Question answering

- **Overview and task definition**
- **History**
- **Open-domain question answering**
- **Basic system architecture**
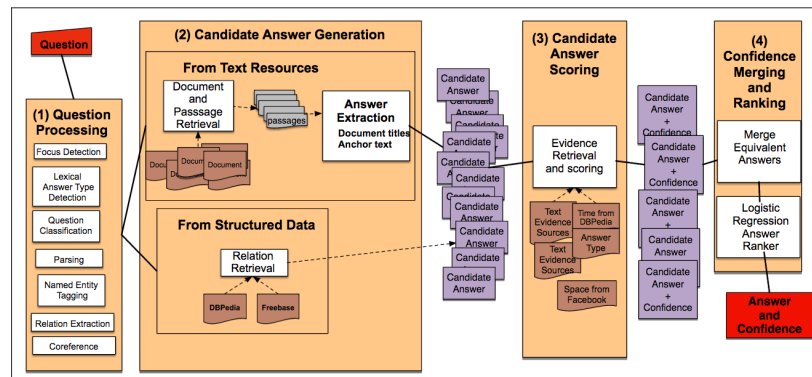 → – Watson architecture

---

Watson

– http://www.youtube.com/watch?v=WFR3lOm_xhE

Statistics

- **Development team:** **25 people**
- **Project duration:** **4 years**
- **Software:** **700K lines of Java**

  300K C++

  ~130 components
- **Hardware:** **90 IBM Power 750 servers**

  2880 Power7 cores @

80+TFLOPS

20 TB memory

10 Gbps network
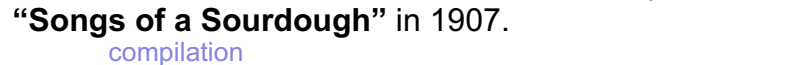

Watson's architecture

Figure 28.9 The 4 broad stages of Watson QA: (1) Question Processing, (2) Candidate Answer Generation, (3) Candidate Answer Scoring, and (4) Answer Merging and Confidence Scoring.

Question processing

- NLP components
  - Parsing (syntactic analysis)
  - Named entity tagging
  - Relation extraction
    - Haven't covered this yet except for *opinions*
    - Examples:
      - Parent-of (x, y)
      - Employee-of (x, y = company)
  - Coreference resolution

Example

Poets and Poetry:

GPE

He was a bank *clerk* in the **Yukon** before he published
"Songs of a Sourdough" in 1907.
compilation

author-of (focus, "Songs of a sourdough")
publish (e1, he, "Songs of a sourdough")
in (e2, e1, 1907)
temporal-link(publish(...), 1907)

Determine focus

- **Focus** = the part of the question that co-refers with the answer
  - aligns with a supporting passage
- extracted by hand-written rules
  - possible due to stylized syntax of Jeopardy questions
  - e.g.
    - **He** was a bank clerk in the Yukon…
    - When hit by electrons, a phosphor gives off electromagnetic energy in **this form**
    - Secretary Chase just submitted **this** to me for the third time; guess what, pal. This time I'm accepting it.
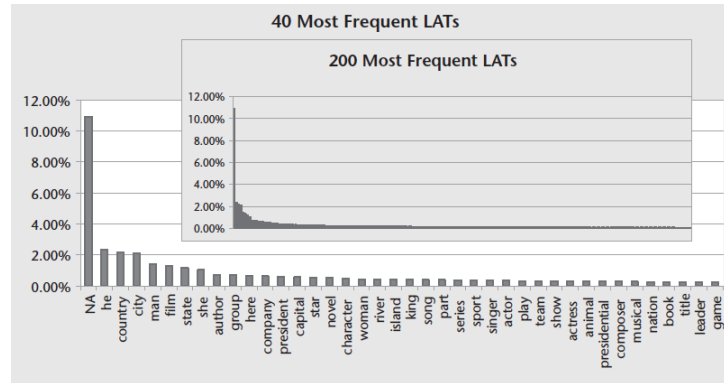
Identify answer type

- IBM calls it the LAT (lexical answer type)
- **word or words in the clue that tell us something about the semantic type of the answer**
- Extracted via hand-crafted rules
- NOT used as filter to rule out answers
- Watson extracts lots of answers, unconstrained by answer type, and a set of answer types, and then later ('candidate answer scoring' phase) scores how well each answer fits the answer
- Just one of many sources of evidence

LATs

- Analyzed 20,000 questions by hand to extract LAT
- ~5000 distinct LATs
- NE tagger with over 100 named entity types covered

less than half the types in these ques- tions.

- LATs for standard Factoid-based QA systems

Tag	Example
ABBREVIATION	
abb	What's the abbreviation for limited partnership?
exp	What does the "c" stand for in the equation E=mc2?
DESCRIPTION	
definition	What are tannins?
description	What are the words to the Canadian National anthem?
manner	How can you get rust stains out of clothing?
reason	What caused the Titanic to sink ?
ENTITY	
animal	What are the names of Odin's ravens?
body	What part of your body contains the corpus callosum?
color	What colors make up a rainbow ?
creative	In what book can I find the story of Aladdin?
currency	What currency is used in China?
disease/medicine	What does Salk vaccine prevent?
event	What war involved the battle of Chapultepec?
food	What kind of nuts are used in marzipan?
instrument	What instrument does Max Roach play?
lang	What's the official language of Algeria?
letter	What letter appears on the cold-water tap in Spain?
other	What is the name of King Arthur's sword?
plant	What are some fragrant white climbing roses?
product	What is the fastest computer?
religion	What religion has the most members?
sport	What was the name of the ball game played by the Mayans?
substance	What fuel do airplanes use?
symbol	What is the chemical symbol for nitrogen?
technique	What is the best way to remove wallpaper?
term	How do you say " Grandma " in Irish?
vehicle	What was the name of Captain Bligh's ship?
word	What's the singular of dice?
HUMAN	
description	Who was Confucius?
group	What are the major companies that are part of Dow Jones?
ind	Who was the first Russian astronaut to do a spacewalk?
title	What was Queen Victoria's title regarding India?
LOCATION	
city	What's the oldest capital city in the Americas?
country	What country borders the most others?
mountain	What is the highest peak in Africa?
other	What river runs through Liverpool?
state	What states do not have state income tax?
NUMERIC	
code	What is the telephone number for the University of Colorado?
count	About how many soldiers died in World War II?
date	What is the date of Boxing Day?
distance	How long was Mao's 1930s Long March?
money	How much did a McDonald's hamburger cost in 1963?
order	Where does Shanghai rank among world cities in population?
other	What is the population of Mexico?
period	What was the average life expectancy during the Stone Age?
percent	What fraction of a beaver's life is spent swimming?
speed	What is the speed of the Mississippi River?
temp	How fast must a spacecraft travel to escape Earth's gravity?
size	What is the size of Argentina?
weight	How many pounds are there in a stone?

Figure 28.4 Question typology from Li and Roth (2002), (2005). Example sentences are from their corpus of 5500 labeled questions. A question can be labeled either with a coarse-

Question classification

- Definition question, multiple-choice, puzzle, fill-in-the-blank, decomposition
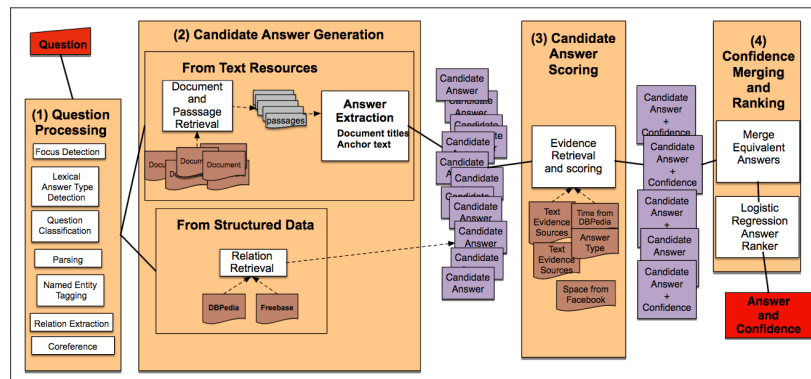- Via pattern-matching regular expressions over words or parse trees.

Decomposition

- Category: "Rap" Sheet
- Clue: This archaic term for a mischievous or annoying child can also mean a rogue or scamp.

- Subclue 1: This archaic term for a mischievous or annoying child.

- Subclue 2: This term can also mean a rogue or scamp.

- Answer: rapscallion

Decomposition

- Category: Diplomatic Relations
- Clue: Of the four countries in the world that the United States does not have diplomatic relations with, the one that's farthest north.

- Inner subclue: the four countries in the world that the United States does not have diplomatic relations with
  - (Bhutan, Cuba, Iran, North Korea)
- Outer subclue: Of Bhutan, Cuba, Iran, and North Korea, the one that's farthest north.
- Answer: North Korea

Watson's architecture

Figure 28.9 The 4 broad stages of Watson QA: (1) Question Processing, (2) Candidate Answer Generation, (3) Candidate Answer Scoring, and (4) Answer Merging and Confidence Scoring.

Content acquisition

- **Depends on answer types**
  - LATs
- **Wide range of**
  - Encyclopedias
  - Disctionaries
  - Thesauri
  - Newswire articles
  - Literary works
  - Taxonomies, ontologies, WordNet
  - Databases
- **Automatic corpus expansion**

Candidate answer generation

- For structured data sources, create query
  - E.g. author-of (focus,"Songs of a sourdough")
- Passage retrieval
- Answer candidates generated for
  - all anchor texts in the document (i.e. the text between <a> and <\a>)
  - extract all noun phrases
  - titles of documents
- Secondary search of Wikipedia
  - return **titles** of highly ranked articles vast majority of Jeopardy! answers are the title of a Wikipedia article.

Candidate answer scoring

- **Uses many sources of evidence**
- Compatibility with the lexical answer type
  - Use DBPedia, WordNet
- Time and space relations extracted from DBpedia or other structured databases
  - E.g. can extract temporal properties of the entity (when was a person born, when died) and then compare to time expressions in the question for compatibility.
- Text retrieval to help retrieve evidence supporting a candidate answer.
- **Output:** set of candidate answers, each with a vector of scoring features

Answer merging

- **Merging, e.g.** *J.F.K.* and *John F. Kennedy*
- **For proper nouns, use automatically generated name dictionaries**
  - large synonym dictionaries that are created by listing all anchor text strings that point to the same Wikipedia page; e.g., *JFK, John F. Kennedy, John Fitzgerald Kennedy, Senator John F. Kennedy, President Kennedy, Jack Kennedy*
- **For common nouns, we can use morphological parsing to merge candidates that are morphological variants.**
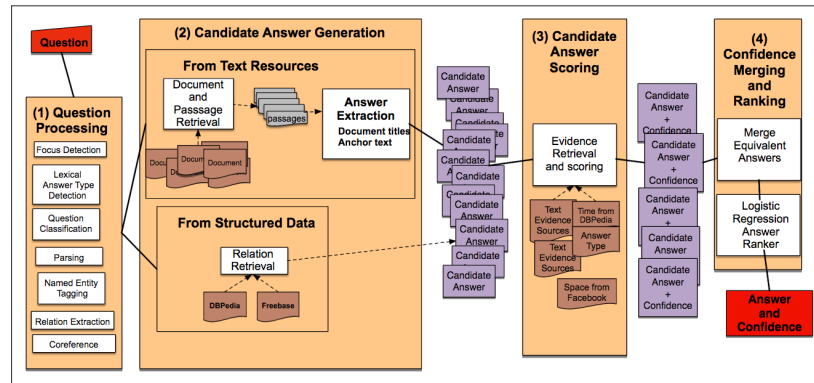
Answer ranking

- **Logistic regression to produce probability that answer is correct**
  - Uses scoring vector
  - Trained with thousands of question/answer pairs
- **Imbalanced data**
  - Many more incorrect answers than correct answers
  - Use *instance weighting*, incorrect answers have weight of 0.5.

The Rest

- **Primary search**
  - Top 250 candidates
- **Soft filtering**
  - A bit mysterious...whittle down to top 100
- **Hypothesis and evidence scoring**
  - "rigorous evaluation process"

Watson's architecture

Figure 28.9 The 4 broad stages of Watson QA: (1) Question Processing, (2) Candidate Answer Generation, (3) Candidate Answer Scoring, and (4) Answer Merging and Confidence Scoring.