Question Answering

What is Question Answering?

Dan Jurafsky
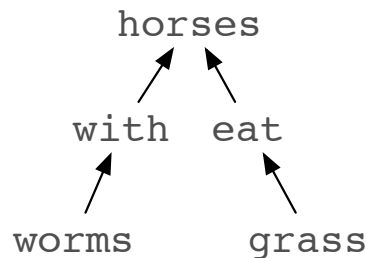
Question Answering

One of the oldest NLP tasks (punched card systems in 1961)
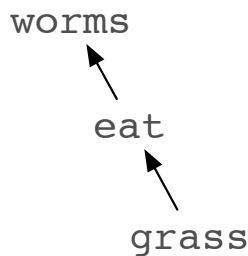
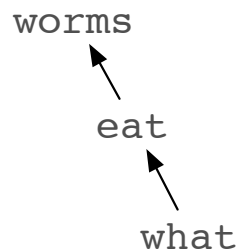Question:

Potential Answers:

```
What do worms eat?        Worms eat grass      Horses with worms eat grass

         worms                    worms                     horses

       eat                      eat                     with    eat

          what                      grass              worms        grass


                          Birds eat worms   Grass is eaten by worms
                                birds                   worms

                              eat                     eat

                                 worms                    grass
```
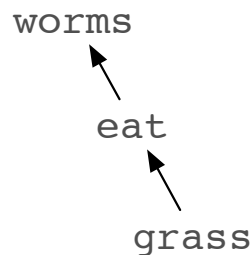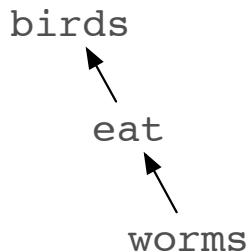
2

Dan Jurafsky

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011!

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

→ Bram Stoker

3

Apple's Siri

4

Dan Jurafsky

WolframAlpha™ computational... knowledge engine

how many calories are in two slices of banana cream pie?

≡ Examples   ⤨ Random

Assuming any type of pie, banana cream | Use pie, banana cream, prepared from recipe or pie, banana cream, no-bake type, prepared from mix instead

Input interpretation:

	amount	2 slices	
pie	type	banana cream	total calories

Average result: Show details

702 Cal (dietary Calories)

5

Types of 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

9

Dan Jurafsky

IR-based Question Answering

Google

Where is the Louvre Museum located?

Search

About 904,000 results (0.30 seconds)

Everything

Images

Maps

Videos

News

Best guess for Louvre Museum Location is **Paris, France**
Mentioned on at least 7 websites including wikipedia.org, answers.com and east-buc.k12.ia.us - Show sources - Feedback

Musée du **Louvre** - Wikipedia, the free encyclopedia
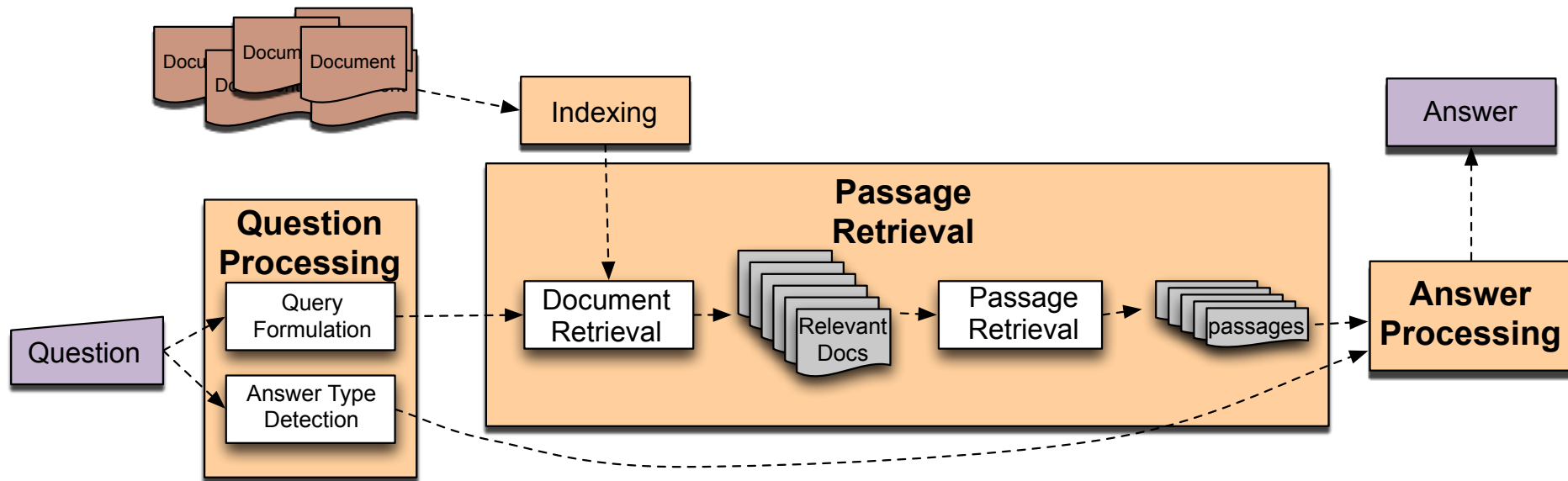en.wikipedia.org/wiki/Musée_du_**Louvre**
Musée du **Louvre** is **located** in Paris. **Location** within Paris. Established, 1793. **Location**,
Palais Royal, Musée du **Louvre**, 75001 Paris, France. Type, Art **museum** ...
Louvre Palace - List of works in the Louvre - Category:Musée du Louvre

Dan Jurafsky

IR-based Factoid QA

11

Dan Jurafsky

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

Dan Jurafsky

Knowledge-based approaches (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

13

Hybrid approaches (IBM 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

Question Answering

What is Question Answering?