Jonas Martinsson
CMSC828D
09/30/12

What is Watson?

After the success of Deep Blue, the chess-playing computer that beat grandmaster Garry Kasparov in 1997, IBM turned to the challenge of trying to beat the best Jeopardy! players. The result of this undertaking was a computer system named Watson, which managed to beat the two most successful participants to date in two exhibition matches in 2011.

To be able to do well in Jeopardy! (Rennie, 2011), contestants must be able to analyze clues from a wide variety of subjects. They also have to consider subtleties in form of for example irony and riddles; and they only have a few seconds to do this. Watson has 16 terabytes of RAM and 2880 processor cores working in parallel. Solving each given clue is divided into a number of subtasks that can be performed independently of each other. Since Watson is not connected to the Internet, all data is stored locally. When under operation, this data is stored in the RAM since reading from a hard drive would be too slow for the purpose. Watsons "knowledge bank" consists of 200 million pages of structured and unstructured content (Jackson, 2011).

An example of how Watson works is shown in Martin (2011). First the clue "The antagonist of Stevenson's Treasure Island." is given (to which the correct answer is Long John Silver). Watson then goes on to perform seven different steps.

1. The clue is parsed with respect to natural language processing, so that it can be logically structured.
2. Assumptions about the semantics of the words in the sentence are generated.
3. Queries based on keywords and the different semantic interpretations are constructed.
4. A multitude of answers based on texts found from the queries in the previous step are presented.
5. For each answer, a new search is done to find evidence for the previous answers. This evidence is then scored so that it can be confirmed or discarded.
6. New searches are done and the answers are scored.
7. All evidence and their respective scores are combined to give a measurable confidence in the answers.

Watson contains 41 different subsystems (Feldman, 2012). The most prominent of these in the example above are natural language processing and machine learning. NLP is used to make a representation of a clue sentence in a way that makes it possible to process it in other domains, such as logic, categorization and search. The role of machine learning in the last step is to assign appropriate weights to the scores of evidence and answers. Prior to the televised matches mentioned above, Watson played over 100 games against other previous Jeopardy! contestants to be able to learn from its mistakes and improve enough to beat any human.

While Watson did succeed in beating the old champions, there are questions it would not be able to answer. One example of this is information that is not explicitly written down, such as "If you're standing, it's the direction you should look to check out the wainscoting." (Martin, 2011).

Aside from answering clues Watson also has algorithms for seeking out daily doubles, and it only decides to hit the buzzer if its confidence level of an answer is over a given threshold. This threshold depends, among other things, on the state of the game, the contestants' scores and how well they are playing.

When viewing Watsons televised Jeopardy! matches, one can notice that some answers looked like they were not entirely processed. By this I mean that they either were in all capital letters, or included pieces of information that was superfluous or unrelated. According to the Watson research team (Martin, 2011) this happened as rarely as in 1% of the answers, but they did not measure if this affected the correctness of the answers at all.

The system was built with the goal to be able to beat the best human Jeopardy! players. However, the ultimate goal was not this but to use it in more scientific contexts. Watson has since its Jeopardy! performance been modified to be able to do work in other fields as well, the main one being healthcare. By looking at the vast amounts of medical data available, the hope is that Watson could be able to recognize subtle patterns more or less impossible for humans to see. The difference from a medically trained human is that Watson would not be able to do deductions based on experience, but a human does not have the time to go through the same amounts of data. Watson could in this context be a complement to human personnel.

One of the current projects utilizing Watson is one including University of Maryland and Nuance Communications (Andrews & Mack, 2011). Nuance has knowledge in speech recognition in the medical domain, and the UMD School of Medicine is investigating how to make medical personnel and technology like Watson interact in the best way.

References

Rennie, John (2011). How IBM's Watson Computer Excels at Jeopardy!. http://blogs.plos.org/retort/2011/02/14/how-ibm%E2%80%99s-watson-computer-will-excel-at-jeopardy/, available 09/30/2012.

Jackson, Joab (2011), IBM Watson Vanquishes Human Jeopardy Foes. http://www.pcworld.com/article/219893/ibm_watson_vanquishes_human_jeopardy_foes.html, available 09/30/2012.

Martin, Erik (2011). IBM Watson Research Team Answers Your Questions. http://blog.reddit.com/2011/02/ibm-watson-research-team-answers-your.html, available 09/30/2012.

Feldman, Susan (2012). IBM's Watson: From Winning Games to Saving Lives. http://www-03.ibm.com/innovation/us/watson/pdf/lcUS23400812.pdf, available 09/30/2012.

Andrews, Chris & Mack, Richard (2011). IBM to Collaborate with Nuance to Apply IBM's "Watson" Analytics Technology to Healthcare. http://www-03.ibm.com/press/us/en/pressrelease/33726.wss, available 09/30/2012.