

Preface

Since its inception less than seven years ago, Wikipedia has become one of the largest and fastest growing online sources of encyclopedic knowledge. One of the reasons why Wikipedia is appealing to contributors and users alike is the richness of its embedded structural information: articles are hyperlinked to each other and connected to categories from an ever expanding taxonomy; pervasive language phenomena such as synonymy and polysemy are addressed through redirection and disambiguation pages; entities of the same type are described in a consistent format using infoboxes; related articles are grouped together in series templates.

As a large-scale repository of structured knowledge, Wikipedia has become a valuable resource for a diverse set of artificial intelligence (AI) applications. Major conferences in natural language processing and machine learning have recently witnessed a significant number of approaches that use Wikipedia for tasks ranging from text categorization and clustering to word sense disambiguation, information retrieval, information extraction and question answering. On the other hand, Wikipedia can greatly benefit from numerous algorithms and representation models developed during decades of AI research, as illustrated recently in tasks such as estimating the reliability of authors' contributions, automatic linking of articles, or intelligent matching of Wikipedia tasks with potential contributors.

The goal of this workshop was to foster the research and dissemination of ideas on the mutually beneficial interaction between Wikipedia and AI. The workshop took place on July 13, 2008, in Chicago, immediately preceding the Annual Meeting of the Association for the Advancement of Artificial Intelligence.

This report contains papers accepted for presentation at the workshop. We issued calls for regular papers, short late-breaking papers, and demos. We received an impressive number of submissions, demonstrating the community interest and the timeliness of the "Wikipedia and AI" research theme: 22 regular papers, 10 short papers and 3 demos were submitted to the workshop. After careful review by our program committee, 10 regular papers, 10 regular papers, 2 short papers and 1 demo were accepted for presentation. Consistent with the original aim of the workshop, the accepted papers address a highly diverse set of problems, in which Wikipedia is seen either as a useful resource, or as a target for algorithms seeking to make it even better. As a rich knowledge source, Wikipedia is shown to benefit applications in information extraction, machine translation, summarization, ontology mining and mapping, and information retrieval. We also learned of interesting applications, most of them using machine learning, that could further enhance the breadth and the quality of Wikipedia, such as predicting the quality of edits, vandalism detection, infobox extraction, crosslingual links creation, and semantic annotation.

We were truly impressed by the high quality of the reviews provided by all the members of the program committee, particularly since deadlines were very tight. All of the committee members provided timely and thoughtful reviews, and the papers that appear here have certainly benefited from that expert feedback.

Finally, when we first started planning this workshop, we agreed that having a high quality invited speaker was crucial. We thank Michael Witbrock not only for his talk, but also for the boost of confidence provided by his quick and enthusiastic acceptance.

– Razvan Bunescu, Evgeniy Gabrilovich, and Rada Mihalcea
July 2008