English

Frank Born

The role of language in the development of AI systems

Bachelor Thesis

GRIN

Bibliographic information published by the German National Library:

The German National Library lists this publication in the National Bibliography; detailed bibliographic data are available on the Internet at http://dnb.dnb.de .

Imprint:

Copyright © 2011 GRIN Verlag
ISBN: 9783668178984

This book at GRIN:

https://www.grin.com/document/318595

Frank Born

The role of language in the development of AI systems

GRIN Verlag

UNIVERSITÄT KASSEL
FACHBEREICH 02 GEISTES- UND KULTURWISSENSCHAFTEN

The Role of Language in the Development of AI Systems

Bachelor Arbeit

Frank Born
09.08.2011

Studiengang: English and American Culture and Business Studies

Content

1. Introduction

Finding relevant information in the vast growing pool of sources is a challenging task. People are confronted with libraries full of books, transcripts, magazines, and numerous other documents; they also have the ability to use local databases, intranets, and the World Wide Web. The Internet gives users access to an incredibly large amount of information, consisting of websites, emails, blogs, eBooks, newspapers, magazines, and so on. This ever growing flood of information can be useful, but often it is overwhelming. For instance, writing a paper allows a person to investigate a specific topic, but the amount of sources available is simply too large. Therefore, only a small portion of all available texts can be investigated and potentially important information may be missed over. Artificial intelligent systems are needed as a tool to find and evaluate useful information. In order to create helpful tools, AI systems need to understand natural language.

The role of language in the development of artificial intelligent systems envelops a broad spectrum of areas. Natural language consists of many facets, has developed over a long time span, and is consistently shifting and changing. Human beings use language ambiguously. The result is a vast amount of overlapping and a large number of possible interpretations of different texts. The development of faster information technologies (e.g. telephone, email, Internet…) catalyzes the expansion of the varieties of natural language data. Therefore, artificial intelligent systems are necessary to use as well as evaluate natural language and make them accessible to people.

AI is a fascinating topic that not only baffled but also inspired the minds of philosophers, scientists, technicians, and even movie-makers. Humans are intrigued by the topic, because the investigation of AI allows people to decipher the complexity of their own minds. By trying to create AI systems people create an insight into human intelligence as well as are able to understand the world that is surrounding them. New technologies enhance the possibilities to further meet human demands concerning computational systems. This reflects the ability of the systems to understand and analyze unstructured text. The largest part of information available is written in unstructured natural language. AI systems can be used to make different types of information available for present day demands.

There are many approaches to advance AI systems. IBM (International Business Machines) has a long history in computation and creates communal interest for the quest of computational advancement with the help of public events. The creation of Deep Blue and the defeat of Garry Kasparov in 1997 in chess signaled a giant step forward in software

engineering. In 2011, IBM developed a computational program, called Watson, to try to defeat the best human players in the game show *Jeopardy!*. The system is able to understand *Jeopardy!* questions which consist of natural language clues. This achievement marks a leap forward in natural language processing (NLP). Natural language is an important factor that will influence the development of AI systems.

In addition, other aspects must be taken into consideration. Computational systems have to show the ability to use learned knowledge. However, what distinguishes computational systems from artificial systems is the potential to be original. The idea behind this investigation is to understand the difference between human and artificial intelligence.

Thoroughly examining Watson will reveal the similarities and differences of the way humans and computational systems understand natural language. This will create insight into the potential and further development of the AI systems. Natural language processing systems have a broad field of applications. The demand of these systems becomes instantly apparent, when investigating various industries such as financial services, call centers, and the medical industry.

Nevertheless, Watson is not the only research program that will influence the future of society. Various smaller software programs will benefit and advance the current development. Also, knowledge representation will have an impact in areas such as the World Wide Web.

One important aspect that should be considered when analyzing projects like Watson is the opportunities that arise with it. In the Art of War, Sun Tzu states: “Know thine enemy better than one knows thyself”. Investigating the *Jeopardy!* challenge characterizes the battle between man and machine. This leads to the conclusion that understanding Watson allows to look at aspects of human intelligence that are still unraveled.

2. IBM's History and the Development of Watson

On February 14, 2011, IBM took on the newest challenge in computer development and natural language processing. Watson, a "supercomputer", was built to compete in *Jeopardy!* against the two most successful *Jeopardy!* players in the United States; Ken Jennings and Brad Rutter. This; however, was not IBM's first challenge. IBM's history can be traced back to the turn of the 19th century. In the 1950s, IBM developed the 701 computers which enabled Convair to create the Atlas missile. The further advanced version, the 704 computer, was applied to satellite tracking. The influence of IBM in the space program is significant, if we consider the various applications of the IBM 704 in the Jupiter missile or considering the capability of two 704 computers to track the Soviet Union's satellite Sputnik. When Explorer I was launched in 1958 an IBM 705 computer was used for guidance and support. This computer was capable of calculating and making more than 1.3 Million logical decisions per minute. Still, this was just the beginning. IBM units are included in the first Apollo missions and the System/360 Model 75s are able to receive and send information to the home base almost instantly. This is an essential leap forward to solve the next problem which was to send a human being to the moon. The National Aeronautics and Space Administration (NASA) and IBM worked closely together and were successful with landing Apollo 11 on the moon ("Space Flight Chronology"). "On July 20, 1969, the human race accomplished its single greatest technological achievement of all time when a human first set foot on another celestial body" (Garber).

Even though technology made a giant leap forward in the 1960's people considered computers far from being able to fulfill human tasks. The thought that computers would one day be able to defeat chess grandmasters was not imaginable. Chess was believed to be a genuine game of human intelligence and no computer would be able imitate that intelligence. In 1997, this was definitely still true; however, the fact that computers cannot defeat chess grandmasters was distorted. After Deep Blue's victory over Garry Kasparov the opinions about the future of chess differed. Deep Blue was now able to defeat the world grand chess champion. This, essentially, infers two possibilities; IBM's Deep Blue is superior to human thought and will be soon greater than human intelligence, and on the other hand, Deep Blue's superiority towards Garry Kasparov can be explained as a memorization advantage. One of the most prominent chess champion Bobby Fischer talks in 2006 about the future of chess and states that "Memorisation is enormously powerful.…It is all just memorisation and prearrangement. It's a terrible game now. Very uncreative" (Chessbase News). Also,

Cognitive Scientist and Pulitzer prize-winner Douglas Hofstadter declares in the *Washington Post*: "My God, I used to think chess required thought. Now I realize it doesn't. It doesn't mean Kasparov isn't a deep thinker, just that you can bypass deep thinking in playing chess" (qtd. in Krauthammer). Subsequently, if chess has to be reconsidered as a genuine game of human intelligence then IBM has achieved an enormous success with Deep Blue. Even though there is a lot of discussion about the impact of Deep Blue it is to say that IBM managed to build a machine that was able to defeat a chess grand champion.

IBM also started a research project on understanding biological processes. In 2008, Blue Gene/L was placed on the list of the top five hundred supercomputers and is used to decipher the human Gene Code. Other applications are "hydrodynamics, quantum chemistry, molecular dynamics, climate modeling and financial modeling" ("Blue Gene").

After all of these developments and achievements IBM set a new target with Watson. Natural language is an entirely different and far more complex subject matter. It seemed that for a long time people did not consider language as a difficult issue for machines to deal with. However, the more that the intention shifted towards NLP, the result became more of the complete opposite.

People communicate in an organic structure. Language evolved over many thousands of years and has not stopped since - especially the numerous varieties of facets in a language. Computers have been used for the longest time for processing, calculating, and searching. Yet, applications such as search are not comparable to open question answering. The difficult task for a computer is to understand human communication. The reason being is that people naturally communicate on different levels. Leader of the Semantic Analysis and Integration Department at IBM's T.J. Watson's Research Center Dr. David Ferrucci says that human beings are:

> very fluently in images, in literature, in writing.…people get natural language because it's a human artifact, they relate those words and those phrases and those ideas back to the way they think.…they ground that information in human cognition, in human experience.…but that is not written in a formal data based language or a formal mathematical language that computers can understand. ("The Next Grand Challenge")

Computers have difficulties understanding the ambiguous meanings in human language. Natural language includes numerous examples of fuzzy concepts, for example "small" or "weak". It is fairly difficult to distinguish and decide if somebody counts as strong or weak. The range is broad and borderline cases are abundant. "Fuzziness of this kind is characteristic

of the human conceptual system" (Bieswanger and Becker 149). Meaning of words can be arbitrary; nevertheless, human beings usually speak, write, and communicate in phrases and sentences. Sentential semantics is a challenge for computers, because "the meaning of a phrase or a sentence is determined by the meaning of its component parts and the way they are combined structurally" (Bieswanger and Becker 151), and following the principle of compositionality. Understanding this concept is especially difficult, when it comes to sentence interpretation. The two essential components of a sentence are syntax and lexical semantics. The syntactic structure of a sentence determines the way one understands a sentence, if the order of the words in a sentence changes, the meaning of this particular sentence can be altered as well.

Example:

(1) John is taller than James.

(2) James is taller than John.

Both sentences are syntactically equivalent; though, it completely changed the semantic of the sentence. In this case the subject and the object are exchanged. Nevertheless, there are sentences in which the word order is unchanged but the sentences are structurally ambiguous, e.g.: John hit the boy with the umbrella.

(1) John chased [the boy with the umbrella].

(2) John chased [the boy] with the umbrella.

Sentence (1) indicates that John chased the boy who was holding the umbrella, whereas sentence (2) states that John was holding the umbrella while chasing the boy (Bieswanger and Becker 154-156). It seems easy for human beings to distinguish these two variants, considering people know the context. Conversely, this is an immensely difficult task for a computer, which lacks on semantic recognizing capabilities.

After Deep Blue, IBM's *Jeopardy!* Challenge is a new step to advance computer technology. The core of the research is not simply to win the game but to improve and make computers more compatible with NLP. Ferrucci understands that the development of Watson "is irresistible to pursue…because as we pursue understanding natural language we pursue the heart of what we think when we think of human intelligence" ("The Next Grand Challenge").

3. *Jeopardy!* and the Potential of QA systems

IBM considered *Jeopardy!* as an appropriate way to test their Watson project, because the structure of the game allows one to advance in question answering technology as well as to create public interest. The current layout of the *Jeopardy!* quiz show has been on television in the United States for over twenty-five years. Since 1984, a broad spectrum of general knowledge questions is asked to three contestants in a three round system. The game includes three important key factors that the contestants must possess which are knowledge, confidence, and speed. In the first two rounds thirty questions are divided into six categories. The dollar values of the questions of the first round ranges from two hundred to one thousand. In the second round they range from four hundred to two thousand. The first player selects a category and a corresponding dollar value question. The question appears and each player can buzz-in to answer the question after the host finished reading the question and a light enables the players to use their signaling devices. The fastest player gets the chance to answer the question, but again speed, knowledge, and confidence are important. The question is phrased as a clue which has to be responded by a question. A player gains the dollar value when his response is correct, otherwise it would be deducted from his total dollar amount. After a player answers a question incorrectly the other players are given the chance to answer the clue. One important aspect of the game is the chance to get the Daily Double. The position of this field is hidden, but if a player uncovers the Daily Double he/she or Watson has to answer the clue and bet a dollar amount from his/her/its total score. The player either wins or loses this amount depending on whether the answer is correct or not. There is one Daily Double in the first round and two in the second. In the final *Jeopardy!* round a category and a clue are revealed. The contestants have thirty seconds to write down their answer and an amount that they want to wage. The player with the most money at the end wins the game (Ferrucci et al. 61).

The host of *Jeopardy!* Alex Trebek explains that the IBM Challenge works in the same fashion except that: “Watson will receive the clues electronically as a text file at the same moment the clues are revealed to Ken and Brad and at the same time I read them. This competition will be a two game total point exhibition match; however, these two games will be played out over the next three days so we can tell the full story” ("Jeopardy! - The IBM Challenge").

The executive producer of *Jeopardy!* Harry Friedman summarizes why this game is so interesting for computer scientists: “*Jeopardy!* really represents natural language. You have to

understand the English language and all the nuances and all the regionalisms and the slang and the short hand to play the game, to get the clues. It's not just a piece of information" ("Why Jeopardy!?").

Watson can help in the development and organization of the Internet and all other kinds of media. The amount of information is increasing dramatically. Therefore, a system is necessary to extract knowledge from a very large repertoire of information. Consequently, Watson is not simply built to play games, but is a system that can help organizing the exponential rising amount of information. One very interesting aspect is the distinction between knowledge acquired by a human being over time versus knowledge a computer system can produce.

Testing these new computer systems (such as Watson in *Jeopardy!*) allows computer developers to experiment and apply new NLP approaches. Even though the results of the Watson project might be valuable for further research, one must not forget that IBM uses *Jeopardy!* as a way to set its own deadlines and goals, and also as a Marketing platform. New developments in NLP and AI are made accessible to a broad public audience. However, behind the *façade* the advantages of a question answering system will leap forward the technology industry. This is another step for computers to interact with humans and probably a way for machines to finally pass other challenges such as the Turing test. Before getting there it is important to understand that "there is an enormous amount of science involved when Watson answers a single *Jeopardy!* question. There is natural language processing, there is machine learning, there is knowledge representation and reasoning, there is deep analytics and it all happens in 3 seconds" ("A System Designed for Answers").

Watson's speed can be explained by computation that is set parallel. IBM uses a computing infrastructure that allows the processors to do many parallel computations at the same time. Senior Vice President and Director of IBM Research Dr. John E. Kelly III. asserts that "the POWER7 system is tuned for very rapid deep analytics of massively parallel problems" ("A System Designed for Answers"). A single central processing unit (CPU) would not be able to process the information given in a *Jeopardy!* clue fast enough to compete against world class *Jeopardy!* contestants. In order "to delivering a single, precise answer to a question requires custom algorithms, terabytes of storage and thousands of POWER7 computing cores working in a massively parallel system" ("A System Designed for Answers").

4. Watson's Appearance

4.1. Watson's Voice

When Watson appeared in the *Jeopardy!* show the creators of Watson thought about different components to make the system more accessible to the audience. Deep Blue was just a computer and his moves were played by a human, whereas Watson received an actual contestant like appearance, including an Avatar with a body/face as well as a voice.

The first approach was a visual read-out of the answer; nonetheless, Andy Aaron, who is responsible for Watson Speech at IBM Research, explains that the use of text-to-speech software would be very useful for shorter phrases, which are mainly used in *Jeopardy!*. Finally, they came to the conclusion that human voice actors have to be auditioned. Countless words and phrases from the English language as well as from various other languages were added to the system. For example, the Latin-American dish *Arroz con pollo* (Rice with chicken) had to be described into the following symbols: '[Oa1ros]'[kcn]'[1po0yo]. The encoded text is equivalent to sounds of the English language. This technique enables Watson to pronounce almost every word in a proper manner. This went so far that numerous people from various countries were interviewed to enable Watson to use the proper pronunciation of particular words. These words where then encoded with particular symbols.

Examples:

(1) édouard '[0eldwar]
(2) Zinzendorf '[ltsln0sln0darf]
(3) Zoe '[lzo0i]
(4) Xinhui '[lSIn]'[lhwe]

The research was not only considered to be used for *Jeopardy!* but for various other applications. The program now contains several thousand words and phrases which can and will be used in future research ("The Face of Watson").

4.2. Watson's Visual Appearance

The visual appearance of Watson went through several considerations. Options such as displaying Watson as a human or in an abstract way were considered. Finally, the decision was made that Watson would be represented by IBM's Smarter Planet logo. The Generative Artist Joshua Davis used variables, put boundaries in position, and was able to create a

sphere. After different approaches Davis decided to generate one leader that would be surrounding the planet icon which would then be followed by forty chasers. If Watson has a high level of confidence the leader and the followers swarm to the top and circle around it. If Watson gets an answer wrong the leader and chasers move to the bottom. In total Davis allows Watson to have 27 various states. An example of this is: Daily Double in which case the leader and the followers go to the top and Watson lights up in a colorful mix of green, white, blue, and light blue. A score loss is characterized by a color scheme consisting of orange, red, and yellow and in which case the chasers follow the leader to the bottom of the planet icon. Score Gain consists of four different greens from dark to light green and the leaders and followers can be seen at the top again. A few other examples are: answered, answering, answered correct, answer revealed, answer wrong, buzzer enabled, buzzer timeout, category selected, and clue revealed. One of the most important aspects that a person should possess for this game is confidence. Therefore, Davis selected four different colors to illustrate four different levels of confidence, whereas, green symbolizes a very high level of confidence ("The Face of Watson").

4.3. Watson's Answer Panel

Watson generates many possible answers for each clue by using a vast amount of different algorithms. All of the possible answers will be narrowed down and ranked accordingly to the confidence level. IBM "researchers added a buzz threshold indicator to Watson's answer panel. This vertical white line shows the minimum level of confidence Watson's top answer must meet in order to trigger a buzz" ("Watson as a competitor"). That means even though Watson generates the correct answer it will not attempt to answer the question. In various situations during the *Jeopardy!* challenge Watson was able to generate a high level of confidence, however, the system was a few seconds too slow to answer the question ("Watson as a competitor"). For instance, in the category "Actors Who Direct" on the last day of the competition Watson was able to generate the right answer with a high confidence for every question, but lost against the human competitors in speed. The answer panel has been used by researchers from the beginning and it allows them to understand as well as analyze Watson's responses, which is essential to improve the system.

5. Aspects of Artificial Intelligence

5.1. Definition of Artificial Intelligence

In order to understand the role of language in the development of artificial intelligent systems it is essential to understand what artificial intelligence means. Philosophers and scientist in various fields came up with countless definitions of what AI actually entails. Yet, the result is that there is no coherent definition. The history of AI will make this phenomenon more accessible and understandable. In human imagination, machines were able for a long time to perform intellectual tasks. *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities* by the English mathematician and philosopher George Boole was one of the first achievements in the nineteenth century which dealt with algebraic logic and, consequently, influenced the ideas behind AI.

The history of artificial intelligence and computers are ultimately linked with each other. The English inventor and mechanical engineer Charles Babbage designed the first "calculating engine". Even though his machine was never completed and was mechanical, it used certain ideas which can be found in present-day computers. In the 1940s and '50s, an increase occurred in the development of electronic computers. The design of ENIAC (Electronic Numerical Integrator and Computer) and later UNIVAC (Universal Automatic Computer) were one of the first important calculating computers. After the success of UNIVAC many companies like IBM saw the potential of computers in various industries. The development of the integrated circuit and later the microprocessor allowed the miniaturization of the computer. The development of software is an essential step for the creation of artificial intelligent systems (Jaffe Productions). Until the 1950s the term artificial intelligence was not used frequently. This changed when the American computer scientist John McCarthy introduced the term in 1955 (Partridge and Hussain 3).

In 1979, Hofstadter stated in his book *Gödel, Escher, Bach: an Eternal Golden Braid* that "one could define AI as coming into existence at the moment when mechanical devices took over any tasks previously performable only by human minds" (601). AI is a flexible term, because the development of new programs and computational systems is continuous and rapid. Therefore, Hofstadter's 32 year old definition is applicable today and is supported by a theorem by computer scientist Larry Tesler which declares that: "AI is whatever hasn't been done yet" (qtd. in Hofstadter 601).

At a point where intelligent computer programs will be able to program and reinvent themselves could be perceived as the creation of AI. American inventor and author Ray

Kurzweil talks in his book *The Singularity Is Near* and in the documentary film *Transcendent Man* about a new age in the development of AI technologies which he calls “Singularity”. Kurzweil defines Singularity as “a future period in which technological change will be so rapid and its impact so profound that every aspect of human life will be irreversibly transformed” (Ptolemaic Productions). The development of technology will result into a stage where there will be a fusion of technological and biological intelligence. The increase of information and the technological development began rather slowly. For instance, the development of the first stone tools occurred ten thousand years ago. During the dark ages the development of new ideas almost stagnated in Europe for almost one thousand years. However, the discovery of the Americas and advances in science as well as technology that followed five hundred years ago, exemplified various improvements. About 150 years ago the discovery of electricity, the usage of the first telephones, and the development of trains had an enormous impact on people’s lives. The rapid developments of the last fifty years including, faster transportation systems, developments of the computer, and the Internet increased the amount of information dramatically. Right now, vast changes in technology and information system occur within six months.

The reason for this fast increase over the last few years can be explained by the fact that the newest technology is used to develop the next technology. Therefore, information technology accelerates over time and rises exponentially. Kurzweil suggests that in the next forty years the acceleration will overtake human understanding and that people have to improve their own intelligence in order to keep up with the development (Ptolemaic Productions). Whether these predictions are reliable is indistinct. Nonetheless, in order to achieve this step of development it is important to understand the idea behind artificial intelligent systems and the way they operate.

5.2. AI and Recursion

Hofstadter sees the answer of understanding AI in the foundation of recursion. This approach seems plausible, because human beings are reluctantly affected by recursion. What needs to be explored is, if artificial intelligent systems are applicable to the principle of recursion. In human cognition, recursion is characterized by nesting, and variations of nesting. Computational programs could benefit and could be enhanced drastically with the use of this principle. Modularization is another aspect that needs consideration. It is used in order to split certain tasks into natural subtasks. This application finds interest and is indeed very useful in

computer science in which a loop enables the computer to perform fixed tasks. The loop then is directed back to perform these operations again. Therefore, loops can be used to "perform some series of related steps over and over, and abort the process when specific conditions are met" (Hofstadter 149). When these fixed rules are applied to standard processes, recursive enumeration could be able to create completely new modes of applications. In human cognition this process can be seen in natural language. People use a set of words, which is defined by the vocabulary of an individual and a fixed set of rules, which allows an individual to use specific grammar to make his utterances understandable. Therefore, humans are able to create a vast amount of sentence structures without necessarily learning these word formation structures deliberately. This allows the human mind to express completely new ideas, because recursion entails an increasing complexity by using a defined sequence of processes.

Recursion applied to a computer program would allow increasing the programs complexity and making it gradually more unpredictable. Complex recursive systems are able to create novel ideas, so if this were to be applied to computational systems, it would be able to overcome their predetermined rules and processes which is the foundation of intelligence. This genuine process of an artificial intelligent system can result in the creation of novel ideas, in the improvement and innovation of itself.

AI is a very broad field including automatic programming, decision-making, natural language processing, pattern, and speech recognition. The development of AI systems lets us understand human intelligence better. Especially, the way humans think when they speak. Computational language systems are nothing like humans. It has to be considered that the creation of AI systems does not have to be equivalent to human intelligence.

5.3. AI and Problem Reduction

Problem reduction can be a conscious or an unconscious process for the human mind. In our society, humans are trained to follow certain steps to achieve their targets. People can think ahead and plan how they want to solve a specific problem. An individual can easily understand that in order to get from A to E, there are several steps (B, C, D) between A and E that need to be solved first. For example, to get to the grocery store across the street an individual has to go from their apartment (A), to the door and open it (B), go down stairs (C), cross the street (D), and enter the grocery store (E). B, C, and D are sub-problems of the actual problem E. These sub-problems also include sub-sub-problems, e.g. unlocking the door (B1) or locking the door (B2). For humans most of these processes are unconsciously and

fairly simple to handle. Animals, in comparison, have difficulties using problem reduction. Most dogs, for instance, have difficulties fetching a bone that landed behind a fence. Most likely, dogs can see and smell the bone, but it does not occur to them that they could use the open gate twenty feet down the fence. In order to take the detour they would have to distance themselves from the bone that is only five feet away. Animals have difficulties breaking a problem into sub-problems. Consequently, things that seem particularly easy for a person, e.g. to go from one side of the fence (A), through the open gate (B) to the bone (C), is very hard to solve for animals, e.g. to go from one side of the fence (A) to go to the other side of the fence (B) (Hofstadter 611).

In order to make a program understand the foundation of sub-goals and sub-sub-goals a technique has to be used that converts main problems into smaller units. Recursion and problem reduction enable computer programs to process specific goals. Nevertheless, the difficulty is to make a system understand and adjust itself to the main problem, meaning that the system has to define its own sub-goals and sub-sub-goals. Hofstadter's approach at solving this problem involves reducing the problem with the help of a "forward motion towards the overall goal" (611) and magnifying the problem with the help of a "backward motion away from the goal" (611). This method includes a usage of various perception diameters which enables a program to define the problem precisely. As well as humans learn from their experience and become more efficient in their judgment, artificial systems are also able to use experience and improve over time by learning rules and procedures. Hofstadter's idea of conceptual space illustrates two possibilities to solve problems:

> (1) try moving away from the goal in some sort of random way, hoping that you may come upon a hidden "gate" through which you can pass….
> (2) try to find a new "space" in which you can represent the problem, and in which there is no abstract fence separating you from your goal – then you can proceed straight towards the goal in the new space. (612)

5.4. AI and Human Intelligence

The human brain processes data in various ways. Intensively thinking about an issue is only one possibility. Occasionally, the brain needs time to reflect on problems. These reflections often occur over night during sleep. In order for a human brain to function effectively it requires sleep which "remains the most reliable predicator of wake state stability

and neurobehavioral functioning" (Minkel, Banks and Dinges 248). A well functioning brain is the crux of intelligence. The processes that occur in the human brain are very complex and completely different than the processes in a computer system. A person and a computer might be able to come up with the same answer in a similar speed. There are various examples in the *Jeopardy!* challenge in which one of the contestants can answer faster or at least at the same speed as Watson. Watson's answers are reflected in the answer panel. In some categories humans can answer a question in less than a second whereas Watson needs two or three seconds. In other categories the opposite can be observed. Nevertheless, the way both participants come up with the same answer is completely different. This could be the point where human intelligence differentiates itself from artificial intelligence. AI systems have difficulties with a technique such as problem reduction, because reducing and magnifying problems is a completely different way of processing data for a program.

In order for computers to process several tasks at once, it needs a huge amount of capacity. Watson for example, is comprised of two units, each consisting of five separate racks and ten IBM Power 750 servers, which is equivalent to 2800 powerful computers that work in a high-speed network ("Jeopardy! - The IBM Challenge").

Apart from Watson, there are other computational systems that are very valuable as research projects. For instance, the Mars Rover "Opportunity", which is currently investigating the surface of the Mars, is programmed to drive independently and to calculate its own route; thereby, making judgments about stone formations that are worth investigating. The idea that a machine can make judgments regarding what it is interested in is a very important factor in Hofstadter's description of artificially intelligent programs.

He understands intelligent programs as versatile systems that are able to solve various problems and by doing so AI systems gain experience that enables them to improve and solve other, more complex problems. This includes that an AI system is able to use the current set of rules and modifies them to benefit its interest (613).

One must be aware that human intelligence is not necessarily equivalent to artificial intelligence. Hofstadter's assumption seems plausible. Nevertheless, this approach is very broad and can be applied to both forms of intelligence or even to intelligence in general:

> The flexibility of intelligence comes from the enormous number of different rules, and levels of rules.…Strange Loops involving rules that change themselves, directly or indirectly, are at the core of intelligence. Sometimes the complexity of our minds seems so overwhelming that one feels that there can be no solution to the problem of understanding intelligence. (27)

A difference between a complex system and an intelligent system can be distinguished by looking at the example of Deep Blue. In 1997 Deep Blue used a high variety of skill sets and memorization to defeat Kasparov. Deep Blue was another stepping stone to develop an artificial intelligent system. Professor of computer science Monty Newborn describes in his article "Deep Blue's contribution to AI":

> Deep Blue combined the search algorithms…refined by numerous computer scientists. It incorporated hardware advances….Deep Blue carried out a parallel search of a chess tree using techniques developed and refined over the final two decades of the century, beginning in 1982….Other parallel systems followed….Deep Blue had knowledge elegantly crafted into its evaluation function….And deep down inside Deep Blue were the endgame databases…that Kasparov knew would play perfect chess if the game ever touched upon the positions they contained. (27)

The system entailed components of intelligence such as learning. Deep Blue went beyond implementing rules and patterns. It developed its own rules and learned from experience. The significance for AI was that "at the most fundamental level Deep Blue's achievement provoked considerable thought on the subject of what intelligence is all about" (27).

5.5. Computers and Learning

There are different approaches to learning, one is based on observation. Visual imagery is important to understand the environment. Humans tend to connect visual imagery to language and this allows people to define their environment. Attempts have been undertaken to apply visual imagery to computational systems. NASA's Mars exploration rover, "Opportunity", is able to locate new stone formations with its wide-angle navigation camera and decide whether it is worth investigating. "Opportunity's" system, AEGIS (Autonomous Exploration for Gathering Increased Science), enables the rover to choose its own route around obstacles, can maneuver its mechanical arm independently, and select its own targets. All this is possible with the help of the AEGIS software which enables "Opportunity" to recognize visual patterns. The visual pattern recognition software differentiates between important stone formations, such as dark and angular objects, from lower priority objects, like light and rounded rocks. Visual imagery helps in aiding certain components of intelligence such as memorization and categorization (Webster).

Another important aspect in regards to language is partitions. Partitions are used to categorize and separate words with similar connotation in multiple languages. Weak partitions

allow overlapping, whereas, strong partitions enable software to work and deal with translations more effectively (Hofstadter 671). All of these components advance people and could; therefore, enable programs to advance in knowledge processing.

5.6. Knowledge and AI

Knowledge is an essential component of intelligence. The way people store and use knowledge is fundamentally different from computational systems today. There are different approaches to knowledge; one being declarative knowledge which is stored in specific places within a program. Modular and non-modular knowledge are two types of ideas that are applied to computer programs. Modular knowledge is a concept in which several modules contain a certain set of rules independently from each other. Non-modular knowledge allows a program to access the different modules and connect them with each other. In addition, accessibility of data is a critical aspect of computational programs. Even though knowledge gained over time by people is completely different from knowledge used by a computer, accessibility in both cases can be compromised. The accessibility of knowledge in a human mind can be compromised by stress, lack of sleep, or complexity of a task; while, a computer can have problems accessing data because of programming errors or overlapping of demands. The issue is that data or information can be part of the active memory or the passive memory. Active memory can be accessed directly, whereas, passive memory is knowledge that is temporarily inaccessible. The issue with knowledge is that it "is not made up of Lego-like building blocks but is a matter of skill and learning" (Janik 54). Knowledge has to be stored, linked, and accessed. Knowledge is not a fixed term; it is always shifting and changing.

Heuristic processing is one way to describe the method people use in broadening their knowledge, which is also an important indicator for computer science. The foundation lies in discovering novel ideas. American psychologist Dr. Clark Moustakas describes the influence on human researchers as follows:

> It refers to a process of internal search through which one discovers the nature and meaning of experience and develops methods and procedures for further investigation and analysis. The self of the researcher is present throughout the process and, while understanding the phenomenon with increasing depth, the researcher also experiences growing self-awareness and self-knowledge. Heuristic processes incorporate creative self-processes and self-discoveries. (9)

This can also be applied to AI systems. Systems like Watson use techniques in its database to find evidence to confirm its hypotheses. By answering more and more clues the system learns and can improve its understanding of an increasing number of questions. Nevertheless, current computational systems are still lacking self-awareness which differentiates them from potentially intelligent systems.

5.7. AI and Natural Language

"Knowing a word involves both knowing the pronunciation and the meaning of the word" (Bieswanger and Becker 142). This is one of the difficulties of the human language. The Swiss linguist Ferdinand de Saussure developed a model of the linguistic sign which includes that words consist of a *signifié* and a *significant*. In natural language, there is a vast amount of words that have the same sound pattern but a different concept. Words with the same sound pattern can be unrelated or related. For example, related words (polysemous words) are a frequent occurrence in natural language. In order to illustrate this phenomenon, dictionaries like *Oxford Advanced Learner's Dictionary* give many examples. The word "center" is definite in nine different ways:

> **centre** (*BrE*)(*NAmE* **center**)/'sentə(r)/ noun, verb
>
> ***-noun***
>
> **MIDDLE 1** [C] the middle point or part of sth: *the centre of a circle* ◊ *a long table in the centre of the room* ◊ *chocolates with soft centres*-picture → CIRCLE
>
> **TOWN/CITY 2** [C] (*especially BrE*) (*NAmE* usually **downtown** [usually sing.]) the main part of a town or city where there are a lot of shops/stores and offices: *in the* ***town/city centre*** ◊ *the centre of town* ◊ *a town-centre car park* **3** [C] a place or an area where a lot of people live; a place where a lot of business or cultural activity takes place: *major urban/industrial centres* ◊ *a centre of population* ◊ *Small towns in South India serve as economic and cultural centres for the surrounding villages.*
>
> **BUILDING 4** [C] a building or place used for a particular purpose or activity: *a* ***shopping/sports/leisure/community centre*** ◊ *the Centre for Policy Studies*
>
> **OF EXCELLENCE 5** [C] **~ of excellence** a place where a particular kind of work is done extremely well
>
> **OF ATTENTION 6** [C, usually sing.] the point towards which people direct their attention: *Children like to be the* ***centre of attention***. ◊ *The prime minister is* ***at the centre of*** *a political row over leaked Cabinet documents.*

-CENTRED 7 (in adjectives) having the thing mentioned as the most important feature or centre of attention: *a child-centred approach to teaching-see also SELF-CENTRED*

IN POLITICS 8 (usually **the centre**) [sing.] a MODERATE (= middle) political position or party, between the extremes of LEFT-WING and RIGHT-WING parties: *a party of the centre*

IN SPORT 9 [C] = CENTRE FORWARD (238)

In all of these possible meanings the core of the word remains and only the context changes. Lexical ambiguity makes it difficult for computational systems to understand human languages. The phrase "she is in the center" can have various meanings. It could mean that she is in the center of a room (definition 1), that she is in a shopping/sports/leisure or community center (definition 4), that she is in the city center (definition 2), and so on. In order to understand this phrase the context is essential. In normal conversation ambiguity rarely causes misunderstandings (Bieswanger and Becker 143). Therefore, it is not enough for an AI system to find keywords, it is mandatory that an AI system understands the context in which specific phrases and words appear. AI has to be able to understand inputs (verbal/textual) and react in the appropriate way. This, however, is a big challenge for computer programs. The ambivalence of the human language makes it fairly difficult for computers to respond, this is why it is so challenging for computers to pass the Turing test, for example.

Syntax builds the foundation of natural language. It creates a detectable and predictable decision procedure; whereas, semantic forms create meaning. Semantic forms include cultural understanding, idioms, slang, and so on. Subsequently, this makes understanding a language demanding. Semantic knowledge is required to understand the arbitrariness of natural language. American professor of computer science Terry Allen Winograd points out that in natural language syntax and semantics are merged together. The distinction of syntax and semantic cannot be separated from the external form of a sentence (qtd. in Hofstadter 631). Understanding language is not simply about understanding words. Language is used to communicate. Hendrickson Professor of Business Phillip G. Clampitt illustrates that "communication is the transmission and/or reception of signals through some channel(s) that humans interpret based on a probabilistic system that is deeply influenced by context" (4). This definition includes several assumptions, one being, that language, which is one main aspect of communication, is transmitted. Therefore, language cannot be transferred and meaning can be compromised. Communication depends on the interpretation of the receiver as well as on the context. Clampitt illustrates the ambiguity of words on the example of the word "run":

> A sprinter can "run" in a race. Yet, politicians "run" races but not exclusively with their legs. Although a horse "runs" with legs, it uses four of them, whereas sprinters use two. A woman can get a "run" in her hose, which is troublesome, but having a "run" of cards is good. However, having a "run" on a bank is bad. "Running" aground is not good at all for a sailor, but a "run" with the wind can be exhilarating. To score a "run" in baseball is different than a "run" in cricket. Hence, we "run" into the ambiguity of language at every turn, even with simple, everyday words. (4)

This play-on-words with the term "run" exemplifies the difficulty of understanding natural language. Humans have almost no difficulty and understand the intended humor instantly, whereas computer systems do not understand the complexity of ambiguous terms easily.

Not only are words ambiguous, but people using language in a very loaded way. Messengers as well as receivers can adjust language to their advantages. Possible implications range from "the sender of a message may purposely use language that has multiple interpretations" (6) to "the receiver may purposely misunderstand[s]" (7) the message. Even though most of the time there is only one interpretation, the likelihood that a message is being altered makes it even more difficult for programmers to understand natural language. Language can be used imprecisely, but can still be understood as long as the context is determined. The contextual understanding is strongly influenced by the culture. Various values play an important role in high-context and low-context cultures. The most important factor is context which builds the foundation of understanding a message (9). "A unique context emerges as people interact, regardless of the culture" (11). This is significantly important for a message. Computational systems are able to use algorithms to analyze questions. However, the sense behind it cannot be grasped by any system so far.

Hofstadter talks in his book, *Gödel, Escher, Bach*, about there being at least three layers that are embedded in every message. Firstly, the frame message has to be perceived as an information bearing content in order to be recognized as a message. Secondly, the outer message includes the structural understanding of the message. Thirdly, the inner message is the actual message which the sender intended to transmit (166). Human beings analyze subconsciously these steps, most of the time. NLP systems on the other hand are following each of these steps individually to use the provided data.

There are different kinds of communication; spoken language that can be transmitted over the telephone or radio, face to face communication, and written communication. Written communication is especially demanding. For example, the tone of voice cannot be heard; therefore, ironic messages could be deceptive. Again, the context is an essential part of

communication. So far computer programs were not able to understand the context and could only reply to the message previously given by a person. Linguistics are essential in developing AI systems, without understanding natural language computer programs can not advance.

Loops and recursions give only a vague idea about what intelligence might entail. The complexity of the human brain and the organic usage of language enable people to use it in various ways and express possibilities, ideas, and/or hopes. Options are plentiful and include; declarative ("I don't know"), conditional ("I would like to know"), emotional subjunctive ("I wish I knew"), and rich counterfactual ("If I knew, I would") sentences. This allows people to communicate about an incredible large amount of variations. This way "human beings [can] organize and categorize their perceptions of the world" (Hofstadter 642).

There are different approaches to understanding natural language. One idea is based on frame theory which is built on constants, parameters as well as variables and is "a computational instantiation of a context....In frame language, one could say that mental representations of situations involve frames nested within each other. "Each of the various ingredients of a situation has its own frame" (644). The hierarchy is nested in a structure which is characterized by frames and sub-frames. Frame theory allows a completely different approach to understand natural language. In spite of this, language is only one component which is part of the construction of artificial intelligent systems.

5.8. Originality of Programs

In order for an AI system to be innovative it has to develop originality. A program is seen as original if the solution to a certain problem was unintended by the programmer. This does not mean that the system is necessarily aware of its originality. It is just another stepping stone that is essential to create an artificial intelligent system. An example of an original program is the elementary Euclidean geometry, written by E. Gelernter (qtd. in Hofstadter 606). The program was able to proof one of the basic theorems of geometry without receiving explicit instruction from the programmer. The programmer was aware of the originality of the program, but the significance is that the program was able to calculate these theorems which were unintended by the programmer. Even if such originality can be traced back to a form of recursion, a human being is still required to analyze and to evaluate the results. Therefore, Hofstadter's assumption that computers are only "a tool for realizing an idea devised by the human" (609) is still valid today. Watson, for example uses a very large repertoire of

vocabulary to generate its answers. The answers, however, are not explicitly programmed. The system can analyze the question and is able to provide the correct answer. Therefore, programmers rely on the software to make the most informed judgments and decisions.

5.9. Creativity and Randomness

Creativity can be a result of learning, memory, knowledge, and understanding of the environment. So far, creativity seems to be organic, whereas, computers have the perception of being mechanic. Gelernter's example shows that computational systems have creative acting abilities. Thereby, it is concluded that randomness is an important factor. Randomness determines the environment and all existence. Therefore, it is not unlikely that creative acts are part of computer technologies and AI systems. Pattern recognition and organization of these patterns are an attempt by humans to bring order into the world. This concept is also used by computers. However, in order for a program to be creative it has to be enabled to work randomly. Only a system that understands these processes can be an AI system. Even though, computers have been developed to be in some way original, and are able to recognize patterns, by playing chess or playing *Jeopardy!*, they have failed to develop creativity. American psychologist Robert Jeffrey Sternberg defines creativity as "the ability to produce work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)….Creativity is a topic of wide scope that is important at both the individual and societal levels for a wide range of task domains" (3). In order to be creative a system must understand the random world that it is surrounded by.

Still, computers work mechanically and that is what distinguishes them from human creativity. Creativity is influenced and reflected by beauty, feelings, love, empathy, and everything else that characterizes human emotions. These issues lie at the core of AI. In order for a program to understand these concepts it has to be able to understand the concept of natural language first. IBM's Watson is in this regard a great achievement. However, Watson understands natural language in a completely different than humans do. For some researchers, the Turing test is a possible experiment to evaluate, if AI systems will be able to understand natural language.

5.10. Turing Test

English mathematician and computer scientist Alan Turing wrote many articles about artificial intelligence. One important article was called *Computing Machinery and Intelligence.* Intelligence, learning, and creativity are very loaded terms. One way to actually test if a machine or program is able to perform an intelligent task is known as the Turing test. The idea behind the Turing test is that if an intellectual task can be performed by a machine and is not distinguishable from a human performance then a machine can be recognized as being intelligent (Tanimoto 9). The actual set up of the Turing test is simply described by the author of *The Elements of Artificial Intelligence* Steven Tanimoto:

> Turing's test is to put a human and a machine in one room, and another human, the "interrogator," in a separate room.…The interrogator may ask questions to either the other human or the machine, referring to one as A and the other as B. However, the interrogator is not told which of A or B is the human or which is the machine. The interrogator cannot see or hear the others but passes messages through an intermediary, which could be an electronic mail system or could be another person. As they respond to questions, A and B each compete with one another to convince the interrogator that he/she or it is the human. If the machine can win, on the average, as often as the human, then it passes the "Turing test," by this particular criterion, can think. In practice, the outcome of such a test would probably depend heavily on the humans involved, as well as the machine. (9-10)

In the 1960s, one of the first computer programs that was able to process natural language was called ELIZA. The program used keywords from the statements of the human participant and generated a response. This simple method was called "template matching" and made a first impact in the development of chat programs. The following is an excerpt from a conversation between ELIZA and a human participant:

SENT:	Men are all alike.
RECEIVED:	In what way?
SENT:	They're always bugging us about something or other.
RECEIVED:	Can you think of a specific example?
SENT:	Well, my boyfriend made me come here.
RECEIVED:	Your boyfriend made you come here.
SENT:	He says I'm depressed much of the time.
RECEIVED:	I am sorry to hear you are depressed.
SENT:	It's true. I am unhappy.

RECEIVED: Do you think coming here will help you not to be unhappy?
SENT: I need some help, that much seems certain.
RECEIVED: What would it mean to you if you got some help?
SENT: Perhaps I could learn to get along with my mother.
RECEIVED: Tell me more about your family. (Christian 77-78)

Words like “boyfriend”, “depressed”, or “unhappy” are used in the response in order to keep the conversation going. This program would have been able to take part in the Turing test. In the 1990s, the Loebner price was introduced to annually test if chat programs are able to pass the Turing test. In 1997, a program called “Converse” which used the name “Catherine” was developed and performed surprisingly well in a conversation with this human participant:

REMOTE: Hi, my name is Catherine, what’s yours?
JUDGE: My name is Elliott.
REMOTE: Well, it’s nice to meet you, Elliott. Did you see that story on CNN last night about the lesbian couple who came out at a White House party on Sunday?
JUDGE: Yes, I did. I think it may all be a publicity stunt for Ellen.
REMOTE: Ellen Degeneres was one of them - she was kissing her lover in front of Bill Clinton. Do you think he was very embarrassed?
JUDGE: I think he want the gay vote. He will do anything to get it and keep it. He does not have a conservative base.
REMOTE: Are you still a Friend of Bill Clinton?
JUDGE: I am not sure if he has any real friends. He has achieved alot and has left alot of people behind.
REMOTE: You really don’t like the guy—well, at least that’s definite—I thought most people were just indifferent or numb by now.
JUDGE: I think we could have a more honest and dynamic leader. One able to face the real dangers facing the country and address them with courage and intelligence. Rather then listening to public opinion polls.
REMOTE: I guess that means you really think there’s serious behind Whitewater, the Special prosecutor and all that stuff?
JUDGE: We have two dead people, Brown and guy who commited suicide, we have lost papers. And we have Clintons people willing to tell all. I think that Hillary will be found guilty of something.
REMOTE: But how do you actually feel about him, now in the number two term?
JUDGE: He is too smart to get caught. He deserved to named hotel manager of the year.

REMOTE: OK, but what happens now—does it all drift on, with everyone waiting to see how much his friends have said about him under threat of lots and lots of jail– or will something dramatic actually happen? (Christian 196-197)

The program used a loop that was able to bring the conversation back to a current event in the news. This human participant was intrigued by the topic and followed the conversation. However, if a human participant did not want to follow "Catherine's" conversational path it was easy to expose "Catherine" as a computer program:

JUDGE: it looks like i've got to go, Catherine. its been nice talking with you
JUDGE: do you need water or something before the 5 minute rounds?
REMOTE: OK, yes on balance … Time to get off this one I think and onto something more down to earth!
JUDGE: like what?
REMOTE: Almost certainly! let's move on
REMOTE: I suppose it depends on where you're coming from, but as the song goes in My Fair Lady, "Why can't a woman be more like a man?" (Christian 198-199)

The difficulty with this test is to create a computer system that can really understand the questions that are being asked and is able to respond to the questions. All these chat bots like "ELIZA" and "Converse" are simple programs that play with the perception of intelligence. A more logical approach to understanding artificial intelligence and consequently human intelligence is the statement from Tanimoto:

> As we grow more sophisticated, we realize that the question of whether a system is intelligent is a shallow one. We should be asking about the kinds, quality and quantity of knowledge in a system, the kinds of inference that it can make with the knowledge, how well-directed its search procedure is, and what means of automatic knowledge acquisition are provided. There are many dimensions of intelligence, and these interact with one another. (10)

Therefore, newly developed systems like Watson show a new intriguing perspective on intelligence. By understanding such systems as well as understanding its way of processing natural language, the focus turns away from the Turing test and more towards the concept of understanding intelligence.

6. Understanding Watson

6.1. Watson's Hardware

On the *Jeopardy!* stage Watson has an observable presence. Watson was developed to compete against human *Jeopardy!* champions and to give confident answers within three seconds. After analyzing hundreds of *Jeopardy!* games the average speed of answering a question settles at approximately 3.5 seconds. Therefore, Watson's final answering speed allows him to answer, on the average, faster than the human competition. The answer panel shows Watson's top three responses. A confidence threshold indicates whether Watson will buzz-in and answer the question. All *Jeopardy!* players including Watson use the same hand buzzer device. In order for Watson to buzz-in, a mechanical device was constructed which is directly linked to Watson. The Avatar, as already mentioned, reflects Watson's presence on stage and the changing colors show in what way the clues are being analyzed and processed ("What powers Watson?").

Watson's Linux operating system consists of "90 IBM Power 750 servers based on the POWER7 processor" ("What powers Watson?"), which include 10 server racks "with associated I/O nodes and communications hubs….The system has a combined total of 16 Terabytes of memory and can operate at over 80 Teraflops (trillions of operations per second)" ("Watson – A System Designed for Answers" 4). It takes one processor core two hours to achieve the same deep analytics that Watson can perform within three seconds. Therefore, 2880 processor cores were combined in a "super high-speed network" ("What powers Watson?"). It is being estimated that Watson, in comparison to Deep Blue, is one hundred times faster. The crux lies in the "POWER7 processors inside the Power 750 [which] is designed to handle both computation-intensive and transactional processing applications – from weather simulations, to banking systems, to competing against humans on Jeopardy!" ("What powers Watson?").

The memory capacity of Watson is relatively small and does not exceed five hundred gigabytes. However, Watson does not store images, videos, or audio files. Text files are far smaller than these kinds of data. This allows Watson to focus on processing documents entirely consisting of natural language, which is the focus of the *Jeopardy!* challenge and of the research. The system includes "Active Memory Expansion" which enables POWER7 technology to maximize and exceed its actual memory capacity. This is due to "innovative compression/decompression of memory content [which] can enable memory expansion up to 100 percent" ("Power 750 Express Server" 1). This innovative technology can expand a

server's capacity from 512 gigabyte to 1 terabyte. The benefit of this technology is that a server can run more partitions as well as enables the partitions to perform more effectively ("Power 750 Express Server" 3).

The Watson system as a whole needs a vast amount of energy. Through adjustments and the new layout of the system, over ninety servers enable Watson to operate faster and reduce the total amount of the energy consumption. The ninety servers with its total of 2880 POWER7 cores run at 3.55 GHz. In order to keep the energy demand as low as possible "all Power Systems include EnergyScale™ technology to reduce energy consumption and provide the ability to manage and customize energy usage" ("What powers Watson?").

Compared to other processors, POWER7 has an advantage, especially considering the intelligent energy. This includes "increased performance and performance per watt" ("Power 750 Express Server" 2) which has ultimately a positive effect on the system as a whole and businesses can benefit from "the first RISC-based ENERGY STAR-qualified servers" ("Power 750 Express Server" 3) which not only reduce the cost of energy but also keeps the emission of greenhouse gases within the guidelines ("Power 750 Express Server" 3).

In order to process and analyze the incoming data, Watson uses a high-speed network that allows the system to operate with 90 x 10,000 megabits. Watson's Ethernet network has a speed of up to ten gigabit. These high performances are necessary to enable Watson to generate answers in the same time as humans do. Watson has to analyze possible answers, buildup confidence, and be faster than his opponents ("What powers Watson?").

Almost half of the energy is used to cool Watson. This is achieved by two cooling units which combined in weight is approximately forty tons. These air conditioning units allow a constant temperature in the server room of 17.8°C.

6.2. Watson's Software

6.2.1. Software Foundations

In order to understand Watson and the influence on NLP it is essential to understand Watson's software and its various components. Considering the following description: "Watson is a workload optimized system based on IBM DeepQA architecture running on a cluster of IBM®POWER7®processor-based servers" ("Watson – A System Designed for Answers" 2). The Power7 processor is part of the investigated hardware, whereas, the DeepQA architecture is part of Watson's software. DeepQA is the foundation of Watson. The

research team had to consider aspects, such as analyzing natural language, by using an enormous data base of sources. After identifying the sources it was essential to generate hypotheses and in order to validate these hypotheses evidence had to be determined which resulted in a hierarchical order of the hypotheses. All of these steps have to be accurate, selected with high confidence, and ultimately have to be performed exceptionally fast to win the *Jeopardy!* challenge. DeepQA has a high variety of potential application, such as in business and medicine. The principles underlying DeepQA that make further application at all possible are:

> 1. Massive parallelism: Exploit massive parallelism in the consideration of multiple interpretations and hypotheses.
> 2. Many experts: Facilitate the integration, application and con-textual evaluation of a wide range of loosely coupled probabilistic question and content analytics.
> 3. Pervasive confidence estimation: No single component commits to an answer; all components produce features and associated confidences, scoring different question and content interpretations. An underlying confidence processing substrate learns how to stack and combine the scores.
> 4. Integrate shallow and deep knowledge: Balance the use of strict semantics and shallow semantics, leveraging many loosely formed ontologies. ("Watson – A System Designed for Answers" 4)

UIMA (Unstructured Information Management Architecture) is the foundation of Watson's architecture. This architecture allows Watson to "analyze unstructured information such as text, audio and images" ("What powers Watson?"). Apache UIMA is set in a parallel structure that allows Watson to process natural language. The clusters in this system create the foundation to perform a broad range of high-speed analytical computations.

UIMA annotators are designed to analyze text that enables Watson to improve over time, and results in better identification of components as well as evaluating of hypotheses. UIMA consists of different parts, for instance, "UIMA-AS, part of Apache UIMA, enables the scale-out of UIMA applications using asynchronous messaging" ("Watson – A System Designed for Answers" 4). A more detailed description of UIMA follows in the next section.

UIMA-AS enables Watson to search, analyze, and evaluate 500 Gb of text files in less than three seconds. Another important framework is the Apache Hadoop framework that "facilitate[s] preprocessing the large volume of data in order to create in-memory datasets used at run-time" ("Watson – A System Designed for Answers" 4). The connection of these two components is that the UIMA annotators are part of Hadoop which enables the

framework to organize and optimize the CPU as well as the processes ("Watson – A System Designed for Answers" 4).

6.2.2. Apache UIMA

The amount of information in the society today is increasing drastically. The Internet is only one example of many; technical reports, voice mails, and other communication tools. The issue with information is that it mainly consists of unstructured natural language components. Therefore, these pieces of information require analysis to retract the desired knowledge. UIMA can be used to decode and analyze the unstructured information such as natural language texts, audio recordings, and videos. Analysis engines are part of the UMIA software which allows abstracting the desired information from a document ("The Knowledge Rush"). This software system is essential for Watson to produce a valid answer. Watson identifies text entities and can generate an answer that lies in the foundation of the UIM application.

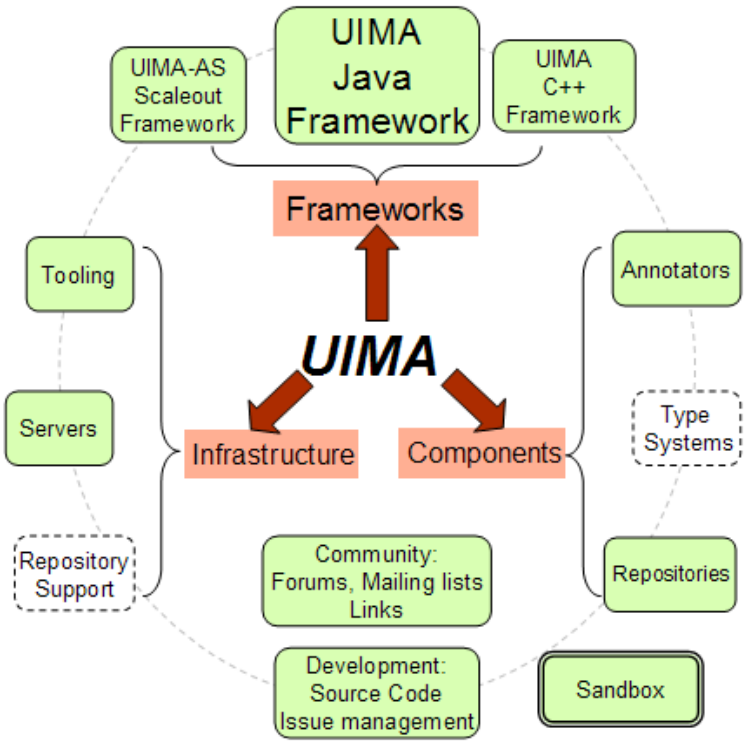

Fig. 1. Visualization of the Unstructured Information Management Application ("Apache UIMA")

UIMA consists of frameworks, infrastructure, and components (see fig. 1). The infrastructure consists of tooling and servers which built the foundation of UIMA. Information is being divided in separate components such as annotators and repositories.

The separate components are written in Java or C++ and managed by the Apache licensed UIMA framework. The annotator components built the foundation of the analysis, whereas the framework is intended to configure the components. This includes the UIMA - Asynchronous Scaleout (UIMA-AS) framework which supports the Java framework. All output that can possibly generate new ideas is placed in the sandbox and is used later when the system requires it for a specific segment. An example would be that UIMA components identify the language of a document which is going to be separated in individual segments. These segments are analyzed independently within the sentence boundary which enables the system to detect the required entity, e.g. date or city ("Apache UIMA").

UIMA is used by many academic projects, but is also very interesting for businesses, such as IBM, and takes an important role in the Watson project. The representation structure underlying UIMA is called Common Analysis Structure (CAS). Analysis engines are part of the structure which entails that they are responsible for the analysis of documents and defining particular sets of these documents. A high performance workflow allows processing the collected data and passing it onto the next analysis engine. This facilitates CAS to generate metadata and enables the analysis engines to be compatible with each other as well as to operate efficiently ("UIMA Architecture Highlights").

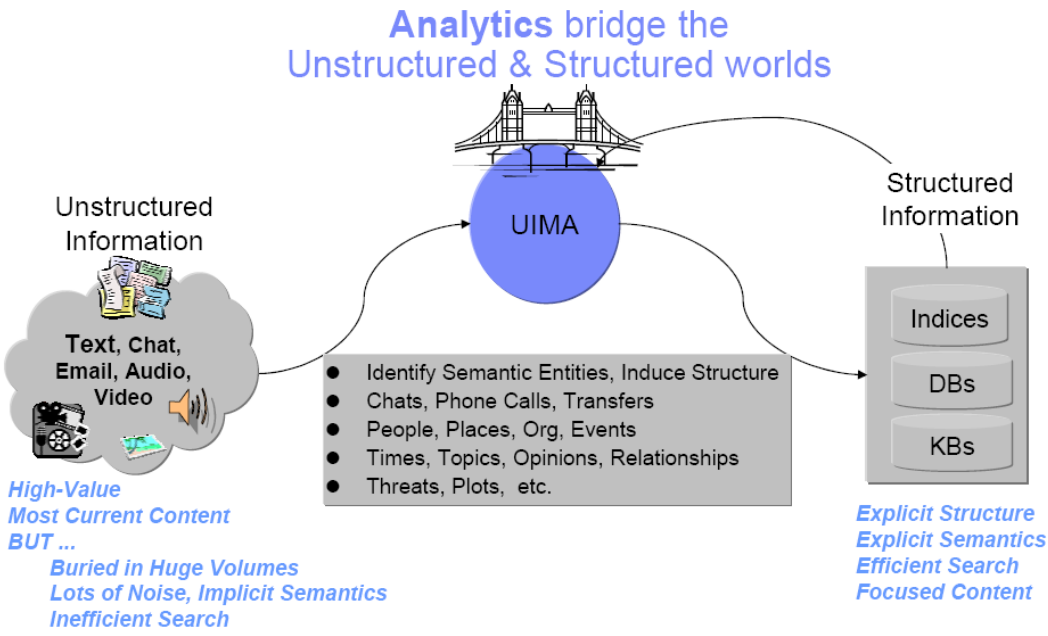

Fig. 2. Visualization of UIMA as a bridging structure ("UIMA Overview & SDK Setup")

UIMA bridges the gap between unstructured information and structured information (see fig. 2). Unstructured information can be any given type of file, such as texts, emails, audio recordings, and images, which are unorganized and exist in large volumes. UIMA acquires information and analyzes the data semantically through language and entity/relation detection, classifications, or translations. The difficulty is that semantics is often implicit in unstructured information. Ontologies, indices, and knowledge bases allow UIMA to structure the information and represent semantics explicitly. The delivery of the structured information follows in semantic search or automated reasoning which enables UIMA to present the inquiry as text files, graphs, and tables. A transformation occurs from an inefficient search towards an explicit structure with a focused content. The result is an efficient search with structured information ("UIMA Architecture Highlights").

An example will illustrate how semantic search works in order to get a better understanding of UIMA. Search engines such as Google or Yahoo detect keywords and list all documents including these keywords, whereas, UIMA analyzes the query and targets the request more effectively. Considering the example that a user is looking for a restaurant with a name that he/she cannot recall, but knows that the restaurant's name includes the word "blue". "Blue" is an ambiguous term and would result in a high number of search results, when the user chooses to use a regular keyword search. UIMA's CAS annotation supports XML Fragments and uses a semantic search ("UIMA and Semantic Search").

> XML Fragment query is written as: <restaurant> blue </restaurant> ("UIMA and Semantic Search")

A named-entity recognizer reduces the result of this query and shortens the list which includes only the word "blue" in phrases that are related to restaurants. A relationship recognizer can be included in the search which enables the user to also look for the "owner of" relationship. CAS is then able to include the relationship annotation with the semantic search ("UIMA and Semantic Search").

> The query is written as: <owner_of>
> <person>blue </person>
> <restaurant>blue </restaurant>
> <owner_of> ("UIMA and Semantic Search")

"Blue" has various meanings and definitions as seen in the *Oxford Advanced Learner's Dictionary*:

> **blue** /blu:/ *adj., noun*
>
> **adj.** (**bluer, bluest**) **1** having the colour of a clear sky or the sea/ocean on a clear day: *piercing blue eyes* ◊ *a blue shirt* **2** (of a person or part of the body) looking slightly blue in colour because the person is cold or cannot breathe easily: *Her hands were blue with cold*. **3** (informal) sad [**SYN] DEPRESSED**: *He'd been feeling blue all week*. **4** films/movies, jokes or stories that are **blue** are about sex: *a blue move*-see also TRUE-BLUE – blueness *noun* [U, sing.]: the blueness of the water [**IDM] do sth till you are blue in the face** (*informal*) to try to do sth as hard and as long as you possibly can but without success: *You can argue till you're blue in the face, but you won't change my mind*.- more at BLACK *adj*., DEVIL, ONCE *adv*., SCREAM *v*.
>
> **noun**-see also BLUES **1** [C, U] 1 the colour of a clear sky or the sea/ocean on a clear day: ***bright/dark/light/pale blue*** ◊ *The room was decorated in vibrant blues and yellows*. ◊ *She was dressed in blue*. **2** [C] (*BrE*) a person who has played a particular sport for Oxford or Cambridge University; a title given to them **3** [C] (*AustralE, NZE, informal*) a mistake **4** [C] (*AustralE, NZE, informal*) a name for a person with red hair **5** [C] (*AustralE, NZE, informal*) a fight [**IDM] out of the blue** unexpectedly; without warning: *The decision came out of the blue*.-more at BOLT *n*., BOY *n*. (156)

When using UIMA with the semantic search application as shown before the query will only look for documents that include the name of the restaurant or the name of a particular person who is the owner of a restaurant. Therefore, possible search results could include:

> "…John Blue, Owner of Blue Lagoon…"
> "…The Owner of Blue Ocean Seafoods, Mr. Blue…"

However, all sentences and/or phrases that are not related to "owner" or "restaurant" would be ranked lower or even be excluded from the search such as ("UIMA and Semantic Search"):

> "...Joe felt blue today..."
> "...Mr. Blue's hands were so cold that they almost looked blue…"

This precision enabled Watson to perform so well in the *Jeopardy!* challenge. Nevertheless, the main purpose of this technology is "to transform unstructured information to structured information by orchestrating analysis engines to detect entities or relations and thus to build

the bridge between the unstructured and the structured world" ("Apache UIMA"). That gives UIMA a broad field of applications.

Ferrucci explains, in the PowerPoint presentation "UIMA and Semantic Search Introductory Overview", the advantage of semantic search over keyword search. A keyword search can find a large number of results including desired matches, undesired contents, or even misses the right content completely.

> Search for the following content:
> [Going rate for leasing a billboard near Triborough Bridge] (2)

When typing this phrase in the Google search bar the top eight hits are concerned with articles about the semantic search and UIMA application. The tenth result is the following:

Wall Street Hotels - Find lodging and hotels ***near*** Wall Street
www.wallstreethotels.net/gettingthere.html - Cached
Wall Street Hotels is committed to finding the best ***rates*** on Wall Street Hotels, **...** flashing, glittering ***billboards*** as they command your eyes upward to take notice. **...** with the Bruckner Expressway (I-278) at the ***Triborough Bridge***. **...** ***Going*** north on FDR Drive takes you into Harlem; traveling south on FDR Drive **...**

Fig. 3. Search result for the query: "Going rate for leasing a billboard near Triborough Bridge" (Google)

This search result includes the keywords "near", "rates", "billboards", "Triborough Bridge", and "going" (see fig. 3). None of the top ten search results is even closely related to the actual search and neither are the following 680 total Google results. There might be helpful search results within these hits; however, it takes valuable time to investigate each individual link. The same results can be found in other search engines, such as Yahoo. In this scenario, these search engines do not seem to be useful. In contrast, semantic search allows knowledge to be extracted from the given search item.

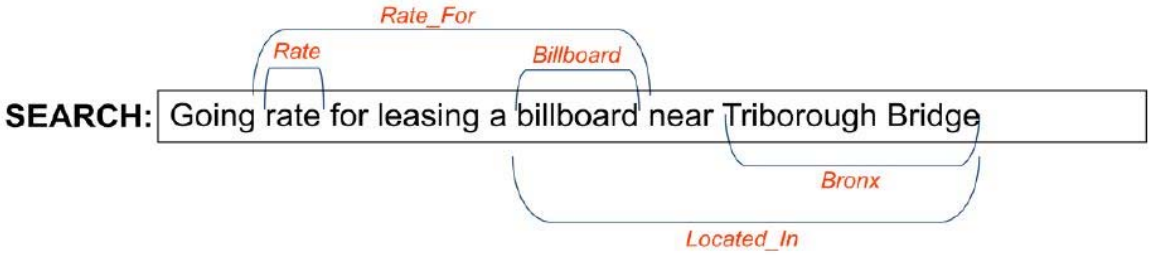

Fig. 4. Semantic components of the query: "Going rate for leasing a billboard near Triborough Bridge" (Ferrucci, UIMA and Semantic Search - Introductory Overview 6)

The semantic search analyzes the semantic types of the phrase which allows better search results (see fig. 4). The phrase is divided into categories including the “Rate”, “Rate_For”, “Billoard", “Located_In”, and the corresponding location, in this case “Bronx”.

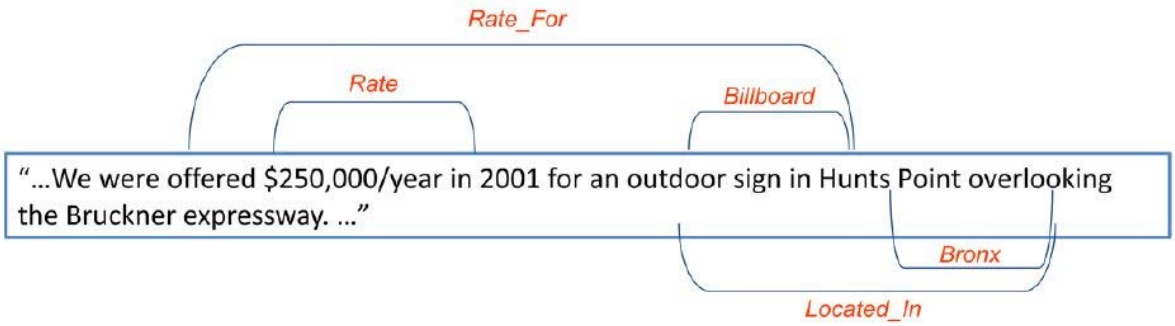

Fig. 5. Search result of the query: “Going rate for leasing a billboard near Triborough Bridge” using semantic search (Ferrucci, UIMA and Semantic Search - Introductory Overview 6)

The search result has no keywords in common; however, all the semantic types like “Rate” and “Rate_For” are included (see fig. 5). Therefore, this hit can be seen as a desired search result and can be used for further research. In comparison a search result that includes common keywords, but where the content is not related is not useful (see fig. 6).

Fig. 6 Search result of the query: “Going rate for leasing a billboard near Triborough Bridge” using keyword search (Ferrucci, UIMA and Semantic Search - Introductory Overview 7)

The keywords: “Bridge”, “rated”, and “Billboard” are in common with the query. Nevertheless, the semantic types: “Song Title”, “Queens”, and “Magazine” are unrelated to the original semantic types of the search phrase. That means it equals an inadequate search result (see fig. 6).

The application of UIMA in Watson allows the system to use semantic search and improves its precision by gathering knowledge. Watson would have never been able to win

the *Jeopardy!* challenge with a keyword search. The key is to use automated annotation to enable the system to learn and improve the precision of the search results.

6.2.3. Watson's System and *Jeopardy!*

The development of Watson took three years. The focus was not only on the research aspect of the DeepQA system, but also to create public interest in this project in a similar way Deep Blue influenced the 1990s. The underlying demand of the *Jeopardy!* challenge is precise answering to natural language. "For researchers, the open-domain QA problem is attractive as it is one of the most challenging in the realm of computer science and artificial intelligence, requiring a synthesis of information retrieval, natural language processing, knowledge representation and reasoning, machine learning, and computer-human interfaces" (Ferrucci et al. 60).

The success of Watson relies on three major aspects; confidence which is inevitable to give an answer at all, but essential to providing enough evidence to generate a high confidence. Having high confidence is the result of precision. Therefore, precision can be generated with the help of systems like UIMA which enables Watson to detect the appropriate answer to the corresponding question. Thirdly, speed is fundamentally important to win a game of *Jeopardy!* and it is necessary for future applications of the system. *Jeopardy!* is another step in creating an applicable QA system that can enrich the field of computer engineering and artificial intelligence.

6.2.3.1. The Jeopardy! Challenge

Jeopardy! uses a broad spectrum of ambiguous questions. Categories include specific subjects (e.g. "Name The Decade"), puns (e.g. "Beatles People"), clues (e.g. "Actors Who Direct"), and various topics (e.g. "Alternate Meanings") ("Jeopardy! - The IBM Challenge" Day 1,3). Therefore, there is not only one ideal algorithm that can answer any given question. The key is to compute confidence in each available component. The combined confidence of each component is necessary to generate a correct answer. To create this confidence is the biggest challenge of DeepQA. All components work together to create the highest possible confidence which is achieved through hierarchical machine-learning methods. If the confidence is high enough the system can attempt to answer the question, otherwise, if the confidence is too low the system will not try to answer the question (Ferrucci et al. 60).

6.2.3.2. Jeopardy! Clues

As already mentioned, the categories are broad and the variety of clues is vast. In order to structure and analyze the clues, different types of classifications are used. One important type of *Jeopardy!* questions are factoid questions. These are questions where the answer consists of facts. The difficulty is to generate the correct fact that is asked (60,62). The following example is part of the first day of the *Jeopardy!* challenge:

Category:	Name The Decade
Clue:	Disneyland opens & the peace symbol is created
Answer:	What are the 1950s? ("Jeopardy! - The IBM Challenge" Day 1)

Another kind of question is called decomposition. Decomposition questions usually consist of different parts that are not related, but have the same answer. These sub-clues usually do not appear in the same sources; therefore, the parts of the question have to be analyzed separately. An answer has to be generated for both clues which correspond to the original question (Ferrucci et al. 62).

For example:	
Category:	Alternate Meanings
Clue:	A piece of wood from a tree, or to puncture with something pointed
Sub-clue A:	A piece of wood from a tree
Sub-clue B:	to puncture with something pointed
Answer:	What is (a) stick? ("Jeopardy! - The IBM Challenge" Day 1)

Deep QA generates decomposition hypotheses for possible interpretations of the clues and sub-clues. Decomposable questions can also consist of sub-clues which are embedded in another sub-clue. If this kind of question is asked, the system answers the question in two steps. The first step includes answering sub-clue A and the answers of sub-clue A will define the answer by embedding it in sub-clue B (Ferrucci et al. 62).

For example:	
Category:	Final Frontiers
Clue:	From the Latin for “end”, this is where trains can also originate
Sub-clue A:	From the Latin for “end”
Sub-clue B:	this is where trains can also originate
Answer:	What is a Terminal? ("Jeopardy! - The IBM Challenge" Day 1)

Another kind of decomposition question includes puzzles. Puzzles are very challenging for people and especially for a computational system that has to processes these various defined clues. There are numerous categories, such as “converting roman numerals”, “Before and After”, and “Rhyme Time” (Ferrucci et al. 62).

For example:

Category:	Before and After Goes to the Movies
Clue:	Film of a typical day in the life of the Beatles, which includes running from bloodthirsty zombie fans in a Romero classic.
Sub-clue 2:	Film of a typical day in the life of the Beatles.
Answer 1:	(A Hard Day’s Night)
Sub-clue 2:	Running from bloodthirsty zombie fans in a Romero classic.
Answer 2:	(Night of the Living Dead)
Answer:	A Hard Day’s Night of the Living Dead (62)

Category:	Rhyme Time
Clue:	It’s where Pele stores his ball.
Sub-clue 1:	Pele ball (soccer)
Sub-clue 2:	where store (cabinet, drawer, locker, and so on)
Answer:	soccer locker (62)

All of these questions can occur in categories; therefore, in order to win the game it is essential that DeepQA can deal with all types of questions. There are only two types of questions that are excluded from the IBM *Jeopardy!* challenge. They are audiovisual questions and special instruction questions. Audiovisual questions include audio recordings, images, or videos which are essential to answer the clue. Special instruction questions are explained verbally. Both of these questions seem very interesting for further developments in computer science and artificial intelligent systems; however, the focus lies on understanding natural language question consisting of text files (62-63).

Before the *Jeopardy!* challenge, IBM collected data from previous *Jeopardy!* questions. Twenty thousand samples were categorized and structured into a lexical answer type (LAT). A word inside the clue was defined to identify the answer. In order to exemplify this process Ferrucci used the following example:

LAT is the string “maneuver.”

Category: Oooh….Chess

Clue: Invented in the 1500s to speed up the game, this maneuver involves two pieces of the same color.

Answer: Castling (63)

This technique does not by far encompass all types of questions. An estimation by IBM suggests that about twelve percent of all questions are not referred to a specific term but to a pronoun (e.g. it, that, this). Therefore, the context has to be understood in order to generate the correct answer (63).

Here is an example:

Category:	The Art Of The Steal
Clue:	Rembrandt’s biblical scene “storm on the sea of” **this** was stolen from a Boston museum in 1990
Answer:	What is Galilee? ("Jeopardy! - The IBM Challenge" Day 2)

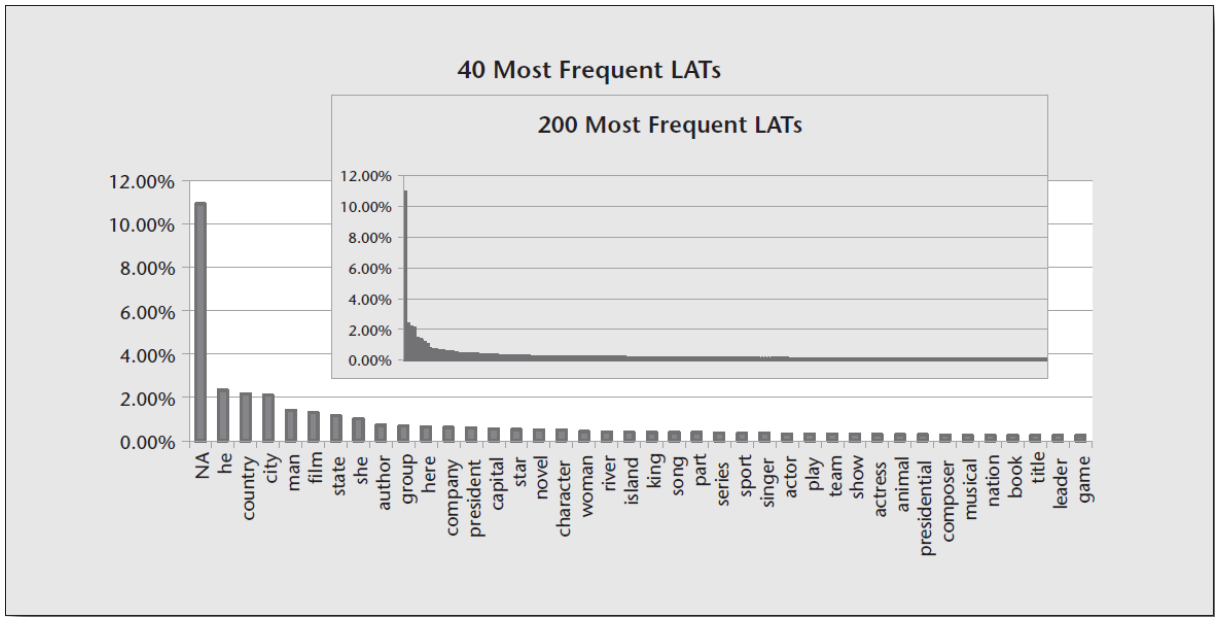

Fig. 7. Lexical Answer Type Frequency (Ferrucci et al. 63)

The LAT chart shows that “the most frequent 200 explicit LATs cover less than 50 percent of the data” (Ferrucci et al. 63) (see fig. 7). However, certain types such as “he”, “country”, and “city” only cover approximately two percent each. Also, “man”, “film”, “state”, and “she” only rank higher than one percent each. Even though these terms are

relatively frequent there are more than 2500 distinct LATs and the column labeled "NA", with a total of more than eleven percent, do not include specific terms. This suggests that knowing the top two hundred LATs will not satisfy or fulfill the desired requirements. This is an opportunity for various applications:

> Our clear technical bias for both business and scientific motivations is to create general-purpose, reusable natural language processing (NLP) and knowledge representation and reasoning (KRR) technology that can exploit as-is natural language resources and as-is structured knowledge rather than to curate task-specific knowledge resources. (63)

Task-specific knowledge resources are not valuable for artificial intelligent systems. In order to deal with the large amounts of information, organizations require systems that are able to process natural language resources and that are able to structure knowledge.

In order to be successful in *Jeopardy!* IBM analyzed almost two thousand *Jeopardy!* games and created a chart that indicates precision and the correctly answered questions (see fig. 8).

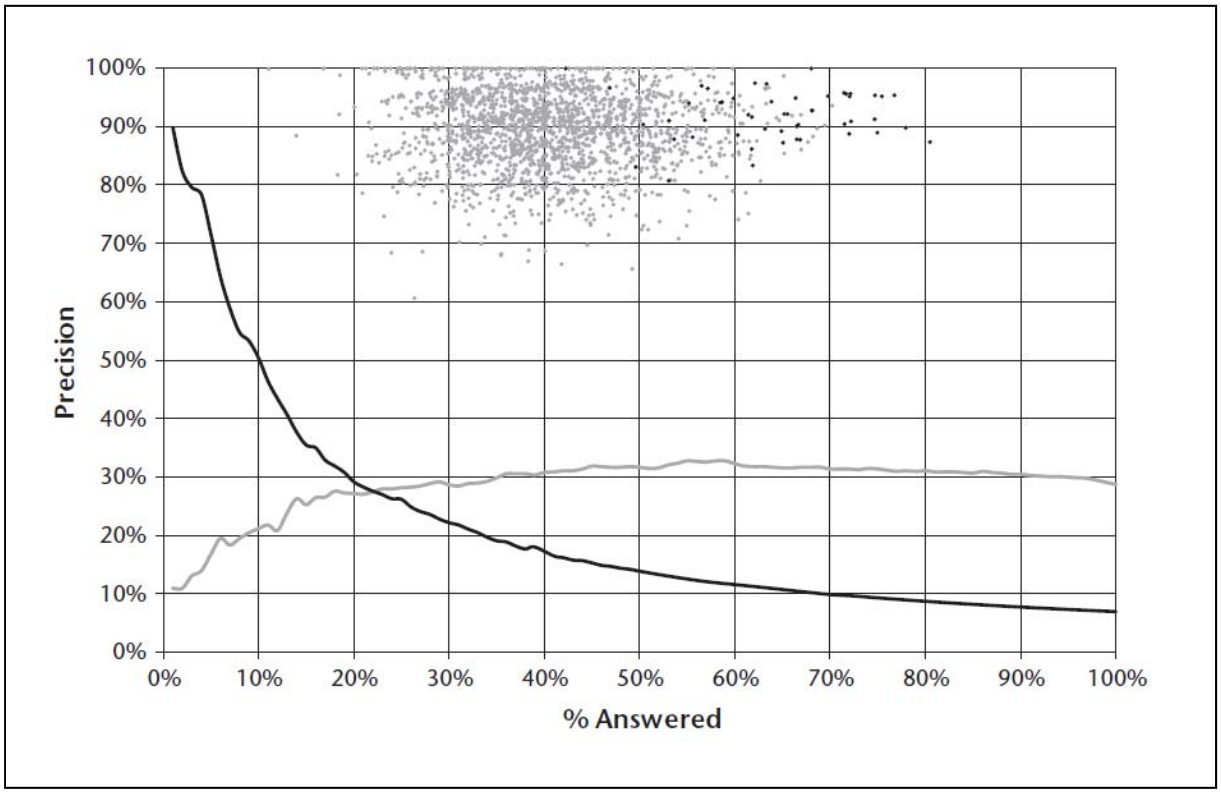

Fig. 8. Chart including "Winners Cloud", "Text Search", and "Knowledge Base Search" (Ferrucci et al. 68)

The ordinate of the chart is labeled "Precision"; this includes the percentage of the questions answered correctly by the candidate who acquired the question. The axis of

abscissas is labeled "% Answered" which includes the total percentage of questions answered by the candidate who acquired the question. The set of gray dots is called "Winners Cloud", because only candidates who won a *Jeopardy!* game were accounted. The average acquired questions lies between forty and fifty percent, whereby, the precision ranging from eighty five to ninety five percent. Slightly darker dots represent Ken Jennings's games, one of Watson's two human competitors in the *Jeopardy!* challenge. His average is approximately sixty-two percent and his answered questions with a precision of ninety-two percent. The chart represents only a guideline for the development of DeepQA. At the beginning the performance measured by DeepQA excludes competition, confidence, speed, and risk management which were all part of the games that are represented in the "Winners Cloud" (65-66).

In order for DeepQA to imitate these performances a wide variation of techniques and algorithms had to be implemented. The task of these systems was to monitor improvements and deteriorations. PIQUANT (Practical Intelligent Question Answering Technology) was one of the first TREC (Text Retrieval Conference) measuring systems that were developed in 2004. This system was not linked to the Internet and the focus was merely on precision and confidence. As already mentioned speed, betting techniques, and clue values play an important part in the *Jeopardy!*. The chart includes two gray baselines. The lighter gray line is based on text search, whereby, terms within the question are used to find the appropriate answer in the database. The darker gray line is based on knowledge search of structured data. The text search graph has a low precision when it comes to answering a small amount of questions. The confidence increases with the number of answered questions. Nevertheless, the graph stagnates at approximately thirty percent. Compared to the knowledge search graph the text search graph performs much better at the one hundred percent mark of answered questions. The knowledge search achieves less than ten percent precision by one hundred percent of answered questions. Still, and this is noticeable, the knowledge search based system has up to ninety percent confidence when the number of answered questions being small. This indicates that both systems are necessary to make DeepQA successful (66-67).

The first approaches with this system were difficult and did not show the desired success. Only when IBM Research began using OAQA (Open Advancement of Question Answers) improvements could be seen. Ferrucci concludes in the research paper "Building Watson: An Overview of the DeepQA Project" that those "system-level advances allowing rapid integration and evaluation of new ideas and new components against end-to-end metrics were essential to our progress" (67). The creation of an architecture that could evolve and

evaluate contexts was the foundation of developing DeepQA which "is a massively parallel probabilistic evidence-based architecture" (68). DeepQA uses more than one hundred techniques and overlaps to process natural language. The complexity of the system and the underlying principles (massive parallelism, many experts, pervasive confidence estimation, and integration of shallow and deep knowledge) allow a broad variety of possible applications (68).

6.2.3.3. Watson's DeepQA Architecture

A closer look at the DeepQA architecture will enable the understanding of NLP. A rich number of algorithms are used to generate hypotheses. In order to evaluate these hypotheses evidence is collected from unstructured and structured databases. The goal is to score the best possible confidence. Thereby, machine learning and reasoning algorithms help the system to assess itself and weigh the algorithms ("How Watson Works").

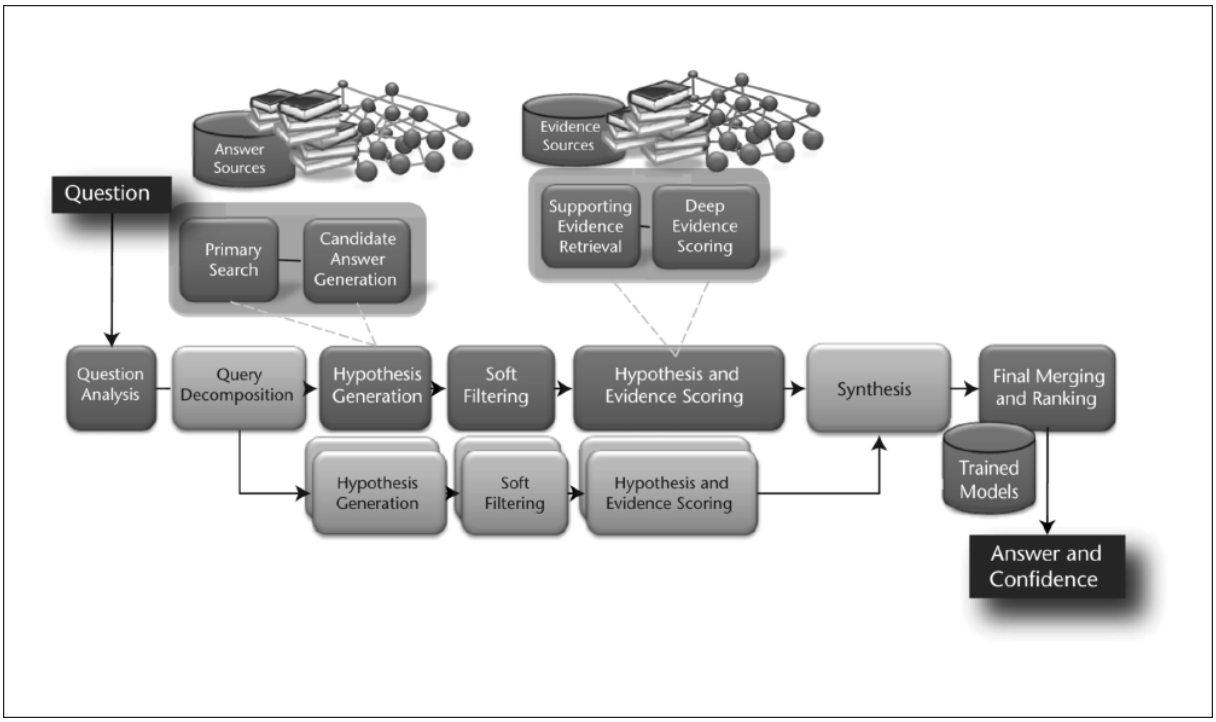

Fig. 9. Visualization of the DeepQA High-Level Architecture (Ferrucci et al. 69)

After a question is revealed it has to be analyzed, however, in order to understand the question the content has to be acquired by using manual and automatic tasks. Manual tasks are for example to analyze example questions, whereas, automatic tasks are domain analyses, as described in the LAT analysis. Watson is not connected to the Internet; therefore, it has to

rely on its own resources which mainly consist of encyclopedias, dictionaries, and so on. The next process consists of four steps.

> Automatic corpus expansion process:
> (1) identify seed documents and retrieve related documents from the web
> (2) extract self-contained text nuggets from the related web documents
> (3) score the nuggets based on whether they are informative with respect to the original seed document
> (4) merge the most informative nuggets into the expanded corpus (Ferrucci et al. 69)

Furthermore, resources are being collected by databases and ontologies. Ontologies include dbPedia, WordNet, and the Yago8 ontology.

American computer scientist Thomas Gruber defines in his publication "A Translation Approach to Portable Ontology Specifications" an ontology as “an explicit specification of a conceptualization” (qtd. in Weller 115). Investigating WordNet will exemplify one of the ontologies and will make the system more comprehensible. WordNet groups specific syntactic categories with similar meaning into sets of synonyms (synsets). Synsets are related with one another and create an interconnection.

Table 1

Examples of relations in WordNet (Nie and Brisebois 430)

Relationship	**Example**
Synonymy	Computer – data processor
Antonymy	Big – small
Hyponymy (IS-A-KIND-OF)	Tree – hyponymy maple
Hypernymy (IA-A)	Maple – hyperonymy tree
Meronymy (HAS)	Computer – meronymy processor
Holonymy (IS-PART-OF)	Processor – holonymy computer

WordNet creates relations between various semantic types (see table 1). Looking at other examples of hyponymy, hyperonymy, and ISA relation illustrates the link between general synsets (e.g. “furniture” and “piece of furniture”) with specific synsets (e.g. “bed” and

“bunk bed”). Therefore, WordNet is able to link hierarchies together, for example, the hyponymy “bunk bed” is a subordinate of the hyperonymy “furniture” (Fellbaum).

Example:
Hyponymy relation: if an armchair is a kind of chair, and if a chair is a kind of furniture, then an armchair is a kind of furniture. (Fellbaum)

“WordNet distinguishes among Types (common nouns) and Instances (specific persons, countries and geographic entities)” (Fellbaum). It is also able to distinguish meronymy which refers to parts of objects, not hierarchies. Therefore, “back”, “seat”, and “leg” can be a part of a “chair”.

Example:
Metonymy relation: if a chair has legs, then an armchair has legs as well. (Fellbaum)

WordNet also uses troponyms to generate a hierarchy between verbs with an increasingly specific relation (Fellbaum).

Examples:
Volume: “communicate”, “talk”, “whisper”
Speed: “move”, “jog”, “run”
Emotion: “like”, “love”, “idolize” (Fellbaum)

WordNet organizes adjectives into antonym pairs. Antonymy are gradable pairs which can signify a polarity.

Examples:
Antonymy relation: “small”, “medium”, “large”
“hot”, “warm”, “tepid”, “cool”, “cold”, “freezing”
(Bieswanger and Becker 140)

Another advantage of WordNet is that it can link, for example antonymy relations with synonymy relations which results in an even broader spectrum of interconnectivity. WordNet is able to relate nouns, verbs, adjectives, and adverbs. Parts of Speech can be, for example a relation between the agent (“painter”) of “paint” and the result (“painting”) (Fellbaum).

Questions have to be analyzed and processed (see fig. 9). Therefore, Watson uses a broad spectrum of applications, such as parses, relations, and logical forms. Parsers are used

to structure unstructured texts. Researcher Adam Lally and Paul Fodor from Stony Brook University investigated various examples of possible applications of parsers.

The following example from their research paper "Natural Language Processing With Prolog in the IBM Watson System" illustrates a way Watson is using this application when looking at *Jeopardy!* clues.

Category:	Poets & Poetry
Clue:	He was a bank clerk in the Yukon before he published "Songs of a Sourdough" in 1907 (1)

This example includes the following base forms (lemma):

Subject:	"he"
Verb:	"publish"
Object:	"Songs of a Sourdough" (2)

This identification enables Watson to use appropriate rules and apply them to the type of the category as well as the structure of the clue.

Focus of the question:	words that refer to the answer ("he")
Lexical answer types:	terms in the question or category that indicate what type of entity is being asked for ("poet")
Relationships:	between entities in either a question or a potential supporting passage (2)

Watson uses Prolog in order to make the language of the category as well as the clue more understandable and accessible. Certain elements can then be used to extract information from the original question to generate an appropriate answer. Prolog facts can include the following numbers which "represent…unique identifiers for parse nodes" (2):

```
lemma(1, "he").
partOfSpeech(1,pronoun).
lemma(2, "publish").
partOfSpeech(2,verb).
lemma(3, "Songs of a Sourdough").
partOfSpeech(3,noun).
```

```
subject(2,1).
object(2,3). (2)
```

After analyzing the components of the sentence, the system then generates the “authorOf relation” (2):

```
authorOf(Author,Composition) :-
        createVerb(Verb),
        subject(Verb,Author),
        author(Author),
        object(Verb,Composition),
        composition(Composition).

createVerb(Verb) :-
        partOfSpeech(Verb,verb),
        lemma(Verb,VerbLemma),
        member(VerbLemma, ["write", "publish",...]). (2)
```

Watson uses its database and can identify a high number of potential text passages, which is useful in order to generate the correct response. However, considering the text passage “Songs of a Sourdough by Robert W. Service” (3) it would end in an error. Lally and Fodor explain that there are “many other clauses of the authorOf relation that match different expressions of the same semantic relation” (3). Therefore, the text passage can be used and described in Prolog as follows:

```
authorOf(Author,Composition) :-
        composition(Composition),
        argument(Composition,Preposition),
        lemma(Preposition, "by"),
        objectOfPreposition(Preposition,Author),
        author(Author). (3)
```

This information enables Watson to combine both relations and to determine that the confident answer to the question is “Robert W. Service”. This exemplifies the diverse techniques that are embedded in Watson. The system is, therefore, able to solve problems including “pattern matching”, “depth-first search”, and “backtracking” (3).

The classification of the questions allows Watson to identify the type of question. Parts of the question, such as ambiguous terms or clauses can be identified and classified.

Through LAT detection the type of question can be classified and Watson can choose the category of the explicit question. The system uses LAT to create an internal network of various types. Watson also has the ability to use candidate answers from previous games to generate a response to a question. This, however, does not suffice to cover a broad spectrum of natural language questions. Therefore, Watson uses relation detection to relate syntactic components and semantic relations within the clue.

Example:

Category: "Church" And "State"

Question: It's New Zealand's second-largest city

Answer: What is Christchurch? ("Jeopardy! - The IBM Challenge" Day 2)

After going through the first few stages, Watson was able to decompose the question in the most reasonable way. The advantage of decomposing a question lies in the ability to collect more evidence to support the hypothesis and to generate a higher confidence. DeepQA uses decomposition to process clues that consist of sub-clues or it embeds sub-clues through an "end-to-end QA system…by a customizable answer combination component" (Ferrucci et al. 70) (see fig. 9).

Following the question analysis the hypothesis generation produces candidate answers, which are labeled with a certain amount of confidence. This step of the architecture is part of the primary search which skims through resources to find distinct contents to answer questions by considering, speed and accuracy. Primary search enables the system to rank the top 250 candidates and generate the correct answers within these candidates with a precision of eighty five percent. Underlying approaches of text search includes document search, passage search, knowledge base search (e.g. SPARQL), which are used for primary search. "The SPARQL query language…supports conjunctions (and also disjunctions) of triple patterns, the counterpart to select-project-join queries in a relational engine" (Neumann and Weikum 647). The following example from the article "RDF-3X: a RISC-style engine for RDF" by the German computer scientists Thomas Neuman and Gerhard Weikum illustrate the search for all movies starring "Johnny Depp":

```
Select ?title Where {
?m <hasTitle> ?title. ?m <hasCasting> ?c.
?c <Actor> ?a. ?a <hasName> "Johnny Depp" } (647)
```

SPARQL is then able to list all the corresponding names of the movies. The results can then be organized in graphs and used for further analyses.

During this stage a high number of candidate answers are being generated which allows further stages to focus more closely on precision. Soft Filtering is based on machine learning and uses scoring algorithms to reduce the number of candidates by filtering them through a threshold. This can be achieved by evaluating the candidates and finding evidence that a candidate is for example a part of LAT. Thereby, the number of candidate answers is limited to approximately one hundred (Ferrucci et al. 72).

Through hypothesis and evidence scoring the remaining candidate answers are evaluated further and additional evidence is used to support or neglect these hypotheses. For example, passage search is an addition to the primary search where the candidate answers are used to retrieve specific evidence from the database which relate to the context of the question. After finding evidence for the hypotheses a score is generated which shows the confidence that the system has towards a candidate answer. Watson uses more than fifty scores to rank hypotheses. These scorers determine the following dimensions: "Taxonomic, Geospatial (location), Temporal, Source Reliability, Gender, Name Consistency, Relational, Passage Support, [and] Theory Consistency" (73).

Example:		
Question:	He was presidentially pardoned on September 8, 1974.	
Answer:	Nixon.	
Retrieved passage:	Ford pardoned Nixon on September 8, 1974.	
Passage scorer A:	counts equal terms between question and passage (e.g. "pardoned", "on", "September", "8", "1974")	
Passage scorer B:	measures longest equal word sequence between question and passage (e.g. "September 8, 1974")	
Passage scorer C:	measures logical alignment of question and passage	
	Logical alignment:	question asks for object
		Nixon in passage is object
		Nixon receives high score (72)

Another scorer is geospatial reasoning which would, for example, give "New York" a higher score than "Tokyo" when asking for an American city. Geo-coordinate information would give "Iceland" a higher score than "Italy" when asking for a country in Europe that is farthest North. Temporal reasoning would give "Clinton" a higher score than "Reagan" when

asking for people living in the White House in the mid-1990s. All these and many more algorithms are used to improve Watson's confidence by evaluating evidence (72-73).

Example:	
Question:	Chile shares its longest land border with this country
Wrong Answer Search Engine:	Bolivia (e.g. more news article about the relation between the two countries)
Correct Answer Watson:	Argentina (73)

Algorithms like geospatial reasoning score Argentina higher than the popular scoring of the search engine. This shows the difference between these two systems. Merging and ranking the right answers is crucial to Watson's success.

One of the last steps is merging and ranking. Merging is an essential step because the different algorithms can score the same or very similar results and it would come to overlapping of surface forms. Therefore, certain answer scores are merged together before ranking them. This happens through matching, co-reference resolution algorithms, and enables the system to combine scores.

The machine-learning approach enables DeepQA to rank the merged scores and to estimate confidences. Watson uses an intermediate model which groups scores in regards to specific domains. The mixture of experts and stacked generalization allows Watson to learn and use deep analytics. Watson's learning ability enables the system to use different techniques for different questions which is essential to deal with the vast field of natural language questions. Therefore, the system requires confidence scoring (Ferrucci et al. 74).

In order to score a high confidence Watson uses an algorithm called LFACS (Logical Form Answer Candidate Scorer) which evaluates evidence and judges if a text "passage provides support for a specific, designated candidate answer" (Murdock 4). LFACS uses the following steps in this approach:

(1) Local Match Construction
(2) Global Map Construction
(3) Candidate Inference Construction
(4) Match Evaluation (4)

NLP is difficult for computational systems, because various algorithms have to be combined to increase the generation of useful hypotheses and proving these hypotheses with

appropriate evidence. Only by a proper application of natural language algorithms will the generation of useful answers enable systems like Watson to be applicable in the future.

6.3. Watson and Natural Language

Watson and his capability to play *Jeopardy!* is a scientific grand challenge which includes not only understanding *Jeopardy!* questions but understanding natural language as a whole. These are the foundations which will have an enormous impact in the near future. The game *Jeopardy!* allows scientist to experiment with NLP, which includes measuring automatic open domain question answering. *Jeopardy!* question are richly formulated and deal with a vast spectrum of knowledge. High precision, an accurate confidence, and speed are all aspects which are relevant for future applications ("How Watson Works").

Computers have compared to humans specific strengths and weaknesses. They are exceptionally good with math, arithmetic, and scientific computation. Considering this example:

Question: What is ln((12,546,798*π)^2)/34,567.46 = ?
Answer: 0.00885 ("How Watson Works")

Computers can generate the answer instantly, whereas, humans have difficulties solving these kinds of mathematical equations. However, humans are exceptionally good in understanding and producing natural language. Natural language is implicit, highly contextual, ambiguous, and often imprecise. Nevertheless, humans are able to decipher most variances of natural language instantly. This is, on the other hand, a very difficult task for a computer ("How Watson Works").

Example:
Question: Where was X born? ("How Watson Works")

In this case “X” can be any person. Therefore, if this exact question is programmed into a database, the computer will have no difficulties to generate the answer.

Table 2

Example of a database ("How Watson Works")

	Person	Birth Place
A.	Einstein	ULM
B.	Nietzsche	RÖCKEN
C.	Columbus	GENOA
…	…	…

When information is structured (see Table 1) computers can compute answers fast and with a high confidence. A problem only occurs when computational systems have to work with unstructured information.

Unstructured information entails usually various sources which the system has to evaluate. An example for a source could be this sentence:

> One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein's birthplace. ("How Watson Works")

English speakers will not have difficulties understanding this sentence and can easily use the information in the statement to answer the question: "Where was Einstein born?". Computers on the other hand, would have problems answering this question by using this statement. For once, there are several names included in this source. Secondly, it does not say born anywhere in the information provided, it only says birthplace. There are many complex aspects in this source and for a computer program it is difficult to generate the correct answer. Watson, however, uses various algorithms which enable the system to use all the information provided to generate a degree of certainty ("How Watson Works").

Example:

Category:	Also On Your Computer Keys
Question:	It's an abbreviation for grand prix auto racing
Correct Answer:	F1
Watson's Answer:	gpc ("Jeopardy! - The IBM Challenge" Day 3)

The type of questions can be specific, but also very vague. Indicators can help the system to decipher the clue but are sometimes deceptive. Watson had strong difficulties with the category "Also On Your Computer Keys" and generated incorrect answers. The example shows that a big database does not enable the system to answer any question. Only a minimal

percentage of the data is covered. The idea behind Watson is, therefore, not to anticipate questions and use gigantic databases, but to understand natural language. NLP technology can be used to analyze any kind of “as-is” texts (e.g. encyclopedia, plays, dictionaries, books). The focus is on creating a system that is smart enough to use existing information. Therefore, a system like Watson is able to analyze unstructured contents and to compute answers as well as confidences, whereas, the foundation to understand these unstructured sources is provided by structured information ("How Watson Works").

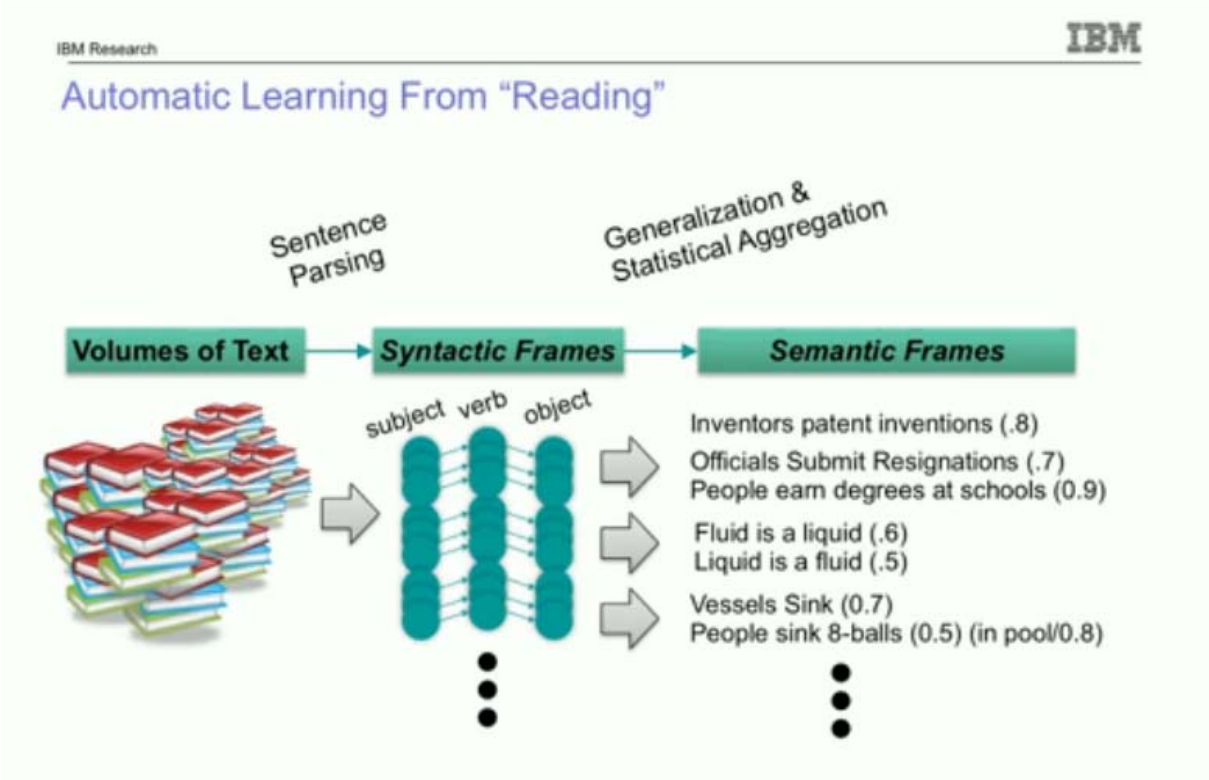

Fig. 10. Visualization of Automatic Learning from “Reading” ("How Watson Works")

Watson is able to process data that the system reads and automatically learns from it (see fig. 10). Watson does not understand the reading context in the depth as humans do. However, the system is able to use structures and the semantics of sentences. It is able to relate to other sentences and recognizes sentence modifications. Through sentence parsing Watson can use syntactic frames. The system can identify subjects, verbs, and objects, and can recognize the relation between them. Watson then produces graphs and statistics by using interrelating algorithms which include all of the generated information. The semantic frames are tagged with a confidence score. Through relating to different sources Watson learns, for example, that “inventors patent inventions” with a confidence of 0.8. That “earn” can be used in the context of: “people earn degrees at school” with a confidence of 0.9. Watson identifies that the term “sink” can stand in relation with “vessel sink” with a confidence of 0.7.

However, “people sink 8-balls” has a higher confidence (0.8) when it is recognized in the context of the game pool, whereas, if the context is disregarded it has only a confidence of 0.5 ("How Watson Works").

Watson’s algorithms do a lot of searches and calculate hypotheses. These hypotheses are then being evaluated. An example from the IBM research will allow a better understanding ("How Watson Works").

Example:
Question: In cell division, mitosis splits the nucleus & cytokinesis splits this liquid cushioning the nucleus. ("How Watson Works")

After receiving the question Watson generates candidate answers. In this example they could be: “organelle”, “vacuole”, “cytoplasm”, “plasma”, “mitochondria”, “blood”, and so on. As already mentioned, Watson generates hypotheses in the question’s context. After the DeepQA architecture evaluated all the information, it generates intermediate hypotheses including confidences.

Example:
Is (“cytoplasm”, “liquid”) = 0.2
Is (“organelle”, “liquid”) = 0.1
Is (“vacuole”, “liquid”) = 0.2
Is (“plasma”, ”liquid”) = 0.7 ("How Watson Works")

By scanning various texts, Watson discovers for example this sentence:

“Cytoplasm is a fluid surrounding the nucleus…” ("How Watson Works")

One problem the system is facing is the question, if a fluid is also a liquid. Many algorithms are used to decipher various kinds of information. In this particular example, the previous investigated ontology WordNet can be used in order to answer the question. WordNet investigates if fluid is also a liquid ("How Watson Works").

WordNet: is_a(Fluid, Liquid)? ("How Watson Works")

The algorithm uses data from the physical knowledge of fluid and liquid. The result is that a liquid is a type of fluid. However, fluid is not a type of liquid. Therefore, WordNet does not have enough evidence to support this hypothesis. Nevertheless, Watson uses also learn resources which are all the information extracted from texts. In this regard, Watson’s process

of learning is similar to the way humans use language. People sometimes consider fluid to be a liquid. This can be seen in this intermediate hypothesis ("How Watson Works"):

Is (“cytoplasm”, “liquid”)= 0.2 ("How Watson Works")

Therefore, Watson learned that:

is_a(Fluid, Liquid) = YES ("How Watson Works")

This is only one example of the way Watson learns information ("How Watson Works"). Watson uses context by having a big database to understand specific questions. This is very similar to the way humans learn and use learned information in different contexts.

Answering questions correctly and learning from sources is dependent on the evidence. Some evidence is more trustworthy than other. IBM researchers exemplify this notion by distinguishing between keyword search and deeper evidence ("How Watson Works"). In this section the previous example of keyword search versus semantic search is going to be extended by using examples of specific algorithms. An example will illustrate this subject:

Example:

Question: In May 1898 Portugal celebrated the 400th anniversary of this explorer’s arrival in India.

Answer: Vasco da Gama ("How Watson Works")

A keyword passage demonstrates the deceptive evidence:

Passage A: **In May**, Gary **arrived in India** after he **celebrated** his **anniversary in Portugal**. ("How Watson Works")

The keyword search indicates seven matching keywords which include “May”, “arrived”, “in”, “India”, “celebrated”, “anniversary”, and “Portugal”. This evidence suggests that “Gary” is the potential answer. However, Watson is able to learn that keyword search can be weak evidence compared to other methods. For instance, keyword search can lead to deceptive evidence, when the system uses this passage to support the previous keyword search result:

Evidence source: And Gary returned home to explore his attic looking for a photo album. ("How Watson Works")

After reading this source, it would classify “Gary” as an explorer. “Gary” is the subject, which is directly related to the verb “explore”. Therefore, this source can be seen as legitimate evidence of the keyword search. However, this evidence does not comply.

Certain algorithms can search for deeper evidence to find the correct answers. Watson could use this passage, for example:

> Passage B: On the 27th of May 1498, Vasco da Gama landed in Kappad Beach. ("How Watson Works")

At first glance, this passage appears to be of minor importance to the system. The keyword search signals only one common word, which is “May”. “May” is in addition an ambiguous term and appears in a large number of texts. Consequently, stronger evidence is more difficult to isolate. This is the idea behind the vast amount of algorithms used in Watson. The combination of algorithms allows Watson to explore many hypotheses and match the evidence ("How Watson Works").

One of the many algorithms is temporal reasoning. Temporal reasoning allows “generating conclusions from time-oriented data based on the latter’s time-oriented attributes (e.g., temporal duration) and their temporal relationships to other data (e.g., temporal order)” (Nguyen et al. 122). This algorithm scans the question and can detect the connection between “May 1989” and “400th anniversary” and relate it to “27th May 1498” of passage B ("How Watson Works").

The statistical paraphrasing algorithm can analyze the words “arrival in” from the question and can related it to “landed in” from passage B. It learns from different sources that these terms can appear in similar contexts. Therefore, a confidence allows the system to assume a similar meaning ("How Watson Works").

Another algorithm is geospatial reasoning. This algorithm can be used in various ways. For instance, it “often takes the form of geographic information systems (GIS) sophisticated systems that combine computational geometry with database techniques to provide powerful abilities to manipulate and visualize vast quantities of digital terrain data” (Forbus, Usher and Chapman 61). However, in systems like Watson geographic reasoning focuses on word files which include geographic regions and cross-reference them to other areas. Therefore, geospatial reasoning can, for example, identify “Kappad Beach” to be a place in “India”. Watson is, therefore, able to generate a hypothesis with a good confidence that “Vasco da Gama” is an explorer and the correct answer to the question ("How Watson Works").

In *Jeopardy!*, as soon as Watson has generated the answer with an appropriate confidence the system can buzz-in and respond to the question. However, only if Watson was able to generate the confidence in time it will do so. The difficulties of the questions are not equal. Sometimes it takes Watson longer to compute the answer and confidence. This also includes betting decisions. If the confidence is low, Watson manages the risk differently. The more questions Watson answers the more the system learns. The method of learning through experience can be exemplified by Watson's ability to learn within a category.
The following example shows how Watson learns within the category "Celebrations of the Month" (see fig. 11).


Category: **Celebrations of the Month**

Clue	Type	Watson's Answer	Correct Answer	
D-Day annivesary & Magna Carta Day	day	Runnymede	June	✗
National Philanthropy Day & All Soul's Day	day	Day of the Dead	November	✗
National Teacher Day & Kentucky Derby Day	day/month(0.2)	Churchill Downs	May	✗
Administrative Professionals Day & National CPAs Goof-Off Day	day/month(0.6)	April	April	✓
National Magic Day & Nevada Admission Day	day/month(0.8)	October	October	✓

Fig. 11. Example of Watson's ability to learn within a category ("How Watson Works")

The category asks for a month, but Watson chooses in the first two clues the type "day" and gives the answers "Runnymede" and "Day of the Dead". The system adjusts and develops its confidence. After the third clue is revealed Watson still gives the wrong answer ("Churchill Downs"), however, the system generated a 0.2 confidence for a month. This confidence increases during the next questions to 0.6 and 0.8. Consequently, Watson is able to answer the last two questions correctly.

Watson is programmed to learn prior to a game, but has also the potential to learn during a game. In this category and in every *Jeopardy!* game Watson is given the correct answer of a question. This enables Watson to learn and adjust itself within the category. Whereas, this is relatively easy for a human, Watson has difficulties in this process. However,

as the example shows over time the system gets smarter with every question ("Learning across categories").

7. Critique on Watson and *Jeopardy!*

Even though Watson did well on *Jeopardy!* there are still some flaws in Watson's performance. The answers generated by Watson seem at the surface like the answers of human candidates. However, by understanding Watson's structure, algorithms, and natural language processing capabilities the difference between the human brain and microprocessors come apparent. For instance, at day two of the *Jeopardy!* challenge Watson answered the final *Jeopardy!* question incorrectly. The difficulty of final Jeopardy! is to bring two information from two different areas together.

Final *Jeopardy!* question:	
Category:	U.S. Cities
Clue:	Its largest airport is named for a World War II hero; its second largest, for a World War II battle
Answer:	What is Chicago?
Watson's answer:	What is Toronto????? ("Jeopardy! - The IBM Challenge" Day 2)

Humans can understand and answer this question relatively easily. The difference to Watson is that the system does not use one big structured database but uses natural language content and tries to understand it. Watson's confidence was very low for this answer, which is signaled by the five question marks. The issue Watson has to deal with is to find enough evidence to support its hypotheses. Humans remarkably connect instantly the category "U.S. Cities" with the clue and can solve the question, whereas, Watson learned that a *Jeopardy!* category does not necessarily relate to the clue.

Watson's fourteen percent confidence for Toronto could be explained with the fact that IBM has not released a detail report about the analysis of this question. One explanation is that there are seven cities with the name Toronto in the United States. Furthermore, it can be assumed that Watson found evidence that the name "U.S." is often referred to "America" or "American". The geographical location of Toronto is in "(North) America". Toronto also has an airport which is named after a World War I hero. An assumption is that the keyword search could have retracted the words, "World", "War", and "hero" from a source. Furthermore, the name Toronto is connected to the United States, because it is for once bordering the U.S. and has a baseball team which plays in the American League (Baker).

A source that would support the hypothesis that Toronto is an American city could be the itinerary of *The Numerati* book tour from the American journalist and author of *The*

Numerati Stephen Baker, which includes American cities, such as Philadelphia, San Francisco, and Toronto. The weakness of Watson is based on its statistical analysis (Baker). However, even though Watson generated the wrong answer, it indicated its weak confidence, which shows that the system was not able to find enough evidence to support its hypothesis.

The success of Watson in *Jeopardy!* has to be seen critically. Luck is relevant when playing *Jeopardy!* including clue selection and betting strategy. As seen in the last game of the *Jeopardy!* challenge, Watson uncovered both Daily Doubles, which it got wrong both times. If one of the human contestants would have been able to find the Daily Double first, their wager could have impacted the game entirely. However, the total earning does not reflect how well a person or Watson can answer questions. This also includes Final *Jeopardy!*, where a good betting strategy can make the difference of winning and losing. Therefore, the *Jeopardy!* challenge does not represent Watson's superiority in the field of QA. It rather generates public interest in Watson and IBM achievements in the development of QA systems. Nevertheless, what is more interesting, are the successes in improving precision, confidence, and speed when dealing with natural language questions. Luck and clue selection have to be disregarded and the focus has to be on the analysis of precision. The ability of the system to choose which question to answer and which not depends on the confidence and the threshold. This improves accuracy and trust in the system (Ferrucci et al. 63-65).

8. Watson's Future

By winning the *Jeopardy!* Challenge and showing its ability to dominate a broad spectrum of natural language questions Watson demonstrated an outstanding performance. Watson has the potential to influence many industries. Applications can envelope many areas of business, e.g. healthcare by giving doctors diagnostic assistance, which includes evidenced-based collaborative medicine. Also, technical support can be used by help desks and contact centers. The government could improve information sharing and security ("How Watson Works").

The society of today is overwhelmed with information. Information is increasing dramatically "at an annual compound rate of 57% and nearly 6 terabytes of information are being exchanged over the Internet every second" ("Power your Planet"). The challenge is to use this data and turn it into useful knowledge. Information that can be found in journals or the web is for the most part unstructured. Watson could be used as a tool which can read as well as understand data and retain information. A system like DeepQA could generate answers to the ever growing amount of questions.

For instance, the information in financial markets rises about seventy percent annually. Intelligent computer systems are needed to analyze information such as news articles and financial blogs, which are all written in natural language. The necessity of such a system can be justified by occurrences like the financial crisis of 2008/09. Executive of financial services Jay Dweck understands the advantages of systems like Watson and states:

> The reason of the financial crisis highlights the problem of sustaining risk. One thing that causes risk is interdependency and the failures that starting to go like dominos and you can use something like Watson to understand what creates those interdependencies. ("Watson after Jeopardy!")

Businesses could monitor financial markets and all economic situations simultaneously which would lead to the ability to predict better strategies for companies. Dweck concludes: "It can put together the logical connections among the desperate pieces of information that it absorbed" ("Finance").

Other beneficiaries are private banking, insurances, and call centers. The idea behind this progress is to bridge the gap between customers and partners of businesses. Dr. Paul Bloom from IBM Research Telecommunications believes that systems like Watson can increase service quality and provide faster answers. Call centers are troubled with a large number of phone calls every day. The number of operators does not suffice and current

automated call center computers work to slowly and often incorrectly, which leads to customer frustration. Watson's ability to understand natural language could improve the situation ("Customer Services").

Watson is not only able to give answers to questions, but the *Jeopardy!* challenge enabled the researches to pay very close attention to the issue of confidence. In areas such as health care, doctors have to have a good amount of confidence to make the right diagnosis. In order to have a high confidence a large amount of information is required. Professor of clinical medicine Dr. Herbert Chase, from Columbia University, estimates that: "At least 30 years is humanly impossible for a physician to master all the material they need to practice at the highest level. Biomedical Literature has doubled in size every seven years but the patients want those facts of the doctor's fingertips" ("Healthcare"). Watson can analyze a lot of resources, such as family history, patient history, medications as well as tests and compare it to texts, journals, and various types of other databases. This enables Watson to generate diagnoses which is basically a set of hypotheses. After analyzing the evidence a confidence score will tell the doctor if the diagnoses can be trusted or not ("How Watson Works"). Watson is a tool which can use encoded data and provide suggestions wish will support the decisions of the medical staff. Also, information like this will be available all around the world which would improve quality and reduce costs. These new tools will make industries more efficient and can improve societal issues ("Watson after Jeopardy!"). Dr. Chase summarizes: "It is the effective and efficient storage, retrieval, analysis, and use of biomedical information to improve health" ("Healthcare").

In order to increase the retrieval from sources like the Internet, emails, reports, and so on, Watson could be used to create knowledge, which is essential to deal with the large amount of information. Other examples are:

- Shaving off just seconds per call to find the right technical documentation in call-centers can save millions.
- Rapidly detecting emerging trends in problem-reports coming in from all over the globe can avoid recalls and save companies and their customers millions if not billions.
- Detecting otherwise unrealized drug interactions through analyzing the linkages in of medical abstracts can help prevent disaster as well as help discover new drugs or cures.
- Analyzing communications linked to terrorist networks in the form of multi-lingual text or other modalities can help uncover plots threatening national security before they happen.
- Analyzing SEC reports to help evaluate corporate financial positions. ("The Knowledge Rush")

All of these applications and many more rely on the retrieval of knowledge from a vast amount of unstructured information which for the most part exists as natural language. Watson is able to analyze as well as relate different sources and use the generated knowledge. The generated knowledge can then easily be used by the applicant. Unstructured information is then available as structured knowledge which can be used for various purposes.

9. AI Research Programs and Knowledge Representation

Many institutions, besides IBM, develop a variety of possible applications for natural language processing AI systems. Researchers from the University of Darmstadt Iryna Gurevych and Mark-Christoph Müller investigating in their publication "Information Extraction with the Darmstadt Knowledge Processing Software Repository" how NLP systems can be used "to create a highly flexible, scalable and easy-to-use toolkit that allows rapid creation of complex NLP pipelines for semantic information processing on demand" (1). The idea behind this development is to retract "different levels of linguistic and application specific processing" (1). Therefore, the Darmstadt Knowledge Processing Software Repository (DKPro) can be used in various fields of NLP applications.

Examples of current applications:	
Semantic Information Retrieval:	in the domain of electronic career guidance, computing semantic, relatedness of words, constructing lexical semantic graphs
Question Answering for eLearning:	question answering by mining FAQs, question paraphrase recognition, automatic quality assessment, comparative analysis of user generated discourse (2)
Examples of future applications:	
DKPro information retrieval:	components supply functionality for all phases of information retrieval, including indexing, retrieval, and (qualitative and quantitative) evaluation
DKPro components for text mining:	include readers for importing text from specialized sites like FAQs, forums like e.g. Nabble, social Q/A sites like YahooAnswers, and Technorati (1)

One of the most important developments in AI, however, is to be concerned with knowledge representation. Researcher and lecturer of information science Katrin Weller describes in her book *Knowledge Representation in the Social Semantic Web* that today's society is dependent on technologies that are able "to structure and store information and, [are able]…to find and retrieve it precisely and effectively" (17). These technologies can be implemented by analyzing and, consequently, improving the usage of the World Wide Web. The advancement of the WWW could be the Semantic Web, because "data should be provided in such a way that not only humans can read it; computers should also be able to

manipulate and recombine the information meaningful" (53). The demand for question answering systems is increasing. A simple keyword search does not suffice to satisfy the demand for the ever growing amount of complex queries. Systems need to be able to combine information from different sources and retrieve the desired answer.

These developments can also be applied to the current transformation of the WWW into the Social Web in which "the borders between 'consumers' and 'producers' of content are blurring" (68). The combination of the Social Web with the Semantic Web can create a Social Semantic Web that can improve the usability of networks. The first attempts of these developments are semantic wikis and semantic blogging. However, these are just the beginning "to enable better access to information by providing a vocabulary for associating documents with content-descriptive keywords" (94).

Another aspect of the Social Semantic Web is ontology engineering. For example, community-base ontology engineering influences the Web already today:

- Can handle broad as well as specific domains,
- Can take over the task of ontology maintenance (missing concepts can be added by any community member),
- Is the key to addressing WWW-wide ontologies,
- And can capture the point of view of the user community. (375-376)

The interesting aspects of this development are the potential applications in the future. Weller gives further examples how people could use tools like ontology engineering in the Web:

- Include different levels of knowledge networking, both social networks and data networks,
- Support ontology engineering, semantic indexing and retrieval within one system,
- Enable semantic upgrades from tags or lightweight semantics up to ontologies,
- Provide incentives for easy user contribution, like playful approaches (gwap), direct profit or feedback for contributors. (376)

Understanding Watson allows examining varieties of algorithms and ontologies that are used in a highly complex system. Yet, research projects are numerous and the combination of these results can improve a large amount of possible areas of application. IBM plays a big role, but it cannot cover all facets of the development of AI systems. Smaller projects like DKPro or community-based ontologies that are able to shape the current layout of the WWW, will play a significant role in the development of AI systems.

10. Conclusion

The definition of AI is inherently difficult, especially by trying to connect it to human intelligence under the consideration of knowledge, learning, randomness, and natural language. The human mind and artificial intelligent systems appear only alike on the surface. The output might be in some cases same or similar, however, the natural language processes are very different. Ideas such as linking words with other words might be related to human thinking processes. They are, however, in the overall composition fairly narrowly connected.

Artificial intelligence is related to human intelligence, but it is not necessarily the same. The complexity of AI systems will increase over time. Further research will allow AI systems to surpass human intelligence. So far, natural language is one of the domains that humans are exceptionally good at. Computers have been proven to be extremely good with mathematical calculation. Now, with the development of Watson, computational systems are being developed that can use information that up until now only humans were able to use.

It is a great achievement that human originality can develop systems that are able to use natural language. Successes like Watson build the foundation to develop independently and autonomously working AI systems. Already today the advantages of such programs can be foreseen. Businesses will become even more efficient, but so will governmental organizations and the medical industry. The advances being achieved today will generate other novel ideas and influence future studies.

The development of systems that are able to use natural language will be the ones that will shape the future of human society. Organizations that do not have access to the new systems will fall behind and will face significant problems. Whereas, organizations that use these systems can gain strategic advantages. The benefits for organizations will also be reflected in the potential usage of AI systems by commercial customers. Systems like Watson will not, however, handle all processing work. These systems can be used as a tool to aid people by evaluating the increasing amount of information.

Watson is a very advanced system that can analyze unstructured documents. Besides natural language texts, there are many other forms of natural language data such as voice and image recognition. Nevertheless, the advances and the successes seen on *Jeopardy!* do not reflect actual artificial intelligence. The system itself shows, however, a few patterns which are related to intelligence such as memory or learning abilities.

Fact is that natural language is one of the most important parts of the development of AI systems and Watson allows people a first glance of the potential achievements of these

systems. That AI systems will influence social life is inevitable. The idea behind the development of these new technologies is maybe the final step to unravel the mystery of human intelligence. The future of the development of AI systems will only have one limitation: “The real advances in intelligent system design are only limited by the imagination of designers in the future” (Aleksander 158).

Works Cited

"A System Designed for Answers". *IBM Watson.* 21 February 2011. Web. 04 July 2011. <http://www-03.ibm.com/innovation/us/watson/what-is-watson/a-system-designed-for-answers.html>.

"Apache UIMA". *Apache UIMA.* 2006-2011. Web. 19 July 2011. <http://uima.apache.org/>.

"Blue Gene". *IBM Blue Gene.* 2 December 2008. Web. 9 July 2011. <http://domino.research.ibm.com/comm/research_projects.nsf/pages/bluegene.index.html>.

"Customer Services". *IBM Watson.* 21 February 2011. Web. 25 July 2011. <http://www-03.ibm.com/innovation/us/watson/watson-for-a-smarter-planet/industry-perspectives/customer-services.html>.

"Finance". *IBM Watson.* 21. February 2011. Web. 25. July 2011. <http://www-03.ibm.com/innovation/us/watson/watson-for-a-smarter-planet/industry-perspectives/finance.html>.

"Healthcare". *IBM Watson.* 21. February 2011. Web. 25. July 2011. <http://www-03.ibm.com/innovation/us/watson/watson-for-a-smarter-planet/industry-perspectives/healthcare.html>.

"How Watson Works". *IBM Watson.* 21 February 2011. Web. 21 July 2011. <http://www-03.ibm.com/innovation/us/watson/building-watson/how-watson-works.html>.

"Learning across categories". *IBM Watson.* 21. February 2011. Web. 25. July 2011. <http://www-03.ibm.com/innovation/us/watson/building-watson/watson-sparring-matches/learning-across-categories.html>.

"Power 750 Express Server". *IBM "Power 750 Express Server".* Somers, 12 April 2011. Web. 21 July 2011.

"Power your Planet". *IBM.* Web. 10. July 2011. <http://www-03.ibm.com/systems/power/news/announcement/20100817_ad.html>.

"Space Flight Chronology". *IBM Space Flight Chronology.* Web. 06 July 2011. <http://www-03.ibm.com/ibm/history/exhibits/space/space_chronology2.html>.

"The Face of Watson". *IBM Watson.* 21 February 2011. Web. 04 July 2011. <http://www-03.ibm.com/innovation/us/watson/what-is-watson/the-face-of-watson.html>.

"The Knowledge Rush". *IBM UIMA.* 27 September 2007. Web. 19 July 2011. <http://domino.research.ibm.com/comm/research_projects.nsf/pages/uima.knowledgeRush.html>.

"The Next Grand Challenge". *IBM Watson.* 21 February 2011. Web. 03 July 2011. <http://www-03.ibm.com/innovation/us/watson/what-is-watson/the-next-grand-challenge.html>.

"UIMA and Semantic Search". *IBM UIMA.* 27 September 2007. Web. 19 July 2011. <http://domino.research.ibm.com/comm/research_projects.nsf/pages/uima.semanticSearch.html>.

"UIMA Architecture Highlights". *IBM UIMA.* 27 September 2007. Web. 19 July 2011. <http://domino.research.ibm.com/comm/research_projects.nsf/pages/uima.architectureHighlights.html>.

"UIMA Overview & SDK Setup". *Apache UIMA Development Community.* July 2007. Web. 19 July 2011. <http://uima.apache.org/downloads/releaseDocs/2.2.0-incubating/docs/html/overview_and_setup/overview_and_setup.html#ugr.ovv.conceptual.uima_introduction>.

"Watson – A System Designed for Answers". *Watson – A System Designed for Answers: The future of workload optimized systems design.* Somers, February 2011. Web. 19 July 2011.

"Watson after Jeopardy!". *IBM Watson.* 21. February 2011. Web. 25. July 2011. <http://www-03.ibm.com/innovation/us/watson/what-is-watson/watson-after-jeopardy.html>.

"Watson as a competitor". *IBM Watson.* 21 February 2011. Web. 25 July 2011. <http://www-03.ibm.com/innovation/us/watson/building-watson/watson-sparring-matches/watson-as-a-competitor.html>.

"What powers Watson?". *IBM Watson.* 21 February 2011. Web. 18 July 2011. <http://www-03.ibm.com/innovation/us/watson/watson-for-a-smarter-planet/watson-schematic.html>.

"Why Jeopardy!?". *IBM Watson.* 21 February 2011. Web. 04 July 2011. <http://www-03.ibm.com/innovation/us/watson/what-is-watson/why-jeopardy.html>.

Aleksander, Igor. *Designing Intelligent Systems: An Introduction.* London: Kogan Page Ltd., 1984. Print.

Baker, Stephen. *Final Jeopardy.* 15 February 2011. Web. 26 July 2011. <http://thenumerati.net/?postID=726&final-jeopardy-how-can-watson-conclude-that-toronto-is-a-u-s-city>.

Bieswanger, Markus and Annette Becker. *Introduction to English Linguistics.* Tübingen: Narr Francke Attempto Verlag GmbH + Co. KG, 2008. Print.

Chessbase News. *Chessbase.* 4 November 2006. Web. 9 July 2011. <http://www.chessbase.com/newsdetail.asp?newsid=3468>.

Christian, Brian. *The Most Human Human.* New York: Doubleday, 2011. E-book.

Clampitt, Phillip G. *Communicating for Managerial Effectiveness.* Thousand Oaks, California: SAGE Publications, Inc., 2010. Print.

Fellbaum, Christiane. „Princeton University.“ 21. June 2011. *WordNet a lexical database for English.* Web. 25. July 2011. <http://wordnet.princeton.edu/>.

Ferrucci, David. *UIMA and Semantic Search - Introductory Overview.* Yorktown Heights, NY, 2006. Web. 06 July 2011.

Ferrucci, David, et al. "Building Watson: An Overview of the DeepQA Project." *AI Magazine* Fall 2010: 59-79. Web. 10 July 2011.

Forbus, Kenneth D., Jeffrey Usher and Vernell Chapman. "Qualitative Spatial Reasoning about Sketch Maps." *AI Magazine* Fall 2004: 61-73. Web. 10 July 2011.

Garber, Steve. *NASA Apollo 11 30th Anniversery.* 20 September 2002. Web. 06 July 2011. <http://www.hq.nasa.gov/office/pao/History/ap11ann/introduction.htm>.

Google. “Going rate for leasing a billboard near Triborough Bridge.” *Google Web Search.* Web. 19 July 2011.

Gurevych, Iryna and Mark-Christoph Müller. *Information Extraction with the Darmstadt Knowledge Processing Software Repository.* Darmstadt, 10 July 2008. Web. 20 July 2011.

Hofstadter, Douglas R. *Gödel, Escher, Bach: an Eternal Golden Braid.* New York: Basic Books, Inc., 1979. Print.

Hornby, A. S. *Oxford Advanced Learner's Dictionary.* Ed. Sally Wehmeier, et al. 7th. Oxford: Oxford University Press, 2005. Print.

Janik, Allan. "Tacit knowledge, Rule-following and Learning." Göranzon, Bo and Magnus Florin. *Artificial Intelligence, Culture and Language: On Education and Work.* Avon: Springer Verlag, 1990. 45-55. Print.

Jeopardy! Dir. Kevin McCarthy. "Jeopardy! - The IBM Challenge". Prod. Harry Friedman. 2011. Video.

Krauthammer, Charles. “Psyched Out by Deep Blue.” *The Washington Post.*The Washington Post, 16 May 1997. Web. 9 July 2011.

Lally, Adam and Paul Fodor. "Natural Language Processing With Prolog in the IBM Watson System." 24 May 2011. Web. 15 July 2011.

Minkel, J. D., S. Banks and D. F. Dinges. "Behavioral Change with Sleep Deprivation."

Moustakas, Clark E. *Heuristic research: design, methodology, and applications.* Newbury Park: Sage Publications, Inc., 1990. *Google Book Search.* Web. 25 July 2011.

Murdock, J. William. "Structure Mapping for Jeopardy! Clues." 2011. Web. 29 July 2011.

Neumann, Thomas and Gerhard Weikum. "RDF-3X: a RISC-style engine for RDF." *Proceedings of the VLDB Endowment* August 2008: 647-659. Web. 02 August 2011.

Newborn, Monty. "Deep Blue's contribution to AI." *Annals of mathematics and artificial intelligence* 2000: 27-30. Web. 02 August 2011.

Nguyen, John H., et al. "Integration of Temporal Reasoning and Temporal-Data Maintenance into a Reusable Database Mediator to Answer Abstract, Time-Oriented Queries: The Tzolkin System." *Journal of Intelligent Information Systems* January/February 1999: 121-145. Web. 30 July 2011.

Nie, Jian-Yun and Martin Brisebois. "An Inferential Approach to Information Retrieval and Its Implementation Using a Manual Thesaurus." *Artificial Intelligence Review* Mai/June 1996: 409-439. Web. 30 July 2011.

Partridge, Derek and Khateeb Hussain. *Artificial Intelligence and Business Management.* Norwood: Ablex Publishing Corporation, 1992. *Google Book Search.* Web. 05 July 2011.

Sternberg, Robert J. *Handbook of creativity* . Cambridge: Cambridge University Press, 1999. *Google Book Search.* Web. 18 July 2011.

Stickgold, Robert and Matt Walker. *The Neuroscience of Sleep.* London: Academic Press, 2009. 241-248. *Google Book Search.* Web. 25 July 2011.

Tanimoto, Steven L. *The Elements of Artificial Intelligence: An Introduction using LISP.* Seattle: Computer Science Press, 1987. E-book.

Thinking Machines - The Creation of the Computer. Dir. Bruce Nash. Jaffe Productions. 1995. Film.

Transcendent Man. Dir. Barry Ptolemy. Ptolemaic Productions. 2009. Film.

Webster, Guy. "NASA Mars Rover Getting Smarter as it Gets Older." 23 March 2010. *Jet Propulsion Laboratory.* Web. 10 July 2011. <http://marsrover.nasa.gov/newsroom/pressreleases/20100323a.html>.

Weller, Katrin. *Knowledge Representation in the Social Semantic Web.* Berlin: Walter de Gruyter GmbH & Co. KG, 2010. Print.