Discussion on Uncertainty Ontology for Annotation and Reasoning (a position paper)

J. Dedek, A. Eckhardt, L. Galambos, P. Vojtas

Charles University in Prague, Department of Software Engineering
{jan.dedek,alan.eckhardt,leo.galambos,peter.vojtas}@mff.cuni.cz

In this position paper we discuss the what, who, when, where, why and how of uncertain reasoning based on achievements of URW3XG [2], our experiments and some future plans.

What and *Why* – improving semantic web practice through uncertain reasoning. This vision is described in the URW3XG charter (see [2]), especially the objective is "*to identify and describe situations* […] *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.*" A crucial point in this is uncertainty annotation of web (extending W3C standards [3]).

Who and *When* - will create, maintain and use this annotation. Will this annotation be done by a human creator using an annotation supporting tool for web page creation? Or will it be done by a third party annotation? For this, we will discuss a refinement of URW3XG use cases. Possible use of this enriched web will be for humans and services.

Where - will be this annotations stored. Our proposal is based on the web crawler Egothor repository [4] (we have crawled data in size of several TB from .cz domain) and an additional semantic repository build on the top using data pile technology [5].

How – to semantically enrich information and how to measure success and/or progress of such enrichment. This problem consists of two parts, namely, a data mining task and an ontology modeling task. Third party annotation of great size can be done only in an automated way and it should be done according to an ontology.

Our annotation ontology grows out of URW3XG uncertainty ontology and extends some features needed for annotation. Below we

show a part of our annotation ontology in Fig. 1. We start here from an assumption that a part of annotation will be done by a web information extraction and that this is the main source of uncertainty.
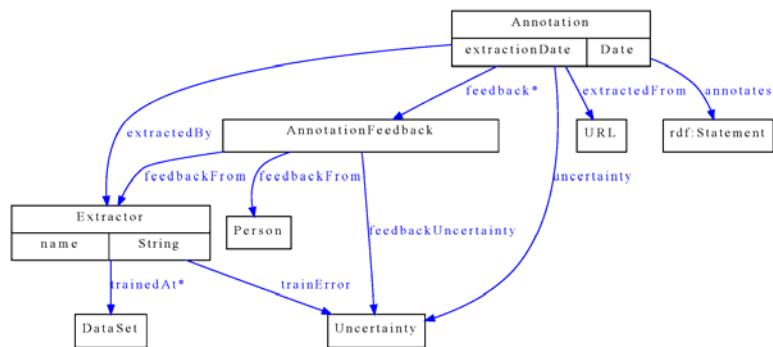

Fig. 1. Part of our uncertainty annotation ontology

Web information extraction splits pages to dominantly tabular and/or textual. Uncertainty issues connected with information extraction (and annotation) from tabular pages were discussed in [1]. Extraction of textual pages will use techniques described in [6]. Both approaches (and any other approach) generate a level of (un)certainty they have about their annotations. Also users, human or agents, can review these uncertainties and provide feedback about them.

Success of this approach can be measured primarily by the advance of semantic web functionalities. This is easier to measure for software agents. More difficult is to design metrics to measure human user satisfaction. All these aspects will be discussed in this presentation.

References.
[1] Eckhardt A., Horváth T., Maruščák D., Novotný R. , Vojtáš P.: Uncertainty Issues in Automating Process Connecting Web and User, in URSW '07 Uncertainty Reasoning for the Semantic Web. CEUR-WS.org/Vol-327/paper9.pdf
[2] Uncertainty Reasoning for the World Wide Web W3C Incubator Group Report 31 March 2008, http://www.w3.org/2005/Incubator/urw3/XGR-urw3/
[3] Search at http://www.w3.org/ for Ruby Annotation, GRDDL, RDFa
[4] Egothor search engine http://www.egothor.org/
[5] Bednárek, D., Obdržálek, D., Yaghob, J., Zavoral, F.: Data Integration Using DataPile Structure, In: Proceedings of the 9th East-European Conference on Advances in Databases and Information Systems, ADBIS 2005, Tallinn, ISBN 9985-59-545-9, 2005, 178-188
[6] J. Dědek, P. Vojtáš. Linguistic extraction for semantic annotation , accepted for IDC 2008, Catania, Sicily, to appear in the Symposium Proceedings, which will be published by Springer as a part of their series Studies in Computational Intelligence