Uncertainty Reasoning for the World Wide Web: Report on the URW3-XG Incubator Group

Kenneth J. Laskey[#]
MITRE Corporation, M/S H305
7515 Colshire Drive
McLean, VA 22102-7508 USA
klaskey@mitre.org

Kathryn Blackmond Laskey
Department of Systems Engineering
and Operations Research
George Mason University
4400 University Drive
Fairfax, VA 22030-4444 USA
klaskey@gmu.edu

Abstract. The Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG) was chartered as a means to explore and better define the challenges of reasoning with and representing uncertain information in the context of the World Wide Web. The objectives of the URW3-XG were: (1) To identify and describe situations on the scale of the World Wide Web for which uncertainty reasoning would significantly increase the potential for extracting useful information; and (2) To identify methodologies that can be applied to these situations and the fundamentals of a standardized representation that could serve as the basis for information exchange necessary for these methodologies to be effectively used. This paper describes the activities undertaken by the URW3-XG, the recommendations produced by the group, and next steps required to carry forward the work begun by the group.

1 Introduction

The Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG) was proposed [1] during the 2006 Uncertainty Reasoning for the Semantic Web workshop [2] as a means to explore and better define the challenges of reasoning with and representing uncertain information in the context of the Semantic Web. In addition, it was intended to identify situations in which the combination of semantics and uncertainty could further the Web Services vision of quickly and efficiently composing services and data resources to address the needs of users in an ever-changing world.

The 2006 workshop included a Use Case Challenge [3] to generate an initial collection of use cases and to gauge the interest of the workshop participants in continu-

[#] The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author.

ing as a W3C XG or through some other collaboration venue. The Use Case Challenge generated a lively interchange of ideas, and the participants overwhelmingly agreed to create the XG to continue the work.

2 W3C Incubator (XG) Process

As noted in [1], the World Wide Web Consortium (W3C) [4] created the Incubator process [5] to provide a formal, yet flexible venue to better understand Web-related challenges and their potential solutions. It encourages a public exploration of issues and potential solutions before the solutions are mature enough for standardization. It also provides a "head start" if the Incubator experimental group, the XG, is able to adequately formulate principles and techniques that gain consensus in the wider community.

The URW3-XG was in operation [6] from 5 March 2007 until its final report [7] was published by the W3C on 22 April 2008. The group included 25 participants from North and South America, Europe, and Australia. Participants came from a range of time zones spanning 18 hours. The group conducted over 20 telecons, with an average duration between 90 and 120 minutes. In addition, face-to-face meetings of subsets of the XG were held at the 5th ISWC (Busan - Korea) and the SUM conference in College Park, Maryland USA. The telecons were supported by the W3C resources (e.g. telecon bridge, IRC, RSSAgent, etc). Meeting results and action items were catalogued in online Minutes [6].

The objectives of the URW3-XG were twofold:

- To identify and describe situations on the scale of the World Wide Web for which uncertainty reasoning would significantly increase the potential for extracting useful information; and,
- To identify methodologies that can be applied to these situations and the fundamentals of a standardized representation that could serve as the basis for information exchange necessary for these methodologies to be effectively used.

3 Results of the URW3-XG Effort

The Final Report [7] was the major deliverable of the URW3-XG. It describes the work done by the XG, identifies elements of uncertainty that need to be represented to support reasoning under uncertainty for the World Wide Web, and provides an overview of the applicability to the World Wide Web of various uncertainty reasoning techniques (in particular, probability theory, fuzzy logic, and belief functions) and the information that needs to be represented for effective uncertainty reasoning to be possible. The report concludes with a discussion on the benefits of standardization of uncertainty representation to the World Wide Web and the Semantic Web and provides a series of recommendations for continued work. The report also includes a Reference List of work relevant to the challenge of developing standardized representations for uncertainty and exploiting them in Web-based services and applications.

A major part of the work was development of a set of use cases illustrating conditions under which uncertainty reasoning is important. Another major effort was the development of an Uncertainty Ontology that was used to categorize uncertainty found in the use cases. These products are described briefly in the following sections. Section 4 then details the conclusions and recommendations from the report.

3.1 The Uncertainty Ontology

The Uncertainty Ontology is a simple ontology developed to demonstrate some basic functionality of exchanging uncertain information. It was used to classify the use cases developed by the URW3-XG with the intent of obtaining a relatively complete coverage of the functionalities related to uncertainty reasoning about information available on the World Wide Web. The top level of the ontology is shown in Figure 1. According to the ontology, uncertainty is associated with sentences that make assertions about the world, and are asserted by agents (human or computer). The uncertainty derivation may be objective (via a formal, repeatable process) or subjective (judgment or guess). Uncertainty type includes ambiguity, empirical uncertainty, randomness, vagueness, inconsistency and incompleteness. Uncertainty models include probability, fuzzy logic, belief functions, rough sets, and other mathematical models for reasoning under uncertainty. Uncertainty nature includes aleatory (chance; inherent in the phenomenon) or epistemic (belief; due to limited knowledge of the agent).

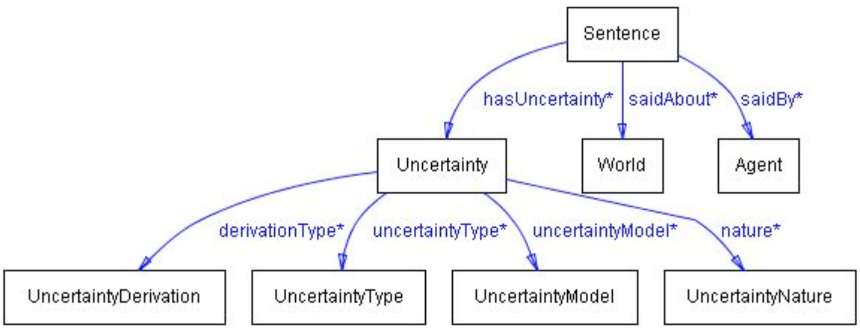

Figure 1 Top level of URW3-XG Uncertainty Ontology

While this ontology served the purpose of focusing discussion of the use cases, allowing use case developers to show examples of annotation of uncertainty, the ontology was only meant to provide a starting point to be refined through an iterative process. Further development of a more complete ontology for annotating uncertainty is one of the XG's recommendations.

3.2 The URW3-XG Uncertainty Use Cases

Building on the work started during the Use Case Challenge, the URW3-XG developed 16 use cases to identify how the representation of uncertainty would help to address issues in Web reasoning that cannot be properly represented with current deterministic approaches. The use cases were developed for the most part using a common template. Occurrences of uncertainty in the use case descriptions were annotated with information from the Uncertainty Ontology. One use case, entitled Buying Speakers, is shown in the Appendix.

The analysis of the use cases indicated that a representation of uncertainty would be required to represent both uncertainty inherent in the data and uncertainty related to the processing of data and the delivery of processing results. This will be discussed further in section 4.

4 Key Conclusions and Recommendations

In automated data processing, we often face situations where Boolean truth-values are unknown, unknowable, or inapplicable. This is true for a wide variety of data and information processing applications, and therefore it should be no surprise that the methodologies considered by the XG are popular in contexts other than the Web. The use cases considered by the XG concerned reasoning challenges specific to the Web, such as discovery of Web Services, order processing via Web Services, and the like. The XG's work confirmed the hypothesis that a unified model for uncertainty annotation of Web resources would provide value for deductive engines, and this could be further facilitated by an ontology characterizing the types and sources of uncertainty.

The work with the Uncertainty Ontology suggested that a finer grained extension may be useful. Such an extension could provide a means to visualize a possible evolution of an upper level Uncertainty Ontology. The conclusions go on to focus especially on finer classification of Machine Agents and uncertainty caused by lack of knowledge of a machine agent.

With respect to the kinds of uncertainty observed in the use cases, it was noted that uncertainty may be an inherent part of the data or may be related to the processing that produces results. In the first case, the standardization should provide a single syntactical system so that people can identify and process this information quickly.

For example, one may want to be able to communicate information that Study X shows that people with property Y have an Z% increased likelihood of disease W. The ability to communicate such information using a common interchange syntax could be extremely useful in a number of web-based applications. Such characterizations of data uncertainty may require something like uncertain extensions to OWL (i.e., probabilistic, fuzzy, belief function, random set, rough set, and hybrid uncertain extensions to OWL).

The second kind of uncertainty involves reasoning on the part of the tools used to access and share web information. For example, if a web service uses uncertainty reasoning to find and rank hotel rooms, the need would be to represent meta-information about the reasoning models and assumptions. This could facilitate the development of trust models, or allow the identification of compatible web services to increase the likelihood that the results are consistent with the user preferences. Here the representation would include determining how to represent the meta-information on processing and deciding how detailed the meta-information would need to be and where it would reside.

The deliberations and conclusions of the URW3-XG led to the following recommendations:

- A principled means for expressing uncertainty will increase the usefulness of Web-based information and a standard way of representing that information should be developed.
- Different use cases appear to lend themselves to different uncertainty formalisms, indicating the standard representation should provide a means to unambiguously identify the formalism providing the context for assigning other uncertainty characteristics and values.
- Different uncertainty formalisms assign values to properties specifically related to the underlying meaning and processing of these values, and the representation should support defining different standard properties for each formalism without requiring changes to the representation itself.
- Sample representations for the most useful formalisms should be developed both as exemplars and for their immediate use, with the ability to expand beyond the initial exemplars as circumstances might indicate to be prudent.
- Given that uncertainty can be present anywhere, the representation should support associating uncertainty with any property or value expressible across the Web.

An open question that remains when considering a standard uncertainty representation is whether existing languages (e.g. OWL, RDFS, RIF) are sufficiently expressive to support the necessary annotations. If so, the development of such annotations might merely require work on a more complete uncertainty ontology and possibly rules; otherwise, the expressiveness of existing languages might need to be extended. As an example of the latter, it might be advisable to develop a probabilistic extension to OWL or a Fuzzy-OWL format or profiles associated with the type of uncertainty to be represented. Further work is required to investigate the adequacy of the existing languages against the compiled use cases.

The means to associate the uncertainty representation with its subject was also beyond the scope of the URW3-XG. The conclusions noted that a mechanism similar to that specified under Semantic Annotations for WSDL and XML Schema (SAWSDL) [8].

5 Considerations for Next Steps

The work of the URW3-XG provided an important beginning for characterizing the range of uncertainty that affects reasoning on the scale of the World Wide Web, and the issues to be considered in designing a standard representation of that uncertainty. However, the work to date likely falls short of what would be needed to charter an effort to develop that representation. Additional work needed includes the following:

- The conclusions note the value of the Uncertainty Ontology developed thus far, but it also notes the value of further work to extend the ontology;
- A representation is needed for uncertainty models but it was beyond the scope of the current effort to decide whether extensions to existing Semantic Web languages (e.g. OWL, RDFS, RIF) will be sufficient or whether new representation standards will be needed;
- As SAWSDL provides a mechanism to associate semantics with certain Web resources, it might also provide a useful model for associating a standard representation of uncertainty information, but the feasibility of such use has not been adequately considered.

The question to be answered is what future venue should be pursued to tackle these issues and others that may become evident. There are several nonexclusive possibilities, among which are

- Continue with the URSW workshop series, using it as a forum to discuss advances in theory and practice;
- Approach other communities, such as those dealing with health care and life sciences, and form a wider collaboration to both continue the research aspects and to provide concrete problems against which to develop solutions;
- Develop a charter for and establish a new XG to work the items recommended by the URW3-XG;
- Investigate funding opportunities to formalize a dedicated effort to pursue the issues and develop implementable solutions and tools in a reasonable time frame.

This paper provides a summary of work to date. As the discussions of the attendees at the 2nd URSW workshop provided the basis for the URW3-XG work, so the 4th

URSW workshop provides the opportunity to discuss these and possibly other options and assess the consensus of the community for its next steps.

References

[1] Laskey, K. J.; Laskey, K. B.; and Costa, P. C. G. (2006) A Proposal for a W3C XG on Uncertainty Reasoning for the World Wide Web. Proceedings of the second workshop on Uncertainty Reasoning for the Semantic Web (URSW 2006), held at the Fifth International Semantic Web Conference (ISWC 2006), 5-9 November 2006, Athens, Georgia, USA. Available at http://c4i.gmu.edu/ursw/2006/files/papers/URSW06_P5_LaskeyCostaLaskey.pdf.

[2] Second workshop on Uncertainty Reasoning for the Semantic Web (URSW 2006), held at the Fifth International Semantic Web Conference (ISWC 2006), 5 November 2006, Athens, Georgia, USA. http://c4i.gmu.edu/ursw/2006/

[3] Laskey, K. J. (2006) Use Case Challenge. Proceedings of the second workshop on Uncertainty Reasoning for the Semantic Web (URSW 2006), held at the Fifth International Semantic Web Conference (ISWC 2006), 5-9 November 2006, Athens, Georgia, USA. Available at http://c4i.gmu.edu/ursw/2006/files/talks/URSW06_UseCaseChallenge.pdf

[4] World Wide Web Consortium (W3C), http://www.w3.org/

[5] W3C Incubator Activity > About XGs, http://www.w3.org/2005/Incubator/about.html

[6] Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG), http://www.w3.org/2005/Incubator/urw3/

[7] URW3-XG Final Report, 31 March 2008. Available at http://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/

[8] Semantic Annotations for WSDL and XML Schema. W3C Recommendation, 28 August 2007. Available at http://www.w3.org/2002/ws/sawsdl/spec/.

[9] Agarwal, S.; and Lamparter, S. (2005) sMART - A Semantic Matchmaking Portal for Electronic Markets. Proceedings of the 7th International IEEE Conference on E-Commerce Technology. Munich, Germany, 2005.

Appendix – Buying Speakers Use Case

1 - Purpose/Goals

Customer needs to make a decision on (1) whether to go to a store today or wait until tomorrow to buy speakers, (2) which speakers to buy and (3) at which store. Customer is interested in two speaker features: wattage and price. Customer has a valuation formula that combines the likelihood of availability of speakers on a particular day in a particular store, as well as the two features. The features of wattage and price are fuzzy. Optionally, Customer gets the formulas from CustomerService, a Web based service that collects information about products, stores, statistics, evaluations.

2 - Assumptions/Preconditions

- Customer either relies on the definitions provided by CustomerService or is knowledgeable in both probability and fuzzy sets.
- Stores provide information to CustomerService. CustomerService keeps information on both probabilistic models and fuzzy models.
- Customer has the capability of either obtaining or defining a combination function for combining probabilistic information with fuzzy.

3 - Required Resources

- Data collected by CustomerService on the availability of items, which in turn depends on restocking and rate of selling.
- Ontology of uncertainty that covers both probability and fuzziness.

4 - Successful End

Customer gets necessary information about the availability and types of speakers from stores. This information is sufficient for customer to compute the required metric.

5 - Failed End

Customer does not get necessary information and thus needs to go to multiple stores, wasting in this way a lot of time.

6 - Main Scenario

1. Customer formulates query about availability of speakers in the stores within some radius.
2. Customer sends the query to the CustomerService.
3. CustomerService replies with information about the availability of speakers. CustomerService cannot say for sure whether a given type of speaker will be available in a store tomorrow or not. It all depends on delivery and rate of sell. Thus CustomerService provides the customer only with probabilistic information.
4. Since part of the query involves requests that cannot be answered in crisp terms (vagueness), CustomerService annotates its replies with fuzzy numbers.
5. CustomerService uses the uncertainty annotated information to compute a metric.
6. Customer uses the resulting values of the metric for particular stores and for particular types of speaker to decide whether to buy speakers, what type and which store.

7. Additional background information or references: This use case was inspired by Agarwal and Lamparter [9].
8. General Issues and Relevance to Uncertainty:
    1. There is known probability distribution on the availability of particular speaker type in particular stores on a particular day in the future. Say there are two stores (not too close to each other) and the probability that speakers of type X will be available in stores A and B tomorrow are Pr(X, A)=0.4 and Pr(X, B)=0.6. The probabilities for all types of speakers are represented in the same way.
        - The uncertainty annotation process (UncAnn) was used.
        - The agent issues a query (a sentence): Sentence. It is a complex sentence consisting of three basic sentences. One related to the availability, one to the wattage and one to the price of speakers.
        - Each of these sub-sentences will have uncertainty Uncertainty associated with it.
        - The uncertainty type related to the availability of particular speaker type in the stores is of type UncAnn - UncertaintyType: Empirical.
        - The uncertainty nature is UncAnn - UncertaintyNature: Aleatory.
        - The uncertainty model is UncAnn - UncertaintyModel: Probability.
    2. The customer has (or obtains from CustomerService) definitions of features of wattage and price in terms of fuzzy membership functions. For wattage, Customer has three such functions: weak, medium and strong. These are of "trapezoid shaped" membership functions. Similarly, for price Customer has three such membership functions: cheap, reasonable and expensive.
        - The uncertainty type related to the features of wattage and price is of type UncAnn - UncertaintyType: Vagueness.
        - The uncertainty nature is UncAnn - UncertaintyNature: Epistemic.
        - The uncertainty model is UncAnn - UncertaintyModel: FuzzySets.
    3. The valuation has three possible outcomes, all are expressed as fuzzy membership functions: bad, fair, good and super.
    4. Customer knows the probabilistic information, since the probabilities are provided by CustomerService. CustomerService uses the Uncertainty Ontology for this purpose.
    5. Customer has (or selects) fuzzy definitions of the features of wattage and price. Again, the six membership functions that define these features are annotated with the Uncertainty Ontology.
    6. Customer has (or uses one suggested by CustomerService) a combination function that computes the decision, d, based upon those types of input. This function can be modified by each customer, however the stores need to give input to CustomerService - the probabilities and the (crisp) values of wattage and price for their products. The features are fuzzified by the customer's client software. Customer uses the Uncertainty Ontology to annotate the fuzziness of particular preferences.