From: AAAI Technical Report SS-02-09. Compilation copyright © 2002, AAAI (www.aaai.org). All rights reserved.

Acquiring (and Using) Linguistic (and World) Knowledge for Information Access or Theory for systems; application for theories

Jussi Karlgren and **Björn Gambäck** and **Pentti Kanerva**

Information access tasks need text understanding

To move forward the research frontier in the general field of information access, one of the bottlenecks we need to address is understanding textual content somewhat better. While full text understanding remains a distant and possibly unattainable goal, advances in content analysis beyond the simple word-occurrence statistics or name-recognition algorithms used today would seem to be desirable.

Information retrieval is a blunt information access task, and information-retrieval systems deliver useful results with a simple text and content model. Much better models are necessary for information access tasks that involve information refinement, meaning tasks that involve processing information in text–and some specific questions in information retrieval proper are fairly knowledge- intensive such as query expansion or questions related to multilinguality.

In addition, the dynamic nature of both information needs and information sources will make a flexible model or set of models a necessity. Models must either be adaptive or easily adapted by some form of low-cost intervention; and they must support incremental knowledge build-up. The first requirement involves acquisition of information from unstructured data; the second involves finding an inspectable and transparent model and developing an understanding of knowledge-intensive interaction.

Text understanding needs a theory

Whatever the type of model, it must be represented in some way. But knowledge modeling, semantics, or ontology construction are areas marked by the absence of significant consensus either in points of theory or scope of application. Even the terminology and success criteria of the somewhat overlapping fields are fragmented. Some approaches to content modeling lay claim to psychological realism, others to inspectability; some are portable, others transparent; some are robust, others logically sound; some efficient, others scalable. There is no explicit standard and not even any accepted practice to adhere to or to deviate from.

Copyright © 2003, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

Information access tasks give focus to modeling

It is too much to hope for a set of standards to emerge from the intellectually fairly volatile and fragmented area of semantics or cognitive modeling. But in our application areas — namely, those in the general field of information access — external success criteria are better established. Compromise from theoretical underpinnings in the name of performance is not only possible but even desirable.

The focus of this workshop are models and model tasks

This workshop aims to bring together researchers that work with any kind of text analysis with the aim of understanding text, and with information- access applications in mind. The idea is to attract participation from diverse fields with the goal of identifying reasonable interfaces between analyses of various type. Participants would be encouraged not only to relate successes their approach has engendered but failures due to lack of knowledge or due to unsatisfactory modeling; other participants would be encouraged to discuss and offer contributions to the goal. Ideally, projects would invite each other to work with material they already are working with and to investigate cross-pollination of different approaches.

Challenge questions

- How does my model adjust to a new domain or to previously unknown material?
- How can the knowledge in my model be inspected, assessed, and hand-edited except in a near-full-scale trial?
- How well does the model perform in a large-scale trial? (By ANY metric!)
- What additional knowledge or theory does my model need to perform better?

Unsatisfactory answers are encouraged if they invite further cooperation between groups!