IBM “Watson”: Beyond playing “Jeopardy!”

Katharine Frase, VP Industries Research, IBM

with thanks to:
David Ferrucci, Principal Investigator, DeepQA Team @ IBM Research

April 24, 2012
Want to Play Chess or Just Chat?

• Chess
  – A finite, mathematically well-defined search space
  – Limited number of moves and states
  – All the symbols are completely grounded in the mathematical rules of the game

• Human Language
  – Words by themselves have no meaning
  – Only grounded in human cognition
  – Words navigate, align and communicate an infinite space of intended meaning
  – Computers can not ground words to human experiences to derive meaning
Informed Decision Making: Search vs. Expert Q&A

Decision Maker
- Has Question
- Distills to 2-3 Keywords
- Reads Documents, Finds Answers
- Finds & Analyzes Evidence

Search Engine
- Finds Documents containing Keywords
- Delivers Documents based on Popularity

Expert
- Understands Question
- Produces Possible Answers & Evidence
- Analyzes Evidence, Computes Confidence
- Delivers Response, Evidence & Confidence

© 2011 IBM Corporation
Easy Questions?

$$\ln((12,546,798 \times \pi) ^ 2 / 34,567.46) = 0.00885$$

Select Payment where Owner=“David Jones” and Type(Product)=“Laptop”,

<table>
<thead>
<tr>
<th>Owner</th>
<th>Serial Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Jones</td>
<td>45322190-AK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Invoice #</th>
<th>Vendor</th>
<th>Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV10895</td>
<td>MyBuy</td>
<td>$104.56</td>
</tr>
</tbody>
</table>

David Jones ≠ Dave Jones
Hard Questions?

Computer programs are natively explicit, fast and exacting in their calculation over numbers and symbols….But Natural Language is implicit, highly contextual, ambiguous and often imprecise.

• Where was X born?

  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.

• X ran this?

  If leadership is an art then surely Jack Welch has proved himself a master painter during his tenure at GE.
The Jeopardy! Challenge: A compelling and notable way to drive and measure the technology of automatic Question Answering along 5 Key Dimensions

- **Broad/Open Domain**
- **Complex Language**
- **High Precision**
- **Accurate Confidence**
- **High Speed**

**$200**
If you’re standing, it’s the direction you should look to check out the wainscoting.

**$1000**
The first person mentioned by name in ‘The Man in the Iron Mask’ is this hero of a previous book by the same author.

**$600**
In cell division, mitosis splits the nucleus & cytokinesis splits this liquid cushioning the nucleus.

**$2000**
Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that’s farthest north.
The Best Human Performance: Our Analysis Reveals the Winner’s Cloud

Each dot represents an actual historical human Jeopardy! game

Top human players are remarkably good.

Computers?

Winning Human Performance

Grand Champion Human Performance

2007 QA Computer System

More Confident

Less Confident

% Answered

Precision
What is Watson?

Decomposition

Hypothesis Generation

Hypothesis & Evidence Scoring

Synthesis

Final Confidence Merging & Ranking

Text Search

- Frame Induction
- Concept Matching
- Deep Parsing
- Entity Extraction

Structured Search

- Sequence Matching
- Relation Extraction
- Graph Matching
- Topic Detection
- Sentence Decomposition

Statistical and Rule-based Analysis

Statistical Learning Framework
(Statistically combines disparate scores from many analytics based on Machine Learning)

UIMA for Text Analytics Standard Interoperability

UIMA-AS (Standard Messaging) for Scale-out and Speed

Linux

Power7

© 2011 IBM Corporation
DeepQA: The Technology Behind Watson
Massively Parallel Probabilistic Evidence-Based Architecture

Generates and scores many hypotheses using a combination of 1000’s Natural Language Processing, Information Retrieval, Machine Learning and Reasoning Algorithms. These gather, evaluate, weigh and balance different types of evidence to deliver the answer with the best support it can find.
Automatic Learning From “Reading”

Sentence Parsing

Volumes of Text

Syntactic Frames

Semantic Frames

Generalization & Statistical Aggregation

Vessels Sink (0.7)
People sink 8-balls (0.5) (in pool/0.8)

Inventors patent inventions (0.8)
Officials Submit Resignations (0.7)
People earn degrees at schools (0.9)

Fluid is a liquid (.6)
Liquid is a fluid (.5)

© 2011 IBM Corporation
QUESTION

Analyze Question

Generate Hypotheses

Collect and Evaluate Evidence

Weigh and Combine for Final Confidences

ANSWER

© 2011 IBM Corporation
EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
Each year the EU selects capitals of culture; one of the 2010 cities was this Turkish "meeting place of cultures".
Each year the EU selects capitals of culture; one of the 2010 cities was Turkish "meeting place of cultures".

Collect and evaluate evidence.

Weigh and combine for final confidences.

Answer.
Each year the EU selects capitals of culture; one of the 2010 cities was this Turkish "meeting place of cultures".

Weigh and combine for final confidences.

© 2011 IBM Corporation
EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
Each year the EU selects capitals of culture; one of the 2010 cities was this Turkish "meeting place of cultures."
Some Basic Jeopardy! Clues

- This **fish** was thought to be extinct millions of years ago until one was found off South Africa in 1938
  - Category: ENDS IN "TH"
  - Answer: **coelacanth**

- When hit by electrons, a phosphor gives off electromagnetic energy in this **form**
  - Category: General Science
  - Answer: **light (or photons)**

- Secy. Chase just submitted **this** to me for the third time--guess what, pal. This time I'm accepting **it**
  - Category: Lincoln Blogs
  - Answer: **his resignation**
Categories are not as simple as they seem.

Watson uses statistical machine learning to discover that Jeopardy! categories are only weak indicators of the answer type.

<table>
<thead>
<tr>
<th>U.S. CITIES</th>
<th>Country Clubs</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. Petersburg is home to Florida’s annual tournament in this game popular on ship decks (Shuffleboard)</td>
<td>From India, the shashpar was a multi-bladed version of this spiked club (a mace)</td>
<td>Archibald MacLeish based his verse play &quot;J.B.&quot; on this book of the Bible (Job)</td>
</tr>
<tr>
<td>Rochester, New York grew because of its location on this (the Erie Canal)</td>
<td>A French riot policeman may wield this, simply the French word for &quot;stick&quot; (a baton)</td>
<td>In 1928 Elie Wiesel was born in Sighet, a Transylvanian village in this country (Romania)</td>
</tr>
</tbody>
</table>
Evaluating Possibilities and Their Evidence

In cell division, mitosis splits the nucleus & cytokinesis splits this **liquid cushioning** the nucleus.

- Organelle
- Vacuole
- Cytoplasm
- Plasma
- Mitochondria
- Blood …

Many candidate answers (CAs) are generated from many different searches.
Each possibility is evaluated according to different dimensions of evidence.

**Just One** piece of evidence is if the CA is of the right type. In this case a “liquid”.

Is(“Cytoplasm”, “liquid”) = 0.2
Is(“organelle”, “liquid”) = 0.1
Is(“vacuole”, “liquid”) = 0.2
Is(“plasma”, “liquid”) = 0.7

“Cytoplasm is a fluid surrounding the nucleus…”

Wordnet $\rightarrow$ Is_a(Fluid, Liquid) $\rightarrow$ ?

Learned $\rightarrow$ Is_a(Fluid, Liquid) $\rightarrow$ yes.
Divide and Conquer
*(Typical in Final Jeopardy!)*

Lyndon B Johnson

In 1968 this man was U.S. president.

When "60 Minutes" premiered this man was U.S. president.

Must identify and solve sub-questions from different sources to answer the top level question

The DeepQA architecture attempts different *decompositions* and recursively applies the QA algorithms
Some Growing Pains (early answers)

THE AMERICAN DREAM
Decades before Lincoln, Daniel Webster spoke of government "made for", "made by" & "answerable to" them

NEW YORK TIMES HEADLINES
An exclamation point was warranted for the "end of" this! In 1918

MILESTONES
In 1994, 25 years after this event, 1 participant said, "For one crowning moment, we were creatures of the cosmic ocean"

THE QUEEN'S ENGLISH
Give a Brit a tinkle when you get into town & you've done this

the People
No One
WW I
a sentence
Apollo 11 moon landing
the Big Bang
Call on the phone
urinate

IBM Watson
Playing in the Winners Cloud

Baseline 12/06

© 2011 IBM Corporation
One Jeopardy! question can take **2 hours on a single 2.6Ghz Core**

Optimized & Scaled out on 2,880-Core Power750 using UIMA-AS,
*Watson* is answering in ~3 seconds

- 90 x IBM Power 750\(^1\) servers
- 2880 POWER7 cores
- POWER7 3.55 GHz chip
- 500 GB per sec on-chip bandwidth
- 10 Gb Ethernet network
- 15 Terabytes of memory
- 20 Terabytes of disk, clustered
- Can operate at 80 Teraflops
- Runs IBM DeepQA software
- Scales out with and searches vast amounts of unstructured information with UIMA & Hadoop open source components
- SUSE Linux provides a cost-effective open platform which is performance-optimized to exploit POWER 7 systems
- 10 racks include servers, networking, shared disk system, cluster controllers

\(^1\) See page 117 for more information.
Watson brings many innovations to apply to business issues

**Watson Capabilities**

1. **Natural language understanding**
2. **Broad domain of unstructured data**
3. **Parallel hypothesis generation and confidence scoring**
4. **Iterative Question/Answering to refine results**
5. **Machine learning**

**Best Fit for Watson**

- **Analysis of unstructured data**
- **Critical questions that require decision support with prioritized recommendations and evidence**
- **High value in decision support**
- **Leverage scale to maximize machine learning and improve outcomes over time**

**Watson Solution Area and Industry Application**

### Assisting Research (e.g., Financial Services)

**Assisting research** of macroeconomic trends, current events and historic data to support investment decisions

**Results:** increased alpha and risk management due to improved investment decision making and customer insights

**Other Applications:** Customer Insights, Knowledge Mgmt

### Assisting Discovery (e.g., Legal Services)

**Assisting research** of historical and current documents to discover relevant references and identify relationships.

**Results:** increased confidence in due diligence, prior art or precedent evaluations, due to context understanding of full passages and learned value of sources, vs “keyword search” based methods.

**Other Applications:** R&D Mgmt, Compliance Mgmt

### Assisting Diagnosis (e.g., Healthcare)

**Assisting diagnosis and suggestions** for patient condition based on patient data, family history, medical journals etc.

**Results:** improved patient care and payer/provider cost management based on more efficient patient outcomes

**Other Applications:** Contact Center, Technical Support

© 2011 IBM Corporation
THANK YOU