Building Watson: A Grand Challenge in Automatic NL Question Answering

David A. Ferrucci
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598, USA
ferrucci@us.ibm.com

Koichi Takeda
IBM Research - Tokyo
Yamato, Kanagawa 242-8502, Japan
takedasu@jp.ibm.com

Extended Abstract

On April 27, 2009, IBM unveiled the details of a project for building an advanced computing system that will be able to compete with humans at the game of Jeopardy! Computer systems that can directly and accurately answer peoples' questions over a broad domain of human knowledge have been envisioned by scientists and writers since the advent of computers themselves. Consider, for example, the computer on Star Trek – how it understands questions and quickly provides accurate, customized answers and can engage in a fluent information seeking dialog with the user. We call this technology open domain question answering and it has tremendous promise for impacting society and business. Applications in business intelligence, health care, customer support, enterprise knowledge management, social computing, science and government would all benefit from such technology. The Project Watson is addressing a grand challenge in Computer Science aimed at illustrating how the integration and advancement of Natural Language Processing (NLP), Information Retrieval (IR), Machine Learning (ML), massively parallel computation and Knowledge Representation and Reasoning (KR&R) can advance open-domain automatic Question Answering to a point where it clearly and consistently rivals the best human performance. An exciting proof-point in this challenge is to develop a computer system that can successfully compete against top human players at the well-known Jeopardy! quiz show. Attaining champion-level performance at the game of Jeopardy! requires a computer system to rapidly and accurately answer challenging open-domain questions, and to predict its own performance on any given category/question. The system must deliver high degrees of precision and confidence over a very broad domain with a 3 second response time. It is highly unlikely that any system will be able to clearly justify all the answer with perfect certainty over such a broad range of natural language questions and content. Computing accurate confidences is an important requirement for determining when to “buzz in” against your competitors and how much to bet. While critical for winning the game, high precision and accurate confidence computations are just as critical for a QA system to provide real value in business settings. The need for speed and for very high precision demands a massively parallel compute platform capable of generating and evaluating 1000’s of hypotheses and their associated evidence. In this talk we will introduce the audience to the Jeopardy! Challenge and describe our technical approach and progress on this grand-challenge problem.