NLP

LCC (e.g. Moldovan et al. 2003)

- Uses logic form transformations

- Uses axioms for inference
  - e.g. Lexical chains

Example:
Heavy selling of Standard & Poor's 500-stock index futures in Chicago relentlessly beat stocks downward.

LF:
heavy_JJ(x1) & selling_NN(x1) & of_IN(x1,x6) & Standard_NN(x2) & &_CC(x13,x2,x3) & Poor_NN(x3) &'s_POS(x6,x13) & 500-stock_JJ(x6) & index_NN(x4) & future_NN(x5) & nn_NNC(x6,x4,x5) & in_IN(x1,x8) & Chicago_NN(x8) & relentlessly_RB(e12) & beat_VB(e12,x1,x9) & stocks_NN(x9) & downward_RB(e12)

Q1394: What country did the game of croquet originate in ?
Answer: Croquet is a 15th-century French sport that has largely been dominated by older, wealthier people who play at exclusive clubs.
Lexical chains:
(1) game:n#3 → HYPERNYM → recreation:n#1 → HYPONYM → sport:n#1
(2) originate_in:v#1 → HYPONYM → stem:v#1 → GLOSS → origin:n#1 → GLOSS → be:v#1

QASM (Radev & al. 2001)

- Noisy channel model
- Convert natural language question into query
  - What country is the biggest producer of tungsten?
  - (biggest OR largest) producer tungsten
- Channel operators
  - DELETE, e.g., delete prepositions, stop words
  - REPLACE, e.g., replace a noun phrase with a WordNet expansion
  - DISJUNCT, e.g., replace a noun phrase with a disjunction

Ravinchandran and Hovy 2002

- Characteristics
  - Automatically learn surface patterns
  - Starts with a seed
  - Query Web
  - Find patterns that contain both the question and the answer terms
- Example
  - Mozart was born in 1756
  - <NAME> was born on <BIRTHDATE>

Watson (Ferrucci et al. 2010)

- Jeopardy winner (2011)

- Architecture
  - Uses "DeepQA": a technology that enables computer systems to directly and precisely answer natural language questions over an open and broad range of knowledge
  - 10 racks of IBM Power 750 servers running Linux
  - 16 terabytes of RAM
  - 2,880 processor cores
  - Capable of operating at 80 teraflops.
  - Mostly in Java but also some C++ and Prolog
  - Integrated using UIMA

Ferrucci et al. 2010. Building Watson: An Overview of the DeepQA Project. AI Magazine. Fall 2010. 59-79.
http://www.pcmag.com/article2/0,2817,2380351,00.asp
http://blog.reddit.com/2011/02/ibm-watson-research-team-answers-your.html

Watson (Ferrucci et al. 2000)

- ## Knowledge sources
  - 200 million pages of structured and unstructured content consuming four terabytes of disk storage
  - Including Wikipedia, WordNet, Yago
- ## Betting strategy
  - Buzz if at least 50% certain
- ## Performance
  - Watson answers 66 correct and 9 incorrect.
  - Watson's two day winning streak was $77,147. Ken Jennings ended with $24,000 and Brad Rutter with $21,600.

http://www.quora.com/What-questions-were-asked-in-the-Jeopardy-episode-involving-Watson

http://theswimmingsubmarine.blogspot.com/2011/02/how-ibms-deep-question-answering.html

http://www.theatlantic.com/technology/archive/2011/02/is-it-time-to-welcome-our-new-computer-overlords/71388/

Watson (Ferrucci et al. 2010)

- Question types
  - 2,500 of them
  - 200 of them are very common

- Jeopardy Question Archive
  - http://j-archive.com/

QA Challenges

- Word Sense Disambiguation
- Co-reference Resolution
- Semantic Role Labeling
- Temporal questions
- Categories on Jeopardy