# Question Answering (and Textual Entailment)

Prof. Sameer Singh

CS 295: STATISTICAL NLP

WINTER 2017

March 14, 2017

# Upcoming...

**Homework**
- Homework 4 was due on **last night**
- Lowest grade of the homeworks will be dropped

**Summaries**
- Paper summaries **due tonight**
- Summary 2 graded

TA/Instructor Evaluations are available!

**Project**
- Final report due in a week: **March 20, 2017**
- Instructions coming soon: ACL style, 5 pages (+references)

# Outline

Question Answering

IR-Based QA System

Other Extensions

Textual Entailment

# Outline

Question Answering

IR-Based QA System

Other Extensions

Textual Entailment

# Questions in Modern Systems

Factoid questions
- ◦ Who wrote "The Universal Declaration of Human Rights"?
- ◦ How many calories are there in two slices of apple pie?
- ◦ What is the average age of the onset of autism?
- ◦ Where is Apple Computer based?

Complex (narrative) questions:
- ◦ In children with an acute febrile illness, what is the efficacy of acetaminophen in reducing fever?
- ◦ What do scholars think about Jefferson's position on dealing with pirates?

# Commercial systems: mainly factoid questions

| | |
|---|---|
| Where is the Louvre Museum located? | In Paris, France |
| What's the abbreviation for limited partnership? | L.P. |
| What are the names of Odin's ravens? | Huginn and Muninn |
| What currency is used in China? | The yuan |
| What kind of nuts are used in marzipan? | almonds |
| What instrument does Max Roach play? | drums |
| What is the telephone number for Stanford University? | 650-723-2300 |

# Paradigms for QA

IR-based approaches
- ◦ TREC;  IBM Watson; Google

Knowledge-based and Hybrid approaches
- ◦ IBM Watson; Apple Siri; Wolfram Alpha; True Knowledge Evi

# Many questions can already be answered by web search



**Google** | What are the names of Odin's ravens?

**Search**          About 214,000 results (0.38 seconds)

Everything

Images

Maps

**Huginn and Muninn - Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/Huginn_and_Muninn
The **names** of the **ravens** are sometimes modernly anglicized as Hugin and Munin. In the Poetic Edda, a disguised **Odin** expresses that he fears that they may ...
Attestations - Archaeological record - Theories - See also

# IR-based Question Answering

# IR-based Factoid QA
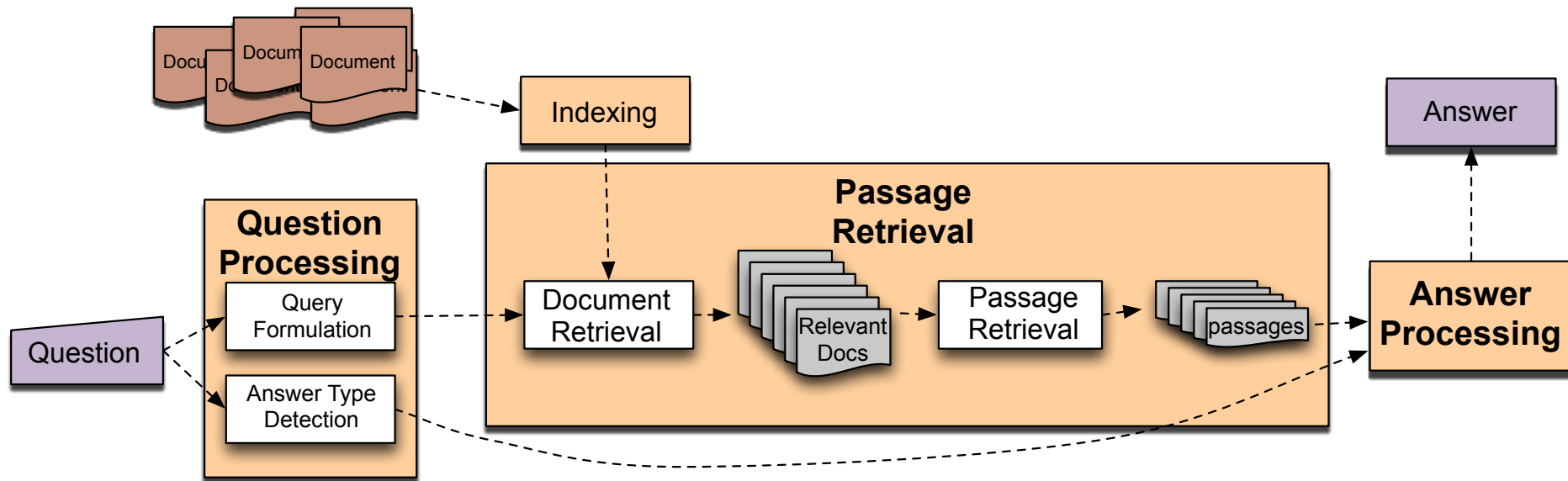
# Knowledge-based QA (Siri)

Build a semantic representation of the query
  ◦ Times, dates, locations, entities, numeric quantities

Map from this semantics to query structured data  or resources
  ◦ Geospatial databases
  ◦ Ontologies (Wikipedia infoboxes, dbPedia, WordNet, Yago)
  ◦ Restaurant review sources and reservation services
  ◦ Scientific databases

# Hybrid approaches (Watson)

Build a shallow semantic representation of the query

Generate answer candidates using IR methods
◦ Augmented with ontologies and semi-structured data

Score each candidate using richer knowledge sources
◦ Geospatial databases
◦ Temporal reasoning
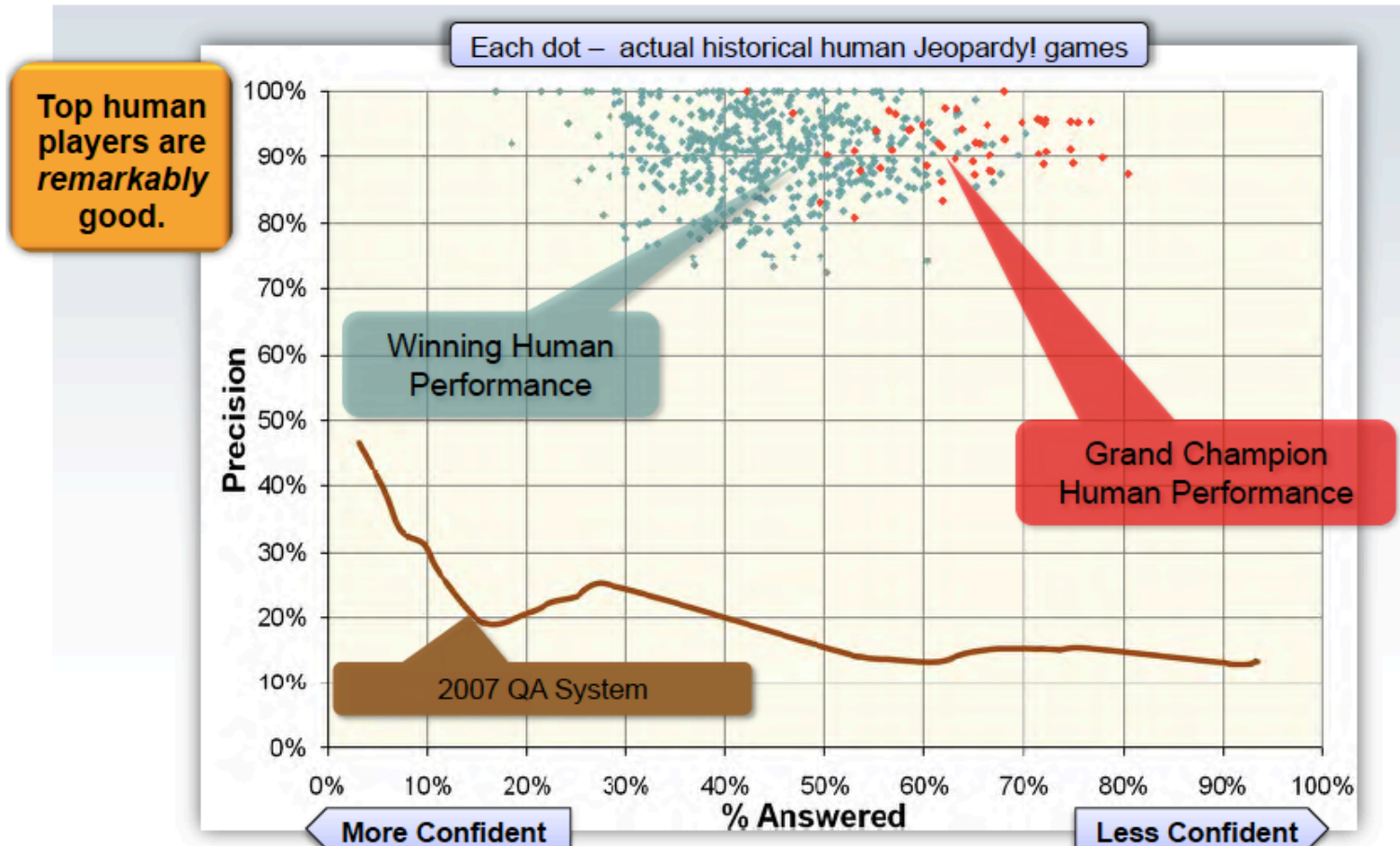◦ Taxonomical classification

# IBM's Watson

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
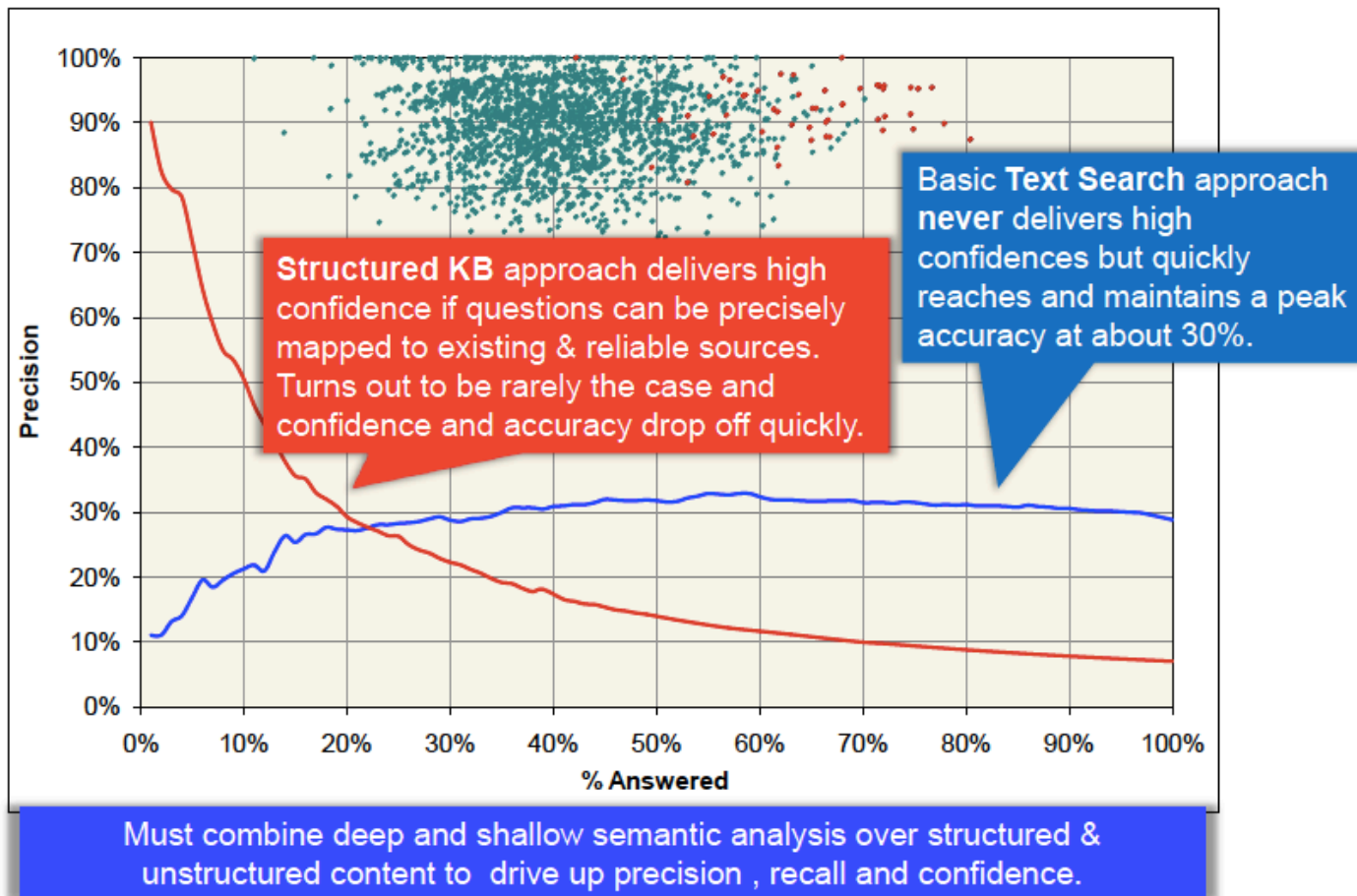INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

→ Bram Stoker

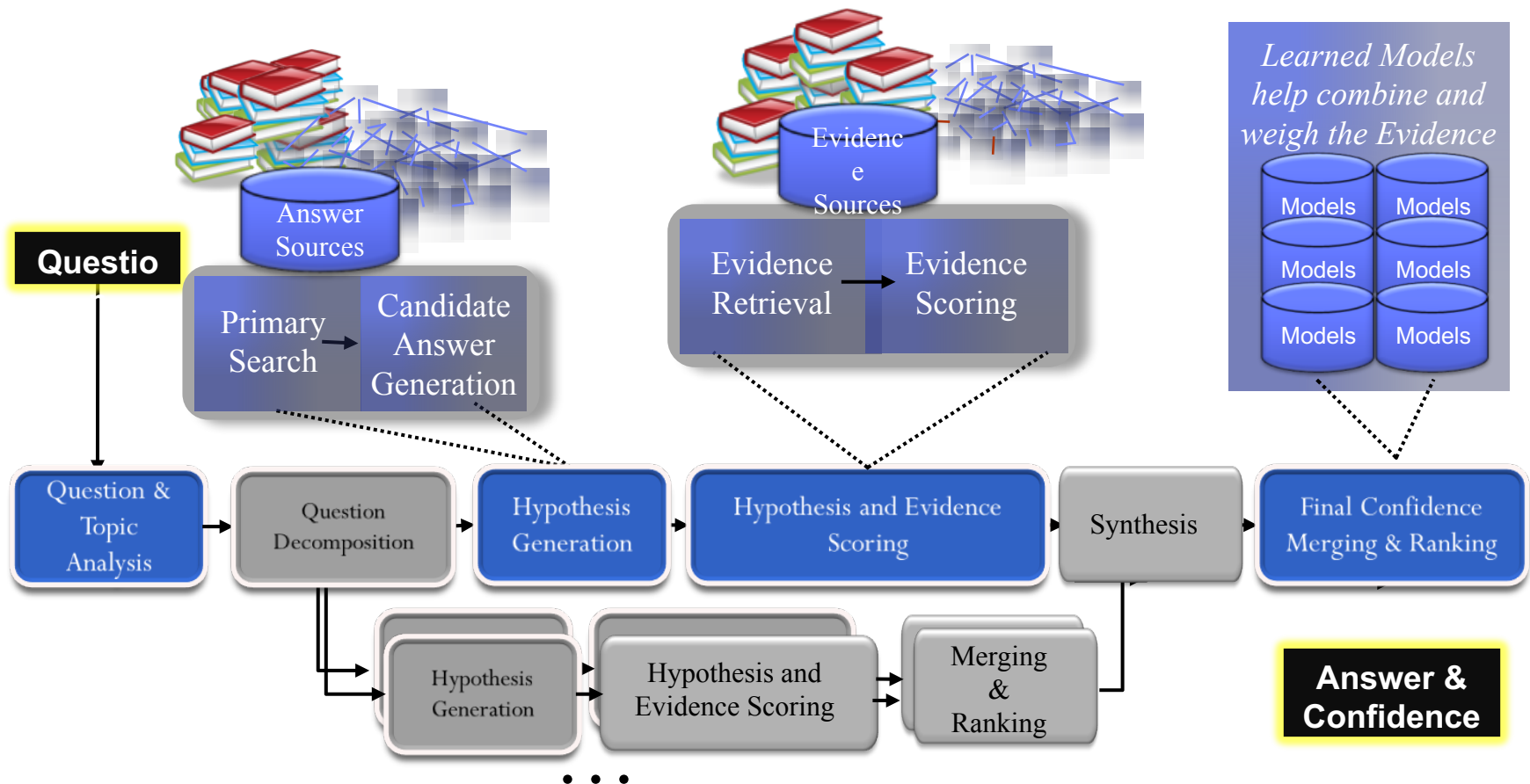Won Jeopardy on February 16, 2011!
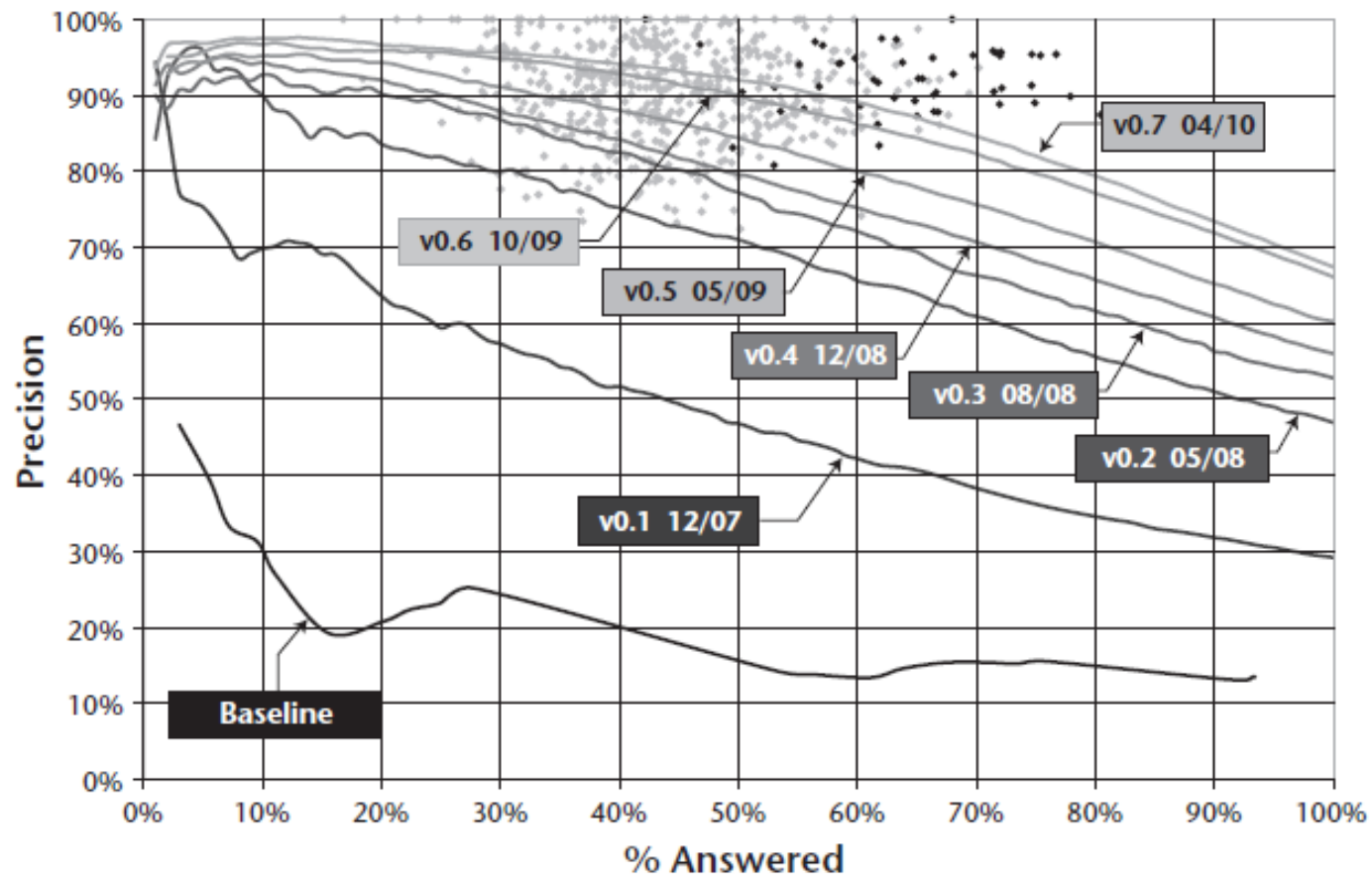
# Motivation for Watson

# Single Source is not Sufficient

# Watson Architecture

# Watson Performance

# Outline

Question Answering

IR-Based QA System

Other Extensions

Textual Entailment

# IR-based Factoid QA
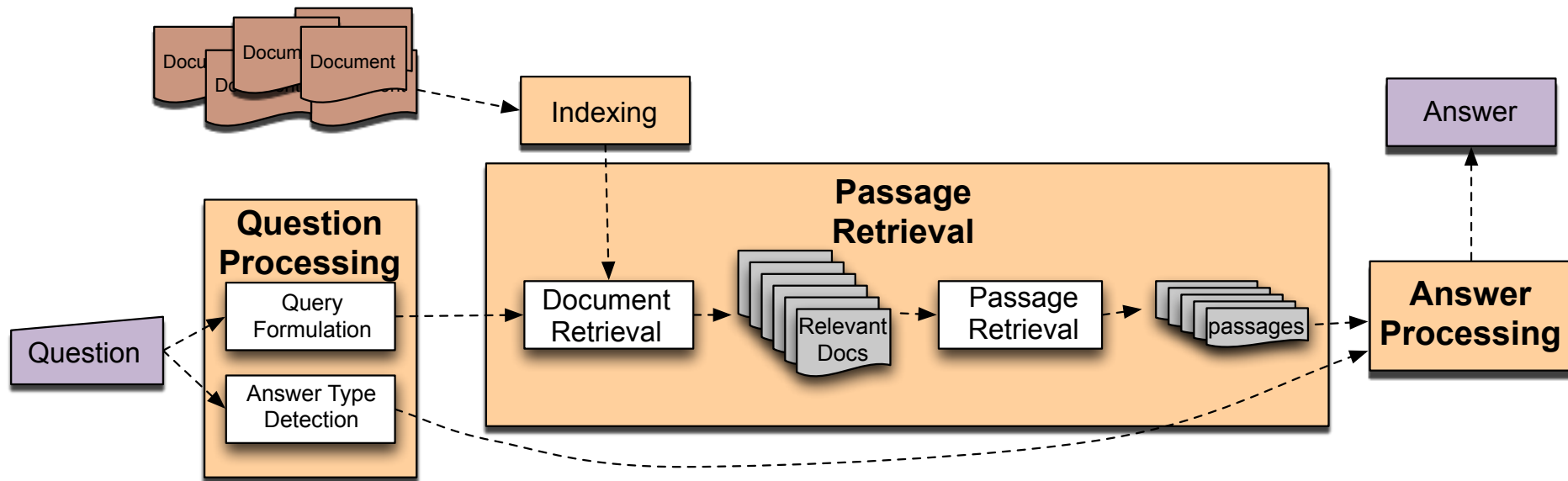
# IR-based Factoid QA

QUESTION PROCESSING
- ◦ Detect question type, answer type, focus, relations
- ◦ Formulate queries to send to a search engine

PASSAGE RETRIEVAL
- ◦ Retrieve ranked documents
- ◦ Break into suitable passages and rerank

ANSWER PROCESSING
- ◦ Extract candidate answers
- ◦ Rank candidates
  - ◦ using evidence from the text and external sources

# Factoid Q/A

# Question Processing

Answer Type Detection
- ◦ Decide the named entity type (person, place) of the answer
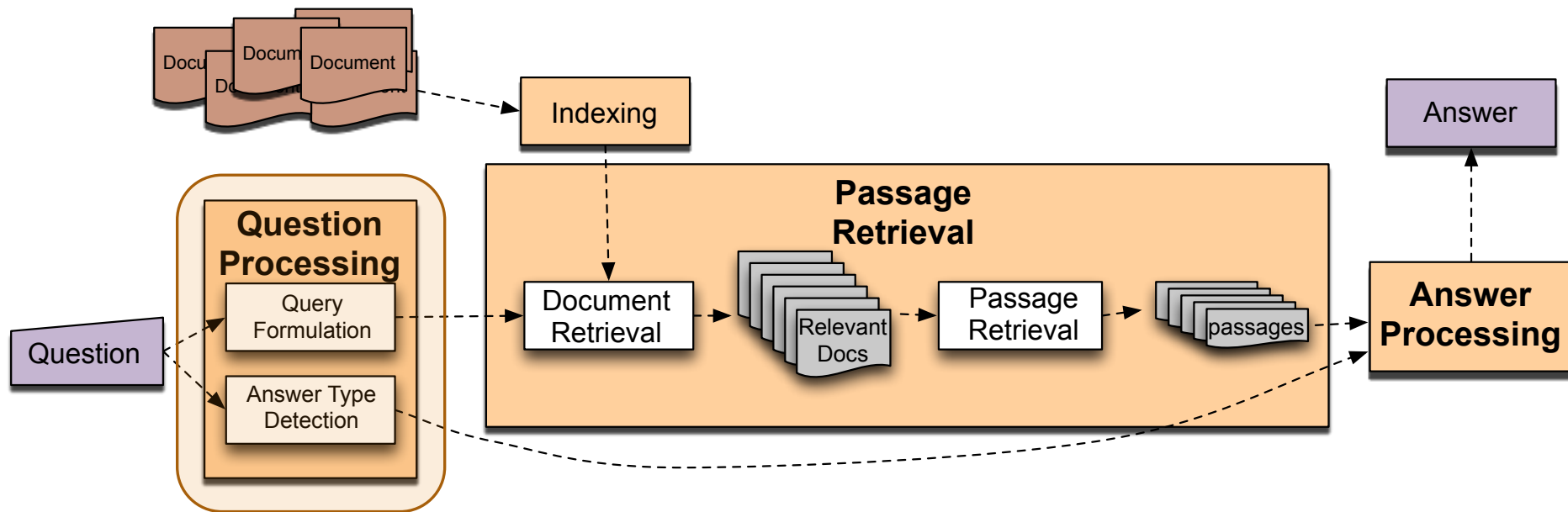
Query Formulation
- ◦ Choose query keywords for the IR system

Question Type classification
- ◦ Is this a definition question, a math question, a list question?

Focus Detection
- ◦ Find the question words that are replaced by the answer

Relation Extraction
- ◦ Find relations between entities in the question

# Question Processing

They're the two states you could be reentering if you're crossing Florida's northern border

Answer Type:  US state

Query:  two states, border, Florida, north

Focus: the two states

Relations:  borders(Florida, ?x, north)

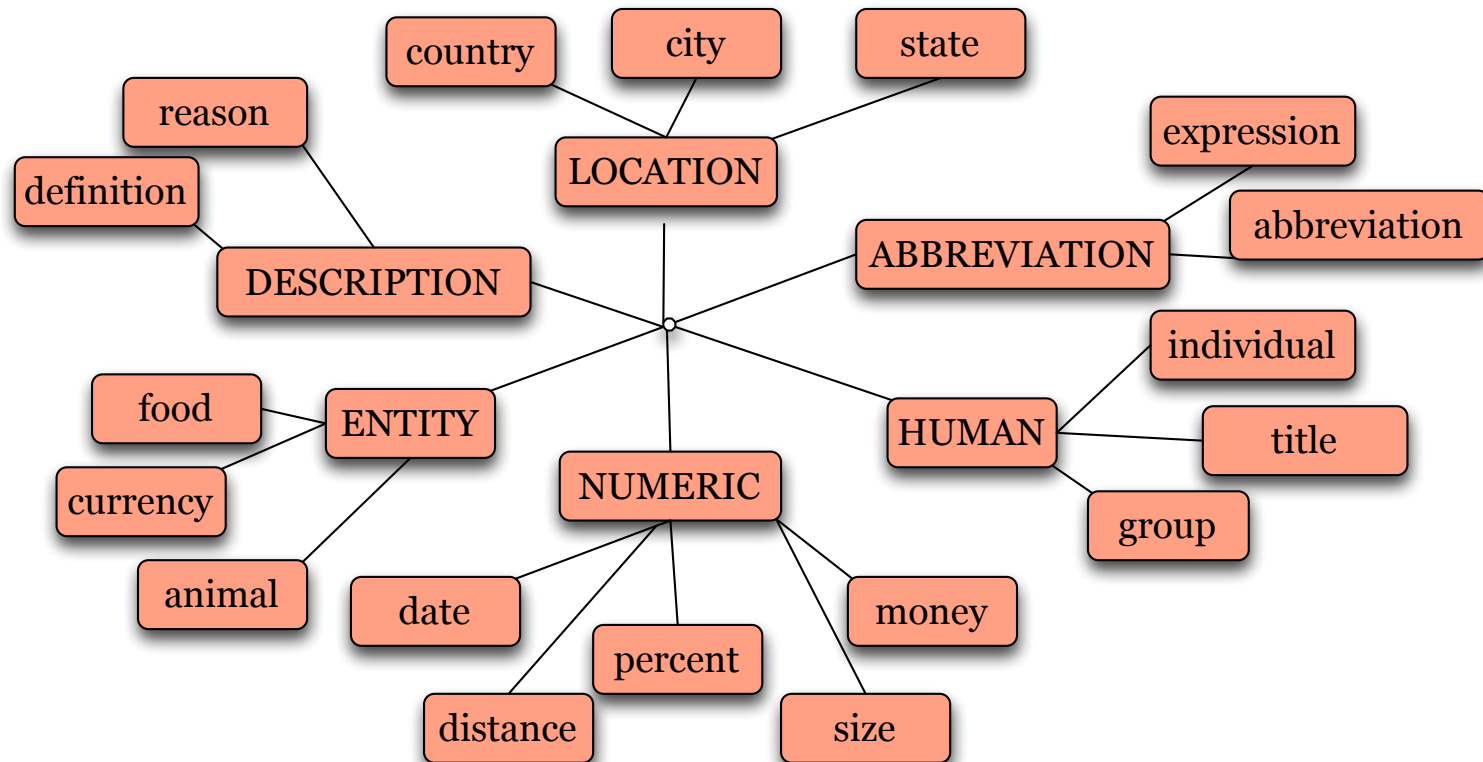# Answer Types: Named Entities

Who founded Virgin Airlines?
PERSON

What Canadian city has the largest population?
CITY

# Part of Answer Type Taxonomy

# Answer Types

| 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? |

# More Answer Types

| 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? |

# Answer types in Watson

2500 answer types in 20,000 Jeopardy question sample
◦ The most frequent 200 answer types cover < 50% of data

The 40 most frequent Jeopardy answer types

he, country, city, man, film, state, she, author, group, here, company, president, capital, star, novel, character, woman, river, island, king, song, part, series, sport, singer, actor, play, team,  show, actress, animal, presidential, composer, musical, nation, book, title, leader, game

Ferrucci et al. Building Watson: An Overview of the DeepQA Project. AI Magazine. 2010

CS 295: STATISTICAL NLP (WINTER 2017)

28

# Answer Type Detection

**Hand-written Rules**

Regular expression-based rules can get some cases:
◦ Who {is|was|are|were} PERSON

Other rules use the question headword:
◦ (the headword of the first noun phrase after the wh-word)
◦ Which city in China has most foreign financial companies?
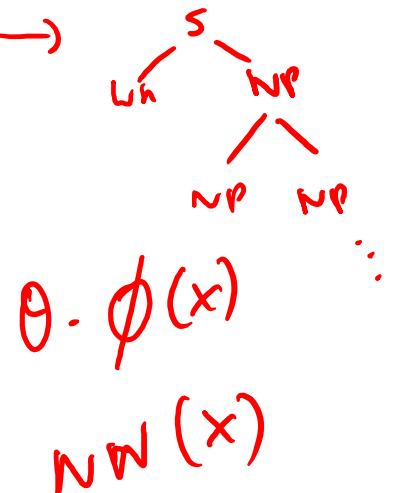◦ What is the state flower of California?

**Machine Learning**

Question words and phrases
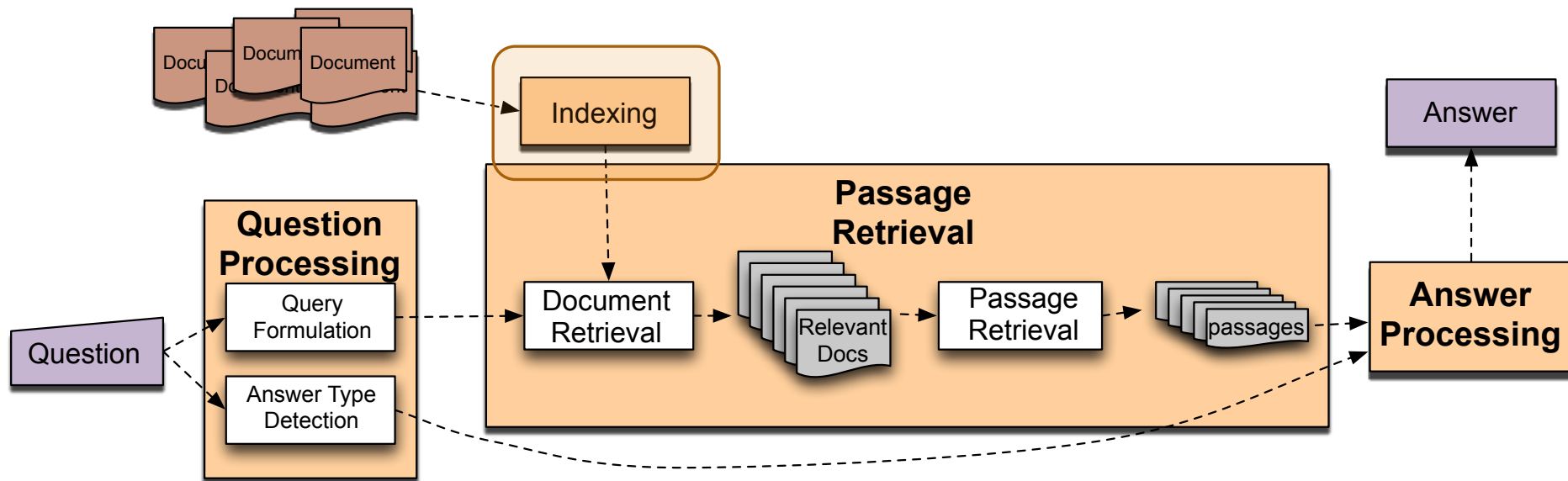Part-of-speech tags
Parse features (headwords)
Named Entities
Semantically related words
+ Rules

$$\theta \cdot \phi(x)$$

$$NN(x)$$

# Factoid Q/A

# Keyword Selection Algorithm

1. Select all non-stop words in quotations

2. Select all NNP words in recognized named entities

3. Select all complex nominals with their adjectival modifiers

4. Select all other complex nominals

5. Select all nouns with their adjectival modifiers

6. Select all other nouns

7. Select all verbs

8. Select all adverbs

9. Select the QFW word (skipped in all previous steps)

10. Select all other words
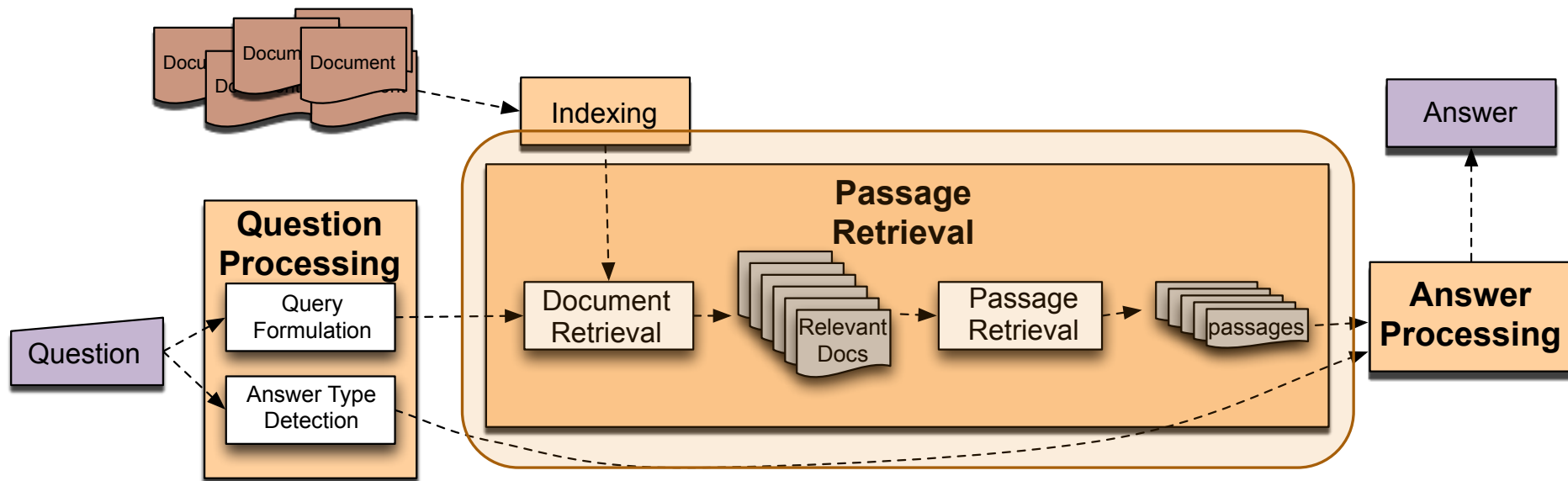
$cos(q, D)$

weighted cosine

# Choosing keywords

Who coined the term "cyberspace" in his novel "Neuromancer"?

| 1 |
| 1 |

| 4 |
| 4 |

| 7 |
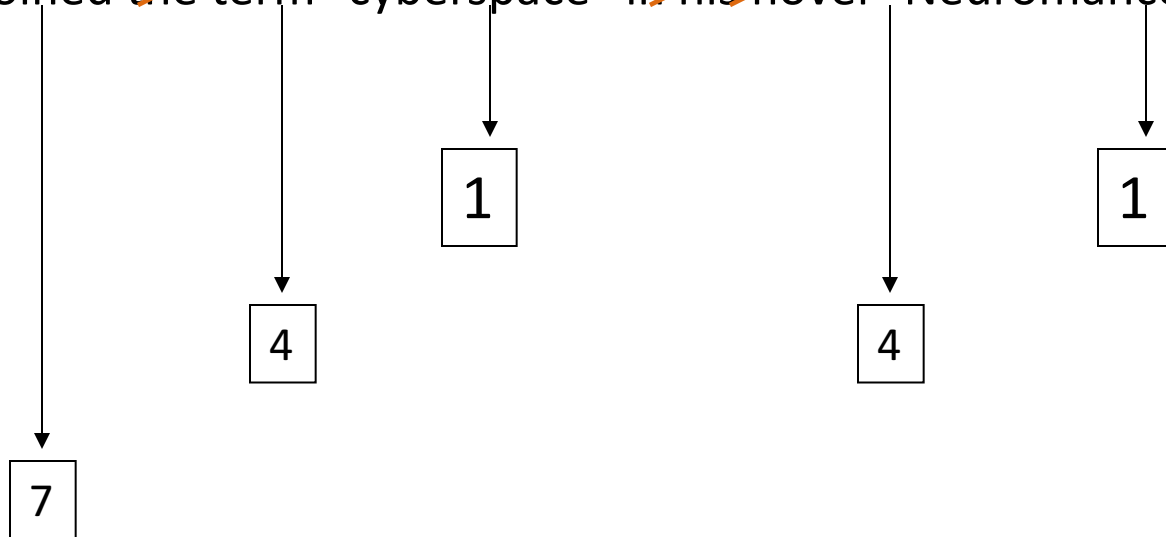
cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7

# Factoid Q/A

# Passage Retrieval

**Step 1**

Retrieve documents using IR
◦ query terms as keywords

**Step 2**

Segment the documents into shorter units
◦ something like paragraphs

**Step 3**

Passage ranking
◦ Use answer type to help rerank passages

# Features for Passage Ranking

Either in rule-based classifiers or with supervised machine learning

Number of Named Entities of the right type in passage
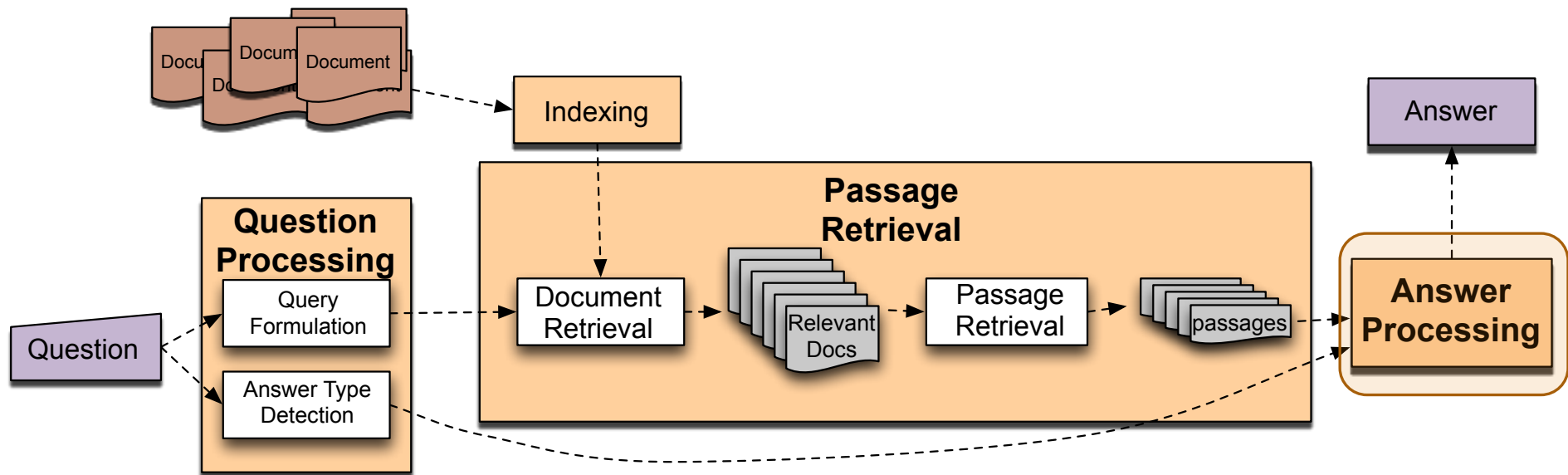
Number of query words in passage

Number of question N-grams also in passage

Proximity of query keywords to each other in passage

Longest sequence of question words

Rank of the document containing passage

# Factoid Q/A

# Answer Extraction

Run an answer-type named-entity tagger on the passages
- Each answer type requires a named-entity tagger that detects it
- If answer type is CITY, tagger has to tag CITY
  - Can be full NER, simple regular expressions, or hybrid

Return the string with the right type:
- Who is the prime minister of India (PERSON)
- Manmohan Singh, Prime Minister of India, had told left leaders that the deal would not be renegotiated.
- How tall is Mt. Everest? (LENGTH)
- The official height of Mount Everest is 29035 feet

# Ranking Candidate Answers

But what if there are multiple candidate answers!

Q: Who was Queen Victoria's second son?

Answer Type:  Person

- Passage:

The Marie biscuit is named after Marie Alexandrovna, the daughter of Czar Alexander II of Russia and wife of Alfred, the second son of Queen Victoria and Prince Albert

# Ranking Candidate Answers

But what if there are multiple candidate answers!

Q: Who was Queen Victoria's second son?

Answer Type:  Person

- Passage:

The Marie biscuit is named after **Marie Alexandrovna**, the daughter of **Czar Alexander II of Russia** and wife of **Alfred**, the second son of **Queen Victoria** and **Prince Albert**

# Features for ML

Answer type match:  Candidate contains a phrase with the correct answer type.

Pattern match: Regular expression pattern matches the candidate.

Question keywords: # of question keywords in the candidate.

Keyword distance: Distance in words between the candidate and query keywords

Novelty factor: A word in the candidate is not in the query.

Apposition features: The candidate is an appositive to question terms

Punctuation location: The candidate is immediately followed by a comma, period, quotation marks, semicolon, or exclamation mark.

Sequences of question terms: The length of the longest sequence of question terms that occurs in the candidate answer.

# Scoring Candidates in Watson

Each candidate answer gets scores from >50 components
- (from unstructured text, semi-structured text, triple stores)
- logical form (parse) match between question and candidate
- passage source reliability
- geospatial location
  - California  is  "southwest of Montana"
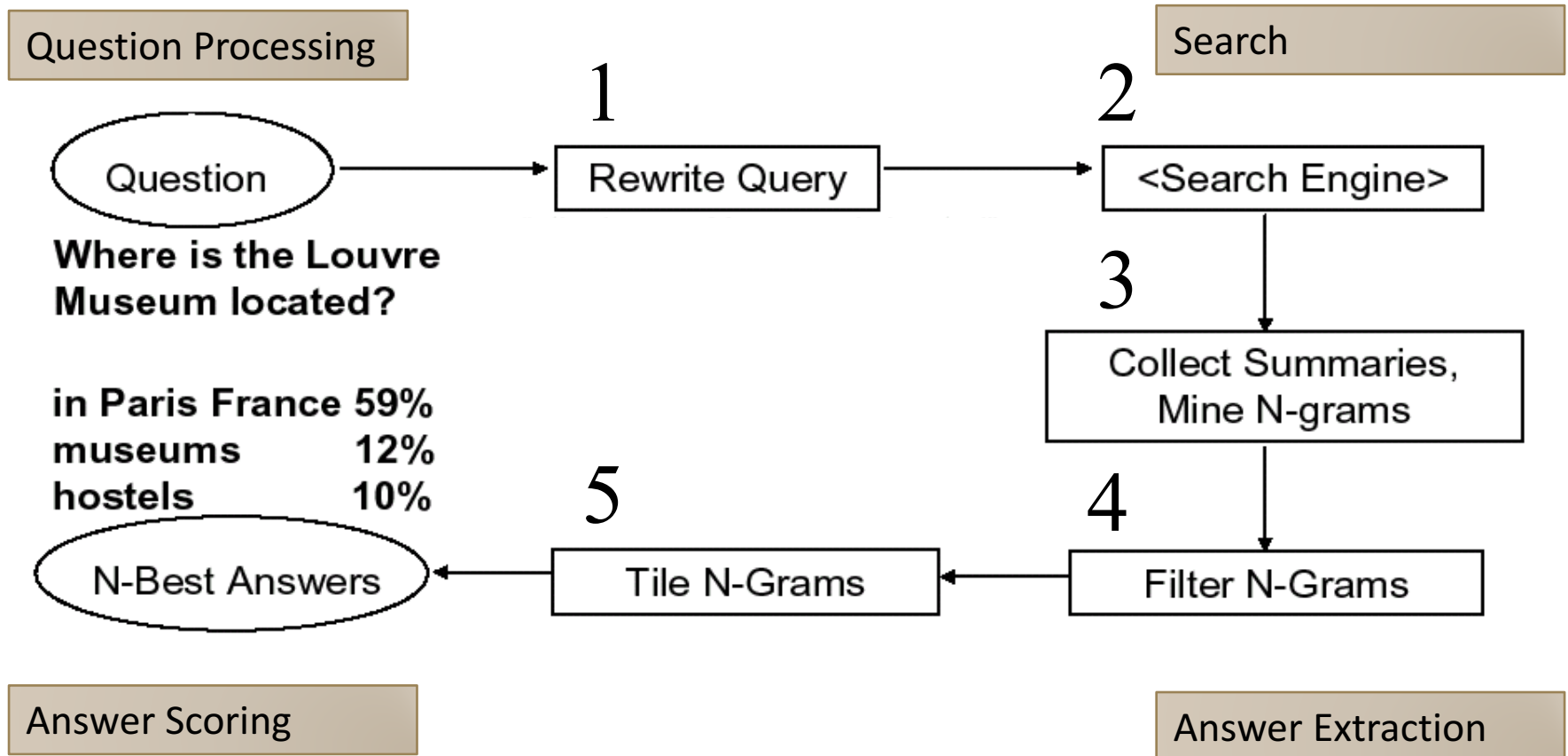- temporal relationships
- taxonomic classification

# Outline

Question Answering

IR-Based QA System

Other Extensions

Textual Entailment

# AskMSR

# Step 1: Rewrite Queries

Intuition: Questions are often syntactically quite close to sentences with the answer
- Where <u>is the Louvre Museum located</u>?
  - <u>The Louvre Museum is located</u> in Paris
- Who <u>created the character of Scrooge</u>?
  - Charles Dickens <u>created the character of Scrooge</u>

"Where <u>is</u> the Louvre Museum located"
- → "<u>is</u> the Louvre Museum located"
- → "the <u>is</u> Louvre Museum located"
- → "the Louvre <u>is</u> Museum located"
- → "the Louvre Museum <u>is</u> located" in Paris
- → "the Louvre Museum located <u>is</u>"

Nonsense, but who cares? It's only a few more queries

# Feedback Loops: FALCON

# Allen AI Science Challenge



Which object is the best conductor of electricity?
(A)  a wax crayon  (B) a plastic spoon
(C) a rubber eraser  (D) an iron nail

# Allen AI Science Challenge



Which object is the best conductor of electricity?
(A)  a wax crayon  (B) a plastic spoon
(C) a rubber eraser  **(D) an iron nail**
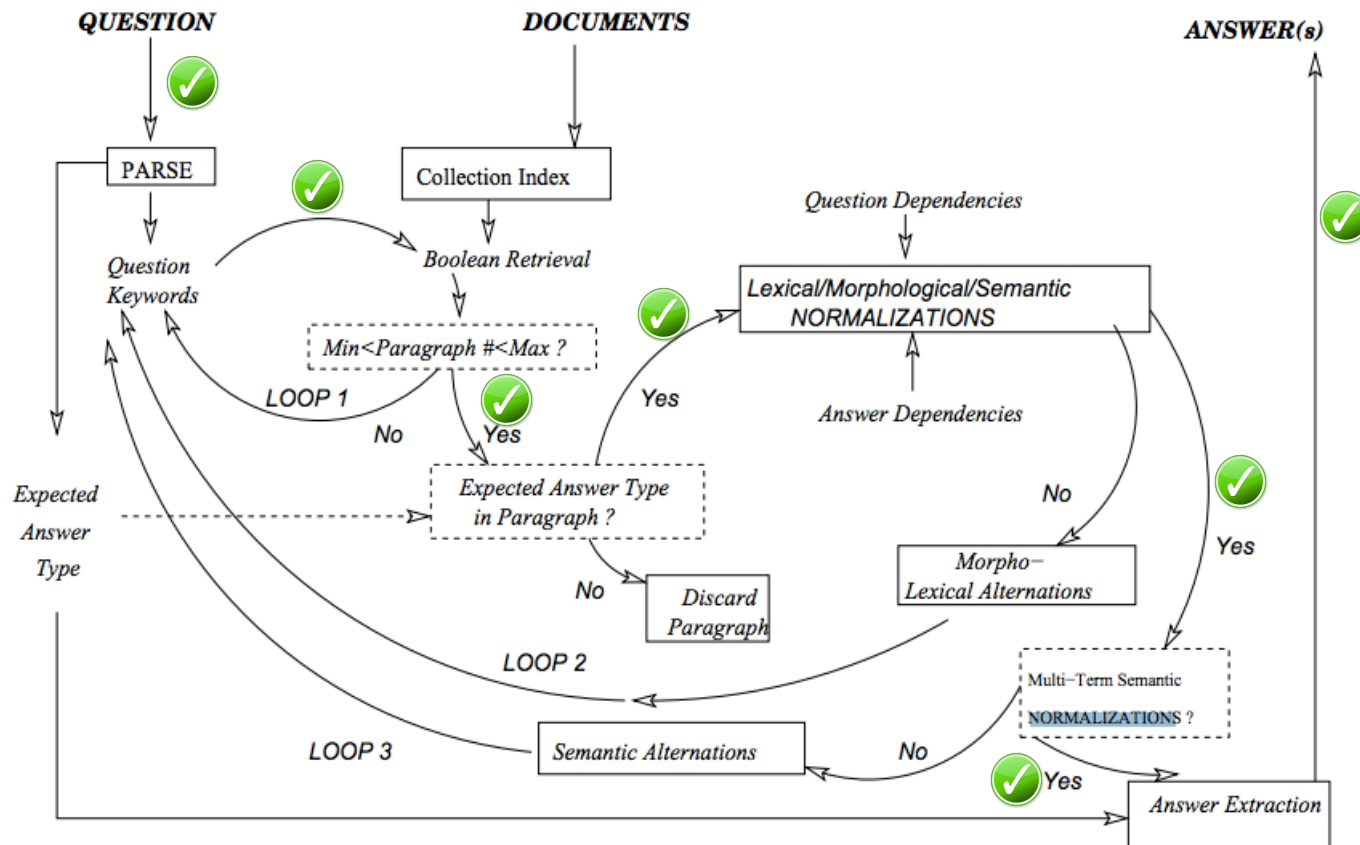
# Allen AI Science Challenge



THE UNIVERSITY OF THE STATE OF NEW YORK

**GRADE 4**

**ELEMENTARY-LEVEL SCIENCE TEST**

**WRITTEN TEST**

**MAY 2004**

Student Name _____

School Name _____

Print your name and the name of your school on the lines above.

The test has two parts. Parts I and II are in this test booklet.

**Part I** contains 30 multiple-choice questions. Record your answers to these questions on the separate answer sheet. Use only a No. 2 pencil on your answer sheet.

**Part II** consists of 11 open-ended questions. Write your answers to Part II in this test booklet.

You will have as much time as you need to answer the questions.

**DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO.**

Copyright 2004
THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
ALBANY, NEW YORK 12234

Fourth graders are planning a roller-skate race. Which surface would be the best for this race?
(A) gravel  (B) sand  (C) blacktop  (D) grass

# Allen AI Science Challenge

Fourth graders are planning a roller-skate race. Which surface would be the best for this race?
(A) gravel  (B) sand  **(C) blacktop**  (D) grass

- Information retrieval methods fail
- Word co-occurrence methods struggle

**Graders** are commonly used in the construction and maintenance of dirt roads and **gravel** roads

Also strong correlations between:
grass ↔ race
gravel ↔ surface

# Allen AI Science Challenge



A student puts two identical plants in the same type and amount of soil. She gives them the same amount of water. She puts one of these plants near a sunny window and the other in a dark room. This experiment tests how the plants respond to
(A) light (B) air (C) water (D) soil

# Allen AI Science Challenge

A student puts two identical plants in the same type and amount of soil. She gives them the same amount of water. She puts one of these plants near a sunny window and the other in a dark room. This experiment tests how the plants respond to
**(A) light** (B) air (C) water (D) soil

**Knowledge** needed (for example):

- near sunny window → receive light
- in a dark room → no light
- test X's response to Y
  → compare X+Y with X+not(Y)

THE UNIVERSITY OF THE STATE OF NEW YORK

**GRADE 4**

**ELEMENTARY-LEVEL SCIENCE TEST**

**WRITTEN TEST**

**MAY 2004**

Student Name _____

School Name _____

Print your name and the name of your school on the lines above.

The test has two parts. Parts I and II are in this test booklet.

**Part I** contains 30 multiple-choice questions. Record your answers to these questions on the separate answer sheet. Use only a No. 2 pencil on your answer sheet.

**Part II** consists of 11 open-ended questions. Write your answers to Part II in this test booklet.

You will have as much time as you need to answer the questions.

**DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO.**

Copyright 2004
THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
ALBANY, NEW YORK 12234

59%

25% random

# Outline

Question Answering

IR-Based QA System

Other Extensions

Textual Entailment

# Natural Language & Meaning
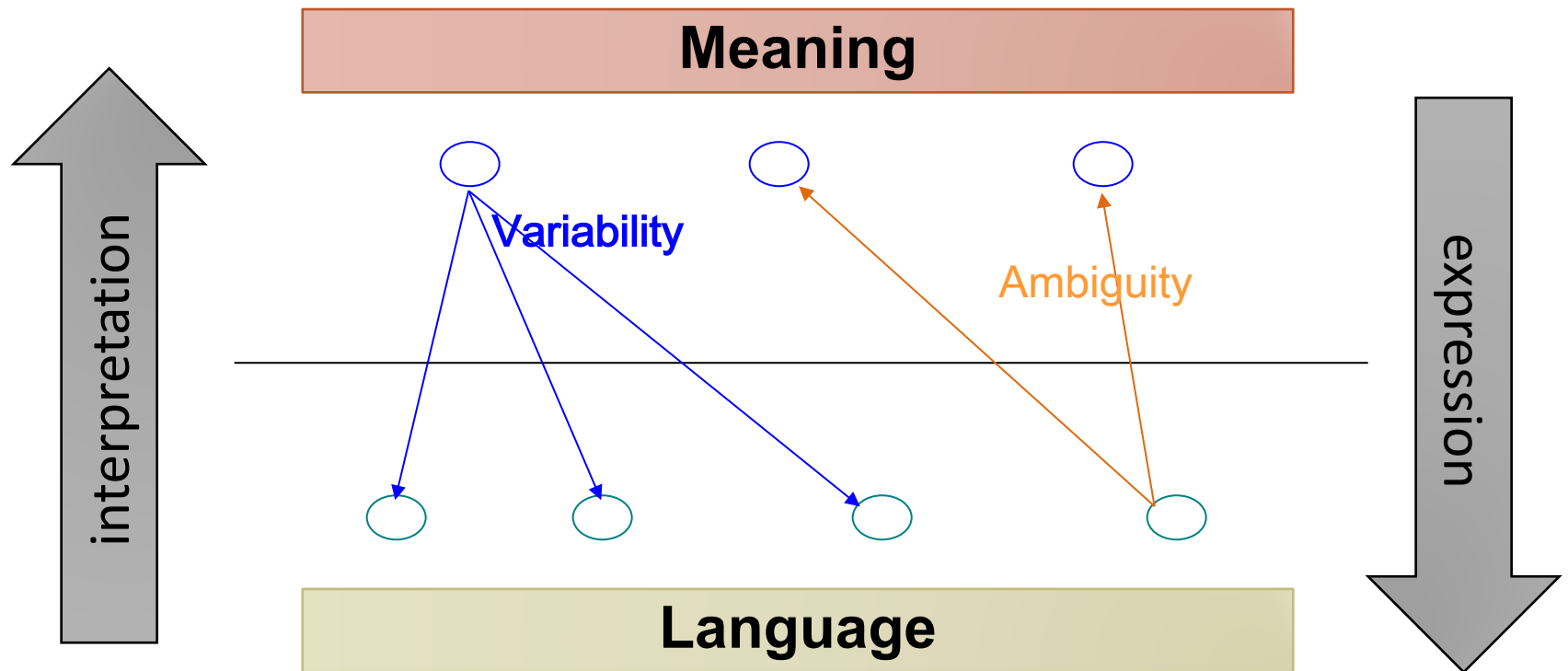
# Inference vs Entailment

- symbolic – logical forms
- statistical – embeddings

Meaning Representation

Natural Language

Inference

Textual Entailment
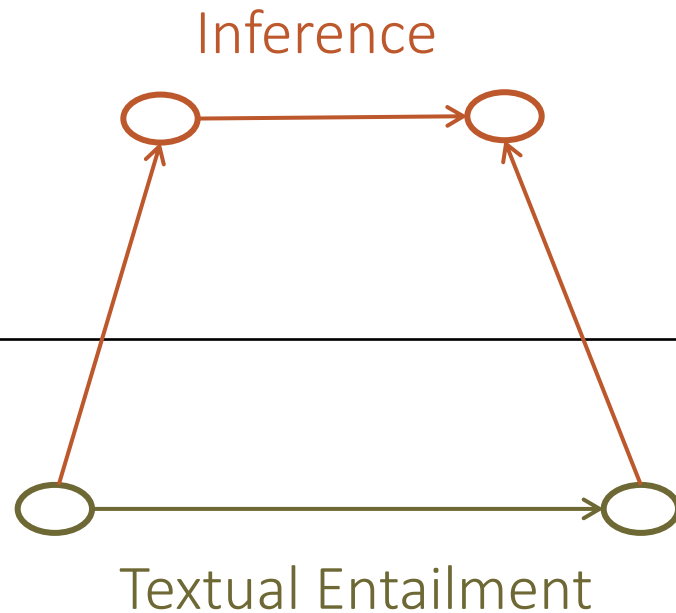
# Textual Entailment

- A directional relation between two text fragments: *Text (t)* and *Hypothesis (h):*

---

***t entails h** ($t \Rightarrow h$)* **if**

**humans** **reading** *t* **will infer that** *h* **is most likely true**

---

- Assuming "common background knowledge" – which is indeed expected from applications

# Example

Text: The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

Hyp 1: BMI acquired an American company. ✓
Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn. ✗
Hyp 3: BMI is an employee-owned concern. ✗

# More Sentence Pairs

1. Some students came to school by car.
   Some students came to school. ✓

2. No students came to school by car.
   Some students came to school. ✗

3. John drove legally.
   John drove. ✓

4. John drove predictably.
   John drove. ✓

5. Legally, John could drive.
   John drove. ✗

# Entailment with Knowledge

**t entails h (t $\Rightarrow$ h) if**

**humans reading t will infer that h is most likely true**

For textual entailment to hold we require:
- *text* AND *knowledge* $\Rightarrow$ *h,* but
- *knowledge* should not entail *h* alone

Systems are not supposed to validate *h*'s truth regardless of *t* (e.g. by searching *h* on the web)

# Example

TEXT:    …While no one accuses Madonna of doing anything illegal in adopting the 4-year-old girl, reportedly named Mercy, there are questions nonetheless about how Madonna is able to navigate Malawi's 18-to-24 month vetting period in just a matter of days or weeks…

HYPOTHESIS:

Madonna is 50 years old.

# Third Label: Contradictions

**_t contradicts h (t $\perp$ h)_ if**

**humans reading _t_ will find the relations/events described by _h_ to be highly unlikely given _t_.**

# Example

**Text:** The purchase of Houston-based LexCorp by BMI for \$2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

**Hyp** 1: BMI acquired an American company. ✓
**Hyp** 2: BMI bought employee-owned LexCorp for \$3.4Bn. ↓
**Hyp** 3: BMI is an employee-owned concern. ✗

# More Examples

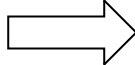| Text | Hypothesis | Label |
|------|-----------|-------|
| John Smith bought a Honda Civic. | John owns a car. | *entails* |
| John Smith bought a Honda Civic. | John sold a car. | *does not entail* |
| John's June 5th purchase of a Honda Civic gave him an alibi. | John did not buy a car. | *contradicts* |
| John Smith bought a Honda Civic for $14,000. | John bought a car for $12,500. | *contradicts* |
| John Smith bought a Honda Civic in Paris. | John bought a car in France. | *entails* |
| John Smith bought a Honda Civic in Paris. | Paris is in France. | *does not entail* |

# Applications

**Question Answering**

*Question*
<u>Who</u> bought Overture?    >>    *Expected answer form*
<u>X</u> bought Overture

**Information Extraction**

Overture's acquisition by <u>Yahoo</u>   *entails* →   <u>Yahoo</u> bought Overture

*text*                                      *hypothesized answer*

**Machine Translation**

He abhorred the men's unctuous ways.

He disliked the men's flattering ways.

**Information Retrieval**

Similarity based on whether document entails the query.