



American Society for Quality (ASQ) — Washington, DC & Maryland Metro Section (509), Software Special Interest Group (SSIG)  
IEEE Computer Society — Washington, DC & Northern Virginia Chapters  
Society for Software Quality (SSQ) — Washington, DC Area Chapter

# Building Watson

## An Overview of the DeepQA Project

by: David A. Ferrucci

Tuesday January 24, 2012

Computer systems that can directly and accurately answer peoples' questions over a broad domain of human knowledge have been envisioned by scientists and writers since the advent of computers themselves. Open domain question answering holds tremendous promise for facilitating informed decision making over vast volumes of natural language content. Applications in business intelligence, healthcare, customer support, enterprise knowledge management, social computing, science and government could all benefit from computer systems capable of deeper language understanding. The DeepQA project is aimed at exploring how advancing and integrating Natural Language Processing (NLP), Information Retrieval (IR), Machine Learning (ML), Knowledge Representation and Reasoning (KR&R) and massively parallel computation can greatly advance the science and application of automatic Question Answering. An exciting proof-point in this challenge was developing a computer system that could successfully compete against top human players at the Jeopardy! quiz show ([www.jeopardy.com](http://www.jeopardy.com)).

Attaining champion-level performance at Jeopardy! requires a computer to rapidly and accurately answer rich open-domain questions, and to predict its own performance on any given question. The system must deliver high degrees of precision and confidence over a very broad range of knowledge and natural language content with a 3-second response time. To do this, the DeepQA team advanced a broad array of NLP techniques to find, generate, evidence and analyze many competing hypotheses over large volumes of natural language content to build Watson ([www.ibmwatson.com](http://www.ibmwatson.com)). An important contributor to Watson's success is its ability to automatically learn and combine accurate confidences across a wide array of algorithms and over different dimensions of evidence. Watson produced accurate confidences to know when to "buzz in" against its competitors and how much to bet. High precision and accurate confidence computations are critical for real business settings where helping users focus on the right content sooner and with greater confidence can make all the difference. The need for speed and high precision demands a massively parallel computing platform capable of generating, evaluating and combing 1000's of hypotheses and their associated evidence. In this talk, I will introduce the audience to the Jeopardy! Challenge, explain how Watson was built on DeepQA to ultimately defeat the two most celebrated human Jeopardy Champions of all time and I will discuss applications of the Watson technology beyond in areas such as healthcare.

**Dr. David Ferrucci** is an IBM Fellow and the Principal Investigator (PI) for the Watson/Jeopardy! project. He has been at IBM's T.J. Watson's Research Center since 1995 where he heads up the Semantic Analysis and Integration department. Dr. Ferrucci focuses on technologies for automatically discovering valuable knowledge in natural language content and using it to enable better decision making.

As part of his research he led the team that developed UIMA. UIMA is a software framework and open standard widely used by industry and academia for collaboratively integrating, deploying and scaling advanced text and multi-modal (e.g., speech, video) analytics. As chief software architect for UIMA, Dr. Ferrucci led its design and chaired the UIMA standards committee at OASIS. The UIMA software framework is deployed in IBM products and has been contributed to Apache open-source to facilitate broader adoption and development.

In 2007, Dr. Ferrucci took on the Jeopardy! Challenge – tasked to create a computer system that can rival human champions at the game of Jeopardy!. As the PI for the exploratory research project dubbed DeepQA, he focused on advancing automatic, open-domain question answering using massively parallel evidence based hypothesis generation and evaluation. By building on UIMA, on key university collaborations and by taking bold research, engineering and management steps, he led his team to integrate and advance many search, NLP and semantic technologies to deliver results that have out-



performed all expectations and have demonstrated world-class performance at a task previously thought insurmountable with the current state-of-the-art. Watson, the computer system built by Ferrucci's team is now competing with top Jeopardy! champions. Under his leadership they have already begun to demonstrate how DeepQA can make dramatic advances for intelligent decision support in areas including medicine, finance, publishing, government and law.

Dr. Ferrucci has been the Principal Investigator (PI) on several government-funded research programs on automatic question answering, intelligent systems and saleable text analytics. His team at IBM consists of 28 researchers and software engineers specializing in the areas of Natural Language Processing (NLP), Software Architecture,

Information Retrieval, Machine Learning and Knowledge Representation and Reasoning (KR&R).

Dr. Ferrucci graduated from Manhattan College with a BS in Biology and from Rensselaer Polytechnic Institute in 1994 with a PhD in Computer Science specializing in knowledge representation and reasoning. He is published in the areas of AI, KR&R, NLP and automatic question-answering.

Tuesday January 24, 2012

5:30 PM – Networking and Pizza(\*)

5:50 - 6:50 PM – Program

(\*) There is no cost to attend at McLean and Silver Spring.

**Locations:**

The presentation will originate from the IBM facility in New York via MeetingPlace, without VTC:

MITRE-2, room **1N100**  
7515 Colshire Drive  
**McLean, VA 22102**  
host: Scott Ankrum  
cell: 240-731-7581

FDA, Bld 66, room **G512**  
10903 New Hampshire Ave  
**Silver Spring, MD 20993**  
host: James Simpson  
cell: 301-996-4976

Your location  
could be here!

MITRE, room **1M306**  
202 Burlington Rd (Rt. 62)  
**Bedford, MA 01730**  
host: Tim Rice  
cell: 978-758-2704

*If you can host another location via VTC, please contact Scott Ankrum (below)*

**TO ATTEND THE MeetingPlace Collaboration CONFERENCE:**

1. Go to: <http://audioconference.mitre.org/> 2. Click on Attend Meeting. If MeetingPlace Collaboration Window does not automatically open, press connect. 3. Dial your telephone to connect to the audio of the meeting.

- Dial 703-983-6338 (x36338) from the Washington DC region.
- Dial 781-271-6338 (x16338) from the Bedford, MA region.

Meeting ID: 509509, when prompted. Meeting Password: 05090509, when prompted.

Visit <http://audioconference.mitre.org> to test your web browser for compatibility with the web conference. Follow this link to the browser test link on the page.

**Registration:**

Registration Website: <http://www.asq509.org/ht/d/DoSurvey/i/26913>

*You must register by noon on Monday, January 23. If you cannot attend at any location, select telephone dial-in when you register. To RSVP for FDA (Silver Spring), please indicate citizenship. If not a US citizen, please provide your title, employer, and address. Allow two business days for registration before the meeting.*

For registration problems or further information contact Scott Ankrum at: [ankrums@mitre.org](mailto:ankrums@mitre.org) or 703-983-6127

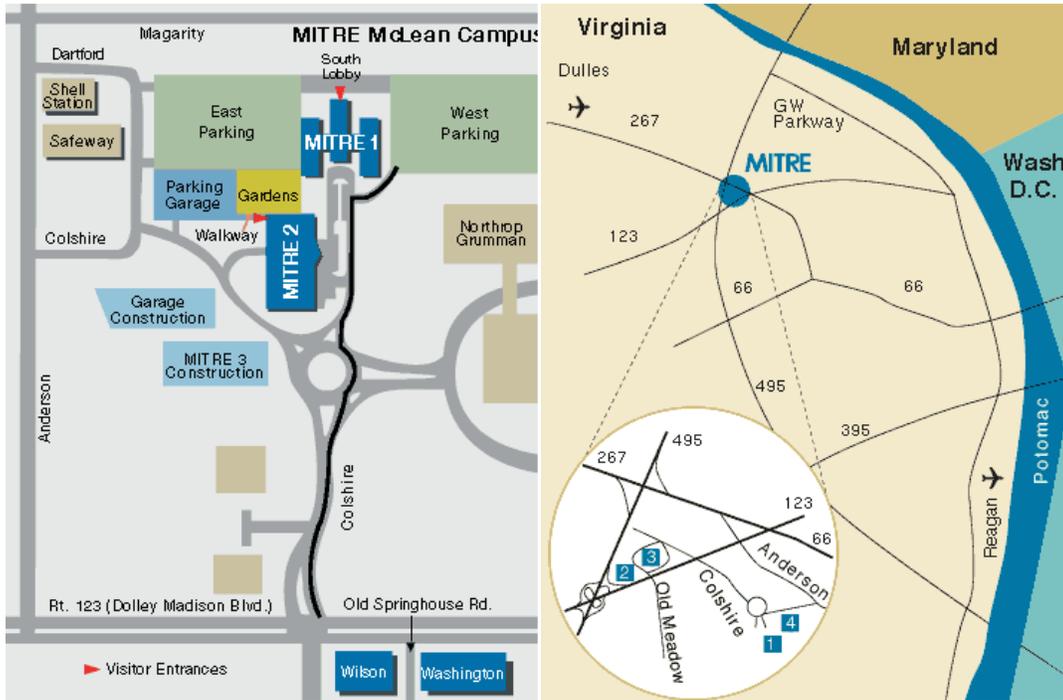
**Software SIG Chairman:** T. Scott Ankrum (ASQ & IEEE)  
**Software SIG Committee:** Richard Eng (ASQ); Chris Jones (ASQ & SSQ); Alfred Kromholz (SSQ); Tim Rice; James M. Simpson (ASQ); Tom Starai (IEEE), Tom Neff (IEEE), Lance Kelson (SSQ & IEEE), Aaron Dagen (IEEE)

**Sponsored Jointly By:** The American Society for Quality (ASQ), Washington DC & Maryland Metro Section (509), Software Special Interest Group (SSIG); IEEE Computer Society, Washington, DC & Northern Virginia Chapters; and Society for Software Quality (SSQ), Washington, DC Area Chapter. Members of the **ASQ SSIG** include software quality professionals, software engineers, and others interested in applying quality principles to the field of software development. See our web page: <http://www.asq509.org/ht/d/sp/i/2499/pid/2499>. We meet every month, usually at the MITRE facility in Tyson's Corner, Virginia, with VTC to other locations.

Next Month: 'Secure SDLC – SW Assurance' by Joe Jarzombek, on Tuesday, February 21, 2012

**Directions to the MITRE-2 Facility in McLean, Virginia:**

Take the Beltway, I-495 to Virginia. Take Exit 46B (McLean, Route 123). Take Route 123 North, (also called Dolley Madison Blvd.) and go to the second traffic light at Colshire Drive. Turn right on Colshire Drive and continue through circle on Colshire. **Park in front of or to the right of the buildings, before passing the gate. (The gates are closed at 7:00).** Additional directions can be found at: [http://www.mitre.org/about/locations/va\\_mclean\\_mitre2.html](http://www.mitre.org/about/locations/va_mclean_mitre2.html)



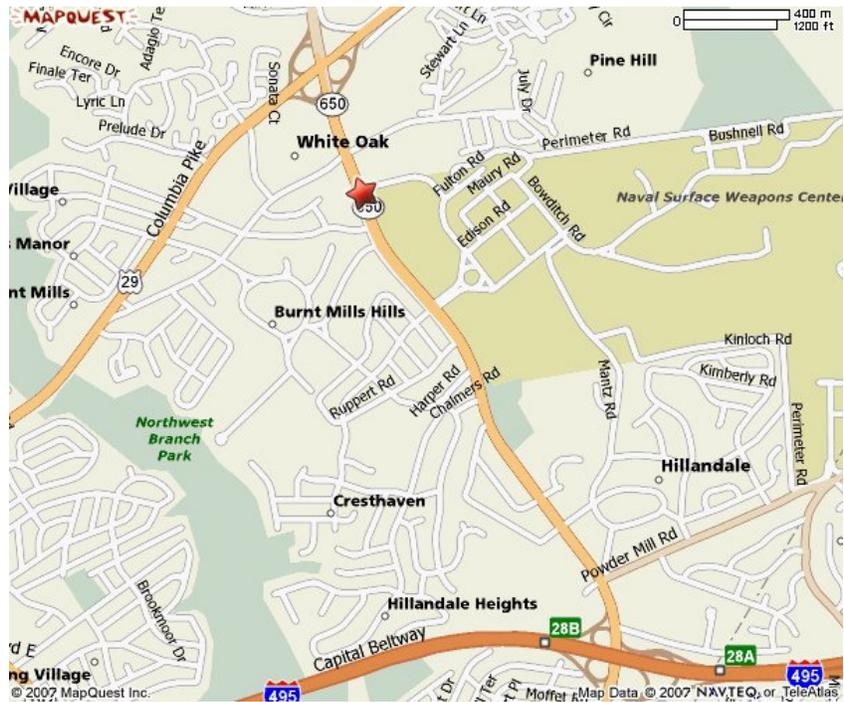
**Directions to the FDA facility in Silver Spring, MD:**

**From the Capital Beltway, I-495,** take New Hampshire Ave, Rt 650 north to Michelson Rd.

**From Columbia Pike, Rt 29,** take New Hampshire Ave, Rt 650 south to Michelson Rd

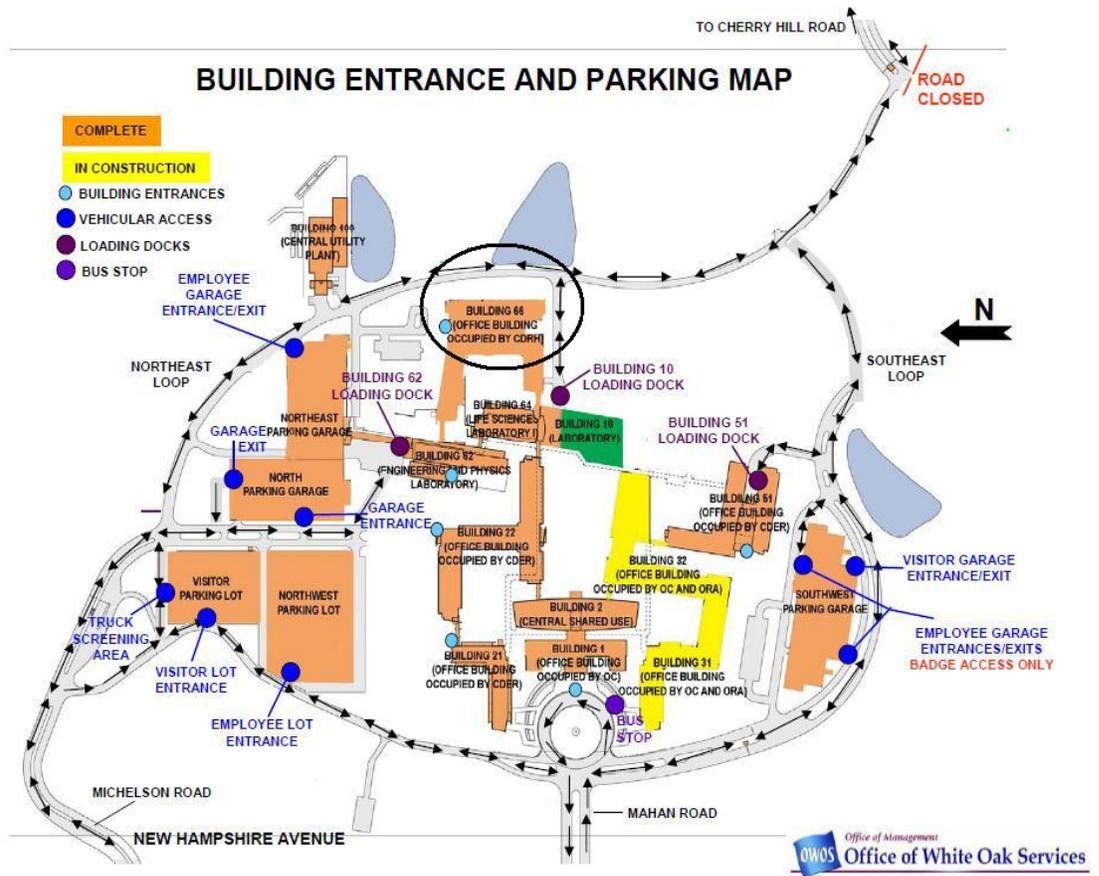
**From the Capital Beltway, I-495,** take New Hampshire Ave, Rt 650 north to Michelson Rd.

**From Columbia Pike, Rt 29,** take New Hampshire Ave, Rt 650 south to Michelson Rd



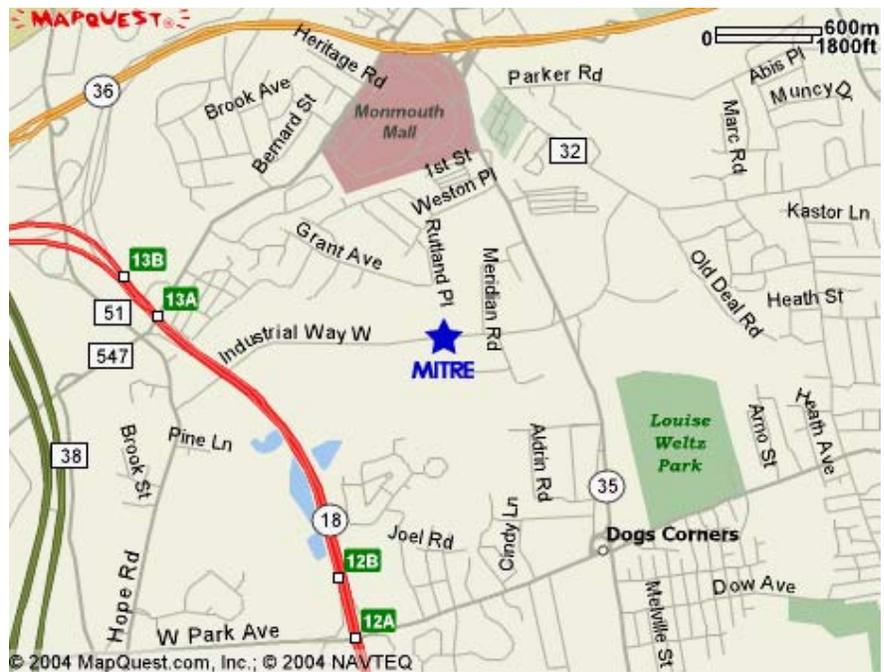
White Oak Building 66 (circled), room G512. This is a large conference room just beyond the Security desk on the right side of the grand atrium.

After 6 PM, visitors may use the surface parking closest to the building 66 entrance which are reserved for commuters with car pools. The north east parking garage is also an option.



## Directions to the MITRE Facility in Eatontown, New Jersey:

**From the New Jersey Turnpike** If traveling from the **SOUTH**, get off the Turnpike at Exit 7A (195 toward Shore Points). Take 195 East towards the ocean and shortly after 195 turns into Route 138 (approximately 35 miles) watch for Route 18 North (Eatontown). Take Route 18 North to Exit 13A (Wayside West/Wyckoff Road). At end of ramp, bear left. At first traffic light (Hope Road) make a left turn. Make second left turn onto Industrial Way. The MITRE Corporation is next to the First Atlantic Credit Union on the left hand side of Industrial Way. There is a MITRE sign out front.



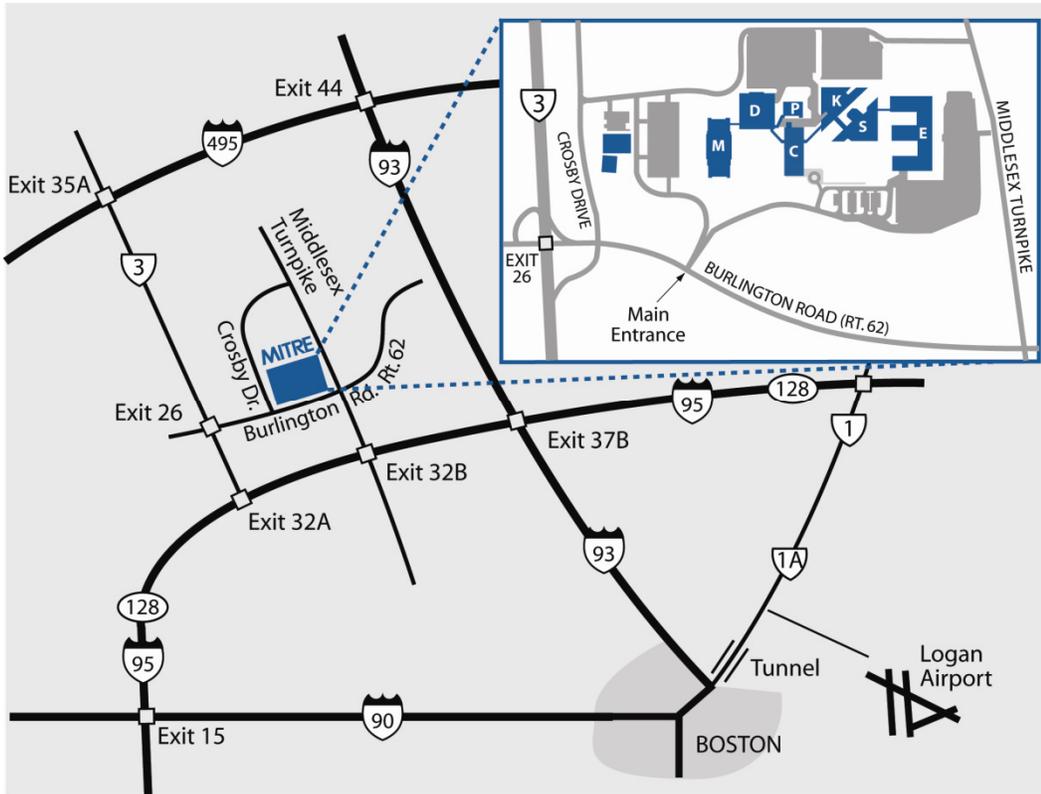
There are two buildings on the left side of Industrial Way. The first building is MITRE, the second building is TYCO. Take a left turn into the parking lot of the first building, which is MITRE.

If traveling from the **NORTH**, get off the Turnpike at Exit 11 (Garden State parkway exit). Follow the Garden State Parkway until Exit 105 (Eatontown). Follow the directions from the Garden State Parkway below.

**From the Garden State Parkway:** Follow the Garden State Parkway from the North or from the South to Exit 105 (the exit numbers increase going from south to north). After paying the toll at Exit 105, make first right turn onto Hope Road. Follow Hope Road to Industrial Way (second left turn after crossing Wyckoff/Shafto Road). Make second left turn onto Industrial Way. The MITRE Corporation is next to the First Atlantic Credit Union on the left hand side of Industrial Way. There is a MITRE sign out front. There are two white buildings on the left side of Industrial Way. The first building is MITRE, the second building is TYCO. Take a left turn into the parking lot of the first building, which is MITRE.

**When you are facing the front of the MITRE building, the entrance to the ASQ meeting will be from the left parking lot opposite First Atlantic bank. Follow the signs to the conference room.**

## Directions to the MITRE complex in Bedford, Massachusetts:



### From New Hampshire to MITRE Complex:

Take Route 3 South to Exit 26 (Route 62)

Turn Left on Route 62. It is a short distance to the MITRE entrance on left--watch for building directories on MITRE grounds

### From Manchester Airport (NH) to MITRE Complex:

Start out going West on Airport Rd. toward parking

Turn slight right onto Brown Ave/NH-3A

Merge onto NH-101 West via the ramp on the left toward Bedford/Nashua

Take the Everett Turnpike South exit toward Merrimack/Nashua

Merge onto Everett Turnpike (toll road)

Everett Turnpike becomes US-3 South

Take exit 26 (Route 62) toward Bedford/Burlington

Turn Left onto (Route 62) Burlington Road

At second set of traffic lights, turn left at MITRE Bedford Campus entrance--watch for building directories on MITRE grounds.

### Logan Airport to MITRE Complex:

Exit airport towards Route 1A South/Summer Tunnel (I-93) Boston.

Proceed through tunnel towards Storrow Drive (I-93 North).

Take I-93 North eleven miles to Exit 37B (Route 128 (I-95) South).

Take Route 128 (I-95) South six miles to Exit 32A.

Take Exit 32A, and proceed on Route 3 North.

Take Route 3 North for two miles to Exit 26 (Route 62).

Bear right on Route 62.

Take Route 62 a short distance to the MITRE Bedford Campus entrance on left--watch for building directories on MITRE grounds.