Ask My Robot:
How Computers Answer Questions

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- )

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
• 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

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

Search Engine

- Finds Documents containing Keywords
- Delivers Documents based on Popularity

Answer System

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

Human

- Has Question
- Distills to 2-3 Keywords
- Reads search hits
- Has Question
  - Natural language question
  - Considers Answer & Evidence

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
- St. Petersburg is home to Florida’s annual tournament in this game popular on ship decks
- What is shuffleboard?

Country Clubs
- From India, the shashpar was a multi-bladed version of this spiked club
- What is a mace?

Authors
- Archibald MacLeish based his verse play “J.B.” on this book of the Bible
- What is Job

- Rochester, New York
- Because of its location on this
- What is the Erie Canal?

- A French riot policeman may wear 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?
The juice of this bog fruit is sometimes used to treat urinary tract infections.

Star chef Mario Batali lays on the lardo, which comes from the back of this animal’s neck.

1948: Johns Hopkins scientists find that this antihistamine alleviates motion sickness.

This title character was the crusty & tough city editor of the Los Angeles Tribune.

Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north.

Questions have dollar values.

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.

$1000: Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north.

$4000: Star chef Mario Batali lays on the lardo, which comes from the back of this animal's neck.

$6000: 1948: Johns Hopkins scientists find that this antihistamine alleviates motion sickness.

$8000: This title character was the crusty & tough city editor of the Los Angeles Tribune.

$10000: Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north.

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

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
Finding Relevant Content

200M pages of text (about 1 million books)

Question & Topic Analysis

Primary Search

Candidate Answer Generation

Candlelight vigil
All-night vigil
Annual vigil...

This type of vigil happens annually outside Graceland

Annual candlelight vigil. Elvis Presley Boulevard is closed off in front of Graceland mansion...

<10KB of relevant content

UNIVERSITY OF HONG KONG
ASK MY ROBOT
© 2013 MICHAEL I. SHAMOS

SOURCE: IBM

Ingesting Web Content

Volumes of Text

Syntactic Frames

Semantic Frames

Sentence Parsing

Generalization & Statistical Aggregation

Inventors patent inventions (.8)

Officials submit resignations (.7)

People earn degrees at schools (0.9)

Fluid is a liquid (.6)

Liquid is a fluid (.5)

Vessels sink (0.7)

People sink 8-balls (0.5) (in pool 0.8)

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

Watson Question Processing

QUESTION

Analyze Question

Generate Hypotheses

Collect and Evaluate Evidence

Weigh and Combine for Final Confidences

ANSWER

SOURCE: IBM
Establishing Confidence

Watson Architecture (Simplified view)

Apply Confidence Ranking

Choose Answer

SOURCE: IBM
IBM Watson Software Architecture

Sample Jeopardy Question

1. **In 1698, this comet discoverer took a ship called the Paramour Pink on the first purely scientific sea voyage.**

   - Wilhelm Tempel (0.71)
   - HMS Paramour (0.84)
   - Christiaan Huygens (0.84)
   - Halley’s Comet (0.12)
   - Edmond Halley (0.33)
   - Pink Panther (0.36)
   - Peter Sellers (0.91)
   - Edmond Halley (0.91)
   - Christiaan Huygens (0.91)

   **Question Analysis**

   - **Keywords:** 1698, comet, discoverer, ship
   - **Answer Type:** comet discoverer
   - **Date:** 1698
   - **Took (discoverer, ship):** Called (ship, Paramour Pink)

   **Evidence Scoring**

   - [0.58 0 1.3 0.97]
   - [0.71 1 13.4 0.72]
   - [0.12 0 2.0 0.40]
   - [0.84 1 10.6 0.21]
   - [0.33 0 6.3 0.43]
   - [0.21 1 11.1 0.92]
   - [0.91 0 -8.2 0.61]
   - [0.91 0 -1.7 0.60]

   **Merging & Ranking**

   - [0.58 0 1.3 0.97]
   - [0.71 1 13.4 0.72]
   - [0.12 0 2.0 0.40]
   - [0.84 1 10.6 0.21]
   - [0.33 0 6.3 0.43]
   - [0.21 1 11.1 0.92]
   - [0.91 0 -8.2 0.61]
   - [0.91 0 -1.7 0.60]

   **Related Content**

   - [Structured & Unstructured]

   **Primary Search**

   - [Search Far and Wide]
   - [Explore many hypotheses]
   - [Find & judge Evidence]
   - [Many inference algorithms]

   **Synthesis**

   - [Temporal Reasoning]
   - [Statistical Paraphrasing]
   - [Geospatial Reasoning]

   **Evidence**

   - [27th of May 1498 Vasco da Gama landed in Kappad Beach]
   - [Vasco da Gama]
   - [Kappad Beach]
   - [27th May 1498]
   - [400th anniversary]
   - [arrival in]
   - [India]
   - [Portuguese explorer]

   **Stronger evidence can be much harder to find and score.**

   **The evidence is still not 100% certain.**

   **Different Types of Evidence**

   - **Temporal**
   - **Statistical**
   - **Paraphrasing**
   - **Geospatial**

   **Learned Models**

   - help combine and weigh the Evidence Sources

   **Answer Sources**

   - Multiple interpretations
   - 100’s Possible Answers
   - 1000’s of Pieces of Evidence
   - 100’s sources
   - 100,000’s Scores from many Deep Analysis Algorithms

   **Candidate Answer Generation**

   1) Edmond Halley (0.85)
   2) Christiaan Huygens (0.20)
   3) Peter Sellers (0.05)

   **Evidence Retrieval**

   - Question Decomposition
   - Hypothesis Generation
   - Evidence Scoring
   - Synthesis
   - Final Confidence Merging & Ranking

   **Answer & Confidence**

   - Question & Topic Analysis
   - Primary Search
   - Candidate Answer Generation
   - Evidence Retrieval
   - Evidence Scoring
   - Merging & Ranking

   **Keyword Evidence**

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

   - Evidence suggests “Gary” is the answer BUT the system must learn that keyword matching may be weak relative to other types of evidence.

   - In May, Gary arrived in India after he celebrated his anniversary in Portugal.
Short Questions Can Be Complex

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

A long, tiresome speech delivered by a frothy pie topping

Diatribe

Harangue

Whipped Cream

Meringue

Answer: Meringue Harangue

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

Watson Easily Beats the Best Humans

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
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

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

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

Argentina

Bolivia

Bolivia is more popular due to a commonly discussed border dispute.

Evidence Profiles Across Sources

SOURCE: IBM