Slide 1

CS 4120: Natural Language Processing

Instructor: Prof. Lu Wang
Northeastern University
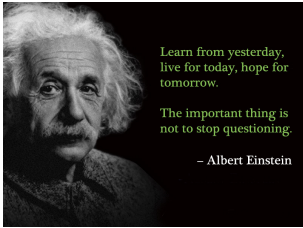Webpage: www.ccs.neu.edu/home/luwang

1

Slide 2

Logistics

- Assignment 2 is due Feb 28, 11:59pm.
- Additional TA office hours for the week:
  - Monday: 4:00 pm - 6:00 pm (162 WVH)
  - Wednesday: 4:00 pm - 6:00 pm (162 WVH)
  - Thursday: 5:00 - 6:00 pm (WVH 1st floor common lab)
  - Friday: 5:30 pm - 7:30 pm (WVH 1st floor common lab)
- Final exam: April 16, 10:30am-12:30pm, location: TBD
  - Open book, any textbook or note is fine, you can even bring laptop
  - But NO Internet access or search, messaging, cellphone usage
  - No heavy calculation, no programming

2

Slide 3

Question Answering

Learn from yesterday, live for today, hope for tomorrow.

The important thing is not to stop questioning.

– Albert Einstein

3

Slide 4

Questions in Search

4

Slide 5

Questions in Search

5

Slide 6

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

6

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
```

7

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
```

8

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

9

Apple's Siri

10

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?*

11

Types of Questions in Modern Systems

• Factoid questions ← **This lecture**
    • *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: ← **Bonus assignment (detailed later)**
    • *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?*

12

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

13

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, but often relies on retrieval AND knowledge)
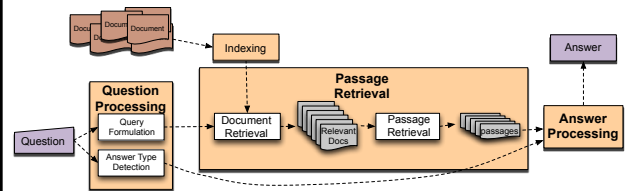
14

Information Retrieval (IR)-based QA

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

15

IR-based Factoid QA

16

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
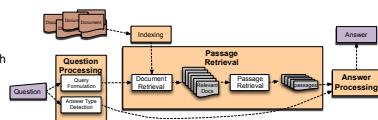
17

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

18

Question Processing

***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)

19

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
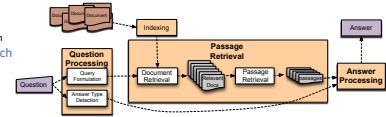
20

Answer Type Detection: Named Entities

- *Who founded Virgin Airlines?*

21

Answer Type Detection: Named Entities

- *Who founded Virgin Airlines?*
  - PERSON
- *What Canadian city has the largest population?*

22

Answer Type Detection: Named Entities

- *Who founded Virgin Airlines?*
  - PERSON
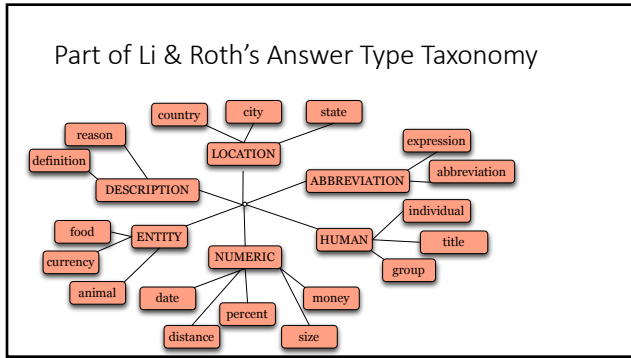- *What Canadian city has the largest population?*
  - CITY

23

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…

24

Part of Li & Roth's Answer Type Taxonomy

```
                 country    city     state
      reason                              expression
  definition            LOCATION              abbreviation
         DESCRIPTION              ABBREVIATION
                                              individual
    food      ENTITY            HUMAN      title
  currency          NUMERIC
    animal    date            money     group
          distance   percent    size
```

25

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?

26

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?

27

Answer types in Jeopardy

Ferrucci et al. 2010. Building Watson: An Overview of the DeepQA Project. AI Magazine. Fall 2010. 59-79.

• 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

28

IR-based Factoid QA
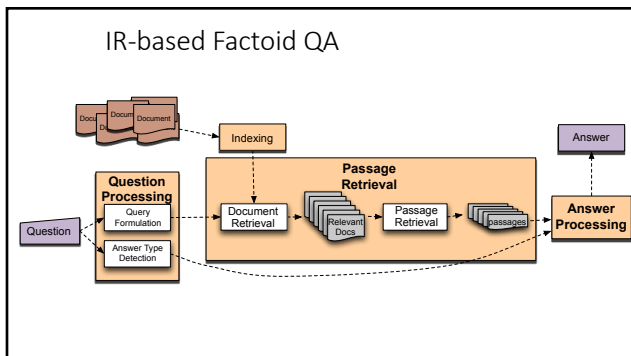
29

Answer Type Detection

• Hand-written rules
• Machine Learning

30

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?

31

Answer Type Detection

- Most often, we treat the problem as machine learning classification
  - **Define** a taxonomy of question types (similar as answer 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!*

32

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?

33

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
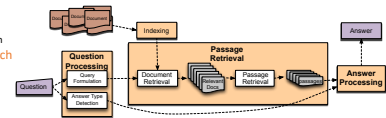
34

Keyword Selection Algorithm

Dan Moldovan, Sanda Harabagiu, Marius Paca, Rada Mihalcea, Richard Goodrum, Roxana Girju and Vasile Rus. 1999. Proceedings of TREC-8.

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

35

Choosing keywords from the query

Slide from Mihai Surdeanu

Who coined the term "cyberspace" in his novel "Neuromancer"?
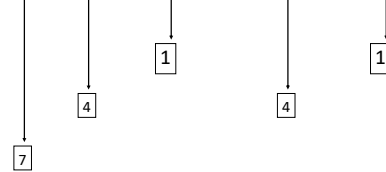
cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7

36

Slide 37 — 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

37

Slide 38 — 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
  - *Think about why? Why not single sentences?*
- Step 3: Passage ranking
  - E.g. use answer type to help rerank passages

38

Slide 39 — 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

39

Slide 40 — 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

40

Slide 41 — 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

41

Slide 42 — 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

42

Adding Analysis Patterns

- "Who is Elvis?"
  - Question type: "who" (answer type is PERSON)
  - 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"

43

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
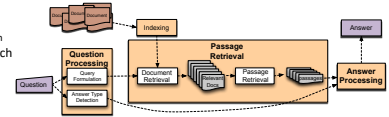
44

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

45

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**

46

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.

47

Candidate Answer scoring in IBM Watson

- Each candidate answer gets scores from >50 components
  - (from unstructured text, semi-structured text, triple stores like knowledge bases)

  - logical form (parse) match between question and candidate
  - passage source reliability
  - geospatial location
    - California is "southwest of Montana"
  - temporal relationships
  - taxonomic classification

48

Information Retrieval (IR)-based QA

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

49

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: ½*
     - *else if third answer is correct: ⅓, etc.*
     - *Score is 0 if none of the M answers are correct*
   - Take the mean over all N queries
   - *Examples*

$$MRR = \frac{\sum_{i=1}^{N} \frac{1}{rank_i}}{N}$$

50

Information Retrieval (IR)-based QA

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

51

Knowledge in QA

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

52

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.?

53

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)
```

54

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)

55

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?"

56

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. → Natural language generation!

57

Information Retrieval (IR)-based QA

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

58

What are recent tasks for QA?

- Reading comprehension (machine reading)

- Popular QA benchmarks

- Visual Question Answering

59

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).

60

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.
A1A2
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?*

61

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.

	CNN			Daily Mail		
	train	valid	test	train	valid	test
# months	95	1	1	56	1	1
# documents	90,266	1,220	1,093	196,961	12,148	10,397
# queries	380,298	3,924	3,198	879,450	64,835	53,182
Max # entities	527	187	396	371	232	245
Avg # entities	26.4	26.5	24.5	26.5	25.5	26.0
Avg # tokens	762	763	716	813	774	780
Vocab size	118,497			208,045		

62

Sample DeepMind Reading Comprehension Test

Original Version	Anonymised Version
Context	
The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." . . .	the *ent381* producer allegedly struck by *ent212* will not press charges against the " *ent153* " host , his lawyer said friday . *ent212* , who hosted one of the most - watched television shows in the world , was dropped by the *ent381* wednesday after an internal investigation by the *ent193* broadcaster found he had subjected producer *ent193* " to an unprovoked physical and verbal attack . " . . .
Query	
Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.	producer **X** will not press charges against *ent212* , his lawyer says .
Answer	
Oisin Tymon	*ent193*

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.

63

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.

Document, Question → LSTM Encoder → Embedding → Answer Extractor → Answer

Incorporated various forms of attention to focus the reader on answering the question while reading the document.

64

Stanford Question Answering Dataset (SQuAD)

- Question: Which team won Super Bowl 50?
- Passage: Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California.
- 100k examples
- Answer must be a span in the passage

65

Stanford Question Answering Dataset (SQuAD)

- Question: Which team won Super Bowl 50?
- Passage: Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California.
- *How do we know which sentence to select?*
- *Furthermore, which phrase is the answer?*

66

No Answer Example

Genghis Khan united the Mongol and Turkic tribes of the steppes and became Great Khan in 1206. He and his successors expanded the Mongol empire across Asia. Under the reign of Genghis' third son, Ögedei Khan, the Mongols destroyed the weakened Jin dynasty in 1234, conquering most of northern China. Ögedei offered his nephew Kublai a position in Xingzhou, Hebei. Kublai was unable to read Chinese but had several Han Chinese teachers attached to him since his early years by his mother Sorghaghtani. He sought the counsel of Chinese Buddhist and Confucian advisers. Möngke Khan succeeded Ögedei's son, Güyük, as Great Khan in 1251. He

When did Genghis Khan kill Great Khan?

67

CoQA: A Conversational Question Answering Challenge

- Input: Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80. Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well. Jessica had . . .

- Q1: Who had a birthday?
- A1: Jessica
- R1 (rational): Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80.

- Q2: How old would she be?
- A2: 80
- R2 (rational): she was turning 80

- Q3: Did she plan to have any visitors?
- A3: Yes
- R3 (rational): Her granddaughter Annie was coming over

68

The NarrativeQA Reading Comprehension Challenge

- Title: Ghostbusters II
- Question: How is Oscar related to Dana?
- Answer: her son
- Summary snippet: . . . Peter's former girlfriend Dana Barrett has had a son, Oscar. . .
- Story snippet:
  - DANA (setting the wheel brakes on the buggy) Thank you, Frank. I'll get the hang of this eventually.
  - She continues digging in her purse while Frank leans over the buggy and makes funny faces at the baby, OSCAR, a very cute nine-month old boy.
  - FRANK (to the baby) Hiya, Oscar. What do you say, slugger?
  - FRANK (to Dana) That's a good-looking kid you got there, Ms. Barrett.

69

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).
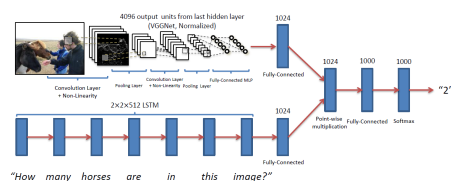
70

VQA Examples

71

LSTM System for VQA

72

Bonus Assignment

- Your task: read a paragraph, create complex questions, and provide their respective answers using the given paragraph
- In total, there are ~31 (+/- 2) paragraphs.
- Roughly requires 2 hours.
- Bonus 2 points (out of 100).

73

Bonus Assignment

- What's the purpose of this bonus assignment?
  - Learning to construct complex questions that need to be answered with one or multiple sentences. (compared to the factoid questions that can be answered with a phrase or a couple of words)

  - Participating in a research study for complex question answering (by contributing your constructed question-answer pairs).

74

Bonus Assignment

- Q1: What types of questions do I ask?
- 1. You may want to ask questions starting with why , how or what. This would result in answers that either tries to reason or explain a concept. (Don't ask who, which, when questions.)
- 2. In addition to this, you can also ask another (why , how or what) question as a follow-up to one (or more) of your questions if possible for the given summary paragraph.
  - These follow-up questions usually would ask more specific information for the question it is related to. You can make multiple follow-ups for a question and even follow-ups to a follow-up question.

75

Good examples! (with format)

Q1:What were the findings of the GAO report?
A1:In September 2014 , GAO reported on the Department of Veterans Affairs ' (VA) Program of Comprehensive Assistance for Family Caregivers (Family Caregiver Program) and found that the program office had limitations with its information technology (IT) system—the Caregiver Application Tracker (CAT) .
Q2:How has the VA attempted to improve the CAT program?
A2:VA has initiated various projects since 2015 to implement a new system , but has not yet been successful in its efforts . (See figure .) Specifically , in July 2015 VA initiated a project to improve the reliability of CAT 's data , called CAT Rescue .
Q2.1:Why did CAT Rescue end in April 2018?
A2.1:However , the department reported in January 2017 that it had identified numerous defects during system testing . The project ended in April 2018 before any new system capabilities were implemented .
Q2.2:Why was the CareT Program unsuccessful?
A2.2:The CareT project was expected to use improved data from CAT Rescue , while also adding new system capabilities . However , the user acceptance testing of CareT identified the need for the department to develop more system capabilities than originally planned .

76

Bad examples!

- Q1:When did GAO report on the Department of Veterans Affairs ' (VA) Program of Comprehensive Assistance for Family Caregivers? (The question is too specific and the response can only be a date or time)

- Q2:Who initiated the CAT Rescue project to improve the reliability of CAT 's data? (To specific question again)

77

Bonus Assignment

- Q2: How should I provide the answers to the questions I ask?
- For each question you make, you should copy-paste the answer span from the given summary paragraph (for which you made the question), that either partially or completely answers that question.
  - Here the answer span means the sentences that you considered to generate the question or the sentences that you think contains your full or part of the intended answer. Note for your answer, copy-paste one or more complete sentence(s), not just words or phrases from the given paragraph. *Please ensure that you do not copy a phrase or word as the answer!*

78

Bonus Assignment

- Q3: How many question answer pairs should I make?
- You should try to make as many questions or their follow-up questions per summary paragraph as you can. The goal is to cover all the information in the given paragraph.
- Try to create at least 4 questions in total, with at least 1 follow-up question per paragraph.
- But if there are only 3 sentences, and you feel like you really can't ask the fourth question. Three questions are fine. But try to come up with one follow-up question.

79

Bonus Assignment

- Q4: How do I access the annotation file and how do I turn my annotations in?
- Instructions will be available on blackboard, including details on the annotation guideline and format.

80

Bonus Assignment

- Q5: When it's due?
- It will be released by this week (~Feb 29), and it's due on March 22.

81