

## GROUP SIX:

### Ontologies: Building Useful Knowledge Bases across Agencies

**Participants:** Eduard Hovy (chair), Shawn Bowers (scribe), Bob MacGregor, Raphael Malyankar, Shawn Kerrigan, Melania Degeratu, Jose Luis Ambite, Steve Reichenbach, Isabelle Cruz, Bill Labiosa.

#### DISCUSSION

The digital government program includes research on technologies to automatically build and populate large ontologies. These conceptual knowledge bases can be used by people to navigate complex terminology both within and across agencies. The goal of this group was to identify both theoretical and practical problems in building ontologies, propose concrete solutions, and explore additional data resources that can be analyzed for ontology construction. We also considered issues of integration with metadata, both existing and proposed.

#### What follows is our discussion.

It seems that the uses of ontologies people are working with include: Data in heterogeneous DB where values need to be integrated. For example, what is a price? Nomenclature, glossaries, and terms ie, do terms mean the same thing? We need some kind of definitional standardization (similar to machine translation of language), and to do reasoning, knowledge representation, infer relationships in knowledge, etc. There are different definitions of ontologies. We do not deal with the pure philosophy notion of ontologies. There is danger of philosophical wars about what they should be.

#### Our top three issues are as follows:

- 1. Problems:** Different words that mean the same thing (e.g., different spellings, or synonyms), but, different words with the same meanings (acronym); acronym lists, hierarchies specific to a discipline; science and engineering are different things; pushing to end user to decide context is important. Co-occurrence analysis, statistical analysis (computational linguistics) patterns of co-occurrence. of words, bank tellers, money, etc; bank, bait, fish, water, etc.; co-occurrence - new research in ontology/s.
- 2. Has anyone actually built ontologies from scratch using text?** There has been some intervention, but in general this can be easier than from text mixtures of domain experts and software engineer's. There is no methodology. Are there tools to support ontologies for creating, browsing, APIs, etc? There are some standards being developed for notation (e.g., KIF) some experience, browsing some standards for specific domains on standardizing ontologies. Several standards and mapping between them is a potentially better approach than coming up with "the standard ontology"
- 3. Building ontologies** The data exists, but there is research being done to build applications that manipulate it an the emphasis on concepts is usually wrong. We can't come up with a single definition for attributes that applies for all cases. Specifying each different definition typically leads to an incomplete list. When people typically model the data (e.g., define a schema), they don't necessarily know the exact definitions anyway. Instead the focus should be on the (binary) relationships. We need to look for usage patterns in the relationships, but don't constrain what you can have. One question to consider is; aren't you just moving the problem to relationships? The answer is no, because we aren't interested in the semantics of concepts, just observe and correlate the relationships. There was a discussion about knowing monthly and yearly salaries. Basically, the discussion was about what the difference is between Entity-Relationship Diagrams/UML Diagrams and ontologies. Parallel ontologies: Are there (declarative) languages to express mapping between parallel ontologies, to traverse back and forth.

We felt it useful to have some survey if it wasn't too difficult to list projects that were being done in government. From simple vocabularies to more complex structures. For example, who created the NAICS ontology of industries? This would be one place to start. We need a list of ontology efforts in government plus contact points. What should have been addressed? We need more on the possibility/trade-offs of creating ontologies programmatically as opposed to by hand? What has been tried and what has failed? What should be the scope of an ontology? An argument can be made that the ontology only needs to be as complex as the problem you need to solve, but you can create large backdrop ontologies, where you can plug in your localized ones (big, general high-level tools)