

# Reminders



FINAL QUIZ ON CHAPTER 22  
(COREF) AND 25 (QA) IS DUE  
BY MIDNIGHT TONIGHT



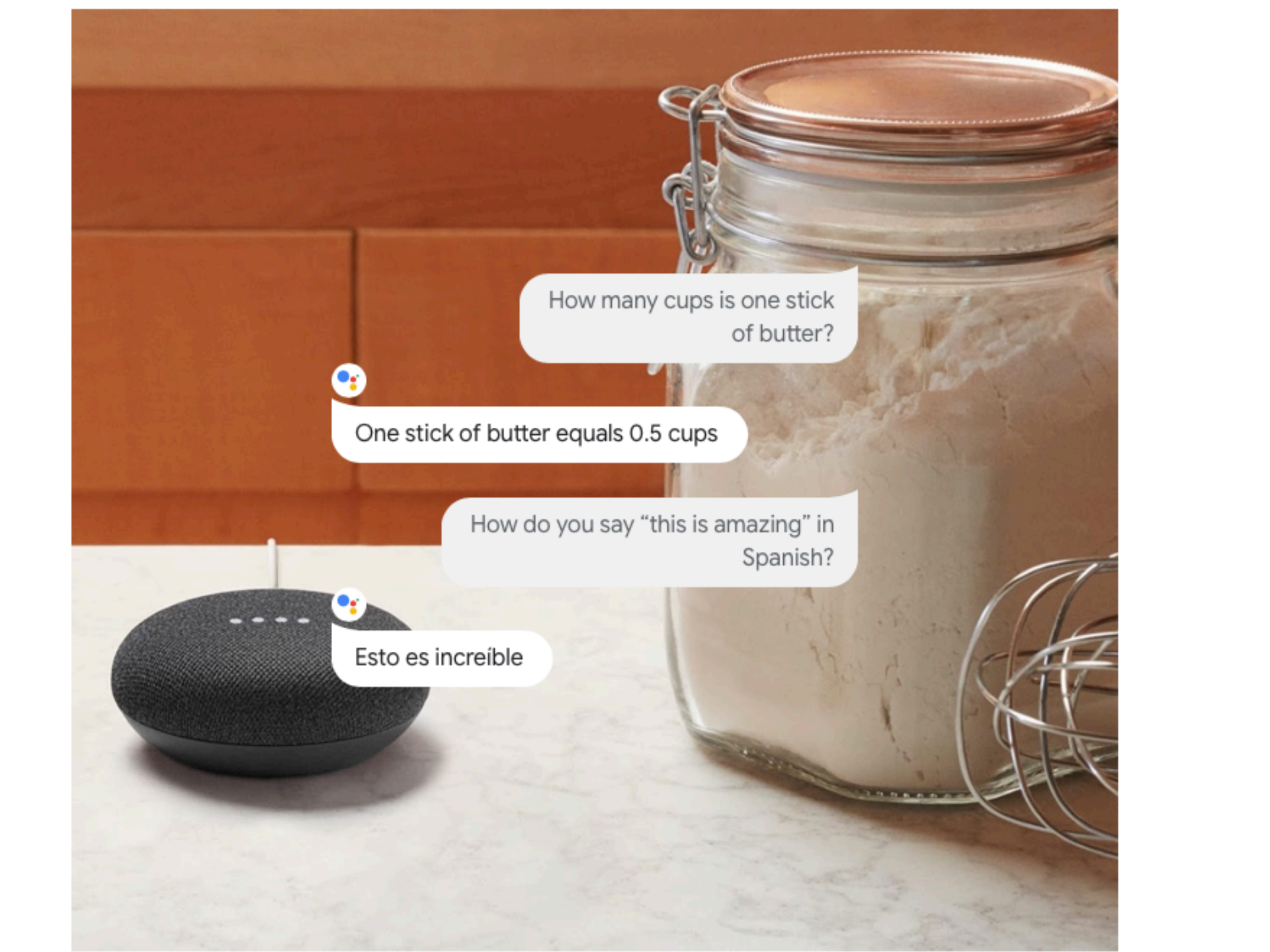
HOMEWORK 12 /  
MILESTONE 4 ARE DUE DUE  
WEDNESDAY.



SIGN-UP FOR A 30-MINUTE  
SLOT TO PRESENT YOUR  
FINAL PROJECT AT  
[ccb.youcanbook.me](http://ccb.youcanbook.me)

# Question Answering

JURAFSKY AND MARTIN CHAPTER 22

A kitchen scene featuring a black smart speaker on a white marble countertop. To the right is a large glass jar with a metal clasp lid, filled with butter. A metal whisk is partially visible next to the jar. The background is a wooden wall. Three speech bubbles are overlaid on the image, showing a conversation between a user and a smart assistant.

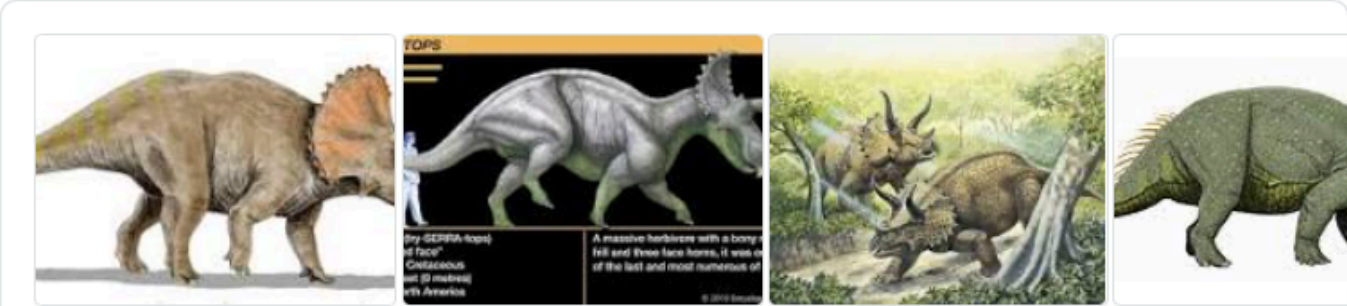
How many cups is one stick of butter?

One stick of butter equals 0.5 cups

How do you say "this is amazing" in Spanish?

Esto es increíble

About 874,000 results (0.58 seconds)



## three horns

Triceratops means "Three Horned Face" in Greek. It is called that because it has **three horns** – two above its eyes and one on its nose. Triceratops horns could be up to 115 cm (45 inches) long. It had a big frill that extended up above the head.

[en.wikibooks.org](https://en.wikibooks.org/wiki/Triceratops) › wiki › Triceratops ▾

[Wikijunior:Dinosaurs/Triceratops - Wikibooks, open books for ...](#)

[? About Featured Snippets](#) [Feedback](#)

### People also ask

What is a Triceratops with one horn? ▾

About 42,600,000 results (0.58 seconds)

Snails / Number of Teeth

**Yes, snails can have between 1000 to 12,000 or more teeth.**

[Feedback](#)

People also ask

Do snails have 14000 teeth?

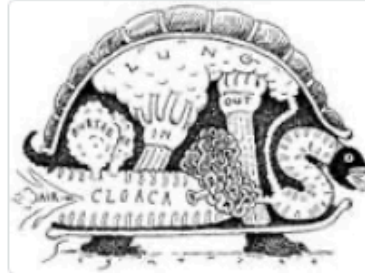
How many teeth do snails have in total?

Do snails have 25000 teeth?

What does a snail teeth look like?

[Feedback](#)

About 326,000 results (0.43 seconds)



View all

**Turtles Breathe Out of Their Butt.** Technically the term is cloacal respiration, and it's not so much **breathing** as just diffusing oxygen in and carbon dioxide out, but the fact remains: when **turtles** hibernate, **their** main source of oxygen is **through their butt**.

Dec 28, 2017

[www.mcgill.ca](http://www.mcgill.ca) › [oss](#) › [article](#) › [did-you-know](#) › [turtles-...](#)

[Turtles Breathe Out of Their Butt | Office for Science and ...](#)

🔍 About Featured Snippets

🗨️ Feedback

### People also ask

Can turtles breathe with their Buttholes? ▼

Do turtles poop out their mouths? ▼

# Feedback Survey for CIS 530 - Spring 2020

Did you enjoy the class?

- Yes
- No

Are you a undergraduate or graduate student?

- Undergraduate
- Undergrad + Submat
- Master's
- PhD
- Other: \_\_\_\_\_

What is your major / graduate degree?

# Factoid Question Answering paradigms

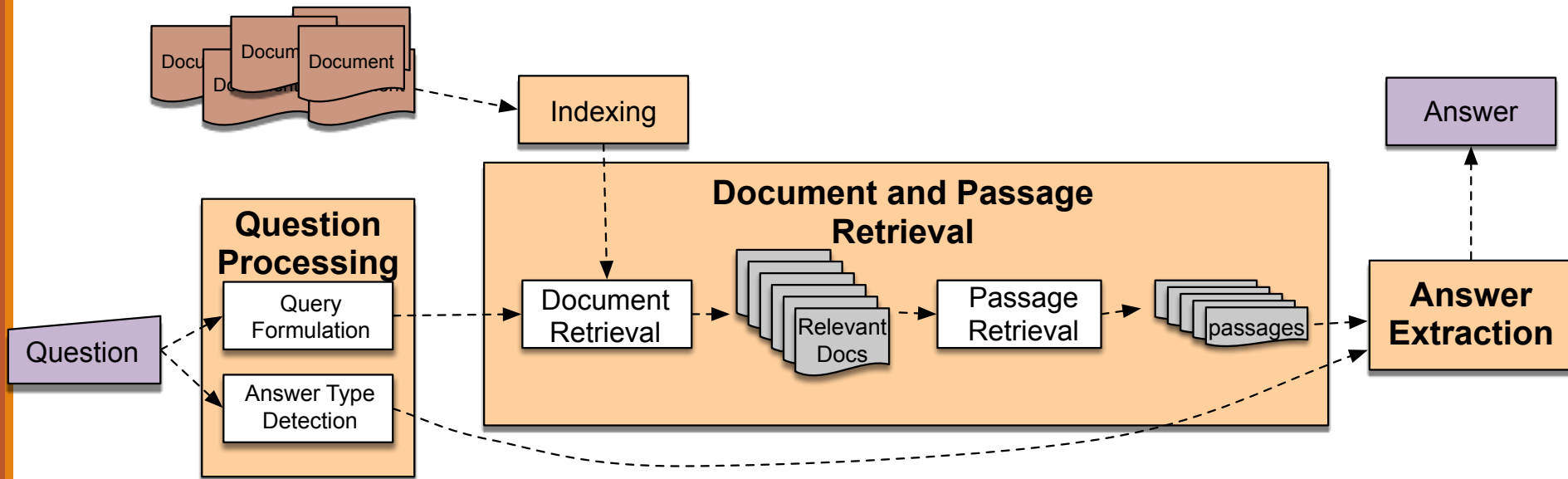
- Information Retrieval based
  - Relies on vast textual information
  - First stage is finding relevant documents and passages
  - Draw an answer
- Knowledge based
  - Build semantic representation of query
  - Query database of facts
- Industrial systems like Watson use a hybrid of both approaches



# Information Retrieval-based Question Answering

Question	Answer
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
What's the official language of Algeria?	Arabic
How many pounds are there in a stone?	14

# Pipeline for IR based factoid answering



# Question Processing

**answer type:** the entity type (person, location, time, etc.)

**focus:** the string of words in the question that are likely to be replaced by the answer in any answer string found.

**question type:** is this a definition question, a math question, a list question?

*Which US state capital has the largest population?*

**query:** “US state capital has the largest population”

**answer type:** city

**focus:** state capital

# Query formulation

Goal: create a query to send to an information retrieval system to retrieve documents that might contain answer strings.

Input	Reformulation Rule	Output
<i>when was the laser invented?</i>	<i>wh-word did A verb B → ...A verb+ed B</i>	<i>the laser was invented</i>
<i>where is the Valley of the Kings?</i>	<i>wh-word did A verb B → ...A verb+ed B</i>	<i>the Valley of the Kings is located in</i>

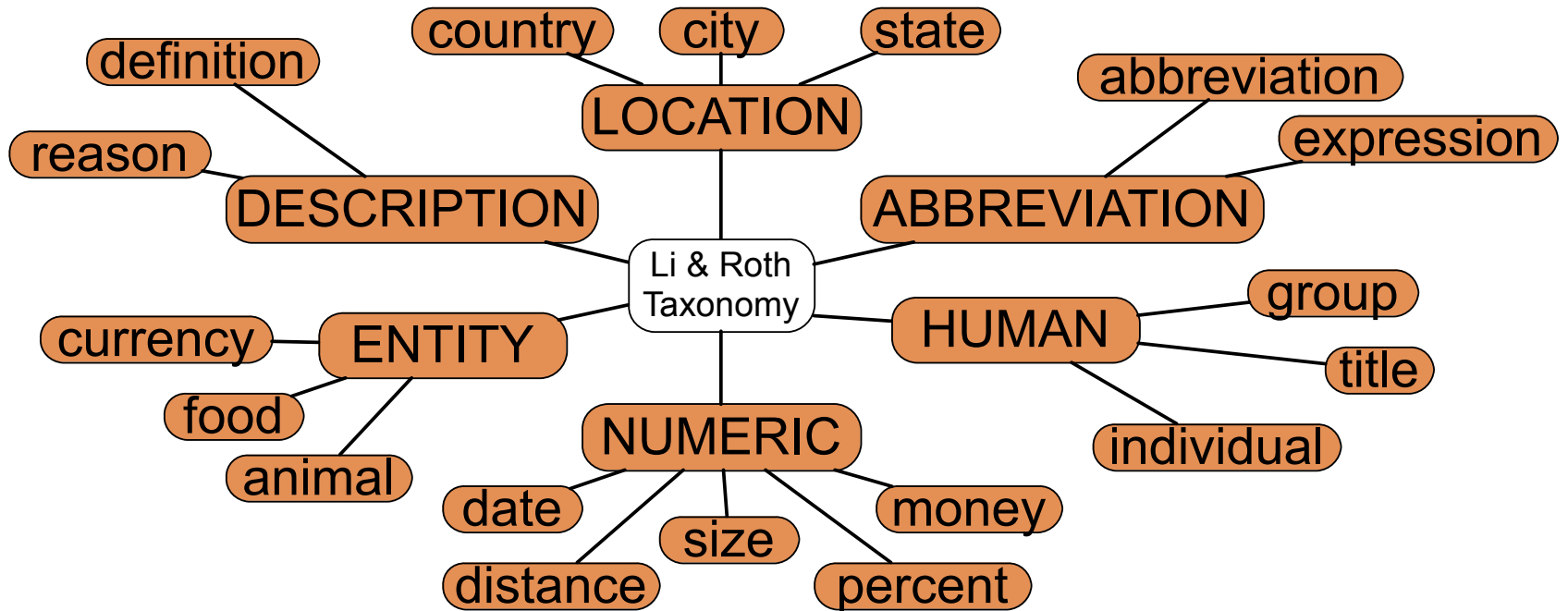
# Answer types

Some systems make use of **question classification**, which is the the task of finding the **answer type**, which is the category of the answer, often based on named entity types.

Question	Expected Answer Type
Who founded Virgin Airlines?	PERSON
Who is Elon Musk?	BIOGRAPHY
Where is the Statue of Liberty located?	LOCATION
What is the state with the largest population?	STATE
What is the state flower of California?	FLOWER

**THE 2018  
SUPER BOWL  
VICTORY PARADE  
ENDED AT  
THE MUSEUM STEPS  
MADE FAMOUS IN  
THIS 1976 MOVIE**

# Answer types



Entity Type	Example
animal	What are kind of birds did the god Odin have?
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?
language	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?



# Document and Passage Retrieval

The question processing stage creates a query to send to an IR system, which returns a set of documents ranked by their relevance to the query.

Perform **passage retrieval** by divide the top ranked documents into smaller passages. Pass these along to the answer extraction phase or filter them based on answer type classification.

In web search, passage retrieval attempts to extract the **snippets** from the retrieved pages so that the relevant passages can be displayed to the user.

# Features used to rank passages

1. The number of named entities of the correct type in the passage
2. The number of question keywords in the passage
3. The longest exact sequence of question keywords that occurs in the passage
4. The rank of the document from which the passage was extracted
5. The proximity of the keywords from the original query to each other
6. The number of n-grams that overlap between the passage and the question

# Answer extraction

After we have performed passage retrieval, we need to extract a specific answer from a passage. This task is commonly modeled by **span labeling**.

A **simple baseline** is to perform NER and retrieve the span with the correct answer type.

*Who is the prime minister of India?*

Prime minister **Narendra Modi** ordered a complete lockdown for India's 1.3 billion people to contain the India's coronavirus outbreak.

*How tall is Mt. Everest?*

The official height of Mount Everest is **29029 feet**

# Answer extraction

Many questions, like DEFINITION questions, do not have a simple named entity as their answer type. Therefore, modern systems tend to use supervised learning.

Pattern	Question	Answer
<b>&lt;answer phrase&gt;</b> such as <i>&lt;question phrase&gt;</i>	What is <i>autism</i> ?	... , <b>developmental disorders</b> such as <i>autism</i> ...
<i>&lt;question phrase&gt;</i> , a <b>&lt;answer phrase&gt;</b>	What is a <i>caldera</i> ?	The Long Valley <i>caldera</i> , a <b>volcanic crater</b> 19 miles long

# Feature-based approaches

Feature	Definition
Answer type match	True if the candidate answer contains a phrase with the correct answer type.
Pattern match	The identity of a pattern that matches the candidate answer
Num matched keywords	How many question keywords are in the candidate answer?
Keyword distance	The distance between the candidate answer and query keywords
Novelty factor	True if at least one word in the candidate answer is not in the query
Apposition features	True if the candidate answer is an appositive to a phrase containing many question terms
Punctuation location	True if the candidate answer is followed punct.
Longest sequences	The length of the longest sequence of question terms that occurs in the candidate answer.
Answer redundancy	How many other candidate passages contain this answer?

# N-gram tiling

Relying on the redundancy of the web, this method begins with the snippets returned from the web search engine, produced by a reformulated query.

## First step

- Every unigram, bigram and trigram occurring in snippet is extracted and weighted
- Weight is a function of number of snippets in which the n-gram occurred, and weight of query reformulation pattern that returned it.

Filtering, based on match to predicted answer type.

## N-gram tiling algorithm

- Concatenate overlapping n-gram fragments into longer answers
- Greedy method: start with highest scoring candidate and try to tile each other candidate. Best scoring concatenation is added to set of candidates.

# Neural Answer Extraction

Given a question  $q$  with  $l$  tokens and a passage  $p$  with  $m$ , compute for each token  $p_i$  in the passage the probability it is the start of the answer span, and the the end of the answer span.

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**"

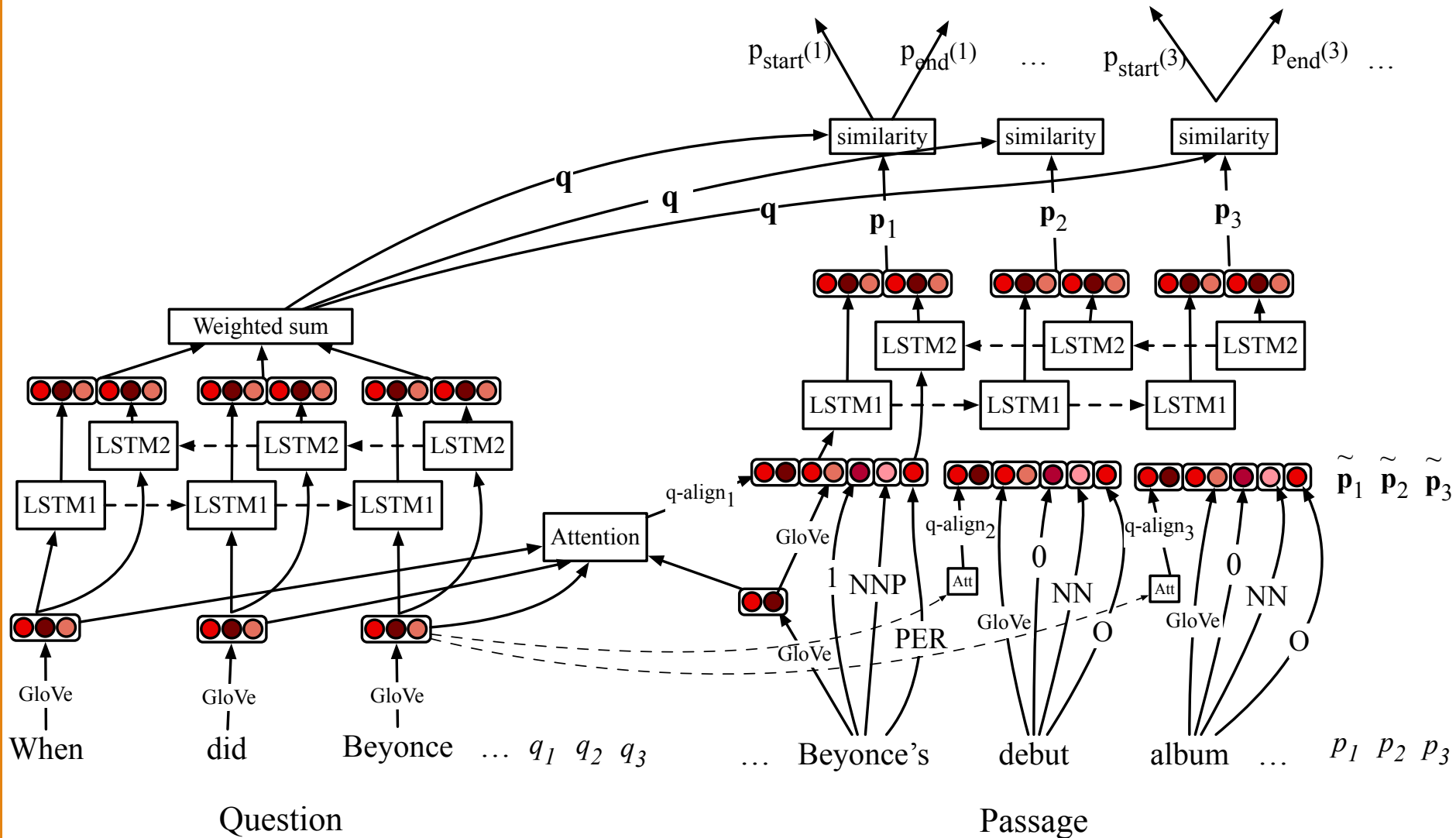
Q: "What areas did Beyoncé compete in when she was growing up?"

A: "**singing and dancing**"

Q: "When did Beyoncé release *Dangerously in Love*?"

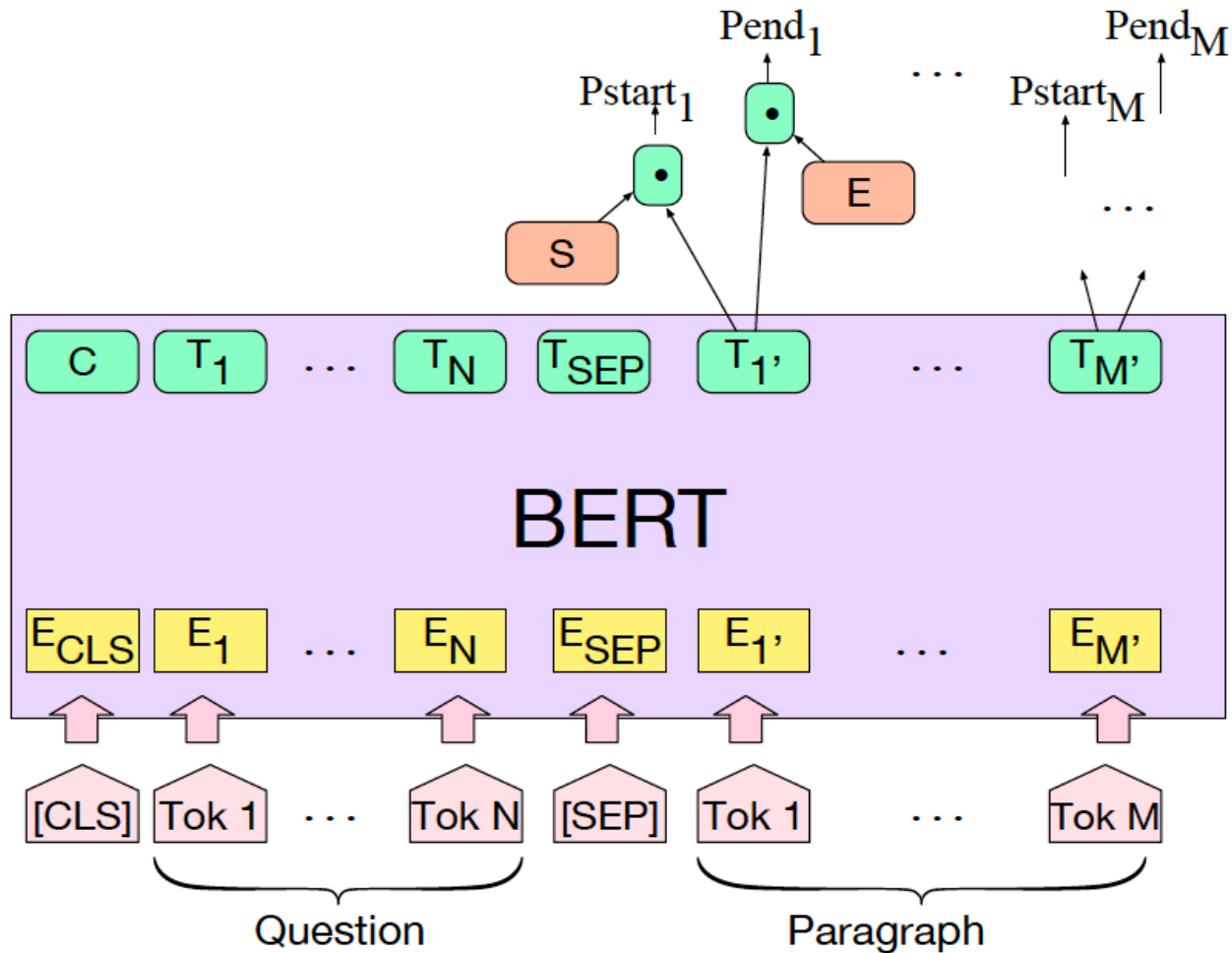
A: "**2003**"

# Neural Answer Extraction: bi-LSTM based approach





# Neural Answer Extraction: BERT based approach



# Knowledge-based Question Answering

Instead of trying to find a span of text on the web, knowledge based QA systems map a natural language question onto a query over a structured database.

Systems for mapping from a text string to any logical form are called **semantic parsers**.

Question	Logical forms
When was Ada Lovelace born?	birth-year (Ada Lovelace, ?x)
What states border Texas?	$\lambda x.state(x) \wedge borders(x,texas)$
What is the largest state?	$argmax(\lambda x.state(x), \lambda x.size(x))$
How many people survived the sinking of the Titanic?	(count (!fb:event.disaster.survivors fb:en.sinking of the titanic))

## BASEBALL: AN AUTOMATIC QUESTION-ANSWERER

Bert F. Green, Jr., Alice K. Wolf, Carol Chomsky, and Kenneth Laughery  
Lincoln Laboratory\*, Massachusetts Institute of Technology  
Lexington 73, Massachusetts

### Summary

Baseball is a computer program that answers questions phrased in ordinary English about stored data. The program reads the question from punched cards. After the words and idioms are looked up in a dictionary, the phrase structure and other syntactic facts are determined for a content analysis, which lists attribute-value pairs specifying the information given and the information requested. The requested information is then extracted from the data matching the specifications, and any necessary processing is done. Finally, the answer is printed. The program's present context is baseball games; it answers such questions as "Where did each team play on July 7?"

### Introduction

Men typically communicate with computers in a variety of artificial, stylized, unambiguous languages that are better adapted to the machine than to the man. For convenience and speed, many future computer-centered systems will require men to communicate with computers in natural language. The business executive, the military commander, and the scientist need to ask questions of the computer in ordinary English, and to have the computer answer questions directly. Baseball is a first step toward this

The program operates in the context of baseball data. At present, the data are the month, day, place, teams and scores for each game in the American League for one year. In this limited context, a small vocabulary is sufficient, the data are simple, and the subject-matter is familiar.

Some temporary restrictions were placed on the input questions so that the initial program could be relatively straightforward. Questions are limited to a single clause; by prohibiting structures with dependent clauses the syntactic analysis is considerably simplified. Logical connectives, such as and, or, and not, are prohibited, as are constructions implying relations like most and highest. Finally, questions involving sequential facts, such as "Did the Red Sox ever win six games in a row?" are prohibited. These restrictions are temporary expedients that will be removed in later versions of the program. Moreover, they do not seriously reduce the number of questions that the program is capable of answering. From simple questions such as "Who did the Red Sox lose to on July 5?" to complex questions such as "Did every team play at least once in each park in each month?" lies a vast number of answerable questions.

Specification List

# BASEBALL: AN AUTOMATIC QUESTION-ANSWERER

Bert F. Green, Jr., Alice K. Wolf, Carol Chomsky, and Kenneth Laughery  
Lincoln Laboratory\*, Massachusetts Institute of Technology  
Lexington 73, Massachusetts

## Summary

Baseball is a computer program that answers questions phrased in ordinary English about stored data. The program reads the question from punched cards. After the words and idioms are looked up in a dictionary, the phras syntactic facts are determined by analysis, which lists attributes specifying the information requested. The request is then extracted from the data base, and any necessary modifications are made. Finally, the answer is printed in the present context is baseball. Examples of such questions as "Where did the Red Sox play on July 7?"

## Introduction

Men typically communicate with computers through a variety of artificial, stylized, unambiguous languages that are better adapted to the machine than to the man. For convenience and speed, many future computer-centered systems will require men to communicate with computers in natural language. The business executive, the military commander, and the scientist need to ask questions of the computer in ordinary English, and to have the computer answer questions directly. Baseball is a first step toward this

The program operates in the context of baseball data. At present, the data are the month, day, place, teams and scores for each game in the American League for one year. In this limited context, a small vocabulary is sufficient, the data are simple, and the subject-matter is

"Where did the Red Sox  
play on July 7?"

Place = ?  
Team = Red Sox  
Month = July  
Day = 7

restrictions were placed on the program so that the initial program was straightforward. Questions are phrased in a single clause; by prohibiting independent clauses the syntactic complexity is considerably simplified. Logical connectives and, or, and not, are prohibited. Instructions implying relations are prohibited. Finally, questions about facts, such as "Did the Red Sox play in a row?" are prohibited. These restrictions are temporary and will be removed in later

versions of the program. Moreover, they do not seriously reduce the number of questions that the program is capable of answering. From simple questions such as "Who did the Red Sox lose to on July 5?" to complex questions such as "Did every team play at least once in each park in each month?" lies a vast number of answerable questions.

Specification List

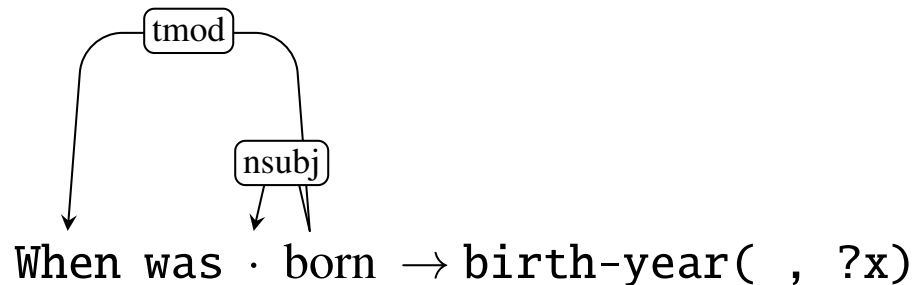
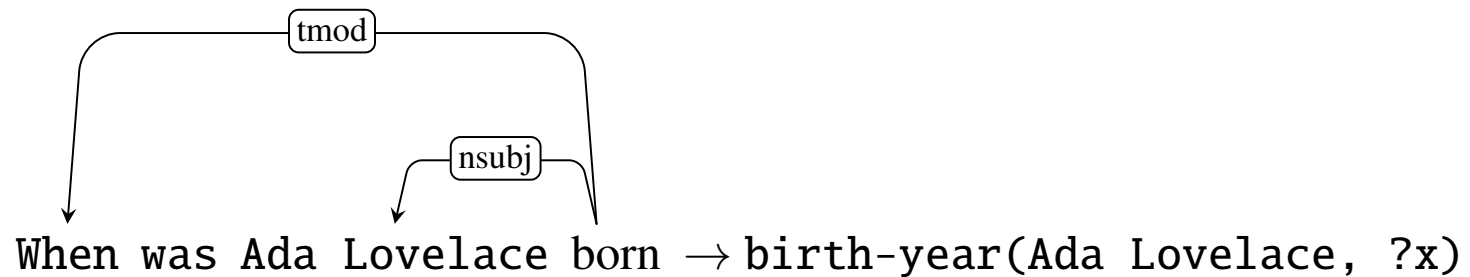
# Rule based systems

For frequent relations, it is worthwhile to write handwritten rules to extract relations from the question.

To extract the birth-year relation, we could write a pattern that searches for the question word **When**, a main verb like **born**, and then extract the named entity argument of the verb.

# Supervised systems

Learn a mapping between questions and database relations by parsing training data, and then learning general patterns.



# Dealing with Variation

Some phrases that align with the Freebase relation  
**country.capital**

<b>capital of</b>	<b>capital city of</b>	<b>become capital of</b>
capitol of	national capital of	official capital of
political capital of	administrative capital of	beautiful capital of
capitol city of	bustling capital of	make capital of
political center of	move its capital to	capital city in
cosmopolitan capital of	remain capital of	modern capital of
federal capital of	beautiful capital city of	administrative capital city of

# Paraphrase databases

**Q: What are the green blobs in plant cells?**

---

*Lemmatized synonyms from PARALEX:*

---

what be the green blob in plant cell?

what be green part in plant cell?

what be the green part of a plant cell?

what be the green substance in plant cell?

what be the part of plant cell that give it green color?

what cell part do plant have that enable the plant to be give a green color?

what part of the plant cell turn it green?

part of the plant cell where the cell get it green color?

the green part in a plant be call?

the part of the plant cell that make the plant green be call?



# Value of Logical Representation of Sentences

Is Barack Obama a US Citizen?

Citizen\_Of(Barack\_Obama, United\_States)

$\forall x \text{ Person}(x) \wedge \text{Born-In}(x, y) \wedge \text{Located-In}(y, \text{United\_States}) \Rightarrow \text{Citizen\_Of}(x, \text{United\_States})$

Person(Barack\_Obama)  $\wedge$

Born-In(Barack\_Obama, Hawaii)  $\wedge$

Located-In(Hawaii, United\_States)

---

Citizen\_Of(Barack\_Obama, United\_States)

**Barack Obama**



**44th President of the United States**

**In office**  
January 20, 2009 – January 20, 2017

**Vice President** [Joe Biden](#)

**Preceded by** [George W. Bush](#)

**Succeeded by** [Donald Trump](#)

---

**Personal details**

**Born** [Honolulu, Hawaii](#)

# Evaluating QA systems

To evaluate a system that returns a **ranked** set answers for a test set consisting of  $N$  questions, the **mean reciprocal rank** (MRR) is defined as

$$\text{MRR} = \frac{1}{N} \sum_{i=1 \text{ s.t. } rank_i \neq 0}^N \frac{1}{rank_i}$$

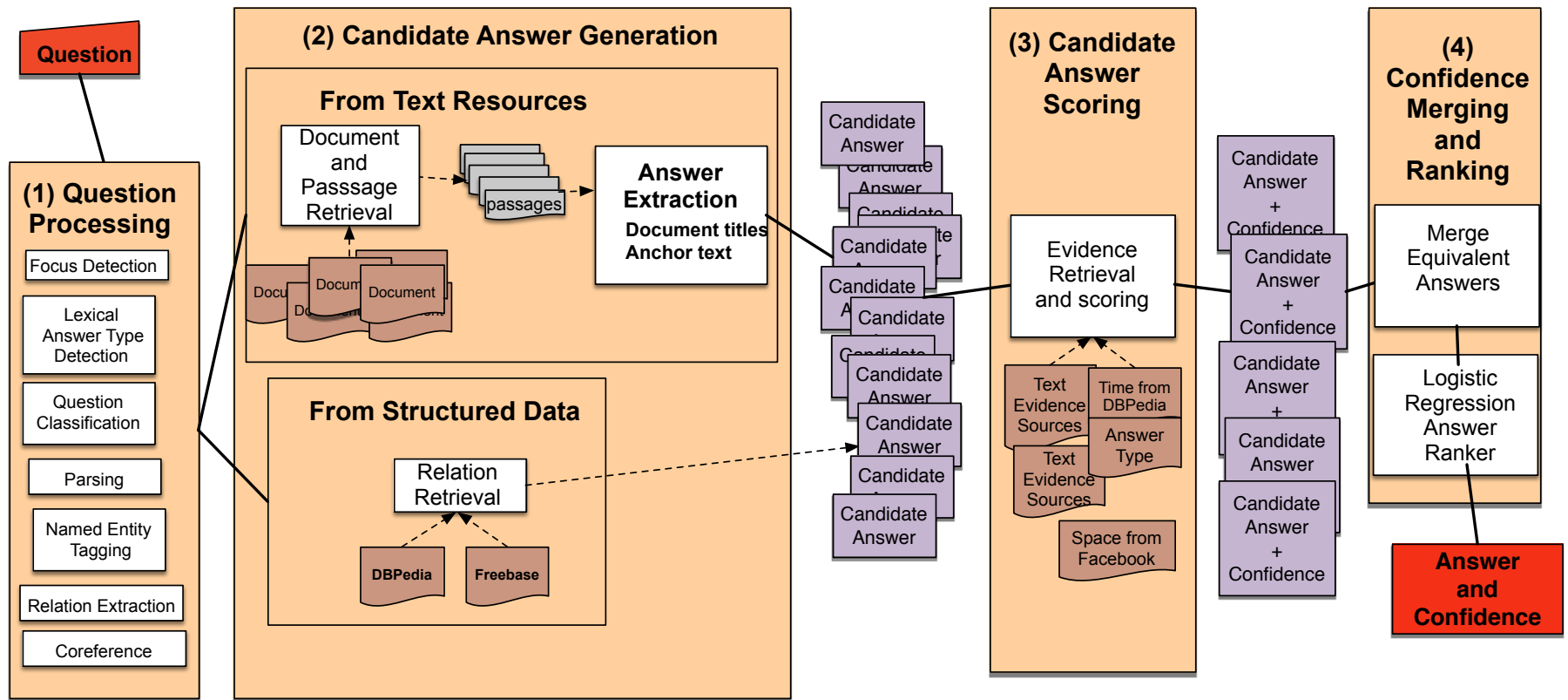
# Evaluating systems on SQuAD

Reading comprehension systems on datasets like SQuAD are often evaluated using two metrics (both ignore punctuation and articles):

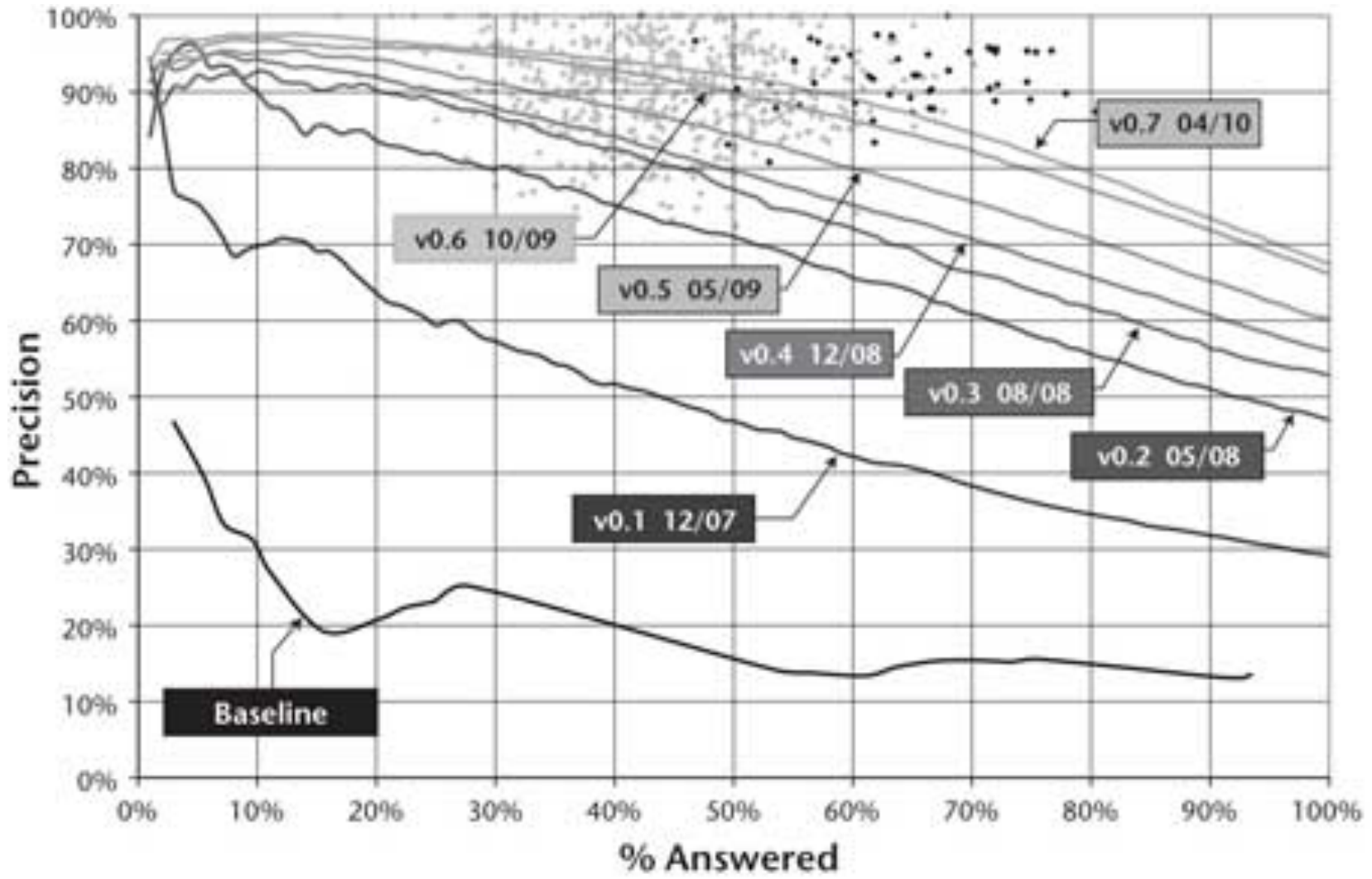
**Exact match:** The percentage of predicted answers that match the gold answer exactly.

**F1 score:** The average overlap between predicted and gold answers. Treat the prediction and gold as a bag of tokens, and compute F1, averaging the F1 over all questions.

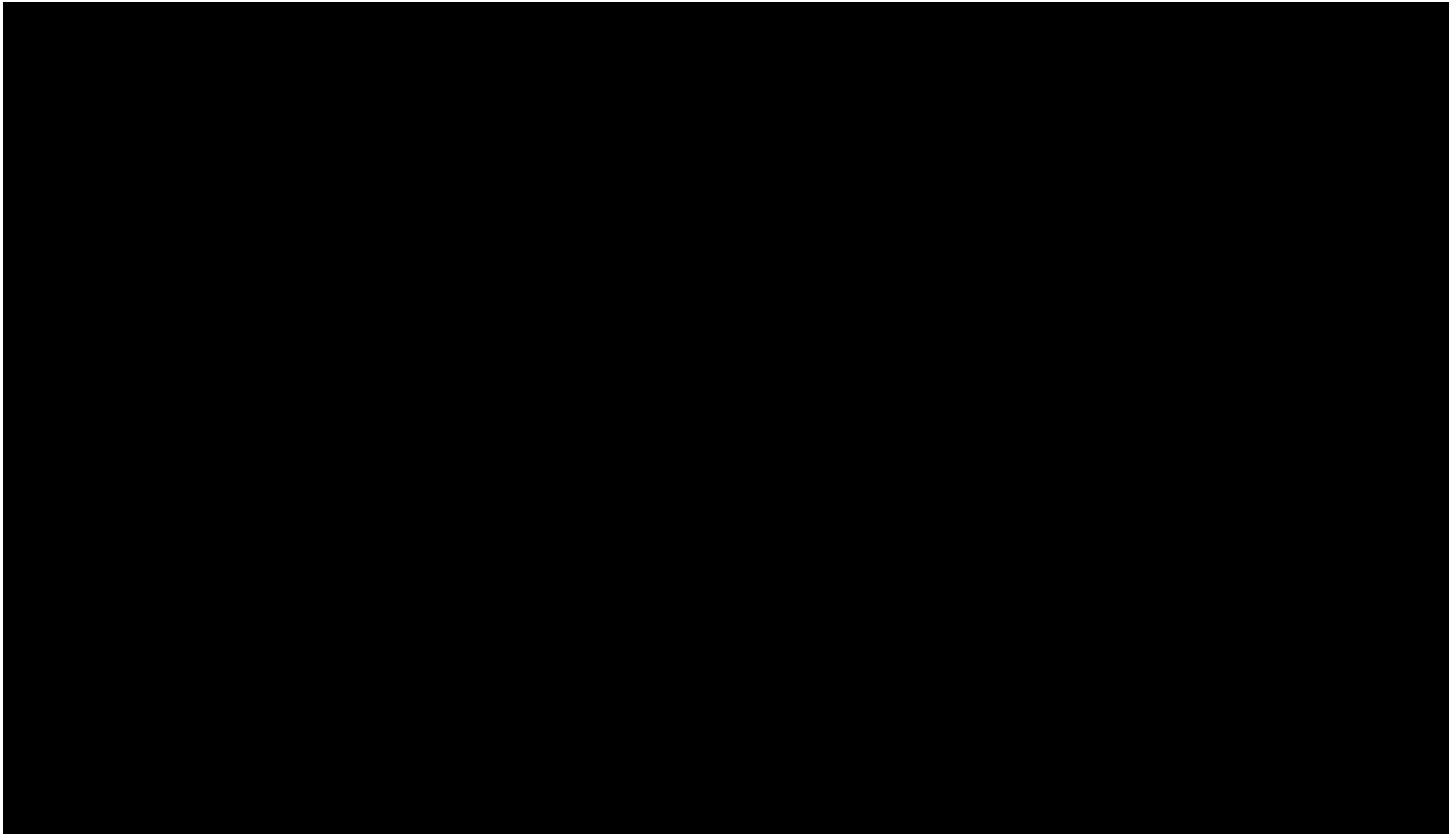
# IBM Watson architecture



# IBM Watson evaluation



# IBM Watson



[https://www.youtube.com/watch?v=IIRN\\_cSNe\\_Y](https://www.youtube.com/watch?v=IIRN_cSNe_Y)

# More challenging QA tasks

AI2 Reasoning Challenge (ARC) (Clark et al., 2018), has questions that are designed to be hard to answer from simple lexical methods:

*Which property of a mineral can be determined just by looking at it?*

- (A) luster**
- (B) mass
- (C) weight
- (D) hardness

# More challenging QA tasks

Khashabi (2018) introduced a challenge set for reading comprehension over multiple sentences

*Which property of a mineral can be determined just by looking at it?*

S1: Most young mammals, including humans, play.

S2: Play is how they learn the skills that they will need as adults.

S6: Big cats also play.

S8: At the same time, they also practice their hunting skills.

S11: Human children learn by playing as well.

S12: For example, playing games and sports can help them learn to follow rules.

S13: They also learn to work together.

What do human children learn by playing games and sports?

A)\* They learn to follow rules and work together

B) hunting skills

C)\* skills that they will need as adult



# More challenging QA tasks

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#) [Read](#) [View source](#) [View history](#)

## Brewing

From Wikipedia, the free encyclopedia


*"Brewer" redirects here. For other uses, see [Brewer \(disambiguation\)](#).*

*This article is about the brewing of beer. For homebrewing, see [Homebrewing](#). For other uses, see [Brewing \(disambiguation\)](#).*

**Brewing** is the production of [beer](#) by [steeping](#) a [starch](#) source (commonly [cereal](#) grains, the most popular of which is [barley](#))<sup>[1]</sup> in water and [fermenting](#) the resulting sweet liquid with [yeast](#). It may be done in a [brewery](#) by a commercial brewer, at home by a [homebrewer](#), or by a variety of traditional methods such as communally by the [indigenous peoples in Brazil](#) when making [cauim](#).<sup>[2]</sup> Brewing has taken place since around the 6th millennium BC, and archaeological evidence suggests that emerging civilizations including [ancient Egypt](#)<sup>[3]</sup> and [Mesopotamia](#) brewed beer.<sup>[4]</sup> Since the nineteenth century the [brewing industry](#) has been part of most western economies.

The basic ingredients of beer are water and a [fermentable](#) starch source such as [malted barley](#). Most beer is fermented with a [brewer's yeast](#) and flavoured with [hops](#).<sup>[5]</sup> Less widely used starch sources include [millet](#), [sorghum](#) and [cassava](#).<sup>[6]</sup> Secondary sources ([adjuncts](#)), such as maize (corn), rice, or sugar, may also be used, sometimes to reduce cost, or to add a feature, such as adding wheat to aid in retaining the foamy head of the beer.<sup>[7]</sup> The proportion of each starch source in a beer recipe is collectively called the [grain bill](#).

Steps in the brewing process include [malting](#), [milling](#), [mashing](#), [lautering](#), [boiling](#), [fermenting](#), [conditioning](#), [filtering](#), and [packaging](#). There are three main fermentation methods, [warm](#), [cool](#) and [spontaneous](#). Fermentation may take place in an open or closed fermenting vessel; a secondary fermentation may also occur in the [cask](#) or [bottle](#). There are several additional [brewing methods](#), such as barrel aging, double dropping, and Yorkshire Square.



A 16th-century brewery

**Contents** [\[hide\]](#)

- 1 [History](#)
- 2 [Ingredients](#)
- 3 [Brewing process](#)
- 4 [Mashing](#)
  - 4.1 [Lautering](#)
- 5 [Boiling](#)
  - 5.1 [Brew kettle or copper](#)

Question: when are hops added to the brewing process?  
Answer: The boiling process