THE UNIVERSITY OF MELBOURNE

Comp90042 Workshop Week 11

May 2022

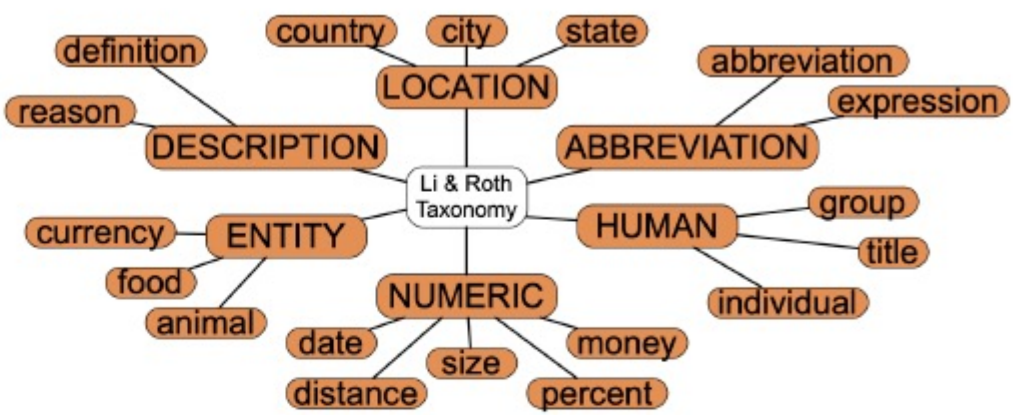
Table of Contents

Table of Contents

Question Answering

1. What is **Question Answering**?

QA is the task of using knowledge — either in terms of raw documents, or in relations that we've already extracted from the documents — to answer questions (perhaps implicitly) posed by a user.



Question Answering

Name	Source	# Questions	Answer
COPA	Q: Weblogs and library	1,000	Multiple choice
MCTest	P: Fictional story	2,000	Multiple choice
WebQuestions	Q: Google suggest API	5,810	KB entity
WikiQA	Q: Bing query logs	3,047	Multiple choice
DMQA	Q & P: CNN & Daily Mail	1,400,000*	Cloze style
SQuAD	P: Wikipedia articles	107,785	Span of words
SQuAD 2.0	P: Wikipedia articles	151,054	Span of words
NewsQA	P: CNN articles	119, 633	Span of words
MS Marco	Q: Bing query logs	1,000,000*	Free answering
SearchQA	Q: Jeopardy	140,000*	Free answering
NarrativeQA	P: Stories	46,765	Free answering
TriviaQA	Q: Trivia quizzes	95,956	Span of words
QAngaroo	P: Wikipedia & PubMed	54,000*	KB entity
ComplexWebQ	Q: WebQuestionsSP	34,689	KB entity
HotpotQA	P: Wikipedia	112,779	Span of words
CommonsenseQA	Q: ConceptNet	12,247	Multiple choice
NaturalQuestions	Q: Google query logs	307,373	Span of words

Table 1: Question answering datasets

Question Answering

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

Extractive question answering/reading comprehension

Question Answering

1. (a) What is **semantic parsing**, and why might it be desirable for QA? Why might approaches like NER be more desirable?

As opposed to syntactic parsing — which attempts to define the structural relationship between elements of a sentence — semantic parsing defines the (meaning–based) relations between those elements.

Example:

Donald Trump is president of the United States.

Syntactic parsing: subject(is, Donald Trump)

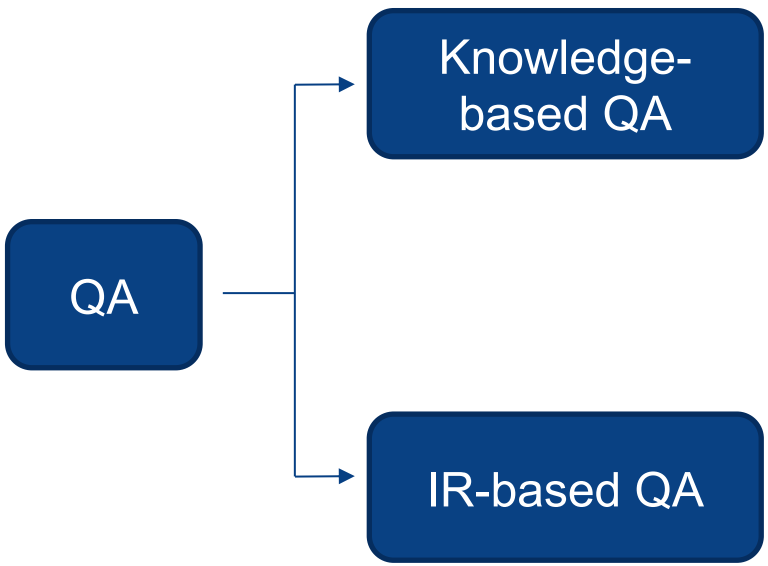
Semantic parsing: is(Donald Trump, president(United States)).

Question: Who is president of the United States?

Representation: is(X,president(United States))

Question Answering

1. (b) What are the main steps for answering a question for a QA system?



Question Answering

1. (b) What are the main steps for answering a question for a QA system?

Knowledge-based QA

– Offline, we process our document collection to generate a list of relations (our knowledge base)

– When we receive a (textual) query, we transform it into the same structural representation, with some known field(s) and some missing field(s)

– We examine our knowledge base for facts that match the known fields

– We rephrase the query as an answer with the missing field(s) filled in from the matching facts from the knowledge base

E.g.:

Knowledge base:
is(Donald Trump, president(United States)).
Is(Ram Nath Kovind, president(India)).
…

Query:
Who is president of the United States?

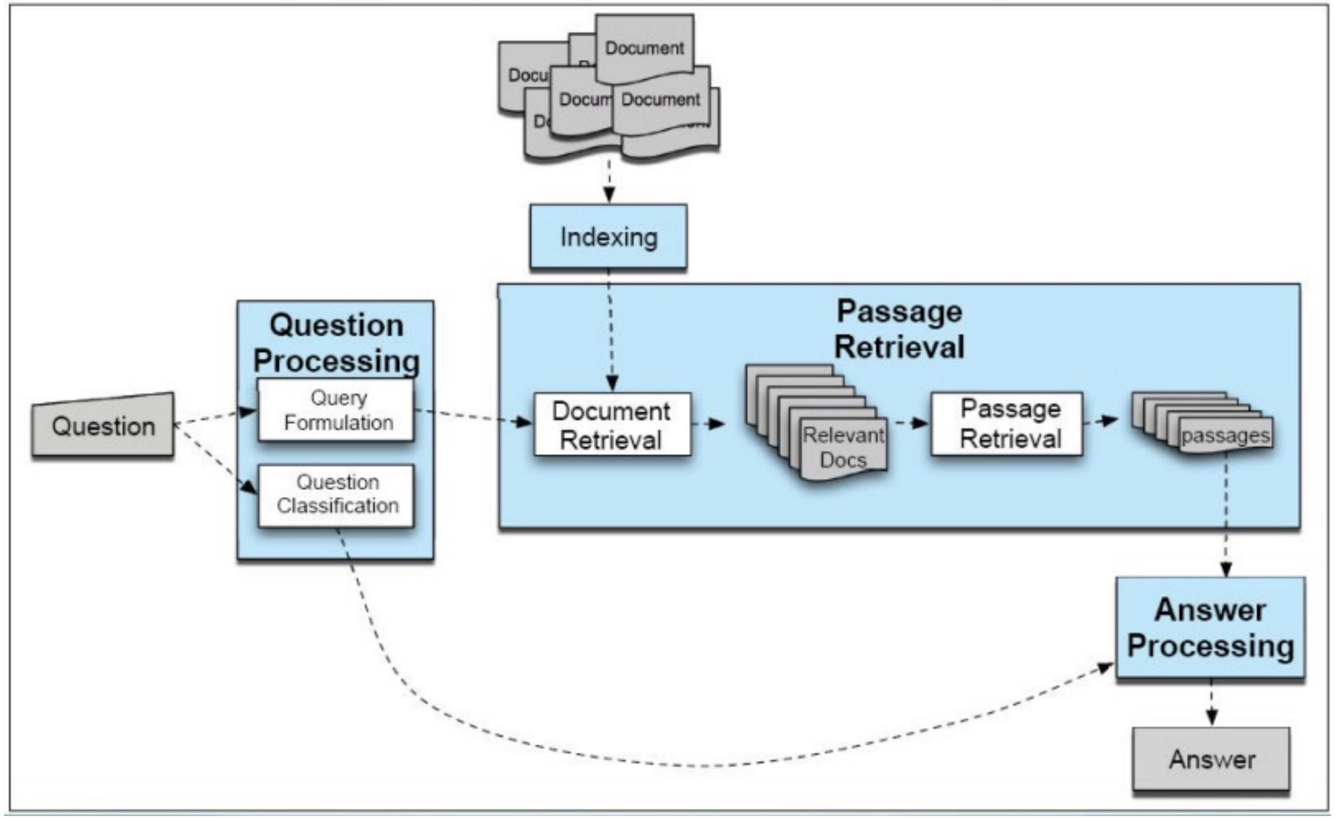
Structural representation:
is(?,president(United States))

Matching facts:
?= Donald Trump

Question Answering

1. (b) What are the main steps for answering a question for a QA system?

IR-based QA



Question Answering

1. (b) What are the main steps for answering a question for a QA system?

IR-based QA

– Offline, we process our document collection into a suitable format for IR querying (e.g. inverted index)

E.g.:

Documents:

Doc1: Donald Trump is the president of United States…

Doc2: Ram Nath Kovind is the president of India.

…

Inverted index:

Words	**Document**
Donald	Doc1
Trump	Doc1
is	Doc1, Doc2
…	…

Question Answering

1. (b) What are the main steps for answering a question for a QA system?

IR-based QA

– Offline, we process our document collection into a suitable format for IR querying (e.g. inverted index)

– When we receive a (textual) query, we remove irrelevant terms, and (possibly) expand the query with related terms

– We select the best document(s) from the collection based on our querying model (e.g. TF-IDF with cosine similarity)

– We identify one or more snippets from the best document(s) that match the query terms, to form an answer

E.g.:

Inverted index: …

Query:
Who is president of the United States?

Cosine similarity:
Query score: cos(doc, query)

Cosine similarity:
Query score: cos(snippet, query)

Table of Contents

Topic Model

2. What is a **Topic Model**?

A topic model is an unsupervised model that discovers latent thematic structure in document collections.

Topic Model

2. (a) What is the **Latent Dirichlet Allocation**, and what are its strengths?

LDA is a particular implementation of topic model. LDA is a probabilistic model that assumes each document has a mixture of topics (in the form of a probability distribution), and each topic has a mixture of words (also a probability distribution).

$$P(t_i) \propto P(t_i|w)P(t_i|d)$$

model probability

$$logP(w_i, w_2, \dots, w_m) = log\sum_{j=0}^{T} P\left(w_1\middle|t_j\right)P\left(t_i\middle|d_{w_1}\right) + \cdots + log\sum_{j=0}^{T} P\left(w_m\middle|t_j\right)P\left(t_j\middle|d_{w_m}\right)$$

Strengths:

Due to its Bayesian formulation (by giving priors to the two aforementioned distributions), LDA is able to infer topics for unseen documents, a capability that its predecessors do not have.

Topic Model

2. (b) What are the different approaches to evaluating a topic model?

- As topic models are unsupervised models, there is no task-based metrics such as accuracy to evaluate them. The best way is to look at the performance of downstream tasks or applications of interest (extrinsic evaluation).

- Other intrinsic evaluation approaches include:
  - Perplexity: a normalized model log probability metric over test data.

$$L = \prod_{w} \sum_{t} P(w \mid t) P(t \mid d_w)$$

$$\text{ppl} = \exp^{\frac{-\log L}{m}}$$

total number of word tokens in test documents

  - Topic coherence: assess how coherent or interpretable the extracted topics are. We can do this manually with word intrusion (injecting random a word into topic and try to guess which is the injected word) or automatically with PMI measures.

Takeaways

1. Question Answering
    1. Knowledge Based QA
    2. IR Based QA
2. Topic Model

THE UNIVERSITY OF MELBOURNE

Thank you