

# IBM Research ~~Beyond Jeopardy!~~

---

## The Implications of IBM's Watson System

Dr. David F. McQueeney  
VP, Software @ IBM Research  
davidmcq@us.ibm.com

Fall Creek Falls Conference, September 2011



## Real Language is Real Hard

### ■ Chess

- Finite, mathematically well-defined search space
- Limited number of moves and states
- Grounded in **explicit, unambiguous** mathematical rules



### ■ Human Language

- Ambiguous, contextual and implicit
- Grounded only in **human cognition**
- Seemingly infinite number of ways to express the same meaning



## What Computers Find Hard

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.

Person	Birth Place
A. Einstein	ULM

Structured

Unstructured

### ■ Where was X born?

*One day, from among his city views of Ulm, Otto chose a watercolor to send to Albert Einstein as a remembrance of Einstein's birthplace.*

Person	Organization
J. Welch	GE

### ■ X ran this?

*If leadership is an art then surely Jack Welch has proved himself a master painter during his tenure at GE.*

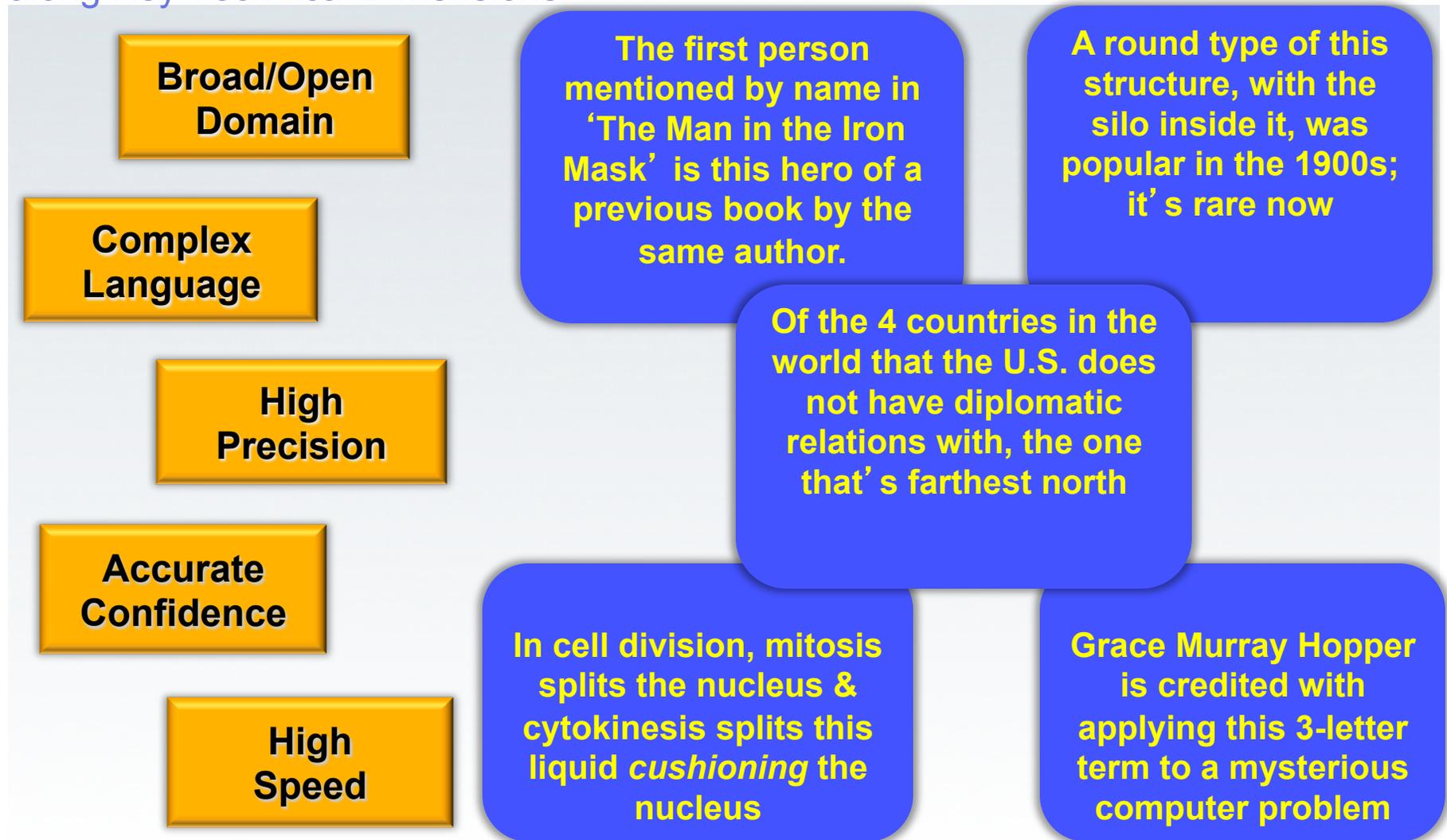
## Automatic Open-Domain Question Answering

*A Long-Standing Grand Challenge in Artificial Intelligence to emulate human expertise*

- Given
  - Rich **Natural Language Questions**
  - Over a **Broad Domain of Knowledge**
  
- Deliver
  - **Precise Answers:** Determine what is being asked & give precise response
  - **Accurate Confidences:** Determine likelihood answer is correct
  - **Consumable Justifications:** Explain why the answer is right
  - **Fast Response Time:** Precision & Confidence in <3 seconds

# The Jeopardy! Challenge

A palpable, compelling and notable way to drive the technology of Question Answering along Key Technical Dimensions



## Some Basic *Jeopardy!* Clues

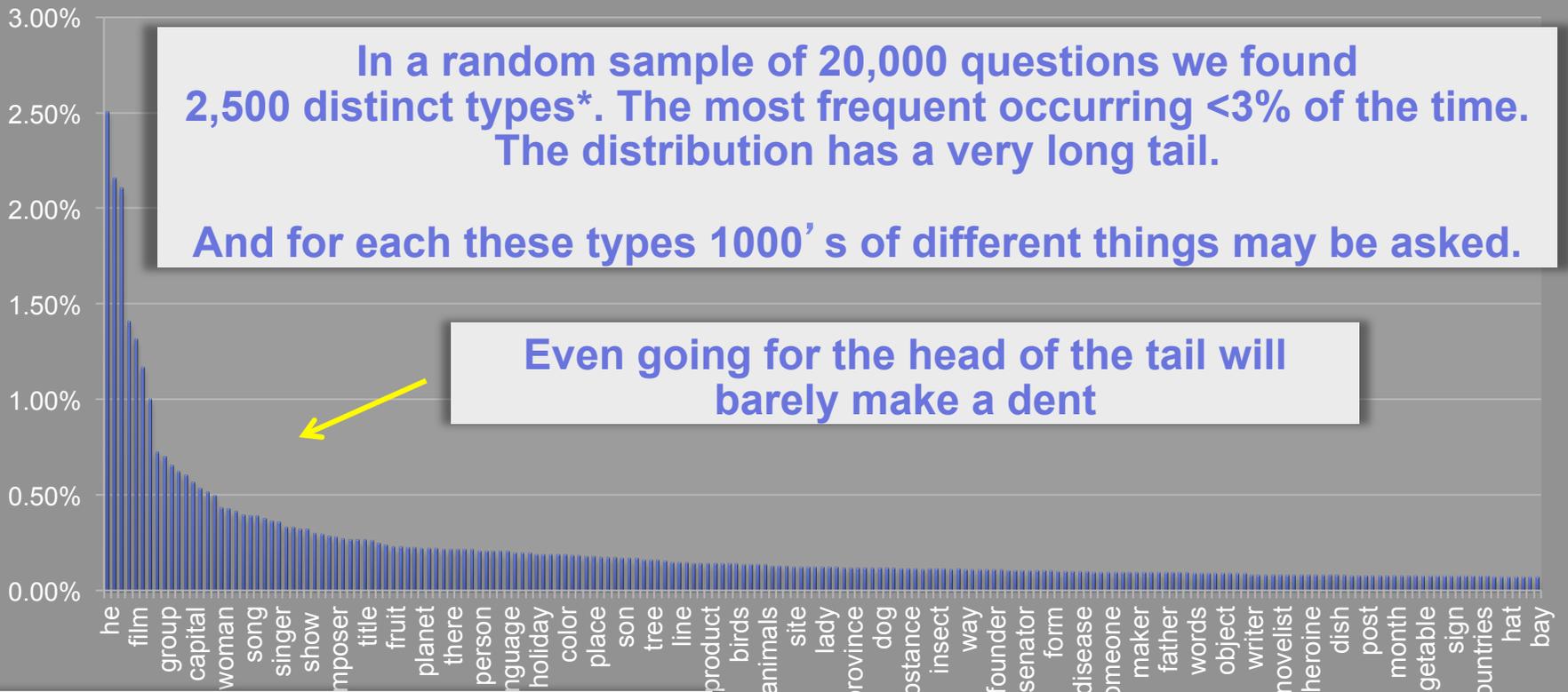
- This **fish** was thought to be extinct millions of years ago until one was found off South Africa in 1938
  - Category: ENDS IN "TH"
  - Answer: **coelacanth**
  
- When hit by electrons, a phosphor gives off electromagnetic energy in this **form**
  - Category: General Science
  - Answer: **light (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**

The *type* of thing being asked for is often indicated but can go from specific to very vague

# Broad Domain

**We do NOT attempt to anticipate all questions and build databases.**

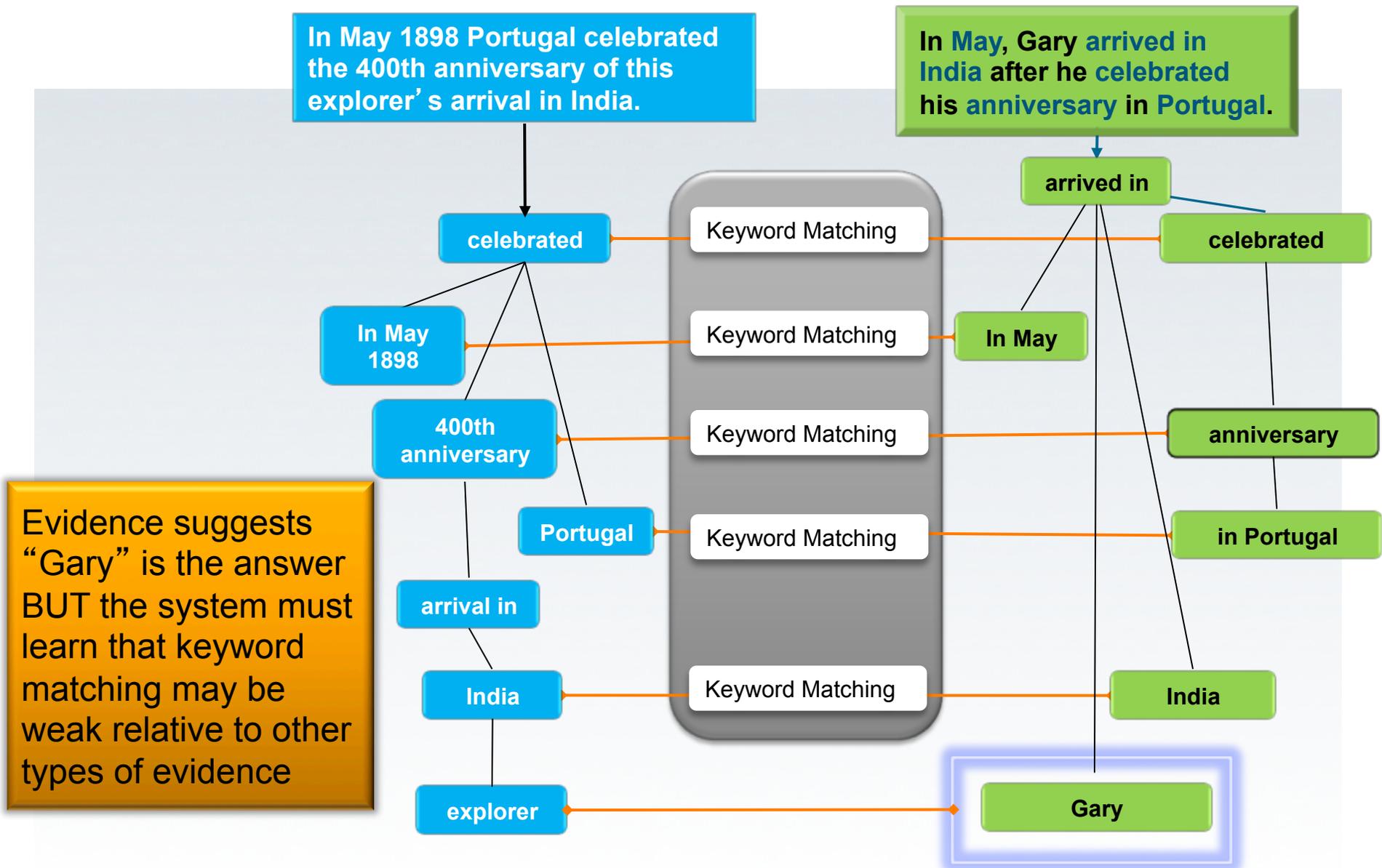
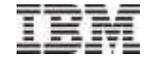
**We do NOT try to build a formal model of the world**



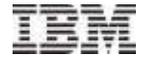
\*13% are non-distinct (e.g, it, this, these or NA)

**Our Focus is on reusable NLP technology for analyzing vast volumes of *as-is* text. Structured sources (DBs and KBs) provide background knowledge for interpreting the text.**

# Different Types of Evidence: Keyword Evidence

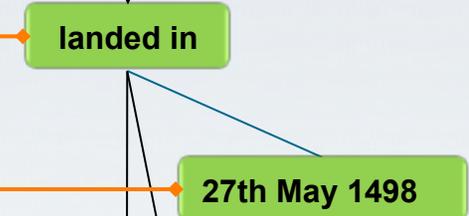
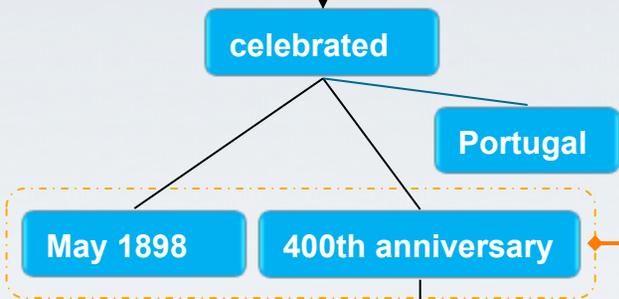
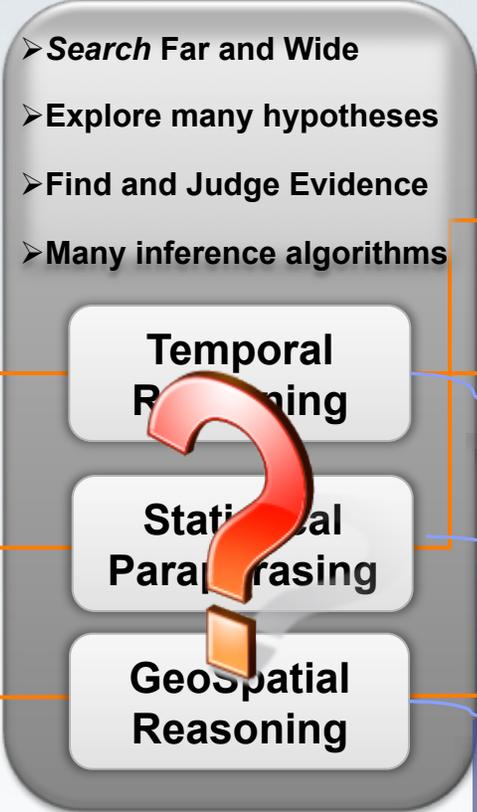


# Different Types of Evidence: Deeper Evidence

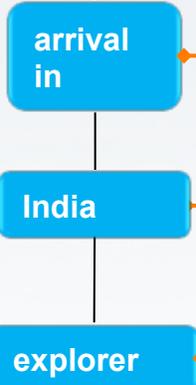


In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India.

On the 27<sup>th</sup> of May 1498, Vasco da Gama landed in Kappad Beach



Stronger evidence can be much harder to find and score.

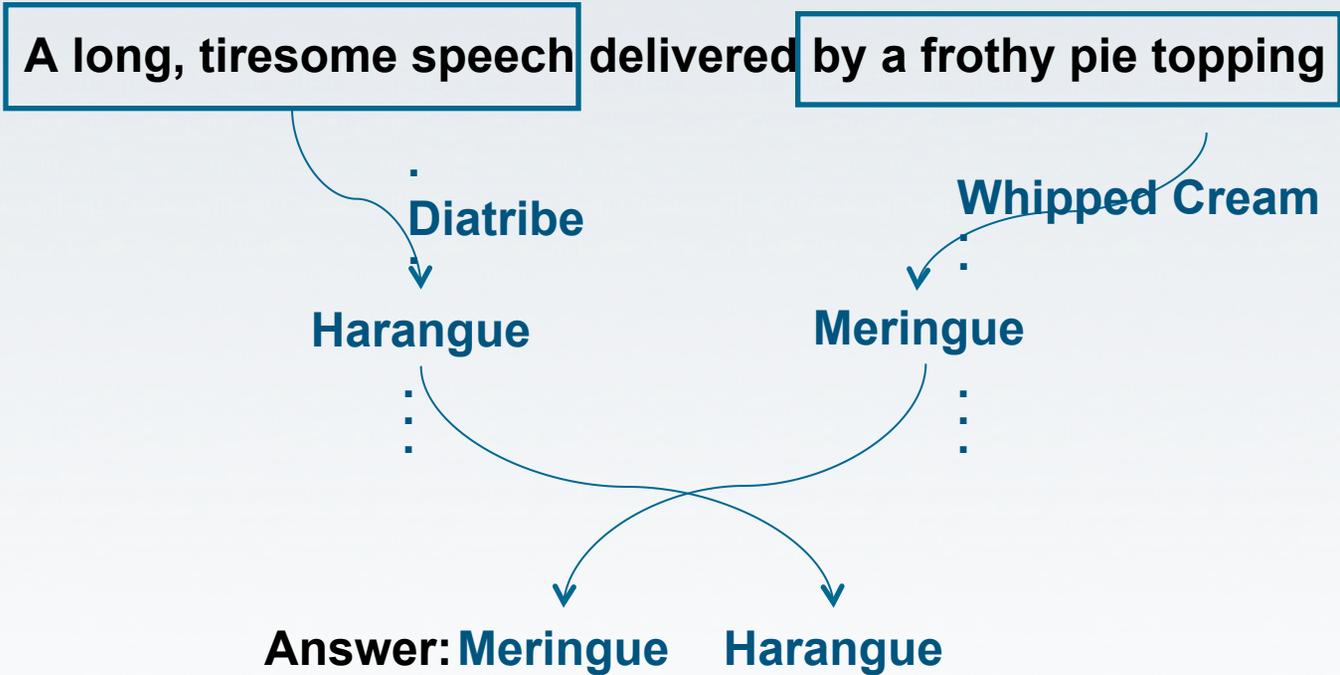


The evidence is still not 100% certain.

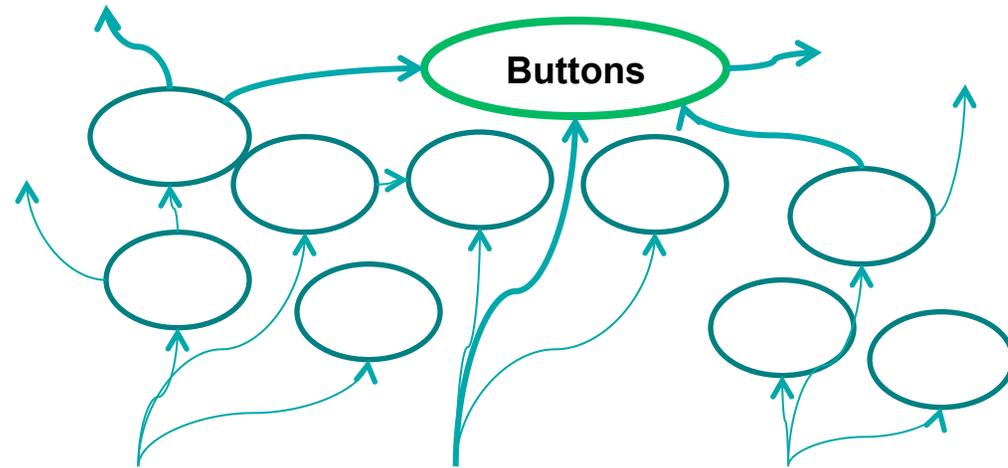
# Not *Just* for Fun

Some Questions require Decomposition and Synthesis

Category: Edible Rhyme Time

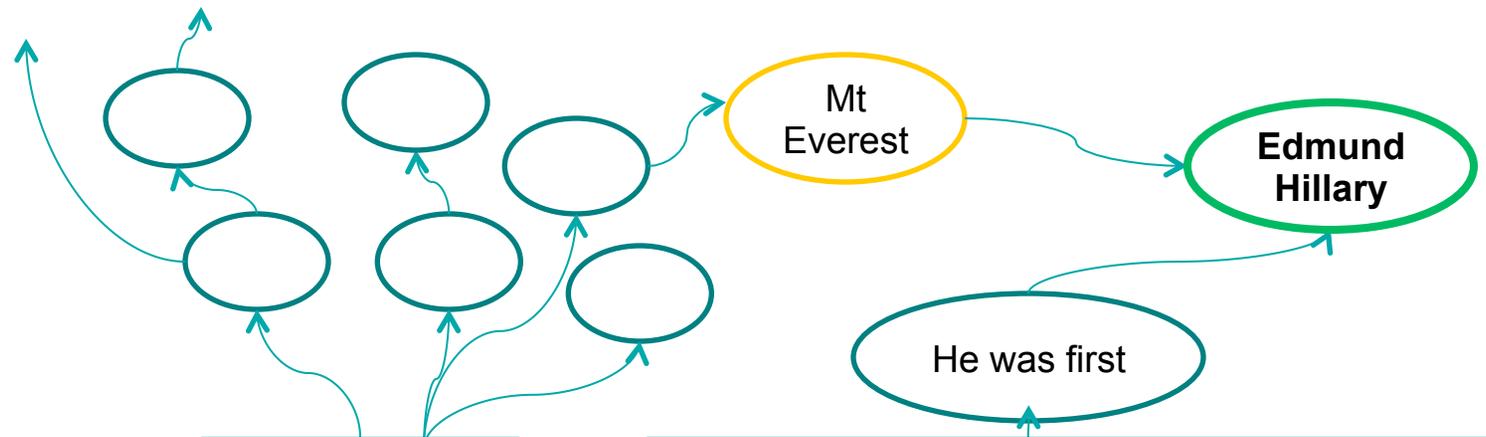


# Missing Links



**Category:** Common Bonds

Shirts, TV remote controls, Telephones

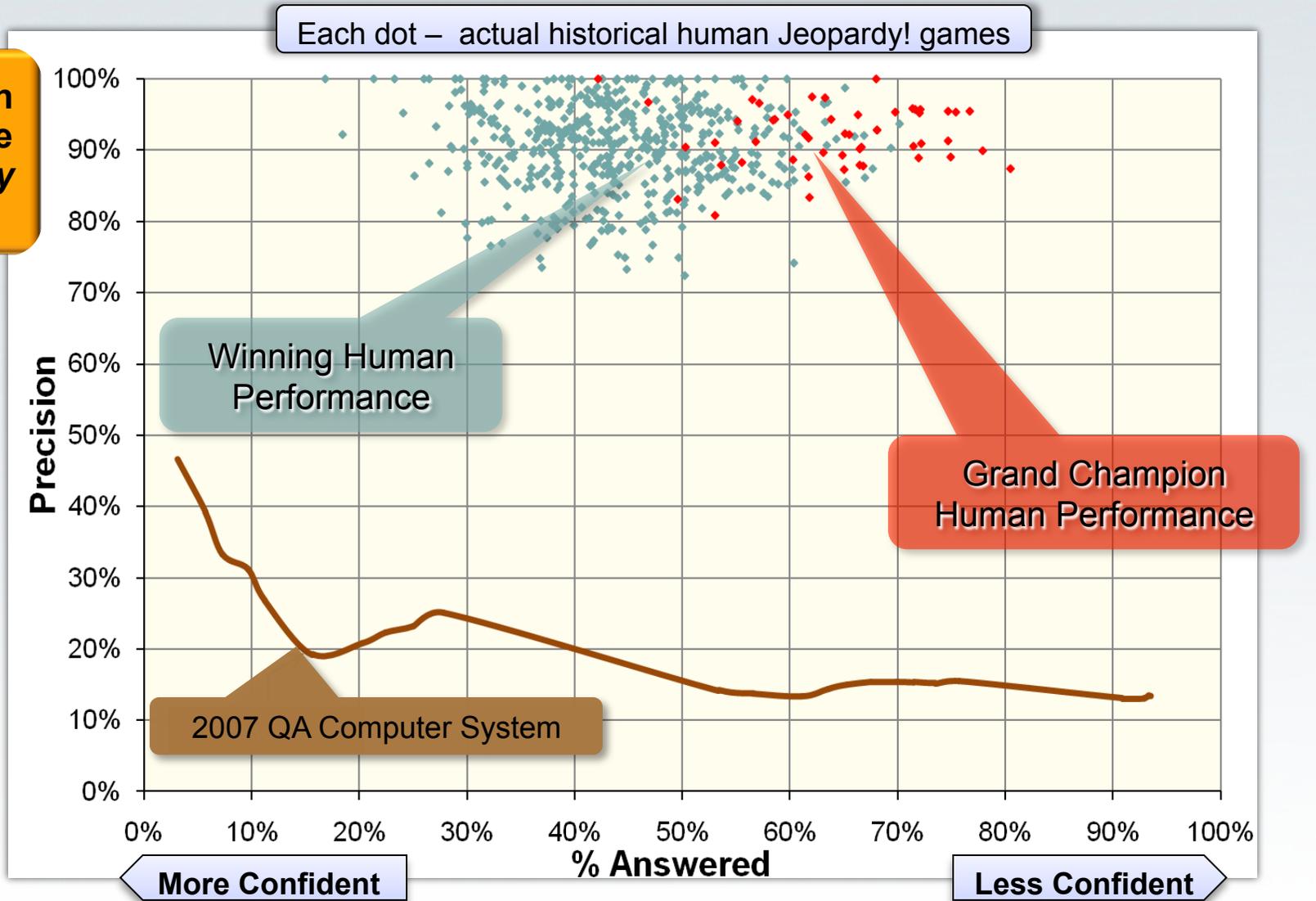


On hearing of the discovery of **George Mallory's body**, **he told reporters he still thinks he was first.**

# What It Takes to compete against Top Human Jeopardy! Players

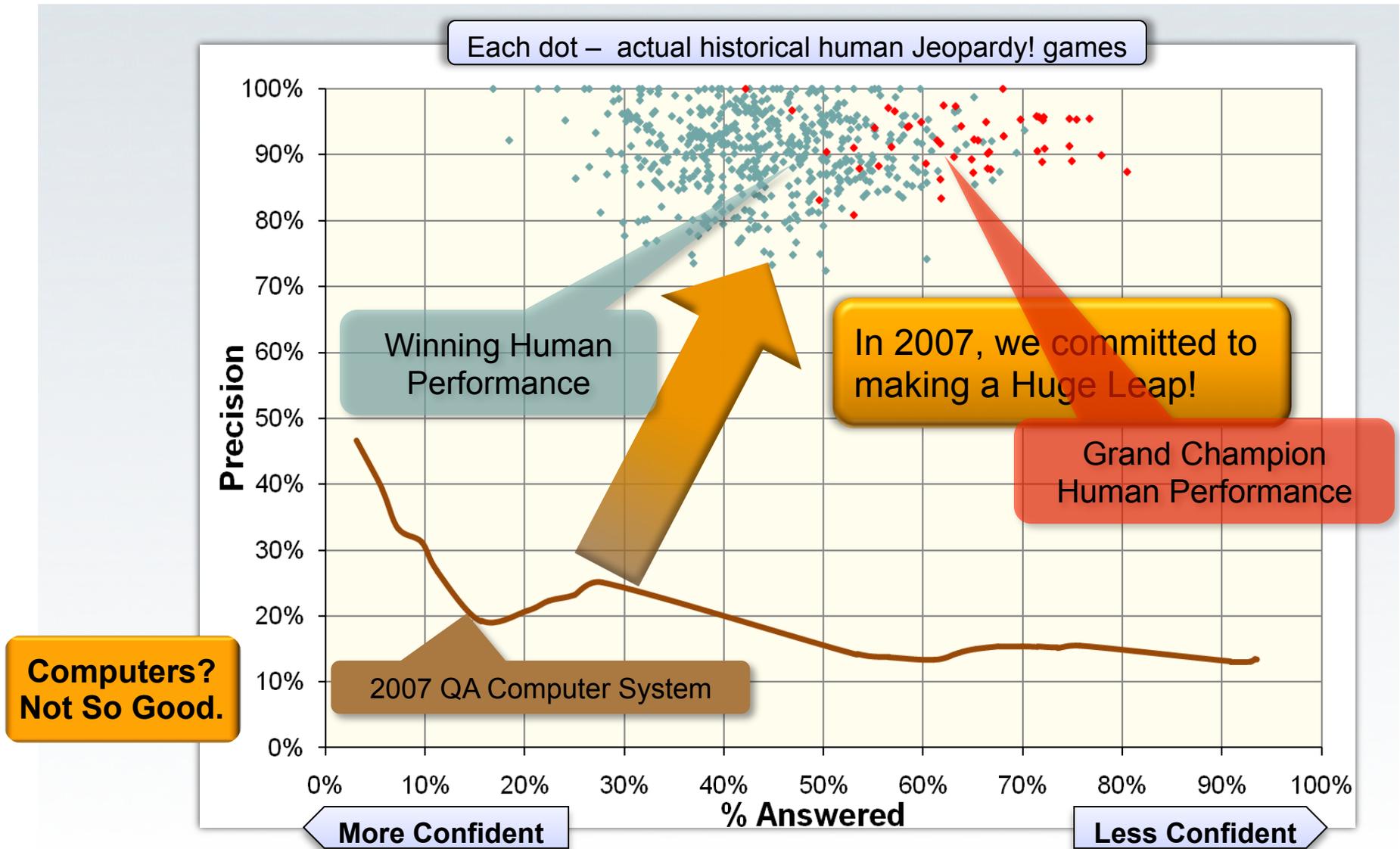
*Our Analysis Reveals the Winner's Cloud*

**Top human players are remarkably good.**



# What It Takes to compete against Top Human Jeopardy! Players

*Our Analysis Reveals the Winner's Cloud*

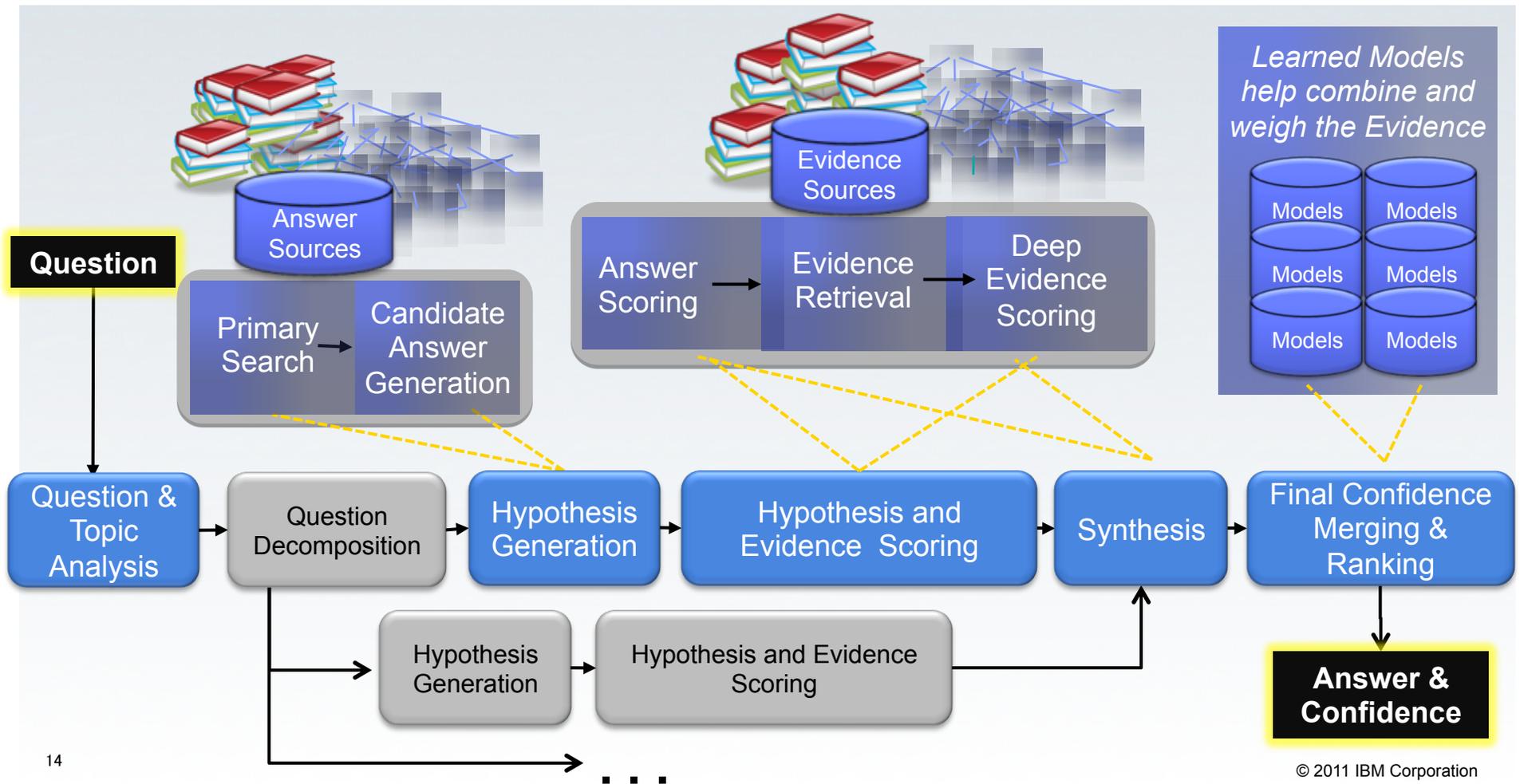


# DeepQA: The Technology Behind Watson



## Massively Parallel Probabilistic Evidence-Based Architecture

DeepQA generates and scores many hypotheses using an extensible collection of **Natural Language Processing, Machine Learning and Reasoning Algorithms**. These gather and weigh evidence over both unstructured and structured content to determine the answer with the best confidence.



# Example Question

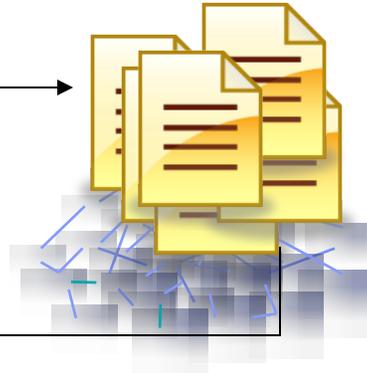
IN 1698, THIS COMET DISCOVERER TOOK A SHIP CALLED THE PARAMOUR PINK ON THE FIRST PURELY SCIENTIFIC SEA VOYAGE

Question Analysis

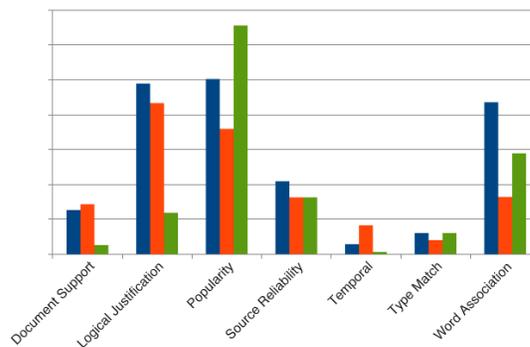
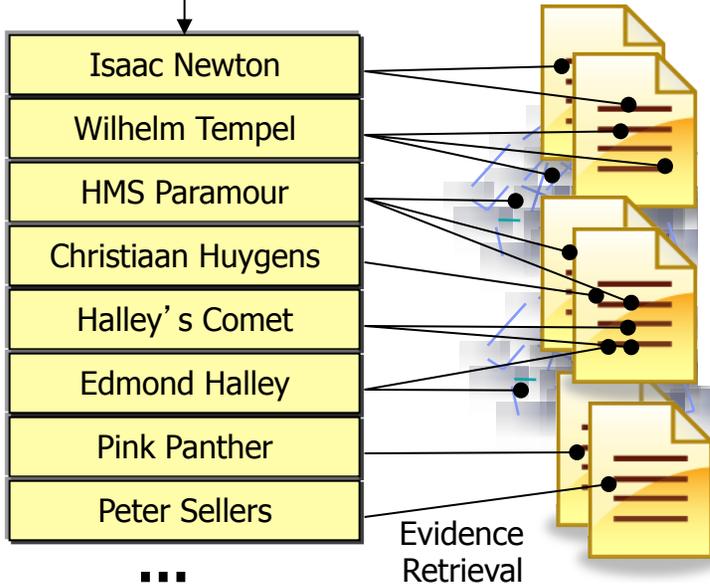
Keywords: 1698, comet, paramour, pink, ...  
 AnswerType(comet discoverer)  
 Date(1698)  
 Took(discoverer, ship)  
 Called(ship, Paramour Pink)  
 ...

Primary Search

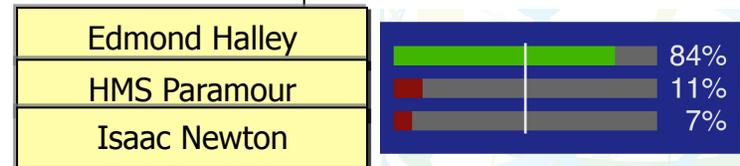
Related Content (Structured & Unstructured)



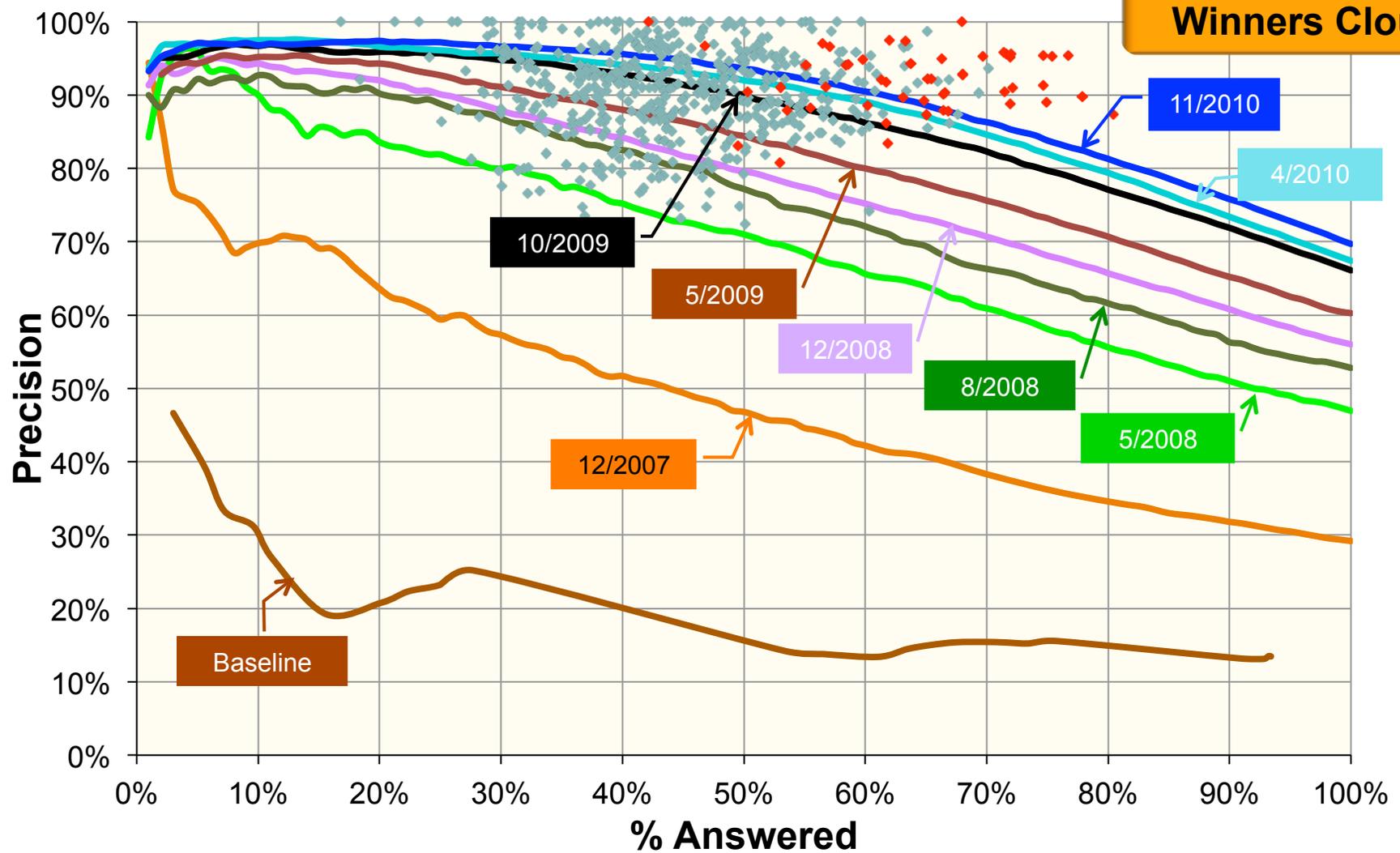
Candidate Answer Generation



Merging & Ranking



# Deep QA: Incremental Progress in Precision and Confidence 6/2007-11/2011



## The real Watson – a Workload Optimized System

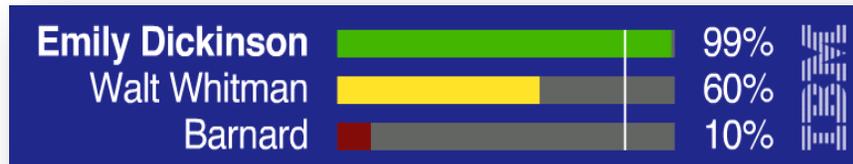
- 90 x IBM Power 750<sup>1</sup> servers
- 2880 POWER7 cores
- POWER7 3.55 GHz chip
- 500 GB per sec on-chip bandwidth
- 10 Gb Ethernet network
- 15 Terabytes of memory
- 20 Terabytes of disk, clustered
- Can operate at 80 Teraflops
- Runs IBM DeepQA software
- Scales out with and searches vast amounts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



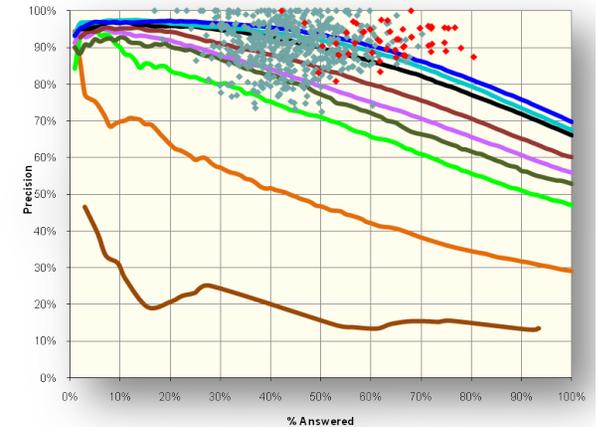
<sup>1</sup> Note that the Power 750 featuring POWER7 is a commercially available server that runs AIX, IBM i and Linux and has been in market since Feb 2011

## Watson: Precision, Confidence & Speed

- **Deep Analytics** – We achieved champion-levels of *Precision* and *Confidence* over a huge variety of expression



- **Speed** – By optimizing Watson's computation for Jeopardy! on **2,880 POWER7** processing cores we went **from 2 hours per question on a single CPU to an average of just 3 seconds** – fast enough to compete with the best.
- **Results** – in 55 real-time sparring against former **Tournament of Champion Players last year**, Watson put on a very competitive performance, winning 71%. In the final Exhibition Match against Ken Jennings and Brad Rutter, Watson won!



Ready to have a little fun with Watson?  
Let's try a few questions and see how we do...



# Potential Business Applications



**Healthcare / Life Sciences:** Diagnostic Assistance, Evidenced-Based, Collaborative Medicine

**Tech Support:** Help-desk, Contact Centers



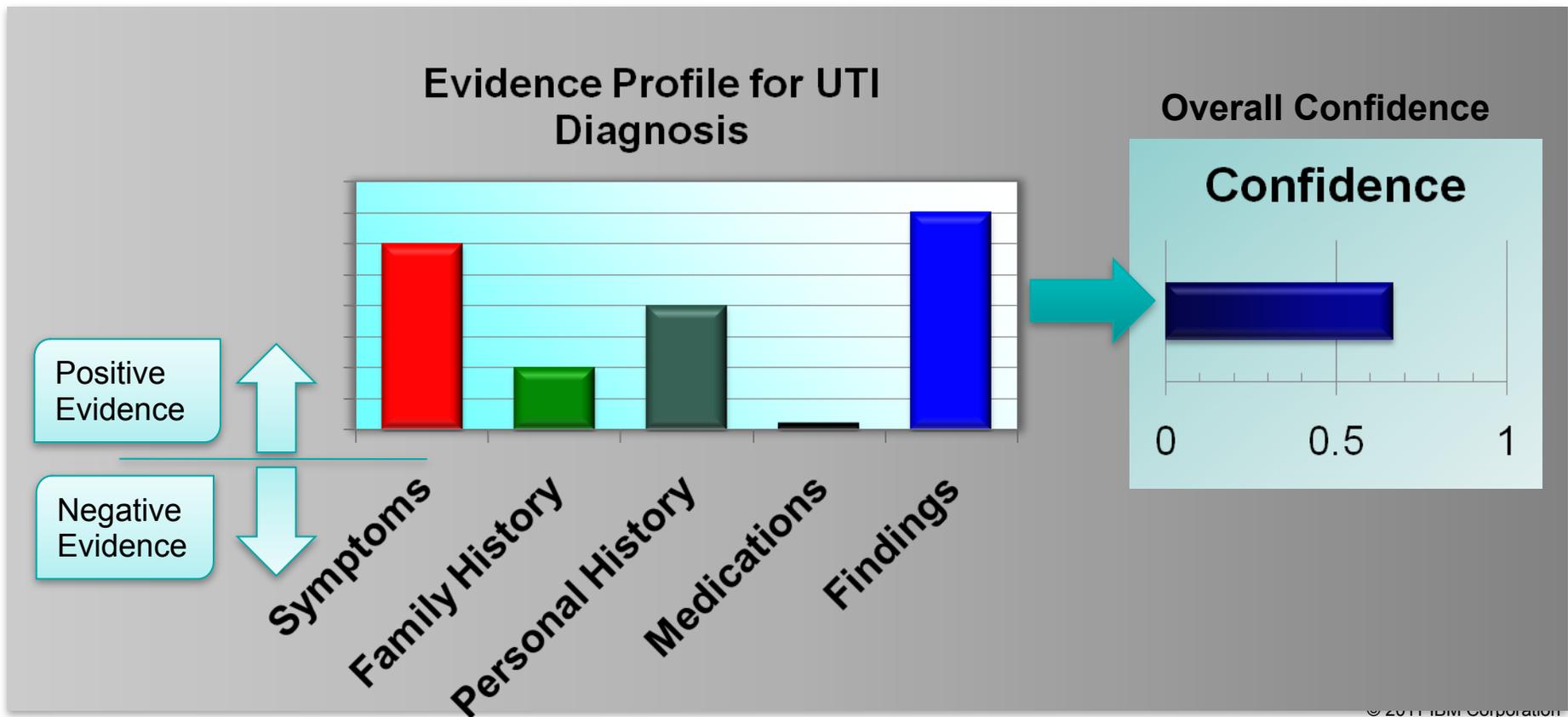
**Enterprise Knowledge Management and Business Intelligence**

**Government:** Improved Information Sharing and Security



## *Evidence Profiles from disparate data is a powerful idea*

- Each dimension contributes to supporting or refuting hypotheses based on
  - **Strength of evidence**
  - **Importance of dimension for diagnosis** (learned from training data)
- Evidence dimensions are combined to produce an overall confidence
- Evidence profiles explain why Watson gives the answers it does



Thank you  
for your attention