Question Answering

COMP90042

Natural Language Processing

Lecture 19

Semester 1 2021 Week 10

Jey Han Lau

THE UNIVERSITY OF MELBOURNE

Introduction

- **Definition**: question answering ("QA") is the task of automatically determining the answer for a natural language question
- Mostly focus on "factoid" questions

Factoid Questions

Factoid questions, have short precise answers:

- What war involved the battle of Chapultepec?
- What is the date of Boxing Day?
- What are some fragrant white climbing roses?
- What are tannins?

Non-factoid Questions

General non-factoid questions require a longer answer, critical analysis, summary, calculation and more:

- Why is the date of Australia Day contentious?
- What is the angle 60 degrees in radians?

Why do we focus on factoid questions in NLP?

- They are easier
- They have an objective answer
- Current NLP technologies cannot handle non-factoid answers
- There's less demand for systems to automatically answer non-factoid questions

2 Key Approaches

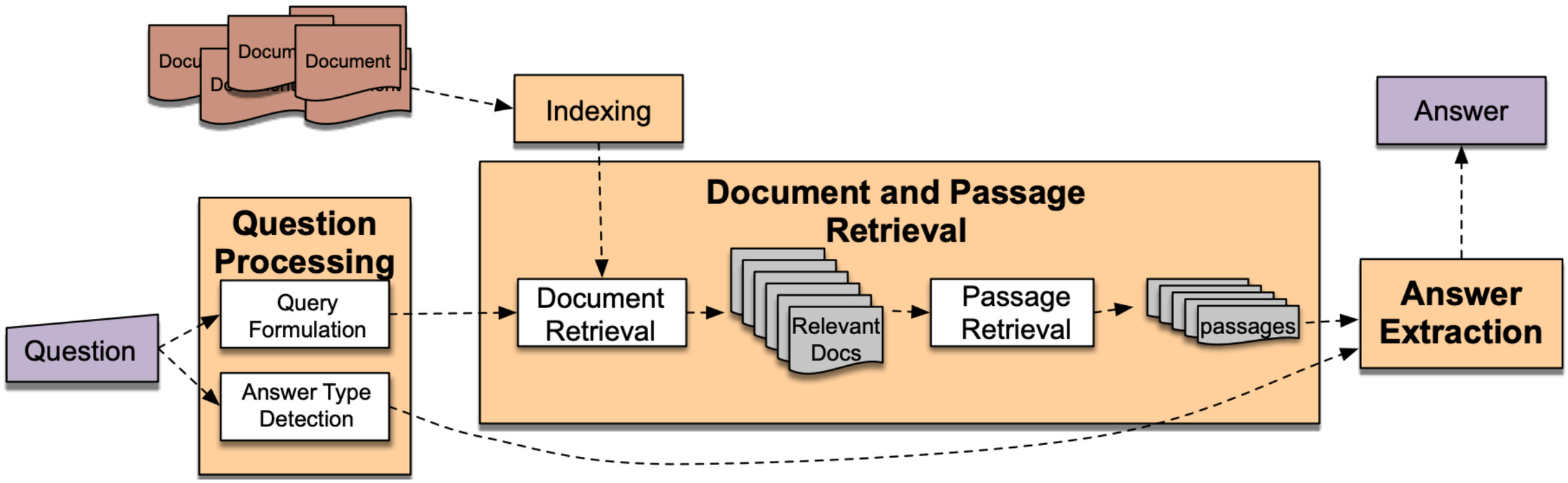
- Information retrieval-based QA
  - Given a query, search relevant documents
  - Find answers within these relevant documents
- Knowledge-based QA
  - Builds semantic representation of the query
  - Query database of facts to find answers

Outline

- IR-based QA
- Knowledge-based QA
- Hybrid QA

IR-based QA

IR-based Factoid QA: TREC-QA



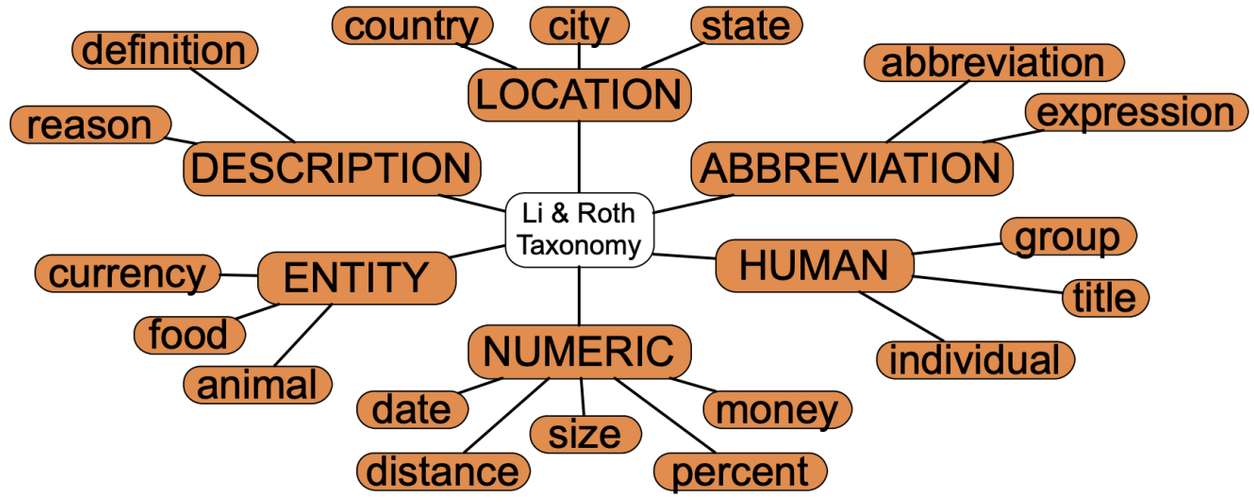
1. Use question to make query for IR engine
2. Find document, and passage within document
3. Extract short answer string

Question Processing

- Find key parts of question that will help retrieval
  - Discard non-content words/symbols (wh-word, ?, etc)
  - Formulate as tf-idf query, using unigrams or bigrams
  - Identify entities and prioritise match
- May reformulate question using templates
  - E.g. “Where is Federation Square located?”
  - Query = “Federation Square located”
  - Query = “Federation Square is located [in/at]”
- Predict expected answer type (here = LOCATION)

Answer Types

- Knowing the type of answer can help in:
  - finding the right passage containing the answer
  - finding the answer string
- Treat as classification
  - given question, predict answer type
  - key feature is question **headword**
  - *What are the **animals** on the Australian coat of arms?*
  - Generally not a difficult task



Tag	Example
ABBREVIATION	
abb	What's the abbreviation for limited partnership?
exp	What does the "c" stand for in the equation E=mc2?
DESCRIPTION	
definition	What are tannins?
description	What are the words to the Canadian National anthem?
manner	How can you get rust stains out of clothing?
reason	What caused the Titanic to sink?
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?
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?
temp	How hot should the oven be when making Peachy Oat Muffins?
speed	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?

Retrieval

- Find top *n* documents matching query (standard IR)
- Next find passages (paragraphs or sentences) in these documents (also driven by IR)
- Should contain:
  - many instances of the question keywords
  - several named entities of the answer type
  - close proximity of these terms in the passage
  - high ranking by IR engine
- Re-rank IR outputs to find best passage (e.g., using supervised learning)

Answer Extraction

- Find a concise answer to the question, as a span in the passage
  - "Who is the federal MP for Melbourne?"
  - *The Division of Melbourne is an Australian Electoral Division in Victoria, represented since the 2010 election by Adam Bandt, a member of the Greens.*
  - "How many Australian PMs have there been since 2013?"
  - *Australia has had five prime ministers in five years. No wonder Merkel needed a cheat sheet at the G-20.*

How?

- Use a neural network to extract answer
- AKA **reading comprehension** task
- But deep learning models require lots of data
- Do we have enough data to train comprehension models?

MCTest

- Crowdworkers write fictional stories, questions and answers
- 500 stories, 2000 questions
- Multiple choice questions

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

SQuAD

- Use Wikipedia passages
- First set of crowdworkers create questions (given passage)
- Second set of crowdworkers label the answer
- 150K questions (!)

- Second version includes unanswerable questions

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, Dangerously in Love (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".
Q: "In what city and state did Beyoncé grow up?" A: "Houston, Texas"
Q: "What areas did Beyoncé compete in when she was growing up?" A: "singing and dancing"
Q: "When did Beyoncé release Dangerously in Love?" A: "2003"

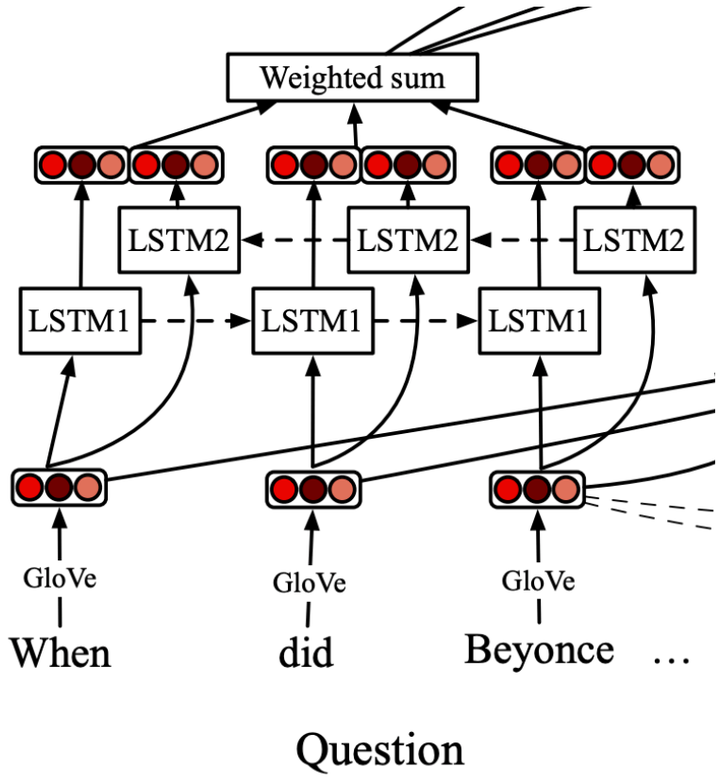
Reading Comprehension

- Given a question and context passage, **predict where the answer span starts and end in passage**?
- Compute:
  - $p_{start}(i)$: prob. of token i is the starting token
  - $p_{end}(i)$: prob. of token i is the ending token

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, Dangerously in Love (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".
Q: "In what city and state did Beyoncé grow up?" A: "Houston, Texas"

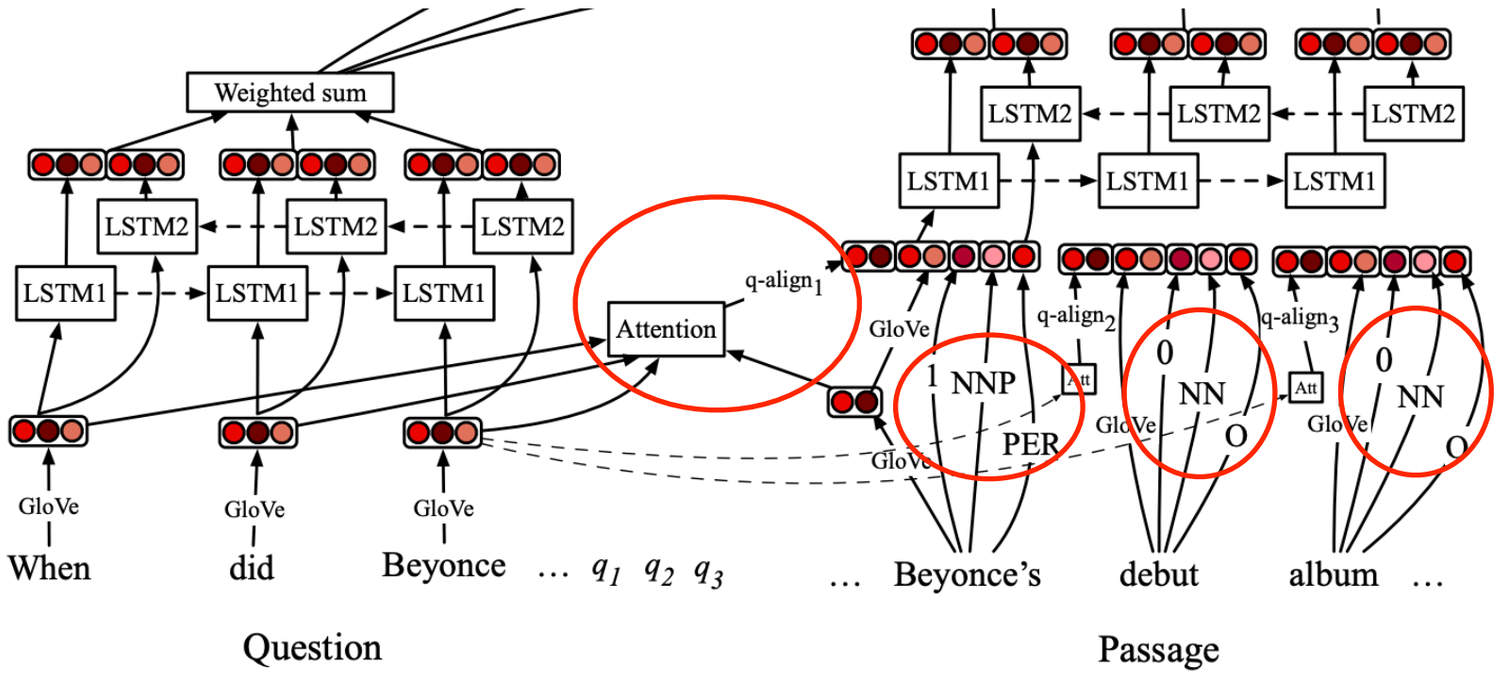
LSTM-Based Model

- Feed question tokens to a bidirectional LSTM
- Aggregate LSTM outputs via weighted sum to produce q, the final question embedding



LSTM-Based Model

- Process passage in a similar way, using another bidirectional LSTM
- More than just word embeddings as input
  - A feature to denote whether the word matches a question word
  - POS feature
  - Weighted question embedding: produced by attending to each question words

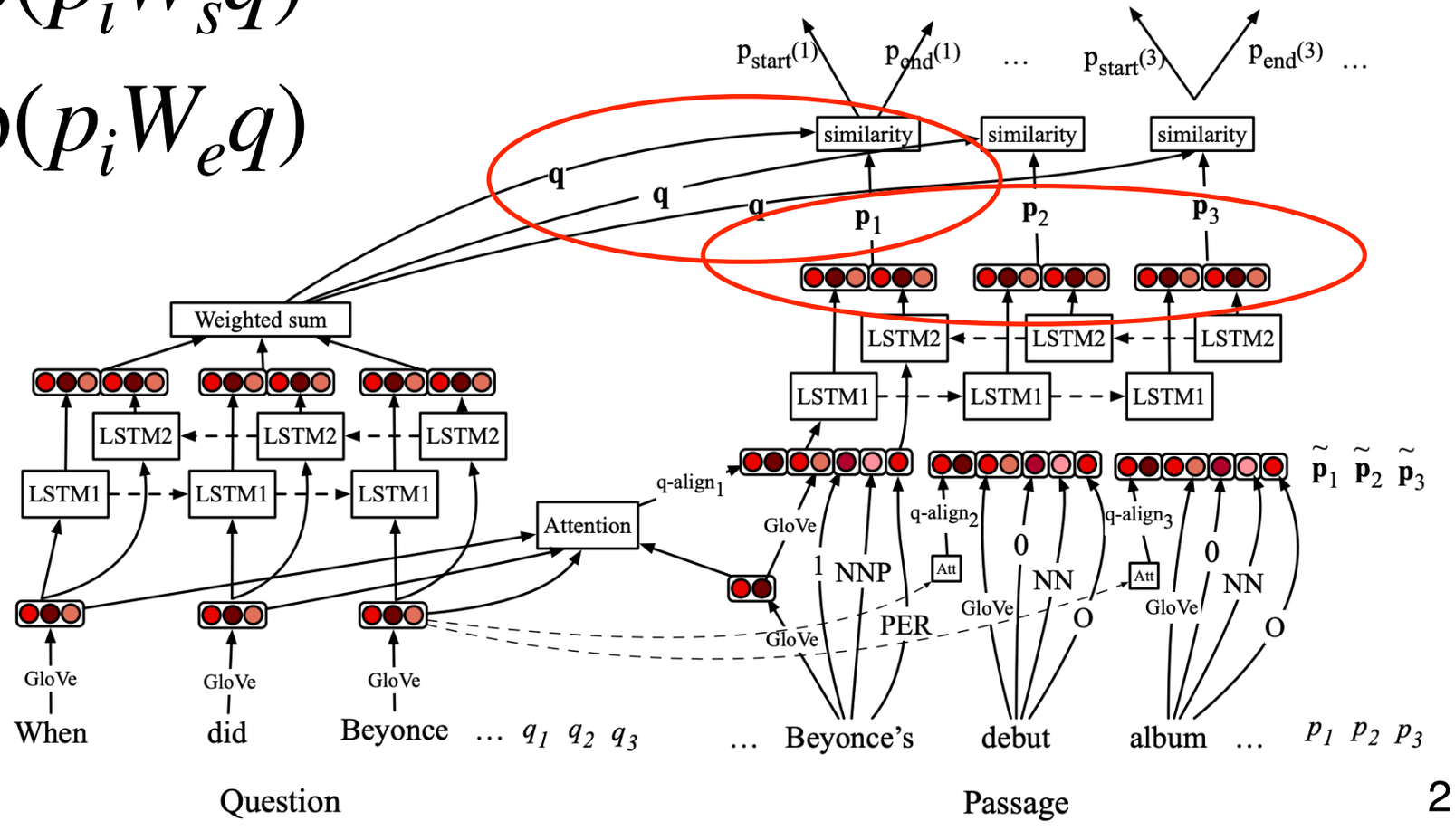


LSTM-Based Model

- $\{p_1, \ldots, p_m\}$: one vector for each passage token from bidirectional LSTM
- To compute start and end probability for each token:

$$p_{start}(i) \propto \exp(p_i W_s q)$$

$$p_{end}(i) \propto \exp(p_i W_e q)$$

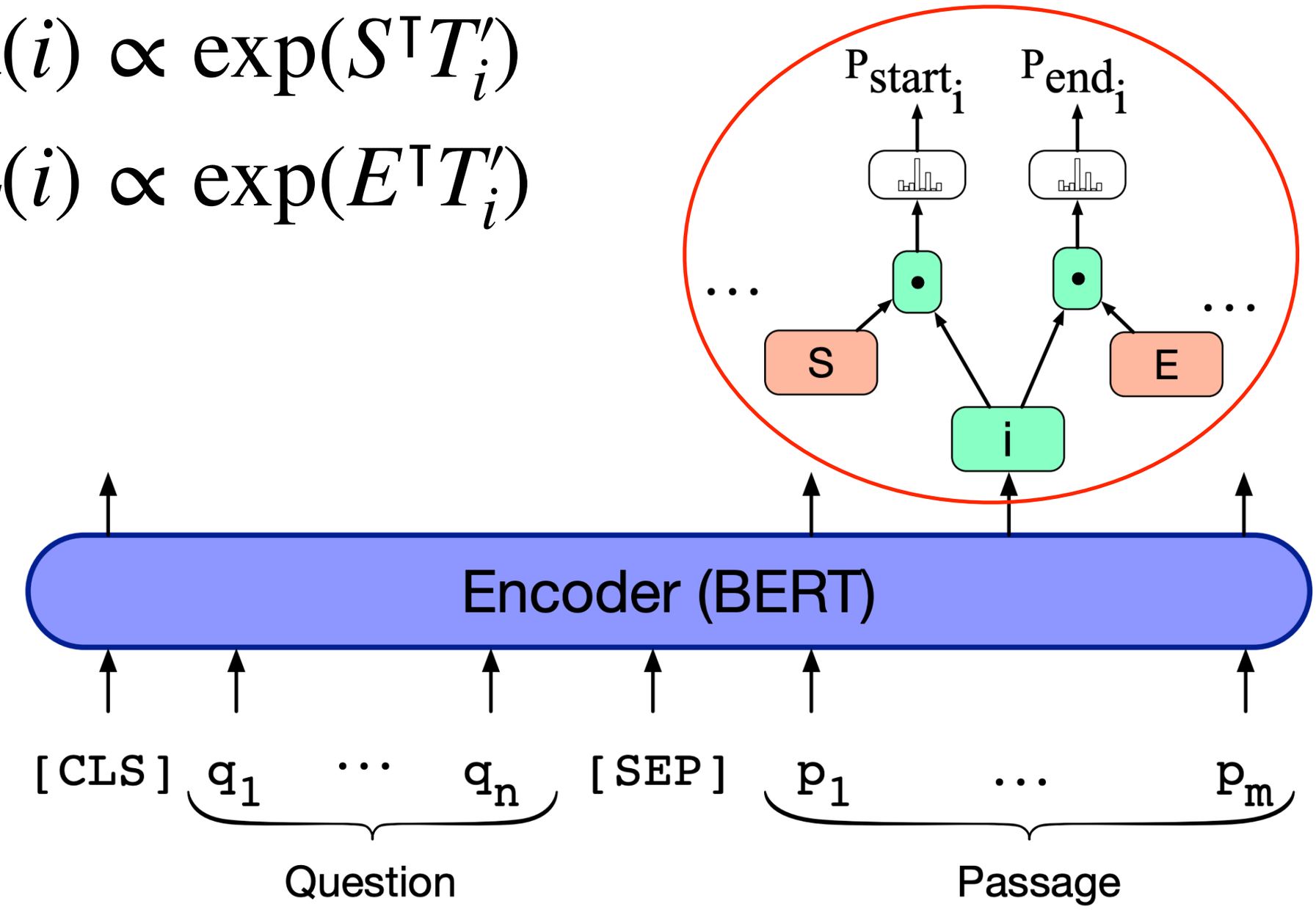


Bert-Based Model

- Fine-tune BERT to predict answer span

$$p_{start}(i) \propto \exp(S^\intercal T_i')$$

$$p_{end}(i) \propto \exp(E^\intercal T_i')$$



Why BERT works better than LSTM?

- It has more parameters
- It's pre-trained and so already "knows" language before it's adapted to the task
- Multi-head attention is the secret sauce
- Self-attention architecture allows fine-grained analysis between words in question and context paragraph

Knowledge-based QA

QA over structured KB

- Many large knowledge bases
  - Freebase, DBpedia, Yago, …
- Can we support natural language queries?
  - E.g. “When was Ada Lovelace born?” → `birth-year (Ada Lovelace, ?x)`
    “What is the capital of England?” → `capital-city(?x, England)`
  - Link “Ada Lovelace” with the correct entity in the KB to find triple *(Ada Lovelace, birth-year, 1815)*

But…

- Converting natural language sentence into triple is not trivial

"When was Ada Lovelace born?" → `birth-year (Ada Lovelace, ?x)`

- Entity linking also an important component
  - Ambiguity: "When was **Lovelace** born?"
- Can we simplify this two-step process?

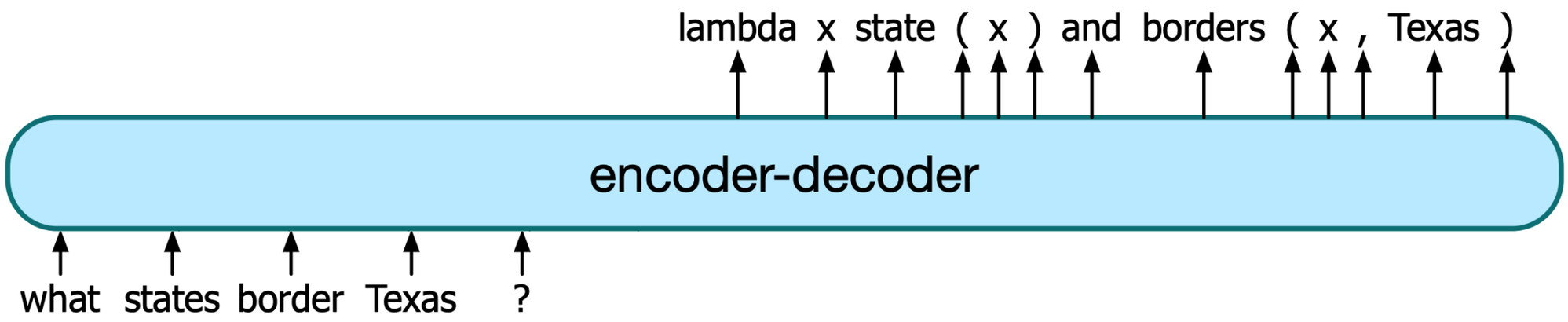
Semantic Parsing

- Convert questions into logical forms to query KB directly
  - Predicate calculus
  - Programming query (e.g. SQL)

Question	Logical form
What states border Texas?	$\lambda x.\text{state}(x) \wedge \text{borders}(x, \text{texas})$
What is the largest state?	$\text{argmax}(\lambda x.\text{state}(x), \lambda x.\text{size}(x))$
I'd like to book a flight from San Diego to Toronto	SELECT DISTINCT f1.flight_id FROM flight f1, airport_service a1, city c1, airport_service a2, city c2 WHERE f1.from_airport=a1.airport_code AND a1.city_code=c1.city_code AND c1.city_name= 'san diego' AND f1.to_airport=a2.airport_code AND a2.city_code=c2.city_code AND c2.city_name= 'toronto'
How many yards longer was Johnson's longest touchdown compared to his shortest touchdown of the first quarter?	ARITHMETIC diff(SELECT num(ARGMAX(SELECT)) SELECT num(ARGMIN(FILTER(SELECT))))

How to Build a Semantic Parser?

- Text-to-text problem:
  - Input = natural language sentence
  - Output = string in logical form
- Encoder-decoder model (lecture 17!)



Hybrid QA

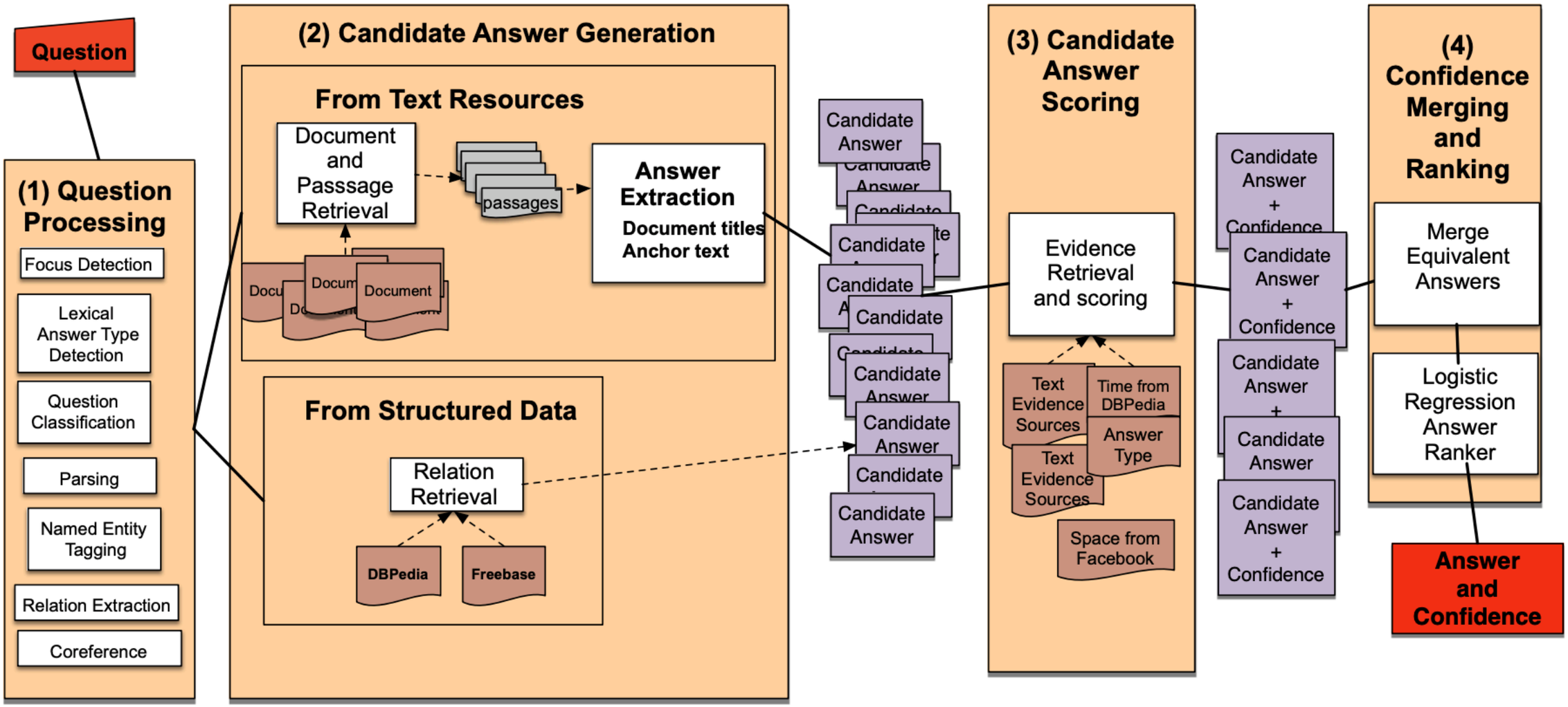
Hybrid Methods

- Why not use both text-based and knowledge-based resources for QA?
- IBM's Watson which won the game show *Jeopardy!* uses a wide variety of resources to answer questions
  - THEATRE: A new play based on this Sir Arthur Conan Doyle canine classic opened on the London stage in 2007.
  - *The Hound Of The Baskervilles*

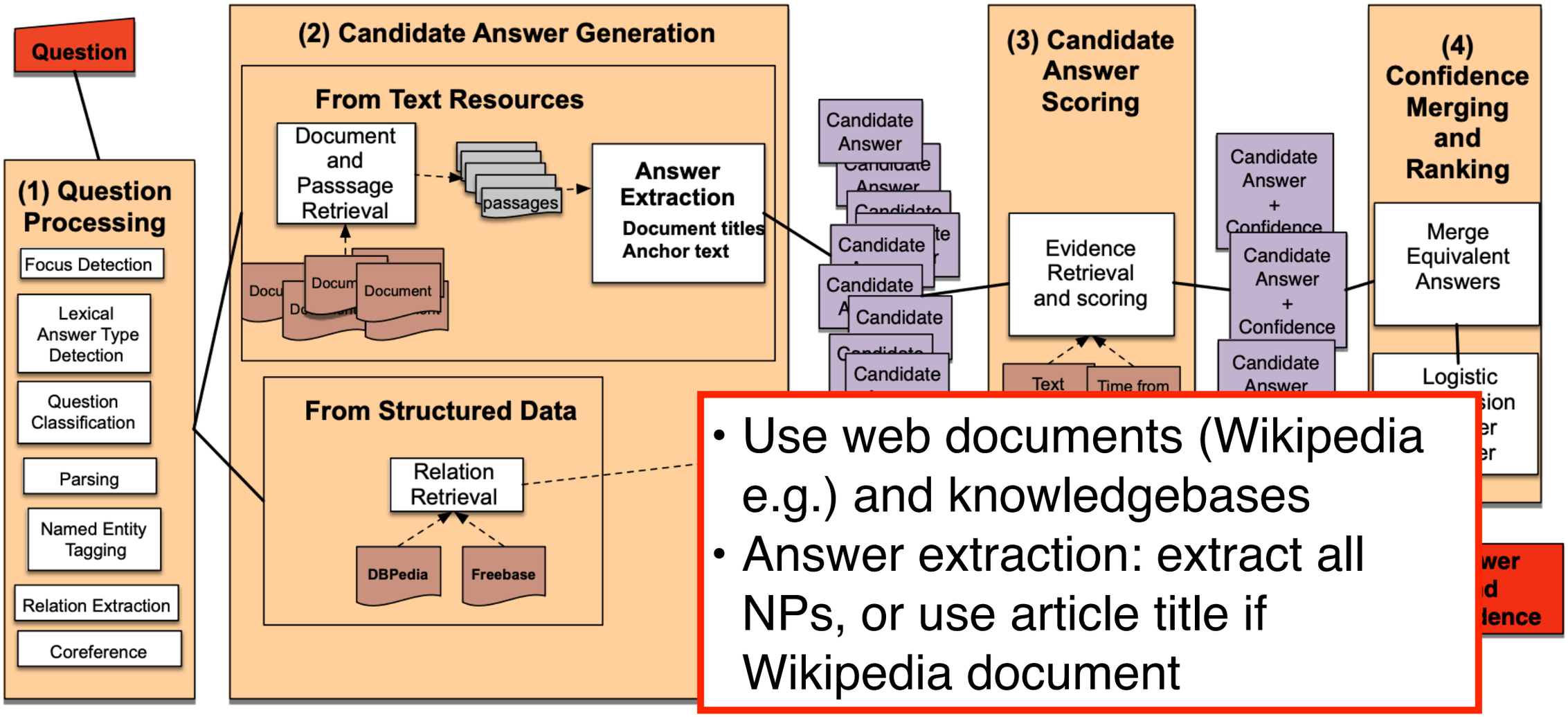
Core Idea of Watson

- Generate lots of candidate answers from text-based and knowledge-based sources
- Use a rich variety of evidence to score them
- Many components in the system, most trained separately

Watson

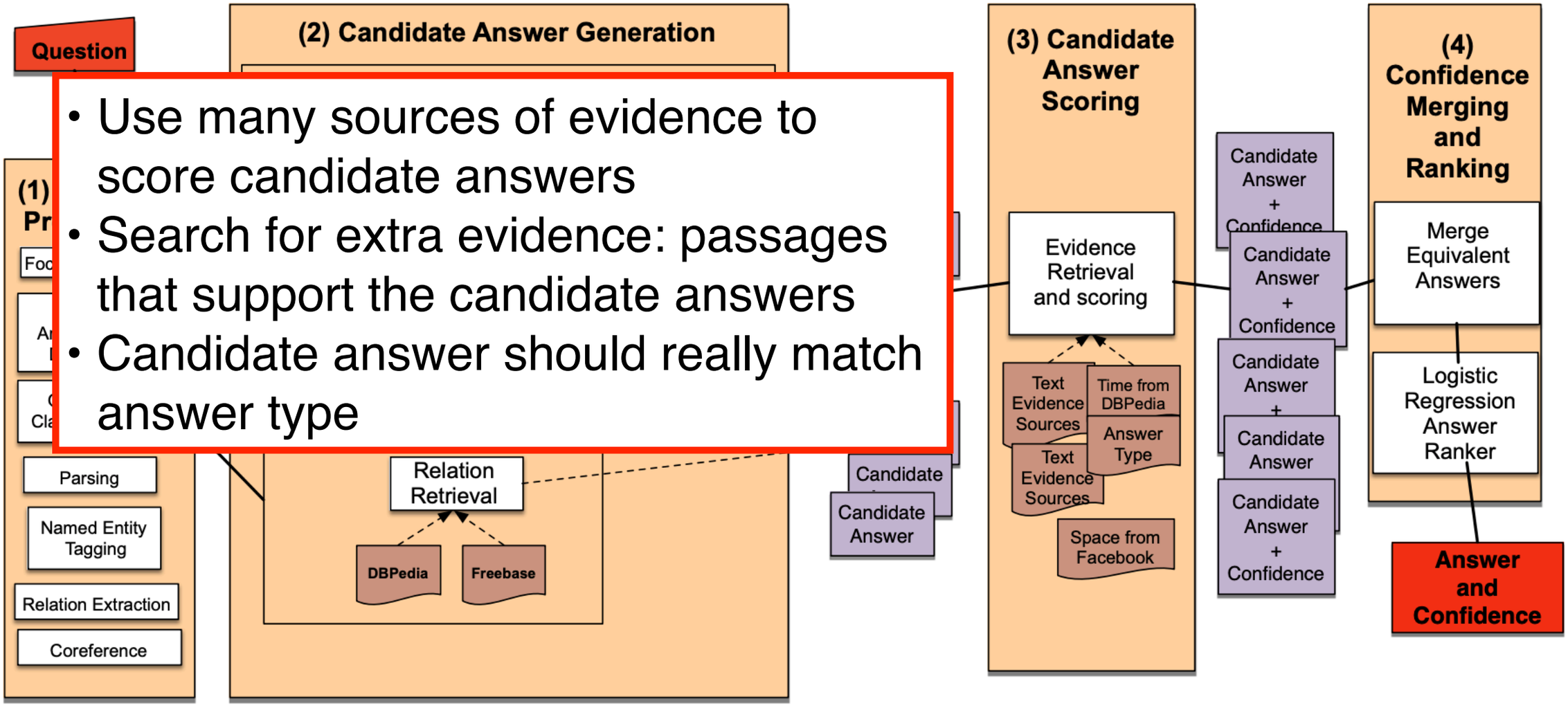

Question
(1) Question Processing
Focus Detection
Lexical Answer Type Detection
Question Classification
Parsing
Named Entity Tagging
Relation Extraction
Coreference
(2) Candidate Answer Generation
From Text Resources
Document and Passsage Retrieval
passages
Answer Extraction
Document titles
Anchor text
Document
From Structured Data
Relation Retrieval
DBPedia
Freebase
Candidate Answer
Candidate
(3) Candidate Answer Scoring
Evidence Retrieval and scoring
Text Evidence Sources
Time from DBPedia
Answer Type
Space from Facebook
Candidate Answer + Confidence
(4) Confidence Merging and Ranking
Merge Equivalent Answers
Logistic Regression Answer Ranker
Answer and Confidence

Watson



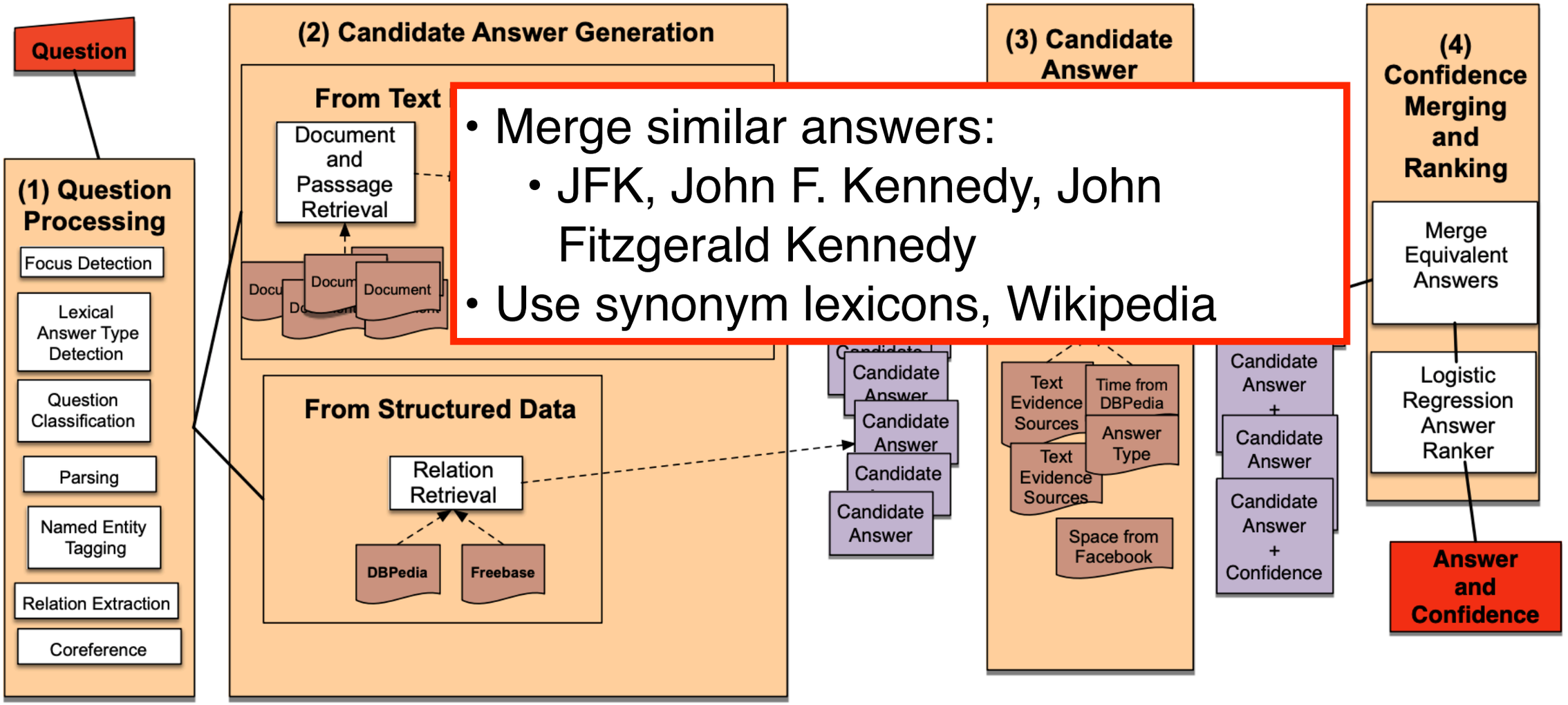
- Use web documents (Wikipedia e.g.) and knowledgebases
- Answer extraction: extract all NPs, or use article title if Wikipedia document

Watson



- Use many sources of evidence to score candidate answers
- Search for extra evidence: passages that support the candidate answers
- Candidate answer should really match answer type

Watson



- Merge similar answers:
  - JFK, John F. Kennedy, John Fitzgerald Kennedy
- Use synonym lexicons, Wikipedia

QA Evaluation

- IR: Mean Reciprocal Rank for systems returning matching passages or answer strings
  - E.g. system returns 4 passages for a query, first correct passage is the 3rd passage
    - MRR = ⅓
- MCTest: Accuracy
- SQuAD: Exact match of string against gold answer

A Final Word

- IR-based QA: search textual resources to answer questions
  - Reading comprehension: assumes question+passage
- Knowledge-based QA: search structured resources to answer questions
- Hot area: many new approaches & evaluation datasets being created all the time (narratives, QA, commonsense reasoning, etc)

Reading

- JM3 Ch. 23.2, 23.4, 23.6