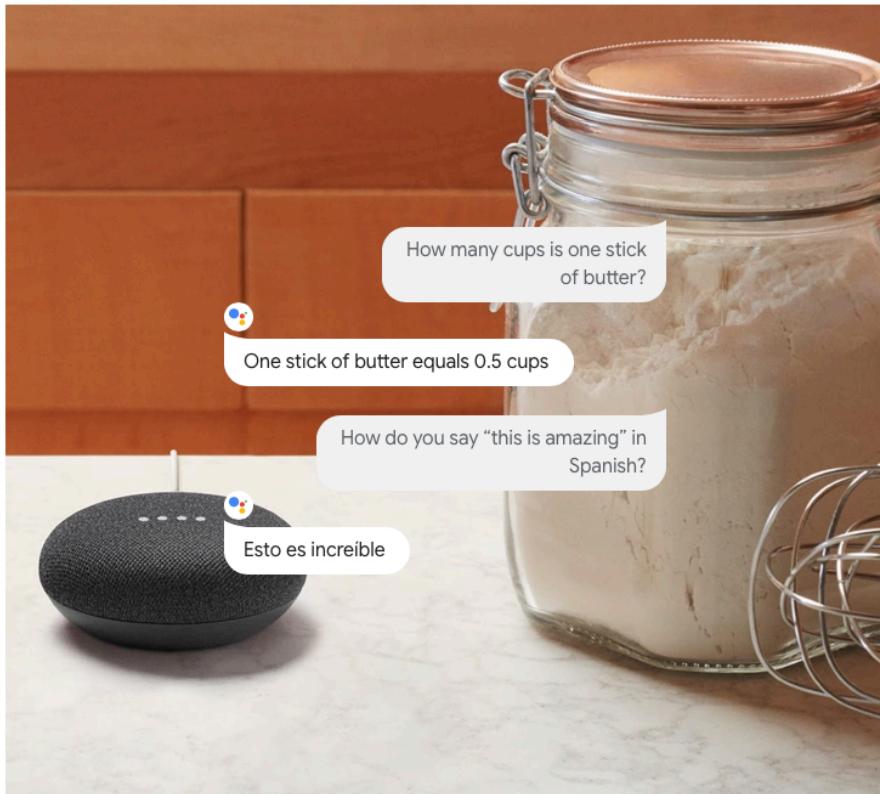


Question Answering

JURAFSKY AND MARTIN CHAPTER 22

Get help around the house from your smart speaker



Get answers

Get answers on sports, weather, finance, calculations, translations and more.

All

Shopping

Images

Videos

News

More

Settings

Tools

About 3,590,000 results (0.65 seconds)

Horses and ponies have about the same **eating** habits. They like to **eat** different grasses. They also like grain and hay. Straw is sometimes mistaken for hay but there is a difference, hay is greener than straw and it has a rich smell.



[Eating Habits of horses and ponys - Horses with Amie](#)
www.horseswithamie.com/horsehealth/eatinghabits.htm

[? About this result](#)[! Feedback](#)

People also ask

What can horses eat list?

What do ponies eat and drink?

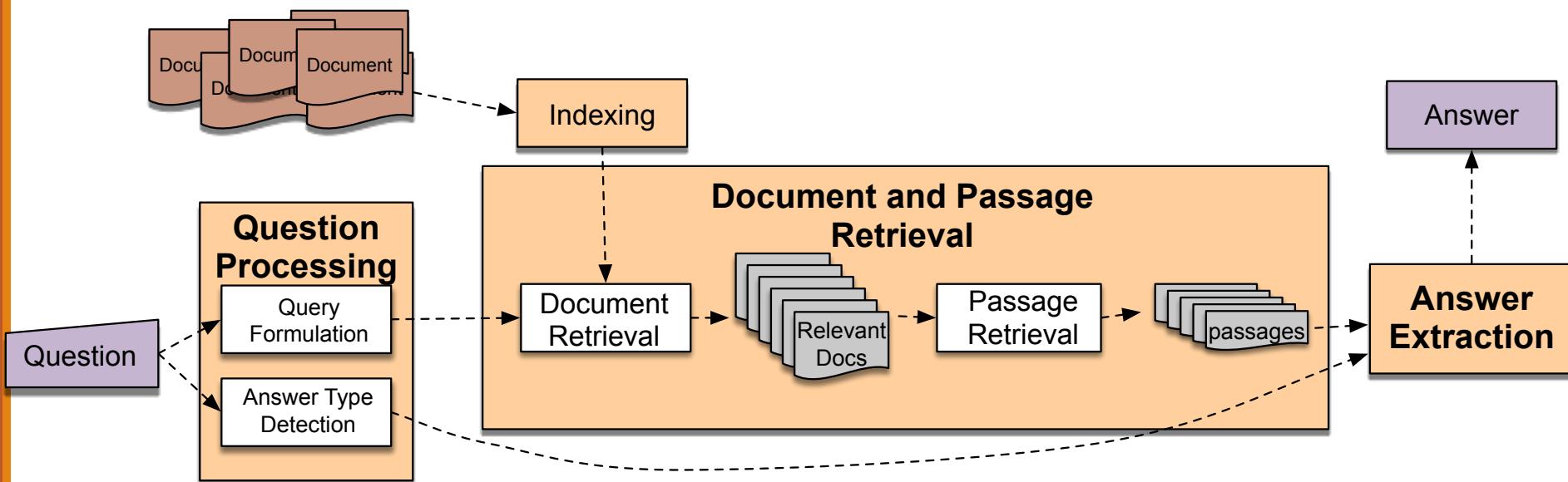
How do horses survive on hay?

What do horses like to eat the most?

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

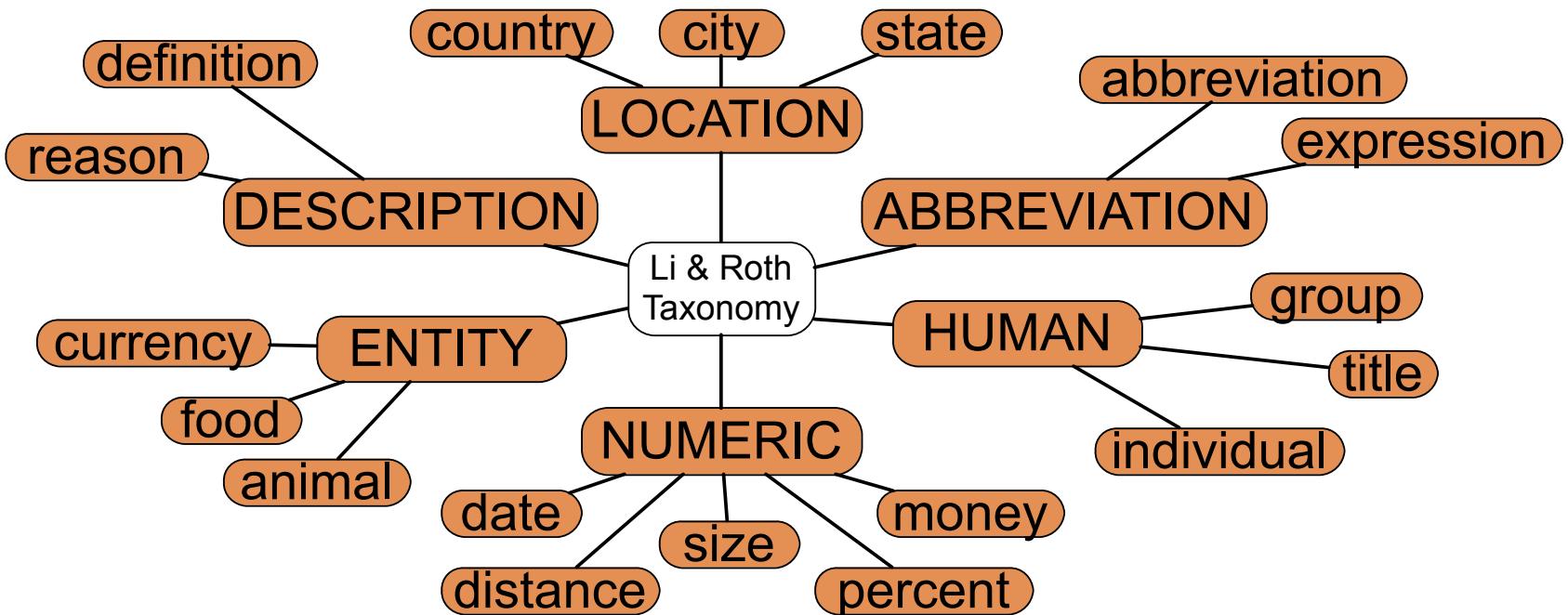
Answer types

Some systems make use of **question classification**, which is 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 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?

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.

Features used to rank passages

The number of named entities of the correct type in the passage

The number of question keywords in the passage

The longest exact sequence of question keywords that occurs in the passage

The rank of the document from which the passage was extracted

The proximity of the keywords from the original query to each other

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 England?

After **David Cameron** stepped down because of his failed campaign to keep the UK in the European Union, **Theresa May** became Britain's prime minister

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
<answer phrase> such as <question phrase>	What is <i>autism</i> ?	... , developmental disorders such as <i>autism</i> ...
<question phrase> , a <answer phrase>	What is a <i>caldera</i> ?	The Long Valley <i>caldera</i> , a volcanic crater 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?

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

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

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 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](#)"

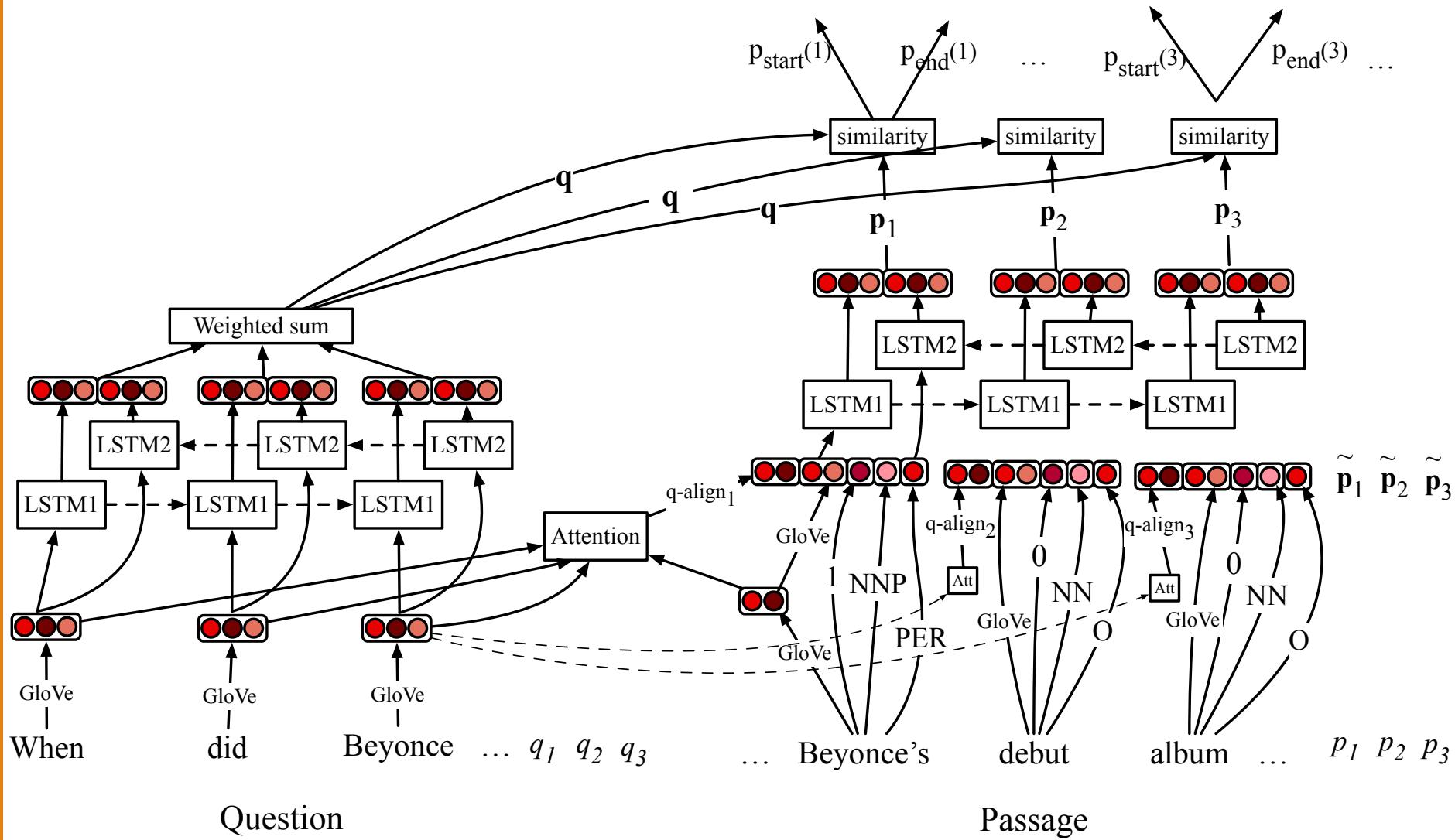
Q: "What areas did Beyoncé compete in when she was growing up?"

A: "[singing and dancing](#)"

Q: "When did Beyoncé release [Dangerously in Love](#)?"

A: "[2003](#)"

Neural Answer Extraction



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.

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

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 phrasal syntactic facts are determined by specifying the information requested. The required facts are then extracted from the data structures, and any necessary inferences are made. Finally, the answer is printed. The present context is baseball, so that such questions as "Where did the Red Sox play on July 7?"

Introduction

Men typically communicate 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

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

Place = ?

Team = Red Sox

Month = July

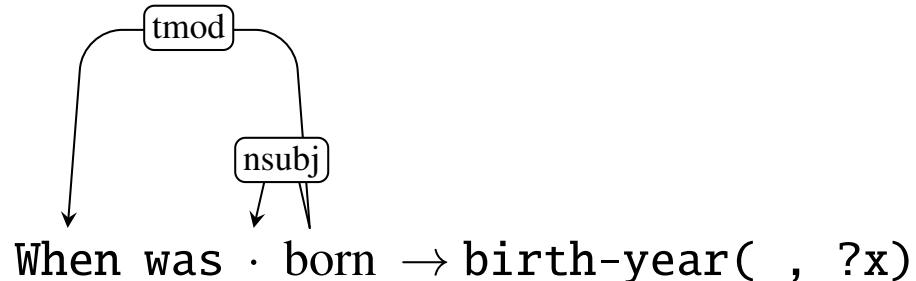
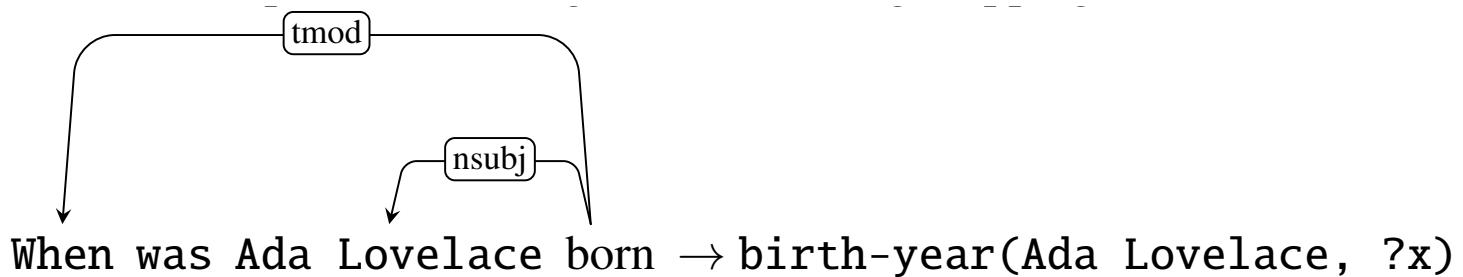
Day = 7

restrictions were placed on the language so that the initial program was straightforward. Questions involving multiple clauses; by prohibiting relative clauses the syntactic structure is considerably simplified. Logical operators and, or, and not, are prohibited. Finally, questions involving relations between entities. Finally, questions involving temporary facts, such as "Did the Red Sox win all their games 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.

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`

capital of	capital city of	become capital of
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?

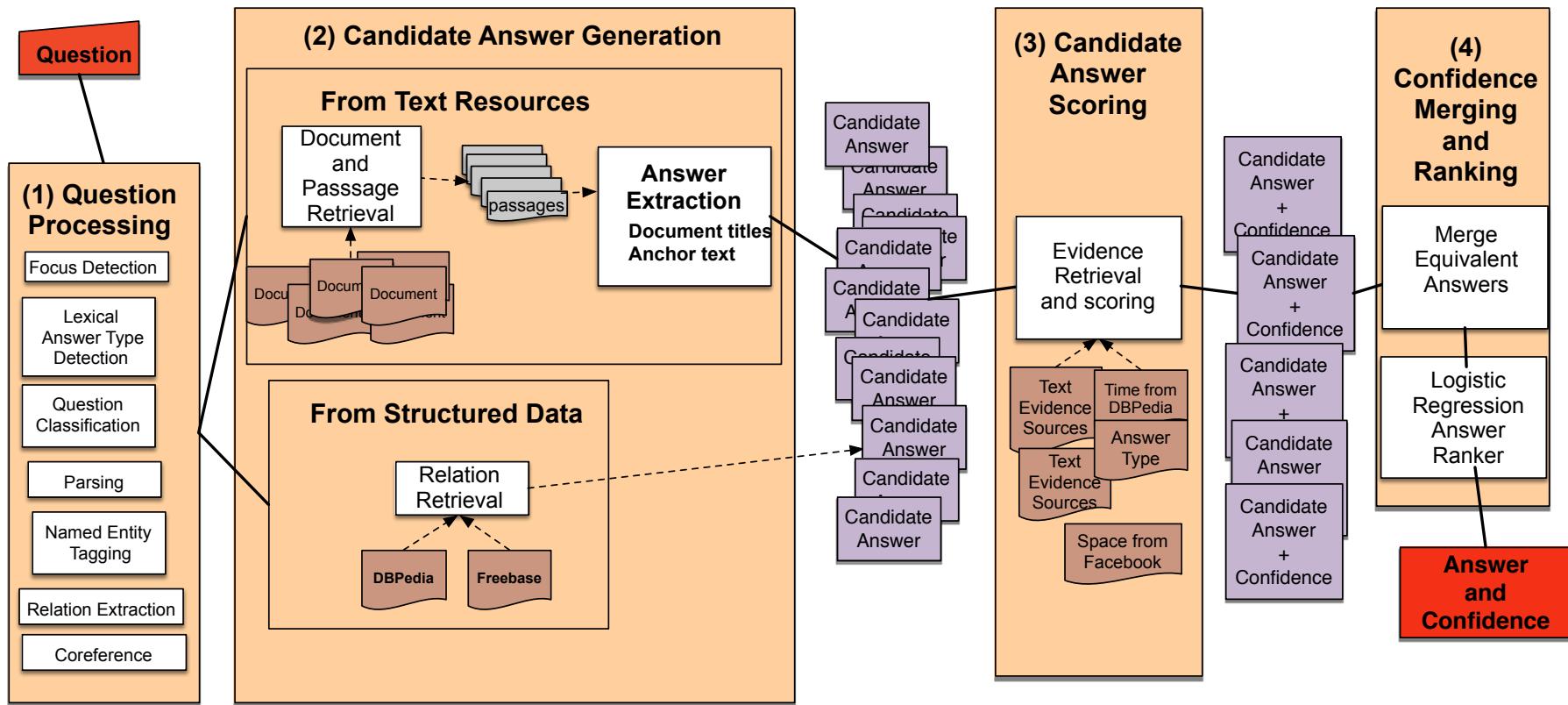
what part of the plant cell turn it green?

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

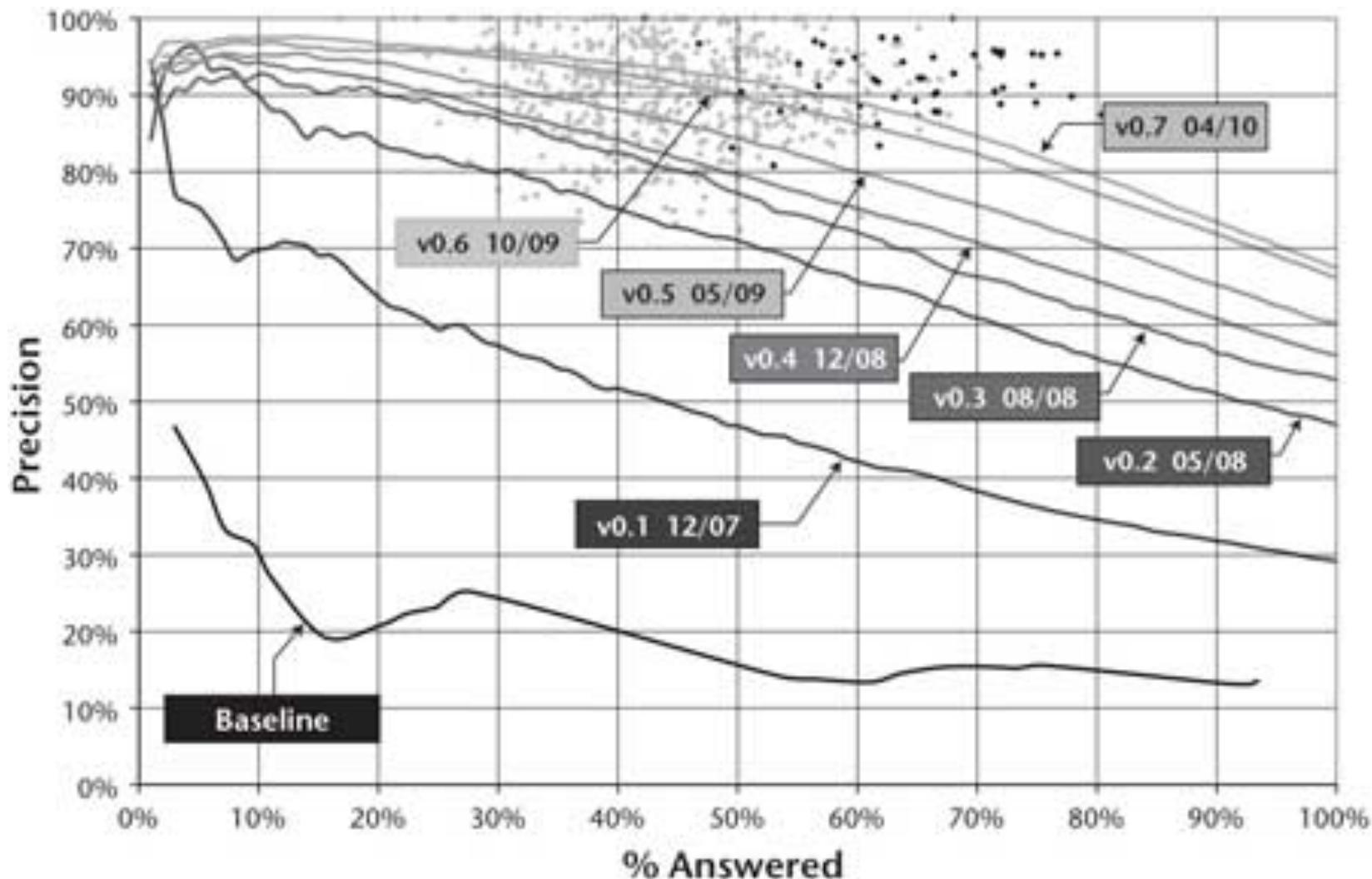
the green part in a plant be call?

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

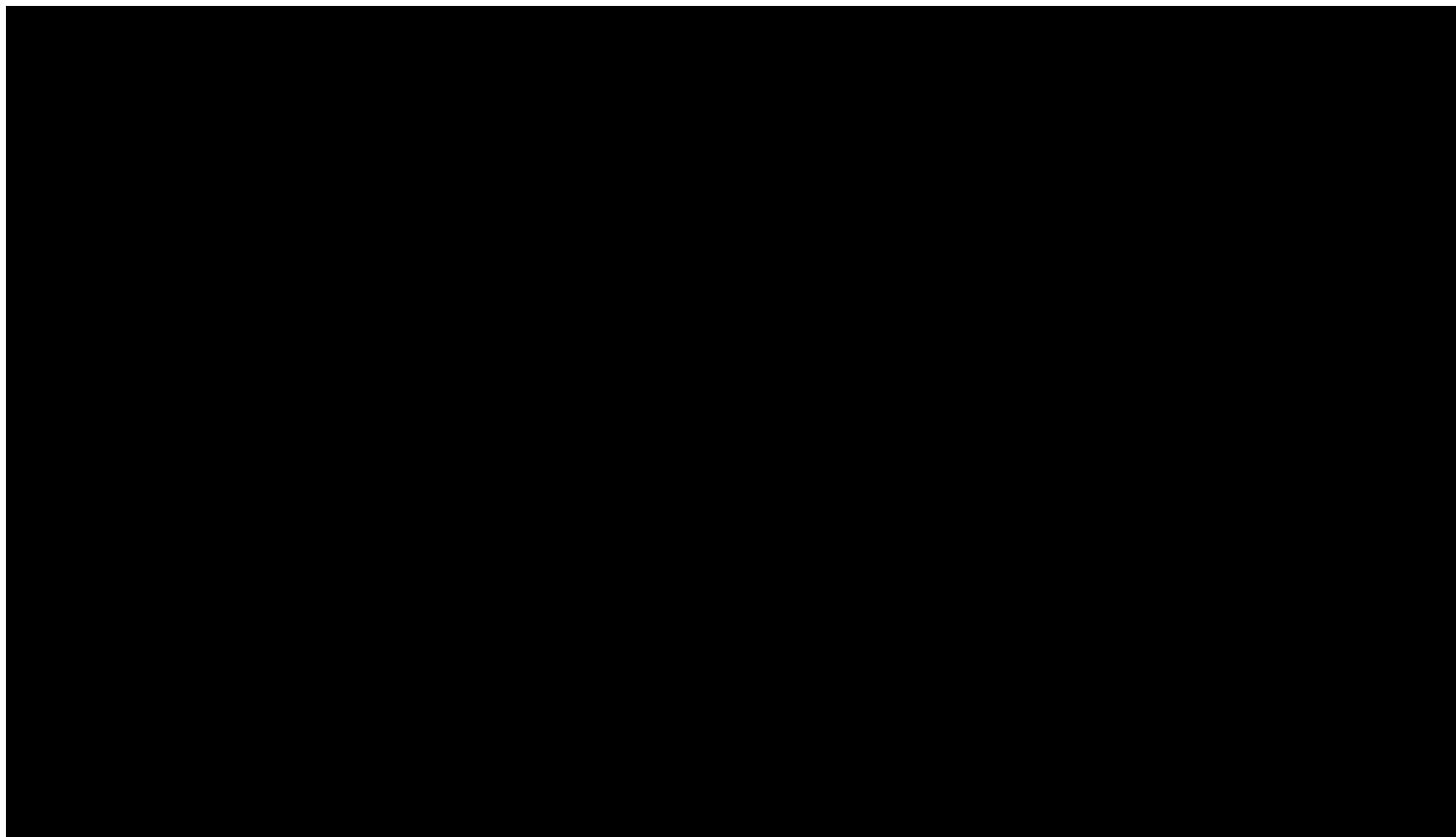
IBM Watson architecture



IBM Watson evaluation



IBM Watson



https://www.youtube.com/watch?v=llrN_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