# Watson and Siri: The Rise of the BI Smart Machine

**Troy Hiltbrand** is a technology strategist for Idaho National Laboratory.
troy.hiltbrand@inl.gov

**Troy Hiltbrand**

## Abstract

**The past few years have seen a significant evolution in human-computer interaction. The era of smart machines is upon us, with automation taking on a more advanced role than ever before and permeating areas that have traditionally been unique to human interaction. This movement has the potential to fundamentally alter the way business intelligence (BI) is executed and deployed across industries as well as the role BI may play in all aspects of decision making.**

## Watson: Human versus Machine in a Battle for Logical Supremacy

In 2011, at the commencement of a special episode of *Jeopardy* pitting man against machine, host Alex Trebek indicated that "you are about to witness what may prove to be an historic competition." He was right.

In this competition, IBM Research charged forward to take the next step in the evolution of computational leadership. This was the follow-up to a groundbreaking 1997 chess match in which IBM's supercomputer Deep Blue faced off with Garry Kasparov, chess grandmaster. That contest proved that a supercomputer could apply programmatic logic to outperform a human master, in this case at the game of chess.

With this new competition, IBM's team was faced with profound and new levels of challenges. With chess, there are predefined rules of movement. Deep Blue focused on analyzing all of the possible outcomes and probabilistically determining the most optimal next move to counter its challenger. It took into consideration patterns associated with Kasparov's past play along with the patterns of many other great chess players.

The *Jeopardy* challenge was inherently different. It required that the machine think like a human *and* interpret language like a human. To up the ante, IBM didn't take on just any *Jeopardy* competitors in their demonstration of computing excellence; they took on Brad Rutter and Ken Jennings, the two most successful champions who had ever played the game. The bar was set high; the team needed to develop a system that would interpret, solve, and respond to clues that spanned many topics presented in various formats.

To accomplish this, researchers developed Watson using natural language processing and text analytics to develop the basis for the human-computer interaction layer as well as a probabilistic approach to identify the best answer for each specific clue. The result displayed not only the answer but also a graphical representation showing the top three potential answers and the probability of being correct. Unlike the chess match, which allowed Deep Blue the traditional chess timing rules to do its analysis and return a result, Watson was under a time crunch to perform all of its computational processing more quickly than its two all-star competitors.

With the complexity of finding the right answer for the given clue paired with the relative messiness of human language inherent within the clues themselves, Watson proved that even with high-power systems and engineering genius on the back-end, it was still a complex challenge. Watson's performance was not without quirks, resulting in a tie at the end of the first day of play.

By the end of the three-day exhibition, Watson came out on top, earning more than $77,147 compared to the $24,000 and $21,600 of its competitors. Ken Jennings, ever a good sport, bowed to the new *Jeopardy* champ. "I for one welcome our new computer overlords," he wrote on his video screen, quoting an episode of *The Simpsons*.

### Siri: Bringing Artificial Intelligence to the Consumer

On October 4, 2011, Apple raised the stakes in the battle for mobile supremacy with its launch of Siri on the iPhone 4s. This innovative feature distinguished the iPhone 4s from its competitors and laid the groundwork to become the digital personal assistant of the future. Siri provided a mechanism for end users to push a button and ask a question, which would then be processed against a multitude of applications on the device (including reminders, calendars, messaging, e-mail, notes, music, clocks, maps, and Web browsers) to either perform a function or return related content. This elevated the mobile phone from a portable computer to a personal digital assistant, freeing the user from needing to know which application should perform the requested function. Now users could speak a simple command and have the phone perform the majority of the processing needed to respond accordingly.

The base technology supporting Siri was important because it went beyond simply doing speech recognition to execute a command. It married recognition with natural language understanding to determine what type of action the end user intended, identify the relevant functionality, execute the command, and return a response within the context of the request. This was significant because it was built into a mobile phone intended to be carried around and provide on-demand access wherever and whenever the end user needed it—the embodiment of computational mobility. Watson, although much more capable in terms of its processing potential and its sub-second response time, required a cluster of 750 computers with 2,880 processor cores on 10 server racks to function, which significantly limited its portability.

### Smart Machines for BI

Watson and Siri have demonstrated that natural language understanding has the potential to fundamentally change how end users interact with computational processing. These same trends also have the potential to fundamentally alter how users engage business intelligence systems in the decision-making process.

Traditionally, business intelligence suites have focused on search and navigation as the mechanism for providing content to end users within a business system's repository. Both of these focus on metadata attached to predefined reports and dashboards. This metadata includes report titles and descriptions, but it is limited in providing a way to find specific answers to questions. This is where natural language understanding bridges the information gap to support business intelligence. Instead of end users typing "sales" into a search bar and needing to know whether they want the report "sales by date" or the report "sales by market segment," users would prefer to type exactly what they're looking for—"Who has sales growth of 10 percent or more?" or "List sales growth at least 10 percent"—and have the engine display a dashboard of sales filtered to show only those segments of the business that have sales of 10 percent or greater. Taking it one step further, users would like to dictate their search requests, no keyboard required.

Complicating the engine's job is that the user's engagement is context sensitive. Unlike the search engine, which can index report metadata in the same fashion for every company, a natural language engagement requires much more context about the content to be effective. For Siri to be effective at answering questions, it must interact with multiple distinct content stores such as maps, calendars, the Web, and so on. The engine has to determine the most probable purpose of the engagement and invoke the mechanism to call that functionality.

> Business intelligence tools of the future can learn from Siri's simplicity as they strip away the complexity associated with knowing how and where to find information and provide a simple and universal interface for users.

To succeed in applying natural language understanding and advancing human computer interaction for business intelligence, engines must address three aspects of the problem: consumption, understanding, and response. These three represent the input, processing, and output stages of system design.

## Consumption

As organizations move into this new paradigm, the first area to address is request consumption.

The traditional method of end users' interaction with a BI suite is to type one or more terms related to the request or to navigate to a predefined location where known information is located. This usually requires multiple steps and in some cases requires end user training to ensure that users understand how to move through the business systems.

To simplify this interface on the iPhone, Apple introduced an advanced voice-to-text system that takes a request in the form of human speech and translates it into a string of text that characterizes the request. Apple removed the barriers of training and complex system interaction by boiling down the interface to the single push of a button. In response to this simple action, Siri is ready to accept any command that the user desires. This opens a conversation stream between human and machine.

Business intelligence tools of the future can learn from this simplicity as they strip away the complexity associated with knowing how and where to find information and provide a simple and universal interface for users. The BI tool must facilitate a request in the language of choice and have the system perform the heavy lifting—converting the request into a set of systematic processes that will supply the user's desired objectives.

Watson didn't use voice-to-text processing, but instead had the clues fed to it at the same time that Alex Trebek read them to the other competitors. This is similar to the way end users naturally inquire with respect to questions about business analytics. It is much more familiar for an end user to ask, "What is happening to the sales in a certain region since we started our marketing campaign?" than to formulate a complex SQL (structured query language) statement, a MDX (multidimensional expressions) statement, or visit a series of screens that will dynamically create the back-end SQL or MDX statement(s).

It is not optimal for an executive with a question to navigate to a portal and navigate to the right report to answer questions. Many organizations have created business intelligence competency centers where executives can send a request and have a team of BI analysts extract and return an answer.

As technology advances in natural language understanding, the process of engaging an analyst to research the question and provide an answer could become a thing of the past. The executive could e-mail or send a text message to a virtual assistant directly and the system would interpret the objective, perform the analysis, and return an answer without the delay of human intervention.

With more public-facing business intelligence solutions enabling customers to perform self-service information gathering, this concept can be extended to social media venues. As questions or requests are made through Facebook, Twitter, Instagram, or a myriad of other social media or communication platforms, these requests can be parsed and their context identified; the request can then be answered without needing a human customer service representative.

This evolution in how requests are consumed and fulfilled will fundamentally change how businesses will work in the future and will have a dramatic impact on the economy as a whole. In the fall of 2013, Gartner predicted that the rise of smart machines will have a deep and widespread impact on businesses through 2020. This prediction includes the potential widespread elimination of millions of middle-class jobs; those employees focused on providing this middleman service may be replaced by smart machines (Gartner, 2013). Although this has serious implications for the state of the economy as a whole, it also means that companies that can get in front of the wave and power the coming evolution will win in the end.

## Understanding

The greatest technical challenge comes after a user makes a request. This includes bringing a level of context and understanding through elements of natural language processing. Language is messy; the same fundamental request can be conveyed in multiple ways, using different words and phrases and through various communication channels. Add the global connectedness of business transactions and the need to support multiple languages and dialects, and the challenge increases dramatically.

To overcome this, natural language processing doesn't try to develop a prescriptive set of rules to follow under every circumstance, but instead uses machine learning and statistical probability to find patterns of speech and likely meanings.

The first step in natural language understanding is taking a string of characters and determining how to break it into parts that can be used to drive processing. These pieces, whether words or phrases, are the basis for interpreting the request.

Parsing a sentence can be as simple as identifying where the spaces are in a sentence and breaking the sentence at these spaces. As punctuation is factored in, the process becomes more challenging. When evaluating a period, the parser needs to distinguish between multiple instances of usage. When a period falls at the end of a sentence, it is not attached to a word but to a sentence and has no relevance to the adjacent word. If it is attached to a word inside a sentence (e.g., Mr. or Dr.), it is associated with the word and not the sentence; it might or might not be able to be stripped away without changing the meaning of the word. If the period falls inside the word, stripping it out might have a more significant impact. For example, "U.S." (with periods) represents a country, whereas without periods it is a pronoun representing the speaker and others.

These complexities apply to other punctuation characters as well. Capitalization can be used to help define sentence boundaries but brings similar challenges in the form of acronyms, mixed case names, and other use of capital letters that are not the norm.

The next step in natural language processing involves identifying the parts of speech of the identified words. Different parts of speech distinctly affect the meaning of the sentence. Nouns represent the entities of concern, adjectives are used to provide additional context to the nouns, verbs are often associated with the action, and prepositional phrases provide context to the sentence as a whole.

Parsers often rely on predefined corpuses of text that have been hand annotated by experts and help define statistical models for determining the part of speech of specific words in specific positions. These corpuses are used to train models that can be applied to an unknown set of text to determine the most probable part of speech for each word in the sentence.

Words are often based on the same root and multiple forms have closely related meanings. Plurals (e.g., tree/trees, ox/oxen, sheep/sheep), gerunds (e.g., water-ski/water-skiing, write/writing, find/finding), and other grammatical vehicles take a word with very similar meaning and mask it so that it fits into a sentence in a different way. Comparing these words in their raw format could miss the fact that the words are meant to achieve the same request. Natural language processing can identify the common root among the different variations of a word, thus identifying the words as related.

Once words are split, parts of speech are identified, and words are "stemmed" to identify commonalities, it is often important to group these words from the sentence back together so that phrases take on their intended purpose. This is because the meaning exists only with the phrase and not with each of the individual words. For example, the phrase "the United States of America" has a specific meaning that the words *united*, *america*, *states*, *the*, and *and* don't have independently.

With words and phrases broken into parts, the next step is to attach meaning to the words. With many words, there are multiple meanings and the instance in question relies on the context of the words surrounding it. For example, the word *bank* could mean a financial institution or the bank of a river. If other words in the sentence indicate money or a physical building, the meaning is likely the financial institution. If the words in the sentence relate to foliage and water, it is probably the bank of a river. These clusters of similarity allow for probabilistic determination of the best meaning of a word or phrase. Processing has to be able to disambiguate the words and phrases based on probabilistic modeling to identify the context.

Words that can be identified as named entities can have even greater importance for meaning. These named entities can include the actual names of people, places, and organizations. A named entity associates a wealth of additional context to the term, expanding the meaning of the phrase. With an organization, this additional context includes attributes such as leadership hierarchy, revenues, stock market information, or a public reputation. With a place, this includes attributes such as population information, information about major exports and imports, geographic location, and a wealth of history. With people, there exist inherent relationships with other individuals, talents, skills, and educational background.

This additional context can be as important to the request as the words in the request. If an end user is looking to see sales for the north region, the phrase *north region* is a geographic designation that is composed of multiple areas. To process this request, the system has to be knowledgeable about this relationship and be able to associate sales for each of the areas and group them into a higher level representing the region.

With a fundamental understanding of the words and phrases associated with the request, the next step is to associate the related task in the system. The system will match multiple facets of the request with the metadata associated with the data to successfully respond to the end user's request. This includes the facets of the five Ws: who, what, when, where, and why.

### Who

From a security perspective, the system must be able to interpret who is making a request and whether that user has permission to view the information. If an executive sends an e-mail message or sends a request in a text message, the system has to associate an identity with the credentials, perform authorization, and determine the accessibility of the information requested.

It must also be able to determine whether the entire breadth of information requested falls within the domain of the user's privileges. If a manager requests information for the entire company, but only has access to regional data, this restriction needs to be considered.

With new forms of interaction at the input stage, new forms of interaction need to be in place at the output stage as well.

If the system is open to the public, the system has to ensure that the information is segmented such that sensitive business information will not be provided in response to a malicious request. If the information is sensitive to one individual, the system must ensure that the information will be provided to that individual but not to others.

### What

Data sources must be tagged with metadata so they can be systematically matched to the "what" of the request. This entails mature metadata management and master data management programs within the organization. As the technology matures, information management will become paramount and organizations that bypass this step will be forced to return and look closely at how to bring it up to par.

### Where

If the information is located across multiple data sources or in multiple data marts, the system must associate the end user's request with the correct data source. This forms a logical mapping that acts as an information traffic director, sending the request in the right direction based on its content and context.

### When

Time frames within requests are often stated in relative terms, not absolutes. The system has to be time aware and must be able to translate the terms "yesterday" or "last week" into instructions that will retrieve the correct data.

### Why

This facet is the most difficult but can also provide the greatest value for users. If the system can apply natural language processing to understand what the user is requesting as well as understand *why* the user is requesting it, then the additional context can increase the breadth of the response and the business value.

If a user sends an e-mail that says, "I am trying to find the sales by region over the last year to determine where to invest marketing dollars," a system that can return sales by region over the past year is valuable. However, a system that can interpret the request as an evaluation of historical marketing campaign investments—and can evaluate the marketing returns by region—would be remarkable.

With a business intelligence competency center, analysts can often read between the lines and understand the *why* associated with the question, allowing them to go beyond providing the simple answer to the question asked. This is where the concept of the smart machine surpasses simply enhancing current BI search using natural language processing. It is here that organizations will find the greatest efficiency in automating knowledge-based work.

## Respond

Finally, once consumption and understanding are in place, the system must respond. With new forms of interaction at the input stage, new forms of interaction need to be in place at the output stage as well.

If a request arrives in an e-mail or text message, returning a traditional graph or chart might not be possible. The information might need to be synthesized into a statement that is as compact as the request. If an executive is using a mobile device in a remote location and sends a request in a text message, that request contains context as to the limitations of the medium for the response. It might be ideal to text back a simple answer while simultaneously sending a heavier, richer set of information via e-mail so the executive can review the information further when a device with a richer interface or larger display is available.

Privacy must be considered when working with a public that is making requests via social media. If an end user uses Twitter to request personal information, there might be risk associated with responding through the same media channel. The mechanism might need to change from a public tweet to a private message, still on the Twitter platform but with different settings for visibility by other users.

An organization needs to clearly understand the risk profile associated with different forms of communication and the level of sensitivity and security of the transmitted information to be able to communicate the right message through the right medium.

Business intelligence organizations and competency centers need to prepare now for the impact they'll experience as smart machines permeate industry.

Just as Apple used speech-to-text as a mechanism for inquiry retrieval, the company also used text-to-speech as a delivery mechanism. This increased the efficacy of the user engagement and increased the perceived value of the response.

## Final Thoughts

Because language is not exact and user intentions are not always clearly articulated, the user engagement might require a multi-step process where the system responds not with an answer but with clarifying questions to provide the necessary context needed to generate the most relevant response.

Watson and Siri whet the appetite of end users for the type of interaction they may have in the future. With recent technological innovation in natural language understanding and human computer interaction, organizations are preparing for fundamental shifts in the way they do business. Business intelligence organizations and competency centers need to prepare now for the impact they'll experience as smart machines permeate industry and drive the economy forward. BI practitioners must determine what solutions they need to put in place to be ready for that coming wave and the state of business interaction it will enable. ■

## Reference

Gartner [2013]. "Gartner Says Smart Machines Will Have Widespread and Deep Business Impact Through 2020," press release, October 10. http://www.gartner.com/newsroom/id/2605015