Semantic Web and Best Practice in Watson

Chris Welty

IBM Research

1 About the Speaker

Dr Chris Welty is a Research Scientist at the IBM T.J. Watson Research Center in New York. His principal area of research is Knowledge Representation, specifically ontologies and the semantic web, and he spends most of his time applying this technology to Natural Language Question Answering as a member of the DeepQA/Watson team. He is best known as a co-developer of the OntoClean methodology with Nicola Guarino, and as the co-chair of the W3C RIF working group.

2 Speech Abstract

IBM's revolutionary Watson system has successfully beaten human Jeopardy champions, and is now being extended and used in other domains, such as healthcare question answering, and financial data analysis. It is a common misconception that Watson is a through-and-through formally semantic system, which translates questions into formal language queries, and returns answers by executing those queries over a large knowledge base. In actual fact, Watson uses a variety of technologies to produce candidate answers to each question, and semantic technologies are primarily used in the subsequent candidate answer ranking components. In particular, linked data sources are used to provide typing evidence for candidate answers, but also several other answer ranking components rely more or less on semantics and linked data.

In this speech Dr Welty discusses these semantic components and the data upon which they operate, giving examples of expected and unexpected behaviours, and how these affect the resulting answers returned by the system. He also touches upon the methods and practices employed in developing and testing Watson, giving useful suggestions for practitioners building real world large-scale cognitive systems.