



Question Answering and Summarization

Natalie Parde, Ph.D.

Department of Computer
Science

University of Illinois at
Chicago

CS 421: Natural Language
Processing
Fall 2019

Many slides adapted from Jurafsky and Martin
(<https://web.stanford.edu/~jurafsky/slp3/>).

What is question answering?

- The process of **automatically retrieving** compact quantities of correct, relevant **information** in response to a user's **query**

We use question answering systems everyday.

Google Where is UIC located? Sign in

All Maps Images News Shopping More Settings Tools

About 2,530,000 results (1.65 seconds)

1200 W Harrison St, Chicago, IL 60607 University of Illinois at Chicago, Address Feedback

People also ask

What part of Chicago is UIC in?
Is UIC public or private?
How do I get into UIC?
What major is UIC known for?

Feedback

University of Illinois at Chicago
<https://www.uic.edu> ▾ Located in the heart of one of the world's great cities, the University of Illinois at ... UIC is proud to be recognized as having one of the most ethnically and ...
[Visit & Directions](#) · [Directory](#) · [UIC Map](#) · [Admissions & Aid](#)

Visit & Directions - University of Illinois at Chicago
<https://www.uic.edu/about/about.visit-directions> ▾ Start your tour of campus at the UIC Visitors Center located in our Student Services Building. The east and south sides of campus are also home to the majority ...

See photos See outside

University of Illinois at Chicago

Website Directions Save

Public university in Chicago, Illinois

BUY TICKETS

The University of Illinois at Chicago is a public research university in Chicago, Illinois. Its campus is in the Near West Side community area, adjacent to the Chicago Loop. [Wikipedia](#)

Address: 1200 W Harrison St, Chicago, IL 60607

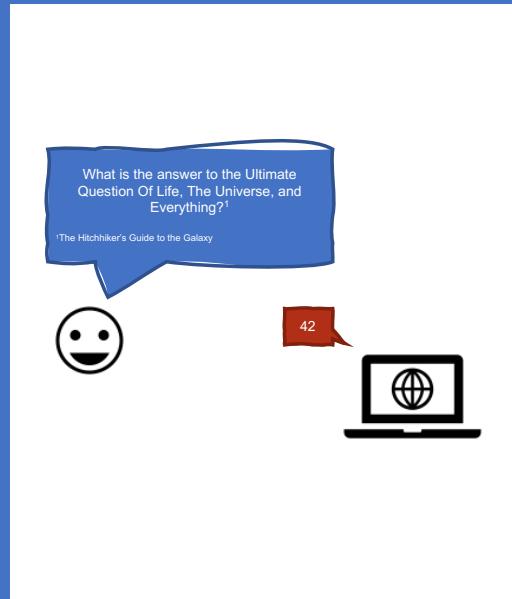
Undergraduate tuition and fees: In-state 13,664 USD, Out-of-state 26,520 USD (2016–17)

Acceptance rate: 73.6% (2016–17)

Typical SAT scores: Reading and Writing 480-580, Math 510-655 (2016–17)

Total enrollment: 29,120 (2016)

Suggest an edit



People have been interested in question answering systems nearly as long as computers have existed.

Technology > TEDx

How did supercomputer Watson beat Jeopardy champion Ken Jennings? Experts discuss.

Posted by: [Kate Torgovnick May](#) April 5, 2013 at 1:59 pm EDT



<https://blog.ted.com/how-did-supercomputer-watson-beat-jeopardy-champion-ken-jennings-experts-discuss/>

Question answering systems have even won game shows!

Question Answering Systems

- Typically focus on **factoid questions**
 - **Factoid Questions:** Questions that can be answered with simple facts expressed in short texts

When was UIC founded?

How far is UIC from the University of Chicago?

What is the average CS class size?

Question Answering Systems

- Two major paradigms:
 - **Information retrieval-based** question answering
 - **Knowledge-based** question answering

Information Retrieval-based Question Answering

- Relies on text from the web or from large corpora
- Given a user question:
 1. Find relevant documents and passages of text
 2. Read the retrieved documents or passages
 3. Extract an answer to the question directly from spans of text

Knowledge-based Question Answering

- Builds a semantic representation of the user's query
 - When was UIC founded? → founded(UIC, x)
- Uses these representations to query a database of facts

Large industrial systems are often hybrids of these two paradigms.

- DeepQA (the question answering system in IBM's Watson):
 - Finds candidate answers in both knowledge bases and text sources
 - Scores each candidate answer
 - Returns the highest scoring answer

Information Retrieval-based Question Answering

Goal: Answer a user's question by finding short text segments containing the requested information

QUESTION

Where is UIC located?

ANSWER

in Chicago, Illinois

What does UIC stand for?

University of Illinois at Chicago

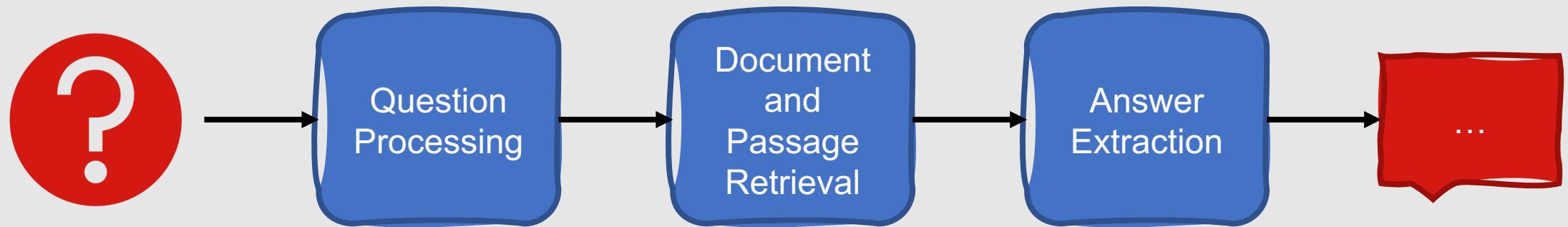
Who taught CS 421 in Fall 2019?

Natalie Parde

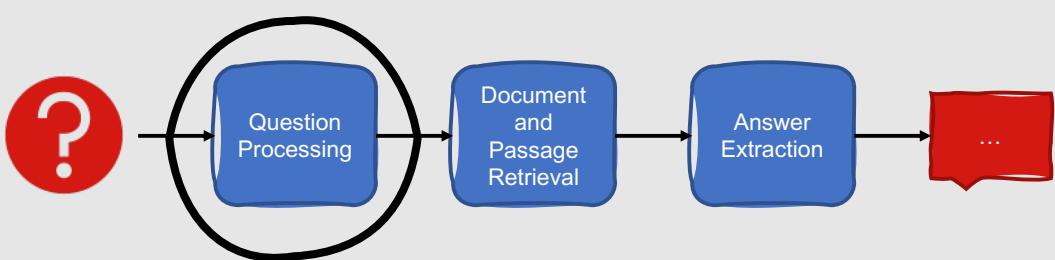
How many grad students are in CS 421?

25

Information Retrieval-based Question Answering



Question Processing



- **Goal: Extract the query**
 - What keywords are needed to match relevant documents?
 - What type of entity should be in the answer (person, location, etc.)?
 - What is the focus of the question (which string of words will likely be replaced by the answer)?
 - What type of question is this (definition, math, list, etc.)?

Question Processing

- Two most common subtasks involved in question processing:
 - Query formulation
 - Answer type detection

When was UIC's Department of Computer Science created?

Query: UIC Department of Computer Science created
Answer Type: Time

Query Formulation



- The task of creating a query to send to an information retrieval system
 - Should contain keywords necessary to obtain relevant documents
- Simple strategy: Pass the entire question as a query
 - Only works with very large corpora (e.g., the web)
- More complex strategy for smaller corpora (e.g., corporate websites or Wikipedia): Use an IR engine to search and index documents

Common Information Retrieval Techniques

TF*IDF matching

- Which document has the highest cosine similarity with the query?

Query expansion

- Add query terms in hopes of matching an answer in one of its many possible forms

Query reformulation

- Rephrase the question to make it look like a substring of possible answers
 - **When was UIC founded? → UIC was founded in**

Answer Type Detection

- The task of determining what type of named entity is needed for the answer
 - Who was the first head of UIC's Department of Computer Science? → PERSON
 - In what city is UIC located? → CITY
- In addition to named entity types, answers can also fall under other categories in a larger, hierarchical, answer type taxonomy
 - PERSON:INDIVIDUAL
 - PERSON:GROUP

ABBREVIATION	abb exp	What's the abbreviation for limited partnership? 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?	

Figure 25.4 Question typology from Li and Roth (2002), (2005). Example sentences are from their corpus of 5500 labeled questions. A question can be labeled either with a coarse-grained tag like HUMAN or NUMERIC or with a fine-grained tag like HUMAN:DESCRIPTION.

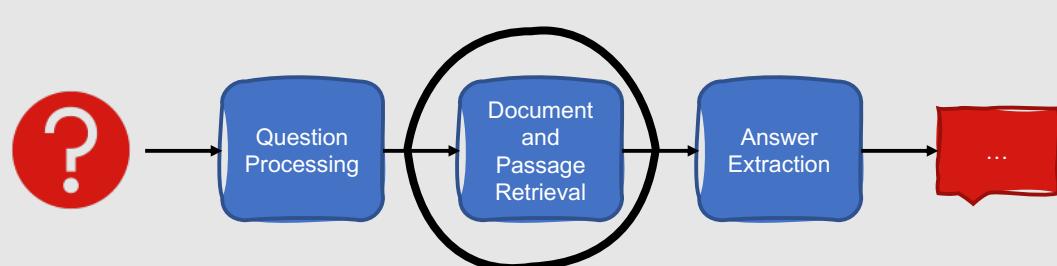
Answer Type Detection

- Hierarchical answer type taxonomy
 - Coarse-grained categories:
 - Abbreviation
 - Description
 - Entity
 - Human
 - Location
 - Numeric
 - Finer-grained subcategories of each

How are answer types detected?

- **Handwritten rules**
 - Who {is | was} the first head of ORGANIZATION → PERSON
- **Supervised machine learning**
 - In general, detecting answer types like PERSON, LOCATION, and TIME is easier; detecting other types is more complex

Document and Passage Retrieval

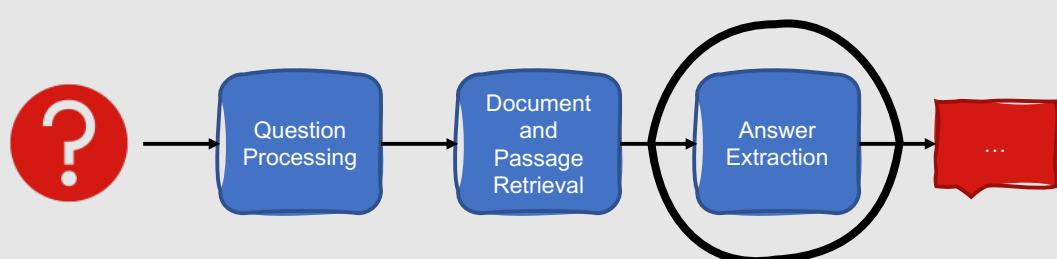


- Ranks a set of documents based on their relevance to the query
- Divides the top n documents into smaller passages
- Pass some or all of those passages along to the next stage

Which passages are passed along to the next stage?

- Simplest approach: Pass along every passage from the top n documents to the next stage
- More sophisticated approaches:
 - Filter the passages based on whether they contain a named entity of the type specified by the question
 - Rank the passages using supervised machine learning and return the subset of highest-ranked passages

Answer Extraction



- Extracts a specific answer from a passage
 - **Span Labeling:** Given a passage, identify the span of text which constitutes an answer

How can we extract answers from passages?

- Simple approach: Run a named entity tagger on the candidate passage, and return whatever entity corresponds to the desired answer type
- However, the answers to many questions may not require a specific named entity type!
 - **What is natural language processing? → The subfield of artificial intelligence that focuses on automatically interpreting and generating natural language**
- Thus, more sophisticated answer extraction systems tend to use supervised machine learning

In what city is UIC located?

UIC, the largest university in **Chicago**....

Feature-based Answer Extraction

Answer type match

- Does the candidate answer contain a phrase with the correct answer type?

Number of matched keywords

- How many keywords from the question are included in the candidate answer?

Text similarity

- What is the cosine similarity between the candidate answer and the query keywords?

Novelty factor

- Does the candidate answer contain a word that was not in the query?

Apposition features

- Is the candidate answer appositive to a phrase containing many question terms?
 - The professor, **Natalie Parde**, is in her office making slides.

Punctuation location

- Is the candidate answer immediately followed by punctuation?

Sequences of question terms

- How long is the longest sequence of question terms in the candidate answer?

N-gram Tiling Answer Extraction

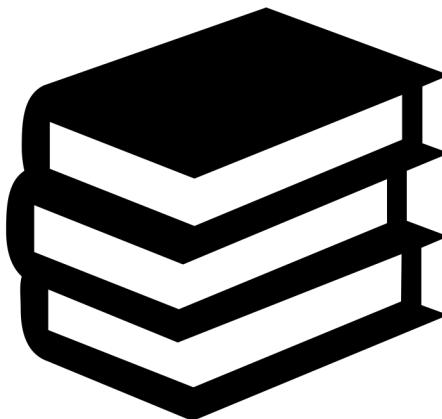
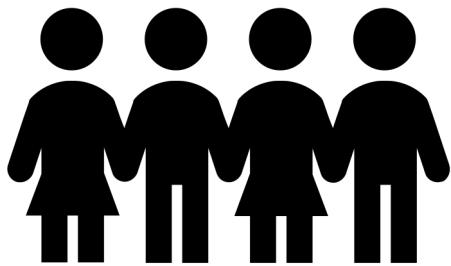
Relies on the redundancy of the web

Works by:

- Starting with the **text snippets returned from a web search engine**
- **Extracting all of the unigrams, bigrams, and trigrams** from each snippet
- **Weighting those n-grams**
 - Based on their frequency and the weight of the patterns that returned them
 - **Scoring those n-grams** based on how well they match the predicted answer type
 - **Concatenating overlapping n-grams** into longer answers
 - **Adding the best concatenation to the list of candidate answers**, and removing lower-scoring candidates

Neural Answer Extraction

- Relies on the intuition that a question and its answer are semantically similar
- Works by:
 - Computing an embedding for the question
 - Computing an embedding for each token of the passage
 - Selecting spans from the passage whose embeddings are closest to the question embedding
- Often designed in the context of **reading comprehension**



Reading Comprehension

- A task designed to measure natural language understanding performance
- Basic premise: Take children's reading comprehension tests, and use them to evaluate text comprehension algorithms

Prime_number

The Stanford Question Answering Dataset

A prime number (or a prime) is a natural number greater than 1 that has no positive divisors other than 1 and itself. A natural number greater than 1 that is not a prime number is called a composite number. For example, 5 is prime because 1 and 5 are its only positive integer factors, whereas 6 is composite because it has the divisors 2 and 3 in addition to 1 and 6. The fundamental theorem of arithmetic establishes the central role of primes in number theory: any integer greater than 1 can be expressed as a product of primes that is unique up to ordering. The uniqueness in this theorem requires excluding 1 as a prime because one can include arbitrarily many instances of 1 in any factorization, e.g., $3, 1 \cdot 3, 1 \cdot 1 \cdot 3$, etc. are all valid factorizations of 3.

What is the only divisor besides 1 that a prime number can have?

Ground Truth Answers: itself itself itself itself itself

What are numbers greater than 1 that can be divided by 3 or more numbers called?

Ground Truth Answers: composite number composite number composite number primes

What theorem defines the main role of primes in number theory?

Ground Truth Answers: The fundamental theorem of arithmetic fundamental theorem of arithmetic

Any number larger than 1 can be represented as a product of what?

Ground Truth Answers: a product of primes product of primes that is unique up to ordering primes primes primes that is unique up to ordering

Why must one be excluded in order to preserve the uniqueness of the

- Stanford Question Answering Dataset (SQuAD)
 - Passages from Wikipedia
 - Associated questions
 - Many have answers that are spans from the passage
 - Some are designed to be unanswerable
 - <https://rajpurkar.github.io/SQuAD-explorer/>
- NewsQA Dataset
 - Question-answer pairs from CNN news articles

Reading Comprehension Datasets

Bidirectional LSTM-based Reading Comprehension

- Low-level goal: Compute, for each token, the probability that it is:
 - The start of the answer span
 - The end of the answer span

How many grad students are in CS 421?

Dr. Parde emailed the 25 grad students in CS 421 to remind them that the final project was only optional for undergrads.

$P_{\text{start}}("25")$

$P_{\text{end}}("25")$

Bidirectional LSTM-based Reading Comprehension

- Learn representations for each question and each word in a passage using bidirectional LSTMs
- Learn classifiers to predict the two probabilities for each word in the passage

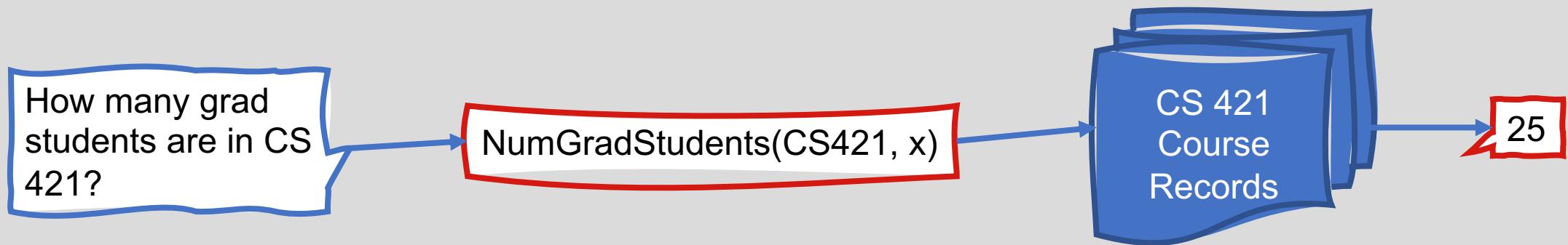
1	ALBERT + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	90.002	92.425
Nov 06, 2019			
2	ALBERT (ensemble model) Google Research & TTIC https://arxiv.org/abs/1909.11942	89.731	92.215
Sep 18, 2019			
3	XLNet + DAAF + Verifier (ensemble) PINGAN Omni-Sinitic	88.592	90.859
Jul 22, 2019			
3	ALBERT (single model) Google Research & TTIC https://arxiv.org/abs/1909.11942	88.107	90.902
Sep 16, 2019			
3	UPM (ensemble) Anonymous	88.231	90.713
Jul 26, 2019			
4	XLNet + SG-Net Verifier (ensemble) Shanghai Jiao Tong University & CloudWalk https://arxiv.org/abs/1908.05147	88.174	90.702
Aug 04, 2019			
5	XLNet + SG-Net Verifier++ (single model) Shanghai Jiao Tong University & CloudWalk https://arxiv.org/abs/1908.05147	87.238	90.071
Aug 04, 2019			
6	UPM (single model) Anonymous	87.193	89.934
Jul 26, 2019			
7	BERT + DAE + AoA (ensemble) Joint Laboratory of HIT and iFLYTEK Research	87.147	89.474
Mar 20, 2019			
7	RoBERTa (single model) Facebook AI	86.820	89.795
Jul 20, 2019			
8	RoBERTa+Span (ensemble) CW	86.651	89.595
Sep 12, 2019			
8	BERT + ConvLSTM + MTL + Verifier (ensemble) Layer 6 AI	86.730	89.286
Mar 15, 2019			
9	Xlnet+Verifier ensemble model	86.719	89.210
Oct 26, 2019			
10	BERT + N-Gram Masking + Synthetic Self-Training (ensemble)	86.673	89.147
Mar 05, 2019			

Many other neural approaches to question answering also exist!

- Many recent methods incorporate BERT embeddings
 - Contextual representations learned using Transformers

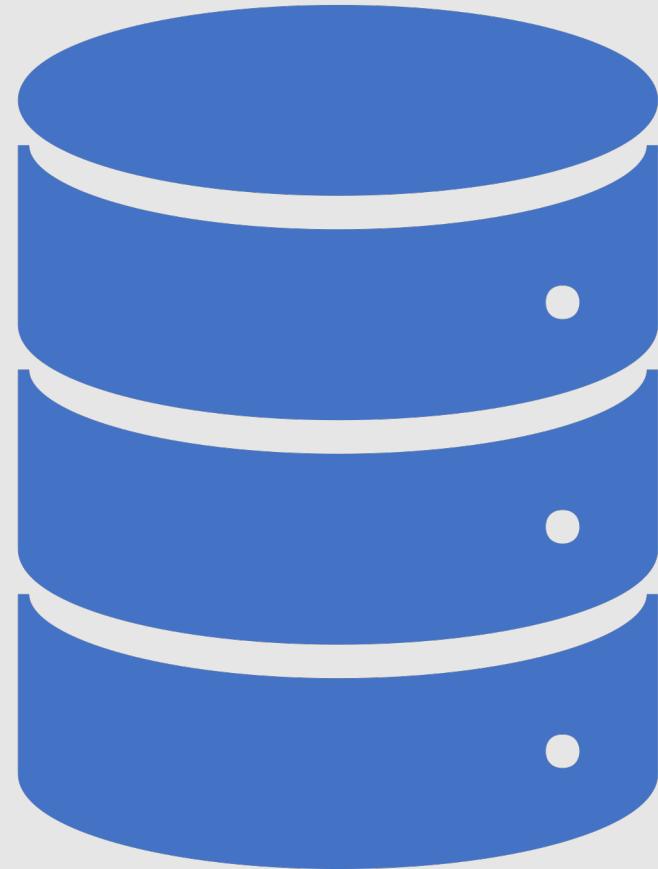
Knowledge-based Question Answering

- Answers questions by mapping them to queries over structured databases



How are text strings typically mapped to logical form?

- Semantic parsers
- Typically map text to:
 - Some form of predicate calculus (e.g., first-order logic)
 - Some type of query language
 - SQL
 - SPARQL
- This means that the question ends up either in the form of a database search query, or in a form that can be easily converted to one



What does the database look like?

- Differs depending on the resource
- Might be:
 - Full relational database
 - Simpler structured database
 - Sets of RDF (subject, predicate, object) triples
- Popular ontologies:
 - Wikidata:
https://www.wikidata.org/wiki/Wikidata:Main_Page
 - DBpedia: <https://wiki.dbpedia.org/>

Simple Knowledge-based Question Answering Task

- Answer factoid questions that ask about one of the missing arguments in a triple

subject	predicate	object
Ada Lovelace	Birth-year	1815

When was Ada Lovelace born?

Birth-year("Ada Lovelace", x)

1815

Rule-based Methods for Knowledge-based Question Answering

Write patterns to extract frequent relations

- When . + born → birth-year

Pros:

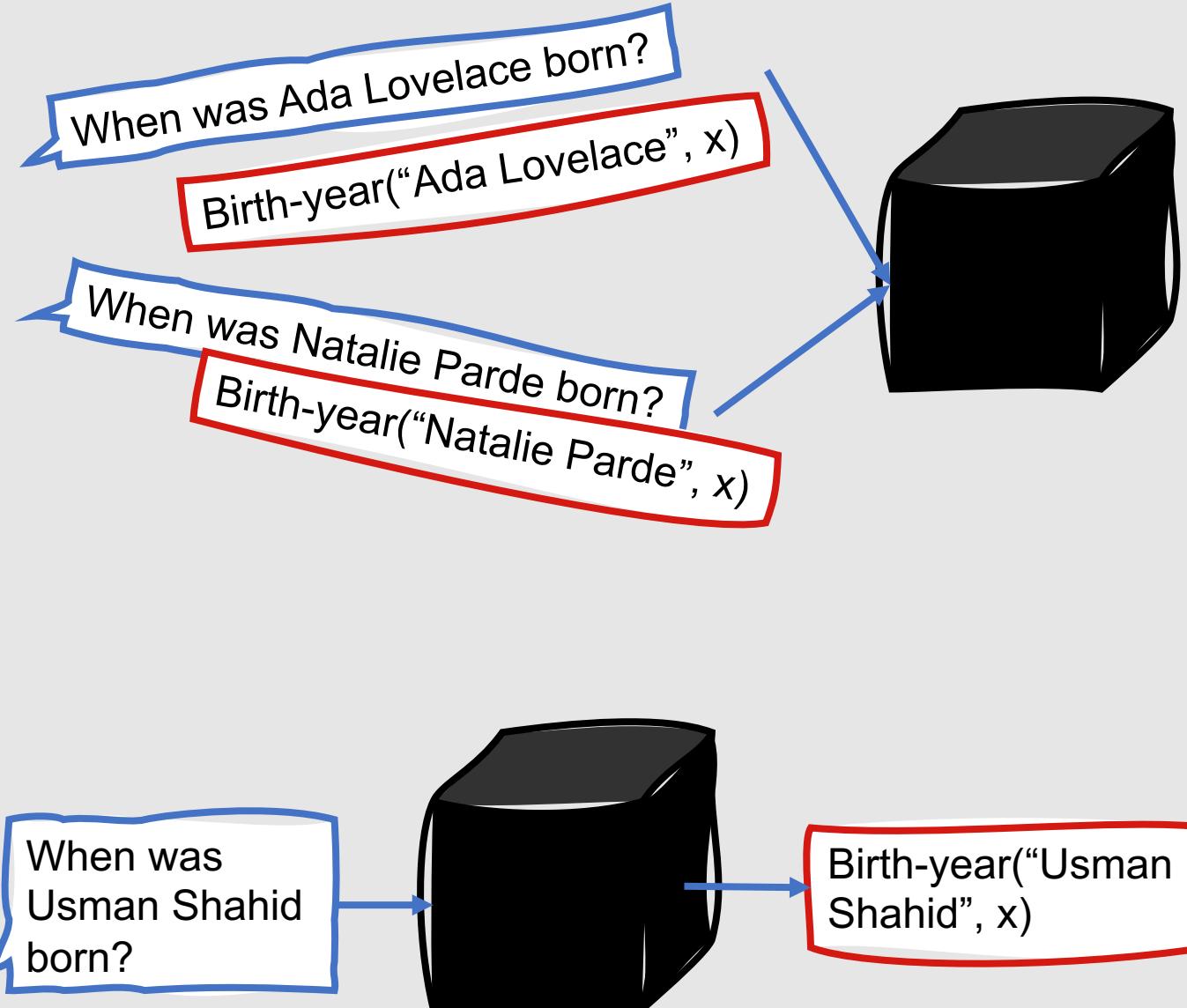
- Simple
- Precise

Cons:

- Not scalable
- Low recall

Supervised Methods for Knowledge-based Question Answering

- Learn from pairs of training questions and their correct logical forms
- Produce a system that maps from new questions to their logical forms



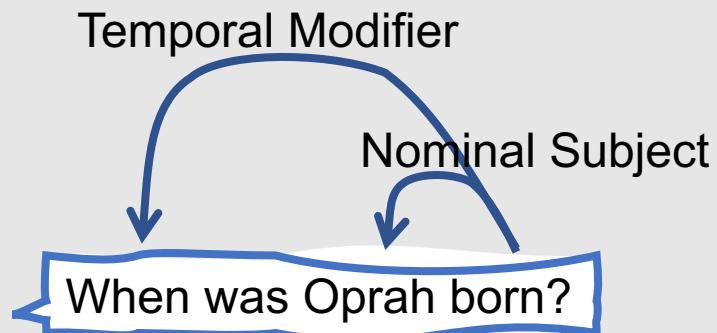
How do most systems do this?

- First, parse the questions
- Then, align the parse trees to a logical form
- Often employ **bootstrapping**
 - Small set of rules for building the mapping
 - Small initial lexicon

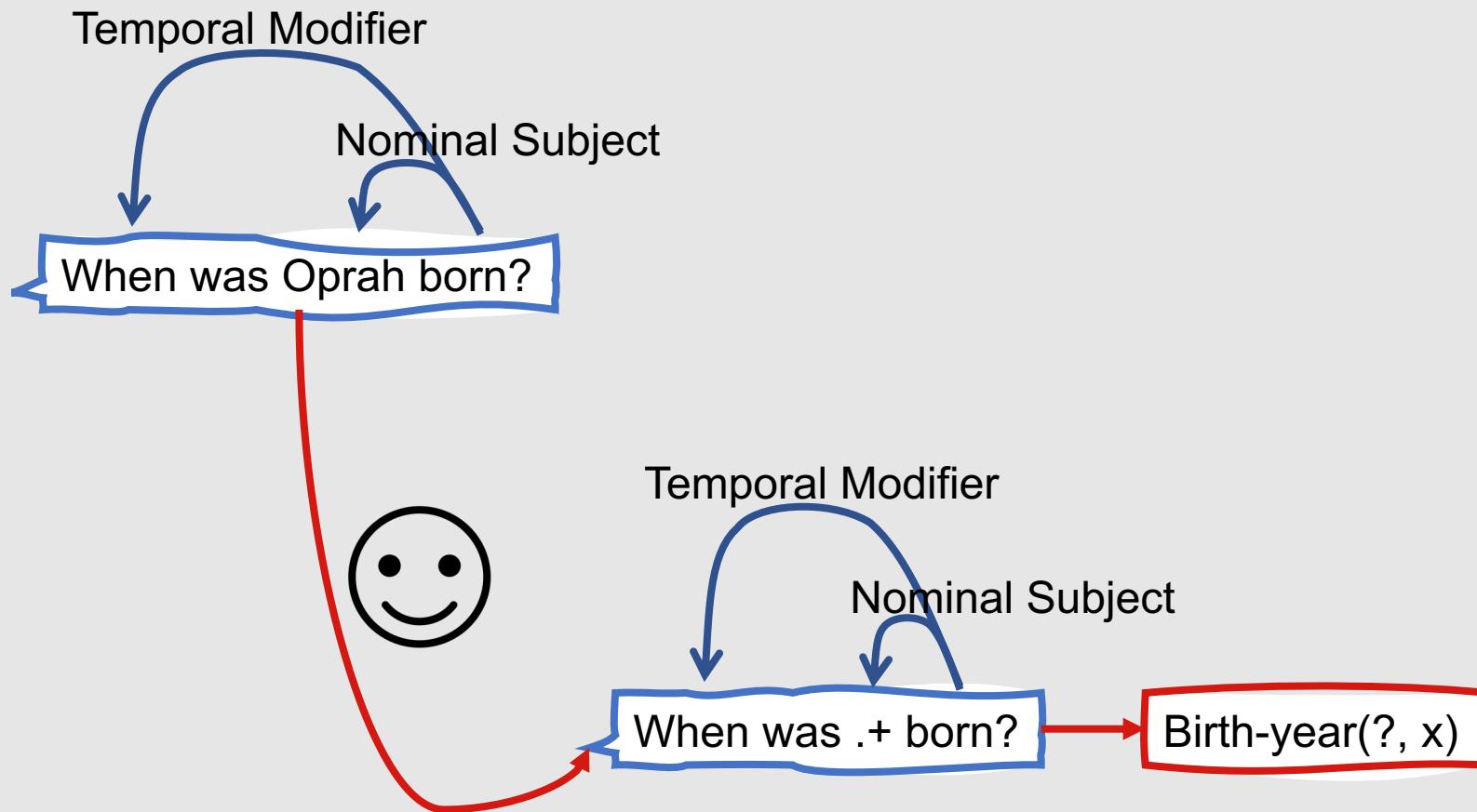
What would this look like?

When was Oprah born?

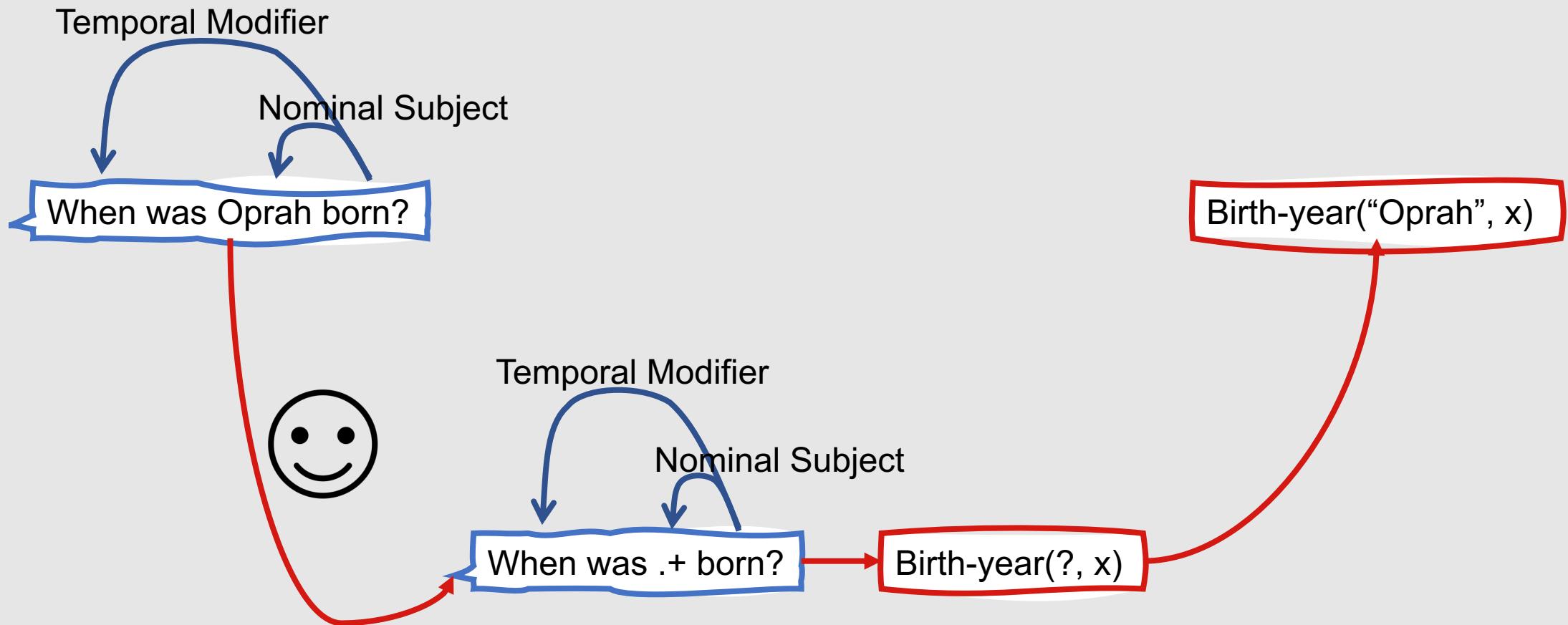
What would this look like?



What would this look like?



What would this look like?



Supervised approaches can be extended to handle more complex questions.

- More complex default rules can be used
- More complex logical forms can be used
- Training samples can be broken down into smaller tuples and then recombined to parse new sentences

What is the biggest state bordering Illinois?

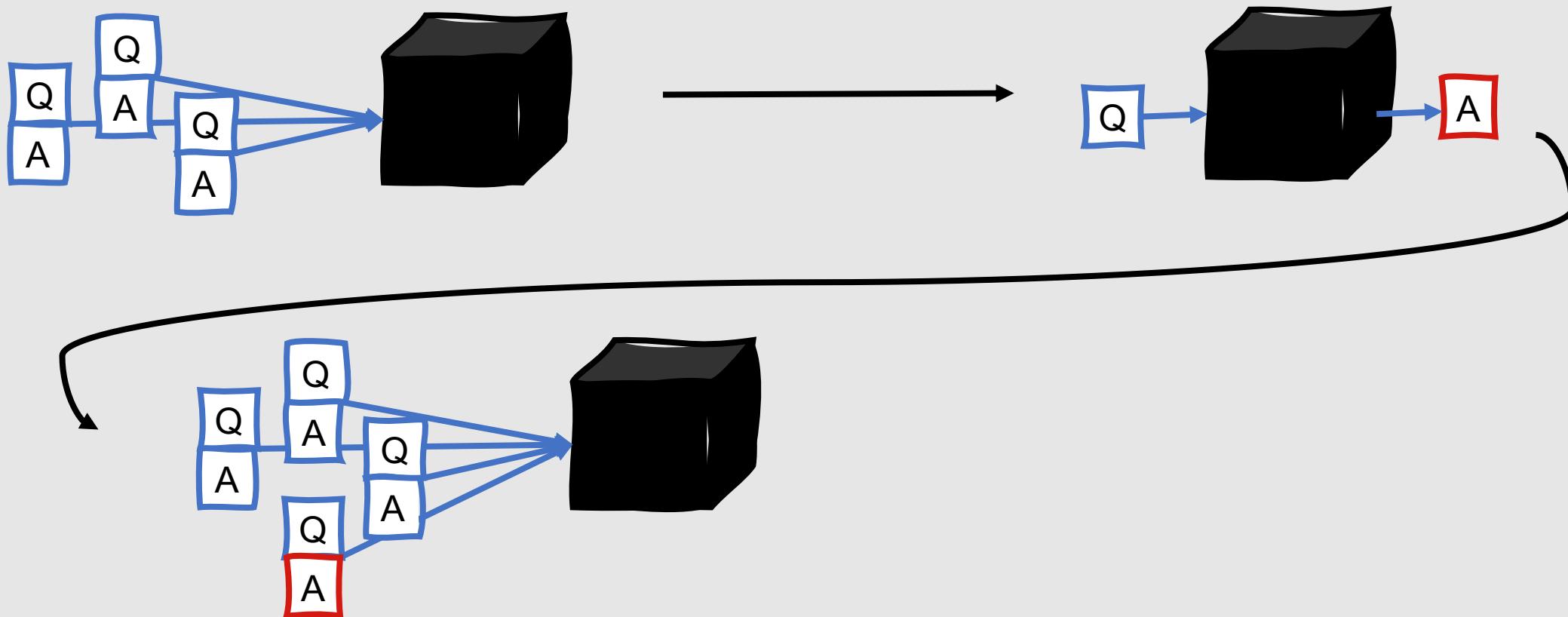
How many more undergrads are there than grad students in CS 421?

Is Chicago closer to Dallas or Denver?

Semi-Supervised Methods for Knowledge-based Question Answering

- What is semi-supervised learning?
 - A form of machine learning that makes use of both labeled and unlabeled data for training
 - Example: Bootstrapping

Semi-Supervised Methods for Knowledge-based Question Answering



Why used semi-supervised learning?

- Even though factoid questions may seem simple, it is difficult to build supervised datasets that comprehensively cover all of their different forms!

When was Oprah born?

What is Oprah's birth year?

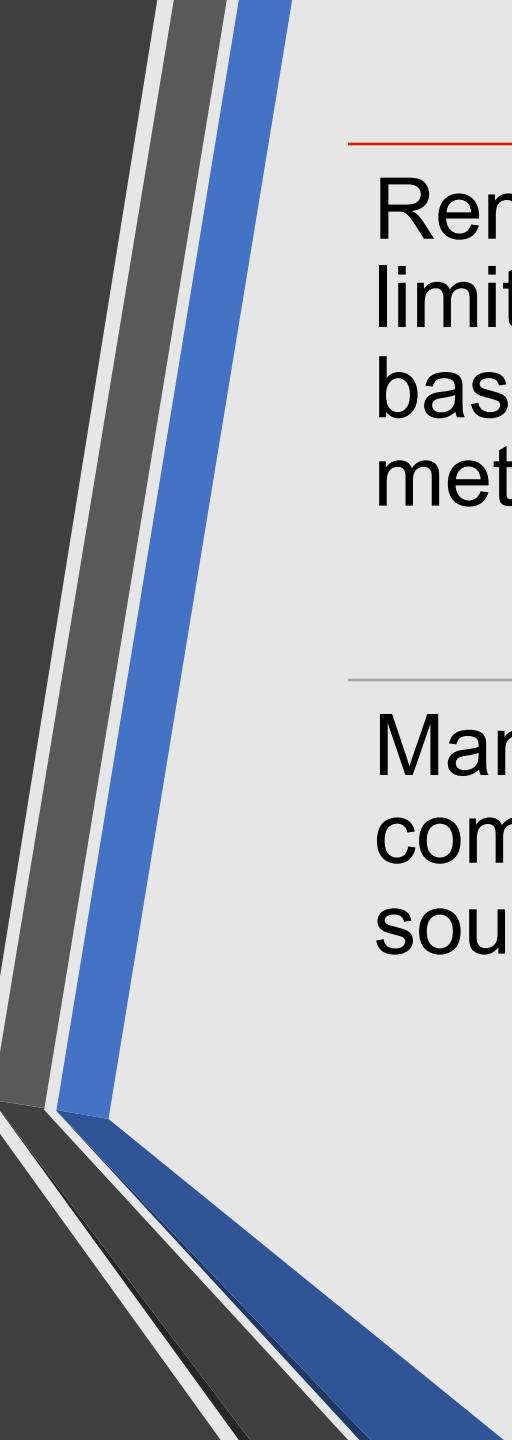
What year was Oprah born?

In what year was Oprah born?

Semi-supervised methods allow us to efficiently make use of textual redundancy.

phrase	relation	phrase	relation	phrase	relation
Capital of	Country.capital	Capital city of	Country.capital	Become capital of	Country.capital
Capitol of	Country.capital	National capital of	Country.capital	Official capital of	Country.capital
Political capital of	Country.capital	Administrative capital of	Country.capital	Beautiful capital of	Country.capital
Capitol city of	Country.capital	Remain capital of	Country.capital	Make capital of	Country.capital
Political center of	Country.capital	Bustling capital of	Country.capital	Capital city in	Country.capital
Cosmopolitan capital of	Country.capital	Move its capital to	Country.capital	Modern capital of	Country.capital
Federal capital of	Country.capital	Beautiful capital city of	Country.capital	Administrative capital city of	Country.capital

Combining Information Sources



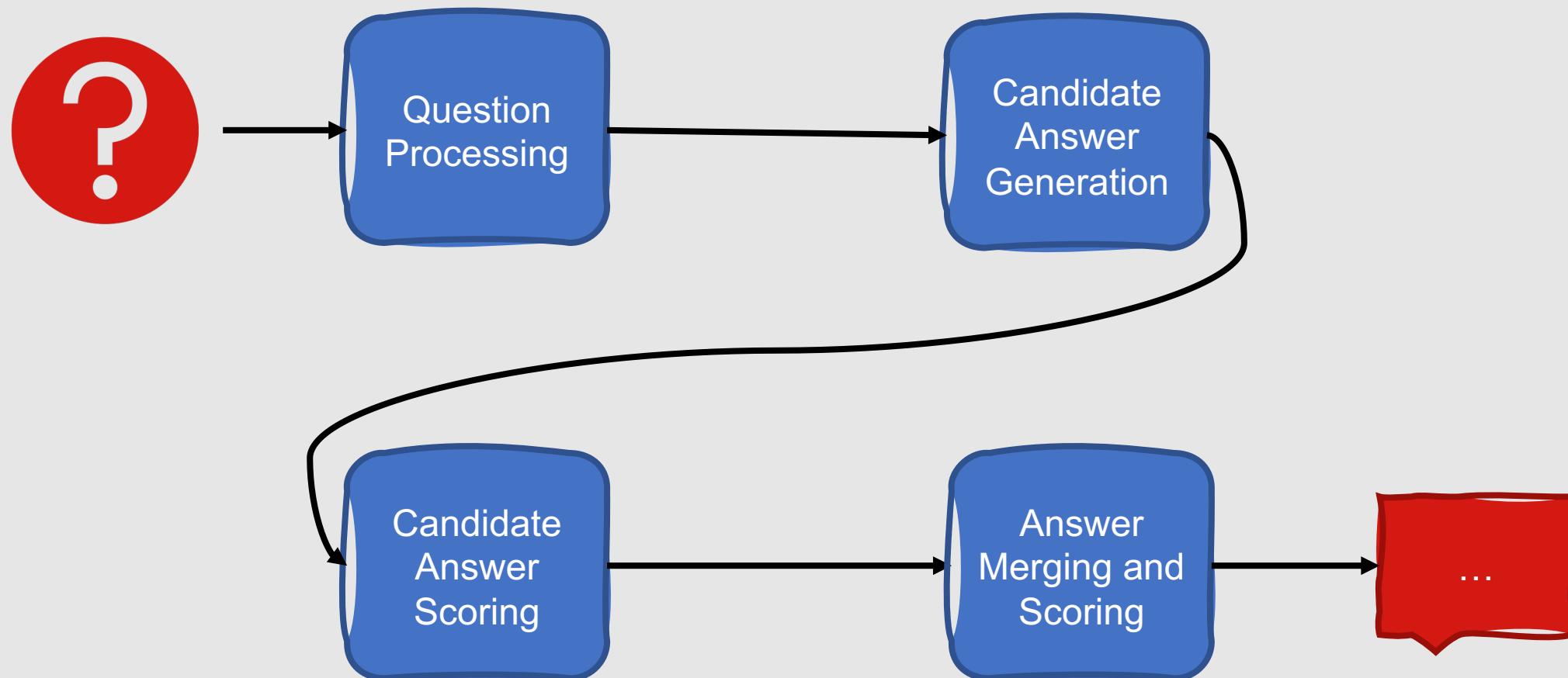
Remember ...there's no need to limit a system to using *only* text-based or *only* knowledge-based methods!

Many high-performing systems combine these two information sources

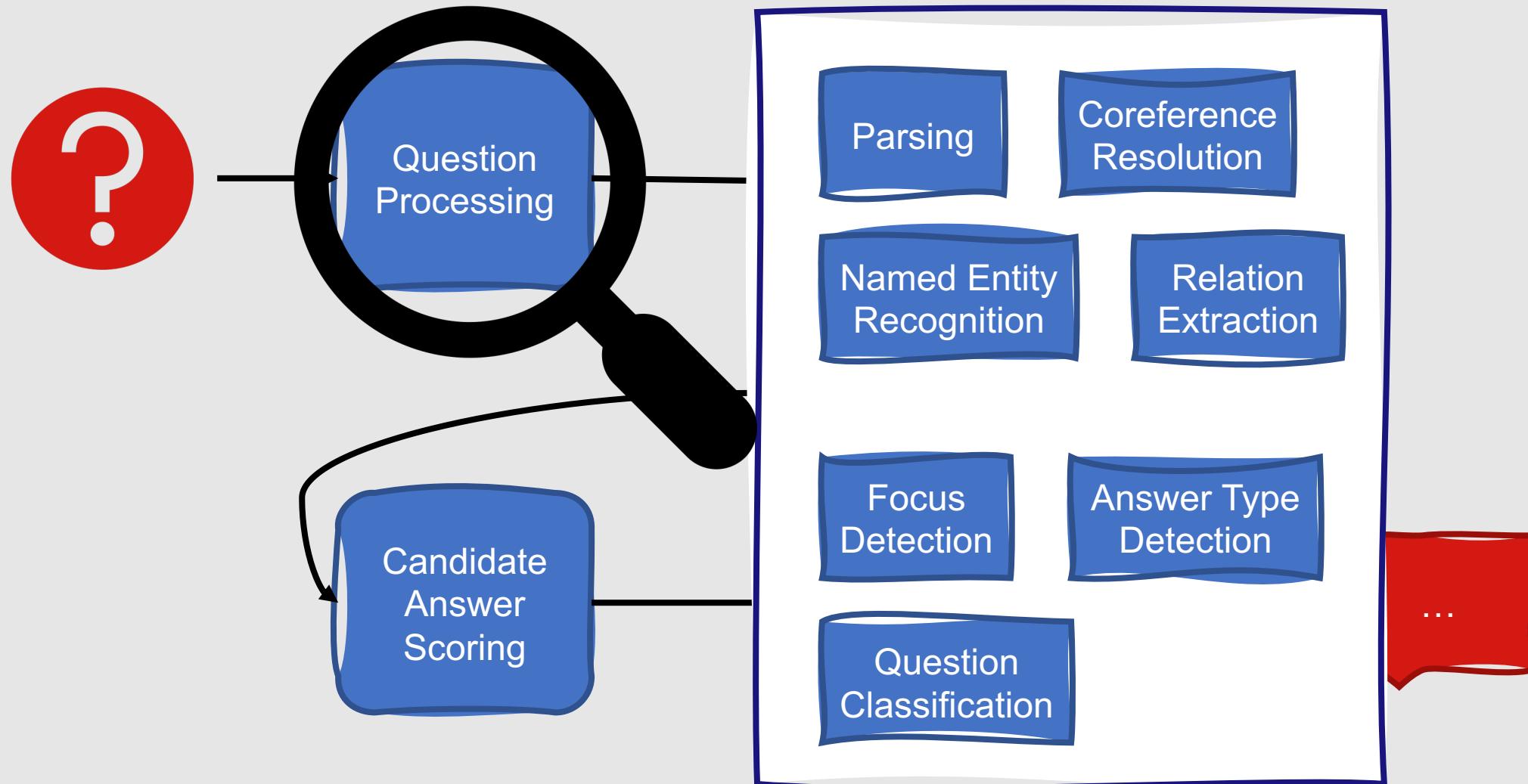
Case Example: DeepQA

- Question answering component of Watson
- Four stages:
 1. **Question processing**
 2. **Candidate answer generation**
 3. **Candidate answer scoring**
 4. **Answer merging and scoring**

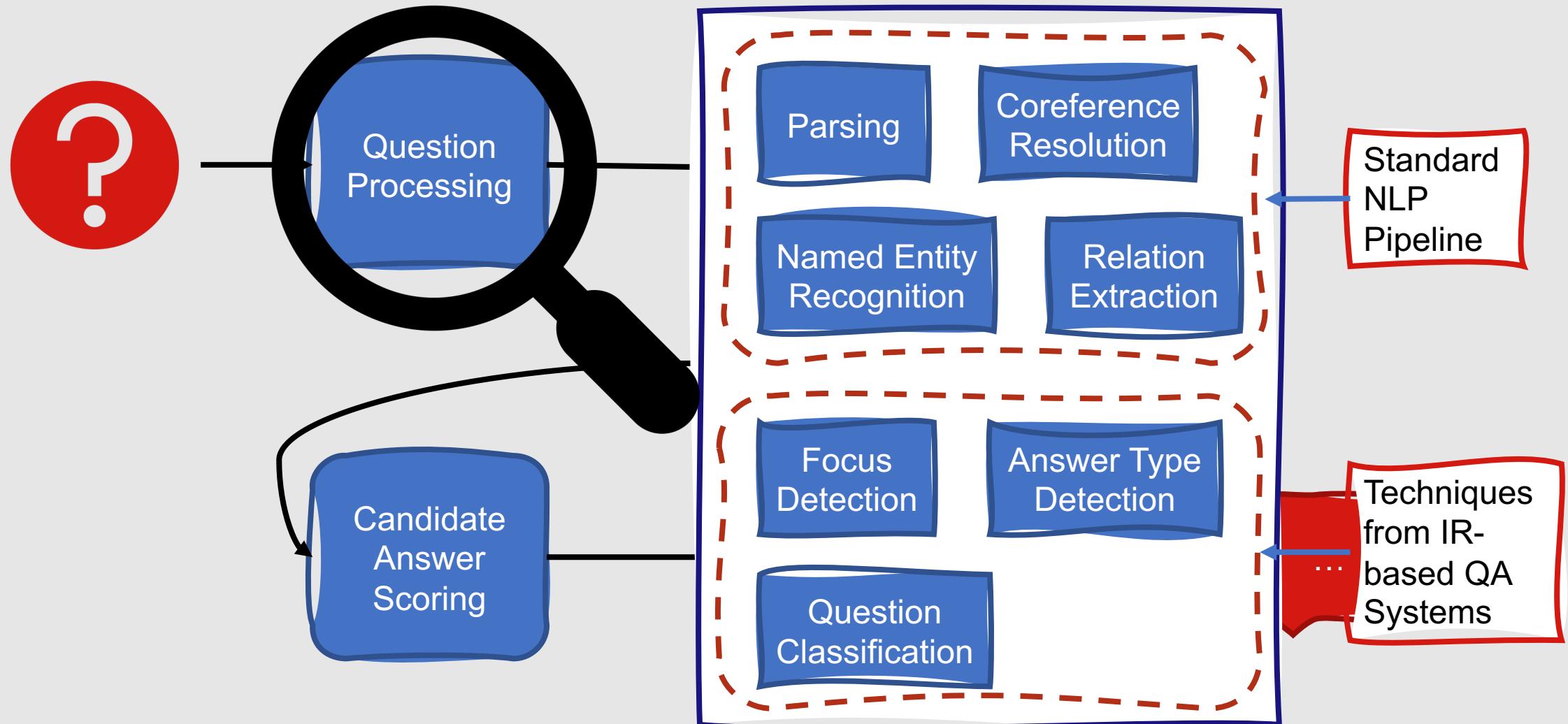
Case Example: DeepQA



Stage 1: Question Preprocessing



Stage 1: Question Preprocessing



Stage 1: Question Preprocessing

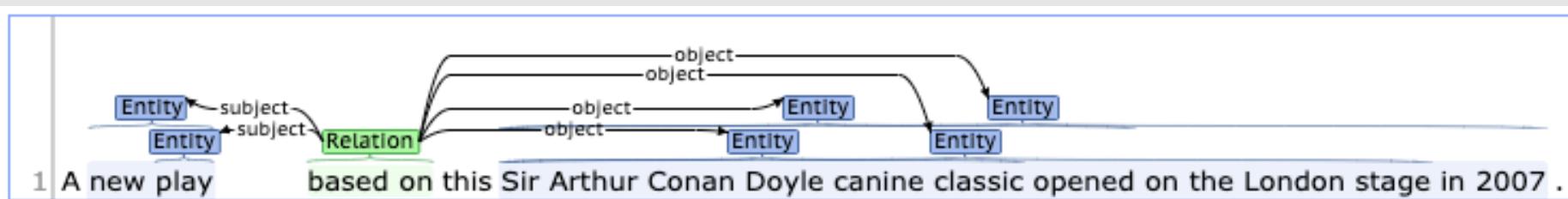
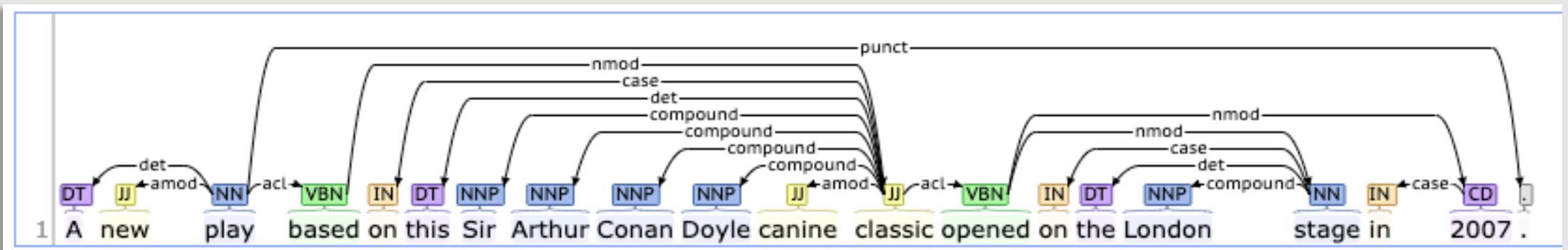
Jeopardy! Example:

A new play based on this Sir Arthur Conan Doyle canine classic opened on the London stage in 2007.

Stage 1: Question Preprocessing

Jeopardy! Example:

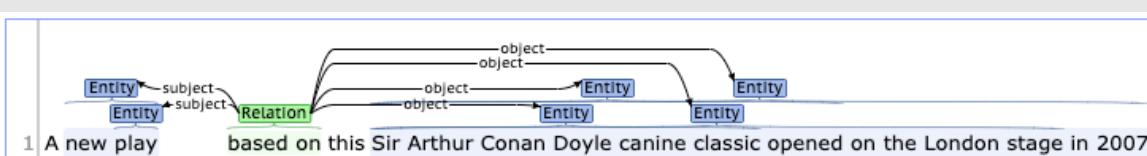
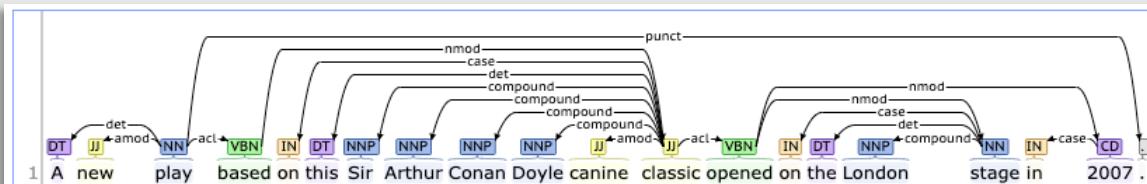
A new play based on this Sir Arthur Conan Doyle canine classic opened on the London stage in 2007.



Stage 1: Question Preprocessing

Jeopardy! Example:

A new play based on this Sir Arthur Conan Doyle canine classic opened on the London stage in 2007.



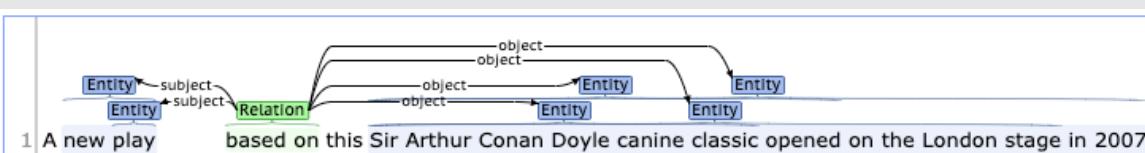
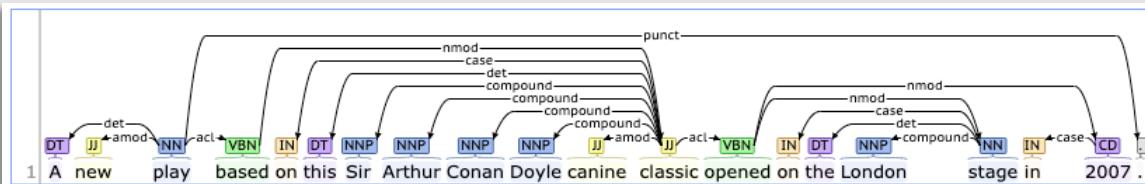
Focus Detection: Which part of the question co-refers with the answer?

Extracted using handwritten rules in DeepQA

Stage 1: Question Preprocessing

Jeopardy! Example:

A new play based on this Sir Arthur Conan Doyle canine **classic** opened on the London stage in 2007.



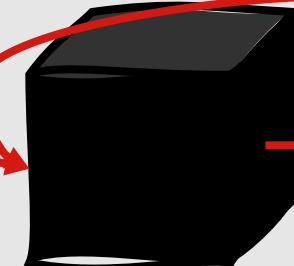
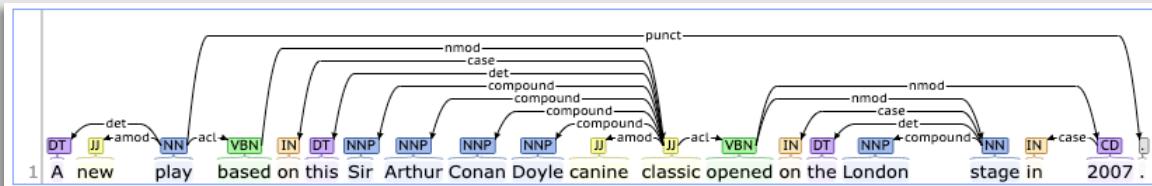
Answer Type Detection: Which word tells us about the semantic type of answer to expect?

DeepQA extracts roughly 5000 possible answer types (some questions may take multiple answer types), using a rule-based approach

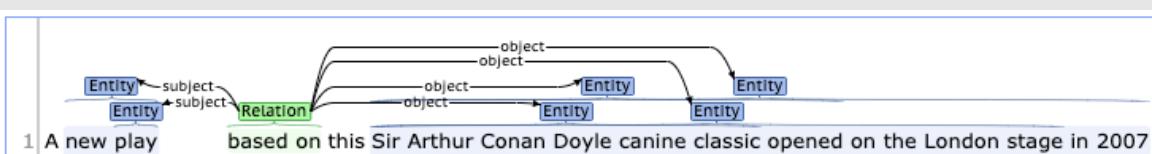
Stage 1: Question Preprocessing

Jeopardy! Example:

A new play based on this Sir Arthur Conan Doyle canine **classic** opened on the London stage in 2007.



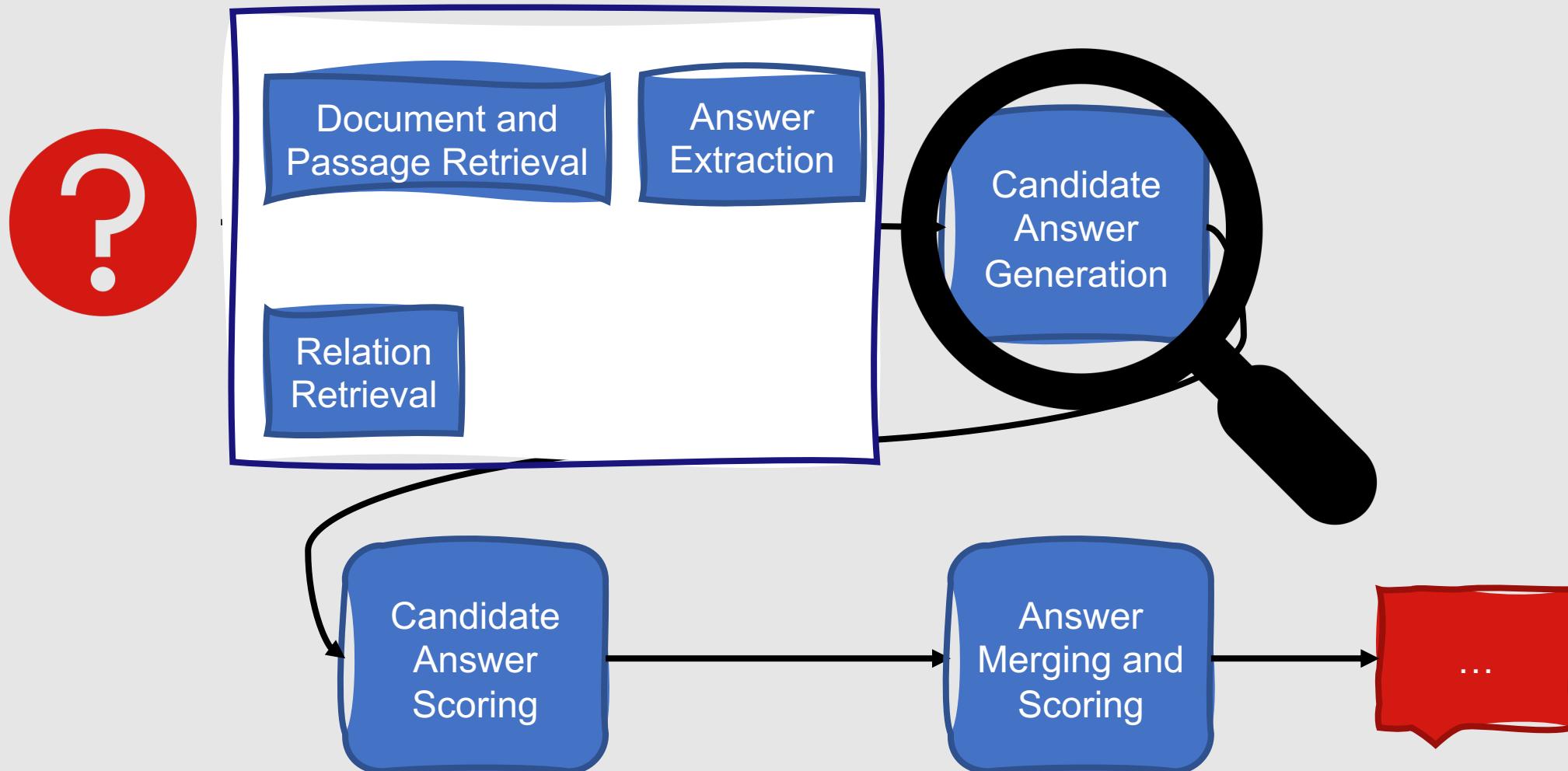
Definition



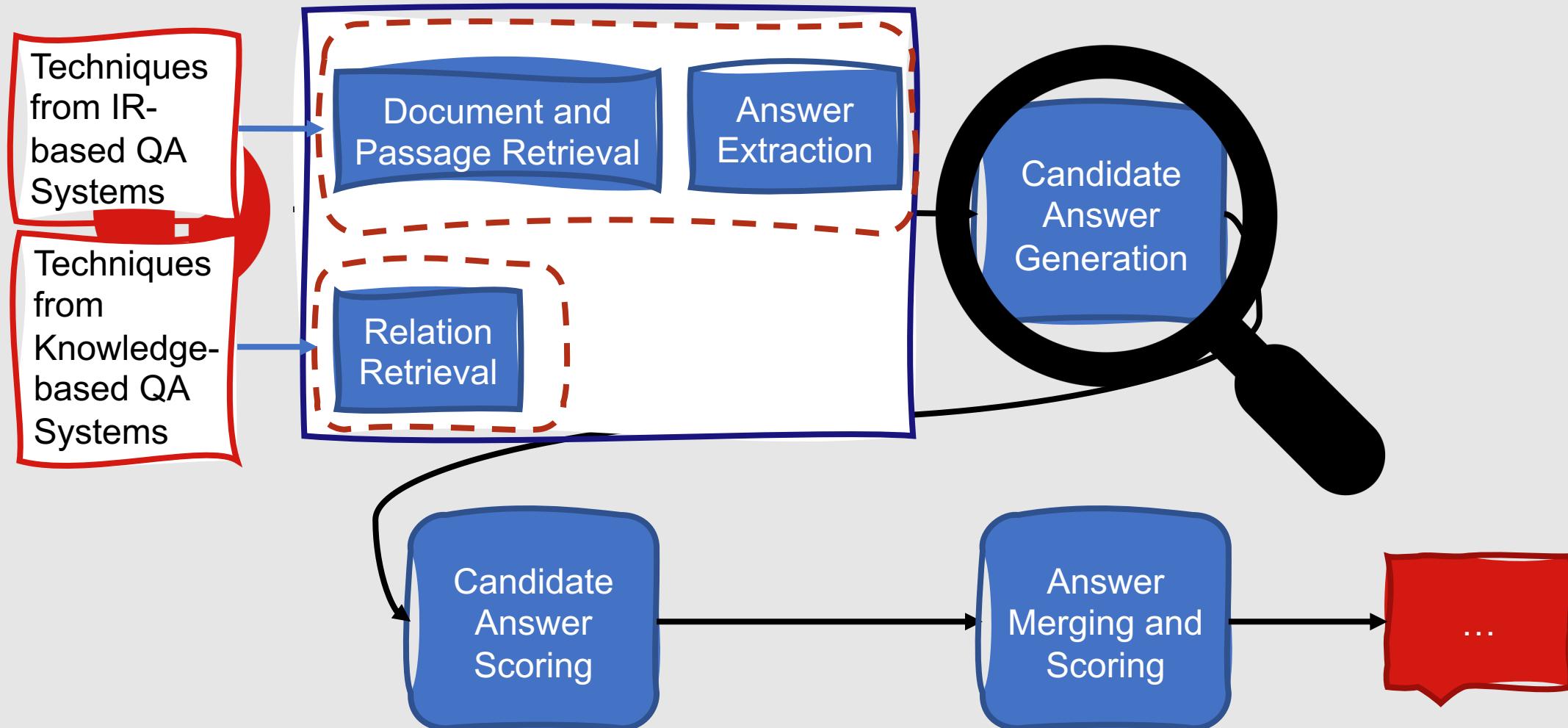
Question Classification: What type of question is this (multiple choice, fill-in-the-blank, definition, etc.)?

Generally done using pattern-matching regular expressions over words or parse trees

Stage 2: Candidate Answer Generation



Stage 2: Candidate Answer Generation



Stage 2: Candidate Answer Generation

Jeopardy! Example:

A new play based on **this Sir Arthur Conan Doyle canine classic** opened on the London stage in 2007.

Document and
Passage Retrieval

In 2007, Peepolykus Theatre Company premiered a new adaptation of *The Hound of the Baskervilles* at West Yorkshire Playhouse in Leeds.

The play is an adaptation of the Arthur Conan Doyle's novel: *The Hound of the Baskervilles* (1901).

Stage 2: Candidate Answer Generation

Jeopardy! Example:

A new play based on **this Sir Arthur Conan Doyle canine classic** opened on the London stage in 2007.

Document and Passage Retrieval

In 2007, Peepolykus Theatre Company premiered a new adaptation of *The Hound of the Baskervilles* at West Yorkshire Playhouse in Leeds.

The play is an adaptation of the Arthur Conan Doyle's novel: *The Hound of the Baskervilles* (1901).

Answer Extraction

The Hound of the Baskervilles

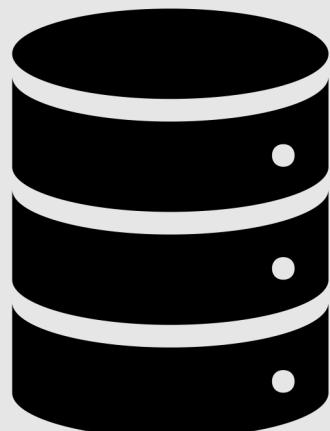
The Hound of the Baskervilles (1901)

Stage 2: Candidate Answer Generation

Jeopardy! Example:

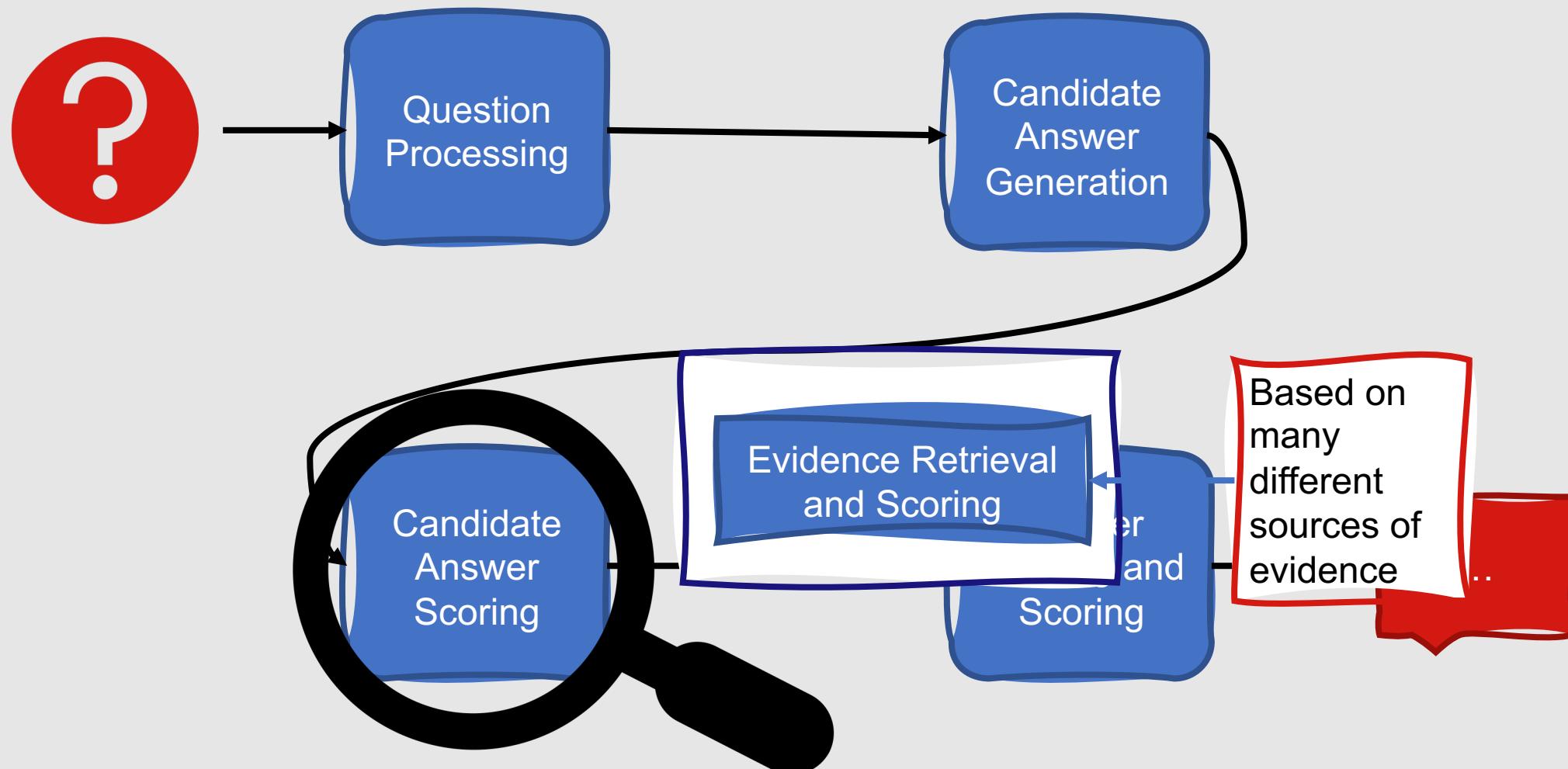
basedOn(x, "Sir Arthur Conan Doyle canine classic")

Relation Retrieval



The Hound of the Baskervilles

Stage 3: Candidate Answer Scoring



Stage 3: Candidate Answer Scoring

The Hound of the Baskervilles

The Hound of the Baskervilles

The Hound of the Baskervilles (1901)

Stage 3: Candidate Answer Scoring

The Hound of the Baskervilles

Expected Answer Type: BOOK

The Hound of the Baskervilles

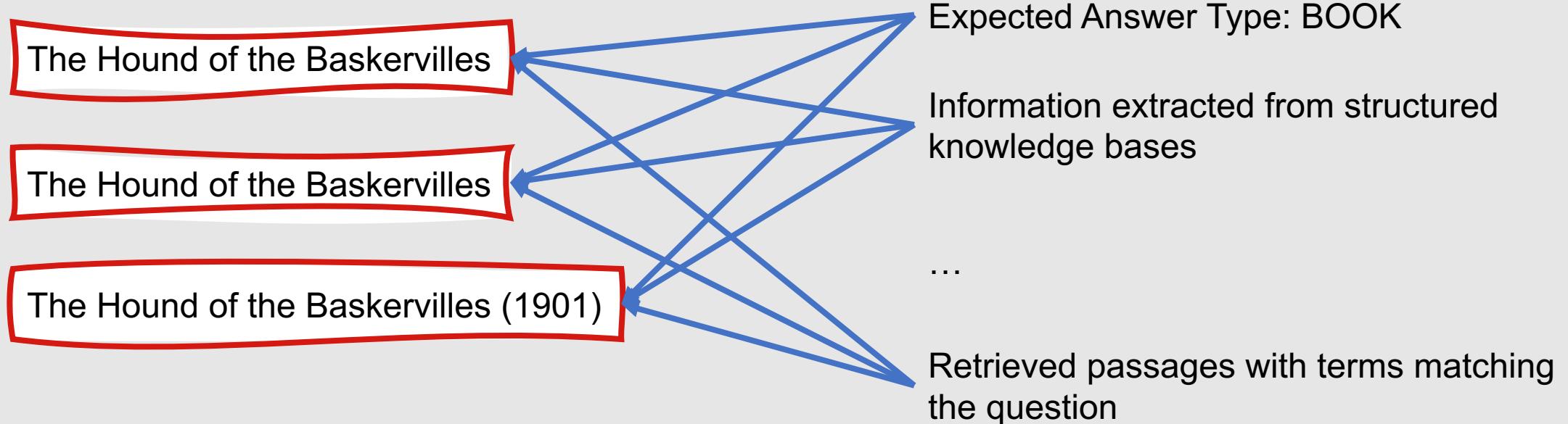
Information extracted from structured knowledge bases

The Hound of the Baskervilles (1901)

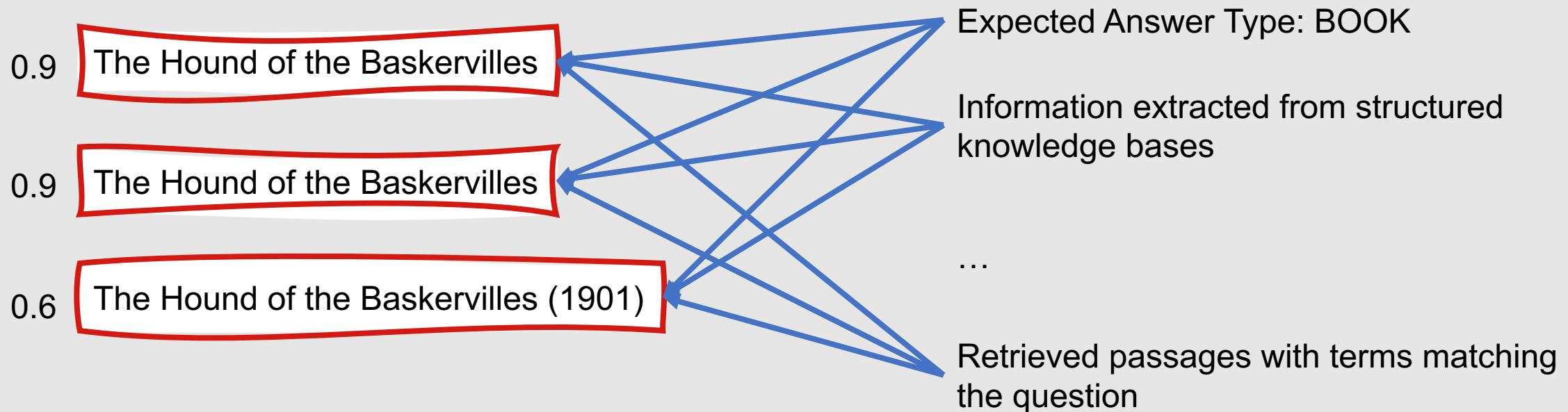
Retrieved passages with terms matching the question

...

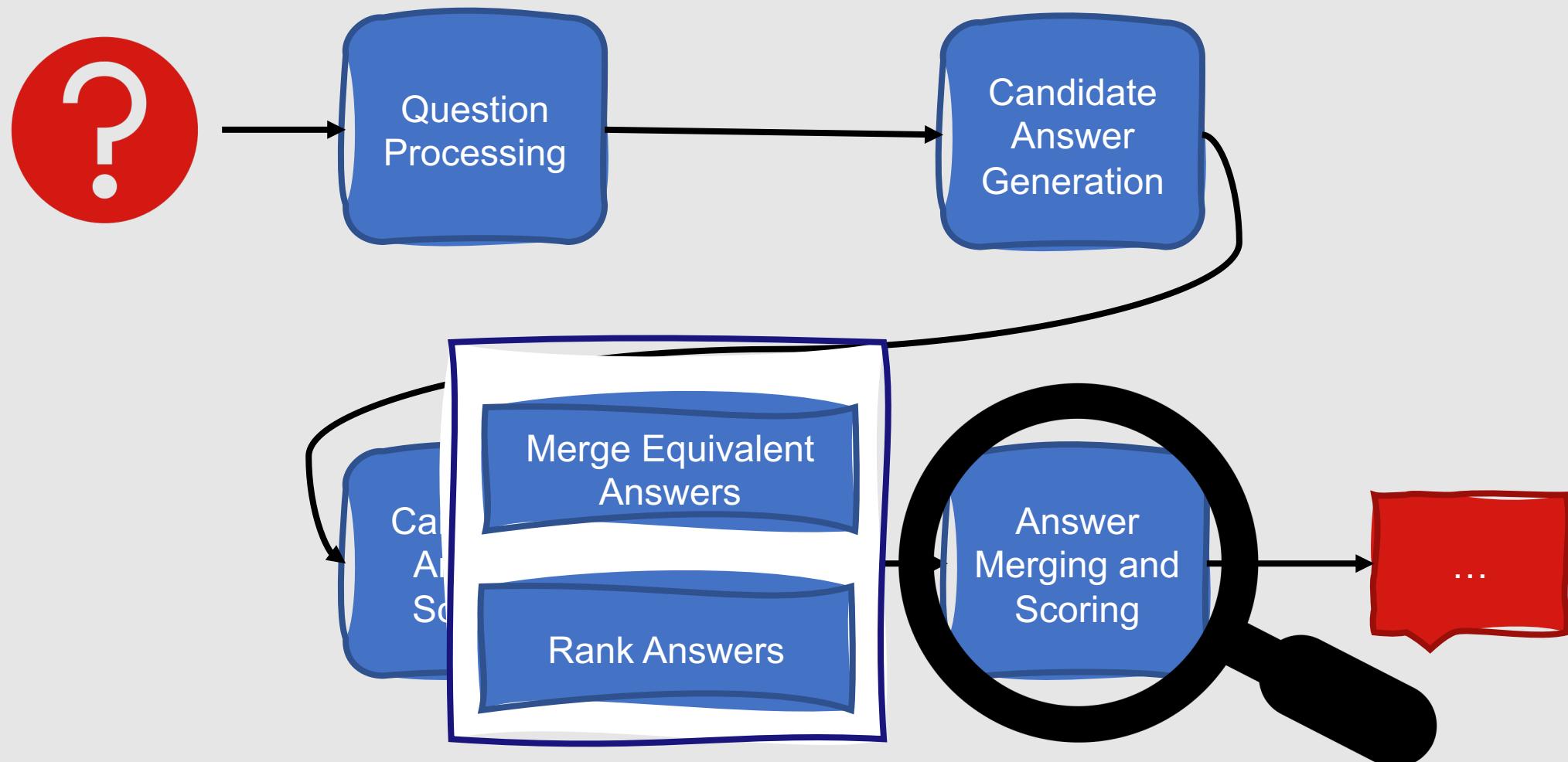
Stage 3: Candidate Answer Scoring



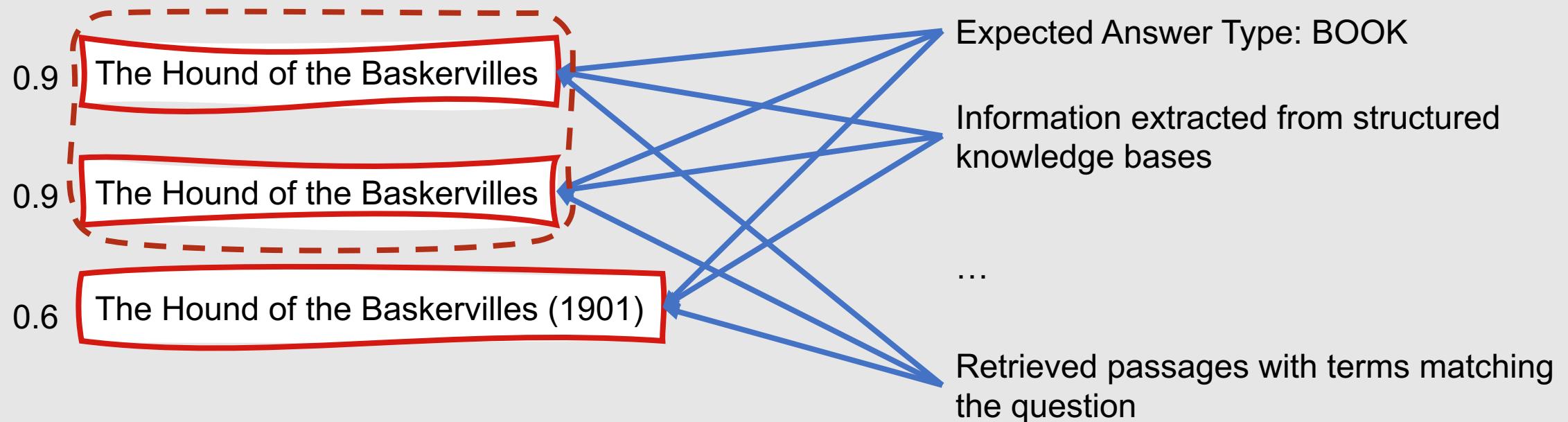
Stage 3: Candidate Answer Scoring



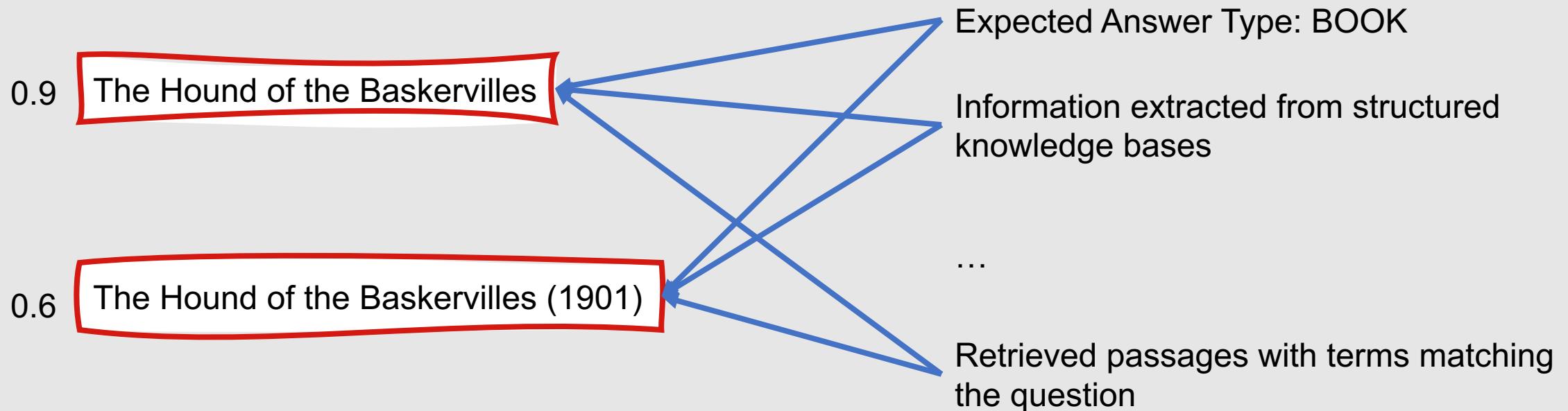
Stage 4: Answer Merging and Scoring



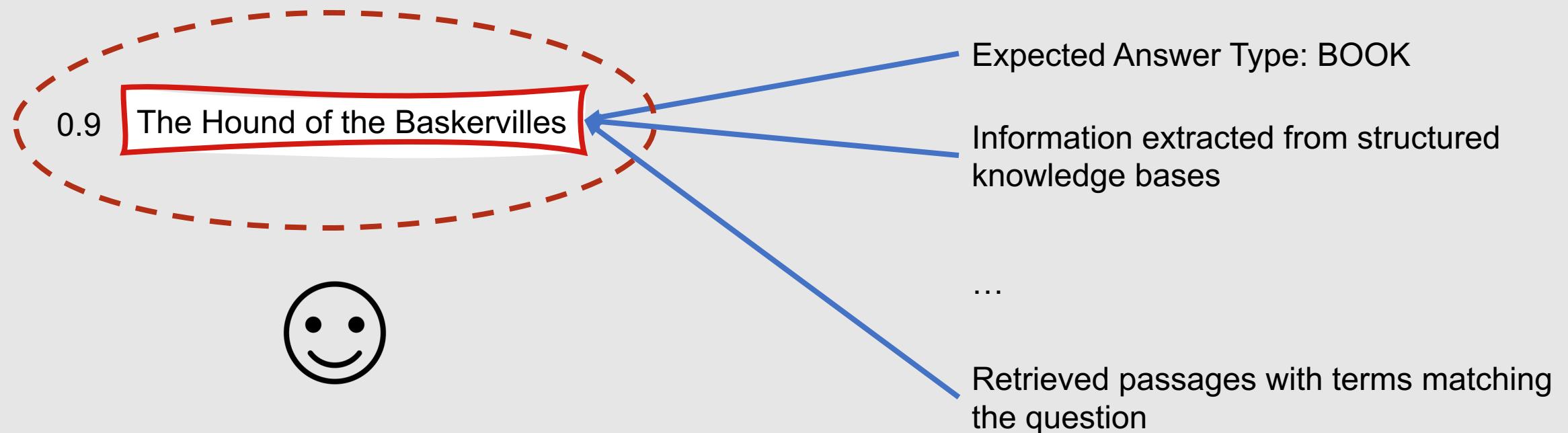
Stage 4: Answer Merging and Scoring



Stage 4: Answer Merging and Scoring



Stage 4: Answer Merging and Scoring



Watson is just one of many question answering architectures!

- Most high-performing QA systems will follow the same intuition:
 - Propose a large number of candidate answers using both IR-based and knowledge-based techniques
 - Develop a variety of IR-based and knowledge-based features to score the candidates

Summary: Question Answering (Part 1)

- **Question answering** is the process of automatically retrieving short spans of correct, relevant information in response to a user's **query**
- Most question answering systems focus on **factoid** questions
- There are two major types of question answering systems:
 - **Information retrieval-based**
 - **Knowledge-based**
- These two types of question answering systems are often combined, as seen in Watson's DeepQA architecture