Knowledge Base Evaluation for Semantic Knowledge Discovery

James Mayfield, Bonnie Dorr, Tim Finin, Douglas Oard and Christine Piatko
Human Language Technology Center of Excellence

Semantic knowledge discovery has traditionally been evaluated at the text level. For example, NIST evaluations such as MUC and ACE evaluate the information extraction of particular types of semantic roles and relations primarily at the mention level. We suggest that evaluating at the level of a knowledge base (KB) extracted from the text has some significant advantages over evaluation at the text level. By knowledge base, we mean the combination of a database, a descriptive schema for the contents of the database, a collection of background knowledge, and an inference capability.

A knowledge base might be evaluated along a variety of dimensions, each corresponding to an aspect of semantic knowledge discovery:

- **Accuracy**: What is the precision of the extracted information? What is the system's level of confidence that the information was properly extracted?
- **Usefulness**: Is the extracted information relevant to the target task? Was all of the relevant information discovered? Is redundancy removed? Is the granularity of the extracted information at the appropriate level?
- **Augmentation**: How does the knowledge discovery augment information already in the knowledge base? Does it exploit that information to improve its performance?
- **Explanation**: What is the provenance of the induced information? Are attribution and committed belief properly expressed?
- **Adaptation**: Can the techniques be applied to new languages, genres, and domains? How much tuning is required to make such a leap?
- **Temporal qualification**: Are the assertions in the knowledge base accurately marked with when they are true? Is information that is currently true properly distinguished from information that was true at some time in the past but that is no longer true?

Each of these evaluation axes might be used to assess the output of a semantic knowledge discovery tool directly, without reference to a knowledge base. However, performing the evaluation over the resultant knowledge base has several advantages:

- It forces the evaluation away from linguistics and fully into the realm of semantics.
- It allows inference rules to be applied in performing the evaluation.
- It enables the use of existing knowledge bases as ground truth, possibly reducing annotation costs significantly.

Note that a KB evaluation approach is applicable not just to situations where fixed known fields are to be extracted, but also to situations where the kinds of information to be extracted are induced from the data themselves [Sekine 2006]. In such cases, the KB approach can be quickly adapted to new kinds of information by writing inference rules that relate the new types of information to be extracted to the ontology that serves as the schema for the knowledge base.

Sekine, S. 2006. On-demand information extraction. In *Proceedings of the COLING/ACL* Annual Meeting of the ACL. Association for Computational Linguistics, Morristown, NJ, pp. 731-738.