

Question Answering

COMP90042 Lecture 14



THE UNIVERSITY OF
MELBOURNE

Introduction

- **Definition:** question answering (“QA”) is the task of automatically determining the answer (set) for a natural language question
- Main focus on “factoid” QA, e.g.,
 - * Who is the prime minister of the United Kingdom?
→ Teresa May

Examples

Factoid questions, have short precise answers:

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

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?

Two Key Approaches

- 1) Retrieval, find answer question via string/text passage in a document (collection)
- 2) Natural Language Interface to Database (“NLIDB”), automatically construct a query, and answer question relative to fixed KB

IR-based Factoid QA: TREC-QA

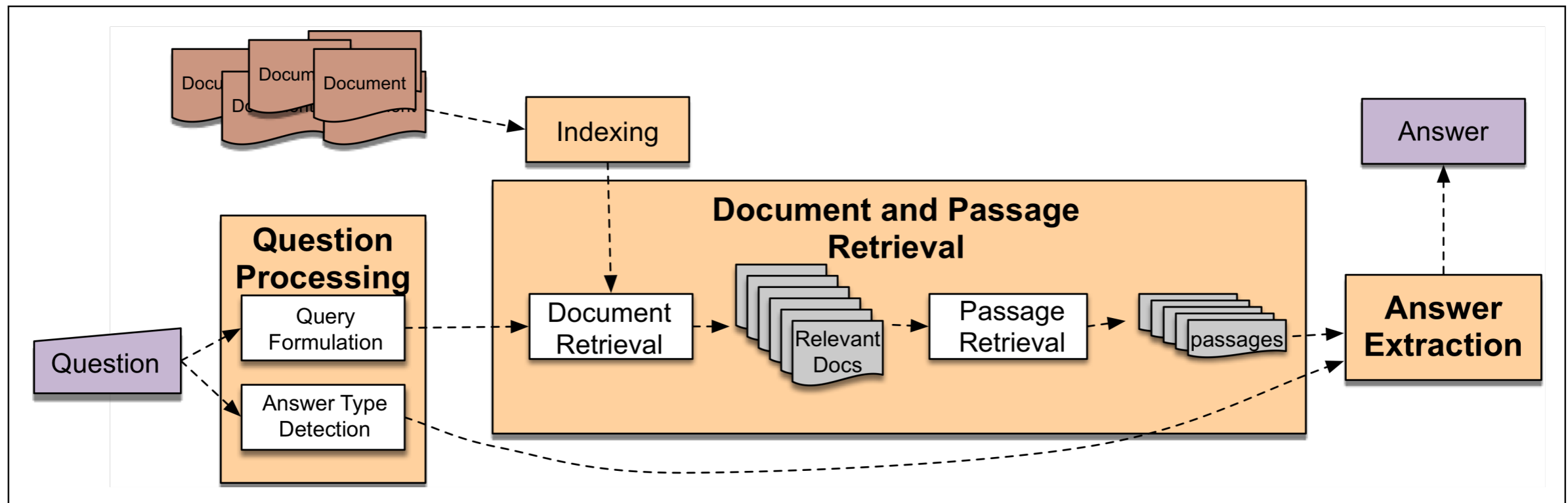


Figure 23.2 IR-based factoid question answering has three stages: question processing, passage retrieval, and answer processing.

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 structural parts (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**

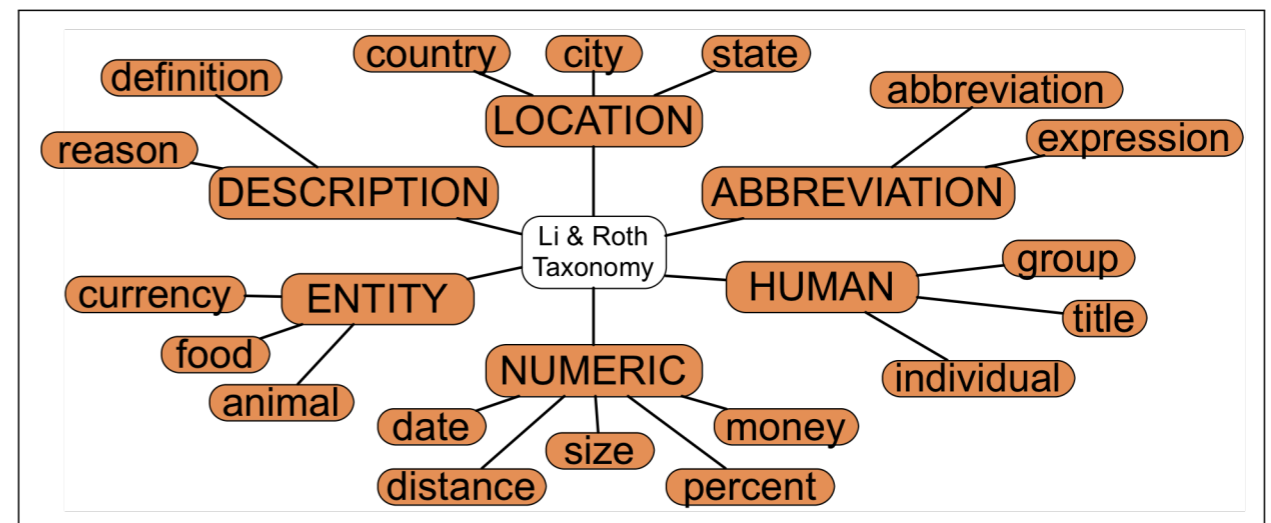


Figure 23.3 A subset of the Li and Roth (2005) answer types.

What are the **animals** on the Australian coat of arms?

Retrieval

- Find top n documents matching query (standard IR)
- Next find passages (paragraphs or sentences) in these documents
- 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; etc
- 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 text
 - * “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.*
- Framed as classification
 - * various features based on match to question, expected entity type match, specific answer patterns

~~Two~~ 2.5 Key Approaches

- 1) Retrieval, find answer question via string/text passage in a document (collection)
- 1.5) Deep learning for answering question from passage**
 - 2) Natural Language Interface to Database (“NLIDB”), automatically construct a query, and answer question relative to fixed KB

Answer Extraction as Reading Comprehension

- Related work on **reading** comprehension, e.g., MCTest
James the turtle was always getting into trouble. Sometimes he'd reach into the freezer and empty out all the food . . .
Q: What is the name of the trouble making turtle?
(a) Fries (b) Pudding (c) James (d) Jane
- **Cloze** questions (filling in the blanks), e.g., Who-did-What
Tottenham won 2-0 at Hapoel Tel Aviv in UEFA Cup ... impressed Spurs skipper Robbie Keane. ... Keane scored the first goal ... with Dimitar Berbatov, who insisted earlier on Thursday he was happy at the London club, heading a second. The 26-year-old Berbatov admitted the reports linking him with a move had affected his performance ... Spurs manager Juande Ramos has ...
Q: Tottenham manager Juande Ramos has hinted he will allow _____ to leave if the Bulgaria striker makes it clear he is unhappy.
(1) Robbie Keane (2) Dimitar Berbatov

SQUAD (Rajpurkar et al., 2016)

- Using Wikipedia text have crowdworkers: (a) create questions which are answerable from a paragraph; and (b) answer questions created by others on that paragraph

The economy of Victoria is highly diversified: service sectors including financial and property services, health, education, wholesale, retail, hospitality and manufacturing constitute the majority of employment. Victoria's total gross state product (GSP) is ranked second in Australia, although Victoria is ranked fourth in terms of GSP per capita because of its limited mining activity. ...

- * What kind of economy does Victoria have?
- * Where according to gross state product does Victoria rank in Australia?
- 150k questions in total, now including unanswerable qs

DrQA (Chen et al, 2017)

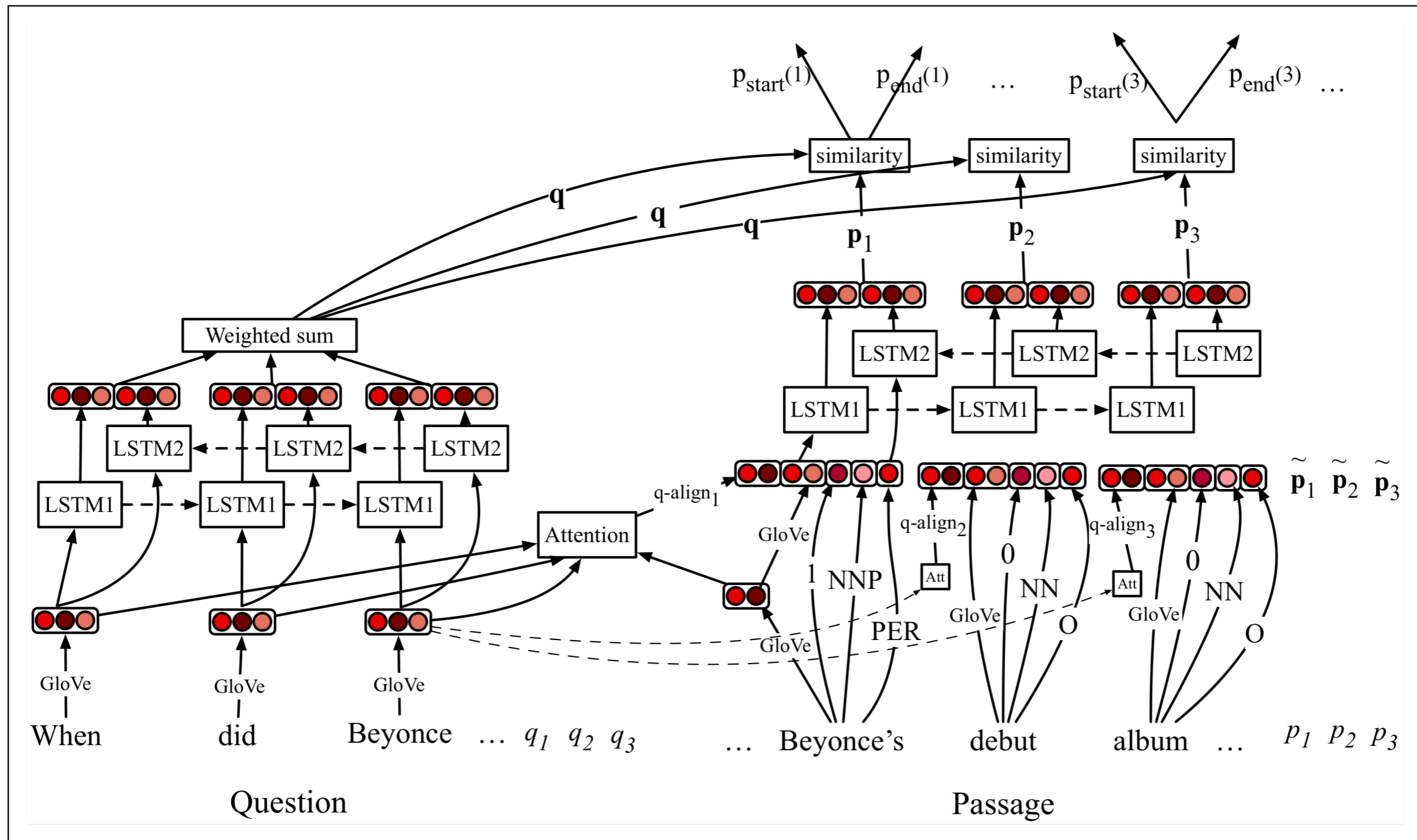


Figure 23.8 The question answering system of Chen et al. (2017), considering part of the question *When did Beyoncé release Dangerously in Love?* and the passage starting *Beyoncé's debut album, Dangerously in Love (2003)*.

~~Two~~ 2.5 Key Approaches

- 1) Retrieval, find answer question via string/text passage in a document (collection)
 - 1.5) Deep learning for answering question from passage
- 2) **Natural Language Interface to Database (“NLIDB”), automatically construct a query, and answer question relative to fixed KB**

QA over structured KB

- Many large knowledge bases
 - * Sports statistics, Moon rock data, ...
 - * Freebase, DBpedia, Yago, ...
- Each with own query language SQL, SPARQL etc.
- 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)`

- * Answer by processing query against KB; i.e., find RDF triple (Ada Lovelace, birth-year, 1815) to provide answer = 1815.

Semantic Parsing

- Based on aligned questions and their logical form, e.g., GeoQuery (Zelle & Mooney 1996)
 What is the capital of the state with the largest population?
 $\text{answer}(C, (\text{capital}(S,C), \text{largest}(P, (\text{state}(S), \text{population}(S,P)))))$.
- Can model using parsing (Zettlemoyer & Collins 2005) to build compositional logical form

$$\begin{array}{c}
 \begin{array}{ccc}
 \text{What} & \text{states} & \text{border} & \text{Texas} \\
 \hline
 (S/(S \setminus NP))/N & N & (S \setminus NP)/NP & NP \\
 \lambda f. \lambda g. \lambda x. f(x) \wedge g(x) & \lambda x. \text{state}(x) & \lambda x. \lambda y. \text{borders}(y, x) & \text{texas} \\
 \hline
 S/(S \setminus NP) & & (S \setminus NP) & \\
 \lambda g. \lambda x. \text{state}(x) \wedge g(x) & & \lambda y. \text{borders}(y, \text{texas}) & \\
 \hline
 S & & & \\
 \lambda x. \text{state}(x) \wedge \text{borders}(x, \text{texas}) & & &
 \end{array}
 \end{array}$$

Sempre (Berant et al 2013)

- Train a NLIDB based solely on (question, answer) pairs, without any given logical forms?
- Instead predict the logical form using a parser
 - * apply logical form to KB to find a candidate answer
 - * learn a model to reward forms giving the correct answer
- Key questions:
 - * how to formulate parser with no data? (lots of tricks!)
 - * how to parameterise model? (relate question to logic)
 - * deal with uncertainty of mapping (use 100s of logical forms for each question)

IBM's WATSON (Ferrucci et al, 2010)

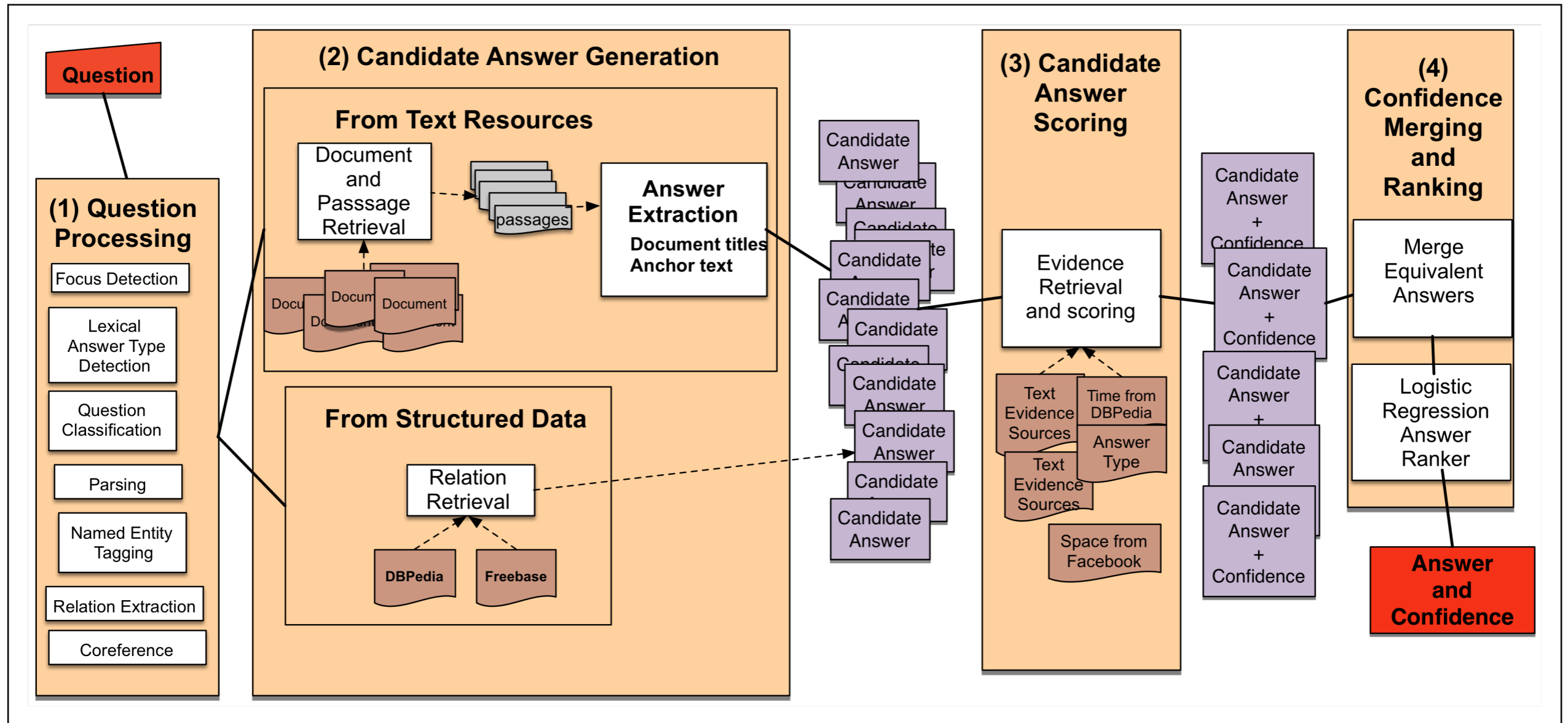


Figure 23.11 The 4 broad stages of Watson QA: (1) Question Processing, (2) Candidate Answer Generation, (3) Candidate Answer Scoring, and (4) Answer Merging and Confidence Scoring.

Evaluation

- TREC-QA: Mean Reciprocal Rank for systems returning matching passages or answer strings
- SQuAD:
 - * Exact match of string against gold answer
 - * F1 score over bag of selected tokens
- Reading comprehension: Accuracy over multiple choice datasets

A final word

- Question answering against text framed as a retrieval pipeline: forming query, document and passage ranking, answer selection
- Answer selection versus reading comprehension
- Natural language querying against knowledge bases using question parsing and logical inference
- Very hot area: Many new approaches & evaluation datasets being created all the time (narratives, QA, commonsense reasoning, etc)

Reading

- JM3 Ch. 23 (skip 23.1.7, 23.2.3, 23.3)
- E18 17.5.2 covers machine reading problems (skip “methods”)
- Cited papers (not required reading)
 - * Berant, J., Chou, A., Frostig, R., and Liang, P. (2013). Semantic parsing on freebase from question-answer pairs. In *EMNLP 2013*.
 - * Chen, D., Fisch, A., Weston, J., and Bordes, A. (2017). Reading wikipedia to answer open-domain questions. In *ACL 2017*.
 - * Ferrucci, D., E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, et al. (2010). Building Watson: An overview of the DeepQA project. *AI magazine* 31(3), 59–79.
 - * Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. In *EMNLP 2016*.
 - * Zettlemoyer, L. and Collins, M. (2005). Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *UAI*.
 - * Zelle, J. M. and Mooney, R. J. (1996). Learning to parse database queries using inductive logic programming. In *AAAI*.