Semantic Science: Ontologies, Data and Probabilistic Theories

David Poole[1], Clinton Smyth[2], and Rita Sharma[2]

[1] Department of Computer Science,
University of British Columbia
http://www.cs.ubc.ca/spider/poole/
[2] Georeference Online Ltd.,
http://www.georeferenceonline.com/

Abstract. This chapter overviews work on *semantic science*. The idea is that, using rich ontologies, both observational data and theories that make (probabilistic) predictions on data are published for the purposes of improving or comparing the theories, and for making predictions in new cases. This paper concentrates on issues and progress in having machine accessible scientific theories that can be used in this way. This paper presents the grand vision, issues that have arisen in building such systems for the geological domain (minerals exploration and geohazards), and sketches the formal foundations that underlie this vision. The aim is to get to the stage where: any new scientific theory can be tested on all available data; any new data can be used to evaluate all existing theories that make predictions on that data; and when someone has a new case they can use the best theories that make predictions on that case.

1 Introduction

The aim of the semantic web (Berners-Lee and Fischetti, 1999; Berners-Lee et al., 2001) is that the world's information is available in a machine-understandable form. This chapter overviews what we call *semantic science*, the application of semantic technology and reasoning under uncertainty to the practice of science. Semantic science requires machine-understandable information of three sorts: ontologies to define vocabulary, data about observations of the world, and theories that make predictions on such data.

Our idea of semantic science is that scientists can publish data and theories that can inter-operate by virtue of using common ontologies. The theories can be judged by how well they predict unseen data and can be used for new cases.

An ontology (Smith, 2003b) is a formal specification of the meaning of the vocabulary used in an information system. Ontologies are needed so that information sources can inter-operate at a semantic level.

There has been recent success in publishing scientific data that adheres to ontologies (McGuinness et al., 2007). Publishing data with respect to well-defined ontologies can allow for semantic inter-operation of the data sets. Meaningful queries can be made against multiple data sets that were collected separately.

Data repositories include the Community Data Portal (http://cdp.ucar.edu/) and the Virtual Solar-Terrestrial Observatory (http://vsto.hao.ucar.edu/index.php).

Science operates by making refutable theories (Popper, 1959). These theories[3] are judged by their predictions, by their usefulness, and by their elegance or plausibility. Theories make (probabilistic) predictions about new cases. Theories may require arbitrary computations to make predictions; indeed many real theories need enormous computational resources. Semantic science aims to provide an infrastructure to test theories on data, and to make theories available for new cases.

Theories need to refer to ontologies as they need to inter-operate with data. Theories specify what data they can make predictions about, and make predictions that can be checked against the relevant data and applied to new cases. It is the ontologies that allow the inter-operation of the data and the theories. Theories can be tested against all of the relevant data sets, and data can be used to discriminate theories.

Given access to the theories, and information about how they perform on the available data sets, practitioners can use the best theories to make predictions on new cases. This thus promises to form a new basis for expert systems.

We have been working on two instances of the semantic science framework in two domains in earth sciences (Smyth et al., 2007), namely minerals exploration in the MINEMATCH® system (http://www.georeferenceonline.com/minematch/) and landslides in the HAZARDMATCH[TM] system. MineMatch contains about 25,000 descriptions of mineral occurrences (called *instances*) that are described at various levels of abstraction and detail using multiple taxonomies, including the British Geological Survey rock classification scheme (http://www.bgs.ac.uk/bgsrcs/) and the Micronex taxonomy of minerals (http://micronex.golinfo.com). We are currently moving to OWL representations of the ontologies. We also work with more than 100 deposit models (these form the theories about where to find particular minerals), including those described by the US Geological Survey (http://minerals.cr.usgs.gov/team/depmod.html) and the British Columbia Geological Survey (http://www.em.gov.bc.ca/Mining/Geolsurv/MetallicMinerals/MineralDepositProfiles/). Similarly, HazardMatch uses tens of thousands of spatial instances (polygons) described using standard taxonomies of environmental modeling such as rock type, geomorphology and geological age. There are currently about 10 models of different landslide types that are derived from published models. We can compare the prediction of the models to known cases and new cases.

Semantic science allows for a diversity of theories. Each theory will specify what data it is prepared to make predictions about. Some theories may be competing and some may be complementary. For example, there may be multiple theories that predict whether a patient has cancer. If they make different predictions in some cases, they can be compared by how well they predict the available

---

[3] Theories are often called hypotheses, laws or models depending on how well established they are. This distinction is redundant in the semantic science realm where we can test how well these actually perform on data.

data. There may be other theories that make predictions about the type(s) of cancer for patients with cancer. These theories are not applicable for patients who don't have cancer. When making predictions, a doctor may use an ensemble of multiple complementary theories: e.g., one to predict whether the patient has cancer and another to predict the type of cancer if cancer is present.

Theories can make predictions in different forms. A theory could make, e.g., a definitive prediction, a probabilistic prediction, a range prediction, or a qualitative prediction. Users can use whatever criteria they like to judge the theories, and use whichever theory or mix of theories they like. For different evaluation criteria, there will be ways to judge the theories on the criteria. We anticipate that probabilistic predictions will be the most useful, as it is probabilities that one gets from data, and probabilities are what is needed (with utilities) to make decisions. However, there are many cases where users will be reluctant to use probabilistic theories (see below). Scientists who wish to judge a theory by elegance or simplicity, as well as fit to data, are free to do so; they can use published data to determine its accuracy and whatever criteria they like to evaluate elegance or simplicity.

We mean science in the broadest sense. We can imagine having theories about what apartment someone would like, or theories about what companies will make the best investments, or theories about diseases and symptoms. Search engines such as Google are being used for diagnosis (Tang and Ng, 2006). It is arguably better to be able to specify symptoms unambiguously using an ontology. Measures such as pagerank (Page et al., 1999) measure popularity. Fortunately, searches for diagnostic tend to return authoritative sites. Scientists, however, should be suspicious of popularity and authority as a basis for prediction. We should base our predictions on the empirical evidence. Building an infrastructure for this is the aim of semantic science.
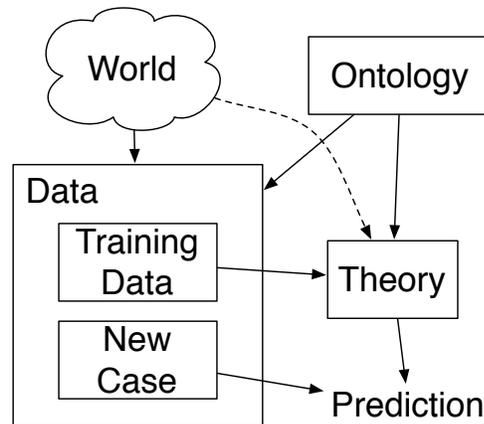

Fig. 1. Ontologies, Data and Theories in Semantic Science

Figure 1 shows the relationship between ontologies, data and theories. The data depends on the world and the ontology. The theories depend on the ontology, indirectly on the world (if a human is designing the theory), and directly on some of the data (as we would expect that the best theories would be based on as much data as possible). Given a new case, a theory can be used to make a prediction. The real world is more complicated, as there are many theories, many ontologies, and lots of data, and they all evolve in time.

This work is complementary to providing services and other tools to scientists, e.g., using the Semantic Grid (De Roure et al., 2005). We expect that the semantic grid will be important for implementing the ideas in this paper.

This chapter is based on Poole et al. (2008).

2 Background

2.1 Ontologies

In philosophy, *ontology* is the study of what exists. In AI, *ontology* (Smith, 2003b) has come to mean a specification of the meaning of the symbols (or of the data) in an information system. In particular, an ontology makes a commitment to what entities and relationships are being modelled, specifies what vocabulary will be used for the entities and relationships, and gives axioms that restrict the use of the vocabulary. The axioms have two purposes: to rule out uses of the terms that are inconsistent with the intended interpretation, and to allow for inference to derive conclusions that are implicit in the use of the vocabulary.

An ontology can be any specification, formal or informal, of the meaning of the symbols. This can be in the head of the person who created the data, or can be stated in some language. Without an ontology, we do not have information, but just a sequence of bits. The simplest form of an ontology is a database schema with informal natural language descriptions of the attributes and the constants. Formal ontologies allow machine understandable specifications.

An ontology written in a language such as OWL (McGuinness and van Harmelen, 2004) specifies individuals, classes and relationships and the vocabulary used to express them. Sometimes classes and relationships are defined in terms of more primitive classes and relationships, but ultimately they are grounded out into primitive classes and relationships that are not actually defined. For example, an ontology could specify that the term "building" will represent buildings. The ontology will not *define* a building, but will give some properties that restrict the use of the term.

Ontologies date back to Aristotle (350 B.C.), who defined terms using what has been called an Aristotelian definition (Berg, 1982; Smith, 2003a). An Aristotelian definition of A is of the form "An A is a B such that C", where B is the immediate super-class of A and C is a condition that defines how A is special. Aristotle called the B the *genus* and C the *differentia* (Sowa, 2000, p. 4).

To build Aristotelian definitions, we will use what we call the *multi-dimensional* design pattern (Alexander et al., 1977), where the differentia in the Aristotelian

definition are built from multiple properties. To define the conditions for a class, we need to think about what properties distinguish this class from the other subclasses of the super-class. Each of these properties defines a (local) dimension. The domain of each property is the most general class for which it makes sense. In the multi-dimensional design pattern, classes are only defined in terms of values of properties. The subclass relation can be derived from this.

There is not a fixed number of dimensions that distinguish all individuals. Rather, dimensions come into existence at different levels of abstraction. For example, the dimensions size and weight may appear for physical individuals, but are not applicable for abstract concepts. "Number of units" may be a dimension for apartment buildings but may not be applicable for other buildings such as sewage plants, where other dimensions may be applicable.

This idea is due to Aristotle:

> "If genera are different and co-ordinate, their differentiae are themselves different in kind. Take as an instance the genus 'animal' and the genus 'knowledge'. 'With feet', 'two-footed', 'winged', 'aquatic', are differentiae of 'animal'; the species of knowledge are not distinguished by the same differentiae. One species of knowledge does not differ from another in being 'two-footed'." (Aristotle, 350 B.C.)

Example 1. Geologists define rocks along three major dimensions: genesis (sedimentary, igneous or metamorphic), composition and texture (Gillespie and Styles, 1999). Particular rocks, such as granite and limestone, are defined by particular values in each dimension (or some subset of the dimensions). Rock taxonomies built using this approach that commit to splitting rock sub-type based on these dimensions in a certain order (usually genesis first, then composition, then texture) do not conveniently represent the sub-types that occur in real data (Struik et al., 2002). For example, if the aforementioned order of splitting the taxonomy is used, there is no convenient single place in the taxonomy for the class of rocks with a particular texture, independent of its members' genesis or composition. The multi-dimensional ontologies seem to be the natural specification, and they also integrate well with probabilities (see Section 4.2).

2.2   Data and Ontologies

Scientists produce lots of data, and science cannot be carried out without data. By data, we mean information about a domain that is produced from sensing.

In linguistics the Sapir-Whorf Hypothesis (Sapir, 1929; Whorf, 1940), says essentially that people's perception and thought are determined by what can be described in their language. The Sapir-Whorf Hypothesis is controversial in linguistics, but a stronger version of this hypothesis should be uncontroversial in information systems:

> What is stored and communicated by an information system is constrained by the representation and the ontology used by the information system.

The reason that this should be less controversial is that the representation and the ontology represent the language of thought or mentalese (Fodor, 1975; Pinker, 1994), not just the language of communication.

As an example, suppose the world produces a deterministic sequence of coin tosses: head, tail, head, tail, head, tail, etc. If the representation and the ontology does not specify the time of each observation or which is the next coin toss in the sequence, that information will have been lost in translating the observation into the internal representation. The best prediction would be to predict heads with probability of 0.5. As another example, if some data adheres to an ontology that specifies that a house is a residential building, then, by definition, all of the observed houses are residential buildings, and so the data cannot refute the fact that houses are residential buildings.

This hypothesis has a number of implications:

- An ontology mediates how perceptions of the world are stored and communicated.
- If there is no distinction in the ontology, there will be no distinction in the data. For example, if an ontology does not have any sub-types of "granite", and does not record the information needed to distinguish between types of granite, the data will not record any sub-types of granite and none can be discovered.
- Ontologies must come before data. This may be confusing as much work is done on building ontologies for existing data sets. This activity should be seen as reconstructing the ontology that was used to create the data set.

  Note that this does not imply that finding regularities in data cannot be used to evolve ontologies; we are claiming that the ontology for each data set comes logically before that data set. This frequently occurs in research when a data set may record the output of a sensor where it is unknown what the senor actually measures (i.e., the meaning of the sensor report is unknown). The initial ontology will then specify the meaning is just a real number, perhaps with some range and precision. Later ontologies may give the output a name.

Some people have argued that uncertainty should be explicitly represented in an ontology because of the inherent uncertainty in data (Pool et al., 2005; da Costa et al., 2005; Laskey et al., 2007). While we believe that it is essential to model the uncertainty in data, we don't believe actual probability values should be in the ontology[4]. The main reason is the ontology is logically prior to the data, but the models of uncertainty in the data are logically posterior to the data: it is only by seeing (some of) the data, that we can estimate the uncertainty (i.e., we want the uncertainty to reflect the posterior distribution after we have seen some data). Because the probabilities are posterior to the data, they should change

---

[4] An ontology will contain the vocabulary to express probability distributions. We need the vocabulary to express continuous and discrete conditional probability distributions, e.g., using PR-OWL (da Costa et al., 2005). The ontologies need to be rich enough to express what scientists want to state in theories.

as data comes in, and so should not be part of the stable foundation of the data that an ontology needs to be. Another way to think about it is that the ontologies define the vocabulary; they do not make empirical claims. Saying that a granite is an igneous, felsic, course rock is not an empirical claim, it just defines what a granite is. Theories make empirical (testable) claims. A specification of a probability is an empirical claim, even if the probability is theory-based (e.g., based on symmetries) and not data-summaries. Thus the probability should not be in the ontology. Note that our claim that probabilities do not belong in definitions is an empirical claim, and is not part of the definition of semantic science.

2.3   Theories

We would argue that theories are best described in terms of probabilities (Polya, 1954) for two main reasons:

- Probabilities summarize the empirical content of data. In particular, we want predictions that can be evaluated against the empirical evidence, and so can be optimized with respect to the evidence. Probability distributions optimize most of the common evaluation criteria, and other predictions (such as the mean or the mode) can be derived from the probability distribution.
- Probabilities, together with utilities, are what is needed to make decisions.

Like data, theories need to adhere to ontologies. There are a number of reasons:

- Theories make predictions on data that adhere to an ontology. To allow semantic interoperability between the data and the theories, they should adhere to a common ontology.
- People should be allowed to disagree about how the world works without disagreeing about the meaning of the terms. If two people have different theories, they should first agree on the terminology (for otherwise they would not know they have a disagreement)—this forms the ontology—and then they should give their theories. Their theories can then be compared to determine what their disagreement is. It is by creating these disagreements, and testing them on data, that science progresses.

Theories can expand the ontology by hypothesizing unobserved objects or properties (hidden variables) that help explain the observations. By expanding the ontology, other theories can refer to the theoretical constructs, and they could appear in data. For example, a theory could postulate that the data is better explained by having a new form of cancer; other theories could refer to this type of cancer and this new type of cancer could even be recorded in data. In this way the theories and the vocabulary can evolve as science advances.

Semantic interoperability can only be achieved by adhering to common ontologies. A community needs to agree on an ontology to make sure they use the same terminology for the same things. However, a community need not, and we

argue should not, agree on the probabilities, as people may have different prior knowledge and have access to different data, and the probabilities should change as more data comes in.

To make a prediction, we usually use many theories. Theories that individuals produce are typically very narrow, only making predictions in very narrow cases. The theories that are put together to make a predictions form a theory *ensemble*. We judge individual theories by how well they fit into ensembles. An example of a theory ensemble is "when the speed of the objects involved is less than 70% of the speed of light, use Newtonian mechanics, otherwise use Einstein's theory of relativity". This ensemble is another theory that may work better than either of the composite theories in practice. Producing such ensembles is of a different sort than producing the base theories, and so should be separated. Rather than dismissing these theories as trivial, they form the basis of prediction for new cases. Virtually all predictions in complex cases will rely on theory ensembles.

The structure of probabilistic theories does not necessarily follow the structure of the ontology. For example, an ontology of lung cancer should specify what lung cancer is, but whether someone will have lung cancer depends on many factors of the particular case and not just on other parts of ontologies (e.g., whether they have other cancers and their work history that includes when they worked in bars that allowed smoking). As another example, the probability that a room will be used as a living room depends not just on properties of that room, but on the properties of other rooms in an apartment.

There are major challenges in building probabilistic theories using ontologies based on languages such as OWL. The main challenge is that OWL sees the world in terms of individuals, classes and properties, while probability theory is in terms of random variables. Section 4.2 discusses how to construct random variables from ontologies.

3   Pragmatic considerations

The MineMatch and HazardMatch systems we have been developing have multiple instances that describe entities and their properties at particular locations on Earth, and models (theories) that make predictions about these locations. The systems are used in various modes:

– In instance-to-models matching, one instance is compared to multiple models. Finding the most likely models for the instance can be used to determine what is the most likely mineral to occur at a location or what types of landslides are predicted to occur at a particular place. In both of these cases, the instance is a place whose description is compared to the models.
– In model-to-instances matching, one model is compared to multiple instances. This can be used to find the location(s) that are most likely to have landslides or contain particular minerals.
– In instance-to-instances matching, one instance is compared to multiple instances to find which other instances are most like this instance.

– In model-to-models matching, one model is compared to multiple models to find which other models are most like this model.

These applications have a number of features that we believe will be shared by many scientific disciplines:

– The instances are heterogeneous, described at various levels of abstraction (using more general or less general terms) and detail (described in terms of parts and sub-parts or not). Similarly, the models use various levels of abstraction and detail. Sometimes the distinctions that are in the instance descriptions are not required by the models, and sometimes the instance descriptions do not make distinctions that are needed by the models.
– The experts often do not publish probabilities in their models, and are reluctant to have probabilities in the system. There are a number of reasons for this. First, they may have very few data points for any model, so that the probabilities will not be based on anything meaningful. Second, the people who want to make decisions (those who want to decide whether to try to mine an area profitably, or insurance companies that decide on insurance premiums) will want to use their own prior probabilities, and may take into account more information than is used in the system.
– The problem domains are afflicted by combinatorial complexity; there many possible model combinations, and very large data collections for assessment. It is difficult to find those few areas that are most likely to contain ore-grade minerals or be susceptible to landslides, and to provide explanations that can be used for further analysis.
– The models are "positive"; there are models of where to find a particular mineral, but people do not publish models of where the mineral is absent. Similarly for landslides; there are models of where particular types of landslides are likely to occur, but not models of where landslides are unlikely to occur.
– The models are neither covering, disjoint nor independent. Often the models are variants of each other. Starting from one model, people produce variants of that model to suit their own purpose. A model does not include all of the cases where the phenomenon it is modelling may occur; it only about a specific context.

4   Foundations of Probabilistic Theories

In this section, we describe the logical and probabilistic foundations for building theories, and relate them to pragmatic choices that we have used in our fielded systems.

4.1   Role of Models in Decision Making

The Bayesian view of using models for decision making is that we would like to make a probabilistic prediction of x for a new case based on a description

d of that case. Thus we want $P(x|d)$. The role of the models is to provide a framework for this prediction.

In terms of probabilities, we can use models as intermediaries:

$$P(x|d) = \sum_{m \in Models} P(x|m \wedge d)P(m|d)$$

where $Models$ is a set of mutually exclusive and covering hypotheses. Thus, for each model, we need to decide what it predicts, and how likely it is based on the description, d, of the current case. Typically models are rich enough to convey the information about the rest of the description, and so we assume $P(x|m \wedge d) = P(x|m)$.

In Bayesian modelling, we try to determine what features best predict (in unseen data) the phenomenon of interest, and then build probabilistic models in terms of these features.

Typically, we do not have $P(m|d)$ which specifies how likely the model is given the description, but instead have predictions of the model, i.e., $P(d|m)$. These two quantities are related by Bayes' theorem:

$$P(m|d) = \frac{P(d|m)P(m)}{P(d)}$$

That is, we often have causal or consequential knowledge and want to do evidential reasoning. For example, we model the symptoms of chicken pox with $P(fever|ch\_pox)$ but want $P(ch\_pox|fever)$. These are related by Bayes' theorem:

$$P(ch\_pox|fever) = \frac{P(fever|ch\_pox) \times P(ch\_pox)}{P(fever)}$$

The reason that we want to store causal or consequential knowledge is that it is more stable to changing contexts. You would expect the symptoms of chicken pox to be stable; they would be the same whether the patient was at home, in a school or in a hospital. However, the probability that someone with a fever has chicken pox would be different in these three contexts, as the prevalence of fever and chicken pox is different in these three contexts.

This has an impact on how diagnostic a feature is. Suppose $fever$ and $spots$ are common given chicken pox, e.g., $P(fever|ch\_pox) = 0.9$, $P(spots|ch\_pox) = 0.9$. Suppose $fever$ has many causes and $spots$ has few. Then spots is more diagnostic of $chicken\ pox$, i.e., $P(ch\_pox|spots) > P(ch\_pox|fever)$, as $P(fever) > P(spots)$.

Note also that the probabilities needed for the prediction, namely $P(x|m)$ are of the same form as $P(d|m)$—they all specify what the model predicts. Rather than making a model to be for a particular feature, a model makes predictions about all of its features.

4.2 Probabilities, Ontologies and Existence

There seems to be a fundamental mismatch between the random variable formalization of probability theory and the formalization of modern ontologies in terms

of individuals, classes and properties. Probabilistic models typically assume we know what random variables exist at modelling time, but what individuals exists is often unknown at modelling time. Interestingly, a large body of research on Bayesian modelling (e.g., Bayesian networks) and modern research into ontologies both have their roots in the expert systems of the 1970's and 1980's (Henrion et al., 1991). Both fields have advanced our understanding of reasoning, and part of our research is to bring these together.

We can reconcile these views by having properties of individuals correspond to random variables. This complicates the probabilistic modelling as the individuals typically only become known at run-time, and so the random variables are unknown at modelling time. This has spurred a body of research in first-order probabilistic models or relational probabilistic models (e.g., Poole (1993), Getoor and Taskar (2007), Kersting and De Raedt (2007), Laskey (2008), Lukasiewicz (2008)). It is even possible to be unsure about the existence of an individual, and so unsure about the existence of a random variable (Poole, 2007).

When dealing with probabilities and individuals we need to deal with three types of uncertainty:

- the probability of existence (Poole, 2007) — the probability that an individual that fits a description actually exists.
- the probability distribution over the types of an individual. This is complicated when there are complex interrelations between classes that can be the types of the individuals.
- the probability of property values. Functional properties give a random variable for each individual with a non-zero probability of being in the class that is the domain of the property. Non-functional properties have a Boolean random variable for each value in the range and each individual with a non-zero probability of being in the domain of the property.

Aristotelian definitions, where a class is defined in terms of its immediate superclass and differentia, provide a way to reduce the second case to the third case. The differentia are described in terms of property values with appropriate domains. By having a probability distribution over the values of the properties (perhaps conditioned on other variable assignments), we can induce a probability distribution over the classes. Note that Aristotelian definitions are general: any class hierarchy can be represented by Aristotelian definitions by introducing new properties.

For example, a granite can be defined as a rock with the property *genesis* having value *igneous*, property *composition* having value *felsic*, and *texture* is *coarse*. By having a probability distribution over the values of *genesis*, a probability distribution over the value of *composition*, and a probability distribution over the values of *texture*, we can determine the prior probability that a rock is a granite.

Note that the probabilistic formulation is complicated by existence prerequisites: only individuals that exist have properties, and only individuals in the class that is domain of a property can have values for that property.

4.3 Bayesian modelling meets pragmatism

Bayesian modelling of scientific reasoning seems like the appropriate formulation of the role of theories or models in science. However, the pragmatic considerations discussed above lead us to not adopt it directly, although it remains the gold standard. The theories (or models) in our fielded systems are based on qualitative probabilistic matching (Smyth and Poole, 2004; Poole and Smyth, 2005; Lukasiewicz and Schellhase, 2007), with the following properties:

- Rather than using probabilities that experts do not want to give, and cannot judge the output from, we use qualitative probabilities, using a 5-point scale (*always, usually, sometimes, rarely, never*) that is derived from the terminology used in published papers. These qualitative probabilities act like log-probabilities, where the values add rather than multiply (Pearl, 1989; Darwiche and Goldszmidt, 1994).
- The models need to be fleshed out for each instance. Models refer to multiple individuals, but they do not refer to the named individuals in the instances. Models specify roles that can be filled by the instance individuals. The predictions of the model for an instance can only be determined given a role assignment that specifies which instance individuals fill the roles in the model.
- Rather than averaging over all possibilities and role assignments, we choose the most likely ones.
- We allow for diverse data about instances and models at multiple levels of abstraction and detail. We also require prior probabilities of the descriptions; we do not assume that we can get the probability of a description from the set of models (as we could if the models were exclusive and covering).
- The explanations for the answers are as important as the answers themselves.

5 Conclusions

This paper has presented the big picture of what we see as semantic science as well as the pragmatic considerations that have gone into our fielded systems that are a first try at realizing our vision. This view of semantic science is meant to complement other views that provide ontologically-based views of data (McGuinness et al., 2007) and ontology-based services (De Roure et al., 2005).

There are many challenges in building the semantic science vision, including how to construct theories, how to determine what theories are useful in making predictions in a particular case, and in finding the data about which a theory makes predictions. The growing interest in scientific ontologies, the desire for scientists (and their funders) to make their data and theories as widely used as possible, and the desire for users to have the best predictions, indicates that this semantic science vision should succeed.

Bibliography

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., and Angel, S. (1977). *A Pattern Language*. Oxford University Press, New York.

Aristotle (350 B.C.). *Categories*. Translated by E. M. Edghill, http://www.classicallibrary.org/Aristotle/categories/.

Berg, J. (1982). Aristotle's theory of definition. In *ATTI del Convegno Internazionale di Storia della Logica*, pp. 19–30. San Gimignano. http://ontology.buffalo.edu/bio/berg.pdf.

Berners-Lee, T. and Fischetti, M. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web, by its inventor*. Harper Collins, San Francisco, CA.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, pp. 28–37.

da Costa, P.C.G., Laskey, K.B., and Laskey, K.J. (2005). PR-OWL: A Bayesian ontology language for the semantic web. In *Proceedings of the ISWC Workshop on Uncertainty Reasoning for the Semantic Web*. Galway, Ireland. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-173/.

Darwiche, A. and Goldszmidt, M. (1994). On the relation between kappa calculus and probabilistic reasoning. In *UAI-94*, pp. 145–153.

De Roure, D., Jennings, N.R., and Shadbolt, N.R. (2005). The semantic grid: Past, present and future. *Procedings of the IEEE*, 93(3): 669–681. http://www.semanticgrid.org/documents/semgrid2004/semgrid2004.pdf.

Fodor, J.A. (1975). *The Language of Thought*. Harvard University Press, Cambridge, Massachusetts.

Getoor, L. and Taskar, B. (Eds.) (2007). *Introduction to Statistical Relational Learning*. MIT Press, Cambridge, MA.

Gillespie, M.R. and Styles, M.T. (1999). BGS rock classification scheme, volume 1: Classification of igneous rocks. Research Report (2nd edition) RR 99-06, British Geological Survey. http://www.bgs.ac.uk/bgsrcs/.

Henrion, M., Breese, J., and Horvitz, E. (1991). Decision analysis and expert systems. *AI Magazine*, 12(4): 61–94.

Kersting, K. and De Raedt, L. (2007). Bayesian logic programming: Theory and tool. In L. Getoor and B. Taskar (Eds.), *An Introduction to Statistical Relational Learning*. MIT Press.

Laskey, K.B., Wright, E.J., and da Costa, P.C.G. (2007). Envisioning uncertainty in geospatial information. In *UAI Applications Workshop 2007 (The 5th Bayesian Modeling Applications Workshop*. http://ite.gmu.edu/~klaskey/uai07workshop/AppWorkshopProceedings/UAIAppWorkshop/paper3.pdf.

Laskey, K.B. (2008). MEBN: A language for first-order Bayesian knowledge bases. *Artificial Intelligence*, 172(2-3). doi:doi:10.1016/j.artint.2007.09.006. http://www.sciencedirect.com/science/article/B6TYF-4PTMXXP-1/2/ce6bcf1c5a5fecfd805501056e9b62a1.

Lukasiewicz, T. (2008). Expressive probabilistic description logics. *Artificial Intelligence*, 172(6-7): 852–883.

Lukasiewicz, T. and Schellhase, J. (2007). Variable-strength conditional preferences for ranking objects in ontologies. *Journal Web Semantics*, 5(3): 180–194.

McGuinness, D., Fox, P., Cinquini, L., West, P., Garcia, J., Benedict, J.L., and Middleton, D. (2007). The virtual solar-terrestrial observatory: A deployed semantic web application case study for scientific research. In *Proceedings of the Nineteenth Conference on Innovative Applications of Artificial Intelligence (IAAI-07)*. Vancouver, BC, Canada. http://www.ksl.stanford.edu/KSL_Abstracts/KSL-07-01.html.

McGuinness, D.L. and van Harmelen, F. (2004). OWL web ontology language overview. W3C Recommendation 10 February 2004, W3C. http://www.w3.org/TR/owl-features/.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report SIDL-WP-1999-0120, Stanford InfoLab. http://dbpubs.stanford.edu/pub/1999-66.

Pearl, J. (1989). Probabilistic semantics for nonmonotonic reasoning: A survey. In R.J. Brachman, H.J. Levesque, and R. Reiter (Eds.), *KR-89*, pp. 505–516. Toronto.

Pinker, S. (1994). *The Language Instinct*. Harper Collins, New York.

Polya, G. (1954). *Mathematics and Plausible Reasoning*, volume II Patterns of Plausible Inference. Princeton University Press.

Pool, M., Fung, F., Cannon, S., and Aikin, J. (2005). Is it worth a hoot? Qualms about OWL for uncertainty reasoning. In *Proceedings of the ISWC Workshop on Uncertainty Reasoning for the Semantic Web*. http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-173/.

Poole, D. (1993). Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64(1): 81–129.

Poole, D. (2007). Logical generative models for probabilistic reasoning about existence, roles and identity. In *22nd AAAI Conference on AI (AAAI-07)*. http://www.cs.ubc.ca/spider/poole/papers/AAAI07-Poole.pdf.

Poole, D. and Smyth, C. (2005). Type uncertainty in ontologically-grounded qualitative probabilistic matching. In *Proc. European Conference on Symbolic and Qualitative Reasoning About Uncertainty (ECSQARU)*, LNAI 3571, pp. 763–774. Springer-Verlag, Barcelona, Spain. http://www.cs.ubc.ca/spider/poole/papers/Poole-Smyth-ecsqaru2005.pdf.

Poole, D., Smyth, C., and Sharma, R. (2008). Semantic science and machine-accessible scientific theories. In *AAAI Spring Symposium on Semantic Science Knowledge Integration*. Stanford, CA.

Popper, K. (1959). *The Logic of Scientific Discovery*. Basic Books, New York, NY.

Sapir, E. (1929). The status of linguistics as a science. *Language*, 5(209).

Smith, B. (2003a). The logic of biological classification and the foundations of biomedical ontology. In D. Westerståhl (Ed.), *Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science*. Elsevier-North-Holland, Oviedo, Spain. http://ontology.buffalo.edu/bio/logic_of_classes.pdf.

Smith, B. (2003b). Ontology. In L. Floridi (Ed.), *Blackwell Guide to the Philosophy of Computing and Information*, pp. 155—166. Oxford: Blackwell. http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf.

Smyth, C. and Poole, D. (2004). Qualitative probabilistic matching with hierarchical descriptions. In *KR-04*. Whistler, BC, Canada. http://www.cs.ubc.ca/spider/poole/papers/KR04SmythC.pdf.

Smyth, C., Poole, D., and Sharma, R. (2007). Semantic e-science and geology. In *AAAI-07 Semantic e-Science workshop*. http://www.cs.ubc.ca/spider/poole/papers/SmythPooleSharmaSemSci2007.pdf.

Sowa, J.F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, CA.

Struik, L., Quat, M., Davenport, P., and Okulitch, A. (2002). A preliminary scheme for multihierarchical rock classification for use with thematic computer-based query systems. Current Research 2002-D10, Geological Survey of Canada. http://daks.ucdavis.edu/~ludaesch/289F-SQ06/handouts/GSC_D10_2002.pdf.

Tang, H. and Ng, J.H.K. (2006). Googling for a diagnosis–use of google as a diagnostic aid: internet based study. *BMJ*, doi:10.1136/bmj.39003.640567.AE.

Whorf, B.L. (1940). Science and linguistics. *Technology Review*, 42(6): 229–31, 247–8.