

# Knowledge Engineering for Intelligent Decision Support

**Maria Vanina Martinez**

Departamento de Ciencias e Ingenieria de la Computacion, Universidad Nacional del Sur (UNS)  
Institute for Computer Science and Engineering (UNS-CONICET)  
San Andrés 800, (8000) Bahia Blanca, Argentina  
mvm@cs.uns.edu.ar

## Abstract

Knowledge can be seen as the collection of skills and information an individual (or group) has acquired through experience, while intelligence as the ability to apply such knowledge. In many areas of Artificial Intelligence, we have been focusing for the last 40 years on the formalization and development of automated ways of finding and collecting data, as well as on the construction of models to represent that data adequately in a way that an automated system can make sense of it. However, in order to achieve real artificial intelligence we need to go beyond data and knowledge representation, and deeper into how such a system could, and would, use available knowledge in order to empower and enhance the capabilities of humans in making decisions in real-world applications. From my point of view, an AI should be able to combine automatically acquired data and knowledge together with specific domain expertise from the users that the tool is expected to help.

## 1 Introduction

Nowadays, large repositories of data are readily accessible and used daily as knowledge bases (KBs) that feed a wide variety of information systems. One important type of such systems are those that support decision making processes—humans use them to interpret the data contained in a database in order to make real world decisions. Examples can be found across all kinds of everyday applications, from retail recommendations to medical, financial, behavioral advice and forecasting applications, etc. Users of these applications leverage the data to make inferences and take actions in the world; sometimes data and actions are about the users themselves, but many times they are about, and affect, other people's lives.

Many of these applications or tasks were originally performed solely by humans: specialists or analysts would go through all the data to make sense of it. This was a quite personalized process, but also an expensive, slow, inefficient, and domain-dependent one. Besides the large amounts of available data, it is important to note the shift in the expectations that users have about the services they consume. Today, *personalization* of services is of paramount importance—it is

clear that if your service/system does not deliver a customized product or experience, you will probably lose your client to a competitor. It is no longer possible, or even desirable, to rely on decision making performed manually; the involvement of an automated system is becoming a requirement.

However, in the modelling of intelligent support tools, it is important to understand that, as part of the decision making process, humans consider far more knowledge than that contained in the database; they primarily incorporate into the process their domain expertise and requirements, as well as their expectations. This is a key point for the construction of an actually useful, robust, and intelligent system that empowers users and enhances their capabilities to reach useful conclusions. Machines therefore need to *understand people*, which could be application users or the subjects of the task itself if the application domain requires reasoning about people's behavior; they need to understand people's goals, what they already know (and their *know-how*), but also their personality traits, preferences, and (individual and social) behavior patterns. The automated system needs to work with the user as a partner that is trying to help solve a problem, creating a synergy that can enhance the decision making process.

During this first of part my career I focused on tackling different problems that I believe stand in the way of reaching these ambitious goals. The three pillars of Intelligent Decision Support development are, from my point of view: (i) Representation and modelling of individual (and group) preferences, together with reasoning tools that allow to understand and use them in a fruitful manner; (ii) Sensible and personalized management of conflicting information at different levels of abstraction; and (iii) Understanding and representing people's knowledge and skills as they evolve over time, in order to predict or consider ahead of time the consequences of this evolution for the decision making process. On a broader scale, my view of the overall task of building intelligent systems—and in particular intelligent decision support ones—is summarized in Figure 1. An enormous amount of information that is vital for feeding these systems is readily available; most of it in a flat and unstructured format that cannot be easily interpreted, neither by humans nor machines. It is the combination of data-driven (or machine learning) tools together with knowledge and logic-based frameworks (or KR) that can structure this data and formalize it into models of knowledge that are amenable to being processed by

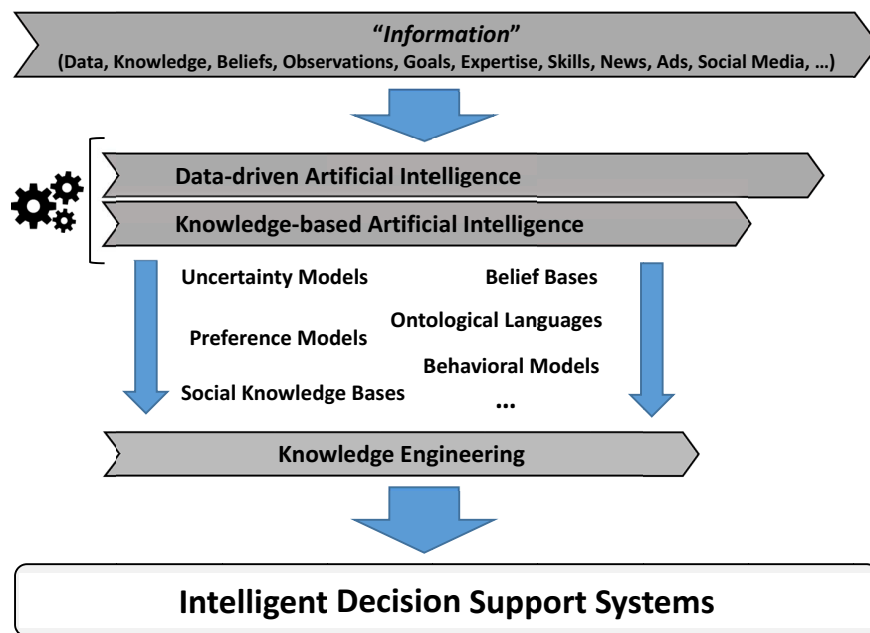


Figure 1: Construction of an Intelligent Decision Support System.

information systems. These kinds of rich models—and actually the full potential of AI—can only be reached by an effective combination of these two approaches. The knowledge engineering step contains all the technical, scientific, and social aspects that are involved in building such systems; it is here where the synergy between the automated decision support system and the human users arises—even assuming that the more basic problems tackled in the previous steps are solved, concocting a final product with these properties is a formidable task.

## 2 Modeling Preferences

Understanding and being able to model people’s preferences is of utmost importance for an intelligent decision support system. As a result, many information systems use preferences from their users in order to be able to provide an experience that is more adequate to their needs and expectations. There has been a wide variety of work in the study and modeling of preferences in areas as varied as philosophy, logic, and economics. In the philosophical tradition, preferences are usually expressed over mutually exclusive “outcomes”, such as truth assignments to formulas; alternatively, in databases preferences are expressed over tuples (or ground atoms). Preferences can be explicitly defined by the user or automatically obtained from their behavior and/or usage of computing tools, such as social networks, blogs, e-commerce, etc. Independently of where they come from, it is inevitable for them to be incomplete (underspecified), inconsistent (overspecified), or inherently uncertain.

One closely related research line in which I have worked on is the incorporation of preference models into the Datalog+/- family of ontology languages for single users [Lukasiewicz *et al.*, 2013], as well two extensions: (i) In [Lukasiewicz *et al.*,

2015b], we explore the combination of Datalog+/- ontologies with both individual user preferences and (probabilistic) uncertainty with which inferences are made. Assuming that more probable answers are in general more preferable, a basic problem is how to rank answers to a user’s queries, since the preference model may be in conflict with the preferences induced by the probabilistic model—it is thus necessary to carefully combine them. (ii) In [Lukasiewicz *et al.*, 2014b], we extend the preference model to contemplate groups of users, and explore how queries can be answered under different operators to ensure that some kind of fairness is achieved.

Richer models of preferences require reasoning mechanisms that present challenges from the computational complexity point of view. In [Lukasiewicz *et al.*, 2014a], we proposed a novel model for reasoning with preferences *à la* databases under probabilistic uncertainty. Similar to Markov logic, the semantics is based on an underlying Markov random field, but possible worlds consist of permutations of the elements of interest rather than truth assignments. This basic difference causes sources of computational intractability to have far-reaching ripple effects, so new algorithms for query answering are needed; by leveraging results in the mathematical branch of Order Theory, we developed both exact and approximation algorithms that are tractable (and FPTAS, respectively) under fixed-parameter assumptions. Later, in [Lukasiewicz *et al.*, 2016] we tackled the problem of adapting variable elimination strategies for conventional probabilistic models to this setting; since, as mentioned above, their computational properties are very different due to the underlying linear ordering of possible worlds, standard methods that exploit model structure to leverage conditional independence (such as in Bayesian networks) are not directly applicable. We thus developed novel variable elimination techniques for different special cases, with empirical evaluations show-

ing how they scale in practice.

Finally, in [Di Noia *et al.*, 2015] we explore how the tastes of users can be represented using qualitative preferences, in particular using ontological knowledge expressed via existential rules combined with CP-theories to (i) represent qualitative preferences along with domain knowledge, and (ii) perform preference-based answering of conjunctive queries. The query answering semantics for the resulting *OCP-theories* is defined through the notion of skyline and  $k$ -rank answers, based on an encoding of the users preferences.

### 3 Managing Conflicting Information

The notion of inconsistency has been extensively studied in many contexts; the most well-known one is classical logic, in which an inconsistent set of formulas entails every formula (known as *triviality*). Many of the approaches to reason about inconsistent KBs are based on the assumption that there is some “epistemically correct” way of resolving inconsistencies or reasoning in the presence of this kind of uncertainty. During this first part of my career, my take on inconsistency management has been that it is a task that cannot be managed or resolved locally, but instead should be addressed sensibly according to the application domain and the user’s context, considering both the system KB’s contents and the knowledge that is important to the user when making decisions about the real world that the KB intends to model.

A significant part of my research efforts have been dedicated to the problem I called *Personalizable Inconsistency Management*, which was the topic of my doctoral work. In [Martinez *et al.*, 2014], we developed the notion of *Inconsistency Management Policy* (IMP)—by means of IMPs, users can specify the exact set of actions to perform in case of inconsistency in a relational database, in combination with traditional query answering languages. IMPs generalize previous work in *database repairing* and *consistent query answering* (CQA) [Arenas *et al.*, 1999], allowing, for instance, record modification or part of the inconsistency to persist.

Further research efforts led me to consider more expressive knowledge representation languages [Martinez *et al.*, 2013; Lukasiewicz *et al.*, 2012; Gottlob *et al.*, 2013] in order to develop models that could be applied to a wider variety of domains. As most of the proposals for inconsistency management in the literature provide an a priori conflict resolution mechanism, the reality is that users are forced to use the semantics *already implemented* in the system. In [Martinez *et al.*, 2013], we developed a unified formalism that allows end-users to bring their domain knowledge to bear in the process of dealing with inconsistency. The need for such a general framework is motivated by analyzing examples of inconsistent theories using different monotonic logics based on Tarski’s abstract logic.

In the last years there has been much interest on defining inconsistency-tolerant semantics for query answering in Description Logics (DLs) [Lembo *et al.*, 2010; Bienvenu, 2012]. The most widely accepted semantics is that of *consistent answers* based on the notion of *data repairs* (maximal consistent subsets of an inconsistent set of assertional axioms). Consistent answers for DLs are an adaptation of

the work of [Arenas *et al.*, 1999], and are the answers to a query that can be obtained from *every possible* repair. It is highly debatable that this semantics is adequate for every kind of application and, furthermore, computing such answers is expensive—it is computationally intractable even for very simple languages. Though some approximations to this semantics have been defined [Lembo *et al.*, 2010; Bienvenu, 2012], the quality of the answers can be compromised and the gain in tractability is not that significant.

After this realization, the main focus of my research on inconsistency management turned towards rethinking inconsistency-tolerant semantics for ontological languages with the goal of defining reasonable semantics and efficient methods of computation, focusing on Datalog+/-, which is in particular useful for representing and reasoning over lightweight ontologies in the Semantic Web [Cali *et al.*, 2012]. In this line of research, in [Lukasiewicz *et al.*, 2012] we developed a general framework for inconsistency management based on the concept of *incision functions* from belief revision [Hansson, 1994], in which we can characterize several query answering semantics as special cases; we also proposed the notion of *lazy consistent answers*, which provides an alternative semantics that offers a good compromise between quality of answers and computation time. Lazy answers are based on a *budget* that restricts the size of removals that need to be made in a set of facts in order to make it consistent—if the budget is large enough, then we go to the trouble of considering all possible ways of solving the conflicts within the budget, but if it is not enough then we get rid of all the sentences that are involved in that particular conflict. If we think of the problem of querying inconsistent KBs as a reasoning task for an intelligent agent, then the value of the budget would be a bound on its reasoning capabilities (more complex reasoning can thus be afforded with higher budgets). Considering a more complex setting where ontologies can be uncertain, in [Gottlob *et al.*, 2013] we investigated alternative inconsistency-tolerant semantics that take into account probabilistic information. Finally, the work in [Lukasiewicz *et al.*, 2015a] studies in depth different types of computational complexity of CQA for the most tractable fragments of Datalog+/-.

### 4 Understanding Knowledge Dynamics

Knowledge is not static, it changes over time as a result of different phenomena: acquisition of information, learning, interaction with other individuals and public opinion, etc. The same happens with beliefs, opinions, preferences, and other personality traits. As we strive towards modeling humans, these models must evolve accordingly and the dynamics of these processes need to be structured into formal frameworks within intelligent systems.

Belief revision (BR) is the research area within AI that focuses on the formalization of the dynamics of intelligent agents’ epistemic states. The foundational work in BR is the AGM model [Alchourrón *et al.*, 1985], which defined three basic operations of belief change: *contraction*, in which a belief knowledge base  $K$  is superseded by another that is a subset of  $K$  not containing a particular sentence  $p$ ; *expan-*

sion, where a sentence  $p$  is added to a KB  $K$  and nothing is removed (that is,  $K$  is replaced by a set that is the smallest logically closed set that contains both  $K$  and  $p$ ); and finally, *revision*, in which a sentence  $p$  is added to  $K$  and at the same time other sentences are removed if needed to ensure that the resulting knowledge base is consistent. In a dual approach to the AGM model, [Hansson, 1994] proposes belief change operations based on the application of *kernel incision functions* that minimally select formulas to delete from minimal inconsistent sets of belief—this framework was developed for non logically-closed sets of beliefs, while AGM assumes logical closure. Within this setting, we investigated ways in which such operations can be defined so they produce more meaningful results by permitting a more global consideration of conflicts. The work in [Deagustini *et al.*, 2014] introduces an approach to *consolidation* (contraction by falsum) based on a refinement of kernel contraction that accounts for the relation among kernels by leveraging logical structures that try to capture a global view of conflicts instead of the local one represented by kernels or minimal inconsistent sets—we call these structures *clusters*. Cluster contraction-based consolidation operators avoid some unnecessary deletions that are otherwise not identifiable.

An important concept that has not received that much attention in the literature is that of *incoherence* or unsatisfiability of a set of constraints. In [Deagustini *et al.*, 2016], we study incoherence in Datalog+/- ontologies; the main contribution of this work is a novel approach to restoring both consistency and coherence based on cluster incision functions, where clusters are defined both over facts and rules. Finally, the work in [D. *et al.*, 2016] studies how incoherence affects inconsistency-tolerant semantics for query answering in Datalog+/- ontologies, showing that in the presence of this phenomenon most inconsistency-tolerant semantics for query answering do not yield meaningful results.

#### 4.1 The Effects of Social Interactions

It is undebatable that social interactions mold, to different degrees, people’s knowledge, beliefs, personalities, and behavior. The formalization of these interactions together with their consequences into the models that intelligent decision support systems work with is an important—and quite challenging—research and development goal.

Many people today participate in more than one social network on a daily basis; such networks can be seen as dynamic environments containing knowledge about their members. From a KR perspective, these environments are complex in many dimensions; for instance, it is not clear how to treat news being communicated through the network, especially when it contradicts a user’s own knowledge, or when contradicting news is received from different sources. In [Gallo *et al.*, 2017], we focused on reasoning about the *diffusion of beliefs* in social media. A Network Knowledge Base, with an underlying complex network, models all sources of social communication among users, who also have access to a stream of news items that are produced by others. We define and formalize revision operations on a global and local level: local revisions are performed first, where each user responds to news items that show up in their feeds; since each local

revision is carried out in parallel, the result of this stage could violate global integrity constraints. A global revision process is then carried out in order to return to a consistent state.

The work described above abstracts an agent’s epistemic state into a set of logical sentences. However, we can argue that different components of such state deserve particular attention since there are many variables that define people’s personalities beyond knowledge. One such component are preferences, which are dynamic and can be affected greatly by social interactions; the work in [Lukasiewicz *et al.*, 2014b] develops a model that allows for the management of preferences of groups of users. For this work, we developed a prototype implementation of the query answering system that was validated over real-world ontological data and preference models (derived from information gathered from real users), showing in particular that our approach is feasible in practice.

### 5 Future Directions

The main focus of my research career has been so far on formalizing solutions to several problems that are central to the characterization and development of intelligent decision support tools that understand people and can work with them in cooperation so that the capabilities of humans for decision making processes can be enhanced. These solutions fall within the general areas of modeling preferences, managing conflicting information, and understanding social dynamics.

The problems described here are by no means closed or completely solved, much in the same way as many other fundamental problems in AI for which radical breakthroughs will be required. It is important to note that the usability of existing frameworks, such as the ones described here, require the availability of—so far assumed to exist—inputs such as human expertise. Other kind of information needed to understand humans is already available, but even so an engineering effort that combines machine learning tools with knowledge-based approaches is needed in order to produce knowledge that can effectively be used in the models. More specifically from the point of view of decision support systems, solutions to problems such as how to adequately and automatically explain obtained inferences to humans, and how to create the required human-machine synergy and cooperation, still deserve much attention and study in both the basic research and software development realms.

### Acknowledgments

This work has been partially supported by EU H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 690974 for the project MIREL: Mining and Reasoning with Legal texts, and by funds provided by CONICET, Dep. de Ciencias e Ing. de la Computación, Universidad Nacional del Sur, and Agencia Nacional de Promoción Científica y Tecnológica, Argentina.

### References

- [Alchourrón *et al.*, 1985] Carlos Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.*, 50(2):510–530, 1985.

- [Arenas *et al.*, 1999] Marcelo Arenas, Leopoldo Bertossi, and Jan Chomicki. Consistent query answers in inconsistent databases. In *Proc. ACM PODS*, pages 68–79, 1999.
- [Bienvenu, 2012] Meghyn Bienvenu. On the complexity of consistent query answering in the presence of simple ontologies. In *Proc. AAAI*, 2012.
- [Calì *et al.*, 2012] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *J. Web Sem.*, 14:57–83, 2012.
- [D. *et al.*, 2016] Cristhian Ariel D. Maria Vanina Martinez, Marcelo Falappa, and Guillermo R. Simari. How does incoherence affect inconsistency-tolerant semantics for Datalog+/-? *AMAI*, pages 1–26, 2016.
- [Deagustini *et al.*, 2014] Cristhian A.D. Deagustini, Maria Vanina Martinez, Marcelo Falappa, and Guillermo R. Simari. Inconsistency resolution and global conflicts. In *Proc. ECAI*, pages 991–992, 2014.
- [Deagustini *et al.*, 2016] Cristhian Ariel D. Deagustini, Maria Vanina Martinez, Marcelo Falappa, and Guillermo R. Simari. Datalog+/- ontology consolidation. *J. Artif. Intell. Res.*, 56:613–656, 2016.
- [Di Noia *et al.*, 2015] Tommaso Di Noia, Thomas Lukasiewicz, Maria Vanina Martinez, Gerardo I. Simari, and Oana Tifrea-Marcuska. Combining existential rules with the power of cp-theories. In *Proc. IJCAI*, pages 2918–2925, 2015.
- [Gallo *et al.*, 2017] Fabio R. Gallo, Gerardo I. Simari, Maria Vanina Martinez, Natalia Abad Santos, and Marcelo A. Falappa. A first approach to belief dynamics in complex social networks. In *Proc. AMW*, 2017.
- [Gottlob *et al.*, 2013] Georg Gottlob, Thomas Lukasiewicz, Maria Vanina Martinez, and Gerardo I. Simari. Query answering under probabilistic uncertainty in Datalog+/- ontologies. *Ann. Math. Artif. Intell.*, 69(1):37–72, 2013.
- [Hansson, 1994] Sven Ove Hansson. Kernel contraction. *J. Symb. Log.*, 59(3):845–859, 1994.
- [Lembo *et al.*, 2010] Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, Marco Ruzzi, and Domenico Savo. Inconsistency-tolerant semantics for description logics. In *Proc. RR*, pages 103–117, 2010.
- [Lukasiewicz *et al.*, 2012] Thomas Lukasiewicz, Maria Vanina Martinez, and Gerardo I. Simari. Inconsistency handling in Datalog+/- ontologies. In *Proc. ECAI*, pages 558–563, 2012.
- [Lukasiewicz *et al.*, 2013] Thomas Lukasiewicz, Maria Vanina Martinez, and Gerardo I. Simari. Preference-based query answering in Datalog+/- ontologies. In *Proc. IJCAI*, pages 1017–1023, 2013.
- [Lukasiewicz *et al.*, 2014a] Thomas Lukasiewicz, Maria Vanina Martinez, and Gerardo I. Simari. Probabilistic preference logic networks. In *Proc. ECAI*, pages 561–566, 2014.
- [Lukasiewicz *et al.*, 2014b] Thomas Lukasiewicz, Maria Vanina Martinez, Gerardo I. Simari, and Oana Tifrea-Marcuska. Ontology-based query answering with group preferences. *ACM Trans. Internet Techn.*, 14(4):25:1–25:24, 2014.
- [Lukasiewicz *et al.*, 2015a] Thomas Lukasiewicz, Maria Vanina Martinez, Andreas Pieris, and Gerardo I. Simari. From classical to consistent query answering under existential rules. In *Proc. AAAI*, pages 1546–1552, 2015.
- [Lukasiewicz *et al.*, 2015b] Thomas Lukasiewicz, Maria Vanina Martinez, Gerardo I. Simari, and Oana Tifrea-Marcuska. Preference-based query answering in probabilistic Datalog+/- ontologies. *J. Data Semantics*, 4(2):81–101, 2015.
- [Lukasiewicz *et al.*, 2016] Thomas Lukasiewicz, Maria Vanina Martinez, David Poole, and Gerardo I. Simari. Probabilistic models over weighted orderings: Fixed-parameter tractable variable elimination. In *Proc. KR*, pages 494–504, 2016.
- [Martinez *et al.*, 2013] Maria Vanina Martinez, Cristian Molinaro, V.S. Subrahmanian, and Leila Amgoud. *A General Framework for Reasoning On Inconsistency*. Springer Briefs in Computer Science. 2013.
- [Martinez *et al.*, 2014] Maria Vanina Martinez, Francesco Parisi, Andrea Pugliese, Gerardo I. Simari, and V.S. Subrahmanian. Policy-based inconsistency management in relational databases. *Int. J. Approx. Reasoning*, 55(2):501–528, 2014.