Letters to the Editor

Ontologies

I recently read Martin Hepp's article "Possible Ontologies: How Reality Constrains the Development of Relevant Ontologies" in the January/February 2007 issue. Although normally *IC* is well known for its good-quality articles, and I do enjoy reading it, I'd like to express my disappointment about this one.

Beside the "technical" flaws of the article (listed later) it raises too many issues without proposing solutions or even concluding on the work. I would have very much liked to see what we can do to improve the situation based on Hepp's excellent analysis, rather than just reading that "we need more and better ontologies."

I do appreciate Hepp's work, and I am absolutely convinced that the issues raised are worth being addressed — by and large I can acknowledge his observations. However, I still have some complaints:

- Friend-of-a-Friend (FOAF) is, as far as I know, not a W3C recommendation, although some of the authors also participate in W3C activities.
- To count RDF as an ontology is not correct. Table 1 seems to mix apples and oranges (RDF, RDF-S, and FOAF).
- Figure 4 is very confusing; it might be that I missed something there, but how was this figure generated?
- On page 95, Hepp states that "the experiment details are described elsewhere," but the reference points to citation 6, which is obviously not the correct source — I'd be very interested in the original work that Martin is referring to, there.
- To take the sheer size of an ontology as a metric for its expressivity is (carefully stated) questionable. In case an ontology is well documented — has, for example, many `<rdfs:comments>` in — it will get big, no matter how many concepts or props you've defined in there.
- The CPU use case didn't add much. I don't see the message in this paragraph.

So, concluding, it is a worthwhile topic, but the presentation was not optimal, in my feeling.

—Michael Hausenblas
Joanneum Research
Forschungsgesellschaft

Martin Hepp responds:
First of all, the explicit purpose of the Peering column is to provide a venue for raising an opinion on current Web topics. My column goes beyond the standards of this format and even provides some preliminary evidence supporting the arguments. The main and clearly stated goal of my article is to point to technical, social, economical, and legal effects, which put a brake on the development of practically useful ontologies, and which may explain the current shortage of ontologies. As per the nature of this format, the article does not aim at solving the problems, but at bringing them to attention.

I am thankful for the feedback, for it allows me to clarify my position, but feel at the same time that the author of the letter does not challenge the column's core contribution. In particular, I read from the author's comments that he regards the paper as a good analysis of a problem that he considers relevant, that is confirmed by his own observations, and that has not yet been described elsewhere. Now to address the technical concerns.

First, that FOAF was not a W3C recommendation. The article says,

> An example of the former is to say, 'I believe the W3C that their definition of `foaf:knows` in the Friend-of-a-Friend vocabulary specification is compatible with my definition; if there are discrepancies, I'm willing to take the consequences.'

My intention was to give an example for committing to an ontology solely by trusting its creators (or by trusting a body endorsing it), instead of by reviewing the specification of the ontology. Although someone might read it this way, I did not intend to say that FOAF is actually defined by the W3C, just as I did not intend to say later in the article that the United Nations is actually providing an ontology of countries.

Next, Hausenblas says,

> To count RDF as an ontology is not correct. Table 1 to me seems to mix apples and oranges (RDF, RDF-S, and FOAF).

As explained in the "Reality Check" section (pp. 94–95), the table takes Swoogle Semantic Web Ontology (SWO)

documents as ontologies because that was the most comprehensive data on Web ontologies and their usage in Web documents immediately available (I would be happy to learn of better data sources). Taking this approximation for a reality check also seems to be valid for me given that, as the article states and as Hausenblas confirms, there is a shortage of widely used ontologies.

Second, RDF, RDF-S, and FOAF have most, if not all, of the characteristics that constitute an ontology: they specify the elements of a domain of discourse, constrain the interpretation of those elements by means of formal semantics or other modalities, and are widely accepted — that is, they are consensual. In the case of RDF and RDF-S, for instance, a W3C recommendation dated February 2004 specifies the interpretation of RDF elements and associated data.[1]

Third, and most important, the problems inhibiting the diffusion of ontologies, as described in the article, can be obviously found in RDF and basically in most W3C recommendations (and other standards alike):

- The effort for reviewing the specification for potential adopters increases with the size of the specification document.
- The effort for maintaining a specification in a quickly evolving domain increases with the amount of detail per element.
- The effort for reviewing an updated version of the standard prior to migrating to this new version increases with the specification's overall size and thus limits the amount of people who are willing to spend the respective effort.

Thus, although these aren't domain ontologies, I think it's valid to regard RDF and RDF-S as ontologies of ontology specifications — particularly for the purpose of testing a hypothesis on the social and economic effects of ontology creation and adoption, which is what the article is about.

Regarding Figure 4, it depicts a three-sided, nonlinear trade-off problem. It illustrates the prediction that the space of possible ontologies is constrained by nonlinear trade-offs between conceptual dynamics, the amount of detail and expressivity of its elements, and the number of users adopting the respective ontology. We know similar patterns and graphs from many problems in economics. Because this is a prediction of a structural pattern (and is clearly marked as such), the figure was naturally not created from experimental data, but from arbitrarily chosen parameters for such a problem of the kind $z = \gamma\sqrt{(N - x^\alpha - y^\beta)}$, where z = degree of detail, x = conceptual dynamics, y = size of the user community, $N > 0$ is an arbitrary calibrating value, and $0 < \alpha, \beta, \gamma \leq 1$. Because an ontology below a particular size and amount of detail is irrelevant (that is, there is a z_{min} below which there is no useful ontology), the tail of the pane on the right side of the original plot ($x = x_{max}, y = y_{max}$) is cut off in Figure 4 of the original article.

My apologies for the mis-cited reference 6 — it should be reference 7.

Next, Hausenblas says, "To take the sheer size of an ontology as a metric for its expressivity is (carefully stated) questionable." The text says, "I used the ontology-specification file size as an approximation for the level of detail *and* expressiveness and the number of Semantic Web documents as an approximation for the size of the community using the ontology."

Although I admit that I use "expressiveness" in a broader sense than a logician here, there are good reasons to do so. One of the main things I'm trying to deliver in the article is that any useful ontology contains a formal part (which constrains the interpretation of conceptual elements by means of logic) and an informal part (such as labels and natural language definitions). Quite clearly, it's

impossible to assert enough in any language about the elements in an ontology to narrow down the interpretations to a single possible world.[1] In practice, we need ontologies that define elements with a narrow, real-world meaning for most of the problems that the Semantic Web community is considering as candidate applications of ontologies. For example, we may need ontologies with classes such as `Portable Color TV ⊆ TV Set ⊆ Media Device`. In such cases, the "practical" ontological commitment goes way beyond `A ⊆ B ⊆ C`.

In effect, anyone who considers adopting this ontology has to review both the formal part of the ontology and the informal part. Thus, when it comes to assessing the effort needed for reviewing the ontology specification prior to using the respective ontology, or the effort for updating an element in the ontology due to changes in the domain, *both the formal and the informal content* of the ontology count.

In my reality check, I use the file size as an approximation of the total amount of detail and expressivity of the formal and informal parts of the ontology.

As for the last comment, the CPU example shows how an ontology of Intel CPUs — which would be needed, for example, for many e-business scenarios and catalog data integration in the computer equipment domain — would often contain less than 40 percent of the current CPU models, assuming that the ontology was updated only once per year. This simulation experiment supports the problem shown in Figure 1.

In a nutshell, I join the author of the letter in his call for starting to work on finding solutions for the practical obstacles identified in the paper. 🔲

Reference

1. *RDF Semantics*, World Wide Web Consortium (W3C) recommendation, 10 Feb. 2004; www.w3.org/TR/2004/REC-rdf-mt-20040210/.