IBM's Watson Supercomputer: A Brief Overview

Stefan Gloutnikov
Computer Science Department
San Jose State University
San Jose, CA 95192
408-924-1000
sgloutnikov@gmail.com

1. ABSTRACT

Artificial intelligence has always been a challenging and important topic in the Computer Science field, and remains so to this day. Over the last couple of years IBM has been developing a computer that has edged the field of Artificial Intelligence forward to new heights. IBM's Watson supercomputer is closer to the human thinking process than any other computer system in the world, and has demonstrated this in a game of Jeopardy against two of the game's best contestants. The astonishing work on the Watson supercomputer is definitely a milestone in computer AI research that has set forth new challenging heights for future AI work.

2. INTRODUCTION

Watson's initial goal was to be able to play the game of Jeopardy and win against some of the best opponents. This is a very challenging task for a computer, due to the fact that the computer must be able to analyze the question being asked, and then somehow reach an answer with a significantly large confidence level in order to be able to buzz in. This whole process is of critical importance to Artificial Intelligence researchers, because it could mean that scientists could develop new 'more intelligent' computers to do various different tasks that current computers are just not capable of. And as it stands now, most, if not all computers have very limited capabilities of performing AI. In this paper we will see how the extremely powerful collection of hardware and newly developed software from IBM has made Watson an AI giant, and has led him to beat two of Jeopardy's best players. Though winning in Jeopardy was an extremely challenging task, we will also see how Watson can possibly be used in future endeavors in the computer science realm.

3. WATSON'S HARDWARE

In order to achieve such flawless performance Watson must run on very powerful hardware that is currently unavailable to the regular consumer (large clusters). Watson takes up roughly a room size when all the components are assembled. As with almost any computer running today, another critical component to this large set-up is the cooling needed in order to cool the hardware so that it can function properly. And this is why all the cooling fans and air conditioning contribute most of the noise inside the room where Watson is positioned. [2, 5, 7, 8]

Figure 1. IBM POWER 750 server.

3.1 Processing Power

The core behind IBM's Watson computer lies in the 90 IBM POWER 750 servers which are clustered together to work as one unit. The servers are laid out on 10 racks, and were primarily chosen for the fact that they can deliver powerful and massive parallelism to the complex processes that Watson will be running. One such feature inside the Power7 processor is what is known as "Intelligent Threads". This technology allows processing threads to vary in their execution dynamically, based on the current workload of the machine. This allows the Power7 to provide a greater total capacity of processing power because more tasks are completed in parallel. Watson will benefit greatly for such architectural design because along with accuracy of the question/answer, speed is as of equal importance. Watson will need to answer a question in roughly three seconds in order to compete with the top Jeopardy players. Each POWER 750 server utilizes four processor sockets, and each socket with four Power7 processors, and four threads per core. [2, 5] With all this combined, Watson is able to operate 2880 processing cores at the speed of 80 trillion teraflops per second. Single POWER 750 servers and Power7 processors are available to public consumers from IBM, but the beauty of Watson is the power that the computer can harness from all ninety servers clustered together. Because the hardware was not specifically engineered for the sole purpose of Watson in mind, costs of the supercomputer have been significantly reduced. The extremely powerful hardware gives Watson the ability to process a question in seconds—something that no other computer in the world can accomplish today. [7, 8] These POWER7 systems were invented with versatility in mind. They can handle traditional on-line shopping transactions, business analytics, data warehousing and many more. This makes the POWER7 systems the most versatile and yet powerful systems in the world, that can adapt into many different environments.

The Power7 processors and Series 750 servers are widely used in the world today at businesses, universities and other scientific facilities. Their powerful architecture and energy efficiency allows for much more cost-effective computing, where other processors lack to bring this exact combination. Researchers and students at Rice University in Houston are using IBM's latest technology to fight cancer. They use the systems to find patterns in extremely large collections of data, which someday might in turn lead to possible cures for cancer. [2, 5]

Figure 2. IBM's Watson.

3.2 Memory and Storage

In order to harness all the power, the Power7 processors and Watson are accompanied by fast memory, and lots of it. In total Watson runs on sixteen terabytes of memory along with roughly twenty-one terabytes of storage. Even through the storage capacity of Watson is so large, the system uses around one terabyte of data in order to generate its Jeopardy answer. Watson pulls most of its information from already published texts in the form of encyclopedias, articles and news events. This is due to the fact that Watson too, like his competitors does not have access to the internet. Also researchers have found that even with internet access, structuring a query to ask a search engine, would be of greater difficulty than to look through published texts. [7, 8]

Both memories, like the processors, run inside the cluster of ninety Series 750 servers. IBM researchers identify that the memory aspect of the machine is just as important to Watson as the processors. Without an efficient and impelling memory system in place, no matter how much computer power you possess will not produce effective and accurate results. [8] Watson's memory architecture consists of an ideal mix of main memory and several levels of cache memory. The architecture is capable of distributing tasks in a way that balances the load among all Power7 processing cores for more efficient computing. In order for Watson to be quick and provide 3-second responses to Jeopardy questions, the entire data set resides on Watson's memory. Like all computers, disk I/O is the slowest operation performed, so engineers have worked around this challenge by storing all data that Watson will need on memory during operation. This way fetching information becomes extremely fast and efficient.

Each of Watson's processors is capable of a bandwidth of 500 GB/s, which is roughly equivalent to one million books per second. The entire cluster powering Watson is capable of bandwidth speeds of around 180,000 GB/s. [8] Even though the supercomputer possesses this magnificent amount of speed, speed alone will not be enough to challenge the best in a game of Jeopardy, and win. It will take another brilliant invention, this time it is in the software department. It's called DeepQA and will accompany Watson's hardware to new AI heights. [8]

3.3 Power7 Energy Consumption

Of course, a powerful processor and system such as the Power7 series which powers Watson will require enormous amounts of electricity. The engineers have cleverly thought out solutions to accommodate the power-hungry system, and to ensure that no energy goes wasted. In order to save energy, the Power7 system is capable of powering on and off different parts of the system and to also adjust processor clock speeds up and down. [2] Based on the tasks which the system needs to perform, and the current temperature conditions in the room, the system is able to self-adjust to accommodate low energy consumption while not disrupting any computation. IBM's Power7 systems are the first four-processor servers that exist today to qualify for EPA's energy star status. Energy star is an US government program that rates power consumption levels of almost all electronics in the United States, including computers and servers. [2]

Figure 3. Energy star is an US government program that rates power consumption levels of almost all electronics in the United States

4. JEOPRADY, CHALLENGES, AND RESTRICTIONS

4.1 Jeopardy

The Jeopardy game is a very popular quiz show in the United States, and is one that requires knowledge in many different categories and topics. There two rounds, in which there are six categories for each round with five questions in each category. In the first round each question is worth $100 through $500 for a total of thirty questions. The second round is worth double, and in each round (two in round two) there is a question that is called the 'Daily Double". These types of questions are hidden from the players and are only revealed after the question has been selected. In case of a daily double question, the player who selected the question automatically has a chance to answer the clue without competition from the remaining players. There is another trick to these types of questions, and that is that the player who selected it must write down an amount they are willing to wager, with $5 being the minimum bet and the maximum bet is the larger of the contestant's current score, and the maximum clue value on the board.

Figure 4. Example of a Jeopardy board.

After the host is finished reading the question, the first one to buzz in gets the opportunity to answer. Contestants have a total of five seconds to give an answer, though most often contestants answer immediately since they must have known the answer in order to buzz-in in the first place. Answers to a question in a game of Jeopardy are given in the form of a question. A correct answer adds to the contestant's total, the amount that the question was worth, and an incorrect answer subtracts it. It is actually possible for a contestant to go in the negative range if he/she has given a series of incorrect answers. This simple little fact makes Watson's work this much harder, because the computer now does not only have to figure out an answer, but is also responsible for figuring out a confidence level of that answer. After a correct answer, the contestant is given the opportunity to select the next available question from the board and this process resumes. On an incorrect answer, the remaining two players are given a chance to attempt answering the question. This is also true even if the second player gets the answer wrong. On no answer, or all three wrong answers, control of the board is returned to the previous player who had control.

At the end of the game there is one category revealed to all three players. After which, they have to write down an amount they are willing to risk on that category and not share it with the rest of the contestants. The maximum is the amount of winnings they currently possess. When this step is complete the actual question is revealed, and the contestants have thirty seconds to write down their answer. A correct answer subsequently adds the amount risked, and an incorrect answer subtracts from the player's total the amount risked. The player with the most winnings at the end wins.

4.1.1 Types of Categories

As already mentioned, each round consists of six categories and five questions in each category. The different categories can be very specific, like school subjects – "Math", "Chemistry", "Physics", and "History". Or less specific, ones like "Before & After", "The Oscars", and "Name's the same". In the less specific categories Watson is unable to jump to an answer straight away, and many different considerations need to be computed. This was only one of the many challenges that faced researchers in making Watson be able to play Jeopardy. [1, 4]

4.1.2 Types of Question

Just like in the category titles, questions too are not very direct and require a lot of analysis before even attempting to compute an answer. Researchers found that most of Jeopardy questions are factoid questions – questions that are based on facts about someone or something, and that has helped them tune the DeepQA software. The questions alone are challenging to even attempt to answer because they only offer small clues as to what an answer might be, and do not ask it directly. Here are some questions that have been asked on Jeopardy before [4]:

Category: General Science

Clue: When hit by electrons, a phosphor gives off electromagnetic energy in this form.

Answer: What is Light?

Category: Lincoln Blogs

Clue: Secretary Chase just submitted this to me for the third time; guess what, pal. This time I'm accepting it.

Answer: What is his resignation?

Category: Head North

Clue: They're the two states you could be reentering if you're crossing Florida's northern border.

Answer: What is Georgia and Alabama?

Some other types of questions can contain multiple clues about the answer within the question, but it is not normal

for both clues to appear together in the same part of any published text that Watson has access to. In this case the question needs to be properly decomposed and analyzed in order to reach a possible answer. Here is an example of that kind of question given by Watson's creators [4]:

Category: "Rap" Sheet

Clue: This archaic term for a mischievous or annoying child can also mean a rogue or scamp.

Sub-clue 1: This archaic term for a mischievous or annoying child.

Sub-clue 2: This term can also mean a rogue or scamp.

Answer: Rapscallion

There are many more forms of questions that IBM Researchers have identified and have defined. They have found various ways of identifying, decomposing, and going about how to answer them, but they go beyond the scope of this paper.

4.2 Challanges

In order for Watson to compete successfully in the game show, there were lots of challenges that had to be met. Some of which include understanding what the question is actually asking, seeking clues gathered from parsing the question, actually seeking information about the clues gathered, computing a confidence level about the answer gathered, and many more. Winning in Jeopardy requires that you are confident in your answers, because a wrong answer will subtract the money amount for which you were playing for. Also, most questions being asked are not direct. Most of them provide hidden clues that will point you towards an answer, but very few will simply state a question that Watson needs to begin seeking an answer for. For this reason Watson needs to be able to break down the clue provided and compute some sort of confidence level from which to decide whether or not to attempt answering. Researchers found out that this process must be completed in one through six seconds, and on average in three seconds. [3, 4] Estimating confidence was a very important task, and a big contributor to shaping the design of the DeepQA software, which runs on Watson. DeepQA factors in that not all components will work perfectly, and provide a perfect answer. The software is able to dynamically learn from its various components some confidence level, and compute a global confidence level at the end for the given question. While it is clearly documented how Watson was designed to compute his confidence level through various statistics collected, it goes beyond the scope of this paper and will be omitted.

Speech synthesis was one of the other many challenges faced by the team at IBM. Watson had to be able to speak properly in order to give proper answers. The way that the team tackled this issue was to break down every possible word in small parts, and have an actor read thousands and thousands of lines in a recording studio. Then when a word would appear, Watson would be able to break the word down, and combine the pre-recorded parts to speak out the word. [10] Watson has an enormous database of words that could most likely never be spoken by the computer. Watson is also capable of making mistakes in speaking out the words, as the text-to-speech synthesis is not yet perfect. A researcher on the sound team at IBM describes Watson's speech capabilities as that of a "young child".

These are only some of the extremely large number of challenges faced by the team developing Watson. They are the ones which I have chosen to inform you about in this paper, and the rest can be found in the various sources cited at the end of the paper.

4.3 Restrictions

Even though Watson is a very powerful and smart computer, some restrictions on the types of questions that could be asked at the Jeopardy game had to be imposed. Questions that required viewing or listening to video and audio had to be excluded from the game at this stage of Watson's development. An example of such a question is [4]:

Category: Picture this: (Contestants are shown a picture of a B-52 bomber)

Clue: Alphanumeric name of the fearsome machine seen here.

Answer: What is the B52 bomber?

IBM Researchers have chosen to throw out any speech-recognition techniques for this current release of Watson. This was done in order to not throw in any errors in the question analysis phase. Competing on Jeopardy was the initial task of IBM's researchers, and errors from speech-recognition software could only get in the way at this stage. [4]

5. DeepQA

DeepQA is the Artificial Intelligence software developed for Watson in order to compete on Jeopardy. QA stands for the typical AI problem of question, answer. DeepQA was written in both Java and C++, and also uses Apache Hadoop for its distributed computing needs. The software runs on Watson's operating system of choice, which is SuSE Linux Enterprise Server 11. There are over one hundred methods that the software uses to analyze clues and searches for possible answers. There is however, a general flow of events that DeepQA goes through in order to generate an answer—described later. Refer to [Figure 5] for a high-level overview of this process. We keep in mind that this entire process is being done all at once, in parallel, over the distributed system which Watson runs on in roughly three seconds.

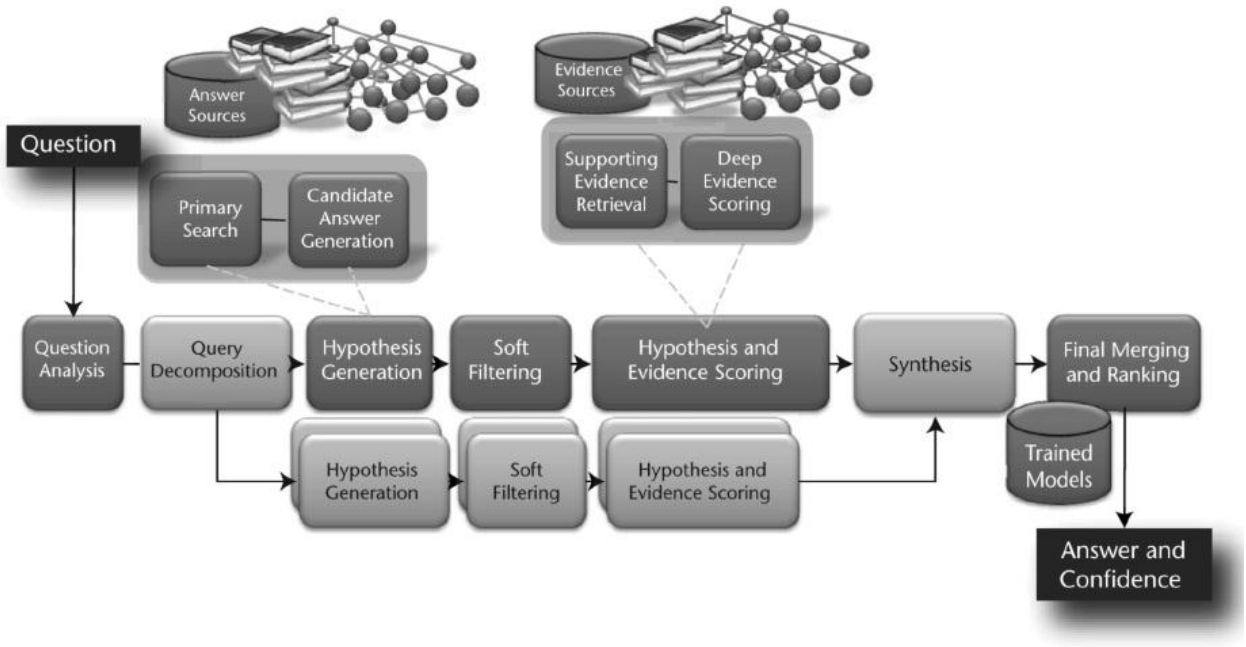

Figure 5. A high level diagram of the DeepQA Process.

5.1 Content Acquisition

In order for any QA software to work, the computer needs to first acquire knowledge that will be used in answering the questions. Watson is initially fed tons of information in the forms of encyclopedia articles, news articles, dictionaries, thesauri, and many more. Watson then runs pre-selected algorithms on the given data to help speed up content acquisition in the future. It is important to note that this step is done only once, at the beginning of powering up Watson. After Watson has read the initial data and ran through its initial algorithms for content acquisition all the working data is moved to memory in order to speed up access to it. [3, 4]

5.2 Question Analysis

In this step Watson begins to process what is actually being asked. When a human is being asked a question, we begin to break down the semantics of the language in order to process the question's meaning. We might not be aware of this, but this is how we understand languages and what they attempt to get across. For any AI however, this is a very challenging task. Watson in this step begins to break down the question in all ways possible, in order to prepare for any text matching a possible answer to the question. There could be hundreds of ways in which Watson could break down a question, and all those hundreds of ways need to be run through the 'knowledge' that Watson possesses from the initial step. Later, another algorithm will filter out bad evidence, low confidence answers, etc. Figure [Figure 6] demonstrates a way that DeepQA has broken down a question.

5.3 Hypothesis Generation

In this step Watson uses the ways that questions were broken down in the previous stage, and begins searching for possible solutions to them. The process of looking through the collection of data for a possible solution is called 'information retrieval' and is something very similar to what Google performs when a user types in a query. Again, like in the previous step, in this stage of DeepQA's process we could have multiple possible answers to a question. Watson stores all possible answers to be ranked in confidence by a later step. At this stage, Watson's configuration allows him to pass forward around 250 possible answers, which will be reduced down even more in the soft filtering step. [4]

5.4 Soft Filtering

This step is also referred to as the 'lightweight filtering' stage. Watson takes in the questions from hypothesis generation, and begins to run simple filters on them in order to eliminate incompatible answers. This step is of critical importance because it ensures speed and eliminates unneeded work. After filtering the possible answers (maximum 250) from the previous step, Watson passes down the remaining sets of answers (maximum 100) for a more intense filter. The step of soft filtering takes significantly less time than intense search, and is how DeepQA tries to eliminate ambiguous work as early as possible. [4]

5.5 Hypothesis and Evidence Sorting

In this stage Watson performs an even deeper and more careful search on available information from the questions passed down from soft filtering. It is assumed that possible

candidates for an answer that passed soft filtering need to be more carefully examined, because one of them will be the answer. In this step Watson also begins to gather more and more evidence to support an answer by running multiple deep sorting analytics. When all evidence is gathered, they need to be run through algorithms to score their credibility. Watson has over fifty different scoring techniques working in parallel, implemented in order to score the generated responses. [4]

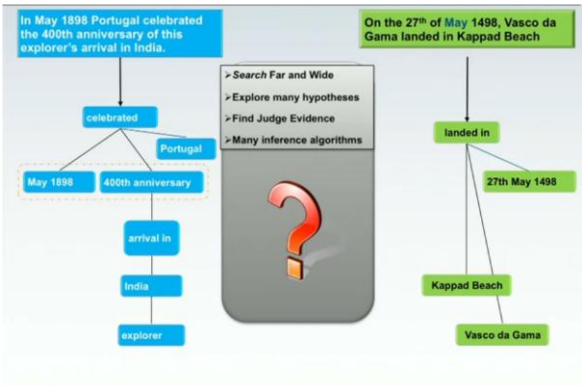

Figure 6. An example of breaking down the question: "In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India."

5.6 Final Merging and Ranking

This is the final step which Watson performs before attempting to answer the question. DeepQA merges selected answers from scoring in the previous step and runs deeper, more complex confidence analysis in order to create a confidence level for a response. Depending on the confidence level of an answer, Watson choses to either buzz-in and attempt to answer or not attempt to answer.

6. STRENGTHS AND WEAKNESSES

6.1 Strengths

Watson's main advantage is that he is a computer, and as a computer he cannot forget anything that is made available to him. That is, all the information provided to the supercomputer in the form of encyclopedias, articles, etc. can always be accessed and referenced to in an attempt to answer a question. A human brain does not have the capacity to operate at such a high level as Watson in this regard. Another strength of Watson as a computer is that a computer has no emotions. A human player can become unconfident in future questions if he/she answers a previous question wrong. This however, is of no importance to Watson because the computer attempts to answer every question in the same manner, regardless of the previous result. Finally, another of Watson's strengths is that the supercomputer is a statistical genius. Every operation performed by the computer is a carefully calculated statistic. Every word in a clue is being bisected by Watson to find every possible answer/combination to the given question, and at the end come up with the one that is statistically best. Even though this is very similar to how the human brain works, the rate at which Watson is capable of operating is much greater than the rate at which the human brain is capable of. [11]

6.2 Weaknesses

Being a computer is not always strength, it can also be counted as a weakness. Watson has not lived or experienced the real world, and thus is unable to draw from real-life experiences unlike humans can. In the event of answering sense-related questions Watson can have trouble answering the question due to the fact that he is a computer. Watson, like humans is capable of learning, but in the event that his algorithm is doing something wrong, Watson will not be able to self-correct himself for future questions (Watson does not care if his answer was right or wrong). Another weakness is that Watson is unable to time his buzz-in time. Humans can see when the host is about to stop reading the question, and thus be allowed to buzz-in. Watson on the other hand can only buzz-in when his software allows him to do so. A human player can know the answer to a question before the host has stopped reading the clues (Watson might too), and time his buzz-in time perfectly with the time when the host finishes the reading and buzzing in. This is an advantage to the human players. [11]

7. FUTURE PROSPECTS FOR WATSON

After Watson's win on Jeopardy, many ideas have come up for Watson's future in our society. It was clear that IBM have developed an astonishing piece of equipment that was able to beat human players in a game of Jeopardy, and that Watson is a giant step forward for Artificial Intelligence.

7.1 Search Engines

One idea that has been floating around after Watson's win is to use its powerful software to improve current search engines. Watson's ability to dissect a question in the game of Jeopardy can mean that it can lead to better search results for humans. Given that a human is searching for something he/she is uncertain about, Watson can use his intelligence to make an accurate prediction to what the user is searching about. In the end, we would receive much more accurate and detailed search results.

7.2 Business Analysis

Watson's ability to recall enormous amounts of data can be of critical importance to business executives and brokers, when they are making critical decisions. Since Watson cannot forget anything, can access and process large quantities of data, the supercomputer can help make projections for the future based on solid facts from the past. This could mean that Watson could possibly prevent future economic meltdowns like the one we recently had in 2008. [12]

7.3 Physician's Assistant

Recently, IBM and the University of Maryland have teamed up to work on bringing Watson to our doctor's office. Watson's question answering skills and large amounts of reference knowledge could mean that Watson can be used to identify potential patient health problems that a physician might miss. Watson could also be used to confirm or dispute a diagnosis given by physicians. In another example given, Watson could possibly be used for prescribing medications and in determining proper treatment options for various different patients. Watson's up to date knowledge on all medical articles could help a physician make a decision on what kind of medication to prescribe, or what kind of alternative treatment might be available. [9]

8. CONCLUSION

It is evidently clear that Watson is a result of an extraordinary piece of engineering in part of the IBM team who developed the supercomputer. Taking on the challenge of beating the best contestants in Jeopardy and succeeding is a giant step forward for Artificial Intelligence. All of the research performed in order to make Watson a reality can be used in future computing projects in order to make an even better and smarter supercomputer. With the endless possibilities to make Watson part of our society, we can live healthier and safer lives.

9. REFERENCES

[1] IBM Watson - http://www-03.ibm.com/innovation/us/watson/

[2] IBM`s Watson: 16 Things You May Not Know About Its Hardware Muscle - http://www.eweek.com/c/a/IT-Infrastructure/IBMs-Watson-16-Things-You-May-Not-Know-About-Its-Hardware-Muscle-491431/

[3] PBS, IBM's Watson: Documentary - http://www.youtube.com/results?search_query=IBM%27s+Watson%3A+Documentary+&aq=f

[4] AI Magazine. Building Watson: An Overview of the DeepQA Project. By David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty

[5] What's the system architecture of the IBM Watson? - http://www.quora.com/IBM-Watson/Whats-the-system-architecture-of-the-IBM-Watson

[6] Building Watson - A Brief Overview of the DeepQA Project - http://www.youtube.com/watch?v=3G2H3DZ8rNc

[7] What makes IBM's Watson run? - http://www.zdnet.com/blog/open-source/what-makes-ibms-watson-run/8208

[8] An Inside Look At the Power Behind IBM Watson - http://www.theinfoboom.com/articles/an-inside-look-at-the-power-behind-ibm-watson/

[9] Watson in the examining room? - http://articles.baltimoresun.com/2011-02-17/health/bs.md-watson-medicine-20110217_1_watson-medical-privacy-patient

[10] IBM Watson: The Face of Watson - http://www.youtube.com/watch?v=WIKM732oEek

[11] 3 Strengths and Weaknesses of the Watson Computer on 'Jeopardy!' - http://www.livescience.com/12859-3-strengths-weaknesses-watson-computer-jeopardy.html

[12] IBM: Watson could boost confidence of financial industry - http://www.computerworld.com/s/article/9215481/IBM_Watson_could_boost_confidence_of_financial_industry