CS 6120/CS 4120: Natural Language Processing

Instructor: Prof. Lu Wang
College of Computer and Information Science
Northeastern University
Webpage: www.ccs.neu.edu/home/luwang
Question Answering
Questions in Search

who is
who is the next bachelorette
who is my congressman
who is my representative
who is kelly clarkson married to
who is snoko
who is lala kent's boyfriend
who is
who is the richest person in the world
who is tom steyer
who is safaree

what is
what is my ip
what is blockchain
what is renal disease
what is bitcoin
what is vero
what is net neutrality
what is daca
what is my ip address
what is dilly dilly
what is lupus

Report inappropriate predictions
Questions in Search

Where is the Louvre Museum located?

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
Question Answering (Some Background)

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

Simmons, Klein, McConlogue. 1964. Indexing and Dependency Logic for Answering English Questions. American Documentation 15:30, 196-204

Question: What do worms eat?

Potential Answers:

- Worms eat grass
- Horses with worms eat grass
- Birds eat worms
- Grass is eaten by worms
What do worms eat?

Worms eat grass

Horses with worms eat grass

Birds eat worms

Grass is eaten by worms

Question Answering

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

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

Question:

What do worms eat?

Potential Answers:

Worms eat grass
Horses with worms eat grass

Birds eat worms
Grass is eaten by worms
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
Apple’s Siri

“Do I need an umbrella tomorrow in San Francisco?”

Yes, San Francisco should get rain tomorrow:

Weekly Forecast

<table>
<thead>
<tr>
<th>TUES</th>
<th>WED</th>
<th>THU</th>
<th>FRI</th>
<th>SAT</th>
<th>SUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>14°</td>
<td>16°</td>
<td>17°</td>
<td>17°</td>
<td>18°</td>
<td>18°</td>
</tr>
<tr>
<td>9°</td>
<td>7°</td>
<td>6°</td>
<td>6°</td>
<td>7°</td>
<td>8°</td>
</tr>
</tbody>
</table>
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?
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

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where is the Louvre Museum located?</td>
<td>In Paris, France</td>
</tr>
<tr>
<td>What’s the abbreviation for limited partnership?</td>
<td>L.P.</td>
</tr>
<tr>
<td>What are the names of Odin’s ravens?</td>
<td>Huginn and Muninn</td>
</tr>
<tr>
<td>What currency is used in China?</td>
<td>The yuan</td>
</tr>
<tr>
<td>What kind of nuts are used in marzipan?</td>
<td>almonds</td>
</tr>
<tr>
<td>What instrument does Max Roach play?</td>
<td>drums</td>
</tr>
</tbody>
</table>
Paradigms for Factoid QA

• Information Retrieval (IR)-based approaches
  • IBM Watson (some parts); Google

• Knowledge-based and Hybrid approaches
  • IBM Watson; Apple Siri; Wolfram Alpha

• Built upon the above two:
  • Data-driven, neural network-based approaches (more recent)
Information Retrieval (IR)-based QA

- Factoid QA pipeline
- Factoid QA evaluation
- Common Knowledge used in QA
- Recent QA tasks
IR-based Factoid QA

Question Processing
- Query Formulation
- Answer Type Detection

Passage Retrieval
- Document Retrieval
- Relevant Docs
- Passage Retrieval

Answer Processing
- Answer
IR-based Factoid QA

- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
    - “Who is the president of US?” -> person
  - Formulate queries to send to a search engine
    - “president of United States”

- **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
Question Processing:
Things to extract from the question

• 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 (if there are more than one entities)
  • Find relations between entities in the question
Jeopardy!: They’re the two states you could be reentering if you’re crossing Florida’s northern border
You should answer: what are the states of Georgia and Alabama?

• Answer Type: US state
• Query Formulation: two states, border, Florida, north
• Focus: the two states
• Relations: borders(Florida, ?x, north)
IR-based Factoid QA

• QUESTION PROCESSING
  • Detect question type, answer type, focus, relations
    • "Who is the president of US?" -> person
  • Formulate queries to send to a search engine
    • "president of United States"

• 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
Answer Type Detection: Named Entities

• *Who founded Virgin Airlines?*
Answer Type Detection: Named Entities

• Who founded Virgin Airlines?
  • PERSON

• What Canadian city has the largest population?
Answer Type Detection: Named Entities

• *Who founded Virgin Airlines?*
  • PERSON

• *What Canadian city has the largest population?*
  • CITY
Answer Type Taxonomy

Xin Li, Dan Roth. 2002. Learning Question Classifiers. COLING'02

• 6 coarse classes
  • ABBREVIATION, ENTITY, DESCRIPTION, HUMAN, LOCATION, NUMERIC

• 50 finer classes
  • LOCATION: city, country, mountain...
  • HUMAN: group, individual, title, description...
  • ENTITY: animal, body, color, currency...
Part of Li & Roth’s Answer Type Taxonomy

- LOCATION
  - country
  - city
  - state

- DESCRIPTION
  - reason
  - definition
  - food
  - currency
  - animal
  - date
  - percent
  - distance

- NUMERIC
  - money
  - size

- ABBREVIATION
  - abbreviation
  - individual
  - title
  - group

- HUMAN
  - expression

# Answer Types

<table>
<thead>
<tr>
<th>ENTITY</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>What are the names of Odin’s ravens?</td>
</tr>
<tr>
<td>body</td>
<td>What part of your body contains the corpus callosum?</td>
</tr>
<tr>
<td>color</td>
<td>What colors make up a rainbow?</td>
</tr>
<tr>
<td>creative</td>
<td>In what book can I find the story of Aladdin?</td>
</tr>
<tr>
<td>currency</td>
<td>What currency is used in China?</td>
</tr>
<tr>
<td>disease/medicine</td>
<td>What does Salk vaccine prevent?</td>
</tr>
<tr>
<td>event</td>
<td>What war involved the battle of Chapultepec?</td>
</tr>
<tr>
<td>food</td>
<td>What kind of nuts are used in marzipan?</td>
</tr>
<tr>
<td>instrument</td>
<td>What instrument does Max Roach play?</td>
</tr>
<tr>
<td>lang</td>
<td>What’s the official language of Algeria?</td>
</tr>
<tr>
<td>letter</td>
<td>What letter appears on the cold-water tap in Spain?</td>
</tr>
<tr>
<td>other</td>
<td>What is the name of King Arthur’s sword?</td>
</tr>
<tr>
<td>plant</td>
<td>What are some fragrant white climbing roses?</td>
</tr>
<tr>
<td>product</td>
<td>What is the fastest computer?</td>
</tr>
<tr>
<td>religion</td>
<td>What religion has the most members?</td>
</tr>
<tr>
<td>sport</td>
<td>What was the name of the ball game played by the Mayans?</td>
</tr>
<tr>
<td>substance</td>
<td>What fuel do airplanes use?</td>
</tr>
<tr>
<td>symbol</td>
<td>What is the chemical symbol for nitrogen?</td>
</tr>
<tr>
<td>technique</td>
<td>What is the best way to remove wallpaper?</td>
</tr>
<tr>
<td>term</td>
<td>How do you say “Grandma” in Irish?</td>
</tr>
<tr>
<td>vehicle</td>
<td>What was the name of Captain Bligh’s ship?</td>
</tr>
<tr>
<td>word</td>
<td>What’s the singular of dice?</td>
</tr>
</tbody>
</table>
# More Answer Types

<table>
<thead>
<tr>
<th><strong>HUMAN</strong></th>
<th><strong>LOCATION</strong></th>
<th><strong>NUMERIC</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>description</td>
<td>Who was Confucius?</td>
<td></td>
</tr>
<tr>
<td>group</td>
<td>What are the major companies that are part of Dow Jones?</td>
<td>About how many soldiers died in World War II?</td>
</tr>
<tr>
<td>ind</td>
<td>Who was the first Russian astronaut to do a spacewalk?</td>
<td>What is the date of Boxing Day?</td>
</tr>
<tr>
<td>title</td>
<td>What was Queen Victoria’s title regarding India?</td>
<td>Where does Shanghai rank among world cities in population?</td>
</tr>
<tr>
<td>city</td>
<td>What’s the oldest capital city in the Americas?</td>
<td></td>
</tr>
<tr>
<td>country</td>
<td>What country borders the most others?</td>
<td>How long was Mao’s 1930s Long March?</td>
</tr>
<tr>
<td>mountain</td>
<td>What is the highest peak in Africa?</td>
<td>How much did a McDonald’s hamburger cost in 1963?</td>
</tr>
<tr>
<td>other</td>
<td>What river runs through Liverpool?</td>
<td></td>
</tr>
<tr>
<td>state</td>
<td>What states do not have state income tax?</td>
<td></td>
</tr>
<tr>
<td>code</td>
<td>What is the telephone number for the University of Colorado?</td>
<td></td>
</tr>
<tr>
<td>count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>date</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>money</td>
<td></td>
<td></td>
</tr>
<tr>
<td>order</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>period</td>
<td></td>
<td></td>
</tr>
<tr>
<td>percent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>temp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Answer types in Jeopardy


• 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
  country, city, man, film, state, 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
Answer Type Detection

• Hand-written rules
• Machine Learning
Answer Type Detection

• Regular expression-based rules can get some cases:
  • Who {is|was|are|were} PERSON
  • PERSON (YEAR – YEAR)

• Other rules use the question headword:
  (the headword of the first noun phrase after the wh-word)

• Which city in China has the largest number of foreign financial companies?
• What is the state flower of California?
Answer Type Detection

• Most often, we treat the problem as machine learning classification
  • Define a taxonomy of question types
  • Annotate training data for each question type
  • Train classifiers for each question class using a rich set of features.
    • features include those hand-written rules!
Features for Answer Type Detection

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

Which city in China has the largest number of foreign financial companies?
What is the state flower of California?
Query Formulation

• **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
    - "Who is the president of US?" -> person
  - Formulate queries to send to a search engine
    - “president of United States”

• **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
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 question focus word (skipped in all previous steps)
10. Select all other words
Choosing keywords from the query

Who coined the term “cyberspace” in his novel “Neuromancer”?

cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7
IR-based Factoid QA

• QUESTION PROCESSING
  • Detect question type, answer type, focus, relations
    • ”Who is the president of US?” -> person
  • Formulate queries to send to a search engine
    • “president of United States”

• 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
Passage Retrieval

• Step 1: IR engine retrieves documents using query terms
• Step 2: Segment the documents into shorter units
  • E.g. paragraphs or consecutive sentences
• 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
Passage Retrieval as Query-focused Summarization

Which country has the largest part of the Amazon rain forest?

[The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists from around the world.] [“That’s some of the most encouraging news about the Amazon rain forest in recent years,” said Thomas Lovejoy, an Amazon specialist.] [“It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon.”]

[Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil.]
Passage Retrieval as Query-focused Summarization

• Decide on a summary length (10% of document length).
• Use standard ad-hoc retrieval algorithm to retrieve top k documents.
• Treat each sentence/paragraph in top N documents as a document itself.
  • Use standard document similarity equations to assign a similarity score to the sentence/paragraph.
• Return highest-scoring sentences/paragraphs as the summary, subject to the length constraint.
Passage Retrieval as Query-focused Summarization

Which country has the largest part of the Amazon rain forest?

[The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists from around the world.] [“That’s some of the most encouraging news about the Amazon rain forest in recent years,” said Thomas Lovejoy, an Amazon specialist.] [“It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon.”]

[Sixty percent of the Amazon, the world’s largest tropical rain forest, lies in Brazil.]
IR-based Factoid QA

• QUESTION PROCESSING
  • Detect question type, answer type, focus, relations
    • “Who is the president of US?” -> person
  • Formulate queries to send to a search engine
    • “president of United States”

• 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
IR-based Factoid QA

- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
    - "Who is the president of US?" -> person
  - Formulate queries to send to a search engine
    - "president of United States"

- **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
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
Adding Analysis Patterns

• “Who is Elvis?”
  • Question type: “who”
  • Named-entity tagging: “Who is <person-name> Elvis</person-name>”
  • Analysis pattern: if question type = “who” and question contains <person-name> then

• Desired answer probably is a description

• Likely answer extraction patterns
  • “Elvis, the X”, e.g., “Elvis, the *king of rock and roll!*”
  • “the X Elvis”, e.g., “the *legendary entertainer* Elvis”
IR-based Factoid QA

• QUESTION PROCESSING
  • Detect question type, answer type, focus, relations
    • "Who is the president of US?" -> person
  • Formulate queries to send to a search engine
    • "president of United States"

• 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
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.
Use machine learning:
Features for ranking candidate answers

**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.
Candidate Answer scoring in IBM 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
Information Retrieval (IR)-based QA

• Factoid QA pipeline
• Factoid QA evaluation
• Common Knowledge used in QA
• Recent QA tasks
Common Evaluation Metrics

1. **Accuracy** (does answer match gold-labeled answer?)

2. **Mean Reciprocal Rank**
   - For each query return a ranked list of M candidate answers.
   - Query score is 1/Rank of the first correct answer
     - If first answer is correct: 1
     - else if second answer is correct: $\frac{1}{2}$
     - else if third answer is correct: $\frac{1}{3}$, etc.
     - Score is 0 if none of the M answers are correct
   - Take the mean over all N queries

\[
MRR = \frac{\sum_{i=1}^{N} \frac{1}{\text{rank}_i}}{N}
\]
Information Retrieval (IR)-based QA

- Factoid QA pipeline
- Factoid QA evaluation
- Common Knowledge used in QA
- Recent QA tasks
Knowledge in QA

• What are other types of knowledge useful for a QA system?
  • Relations
  • Temporal information
  • Dialogue context
Relation Extraction

• Answers: Databases of Relations
  • born-in(“Emma Goldman”, “June 27 1869”)
  • author-of(“Cao Xue Qin”, “Dream of the Red Chamber”)
  • Draw from Wikipedia infoboxes, DBpedia, FreeBase, etc.

• Questions: Extracting Relations in Questions
  Whose granddaughter starred in E.T.?
  (acted-in ?x “E.T.”)
  (granddaughter-of ?x ?y)
Temporal Reasoning

• Relation databases
  • (and obituaries, biographical dictionaries, etc.)

• IBM Watson
  “In 1594 he took a job as a tax collector in Andalusia”

Candidates:
  • Thoreau is a bad answer (born in 1817)
  • Cervantes is possible (was alive in 1594)
Context and Conversation in Virtual Assistants like Siri

• Coreference helps resolve ambiguities
  U: “Book a table at Il Fornaio at 7:00 with my mom”
  U: “Also send her an email reminder”

• Clarification questions:
  U: “Chicago pizza”
  S: “Did you mean pizza restaurants in Chicago or Chicago-style pizza?”
Limitations of Factoid Q/A

• Question must query a specific fact that is explicitly stated somewhere in the document corpus.

• Does not allow aggregating or accumulating information across multiple information sources.

• Does not require “deep compositional” semantics, nor inferential reasoning to generate answer.
Information Retrieval (IR)-based QA

• Factoid QA pipeline
• Factoid QA evaluation
• Common Knowledge used in QA
• Recent QA tasks
What are recent tasks for QA?

• Reading comprehension (machine reading)

• Visual Question Answering
Reading Comprehension Q/A

• Answer questions that test comprehension of a specific document.
• Use standardized tests of reading comprehension to evaluate performance (Hirschman et al. 1999; Rilo & Thelen, 2000; Ng et al. 2000; Charniak et al. 2000).
Sample Reading Comprehension Test

School Kids: Clean Up Creek!

(EVERETT, WASHINGTON, June, 1988) – It has taken five years of hard work. A3
But a group of children have cleaned up Pigeon Creek. A1, A2
In 1983, the creek was so dirty that the fish had all died. A group of kids at Jackson A3
School decided to do something about it. It became a project for their whole class.
First they cleaned out the bottom of the creek, called the creek bed. There were many A4
weeds. They also found lot of trash and litter. Then the students raised fish eggs in a
tank of water called an aquarium. When the eggs hatched, they emptied the new fish
into the creek. They wanted others to stop making the creek dirty. So they made signs A5
that said "Don't Dump". They put the signs near the creek. They also made little books,
called booklets. The booklets told people how to keep the creek clean. Now, the water
is clean and new fish are being born in the creek. Pigeon Creek is a sign of the good A2
things young people can do.

Q1: Who cleaned up the creek?  Q2: What was the name of the creek?  Q3: When did the work begin?
Q4: Where is the bed of a creek?  Q5: Why did the kids make booklets?
Large Scale Reading Comprehension Data

- DeepMind’s large-scale data for reading comprehension Q/A (Hermann et al., 2015).
  - News articles used as source documents.
  - Questions constructed automatically from article summary sentences.

<table>
<thead>
<tr>
<th></th>
<th>CNN</th>
<th></th>
<th></th>
<th>Daily Mail</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>train</td>
<td>valid</td>
<td>test</td>
<td>train</td>
<td>valid</td>
</tr>
<tr>
<td># months</td>
<td></td>
<td>95</td>
<td>1</td>
<td>1</td>
<td>56</td>
<td>1</td>
</tr>
<tr>
<td># documents</td>
<td></td>
<td>90,266</td>
<td>1,220</td>
<td>1,093</td>
<td>196,961</td>
<td>12,148</td>
</tr>
<tr>
<td># queries</td>
<td></td>
<td>380,298</td>
<td>3,924</td>
<td>3,198</td>
<td>879,450</td>
<td>64,835</td>
</tr>
<tr>
<td>Max # entities</td>
<td></td>
<td>527</td>
<td>187</td>
<td>396</td>
<td>371</td>
<td>232</td>
</tr>
<tr>
<td>Avg # entities</td>
<td></td>
<td>26.4</td>
<td>26.5</td>
<td>24.5</td>
<td>26.5</td>
<td>25.5</td>
</tr>
<tr>
<td>Avg # tokens</td>
<td></td>
<td>762</td>
<td>763</td>
<td>716</td>
<td>813</td>
<td>774</td>
</tr>
<tr>
<td>Vocab size</td>
<td></td>
<td>118,497</td>
<td></td>
<td></td>
<td>208,045</td>
<td></td>
</tr>
</tbody>
</table>
Sample DeepMind Reading Comprehension Test

<table>
<thead>
<tr>
<th>Original Version</th>
<th>Anonymised Version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td></td>
</tr>
<tr>
<td>The BBC producer allegedly struck by Jeremy Clarkson will not press charges</td>
<td>the \textit{ent381} producer allegedly struck by \textit{ent212} will</td>
</tr>
<tr>
<td>against the “Top Gear” host, his lawyer said Friday. Clarkson, who hosted one</td>
<td>not press charges against the “\textit{ent153}” host, his</td>
</tr>
<tr>
<td>of the most-watched television shows in the world, was dropped by the BBC</td>
<td>lawyer said Friday. \textit{ent212}, who hosted one of the</td>
</tr>
<tr>
<td>Wednesday after an internal investigation by the British broadcaster found he</td>
<td>most-watched television shows in the world, was dropped by the \textit{ent381}</td>
</tr>
<tr>
<td>had subjected producer Oisin Tymon “to an unprovoked physical and verbal attack.”</td>
<td>\textit{wednesday} after an internal investigation by the \textit{ent180}</td>
</tr>
<tr>
<td></td>
<td>broadcaster found he had subjected producer \textit{ent193} “to an unprovoked</td>
</tr>
<tr>
<td></td>
<td>physical and verbal attack.”</td>
</tr>
<tr>
<td><strong>Query</strong></td>
<td></td>
</tr>
<tr>
<td>Producer X will not press charges against Jeremy Clarkson, his lawyer says.</td>
<td>producer X will not press charges against \textit{ent212}, his lawyer says.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Answer</strong></td>
<td></td>
</tr>
<tr>
<td>Oisin Tymon</td>
<td>\textit{ent193}</td>
</tr>
</tbody>
</table>

Table 3: Original and anonymised version of a data point from the Daily Mail validation set. The anonymised entity markers are constantly permuted during training and testing.
Deep LSTM Reader

- DeepMind uses LSTM recurrent neural net (RNN) to encode document and query into a vector that is then used to predict the answer.

Incorporated various forms of attention to focus the reader on answering the question while reading the document.
Visual Question Answering (VQA)

• Answer natural language questions about information in images.
• VaTech/MSR group has put together VQA dataset with ~750K questions over ~250K images (Antol et al., 2016).
VQA Examples

What color are her eyes?  
What is the mustache made of?

How many slices of pizza are there?  
Is this a vegetarian pizza?

Is this person expecting company?  
What is just under the tree?

Does it appear to be rainy?  
Does this person have 20/20 vision?
LSTM System for VQA

“How many horses are in this image?”