Watson: In Winter 2011 it won at Jeopardy!

- Let’s try it.

Watson playing Jeopardy

Watson Does
- Use IBM’s TTS engine to
  - Speak Clue Selections
  - Speak Responses
- Physically press the button to Ring-in
- Gets the clue electronically when humans see it

Watson Does NOT (today)
- Hear
- See
- Process Audible/Visual Clues
- These are excluded from the contest
- Have a life-line
  - It is a completely self-contained HW/SW system.
  - Highlighting the human vs. machine comparison at this task

Two years later: Not just Jeopardy!

IBM’s Watson Gets Its First Piece Of Business In Healthcare

IBM’s Watson, the Jeopardy-winning supercomputer that cost $100 million to build back in 2011, just got a whole new role. The Watson project is now an enterprise software platform, as IBM grows its healthcare business.

Watson was trained on a large set of public and private data sources, including medical records from the Mayo Clinic, scientific literature, and publicly available resources. It was designed to answer questions that can’t be answered by reviewing a single data set, but require training in a variety of different domains. Additionally, it has access to a vast amount of data in real-time.

Watson will now be able to help doctors and other healthcare providers with their daily tasks.

Creativity in the kitchen?

[http://www.psfk.com/2013/03/ibm-watson-top-chef.html](http://www.psfk.com/2013/03/ibm-watson-top-chef.html)
What is the underlying technology called?
Question Answering (QA)
This is a Natural Language Processing Technology

The Turing Test
- Turing in 1950 published a philosophical paper designed to stop people arguing about whether or not machines could think.
- He proposed that the question be replaced with a test, which is what is now called the Turing Test.

Types of Question Answering
- Factoid
  - Who discovered oxygen?
  - When did Hawaii become a state?
  - Where is Ayers Rock?
  - What team won the World Series in 1992?
- List
  - What countries export oil?
  - Name U.S. cities that have a "Shubert" theater.
- Definition
  - Who is Aaron Copland?
  - What is a quasar?

How long have people been working on it?
- TREC = Text REtrieval Conferences
  - Series of annual evaluations, started in 1992
  - Organized into "tracks"
  - Test collections are formed by "pooling"
  - Gather results from all participants
  - Corpus/topics/judgments can be reused
- TREC has had a QA Track since 1999.
  - http://trec.nist.gov/data/qa/T8_QAdata/development.qa
Types of Question Answering

- http://trec.nist.gov/data/qa/T8_QAdata/development.qa

Roots of Question Answering

- Information Retrieval (IR)
- Information Extraction (IE)

Motivated by the observation that Information Retrieval does not answer your question, it provides a ranked list of documents that “might” contain the answer to your question.

Information Extraction (IE)

- IE systems (originally… but recent advances)
  - Identify documents of a specific type
  - Extract information according to pre-defined templates
  - Place the information into frame-like database records

<table>
<thead>
<tr>
<th>Weather disaster</th>
<th>Type</th>
<th>Damage</th>
<th>Date</th>
<th>Location</th>
<th>Deaths</th>
</tr>
</thead>
</table>

- Templates = pre-defined questions
- Extracted information = answers
- Limitations
  - Templates are domain dependent and not easily portable

Information Retrieval (IR)

- The first versions of search engines
- IR systems
  - Use statistical methods
  - Rely on frequency of words in query, document, collection
  - Retrieve complete documents
  - Return ranked lists of “hits” based on relevance
- Limitations
  - Answers questions indirectly (with a ranked list of documents that may or may not have the answer)
  - Does not attempt to understand the “meaning” of user’s query or documents in the collection

Web search: It Matters How It Works

1. Gather information.
2. Keep copies.
3. Build an index.
4. Understand the query.
5. Determine the relevance of each possible result to the query.
6. Determine the ranking of the relevant results.
7. Present the results.
In five years, QA will be in all web search engines
Remember: you heard it here first!

Question Answering: A type of NLP

Contrast IR/Search with Asking an Expert

What does Jeopardy Host do?

Types of Question Answering
- Factoid
  - Who discovered oxygen?
  - When did Hawaii become a state?
  - Where is Ayers Rock?
  - What team won the World Series in 1992?
- List
  - What countries export oil?
  - Name U.S. cities that have a "Shubert" theater.
- Definition
  - Who is Aaron Copland?
  - What is a quasar?
Central Idea of Factoid QA

- Determine the semantic type of the expected answer
  
  "Who won the Nobel Peace Prize in 1991?" is looking for a PERSON

- Retrieve documents that have keywords from the question
  
  Retrieve documents that have the keywords "won", "Nobel Peace Prize", and "1991"

- Look for named-entities of the proper type near keywords
  
  Look for a PERSON near the keywords "won", "Nobel Peace Prize", and "1991"

An Example

Who won the Nobel Peace Prize in 1991?

- The military junta took power in 1988 as pro-democracy demonstrations were sweeping the country. It held elections in 1990, but has ignored their result. It has kept the 1991 Nobel peace prize winner, Aung San Suu Kyi, leader of the opposition party which won a landslide victory in the poll - under house arrest since July 1989.

- The regime, which is also engaged in a battle with insurgents near its eastern border with Thailand, ignored a 1990 election victory by an opposition party and is detaining its leader, Aung San Suu Kyi, who was awarded the 1991 Nobel Peace Prize. According to the British Red Cross, 5,000 or more refugees, mainly the elderly and women and children, are crossing into Bangladesh each day.

Generic QA Architecture

Question analysis

- Question word cues
  
  - Who → person, organization, location (e.g., city)
  - When → date
  - Where → location
  - What/Why/How → ??

- Head noun cues
  
  - What city, which country, what year...
  - Which astronaut, what blues band, ...

- Scalar adjective cues
  
  - How long, how fast, how far, how old, ...

Extracting Named Entities

Person: Mr. Hubert J. Smith, Adm. McInnes, Grace Chan
Title: Chairman, Vice President of Technology, Secretary of State
City: New York, Rome, Paris, Birmingham, Seneca Falls
Province: Kansas, Yorkshire, Uttar Pradesh
Business: GTE Corporation, FreeMarkets Inc., Acme
University: Bryn Mawr College, University of Iowa
Organization: Red Cross, Boys and Girls Club

What about structured knowledge sources like Wikipedia, online databases etc?
Example: What is the service ceiling of a U-2?

Wikipedia: What knowledge can we get from Wikipedia?


Using WordNet: Online Thesaurus.

- http://wordnetweb.princeton.edu/
- What is the service ceiling of a U-2?
- Can access it FROM a program (not just this interface).

More Named Entities

- Currency: 400 yen, $100, DM 450,000
- Linear: 10 feet, 100 miles, 15 centimeters
- Area: a square foot, 15 acres
- Volume: 6 cubic feet, 100 gallons
- Weight: 10 pounds, half a ton, 100 kilos
- Duration: 10 day, five minutes, 3 years, a millennium
- Frequency: daily, biannually, 5 times, 3 times a day
- Speed: 6 miles per hour, 15 feet per second, 5 kph
- Age: 3 weeks old, 10-year-old, 50 years of age

Answer Type Hierarchy: Sheffield TREC group

This is where technology was when IBM started their project.
Artificial Intelligence’s Great Quest: How did Watson do it?

Slides Courtesy of Jennifer Chu-Carroll
IBM Research

IBM’s big team: Jennifer “harder than I’ve ever worked in my life”

Core team funded by DARPA under TREC for more than 5 years before IBM decided to do this

Open Domain Question Answering

Automatic Open-Domain Question Answering
A Long-Standing 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

Jeopardy seemed like it could be possible

The Jeopardy! Challenge
A plausible, compelling and notable way to drive the technology of Question Answering along 5 Key Dimensions


Enabling Technologies – The Time Was Right

Natural Knowledge
- Large volumes natural language electronic text (e.g., news, wiki, Wikipedia, web, etc.)
- Encodes knowledge and greater linguistic context to better resolve intended meaning

NLP (Text Analysis)
- Entity and Relation Detection, Synonyms & Semantic Parsing
- Statistical NLP: Consider sentence, lower cost information extraction
- Statistical Parsing: Learn ways to express same meaning

Semi-Structured Knowledge
- Large volumes of Thesaurus, Dictionaries, Terminology, Linked Data, and the Semantic Web
- Rapid, community-based construction
- Always new domains - Specialized and General

Compute Power
- Massive parallel compute power
- 1000s of computer cores working simultaneously
- Tens of globally addressable main memory

Answer Type Hierarchy
STEP 1: Get a massive number of Jeopardy questions and analyze them
(if we are going to do this we have to win! How can we win??)

First finding: Cannot Anticipate Answer Types!!

Broad Domain

We DO NOT attempt to anticipate all questions and build specialized databases.

In a random sample of 20,000 questions we found 2,500 distinct types. The most frequent occurring <3% of the time. The distribution has a very long tail.

And for each these types 100% of different things may be asked.

Even going for the head of the tail will barely make a dent.

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

Interesting tradeoff: Knowledge, Precision, Open Domain

Structured knowledge vs. text search (clicker)

What does this graph show?

A. Structured knowledge sources can answer most of the questions with high confidence
B. Text search by itself can never answer more than 30% of the questions
C. Structured knowledge by itself can never answer more than 50% of the questions with 50% confidence
D. Watson must combine deep and shallow semantic analysis (natural language understanding) to win

Does WordNet know about Chocolate Cake?

Chocolate Decadence Cake?: The Web knows all!

WordNet Search - 3.1

Word to search for: Chocolate cake

Display Options: All senses & synonyms

Display for sense (gloss): "an example sentence"

Noun
- indirect chocolate cake (cake containing chocolate)
  - direct hypernym / hyponym
- direct hypernym / hyponym / hypernym / synonym

Chocolate Decadence Cake?: Recipes - Allrecipes.com

This is the first chocolate cake recipe by Allrecipes.com.

Chocolate Decadence Cake Recipe - Allrecipes.com

Ingredients:
- 1 cup all-purpose flour
- 1 cup granulated sugar
- 3/4 cup unsweetened cocoa powder
- 2 cups milk
- 1/2 cup vegetable oil
- 2 large eggs
- 2 teaspoons vanilla extract
- 2 teaspoons baking powder
- 1 teaspoon baking soda
- 1/2 teaspoon salt

Steps:
1. Preheat oven to 350°F (175°C).
2. Lightly grease the bottoms and sides of two 9" x 13" pans.
3. In a large bowl, combine flour, sugar, cocoa, baking powder, baking soda, and salt; stir well.
4. Add milk, oil, eggs, and vanilla; beat with a mixer on medium speed until well combined.
5. Pour batter evenly into prepared pans.
6. Bake for 30 to 35 minutes or until a toothpick inserted in the center comes out clean. Cool completely in pans on a wire rack.

Chocolate Decadence Cake II Recipe - Allrecipes.com

This is a rich, thick cake that is very moist and fudgy. A great dessert for special occasions!

Chocolate Decadence Cake II Recipe - Allrecipes.com

Ingredients:
- 2 cups all-purpose flour
- 3/4 cup unsweetened cocoa powder
- 3/4 cup granulated sugar
- 1 cup packed dark brown sugar
- 1 1/2 teaspoons baking powder
- 1/2 teaspoon baking soda
- 1/2 teaspoon salt
- 1 cup buttermilk
- 1/2 cup vegetable oil
- 2 large eggs
- 1 teaspoon vanilla extract
- 1/2 cup milk

Steps:
1. Preheat oven to 350°F (175°C).
2. Grease two 9" round cake pans.
3. In a large bowl, combine flour, cocoa, sugar, brown sugar, baking powder, baking soda, and salt; stir well.
4. Add buttermilk, oil, eggs, and vanilla; beat well.
5. Stir in milk; beat just until smooth.
6. Pour batter into prepared pans.
7. Bake for 27 to 32 minutes or until a toothpick inserted in the center comes out clean.
8. Cool completely in pans on a wire rack.
How good does it have to be to win?

Incremental Progress

A Few Guiding Principles

This is not easy!
More NLP: techniques to learn knowledge from texts, invented in the last 10 years

Inducing Frames

Evaluating possibilities and their Evidence

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

Big Data: So much stuff on the web!

What is good evidence? Wikipedia?

Grouping Features => “Evidence Profiles”

Creative Question Types: Edible Rhyme
Decomposing into phrases

Question Decomposition: Nested

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

In 1968 this man was U.S. president.

When “60 Minutes” premiered, this man was U.S. president.

The DeepQA architecture attempts different decompositions and recursively applies the QA algorithms.

Incremental Progress


Playing in the Winners Cloud

Watson: In Winter 2011 it won at Jeopardy!

Two years later: Not just Jeopardy!

IBM’s Watson Gets Its First Piece Of Business In Healthcare

Creativity in the kitchen?

QA & NLP: Beyond Boolean Queries

Search Engine words are independent

- Words don’t have to occur together
- Use Boolean queries and quotes
- Logical Operators: AND, OR, NOT
  - monet AND water AND lilies
  - “van gogh” OR gauguin
  - vermeer AND girl AND NOT pearl

http://www.psfk.com/2013/03/ibm-watson-top-chef.html
The types of questions studied in research on question answering include:

A. Definition Questions: e.g. Who is Jared Borgetti?
B. Factoid Questions: e.g. How many feet above sea level is the UCSC Campus?
C. List Questions: e.g. Name the players on Brazil’s World Cup Winning Team in 2002.
D. None of the above
E. All of the above

Factoid QA

* The central idea of Factoid questions is to determine the semantic type of the expected answer. Some of the possible semantic types of expected answers are:

A. Persons and Places
B. Distances, Heights, Lengths and other numerical values
C. Universities, Businesses, Organizations
D. Types of Animals, such as Reptile or Mammal
E. All of the above

This graph from Prof. Walker’s lecture on IBM’s Watson was used to demonstrate that:

A. It is always possible to determine the semantic type of the expected answer
B. Anticipating the question types and building specialized databases would not let Watson win at Jeopardy
C. Most questions have to do with Distances, Heights, Lengths and other numerical values
D. Online encyclopedias have most answers already pre-computed in an easy to extract format
E. All of the above

The potential business applications of QA are:

A. Customer care and help systems
B. Answering Healthcare questions
C. A & B
D. Playing Jeopardy, which can’t make IBM much money
E. None of the above

Information retrieval technology was the start for web search

Web Search Review.
Web search: It Matters How It Works

1. Gather information.
2. Keep copies.
3. Build an index.
4. Understand the query.
5. Determine the relevance of each possible result to the query.
6. Determine the ranking of the relevant results.
7. Present the results.

Its all free?? : Well no. Who Pays for What?

- Users could pay a subscription fee (early AOL and CompuServe)
- Web sites could pay for being indexed.
- The government could pay (taxes!).
- Advertisers could pay.
- And it matters who pays cause it affects how it works

Clickers

- Google responds efficiently to queries by going out and searching the web in real time.

A. TRUE
B. FALSE

Which steps happen in the background?

A. Gather information & Keep copies.
B. Gather information & Keep copies & Build an index.
C. Understand the query.
D. Determine the relevance of each possible result to the query & Determine the ranking of the relevant results.

Clickers: It Matters How It Works

“The search tools that help us find needles in the digital haystack have become the lenses through which we view the digital landscape. Businesses and governments use them to distort our picture of reality.”

A. TRUE
B. FALSE

Building an Index
How many hosts are out there?

A. Every website has to register with Google to get indexed
B. Every website has to pay Google to get indexed
C. Google knows which pages to index because of your digital footprint
D. Google has a list of ‘trusted pages’ and it just follows the links from them

How does it find the pages to index?

A. About 50% on average but some days it’s 100% and some days it’s 30%
B. Less than 5%
C. Best estimates put it at 30% because of all the bots that keep crawlers out.

Clickers: What goes into the index?

A. Every word on the page
B. The words that other sites use when they point to (link to) this page.
C. Only the keywords on the page.
D. Google has a list of special keywords that all pages get indexed by.

Clickers

* The federal government uses your tax dollars to guarantee that Google and other search engine providers like Microsoft (Bing) return the objectively best results of your query.

A. TRUE
B. FALSE
4. Understand the Query

- Steps 1-3 happen in “the background”
- Not much “understanding” in today’s search engines but that could change soon (QA lecture 3/12).
- Advanced search engine features help

49’ers beat Ravens vs. Ravens beat 49’ers

Powerset. Bought by MS.

---

Boolean Queries

Search Engine words are independent

- Words don’t have to occur together
- Use Boolean queries and quotes
- Logical Operators: AND, OR, NOT
  - monet AND water AND lilies
  - “van gogh” OR gauguin
  - vermeer AND girl AND NOT pearl

---

QA & NLP: Beyond Boolean Queries

Search Engine words are independent

- Words don’t have to occur together
- Use Boolean queries and quotes
- Logical Operators: AND, OR, NOT
  - monet AND water AND lilies
  - “van gogh” OR gauguin
  - vermeer AND girl AND NOT pearl

---

How does the index get used at search time by default? (without advanced search)

A. Every website registers with Google exactly which terms to index by and which combinations
B. Google uses “boolean” combinations. The index is made of single words. Google ANDS them together and finds which webpages (URLs) are in the intersection of all the terms
C. Google indexes the pages individually for each person using your digital footprint
D. All of the above.

---

Relevance and Ranking
How does it rank the pages it finds?

A. It uses what is called a “page rank” algorithm, that uses many different factors
B. It depends only on who is willing to pay the most.
C. The government tells it how to rank pages.
D. None of the above.

The information that Page Rank uses includes:

- Keywords in heading or titles and keywords in the body text
- Information about whether the site is “trustworthy”
- Whether the links on this page are to relevant pages
- Whether the links to this page are relevant
- Age of the page
- Quality of the text (e.g., absence of misspellings)

Search engine possible funding models:

A. Users could pay a subscription fee
B. Web sites could pay for being indexed.
C. The government could pay using taxes the same way they pay for roads or police.
D. Advertisers could pay for having their ads featured in the side bar.
E. All of the above are possible.

Search engine current funding model:

A. Users pay a subscription fee
B. Web sites pay for being indexed.
C. The government pays using taxes the same way they pay for roads or police.
D. Advertisers pay for having their ads featured in the side bar.
E. B&D

What makes search attractive to advertisers?

A. Advertisements can be targeted more precisely using your digital footprint.
B. Advertisers only pay when you click on their ad.
C. Millions of people use search every day.
D. All of the above.

Advertiser only pays when you click!
Summary

- Search engines offer unprecedented access to information.
- Search engines place the power to shape what we see into the hands of a few companies.
- Search engines continue to evolve.
  - Recently adding in Google Plus
  - Startups on indexing twitter etc
- Question Answering is the next new thing!