plugin into the graphical visualisation mechanism of the *FusionLab* platform.

One step further would be to get dynamically, into the machine processes, the information (e.g. degree of beliefs, situational hypotheses, utilities, arguments, and preferences) from the user according to his experience and the flow of observations. This could be done again by encompassing the Protégé plugin in the *FusionLab*. However, one of the best options for that would be to directly process hypotheses described in natural language (i.e. plain text). Briefly, this would include tools for speech recognition, for name entity extraction and for automatically populating ontologies (i.e. creating computable "semantic beliefs"). These processes could be carried out by the *WebLab*™platform[1]. This platform, also developed by Cassidian, aims at providing applications that need to process multimedia data (text, image, audio and video). Once these plain text hypotheses would have been encoded in

[1] `http://weblab.ow2.org`

candidate instances, our framework would be directly able to perform the combination and decision processes without having to "normalize" hypotheses. Indeed, observations of the same object or situation might be described on different levels of granularity. The benefit of our semantic beliefs framework is that it enables us to deal with that kind of input in a consistent way to always create a proper frame of discourse.

Thus, this work provides key elements for handling - what is called - soft data fusion. Soft data is data generated by humans which carries a high level of semantics. This is a very recently and promising studied field inside the global information fusion research community.

12.3 Long-term perspectives

Towards a Unified Framework?

One of the issues of future work is to analyze how far can our framework be extended to all types of uncertainty and all mathematical theories. This analysis should deal with both issues concerning a possible unified representation (through an *uncertainty-Ontology* instead of a *DS-Ontology*) and with issues in reusing semantic set operators to create a proper universe of discourse for other mathematical theories.

A total unified framework seems to be somehow utopian thoughts since it is well admitted that there is no formalism to describe all kinds of uncertain information. However, some elementary constructs may be general enough to be shared by any uncertain decision-making approach. According to [Jousselme 2012], the four following notions can be part of these elementary constructs: universe of discourse, instantiated uncertainty representation, reasoning step and decision step. These kinds of elements might be useful to start describing the high level concepts of an *uncertain-Ontology*, which then would be refined according to the specificities of each mathematical theories (see section 3.2).

An ontology such as the *Uncertain-Ontology* would then be interesting to standardize. Indeed, in the same way we have proposed the *DS-Ontology* in our works, other researchers / developers must come up with their own set of constructs for representing uncertainty in ontologies. This is however a major liability in an environment so dependent on interoperability among systems and applications. Still, we recognize that the Semantic Web does not promote standardized overall ontology for every domain, but prefers to consist in a huge collection of different ontologies (see Linked Data community for example `http://linkeddata.org/`). We can continue to argue that in that context, the *Uncertain-Ontology* is not seen as representing domain knowledge but as a way to represent uncertainty in domain knowledge.

Part VI

Appendixes

Appendix A

A Domain Ontology Example - Source Code

```
Prefix(:=<http://www.examples.org/ontologies/DomainOntologyExample.owl#>)
Prefix(owl:=<http://www.w3.org/2002/07/owl#>)
Prefix(rdf:=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(xml:=<http://www.w3.org/XML/1998/namespace>)
Prefix(xsd:=<http://www.w3.org/2001/XMLSchema#>)
Prefix(rdfs:=<http://www.w3.org/2000/01/rdf-schema#>)
Prefix(DomainOntologyExample:=
        <http://www.examples.org/ontologies/DomainOntologyExample.owl#>)


Ontology(<http://www.examples.org/ontologies/DomainOntologyExample.owl>

Declaration(Class(:Aircraft))
SubClassOf(:Aircraft :Vehicle)
Declaration(Class(:Bicycle))
SubClassOf(:Bicycle :LandVehicle)
Declaration(Class(:Book))
Declaration(Class(:Car))
SubClassOf(:Car :LandVehicle)
Declaration(Class(:Color))
Declaration(Class(:Direction))
Declaration(Class(:FireTruck))
SubClassOf(:FireTruck :Truck)
SubClassOf(:FireTruck ObjectHasValue(:hasMainColor :red))
Declaration(Class(:LandVehicle))
SubClassOf(:LandVehicle :Vehicle)
Declaration(Class(:SubsurfaceVessel))
SubClassOf(:SubsurfaceVessel :WaterCraft)
Declaration(Class(:SurfaceVessel))
SubClassOf(:SurfaceVessel :WaterCraft)
Declaration(Class(:Truck))
SubClassOf(:Truck :LandVehicle)
Declaration(Class(:Vehicle))
SubClassOf(:Vehicle owl:Thing)
Declaration(Class(:WaterCraft))
SubClassOf(:WaterCraft :Vehicle)
Declaration(Class(owl:Thing))
Declaration(ObjectProperty(:hasMainColor))
SubObjectPropertyOf(:hasMainColor owl:topObjectProperty)
FunctionalObjectProperty(:hasMainColor)
ObjectPropertyDomain(:hasMainColor :Vehicle)
ObjectPropertyRange(:hasMainColor :Color)
```

```
Declaration(ObjectProperty(:isStoppedNear))
ObjectPropertyDomain(:isStoppedNear :Vehicle)
ObjectPropertyRange(:isStoppedNear :Vehicle)
Declaration(ObjectProperty(:movesFastTowards))
SubObjectPropertyOf(:movesFastTowards :movesTowards)
Declaration(ObjectProperty(:movesSlowlyTowards))
SubObjectPropertyOf(:movesSlowlyTowards :movesTowards)
Declaration(ObjectProperty(:movesTowards))
SubObjectPropertyOf(:movesTowards owl:topObjectProperty)
ObjectPropertyDomain(:movesTowards :Vehicle)
ObjectPropertyRange(:movesTowards ObjectUnionOf(:Vehicle :Direction))
Declaration(ObjectProperty(owl:topObjectProperty))
Declaration(DataProperty(:hasNumberOfDriveWheels))
DataPropertyDomain(:hasNumberOfDriveWheels :LandVehicle)
DataPropertyRange(:hasNumberOfDriveWheels xsd:integer)
Declaration(NamedIndividual(:aircraft))
ClassAssertion(:Aircraft :aircraft)
Declaration(NamedIndividual(:blue))
ClassAssertion(:Color :blue)
Declaration(NamedIndividual(:east))
ClassAssertion(:Direction :east)
Declaration(NamedIndividual(:fireTruck))
ClassAssertion(:FireTruck :fireTruck)
Declaration(NamedIndividual(:green))
ClassAssertion(:Color :green)
Declaration(NamedIndividual(:landVehicle))
ClassAssertion(:LandVehicle :landVehicle)
Declaration(NamedIndividual(:north))
ClassAssertion(:Direction :north)
Declaration(NamedIndividual(:red))
ClassAssertion(:Color :red)
Declaration(NamedIndividual(:redCar))
ClassAssertion(:Car :redCar)
ObjectPropertyAssertion(:hasMainColor :redCar :red)
Declaration(NamedIndividual(:south))
ClassAssertion(:Direction :south)
Declaration(NamedIndividual(:west))
ClassAssertion(:Direction :west)
DisjointClasses(:Aircraft :LandVehicle :WaterCraft)
)
```

Appendix B

DS-Ontology.owl - Source Code

```
Prefix(:=<http://www.owl-ontologies.com/DS-Ontology.owl#>)
Prefix(owl:=<http://www.w3.org/2002/07/owl#>)
Prefix(rdf:=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(xml:=<http://www.w3.org/XML/1998/namespace>)
Prefix(xsd:=<http://www.w3.org/2001/XMLSchema#>)
Prefix(rdfs:=<http://www.w3.org/2000/01/rdf-schema#>)
Prefix(skos:=<http://www.w3.org/2004/02/skos/core#>)


Ontology(<http://www.owl-ontologies.com/DS-Ontology.owl>
Annotation(<http://purl.org/dc/elements/1.1/creator> "Amandine Bellenger")
Annotation(rdfs:comment "The DS-Ontology is a meta-ontology that
once imported in a domain-ontology permits to instantiate
this latter in an uncertain manner.
This ontology is part of the \"Semantic Beliefs\" framework.
It is inspired by the Dempster-Shafer theory of Evidence.")

Declaration(Class(:DS_class))
EquivalentClasses(:DS_class ObjectSomeValuesFrom(:hasDS_hypothesisElement owl:Thing))
SubClassOf(:DS_class :DS_concept)
Declaration(Class(:DS_concept))
AnnotationAssertion(rdfs:comment :DS_concept "A DS_concept instance links a
reporting_source instance with candidate instances and with a degree of belief.
It is subcategorized in DS_class or DS_property when the set of candidate
instances is a set of instances of classes or of properties.")
EquivalentClasses(:DS_concept ObjectUnionOf(DataExactCardinality(1 :DS_belief)
 DataExactCardinality(1 :DS_mass) DataExactCardinality(1 :DS_plausibility)))
EquivalentClasses(:DS_concept ObjectExactCardinality(1 :hasDS_source
 :Reporting_Source))
Declaration(Class(:DS_property))
SubClassOf(:DS_property :DS_concept)
Declaration(Class(:Reporting_Source))
AnnotationAssertion(rdfs:comment :Reporting_Source "It is a person or a
thing that indicates its state of belief.")
Declaration(Class(:Uncertain_class))
EquivalentClasses(:Uncertain_class ObjectSomeValuesFrom(:hasDS_concept :DS_class))
SubClassOf(:Uncertain_class :Uncertain_concept)
DisjointClasses(:Uncertain_class :Uncertain_property)
Declaration(Class(:Uncertain_concept))
AnnotationAssertion(rdfs:comment :Uncertain_concept "For a given phenomenon,
 an Uncertain_concept instance gathers all DS_concept instances
 which are related to that phenomenon.
 It is subcategorized in Uncertain_class or Uncertain_property if the set of
 candidate instances are instances of classes of intances of properties.
 Becareful: For a given Reporting_source instance, an Uncertain_concept
```

```
 instance verifies that the sum of its mass values is equal to 1 (this
 constraint is impossible to represent in OWL 2 DL).")
EquivalentClasses(:Uncertain_concept ObjectUnionOf(:Uncertain_property
 :Uncertain_class))
EquivalentClasses(:Uncertain_concept ObjectSomeValuesFrom(:hasDS_concept
 :DS_concept))
Declaration(Class(:Uncertain_property))
EquivalentClasses(:Uncertain_property ObjectSomeValuesFrom(:hasDS_concept
 :DS_property))
SubClassOf(:Uncertain_property :Uncertain_concept)
DisjointClasses(:Uncertain_property :Uncertain_class)
Declaration(Class(owl:Thing))
Declaration(ObjectProperty(:hasDS_concept))
AnnotationAssertion(rdfs:comment :hasDS_concept "This relation links an
 Uncertain_concept instance with its DS_concept instances.")
SubObjectPropertyOf(:hasDS_concept owl:topObjectProperty)
ObjectPropertyDomain(:hasDS_concept :Uncertain_concept)
ObjectPropertyRange(:hasDS_concept :DS_concept)
Declaration(ObjectProperty(:hasDS_hypothesisElement))
AnnotationAssertion(rdfs:comment :hasDS_hypothesisElement "A DS_class instance
 is related to its candidate instances (hypotheses) through this property.
A candidate instance is an instance of a domain-ontology that is a proposed
 explanation for a phenomenon.")
ObjectPropertyDomain(:hasDS_hypothesisElement :DS_concept)
Declaration(ObjectProperty(:hasDS_source))
AnnotationAssertion(rdfs:comment :hasDS_source "A DS_concept instance
 is related to its Reporting_source instance through this property.")
SubObjectPropertyOf(:hasDS_source owl:topObjectProperty)
FunctionalObjectProperty(:hasDS_source)
ObjectPropertyDomain(:hasDS_source :DS_concept)
ObjectPropertyRange(:hasDS_source :Reporting_Source)
Declaration(ObjectProperty(:hasUncertain_property))
AnnotationAssertion(rdfs:comment :hasUncertain_property "The fixed subject
 of candidate instances of properties are related to an Uncertain_property
 instance through this property.")
SubObjectPropertyOf(:hasUncertain_property owl:topObjectProperty)
ObjectPropertyRange(:hasUncertain_property :Uncertain_property)
Declaration(DataProperty(:DS_belief))
SubDataPropertyOf(:DS_belief :DS_numericRelation)
FunctionalDataProperty(:DS_belief)
DataPropertyDomain(:DS_belief :DS_concept)
Declaration(DataProperty(:DS_mass))
SubDataPropertyOf(:DS_mass :DS_numericRelation)
FunctionalDataProperty(:DS_mass)
DataPropertyDomain(:DS_mass :DS_concept)
Declaration(DataProperty(:DS_numericRelation))
AnnotationAssertion(rdfs:comment :DS_numericRelation "A degree of belief
 is associated to a DS_concept instance thanks to this property.")
DataPropertyDomain(:DS_numericRelation :DS_concept)
DataPropertyRange(:DS_numericRelation :specificUncertaintyDatatype)
Declaration(DataProperty(:DS_plausibility))
SubDataPropertyOf(:DS_plausibility :DS_numericRelation)
FunctionalDataProperty(:DS_plausibility)
DataPropertyDomain(:DS_plausibility :DS_concept)
```

```
Declaration(DataProperty(owl:topDataProperty))
Declaration(AnnotationProperty(<http://purl.org/dc/elements/1.1/creator>))
Declaration(AnnotationProperty(<http://purl.org/dc/elements/1.1/rights>))
Declaration(AnnotationProperty(rdfs:comment))
Declaration(Datatype(:specificUncertaintyDatatype))
DatatypeDefinition(:specificUncertaintyDatatype DatatypeRestriction(xsd:double
 xsd:maxInclusive "1.0"^^xsd:double xsd:minInclusive "0.0"^^xsd:double))
)
```

Appendix C

Uncertain Domain Ontology Example - Source Code

```
Prefix(:=<http://www.examples.org/ontologies/DomainOntologyExample.owl#>)
Prefix(owl:=<http://www.w3.org/2002/07/owl#>)
Prefix(rdf:=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(xml:=<http://www.w3.org/XML/1998/namespace>)
Prefix(xsd:=<http://www.w3.org/2001/XMLSchema#>)
Prefix(rdfs:=<http://www.w3.org/2000/01/rdf-schema#>)
Prefix(DS-Ontology:=<http://www.owl-ontologies.com/DS-Ontology.owl#>)
Prefix(DomainOntologyExample:=<http://www.examples.org/ontologies/DomainOntologyExample.owl#>)


Ontology(<http://www.examples.org/ontologies/DomainOntologyExample.owl>
Import(<http://www.owl-ontologies.com/DS-Ontology.owl>)

Declaration(Class(:Aircraft))
SubClassOf(:Aircraft :Vehicle)
Declaration(Class(:Bicycle))
SubClassOf(:Bicycle :LandVehicle)
Declaration(Class(:Book))
Declaration(Class(:Car))
SubClassOf(:Car :LandVehicle)
SubClassOf(:Car DataHasValue(:hasNumberOfDriveWheels "4"^^xsd:integer))
Declaration(Class(:Color))
Declaration(Class(:Direction))
Declaration(Class(:FireTruck))
SubClassOf(:FireTruck :Truck)
SubClassOf(:FireTruck ObjectHasValue(:hasMainColor :red))
Declaration(Class(:LandVehicle))
SubClassOf(:LandVehicle :Vehicle)
Declaration(Class(:SubsurfaceVessel))
SubClassOf(:SubsurfaceVessel :WaterCraft)
Declaration(Class(:SurfaceVessel))
SubClassOf(:SurfaceVessel :WaterCraft)
Declaration(Class(:Truck))
SubClassOf(:Truck :LandVehicle)
Declaration(Class(:Vehicle))
SubClassOf(:Vehicle owl:Thing)
Declaration(Class(:WaterCraft))
SubClassOf(:WaterCraft :Vehicle)
Declaration(Class(owl:Thing))
Declaration(ObjectProperty(:hasMainColor))
SubObjectPropertyOf(:hasMainColor owl:topObjectProperty)
FunctionalObjectProperty(:hasMainColor)
```

```
ObjectPropertyDomain(:hasMainColor :Vehicle)
ObjectPropertyRange(:hasMainColor :Color)
Declaration(ObjectProperty(:isStoppedNear))
ObjectPropertyDomain(:isStoppedNear :Vehicle)
ObjectPropertyRange(:isStoppedNear :Vehicle)
Declaration(ObjectProperty(:movesFastTowards))
SubObjectPropertyOf(:movesFastTowards :movesTowards)
Declaration(ObjectProperty(:movesSlowlyTowards))
SubObjectPropertyOf(:movesSlowlyTowards :movesTowards)
Declaration(ObjectProperty(:movesTowards))
SubObjectPropertyOf(:movesTowards owl:topObjectProperty)
ObjectPropertyDomain(:movesTowards :Vehicle)
ObjectPropertyRange(:movesTowards ObjectUnionOf(:Vehicle :Direction))
Declaration(ObjectProperty(owl:topObjectProperty))
Declaration(DataProperty(:hasNumberOfDriveWheels))
DataPropertyDomain(:hasNumberOfDriveWheels :LandVehicle)
DataPropertyRange(:hasNumberOfDriveWheels xsd:integer)
Declaration(NamedIndividual(:aircraft))
ClassAssertion(:Aircraft :aircraft)
Declaration(NamedIndividual(:blue))
ClassAssertion(:Color :blue)
Declaration(NamedIndividual(:ds_class_1))
ClassAssertion(DS-Ontology:DS_class :ds_class_1)
ObjectPropertyAssertion(DS-Ontology:hasDS_hypothesisElement :ds_class_1 :redCar)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_class_1 :human)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_class_1 "0.2"^^xsd:double)
Declaration(NamedIndividual(:ds_class_2))
ClassAssertion(DS-Ontology:DS_class :ds_class_2)
ObjectPropertyAssertion(DS-Ontology:hasDS_hypothesisElement :ds_class_2 :fireTruck)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_class_2 :human)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_class_2 "0.4"^^xsd:double)
Declaration(NamedIndividual(:ds_class_3))
ClassAssertion(DS-Ontology:DS_class :ds_class_3)
ObjectPropertyAssertion(DS-Ontology:hasDS_hypothesisElement :ds_class_3 :landVehicle)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_class_3 :human)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_class_3 "0.4"^^xsd:double)
Declaration(NamedIndividual(:ds_class_4))
ClassAssertion(DS-Ontology:DS_class :ds_class_4)
ObjectPropertyAssertion(DS-Ontology:hasDS_hypothesisElement :ds_class_4 :landVehicle)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_class_4 :radar)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_class_4 "0.6"^^xsd:double)
Declaration(NamedIndividual(:ds_class_5))
ClassAssertion(DS-Ontology:DS_class :ds_class_5)
ObjectPropertyAssertion(DS-Ontology:hasDS_hypothesisElement :ds_class_5 :aircraft)
ObjectPropertyAssertion(DS-Ontology:hasDS_hypothesisElement :ds_class_5 :landVehicle)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_class_5 :radar)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_class_5 "0.3"^^xsd:double)
Declaration(NamedIndividual(:ds_class_6))
ClassAssertion(DS-Ontology:DS_class :ds_class_6)
ObjectPropertyAssertion(DS-Ontology:hasDS_hypothesisElement :ds_class_6 :aircraft)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_class_6 :radar)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_class_6 "0.1"^^xsd:double)
Declaration(NamedIndividual(:ds_property_1))
ClassAssertion(DS-Ontology:DS_property :ds_property_1)
```

```
ObjectPropertyAssertion(:movesTowards :ds_property_1 :suspectCar)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_property_1 :radar)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_property_1 "0.4"^^xsd:double)
Declaration(NamedIndividual(:ds_property_2))
ClassAssertion(DS-Ontology:DS_property :ds_property_2)
ObjectPropertyAssertion(:isStoppedNear :ds_property_2 :suspectCar)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_property_2 :radar)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_property_2 "0.3"^^xsd:double)
Declaration(NamedIndividual(:ds_property_3))
ClassAssertion(DS-Ontology:DS_property :ds_property_3)
ObjectPropertyAssertion(:movesFastTowards :ds_property_3 :suspectCar)
ObjectPropertyAssertion(DS-Ontology:hasDS_source :ds_property_3 :radar)
DataPropertyAssertion(DS-Ontology:DS_mass :ds_property_3 "0.3"^^xsd:double)
Declaration(NamedIndividual(:east))
ClassAssertion(:Direction :east)
Declaration(NamedIndividual(:fireTruck))
ClassAssertion(:FireTruck :fireTruck)
Declaration(NamedIndividual(:green))
ClassAssertion(:Color :green)
Declaration(NamedIndividual(:human))
ClassAssertion(DS-Ontology:Reporting_Source :human)
Declaration(NamedIndividual(:landVehicle))
ClassAssertion(:LandVehicle :landVehicle)
Declaration(NamedIndividual(:north))
ClassAssertion(:Direction :north)
Declaration(NamedIndividual(:policeCar))
ClassAssertion(:Car :policeCar)
ObjectPropertyAssertion(DS-Ontology:hasUncertain_property :policeCar :uncertain_property)
Declaration(NamedIndividual(:radar))
ClassAssertion(DS-Ontology:Reporting_Source :radar)
Declaration(NamedIndividual(:red))
ClassAssertion(:Color :red)
Declaration(NamedIndividual(:redCar))
ClassAssertion(:Car :redCar)
ObjectPropertyAssertion(:hasMainColor :redCar :red)
Declaration(NamedIndividual(:south))
ClassAssertion(:Direction :south)
Declaration(NamedIndividual(:suspectCar))
ClassAssertion(:Car :suspectCar)
Declaration(NamedIndividual(:uncertain_class))
ClassAssertion(DS-Ontology:Uncertain_class :uncertain_class)
ObjectPropertyAssertion(:movesTowards :uncertain_class :south)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_class :ds_class_6)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_class :ds_class_2)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_class :ds_class_3)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_class :ds_class_4)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_class :ds_class_5)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_class :ds_class_1)
Declaration(NamedIndividual(:uncertain_property))
ClassAssertion(DS-Ontology:Uncertain_property :uncertain_property)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_property :ds_property_2)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_property :ds_property_3)
ObjectPropertyAssertion(DS-Ontology:hasDS_concept :uncertain_property :ds_property_1)
Declaration(NamedIndividual(:west))
```

```
ClassAssertion(:Direction :west)
DisjointClasses(:Aircraft :LandVehicle :WaterCraft)
)
```

Appendix D

Semantic Set Operators Experimentation - The Questionnaire (in French)

Merci de participer à ce questionnaire ! Cela devrait vous prendre environ 30 minutes. Les résultats sont anonymes et seront utiles pour valider une partie des travaux de ma thèse !

Requis

- Avoir des connaissances basiques sur les ontologies.
- Avoir un éditeur d'ontologies installé sur votre ordinateur (par ex, l'outil Protégé téléchargeable sur `http://protege.stanford.edu/download/download.html`).

Déroulement du Questionnaire

Ce questionnaire comporte trois parties, chacune liée à une ontologie différente. Pour chacune, il y a quatres étapes à réaliser:
- lire la description succincte de l'ontologie,
- copier/coller le lien dans votre navigateur pour télécharger l'ontologie,
- ouvrir et parcourir rapidement l'ontologie dans votre éditeur préféré, et
- évaluer les propositions du questionnaire en vous référant à l'ontologie.

Il est préférable de remplir ce questionnaire dans l'ordre car certaines questions vous sensibiliseront pour d'autres. Ces propositions concernent entre autres les relations sémantiques entre instances d'une ontologie (relations d'inclusions et d'intersections sémantiques). Par exemple, en considérant une ontologie adéquate, le concept de "véhicule terrestre" inclut sémantiquement le concept de "voiture". A l'inverse le concept de "voiture rouge" n'inclut pas "voiture". Enfin, le concept "voiture" n'a pas d'intersection sémantique avec "avion" mais pourrait en avoir une avec le concept "monospace".

Ces propositions sont à évaluer sur une échelle de 1 à 4:
- 1 : la proposition est fausse
- 2 : la proposition me paraît plutôt fausse
- 3 : la proposition me paraît plutôt vrai

- 4 : la proposition est vrai

L'ontologie Pizza

Cette ontologie décrit le domaine de la pizza. Cela passe par la base de la pizza (croûte, épaisseur, etc.), jusqu'à la garniture de la pizza (mozzarella, poivron, tomate, etc.). En s'appuyant sur ces notions, l'ontologie définie des catégories de pizzas (végétarienne, au fromage, etc.) et des noms de pizzas (la Margherita, la Reine, etc.).

Vous pouvez maintenant télécharger l'ontologie sur https://docs.google.com/open?id=0B-mUmmpLfBihZThXNjBvcUJSams (NB: adapté du tutoriel http://www.co-ode.org/ontologies/pizza/2007/02/12/).

Proposition	**1**	**2**	**3**	**4**
1- La classe "SpicyPizza" est une sous-classe de "NamedPizza"				
3- "AmericanHot" est une "NamedPizza"				
3- "pizza" est une instance de la classe "Pizza"				
4- "anotherTopping" est une instance de "PepperTopping"				
5- "anotherTopping" est une instance de "SauceTopping"				
6- "spicyPizza" inclut sémantiquement "pizza"				
7- "crispyVegetarianPizza" est inclus sémantiquement dans "spicyPizza"				
8- "spicyPizza" inclut sémantiquement "americanHotPizza"				
9- "spicyPizza" est plus proche de "crispyVegetarianPizza" que de "tomato"				
10- "spicyPizza" a une intersection sémantique avec "tomato"				
11- "spicyPizza" a une intersection sémantique avec "crispyVegetarianPizza"				

L'ontologie People+Pets

L'ontologie décrit les différentes catégories de personnes (homme, femme, parent, enfant, etc.) et des liens logiques définissant ces catégories entre elles (a pour parents, dame âgée, etc.). Les animaux sont également représentés (chien, chat, girafe, etc.) ainsi que leur relations avec les humains (est l'animal de, etc).

Vous pouvez maintenant télécharger l'ontologie sur https://docs.google.com/open?id=0B-mUmmpLfBihcWoyMkt0WVBtdUU (NB: adapté de http://protege.cim3.net/file/pub/ontologies/people.pets/people+pets.owl).

Proposition	1	2	3	4
1- "Buffy" est de type "old_lady"				
2- "Tibbs" est l'animal de "Fred"				
3- "Buffy" inclut sémantiquement "Minnie"				
4- "Pepet" inclut sémantiquement "Tibbs" and "Tom"				
5- "Minnie" inclut sémantiquement "Buffy"				
6- "Minnie" a une intersection sémantique avec "Tibbs"				
7- "Buffy" a une intersection sémantique avec "Minnie"				
9- "Buffy" a une intersection sémantique avec "Rex"				

L'ontologie du Voyage

Cette ontologie est dédiée au domaine du tourisme. Les concepts hébergements (Hôtel, Bed and Breakfast, etc.), d'activités (aventure, sport, relaxation, etc.) et de types de destinations (plages, campagne, ville, etc.) y sont décrits.

Vous pouvez maintenant télécharger l'ontologie sur https://docs.google.com/open?id=0B-mUmmpLfBihNjQtVm52WEpGOVk (NB: adapté de http://protege.cim3.net/file/pub/ontologies/travel/travel.owl).

Proposition	1	2	3	4
1- "BackpackersDestination" est une sous-classe de "BudgetHotelDestination"				
2- "InstanceB" est un "BackpackersDestination"				
3- "InstanceF" est de type "RuralArea" et de type "QuietDestination"				
4- "InstanceA" inclut sémantiquement "InstanceB"				
5- "InstanceC" inclut sémantiquement "InstanceB"				
6- "InstanceB" est inclus sémantiquement dans "InstanceA"				
7- "InstanceA" a une intersection sémantique avec "InstanceB"				
8- "InstanceC" a une intersection sémantique avec "InstanceB"				
9- "InstanceD" inclut sémantiquement "InstanceE"				
10- "InstanceE" inclut sémantiquement "InstanceD"				
11- "InstanceD" est inclus sémantiquement dans "InstanceF"				

APPENDIX E

Protégé Snapshots of the Ontologies related to Experimentations

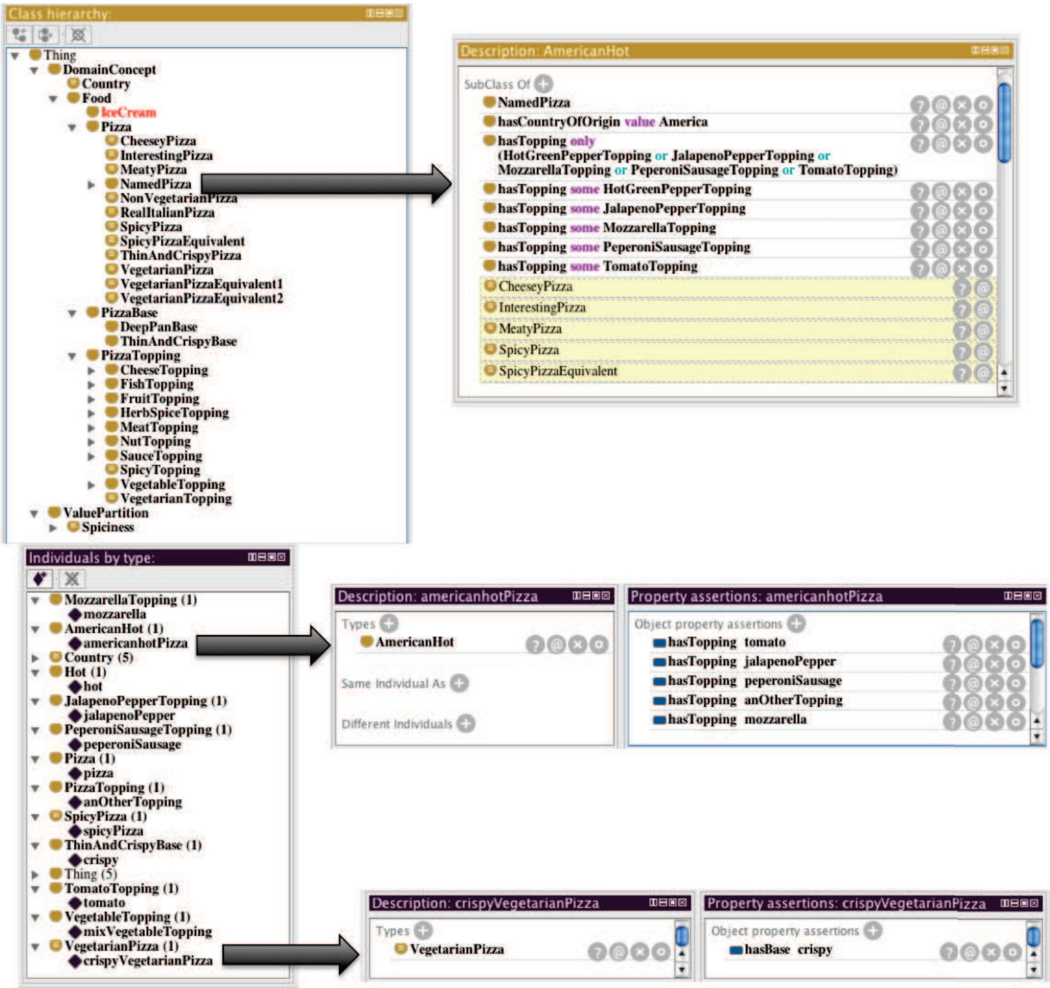

Figure E.1: Protégé Snapshots of the adapted Pizza Ontology.

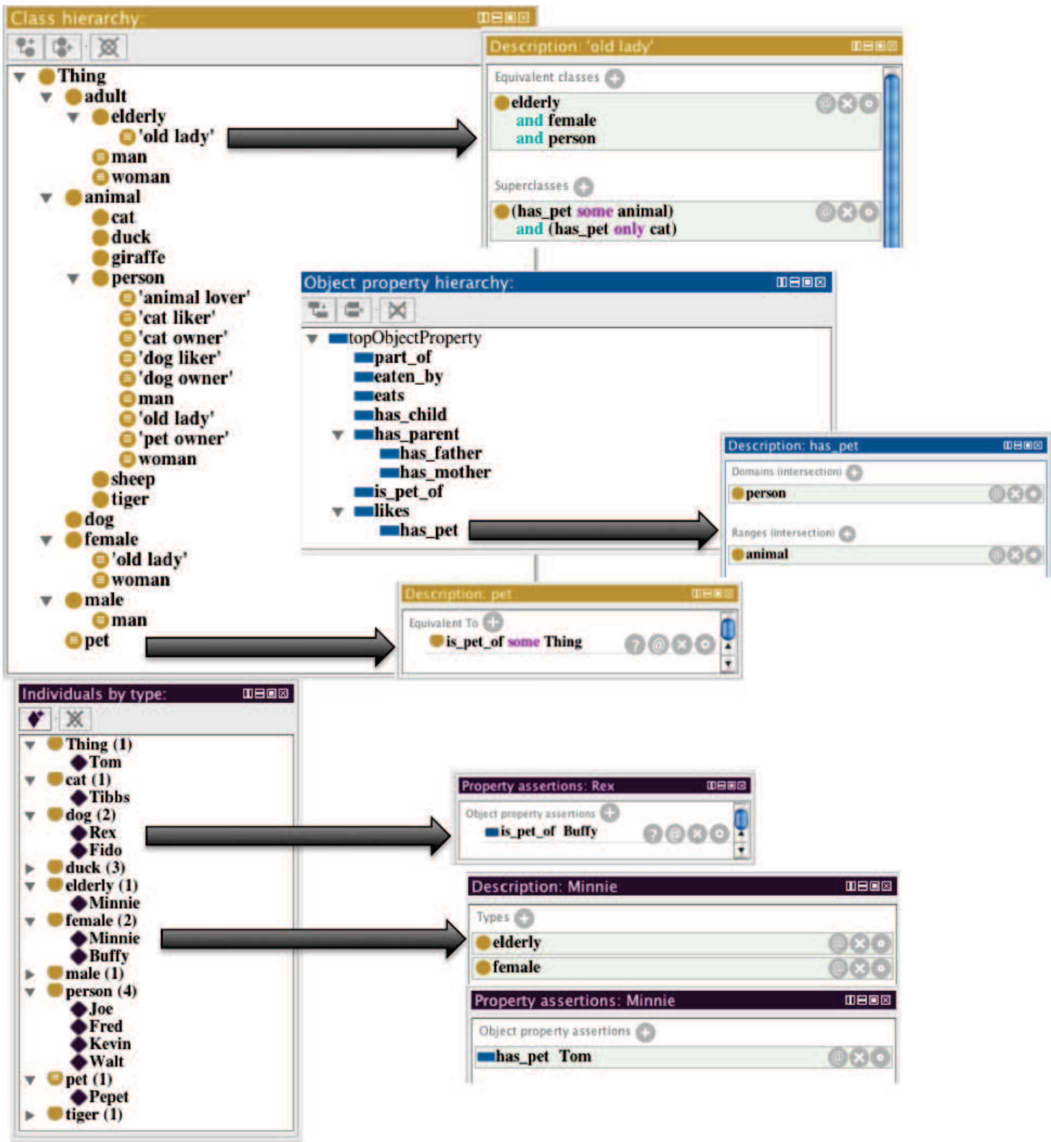

Figure E.2: Protégé Snapshots of the adapted People and Pets Ontology.

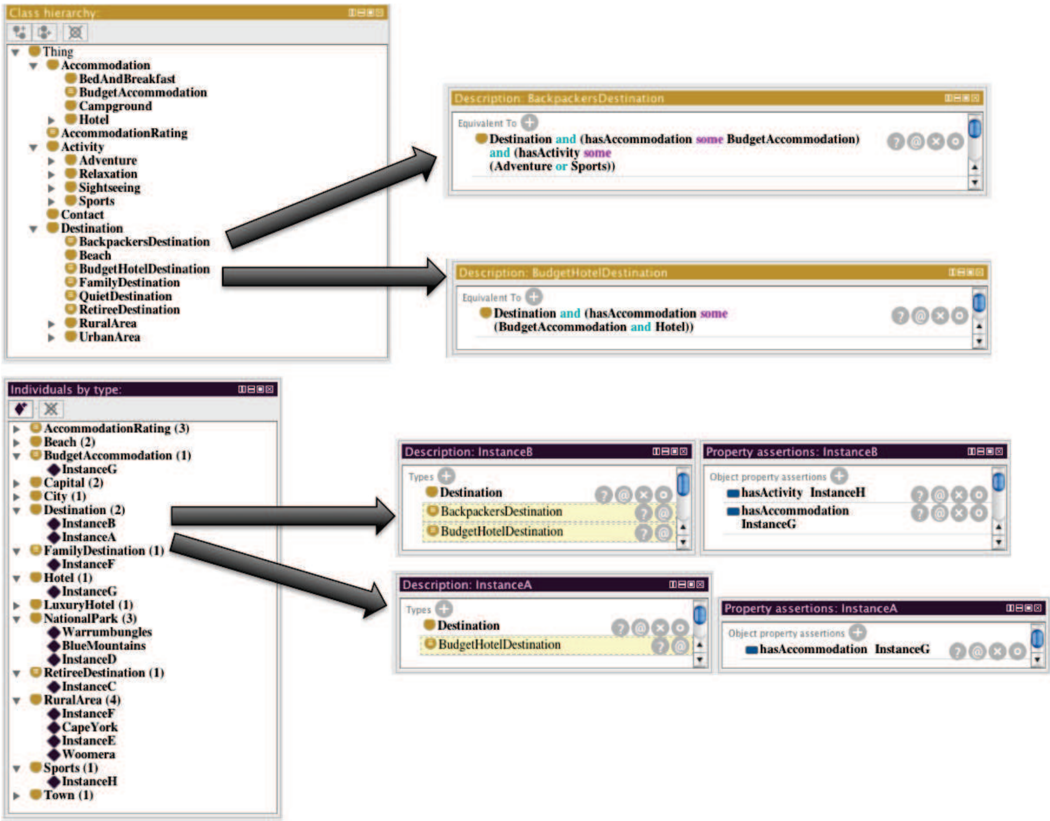

Figure E.3: Protégé Snapshots of the adapted Travel Ontology.

Appendix F

The URREF Ontology Class Definitions

Each criterion listed in the ETURWG's URREF ontology[1] is listed below, and numbered in accordance to its position in the schema.

This ontology is meant to emcompass all the different aspects that must be considered when evaluating uncertainty handling in multi-sensor fusion systems.

F.1 Input Criteria

This general concept encompasses the criteria that directly affect the way evidence is input to the system. It mostly focuses on the source of input data or evidence, which can be tangible (sensing or physical), testimonial (human), documentary, or known missing (Schum, 1994).

F.1.1 Relevance to Problem

Relevance to Problem assess how a given uncertainty representation is able to capture how a given input is relevant to the problem that was the source of the data request. This is a criterion specific to high-level fusion systems that work at levels 3 and above of the JDL model.

F.1.2 Weight or Force of Evidence

Weight or Force of Evidence assess how a given uncertainty representation is able to capture by how much a given input can affect the processing and output of the fusion system. Ideally, this should be an objective assessment and the representation approach must provide a means to measure the degree of impact of an evidence item with a numerical scale. This criterion is especially useful for determining the value of information in systems that must trade-off their ability to capture more evidence with active sensors with the need to avoid being observed. That is, this criterion is especially important to systems that rely on value of information.

F.1.3 Credibility

Also known as believability, it mainly comprises the aspects that directly affect a sensor (soft or hard) in its ability to capture evidence.

[1] Definitions taken from http://eturwg.c4i.gmu.edu/?q=URREF_Ontology (retrieved November 15, 2012)

F.1.3.1 Veracity

This is a measure of the sensor's ability to provide a "truthful" report. That is, a measure of whether the sensor reports what it believes is true. The concept originated with human testimony (deliberate intention to deceive), but can be applied to sensor errors (such as sensor faults) that cause the sensor data to deviate from what would have been reported had the error not existed. Example: - A justice of the peace states the person was aged 35 – higher Veracity - A known criminal and liar states the person was aged 35 – lower Veracity Rationale: the sensor "justice of the peace" is more likely to produce a veritable information than the sensor "known criminal and liar"

F.1.3.2 Objectivity

This is a measure of bias, which applies to all types of sensors.

F.1.3.3 Observational sensitivity

This is a measure of whether the sensor can sense what it claims to have sensed, also precision of measurement. Example 1: - The sober man said the car was yellow – higher Observational Sensitivity - The drunk man said the car was yellow – lower Observational Sensitivity Example 2: - The ANPR camera stated the number plate was YR59 WXT – higher Observational Sensitivity - The witness said the number plate was YR59 WXT – lower Observational Sensitivity

F.1.3.4 Self-Confidence

This is a measure of the information credibility as evaluated by the sensor itself. This is particularly relevant for soft sensors (HUMINT data) as often such sources provide appreciations of the information conveyed (such as it is possible, it is true, etc). The idea behind this measure is that than HUMINT data can potentially convey two "types" of information: the information itself (tomorrow the sun will shine) but also some qualification of this information (it is *possible* because it is summer time and most of the time we have shinny weather). The purpose of this measure is to take advantage of this particularity, and to have a first evaluation of the uncertainty as expressed by the author himself. It is obvious that we are dealing with the credibility of the information: the author is providing us some information, but it is also telling us how much he believes it himself. We are not dealing (at least not directly) with information veracity: even if the author considers the information as possible, and he trusts it, it could be false at the end (even in summer time we can have a cloudy day). The self-confidence is telling us how much the author trust the information, but not necessarily that this information is false or true.

F.2 Representation Criteria

This general concept encompasses the criteria that directly affect the way information is captured by and transmitted through the system. It can also be called as interfacing or transport criteria, as it deals with how the representational model deals with transmitting information within the system.

F.2.1 Evidence Handling

These criteria apply particularly to the ability of a given representation of uncertainty to capture specific characteristics of incomplete evidence that are available to or produced by the system. The main focus is on measuring the quality of the evidence by assessing how well this evidence is able to support the development of a conclusion.

F.2.1.1 Conclusiveness

This is a measure of how well the available evidence will support a definitive conclusion (strongly select a hypothesis).

F.2.1.2 Ambiguity

This is a measure of the extent to which the set of data can be interpreted to support different conclusions.

F.2.1.3 Completeness

This is a measure of the range of the available evidence, in terms of how much is available and how much is unknown. It assesses the variety and eliminative characteristics of the data.

F.2.1.4 Reliability

This is a measure of the overall truthfulness (accordance with reality) of the evidence.

F.2.1.5 Dissonance

This is a measure of the extent to which the evidence is explicitly contradictory or inconsistent. That is, if a conflict exists in the information supporting a hypothesis (e.g. two descriptions of the same event or entity are not compatible).

F.2.2 Knowledge Handling

These criteria is intended to measure the ability of a given uncertainty representation technique to convey knowledge.

F.2.2.1 Expressiveness

This is a measure of the representational power of a given technique.

Assessment It should be practicable for a user of the system to make (and feel comfortable with) all the uncertainty assessments that are needed as input. The system should give some guidance on how to make the assessments. It should be able to handle judgments of various types, including expressions of uncertainty in natural language such as "if A then probably B", and to combine qualitative judgments with quantitative assessments of uncertainty (adapted from Welley 1996). Typical assessment questions include: a) which input can complete a result (help for planning of recce assets)? b) What is the state of the data? Is there enough data? c) What is the reliability of the most probable results? And others.

Adaptability Adaptability criteria encompass the ability of the representational model to allow for different configurations of the model. As an example, an adaptable representational framework would have most of its elements configurable by Subject matter Experts (SME). Typical configuration elements might include: a) changes in basic facts (knowledge); b) adding new rules and classes to the model; c) adding and modeling new input sources; and d) configuration of the possible output of the model.

Simplicity Simplicity criteria are meant to access the level of complexity involved in dealing with the representational framework. In general, a representational model that allows users to execute common operations (e.g. configure the system, enter evidence, proceed with analysis, etc.) without requiring deep knowledge about the inner details of the technique (e.g. the mathematical underpinnings of the inferential process) should meet the simplicity criteria.

Compatibility This is a measure of how compatible a given knowledge representation is to data standards, and should be related to the degree of flexibility it has in being coded with various standards.

F.3 Reasoning Criteria

This general concept encompasses the criteria that directly affect the way the system transforms its data into knowledge. It can also be called as process or inference criteria, as it deals with how the uncertainty model performs operations with information.

F.3.1 Correctness

This is a measure of the inferential process ability to produce correct results. In cases where there is no ground truth to establish a correct answer (including a simulated ground truth), the representation technique can still be evaluated in terms of how its answers align with what is expected from a gold standard (e.g. SMEs, documentation, etc.)

F.3.2 Consistency

This is a measure of the inferential process ability to produce the same results when input with the same data within the same conditions.

F.3.3 Scalability

This is a measure of a representational technique's ability to be used in different magnitudes of data within the same problem. It might be broken down into sub-criteria such as modularity.

F.3.4 Computational Cost

This is a measure of how much of the system's computational resources are required by a given representational technique to produce its results.

F.3.5 Performance

These include metrics to assess how suitable the representational model is to handle the functional requirements of an information fusion system. Other system architecture factors also affect these metrics.

F.3.5.1 Timeliness (from data input to product output)

This measures how long a given uncertainty representation technique takes to produce its results since data is input. Taken from another perspective, it measures whether the representation technique is capable of producing results within the timeframe required by the system's performance goals.

F.3.5.2 Throughput (average / peak rates through the system)

This is a measure of the average and (possibly) peak rate through the system. This differs from timeliness in that a system can have a long timeliness, but still produce a large number of answers in a given amount of time.

F.4 Output Criteria

These criteria are usually related to the system's results and its ability to communicate it to its users in a clear fashion.

F.4.1 Quality

This is a group of criteria meant to assess the informational quality of the system's output. It is common to see in the literature the same concepts with different names. For example, Accuracy sometimes is used as a synonym of precision; sometimes they are used with the exact opposite of their use below.

F.4.1.1 Accuracy

Criteria on accuracy are meant to assess the output of the system in terms of "how right" the answers are. Usual metrics include rate of correct identification/hit, false alarm rate, etc.

F.4.1.2 Precision

Criteria on precision are meant to assess the output of the system in terms of "how good" the answers are. It is a measure of the granularity of the system's output.

Accuracy and precision can be inversely related. As one makes the granularity coarser, one can expect that the system will have a better accuracy. Precision can also be used to put a boundary on the certainty of the reported result.

F.4.2 Interpretation

The output of the system in terms of uncertainty representation and reasoning should have a clear meaning that is sufficiently definite to be used to guide assessment, to understand the conclusions of the system and use them as a basis for action, and to support the rules for combining and updating measures (adapted from Walley 1996).

F.4.3 Traceability

Traceability criteria focus on establishing a correlation between the outcome of the reasoning process with the various input and events computed by the system, so for example one can easily explain why and how the system arrived to a specific answer.

Bibliography

[Aamodt 1995] A. Aamodt and M. Nygard. *Different roles and mutual dependencies of data, information, and knowledge–An AI perspective on their integration*. Data & Knowledge Engineering, vol. 16, no. 3, pages 191–222, 1995. (Cited on page 47.)

[Albertoni 2006] R. Albertoni and M. De Martino. *Semantic similarity of ontology instances tailored on the application context*. On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, pages 1020–1038, 2006. (Cited on page 97.)

[Alberts 2010] D.S. Alberts. *The Agility Imperative: Precis*, 2010. (Cited on page 20.)

[Allemang 2008] D. Allemang and J.A. Hendler. Semantic web for the working ontologist: modeling in RDF, RDFS and OWL. Morgan Kaufmann, 2008. (Cited on pages 52 and 55.)

[Arp 2008] R. Arp and B. Smith. *Function, role, and disposition in basic formal ontology*. Nature Preceedings, pages 1–4, 2008. (Cited on page 69.)

[Bellenger 2010] A. Bellenger and S. Gatepaille. *Uncertainty in Ontologies: Dempster-Shafer Theory for Data Fusion Applications*. In Workshop on the Theory of Belief Functions, 2010. (Cited on page 82.)

[Bellenger 2011a] A. Bellenger, S. Gatepaille, H. Abdulrab and J.P. Kotowicz. *An Evidential Approach for Modeling and Reasoning on Uncertainty in Semantic Applications*. In Proceedings of the 7th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW) at the 10th International Semantic Web Conference (ISWC), volume 778 of *CEUR Workshop Proceedings*, pages 27–38, 2011. (Cited on page 82.)

[Bellenger 2011b] A. Bellenger, X. Lerouvreur, S. Gatepaille, H. Abdulrab and J.P. Kotowicz. *An Information Fusion Semantic and Service Enablement Platform: the FusionLab Approach*. In Proceedings of the 14th International Conference on Information Fusion, 2011. (Cited on pages 133 and 134.)

[Bellenger 2012a] A. Bellenger, X. Lerouvreur, H. Abdulrab and J.P. Kotowicz. *Semantic Beliefs Fusion*. In Advances on Computational Intelligence, volume 297 of *Communications in Computer and Information Science*, pages 491–500. Springer, 2012. (Cited on page 82.)

[Bellenger 2012b] Amandine Bellenger, Sylvain Gatepaille and Habib Abdulrab. *Patent entitled "Procédé permettant la fusion de croyances sémantiques"*, 2012. (Cited on page 120.)

[Bellinger 2004] G. Bellinger, D. Castro and A. Mills. *Data, information, knowledge, and wisdom*. URL: `http://www.systems-thinking.org/dikw/dikw.htm`, 2004. (Cited on page 47.)

[Bengtsson 2001] M. Bengtsson and J. Schubert. *Dempster–Shafer clustering using Potts spin mean field theory*. Soft Computing-A Fusion of Foundations, Methodologies and Applications, vol. 5, no. 3, pages 215–228, 2001. (Cited on page 170.)

[Berners-Lee 2001] T. Berners-Lee, J. Hendler, O. Lassila*et al. The Semantic Web*. Scientific American, vol. 284, no. 5, pages 34–43, 2001. (Cited on page 51.)

[Blasch 2006] E. Blasch, I. Kadar, J. Salerno, M.M. Kokar, S. Das, G.M. Powell, D.D. Corkill and E.H. Ruspini. *Issues and challenges in situation assessment (level 2 fusion)*. Journal of Advances in Information Fusion, vol. 1, no. 2, pages 122–139, 2006. (Cited on page 27.)

[Blasch 2010] E. Blasch, J. Llinas, D. Lambert, P. Valin, S. Das, C. Chong, M. Kokar and E. Shahbazian. *High Level Information Fusion Developments, Issues, and Grand Challenges*. In Proceedings of the 13th International Conference on Information Fusion. IEEE, 2010. (Cited on page 26.)

[Blasch 2012] E. Blasch, P. Costa, K. Laskey, D. Stampouli, G.W. Ng, J. Schubert, R. Nagi and P. Valin. *Issues of uncertainty analysis in High-level information fusion*. In Proceedings of the 15th International Conference on Information Fusion, 2012. (Cited on page 169.)

[Bobillo 2008] F. Bobillo and U. Straccia. *fuzzyDL: An expressive fuzzy description logic reasoner*. In IEEE International Conference on Fuzzy Systems, pages 923–930. IEEE, 2008. (Cited on pages 67, 68, 69 and 71.)

[Boisot 2004] M. Boisot and A. Canals. *Data, information and knowledge: have we got it right?* Journal of Evolutionary Economics, vol. 14, no. 1, pages 43–67, 2004. (Cited on page 47.)

[Borst 1997] W.N. Borst. *Construction of engineering ontologies for knowledge sharing and reuse*. PhD thesis, Universiteit Twente, 1997. (Cited on page 49.)

[Bossé 2007] Éloi Bossé, Jean Roy and Steve Wark. Concepts, models, and tools for information fusion. Artech House, Inc., 2007. (Cited on page 61.)

[Boury-Brisset 2003] A.C. Boury-Brisset. *Ontology-based approach for information fusion*. In Proceedings of the 6th International Conference on Information Fusion, pages 522–529, 2003. (Cited on page 62.)

[Boury-Brisset 2004] A.C. Boury-Brisset, DEFENCE RESEARCH and DEVELOPMENT CANADA VALCARTIER (QUEBEC). *Ontological Approach to Military Knowledge Modeling and Management*. 2004. (Cited on page 26.)

[Carroll 2005] J.J. Carroll, C. Bizer, P. Hayes and P. Stickler. *Named graphs, provenance and trust*. In Proceedings of the 14th international conference on World Wide Web, pages 613–622. ACM, 2005. (Cited on page 68.)

[Carvalho 2010] R. Carvalho, K.B. Laskey and P. Costa. *PR-OWL 2.0–Bridging the Gap to OWL Semantics*. In Proceedings of the 6th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW) at the 9th International Semantic Web Conference (ISWC). Springer, 2010. (Cited on page 70.)

[Costa 2005] P.C.G. Costa. *Bayesian semantics for the Semantic Web*. PhD thesis, 2005. (Cited on page 63.)

[Costa 2006] P.C.G. Costa and K.B. Laskey. *PR-OWL: A framework for probabilistic ontologies*. In Proceeding of the Fourth International Conference on Formal Ontology in Information Systems (FOIS), pages 237–249. IOS Press, 2006. (Cited on pages 68, 70 and 75.)

[Costa 2010] P.C.G. Costa, K.C. Chang, K. Laskey, T. Levitt and W. Sun. *High-level fusion: Issues in developing a formal theory*. In Proceedings of the 13th International Conference on Information Fusion, pages 1–8. IEEE, 2010. (Cited on page 62.)

[Costa 2012] P. Costa, K. Laskey, E. Blasch and A.L. Jousselme. *Towards Unbiased Evaluation of Uncertainty Reasoning: The URREF Ontology.* In Proceedings of the 15th International Conference on Information Fusion, 2012. (Cited on pages 149 and 156.)

[Coucharière 2010] O. Coucharière. *Logique de description ALC possibiliste pour les systèmes d'information opérationnels: théorie, algorithme et applications.* PhD thesis, 2010. (Cited on page 71.)

[Couto 2007] F.M. Couto, M.J. Silva and P.M. Coutinho. *Measuring semantic similarity between Gene Ontology terms.* Data & knowledge engineering, vol. 61, no. 1, pages 137–152, 2007. (Cited on page 97.)

[d'Amato 2007] C. d'Amato. *Similarity-based Learning Methods for the Semantic Web.* PhD thesis, 2007. (Cited on page 97.)

[Das 2011] S. Das. *Unstructured Data Analysis for High-Level Fusion.* Tutorial in the 14th International Conference on Information Fusion, July 2011. (Cited on page 33.)

[DefenceWeb 2011] DefenceWeb. *US military overwhelmed by surveillance data.* `http://www.defenceweb.co.za/index.php?option%3Dcom_content%26view%3Darticle%26id%3D20323:us-military-overwhelmed-by-surveillance-data%26catid%3D48:Information%20%26%20Communication%20Technologies%26Itemid%3D109`, 2011. (Cited on page 21.)

[Denoeux 2006] T. Denoeux. *The cautious rule of combination for belief functions and some extensions.* In Proceedings of the 9th International Conference on Information Fusion, pages 1–8. IEEE, 2006. (Cited on page 43.)

[Denoeux 2011] T. Denoeux. *Methods for building belief functions.* Spring School on Belief Functions Theory and Applications, France, 2011. (Cited on page 38.)

[Dezert 2004] J. Dezert and F. Smarandache. *Advances and applications of DSmT for information fusion.* Am. Res. Press, Rehoboth, pages 61–104, 2004. (Cited on pages 73 and 74.)

[Ding 2004] Z. Ding and Y. Peng. *A probabilistic extension to ontology language OWL.* In Proceedings of the 37th International Conference on System Sciences, pages 10–pp. IEEE, 2004. (Cited on pages 68, 70 and 72.)

[Ding 2006] Z. Ding, Y. Peng and R. Pan. *BayesOWL: Uncertainty modeling in semantic web ontologies.* Soft Computing in Ontologies and Semantic Web, pages 3–29, 2006. (Cited on pages 66, 67 and 70.)

[Dividino 2009] R. Dividino, S. Sizov, S. Staab and B. Schueler. *Querying for provenance, trust, uncertainty and other meta knowledge in RDF.* Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, no. 3, pages 204–219, 2009. (Cited on page 68.)

[Dorion 2005] E. Dorion, C.J. Matheus, M.M. Kokar, DEFENCE RESEARCH and DEVELOPMENT CANADA VALCARTIER (QUEBEC). *Towards a formal ontology for military coalitions operations.* 2005. (Cited on page 132.)

[Dretske 1981] F. Dretske. *The pragmatic dimension of knowledge.* Philosophical Studies, vol. 40, no. 3, pages 363–378, 1981. (Cited on page 47.)

[Dubois 2008] D. Dubois and H. Prade. *A set-theoretic view of belief functions*. Classic Works of the Dempster-Shafer Theory of Belief Functions, pages 375–410, 2008. (Cited on page 43.)

[Dubois 2011] D. Dubois. *Uncertainty Theories: A Unified View*. Spring School on Belief Functions Theory and Applications, France, 2011. (Cited on pages 32, 34 and 37.)

[Endsley 1995] M.R. Endsley. *Toward a theory of situation awareness in dynamic systems*. Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37, no. 1, pages 32–64, 1995. (Cited on pages 24 and 25.)

[Endsley 2000] M.R. Endsley. *Theoretical underpinnings of situation awareness: A critical review*. Situation awareness analysis and measurement, pages 3–32, 2000. (Cited on page 24.)

[Essaid 2009] A. Essaid and B.B. Yaghlane. *BeliefOWL: An Evidential Representation in OWL Ontology*. In Proceedings of the 5th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW) at the 8th International Semantic Web Conference (ISWC), page 77. Citeseer, 2009. (Cited on pages 68 and 72.)

[ETURWG 2012] ETURWG. *Use Case 1: Ship Locating and Tracking*, 2012. (Cited on page 155.)

[Fensel 2001] D. Fensel, F. Van Harmelen, I. Horrocks, D.L. McGuinness and P.F. Patel-Schneider. *OIL: An ontology infrastructure for the semantic web*. Intelligent Systems, vol. 16, no. 2, pages 38–45, 2001. (Cited on page 50.)

[Garcia González 2006] R. Garcia González*et al. A semantic web approach to digital rights management*. PhD thesis, Universitat Pompeu Fabra, 2006. (Cited on page 46.)

[Genesereth 1992] M.R. Genesereth, R.E. Fikes*et al. Knowledge interchange format-version 3.0: reference manual*. 1992. (Cited on page 50.)

[Giroux 2008] P. Giroux, S. Brunessaux, S. Brunessaux, J. Doucy, G. Dupont, B. Grilheres, Y. Mombrun, A. Saval and P.A. des Portes. *Weblab: An integration infrastructure to ease the development of multimedia processing applications*. In International Conference on Software and System Engineering and their Applications (ICSSEA), 2008. (Cited on page 129.)

[Group 2003] Common Logic Working Group*et al. Common logic: Abstract syntax and semantics*, 2003. (Cited on page 50.)

[Gruber 1993] T.R. Gruber*et al. A translation approach to portable ontology specifications*. Knowledge acquisition, vol. 5, pages 199–199, 1993. (Cited on page 49.)

[Haase 2005] P. Haase and L. Stojanovic. *Consistent evolution of owl ontologies*. The Semantic Web: Research and Applications, pages 91–133, 2005. (Cited on page 56.)

[Hitzler 2009] P. Hitzler, M. Krötzsch, B. Parsia, P.F. Patel-Schneider and S. Rudolph. *OWL 2 Web Ontology Language Primer. W3C Recommendation 27 October 2009*. World Wide Web Consortium, 2009. (Cited on pages 54, 57 and 82.)

[Hobbs 2006] J.R. Hobbs and F. Pan. *Time Ontology in OWL. W3C Working Draft, 27 September 2006*. W3C (`http://www.w3.org/TR/2006/WD-owl-time-20060927/`), 2006. (Cited on page 133.)

[Hois 2009] J. Hois. *A Semantic Framework for Uncertainties in Ontologies*. In Twenty-Second International FLAIRS Conference, 2009. (Cited on pages 62 and 67.)

[Jackson 1990] P. Jackson. Introduction to expert systems. Addison-Wesley Longman Publishing Co., Inc., 1990. (Cited on page 151.)

[Jarrar 2002] M. Jarrar and R. Meersman. *Formal ontology engineering in the dogma approach*. On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE, pages 1238–1254, 2002. (Cited on page 50.)

[Jousselme 2011] A.-L. Jousselme. *The applications of belief functions in a situation analysis process*. Spring School on Belief Functions Theory and Applications, France, 2011. (Cited on page 20.)

[Jousselme 2012] A.L. Jousselme and P. Maupin. *A brief survey of comparative elements for uncertainty calculi and decision procedures assessment (ETUR Panel)*. In Proceedings of the 15th International Conference on Information Fusion, 2012. (Cited on page 172.)

[Keet 2010] C. Keet. *Ontology engineering with rough concepts and instances*. 17th International Conference on Knowledge Engineering and Knowledge Management, pages 507–517, 2010. (Cited on pages 66 and 72.)

[Kessler 1992] O. Kessler, K. Askin, N. Beck, J. Lynch, F. White, D. Buede, D. Hall and J. Llinas. *Functional description of the data fusion process*. Office of Naval Technology, Naval Air Development Center, Warminster, PA, 1992. (Cited on page 22.)

[Kifer 1989] M. Kifer and G. Lausen. *F-logic: a higher-order language for reasoning about objects, inheritance, and scheme*. In ACM SIGMOD Record, volume 18, pages 134–146. ACM, 1989. (Cited on page 50.)

[Klinov 2008] P. Klinov. *Pronto: A non-monotonic probabilistic description logic reasoner*. In Proceedings of the 5th European Semantic Web Conference (ESWC), pages 822–826. Springer-Verlag, 2008. (Cited on pages 67, 68, 69, 71 and 75.)

[Klinov 2011] P. Klinov. *Practical reasoning in probabilistic description logic*. PhD thesis, 2011. (Cited on page 63.)

[Kokar 2006] M.M. Kokar, C.J. Matheus, K. Baclawski, J.A. Letkowski, M. Hinman and J. Salerno. *Use cases for ontologies in information fusion*. Technical report, DTIC Document, 2006. (Cited on page 62.)

[Kruger 2008] K. Kruger, U. Schade and J. Ziegler. *Uncertainty in the fusion of information from multiple diverse sources for situation awareness*. In Proceedings of the 11th International Conference on Information Fusion, pages 1–8. IEEE, 2008. (Cited on page 29.)

[Lambert 2003] D.A. Lambert. *Grand challenges of information fusion*. In Proceedings of the Sixth International Conference on Information Fusion. IEEE, 2003. (Cited on page 28.)

[Laskey 2008] K.J. Laskey and K.B. Laskey. *Uncertainty reasoning for the world wide web: Report on the URW3-XG incubator group*. URW3-XG, W3C, 2008. (Cited on pages 33, 67 and 167.)

[Laskey 2011] K.B. Laskey, R. Haberlin, P. Costa and R.N. Carvalho. *PR-OWL 2 Case Study: A Maritime Domain Probabilistic Ontology*. 2011. (Cited on page 67.)

[Laudy 2009] C. Laudy and J.G. Ganascia. *Introducing semantic knowledge in high-level fusion*. In Military Communications Conference, 2009. MILCOM 2009. IEEE, pages 1–7. IEEE, 2009. (Cited on pages 62 and 97.)

[Lesot 2008] M-J Lesot, O. Couchariere, B. Bouchon-Meunier and J-L Rogier. *Inconsistency degree computation for possibilistic description logic: An extension of the tableau algorithm*. In Fuzzy Information Processing Society, 2008. NAFIPS 2008. Annual Meeting of the North American, pages 1–6. IEEE, 2008. (Cited on page 71.)

[Little 2006] E.G. Little and G.L. Rogova. *An ontological analysis of threat and vulnerability*. In Proceedings of the 9th International Conference on Information Fusion, pages 1–8. IEEE, 2006. (Cited on page 62.)

[Little 2008] E. Little, K. Sambhoos and J. Llinas. *Enhancing graph matching techniques with ontologies*. In Proceedings of the 11th International Conference on Information Fusion, pages 1–8. IEEE, 2008. (Cited on page 62.)

[Llinas 2004] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz and F. White. *Revisiting the JDL data fusion model II*. In Proceedings of the 7th International Conference on Information Fusion, volume 2, pages 1218–1230. Citeseer, 2004. (Cited on page 60.)

[Lord 2003] P.W. Lord, R.D. Stevens, A. Brass and C.A. Goble. *Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation*. Bioinformatics, vol. 19, no. 10, pages 1275–1283, 2003. (Cited on page 97.)

[Lukasiewicz 2008] T. Lukasiewicz and U. Straccia. *Managing uncertainty and vagueness in description logics for the semantic web*. Web Semantics: Science, Services and Agents on the World Wide Web, vol. 6, no. 4, pages 291–308, 2008. (Cited on page 65.)

[Martin 2008] A. Martin, C. Osswald, J. Dezert and F. Smarandache. *General combination rules for qualitative and quantitative beliefs*. Journal of Advances in Information Fusion, pages 67–89, 2008. (Cited on pages 42 and 72.)

[Martin 2009] A. Martin. *Modélisation et gestion du conflit dans la théorie des fonctions de croyance*. 2009. (Cited on page 74.)

[Matheus 2006] C. Matheus and B. Ulicny. *On the automatic generation of an OWL ontology based on the Joint C3 Information Exchange Data Model*. In Proceedings of the 12th International Command and Control Research and Technology Symposium, 2006. (Cited on page 132.)

[McGuinness 2000] B. McGuinness and L. Foy. *A subjective measure of SA: the crew awareness rating scale (CARS)*. In Proceedings of the first human performance, situation awareness, and automation conference, 2000. (Cited on page 25.)

[Mitra 2005] P. Mitra, N. Noy and A. Jaiswal. *Omen: A probabilistic ontology mapping tool*. In Proceedings of the 4th International Semantic Web Conference (ISWC), pages 537–547. Springer, 2005. (Cited on pages 66, 68 and 70.)

[Nagy 2006] M. Nagy, M. Vargas-Vera and E. Motta. *DSSim-ontology mapping with uncertainty*. 2006. (Cited on page 72.)

[Nikolov 2008] A. Nikolov, V. Uren, E. Motta and A. De Roeck. *Using the Dempster-Shafer theory of evidence to resolve ABox inconsistencies*. In Proceedings of the 4th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW) at the 7th International Semantic Web Conference (ISWC), pages 143–160. Springer, 2008. (Cited on pages 66 and 72.)

[Niles 2001] I. Niles and A. Pease. *Towards a standard upper ontology.* In Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001, pages 2–9. ACM, 2001. (Cited on page 69.)

[Pan 2005] R. Pan, Z. Ding, Y. Yu and Y. Peng. *A Bayesian network approach to ontology mapping.* In Proceedings of the 4th International Semantic Web Conference (ISWC), pages 563–577. Springer, 2005. (Cited on page 70.)

[Pawlak 1995] Z. Pawlak. *Rough sets.* In Proceedings of the ACM 23rd annual conference on Computer science, pages 262–264. ACM, 1995. (Cited on page 34.)

[Predoiu 2009] L. Predoiu and H. Stuckenschmidt. *Probabilistic Models for the Semantic Web.* The Semantic Web for Knowledge and Data Management: Technologies and Practices, 2009. (Cited on page 65.)

[Price 2004] R. Price, M. Oxenham, J. Percival and D. Lambert. *Information Fusion and Extraction Priorities for Australia's Information Capability*, 2004. (Cited on page 26.)

[Quillian 1968] MR. Quillian. *Semantic Memory.* Semantic Information Processing, Cambridge, MA: MIT Press. Copyright 1968 by MIT Press, Reprinted by permission, page 236, 1968. (Cited on page 48.)

[Rodriguez 2003] M.A. Rodriguez and M.J. Egenhofer. *Determining semantic similarity among entity classes from different ontologies.* Knowledge and Data Engineering, IEEE Transactions on, vol. 15, no. 2, pages 442–456, 2003. (Cited on page 97.)

[Rogova 2010] G.L. Rogova and E. Bosse. *Information quality in information fusion.* In Proceedings of the 13th International Conference on Information Fusion, pages 1–8. IEEE, 2010. (Cited on pages 32 and 33.)

[Rombaut 2001] M. Rombaut. *Fusion: état de l'art et perspectives,.* Rapport DGA Convention DSP 99.60.078, LIS UMR 5083 - INPG, page 47, 2001. (Cited on page 167.)

[RTG-044 2011] NATO IST Panel Research Task Group IST-094 / RTG-044. *Framework for Semantic Interoperablity.* Technical report, North Atlantic Treaty Organization - Research and Technology Agency, 2011. (Cited on page 61.)

[Salerno 2007] J.J. Salerno. *Where's Level 2/3 Fusion–a Look Back over the Past 10 Years.* In Proceedings of the 10th International Conference on Information Fusion, pages 1–4, 2007. (Cited on page 26.)

[Schlobach 2007] S. Schlobach, M. Klein and L. Peelen. *Description logics with approximate definitions: Precise modeling of vague concepts.* In Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI, volume 7, pages 6–12, 2007. (Cited on page 72.)

[Schubert 2008] J. Schubert. *Clustering decomposed belief functions using generalized weights of conflict.* International journal of approximate reasoning, vol. 48, no. 2, pages 466–480, 2008. (Cited on page 170.)

[Schubert 2010] J. Schubert. *Constructing and reasoning about alternative frames of discernment.* In Proceedings of the Workshop on the Theory of Belief Functions, 2010. (Cited on page 167.)

[Shafer 1976] G. Shafer. A mathematical theory of evidence, volume 1. Princeton university press Princeton, NJ, 1976. (Cited on pages 36 and 73.)

[Simou 2007] N. Simou and S. Kollias. *Fire: A fuzzy reasoning engine for impecise knowledge*. In K-Space PhD Students Workshop, volume 14. Citeseer, 2007. (Cited on page 71.)

[Smart 2005] P.R. Smart, N.R. Shadbolt, L.A. Carr and M.C. Schraefel. *Knowledge-based information fusion for improved situational awareness*. In Proceedings of the 8th International Conference on Information Fusion, volume 2, pages 8–pp. IEEE, 2005. (Cited on page 62.)

[Smets 1990] P. Smets. *Constructing the pignistic probability function in a context of uncertainty*. In Uncertainty in artificial intelligence, volume 5, pages 29–39, 1990. (Cited on page 43.)

[Smets 1993] P. Smets. *Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem*. International Journal of Approximate Reasoning, vol. 9, no. 1, pages 1–35, 1993. (Cited on page 43.)

[Smets 1997] P. Smets. *Imperfect information: Imprecision-uncertainty*. Uncertainty management in information systems. from needs to solutions, pages 225–254, 1997. (Cited on page 29.)

[Sølvberg 1999] A. Sølvberg. *Data and what they refer to*. Conceptual Modeling, pages 211–226, 1999. (Cited on page 46.)

[Sowa 1976] J.F. Sowa. *Conceptual graphs for a data base interface*. IBM Journal of Research and Development, vol. 20, no. 4, pages 336–357, 1976. (Cited on page 48.)

[Steinberg 2004] A.N. Steinberg and C.L. Bowman. *Rethinking the JDL data fusion levels*. NSSDF JHAPL, 2004. (Cited on pages 22 and 23.)

[Straccia 2008] U. Straccia. *Managing uncertainty and vagueness in description logics, logic programs and description logic programs*. Reasoning Web, pages 54–103, 2008. (Cited on page 65.)

[Toth 2008] G. Toth, M. Kokar, K. Wallenius, K. Laskey, M. Sudit, M. Hultner and O. Kessler. *Higher-Level Information Fusion: Challenges to the Academic Community*. In Proceedings of the 11th International Conference on Information Fusion, 2008. (Cited on page 27.)

[Walley 1991] P. Walley. Statistical reasoning with imprecise probabilities. Chapman and Hall London, 1991. (Cited on page 34.)

[Waltz 1990] E. Waltz and J. Llinas. Multisensor data fusion. Artech House, 1990. (Cited on page 26.)

[Wartik 2009] S. Wartik. *A JC3IEDM OWL-DL Ontology*. OWL: Experiences and Directions, 2009. (Cited on page 132.)

[Wu 1994] Z. Wu and M. Palmer. *Verb semantics and lexical selection*. In Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics, pages 133–138, 1994. (Cited on page 98.)

[Yager 1987] R.R. Yager. *On the Dempster-Shafer framework and new combination rules*. Information sciences, vol. 41, no. 2, pages 93–137, 1987. (Cited on page 43.)

[Yaghlane 2007] B.B. Yaghlane and N. Laamari. *Owl-cm: Owl combining matcher based on belief functions theory*. In Proceedings of the 2nd International Workshop on Ontology Matching. Citeseer, 2007. (Cited on page 72.)

[Yang 2005] Y. Yang and J. Calmet. *Ontobayes: An ontology-driven uncertainty model.* In International Conference on Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, volume 1, pages 457–463. IEEE, 2005. (Cited on pages 71 and 72.)

[Zadeh 1965] L.A. Zadeh. *Fuzzy sets.* Information and control, vol. 8, no. 3, pages 338–353, 1965. (Cited on page 35.)

[Zadeh 1978] L.A. Zadeh. *Fuzzy sets as a basis for a theory of possibility.* Fuzzy sets and systems, vol. 1, no. 1, pages 3–28, 1978. (Cited on page 36.)

Semantic Decision Support for Information Fusion Applications

Abstract: Underlying much of the information processing applications is the need to fuse information. Moreover, with the necessity to represent and reason over more complex symbolic and semantic information, these applications are turning towards new standardized Semantic Web technologies, such as OWL ontologies.

Within the Defense and Security context, information fusion applications aim to combine data and information coming from a large set of sensors in order to deliver to the decision-maker a better understanding of the observed situation and of its future evolution. Interest for ontologies in this domain is high. However, the question of dealing with uncertainty, which is ubiquitous and inherent to any information fusion applications, is still considered as a major challenge in Semantic Web standards.

The main objective of this thesis is thus to provide a framework for fusion of semantic beliefs. It enables the representation of uncertain and semantic pieces of information and offers a way to reason over it. Through the meta-ontology we proposed - called the *DS-Ontology* -, different sources can populate a same domain ontology, according to their own semantic beliefs, resulting in a set of candidate instances with associated numerical weights on them.

We then propose an automatic process for generating a proper aligned space of hypotheses definition. For that purpose, notions of semantic inclusion and intersection between instances of an ontology are introduced by relying on their semantics (hierarchical structure, constraint axioms and properties defined in the ontology). These operators have been validated by experimentations on human reasoning. This step enables us to make a direct and consistent use of the Evidential theory, also known as the Dempster-Shafer theory, which is an extension of probabilities. It is used to combine our semantic beliefs and support decision on the "best" hypothesis.

This formal framework has been implemented through a prototype. It is used in a set of Protégé (well-known ontology editor in the community) plugins to graphically edit and reason over semantic beliefs. It has also been encapsulated as a service in the *FusionLab* information fusion integration platform, developed within Cassidian - an EADS Company. This service has particularly been used and experimented in a European project for maritime surveillance.

Keywords: Information Fusion, Ontologies, Uncertainty, Evidential Theory, Dempster-Shafer Theory, Beliefs, Semantic Similarities

Aide à la Décision Sémantique pour la Fusion d'Informations

Résumé: Le besoin de fusionner de l'information est sous-jacent à un grand nombre d'applications de traitement de l'information. De plus, avec la nécessité de représenter et de raisonner sur des informations symboliques et sémantiquement de plus en plus complexes, ces applications se tournent dorénavant vers les nouveaux standards des technologies du Web Sémantique, telles que les ontologies OWL.

Au sein du contexte de Défense et Sécurité, les applications dites de fusion d'informations ont pour objectif de combiner des données et informations provenant d'un grand nombre de senseurs afin de fournir aux décideurs une meilleure compréhension de la situation observée et de ses possibles évolutions futures. L'intérêt pour les ontologies dans ce domaine est grand. Cependant, la question de la gestion de l'incertitude, qui est inhérente et omniprésente à n'importe quelle application de fusion d'informations est toujours considérée comme un défaut majeur de ces standards du Web Sémantique.

L'objectif principal de cette thèse est de fournir un framework de fusion de croyances sémantiques. Celui-ci permet la représentation d'informations sémantiques et incertaines et en offre un raisonnement adéquat. Au travers d'une méta-ontologie que nous avons proposée - la *DS-Ontology* -, différentes sources peuvent peupler une même ontologie de domaine, conformément à leurs propres croyances sémantiques, aboutissant à un ensemble d'instances candidates associées à des poids numériques.

Nous proposons ensuite un processus automatique de génération de l'espace de définition des hypothèses. A cette fin, les notions d'inclusion et d'intersection sémantique entre instances d'une ontologie sont introduites en se basant sur leur sémantique (structure hiérarchique, axiomes de contraintes logiques et propriétés définies dans l'ontologie). Ces opérateurs ont été validés via des expérimentations sur le raisonnement humain. Cette étape nous permet d'utiliser directement et de façon consistante la théorie de l'Evidence, aussi appelée la théorie de Dempster-Shafer, qui est une extension des probabilités. Celle-ci est utilisée pour combiner nos croyances sémantiques et aider à la décision de la "meilleure" hypothèse.

Ce framework a été implémenté au travers d'un prototype. Il est utilisé par un ensemble de plugins se basant sur le logiciel *Protégé* permettant d'éditer et de raisonner graphiquement sur les croyances sémantiques. Ce prototype a aussi été intégré en tant que service dans la plateforme d'intégration d'applications de fusion d'informations: le *FusionLab*, développé au sein de Cassidian. Ce service a été employé et expérimenté pour un projet européen dans la surveillance maritime.

Mots clés: Fusion d'Informations, Ontologies, Incertitude, Théorie de l'Evidence, Théorie de Dempster-Shafer, Croyances, Similarités Sémantiques