

Ask My Robot:

How Computers Answer Questions



Michael I. Shamos, Ph.D., J.D.
Language Technologies Institute
School of Computer Science
Carnegie Mellon University



UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Background

- Ph.D., Yale University (computer science, 1978)
- J.D., Duquesne University (law, 1981)
- Carnegie Mellon computer science faculty (1975 -)
 - Institute for Software Research
 - Language Technologies Institute
- Director, Master's Program in eBusiness Technology
- Visiting Professor, University of Hong Kong (2001-)

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Can Machines Answer Questions?

- Can tell humans from machines by asking questions that require human knowledge or “intelligence”
- Example: **CAPTCHA** (Completely Automated Public Turing Test To Tell Computers and Humans Apart)



- Invented by Carnegie Mellon professor Luis Von Ahn
- Sold to Google in 2009

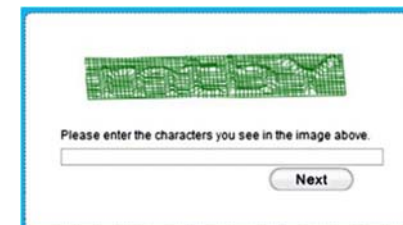
UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Can Machines Answer Questions?

- Problem: these are easy to defeat OR are too hard for humans



UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS



- TicketMaster v. RMG Technologies
- 2007 lawsuit involving captchas
- TicketMaster website sells tickets to entertainment events (sports, theatre, concerts)
- TicketMaster's site contains a "Terms of Use" link:
- "When purchasing tickets on Ticketmaster.com, you are limited to a specified number of tickets for each event (also known as a 'ticket limit'). This amount is included on the unique event page and is verified with every transaction. This policy is in effect to discourage unfair ticket buying practices."

TicketMaster v. RMG Technologies


- "You agree that you will not use any robot, spider or other automatic device, process or means to access the Site ... You agree that you will not take any action that imposes an unreasonable or disproportionately large load on our infrastructure. You agree that you will not access, reload or "refresh" transactional event or ticketing pages, or make any other request to transactional servers, more than once during any three second interval."
- TicketMaster uses Captchas to ensure that it is dealing with real humans.

The TicketMaster Captcha



Verification Code

This step helps prevent unfair use of automated programs.



Enter the verification code as it is shown in the box above.

[Can't see the code?](#)

[What's this?](#)
Some people use automated programs to block other customers from getting tickets. To counteract this practice we ask that you type in the code before we put tickets on hold for you. You do not have permission to access this web site if you are using an automated program.
[How Does Code Verification Work?](#)

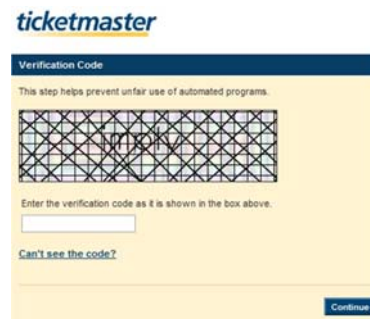
[Continue](#)

TicketMaster v. RMG Technologies

- RMG Technologies of Pittsburgh developed software call the Ticket Broker Acquisition Tool (TBAT) which
 - solves TicketMaster's Captchas
 - floods Ticketmaster with thousands of automated requests to freeze out legitimate customers
 - enables purchase of large numbers of tickets to popular events
- RMGs customers use the software to corner the market on hot tickets enabling them to resell the tickets for thousands of dollars apiece
- One such RMG customer made 425,000 ticket requests in a single day

After the Lawsuit

- Ticketmaster changed its captchas:



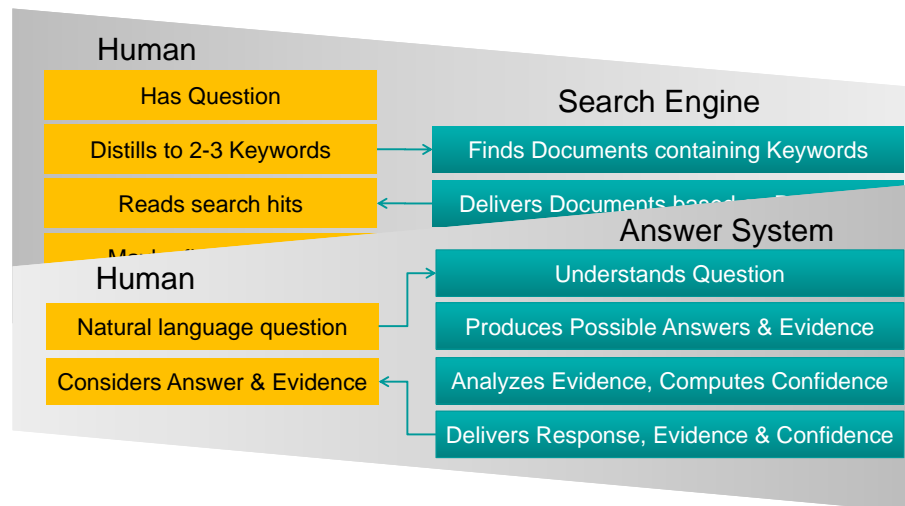
- Now humans have trouble with them!

Better Test: Ask a Question



- This is clever. It requires
 - image processing
 - understanding of metro lines
 - note: one stop west of Smithsonian is to the NORTH
- What does it take for machines to answer questions?

Search Engines vs. Question Answering



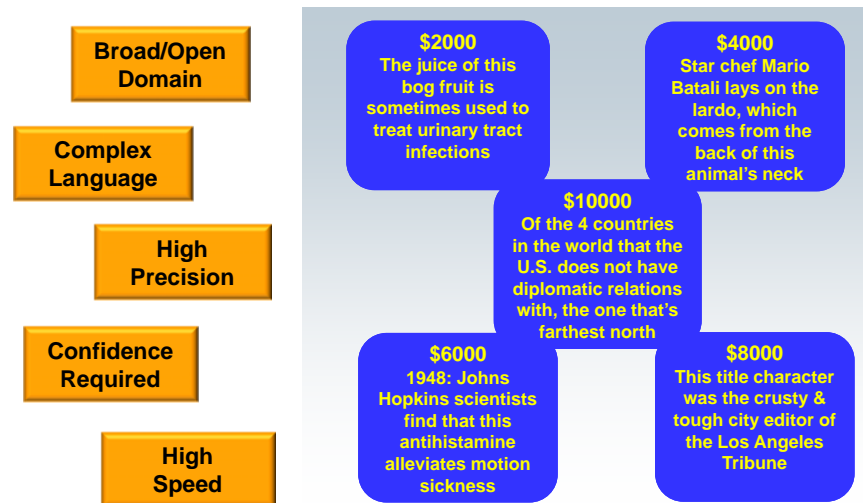
The Jeopardy! Game

- Regarded as a test of human intelligence
- In Jeopardy!, the “answer” is given – the player has to form the question
- Examples:

U.S. CITIES	Country Clubs	Authors
St. Petersburg is home to Florida's annual tournament in this game popular on ship decks (What is Shuffleboard?)	From India, the shashpar was a multi-bladed version of this spiked club (What is a mace?)	Archibald MacLeish based his verse play "J.B." on this book of the Bible (What is Job)
Rochester, New York grew because of its location on this (What is the Erie Canal?)	A French riot policeman may wield this, simply the French word for "stick" (What is a baton?)	In 1928 Elie Wiesel was born in Sighet, a Transylvanian village in this country (What is Romania?)

- This is a massive problem, requiring a huge amount of data, ability to make inferences and weigh competing data sources

The Jeopardy! Game



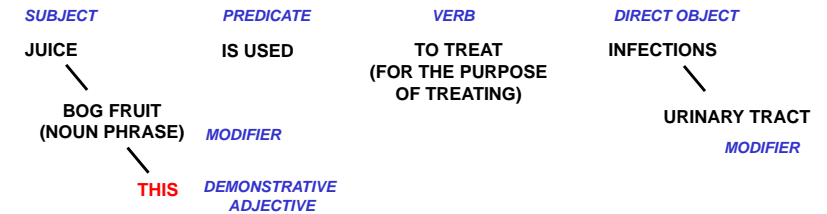
Questions have dollar values

SOURCE: IBM

Steps in Question Answering

- Understand the question
- What is it asking for?
- Parse it

\$2000
The juice of this bog fruit is sometimes used to treat urinary tract infections



- Here the answer is a type of “bog fruit” because of the presence of the demonstrative adjective “**THIS**”

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Answering Strategy

- Answer cannot be “looked up”
- Probably no reference says “The cranberry is a bog fruit sometimes used to treat urinary tract infections”
- Might be an article associating cranberries with urinary tract infections
- Might also be an article about bog fruits that lists cranberries (and others)
- Strategy 1:
 - make a list of bog fruits
 - make a list of urinary tract infection treatments
 - look for a match
 - do this within 3 seconds

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Answering Strategy

- The strategy requires lots of information, such as knowing urinary tract infection treatments
- Where does it come from? The Web
- Must ingest huge amounts of data (web pages) in advance, and index them for fast retrieval

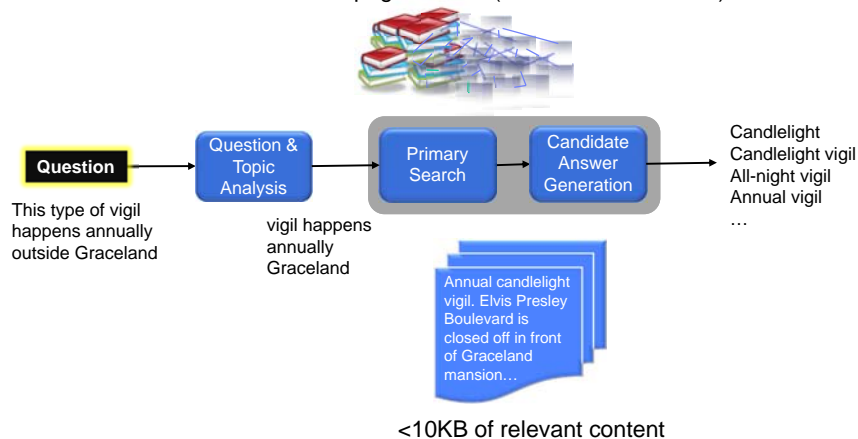
UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Finding Relevant Content

200M pages of text (about 1 million books)



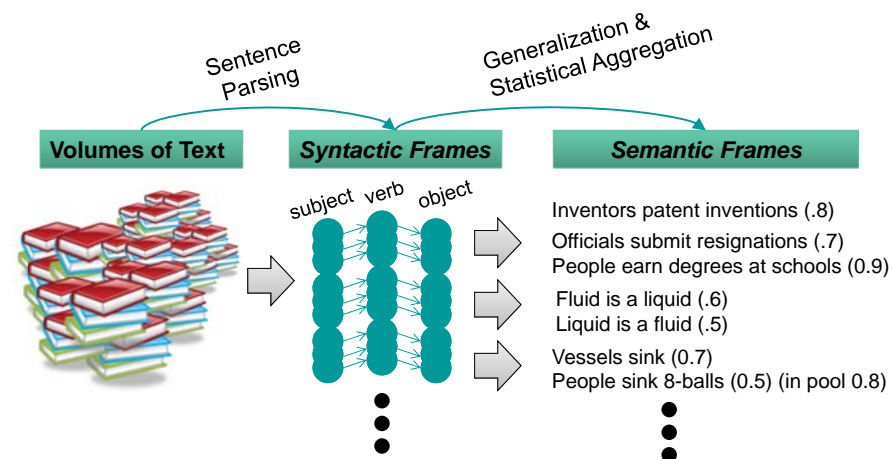
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Ingesting Web Content



SOURCE: IBM

Watson Architecture Phases

- **Analyze Question:** Use language processing algorithms to learn what it being asked
- **Primary Search:** Retrieve content related to the question (both unstructured text and structured knowledge-base entries)
- **Candidate Answer Generation:** From the retrieved content, extract phrases that could be possible answers
- **Evidence Retrieval:** For each Candidate Answer, retrieve more related content
- **Evidence Scoring:** Algorithms determine the degree to which retrieved evidence supports the Candidate Answers.
- **Merging and Ranking:** Consider all the scored evidence to produce a final ranked list of answers with confidences

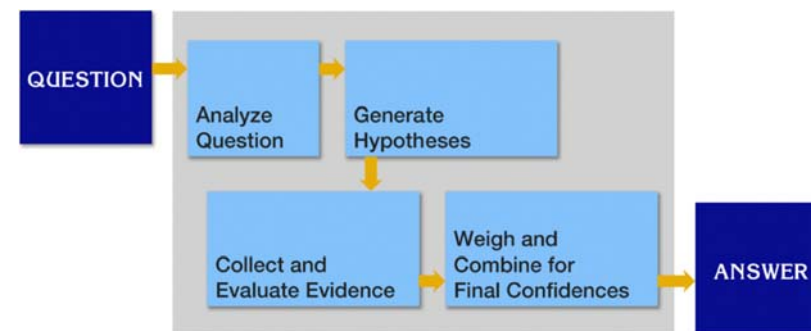
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Watson Question Processing



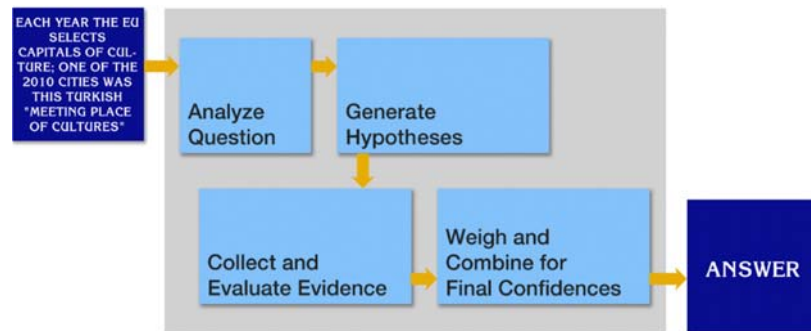
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Watson Question Processing



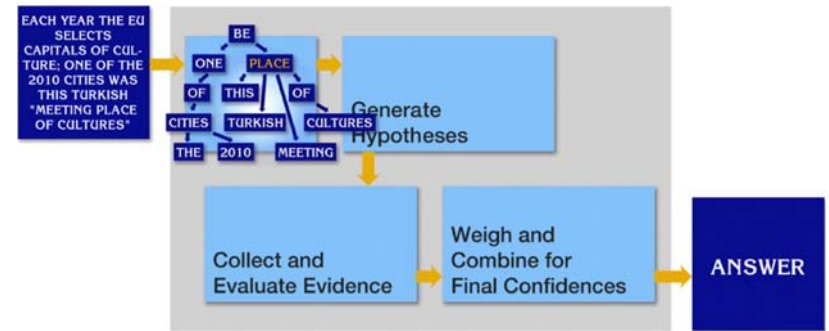
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Watson Question Parsing



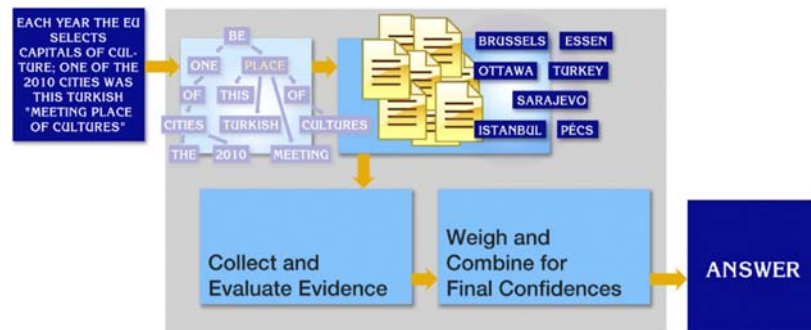
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Generate Hypotheses



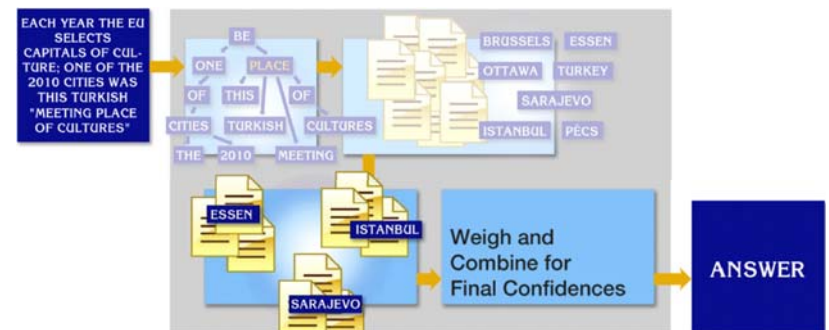
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Evaluate Supporting Evidence



SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Apply Confidence Ranking



SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Choose Answer



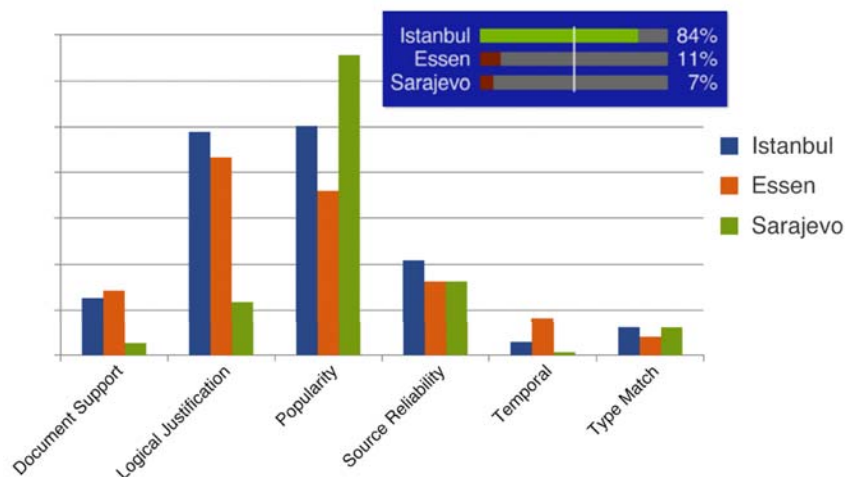
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Establishing Confidence



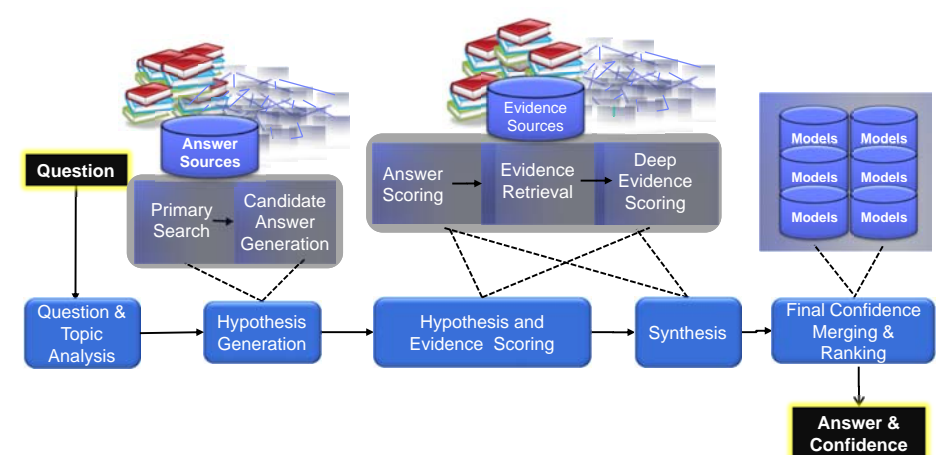
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Watson Architecture (Simplified view)



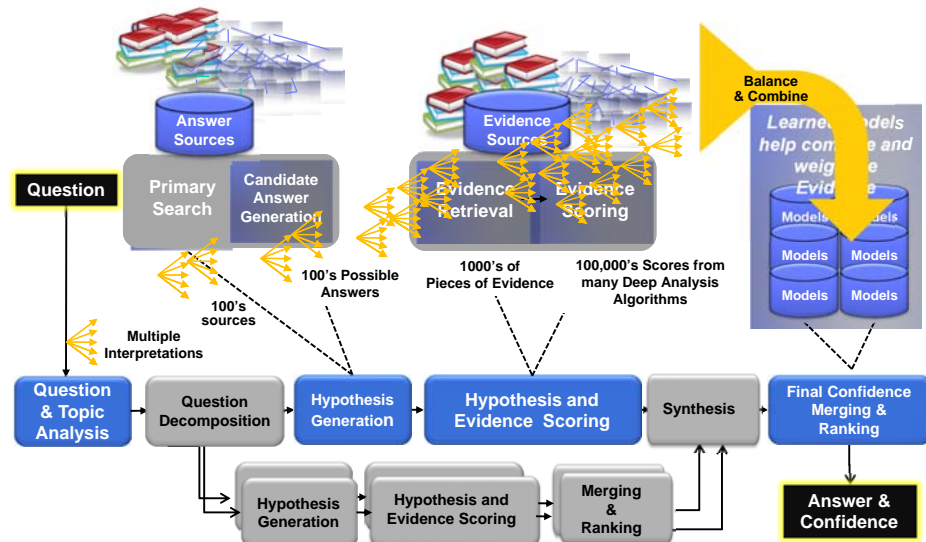
SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

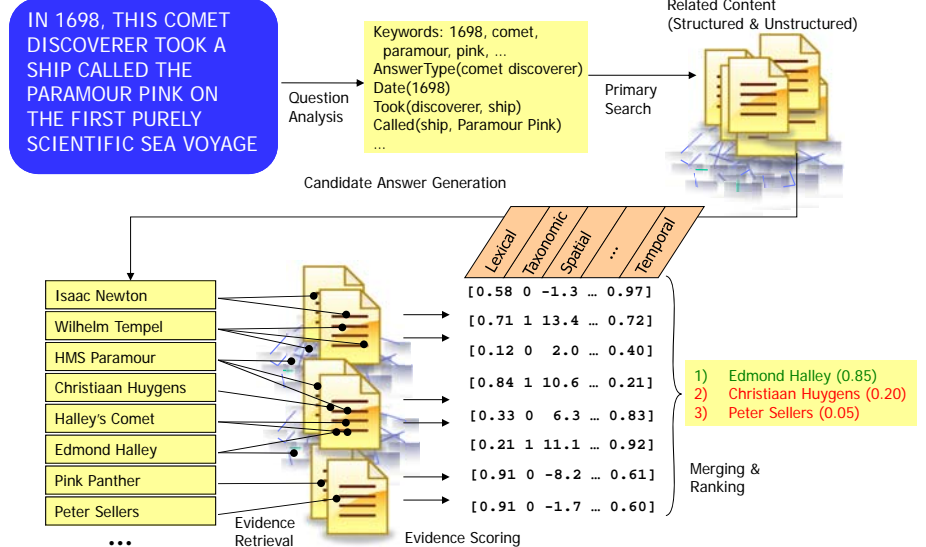
© 2013 MICHAEL I. SHAMOS

IBM Watson Software Architecture



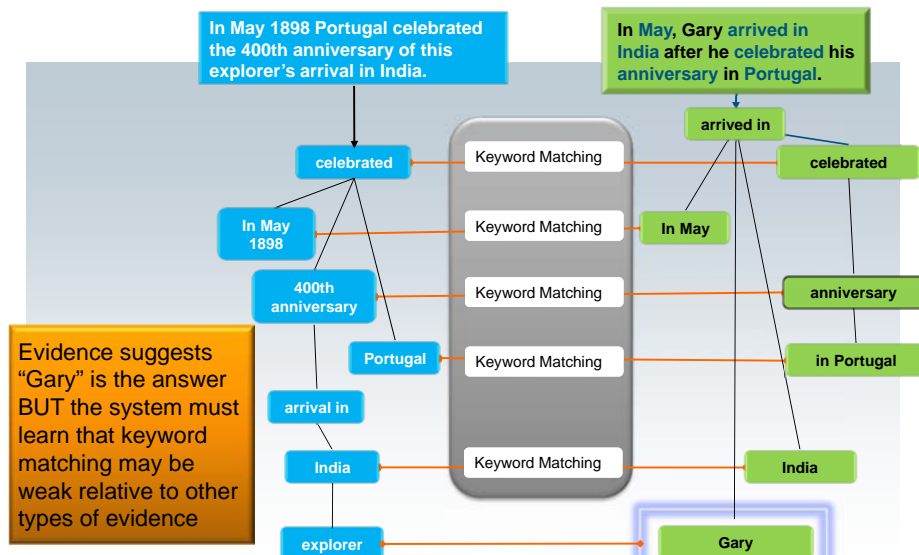
SOURCE: IBM

Sample Jeopardy Question



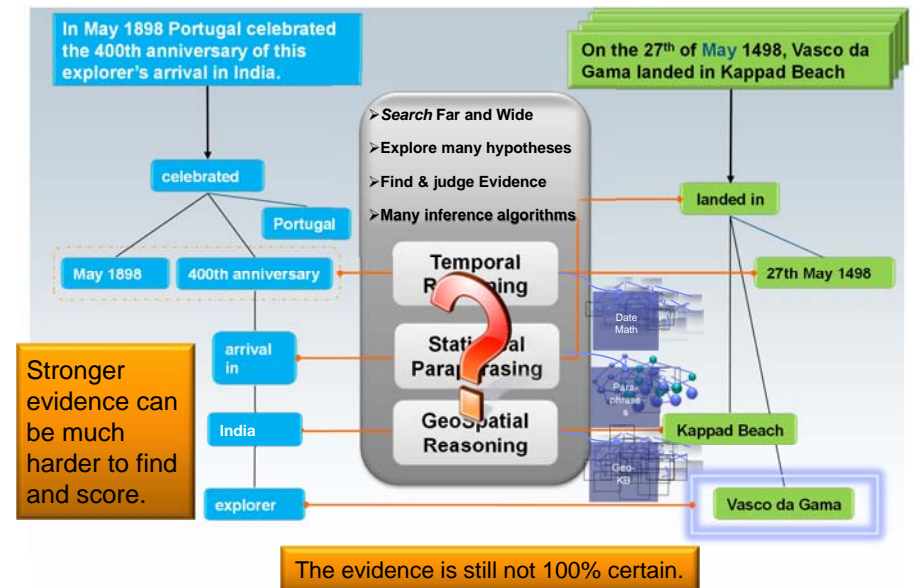
SOURCE: IBM

Keyword Evidence



SOURCE: IBM

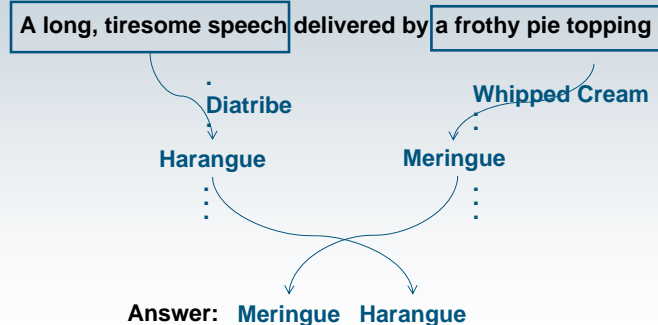
Different Types of Evidence



SOURCE: IBM

Short Questions Can Be Complex

Category: Edible Rhyme Time
(involves food and a rhyme)



SOURCE: IBM

IBM Watson Hardware

- 90 IBM Power 750 servers
- 2880 POWER7 cores
- POWER7 3.55 GHz chip
- 500 GB/sec on-chip bandwidth
- 10 Gb Ethernet network
- 15 Terabytes of RAM
- 20 Terabytes of disk
- Searches vast amounts of text using
 - Unstructured Information Management Architecture (UIMA, open source)
 - Apache Hadoop (open source) distributed file system
- 10 racks with servers, networking, shared disks, cluster controllers



SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Watson Hardware Capabilities

- Operates at 80 teraflops (trillion floating-point operations per second). Human brain processing power: ~ 100 teraflops.
- 200 million times faster than the Space Shuttle's computers
- In 3 seconds parses a 2000-foot shelf of books, picks out the relevant information, and creates an answer



SOURCE: IBM

UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Watson Easily Beats the Best Humans



UNIVERSITY OF HONG KONG

ASK MY ROBOT

© 2013 MICHAEL I. SHAMOS

Future Watson Applications



Government: Improved Information Sharing and Education

Tech Support: Help Desks, Call Centers



Enterprise Knowledge Management, Business Intelligence



Law: Legal Reasoning, Dispute Resolution

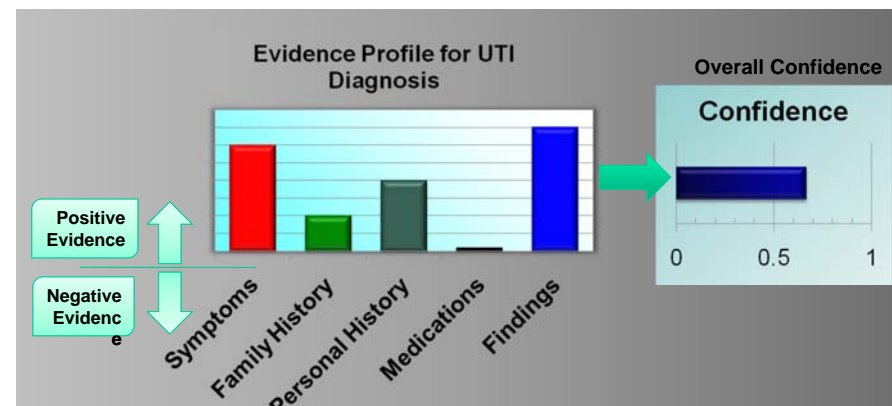


Healthcare: Diagnosis, Evidence-Based Collaborative Medicine

SOURCE: IBM

The Big Challenge: Medical Diagnosis

- Must consider
 - Strength of evidence and
 - Importance of evidence to diagnosis (learned from training data)
- Combine evidence dimensions to produce an overall confidence



SOURCE: IBM

Watson Evidence-Based Diagnosis



Symptoms



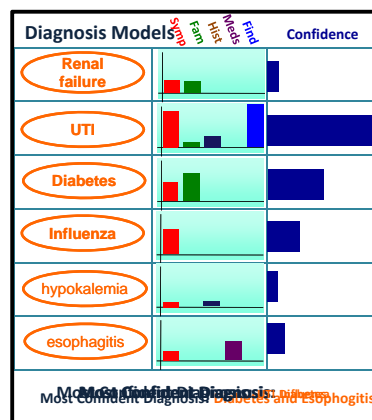
Family History
Patient History
Medications
Tests/Findings



Notes/Hypotheses



Huge Volumes of Texts, Journals, References, DBs etc.



SOURCE: IBM

The Big Idea:

Evidence-Based Reasoning over Natural Language Content

- **Deep Analysis** of clues/questions AND content
- Search for many possible answers based on different interpretations of question
- **Find, analyze and score EVIDENCE** from many different sources (not just one document) for each answer using many advanced NLP and reasoning algorithms
- **Combine evidence** and compute a confidence value for each possibility using statistical machine learning
- Rank answers based on confidence
- If top answer is above a threshold – buzz in else keep quiet

SOURCE: IBM

Q&A

Evidence Profiles Across Sources

Clue: Chile shares its longest land border with this country.

