Natural Language Processing
SoSe 2016

Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Question Answering

Dr. Mariana Neves

July 4th, 2016

Outline

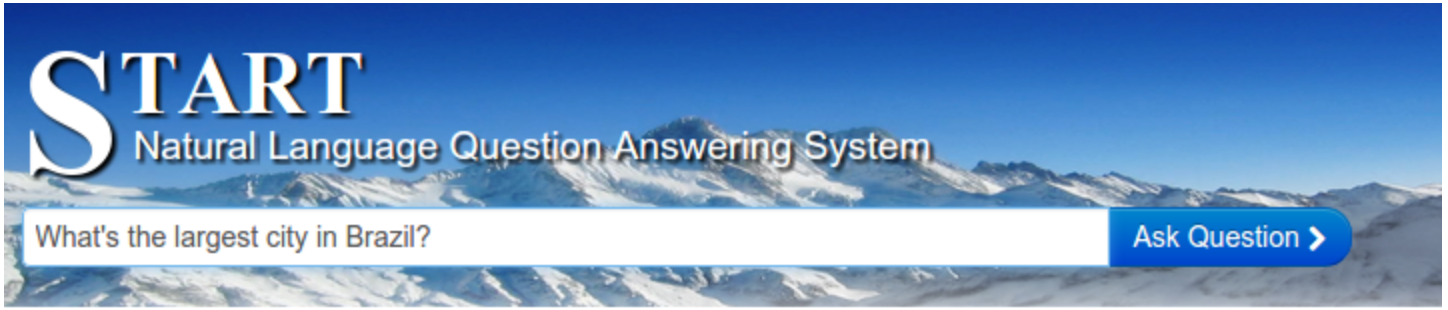
- Introduction
- QA Architecture

Outline

- Introduction
- QA Architecture

Motivation

- Find small segments of text which answer users' questions



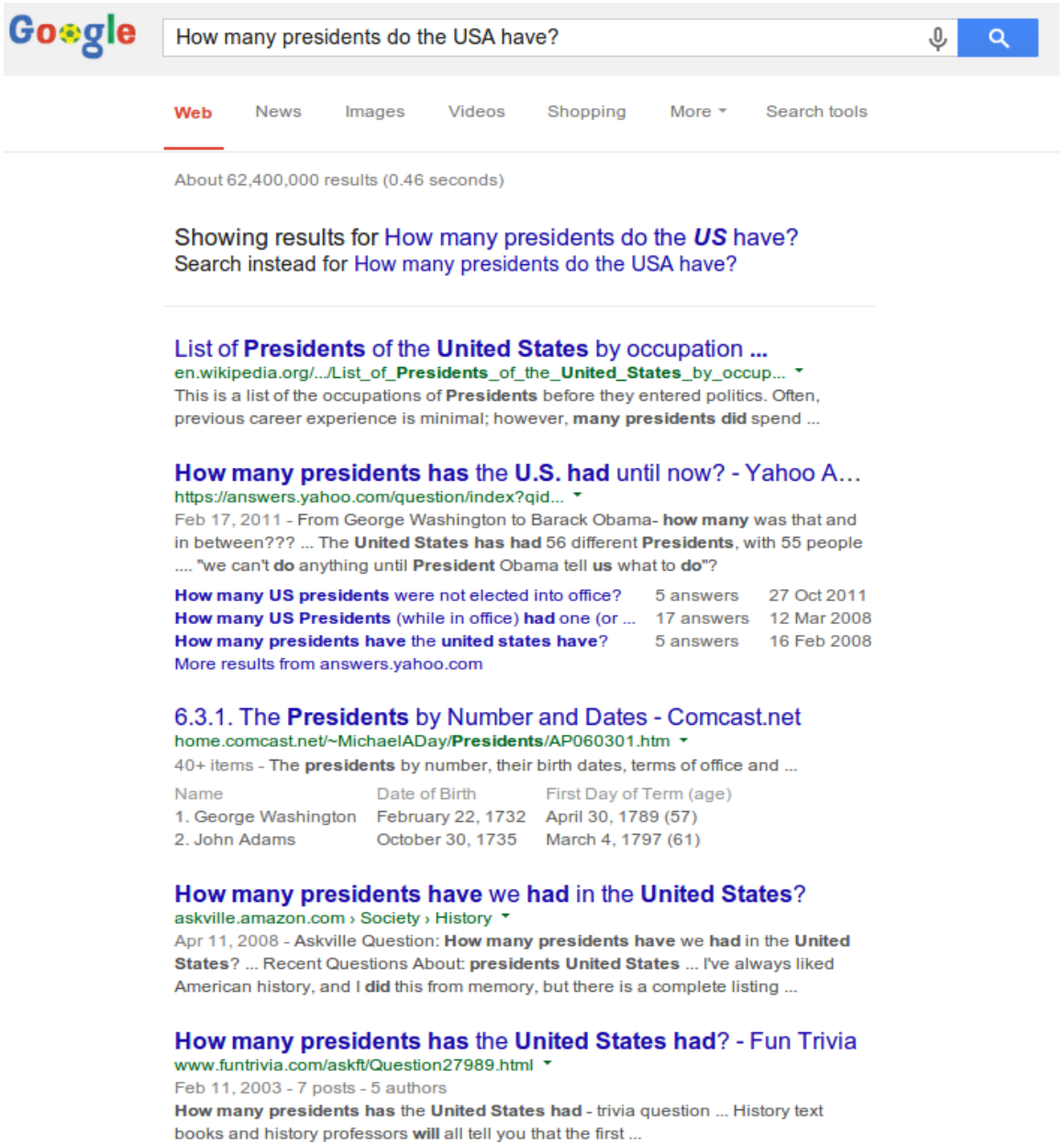
==> What's the largest city in Brazil?

Brazil

Capital (2009 est.): Brasília, 3,789,000

Largest cities: São Paulo, 19,900,000; Rio de Janeiro, 11,836,000; Salvador, 2,590,400; Belo Horizonte, 5,736,000; Recife, 1,485,500; Porto Alegre, 4,034,000

(http://start.csail.mit.edu/)


Google
How many presidents do the USA have?
Web
News
Images
Videos
Shopping
More
Search tools
About 62,400,000 results (0.46 seconds)
Showing results for How many presidents do the US have?
Search instead for How many presidents do the USA have?
List of Presidents of the United States by occupation ...
en.wikipedia.org/.../List_of_Presidents_of_the_United_States_by_occup...
This is a list of the occupations of Presidents before they entered politics. Often, previous career experience is minimal; however, many presidents did spend ...
How many presidents has the U.S. had until now? - Yahoo A...
https://answers.yahoo.com/question/index?qid...
Feb 17, 2011 - From George Washington to Barack Obama- how many was that and in between??? ... The United States has had 56 different Presidents, with 55 people .... "we can't do anything until President Obama tell us what to do"?
How many US presidents were not elected into office? 5 answers 27 Oct 2011
How many US Presidents (while in office) had one (or ... 17 answers 12 Mar 2008
How many presidents have the united states have? 5 answers 16 Feb 2008
More results from answers.yahoo.com
6.3.1. The Presidents by Number and Dates - Comcast.net
home.comcast.net/~MichaelADay/Presidents/AP060301.htm
40+ items - The presidents by number, their birth dates, terms of office and ...
Name Date of Birth First Day of Term (age)
1. George Washington February 22, 1732 April 30, 1789 (57)
2. John Adams October 30, 1735 March 4, 1797 (61)
How many presidents have we had in the United States?
askville.amazon.com › Society › History
Apr 11, 2008 - Askville Question: How many presidents have we had in the United States? ... Recent Questions About: presidents United States ... I've always liked American history, and I did this from memory, but there is a complete listing ...
How many presidents has the United States had? - Fun Trivia
www.funtrivia.com/askft/Question27989.html
Feb 11, 2003 - 7 posts - 5 authors
How many presidents has the United States had - trivia question ... History text books and history professors will all tell you that the first ...

Motivation

- How many presidents do the USA have?
  - 44
  - List of names in chronological order

40	Ronald Reagan[3][4]	Actor & broadcaster, President of the Screen Actors Guild[5]
41	George H. W. Bush	Pilot (Navy Lieutenant, Junior Grade),[6] Businessman (Oil)[7]
42	Bill Clinton	Lawyer, Law lecturer
43	George W. Bush	Businessman (Oil,[8] baseball)[9]
44	Barack Obama	Public Interest professional, Lawyer, Constitutional Law Professor [10]

(http://en.wikipedia.org/wiki/List_of_Presidents_of_the_United_States_by_occupation)

Motivation

- Information retrieval
  - Keywords (short input)
  - Document (long output)

- Question Answering
  - Natural language question (long input)
  - Short answer (short output)

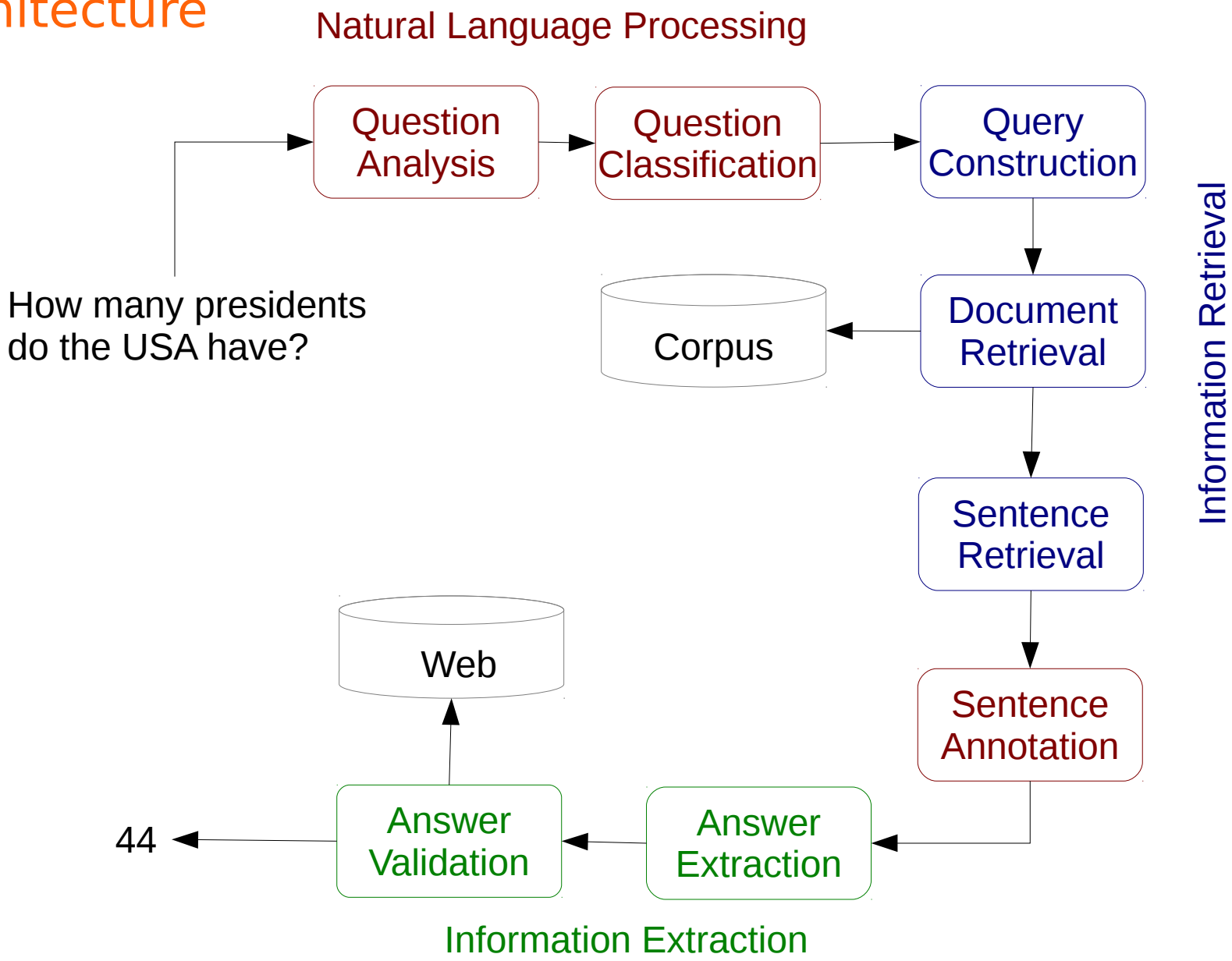
QA Types

- Closed-domain
  - Answer questions from a specific domain
- Open-domain
  - Answer any domain-independent question

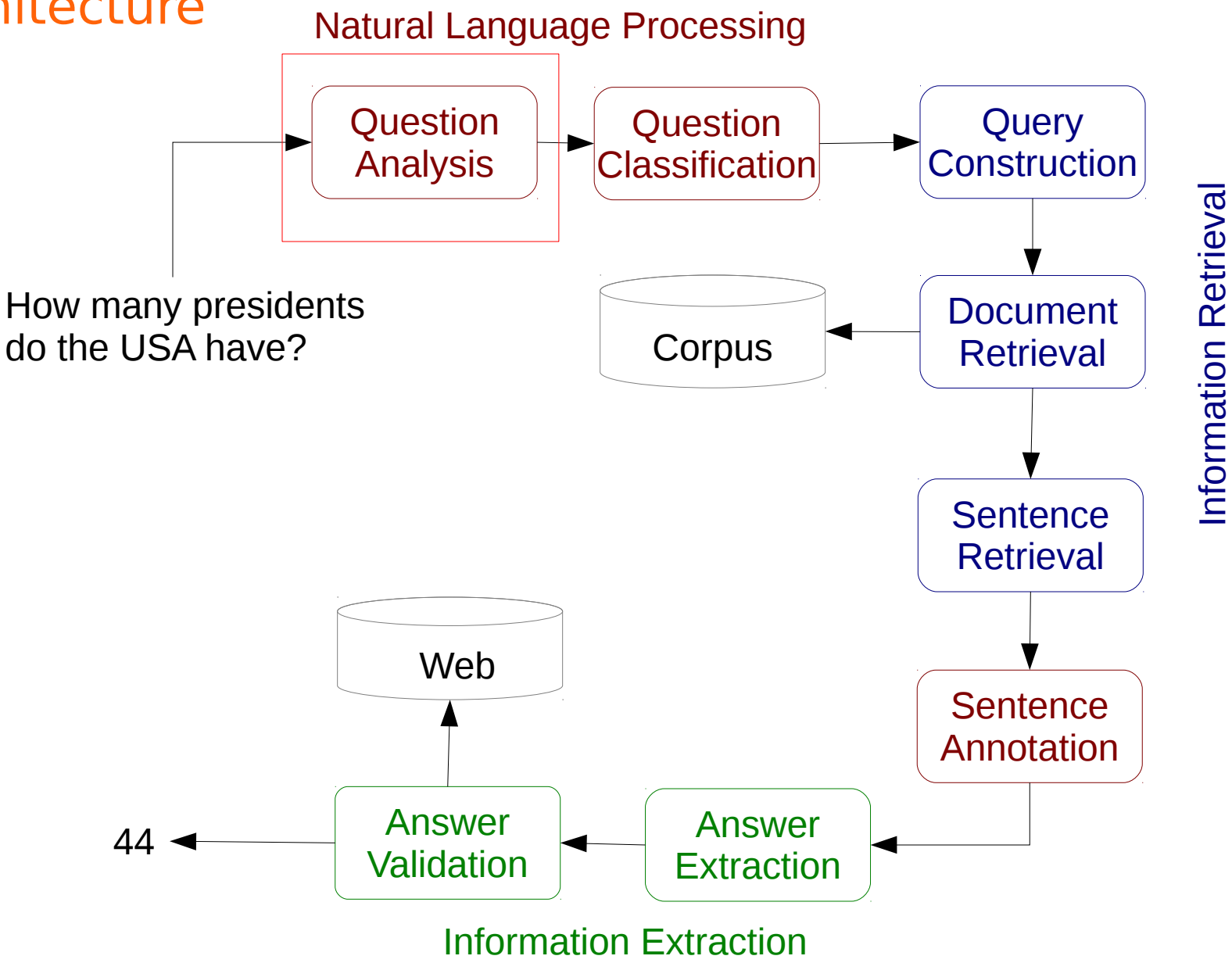
Outline

- Introduction
- QA Architecture

Architecture



Architecture



Question Analysis

- Named-Entity Recognition
- Surface Text Pattern Learning
- Syntactic Parsing
- Semantic Role Labeling

Question Analysis: Named-Entity Recognition

- Recognize the named entities in the text to extract the target of the question
- Use the question's target in the query construction step

- Example:
  - Question: "In what country was Albert Einstein born?"
  - Target: "Albert Einstein"

Question Analysis: Pattern Learning

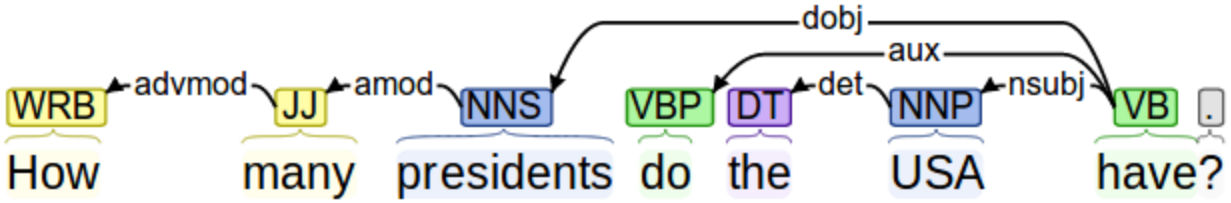
- Extract a pattern from the question
- Match the pattern with a list of pre-defined question patterns
- Find the corresponding answer pattern
- Realize the position of the answer in the sentence in the answer extraction step

Question Analysis: Pattern Learning

- Example:
  - Question: “In what country was Albert Einstein born?”
  - Question Pattern: “In what country was X born?”
  - Answer Pattern: “X was born in Y.”

Question Analysis: Syntactic Parsing

- Use a dependency parser to extract the syntactic relations between question terms
- Use the dependency relation paths between question terms to extract the correct answer in the answer extraction step



(http://nlp.stanford.edu:8080/corenlp/process)

Question Analysis: Semantic Role Labeling

- Finde the question's head verb
- Example:
  - „Who$_{[Buyer]}$ purchased YouTube$_{[Goods]}$?“
- Commerce_buy

Buyer [Byr]

The Buyer wants the Goods and offers Money to a Seller in exchange for them.
Jess BOUGHT a coat.

Lee BOUGHT a textbook from Abby.

Goods [Gds]

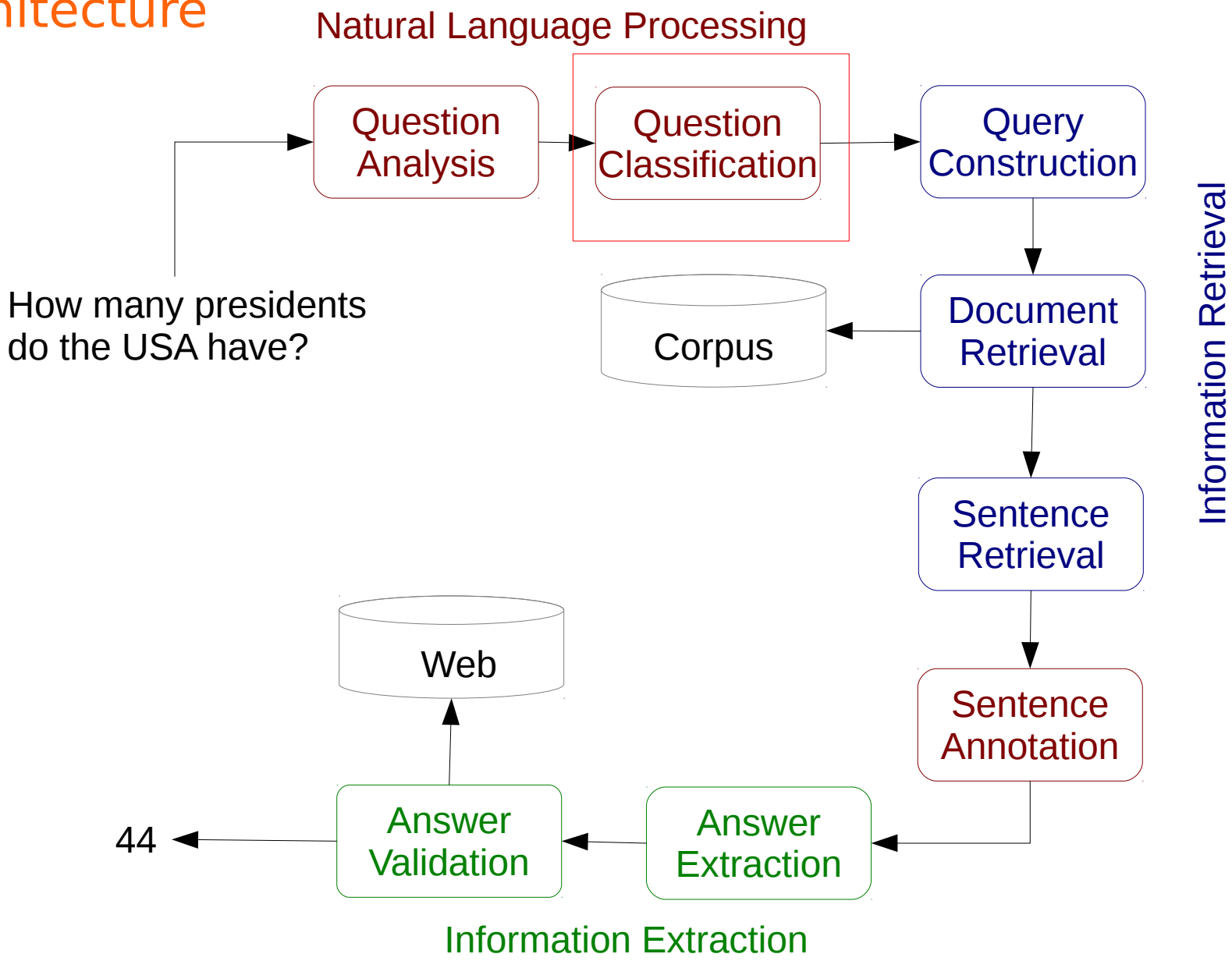
The FE Goods is anything (including labor or time, for example) which is exchanged for Money in a transaction.
Only one winner PURCHASED the paintings

(https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Commerce_buy)

Question Analysis: Semantic Role Labeling

- COMMERCE−BUY
  - buy (v), buyer (n), purchase (n), purchase (v), purchaser (n)
- Identify candidate passages
  - Buyer [Subj,NP] verb Goods [Obj,NP]
  - Buyer [Subj,NP] verb Goods [Obj,NP] Seller [Dep,PP-from]
  - Goods [Subj,NP] verb Buyer [Dep,PP-by]
  - ...
- Example:
  - „In 2006, YouTube$_{[Goods]}$ was purchased by Google$_{[Buyer]}$ for $1.65 billion.“

Architecture



Question Classification

- Classify the input question into a set of question types
- Map question types to the available named-entity labels
- Find strings that have the same type as the input question in the answer extraction step

- Example:
  - Question: “In what country was Albert Einstein born?”
  - Type: LOCATION - Country

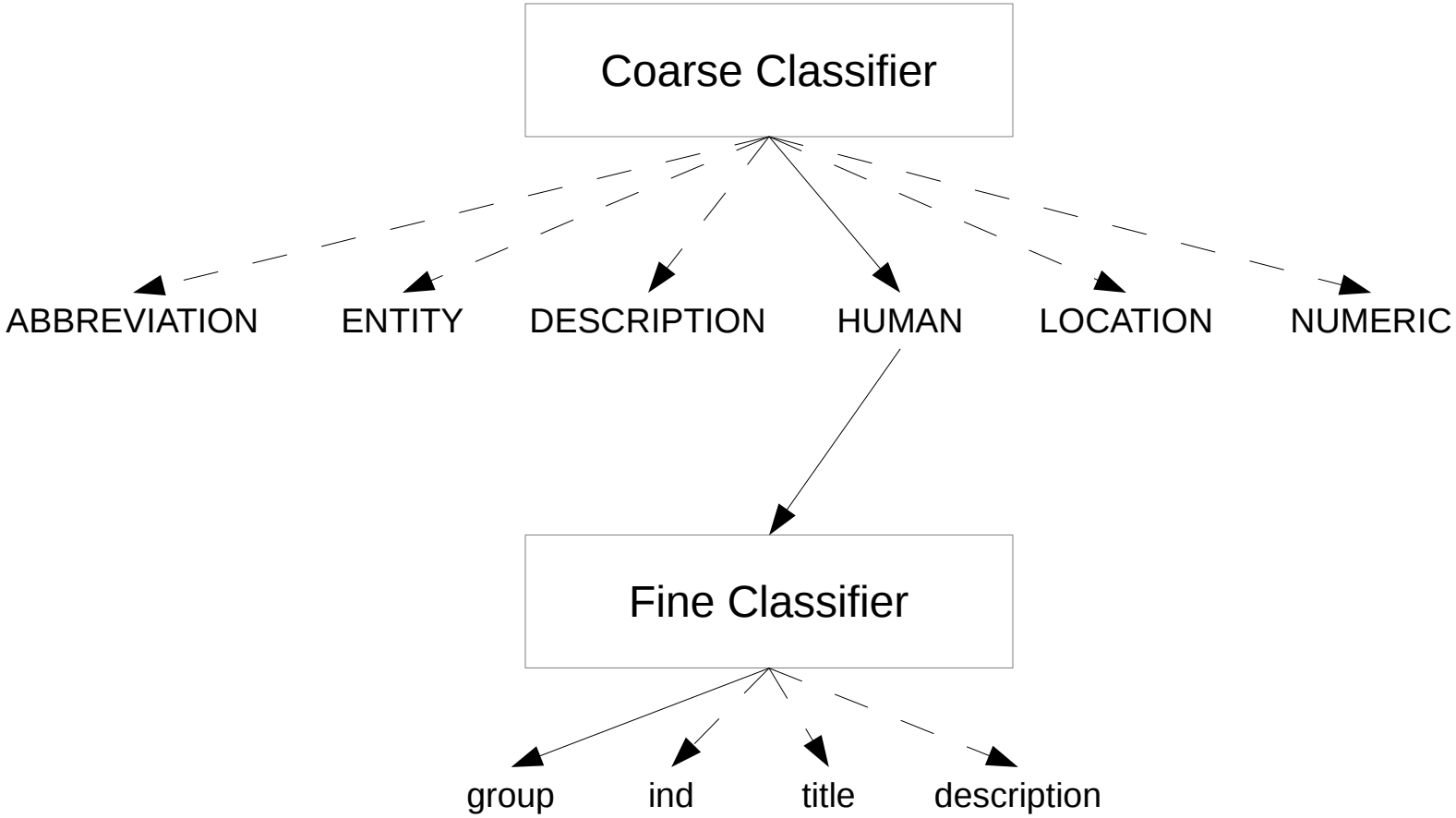
Question Classification

- Example (NER):
  - S1: “ Albert Einstein was born in 14 March 1879 .”
  - S2: “ Albert Einstein was born in Germany .”
  - S3: “ Albert Einstein was born in a Jewish family.”

Question Classification

- Classification taxonomies
  - BBN
  - Pasca & Harabagiu
  - Li & Roth
    - 6 coarse- and 50 fine-grained classes
      - ABBREVIATION
      - ENTITY
      - DESCRIPTION
      - HUMAN
      - LOCATION
      - NUMERIC

Question Classification

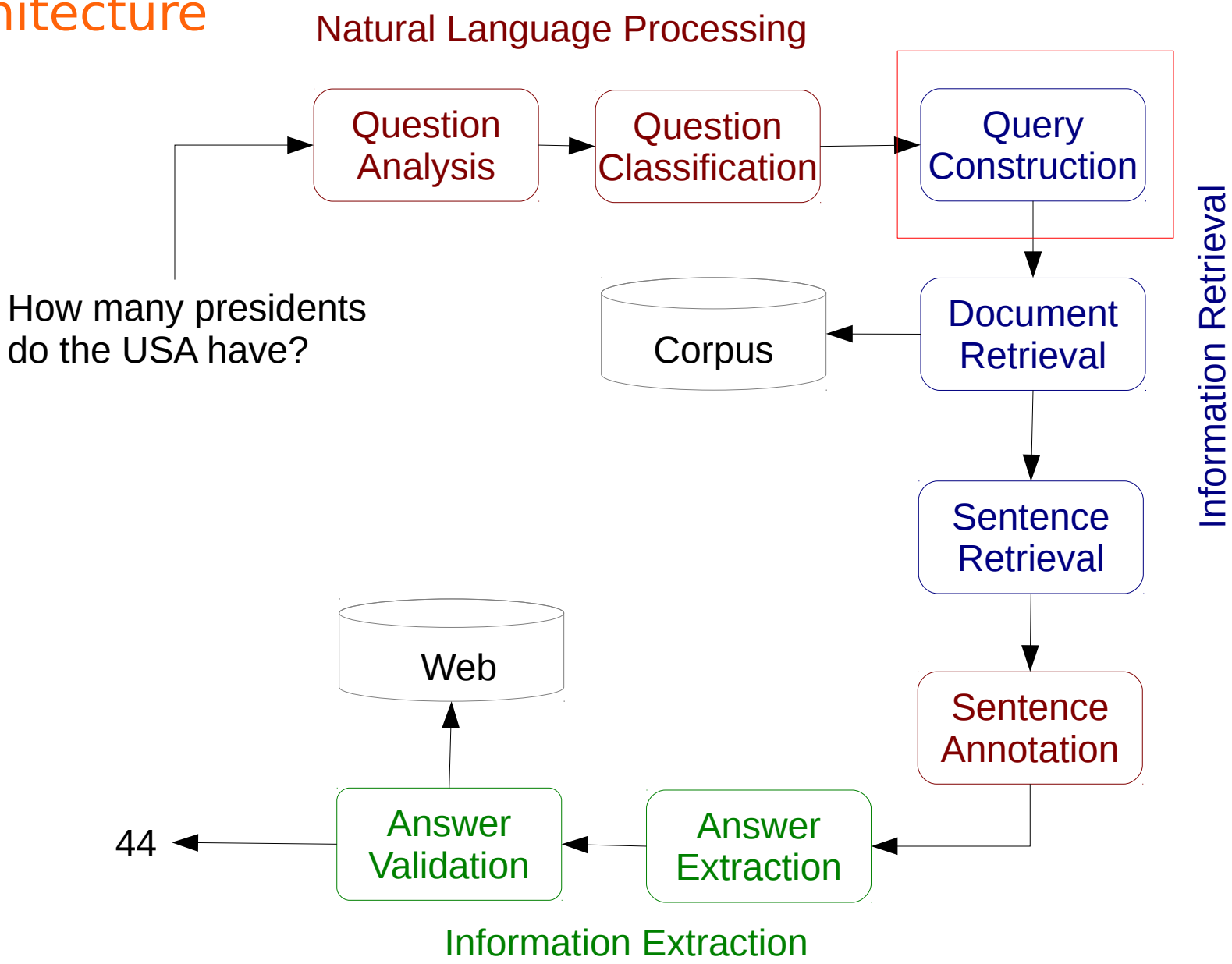


Question Classification

- Rule-based
- Machine learning-based

- Considering the confidence measure of the classification to filter the result

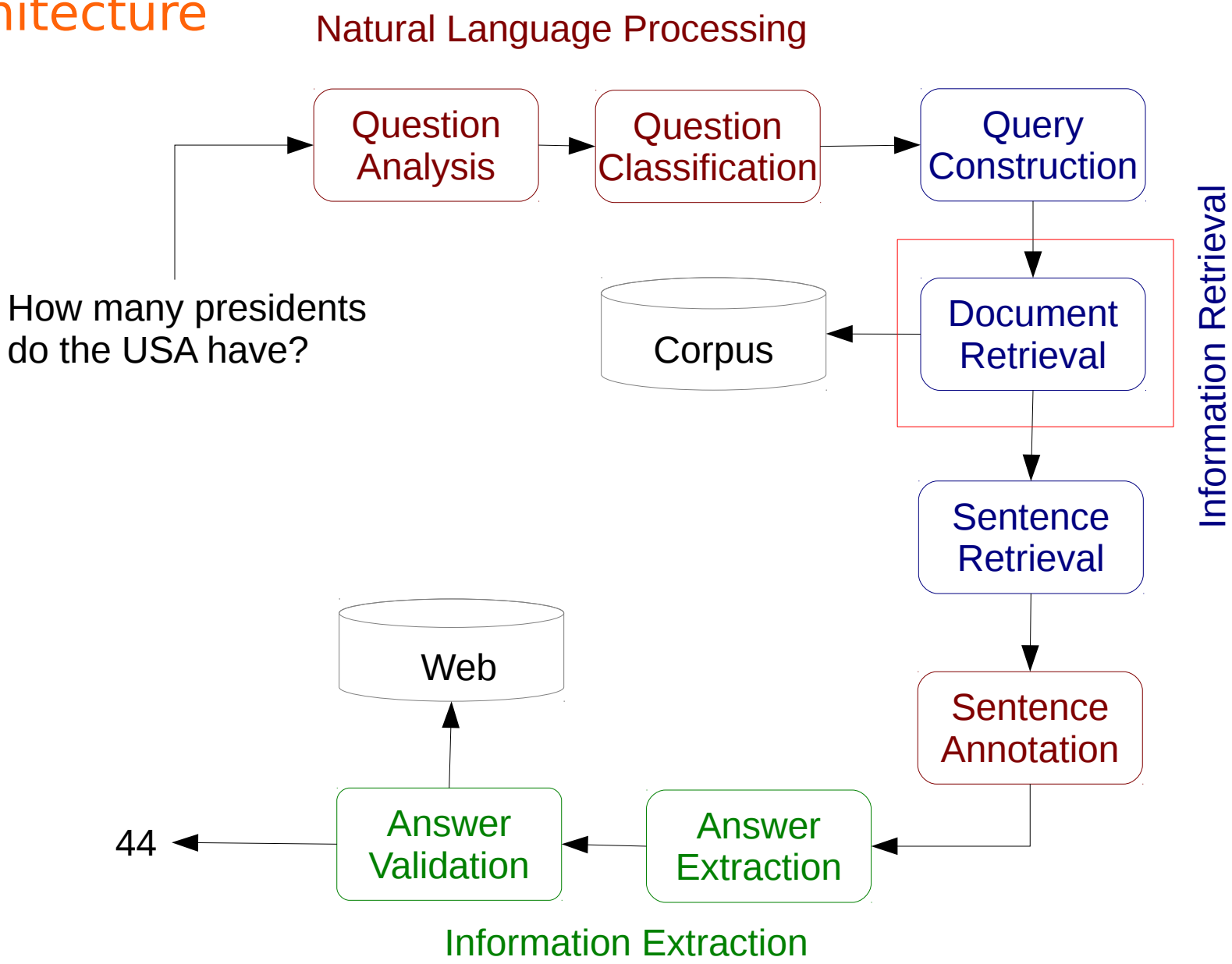
Architecture



Query Construction

- Goal:
  - Formulate a query with a high chance of retrieving relevant documents
- Task:
  - Assign a higher weight to the question’s target
  - Use query expansion techniques to expand the query

Architecture



Document Retrieval

- Importance:
  - QA components use computationally intensive algorithms
  - Time complexity of the system strongly depends on the size of the to be processed corpus

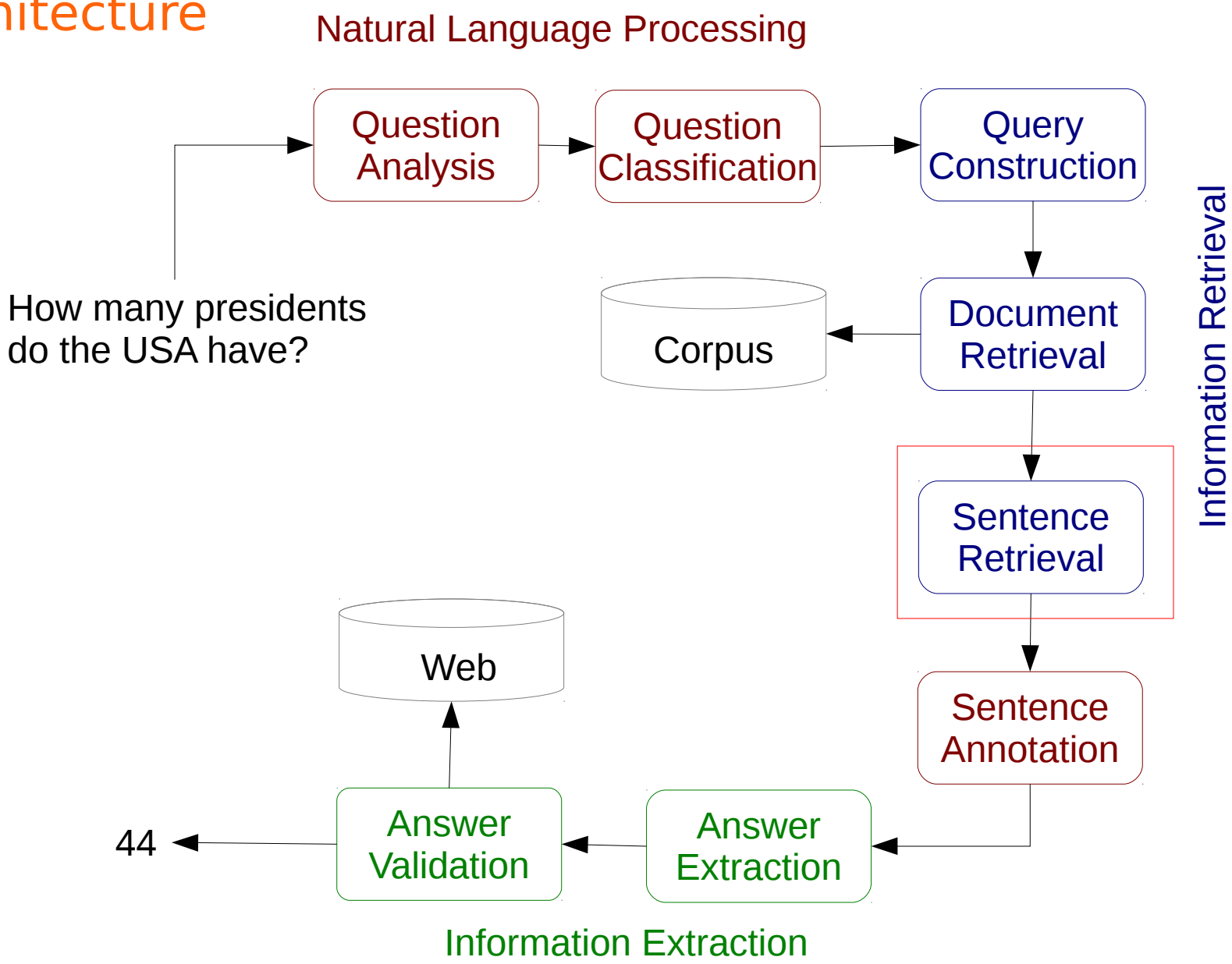
- Task:
  - Reduce the search space for the subsequent components
  - Retrieve relevant documents from a large corpus
  - Select top „n“ retrieved documents for the next steps

Document Retrieval

- Use available information retrieval models
  - Vector Space Model
  - Probabilistic Model
  - Language Model

- Use available information retrieval toolkits: Lucene, Lemur, Sors, ElasticSearch, etc.

Architecture



Sentence Retrieval

- Task:
  - Find small segments of text that contain the answer
- Benefits beyond document retrieval:
  - Documents can be very large
  - Documents span different subject areas
  - The relevant information is expressed locally
  - Sentences simplifies the answer extraction step
- Main problem: sentence brevity

Sentence Retrieval

- Information retrieval models for sentence retrieval
  - Vector Space Model
  - Probabilistic Model
  - Language Model
    - Jelinek-Mercer Linear Interpolation
    - Bayesian Smoothing with Dirichlet Prior
    - Absolute Discounting

Sentence Retrieval

- Support from named-entity recognition and semantic role labeling

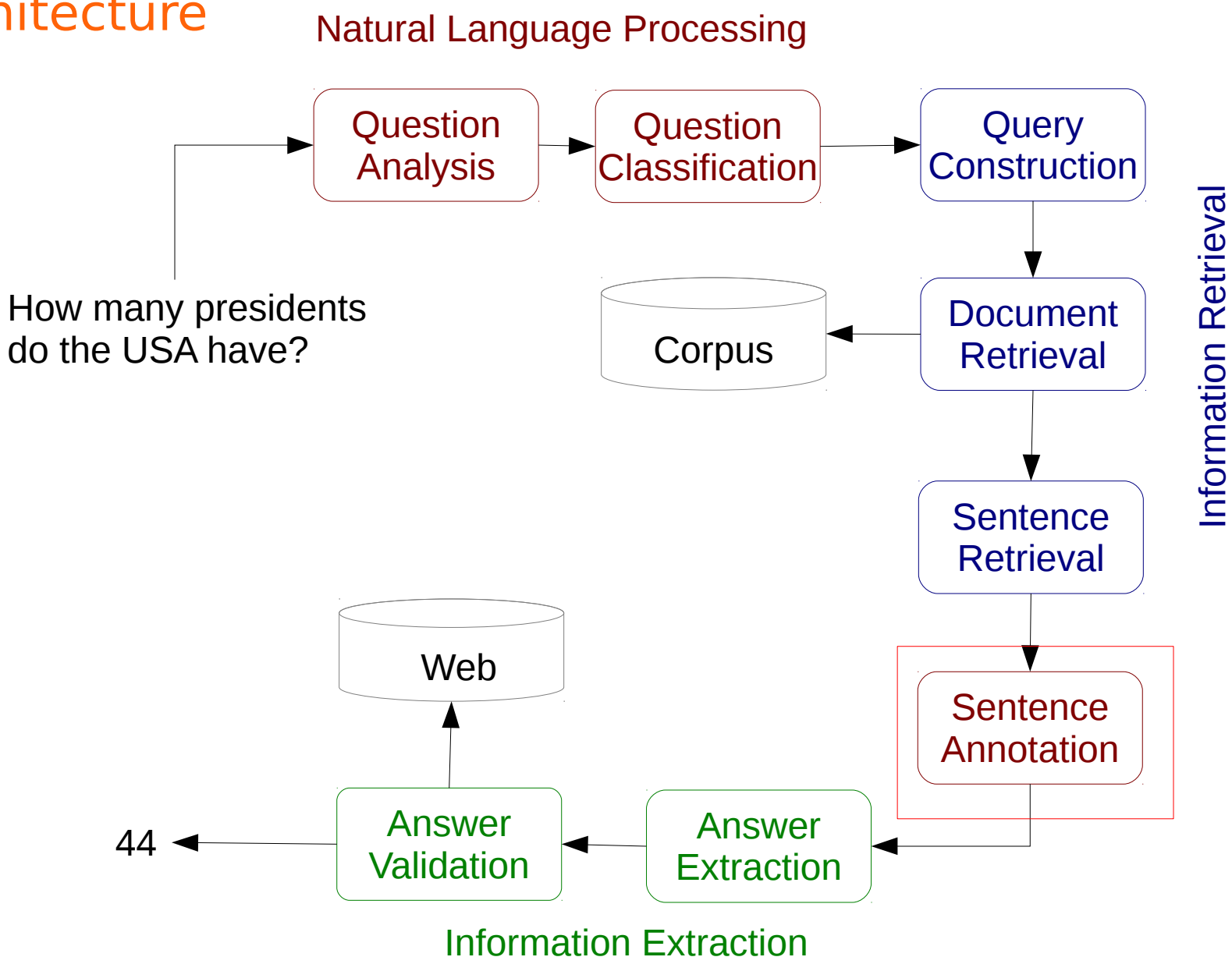
  - “ Albert Einstein was born in 14 March 1879 .”
  - “ Albert Einstein was born in Germany .”
  - “ Albert Einstein was born in a Jewish family.”

  - Buyer [Subj,NP] verb Goods [Obj,NP]
  - Buyer [Subj,NP] verb Goods [Obj,NP] Seller [Dep,PP-from]
  - Goods [Subj,NP] verb Buyer [Dep,PP-by]

Sentence Retrieval

- Supervised learning
  - Features
    - Number of entity types of the right type
    - Number of question keywords
    - Longest exact sequence of question keywords
    - Rank of the document
    - Proximity of the keywords (shortest path)
    - N-gram overlap between passage and question

Architecture

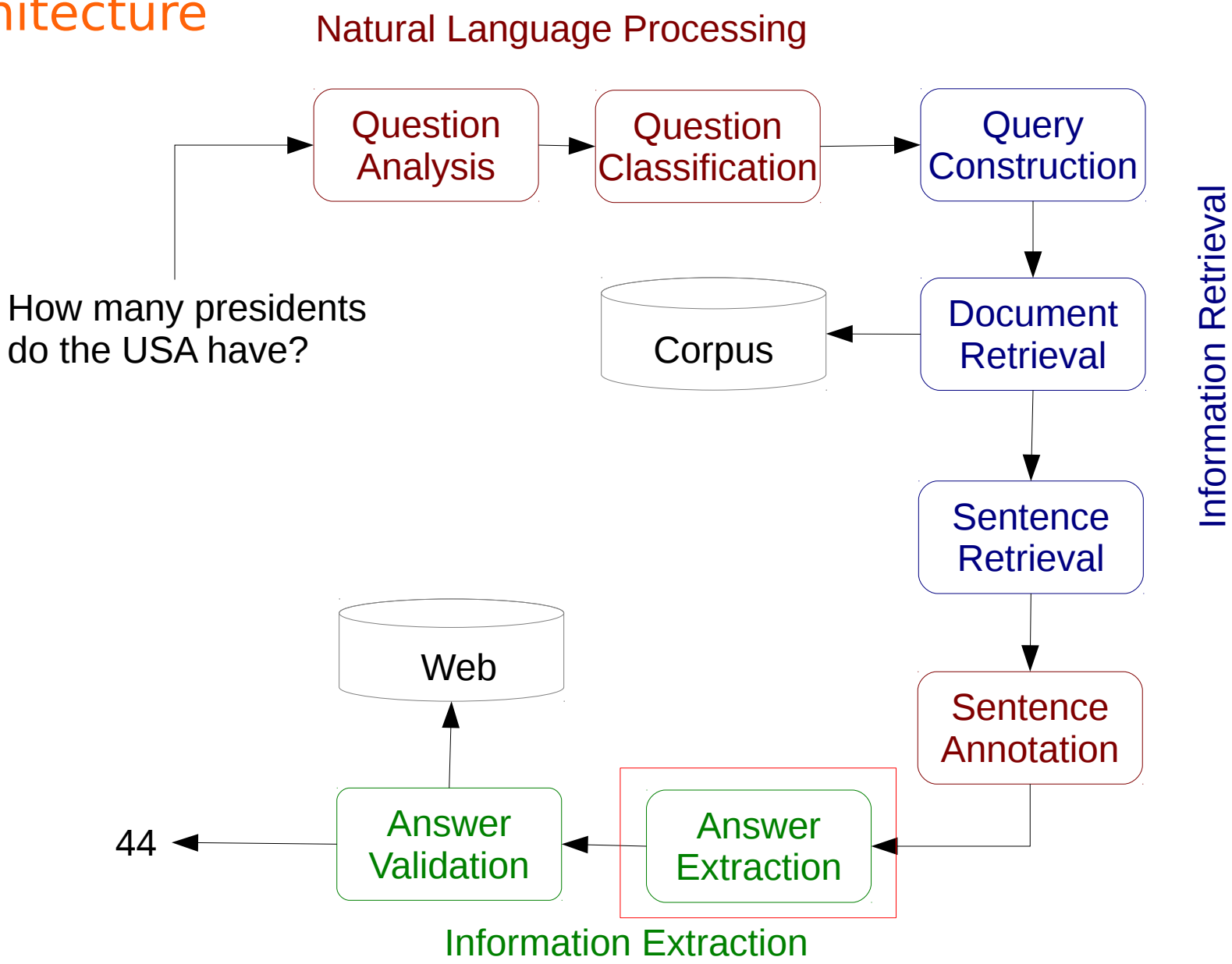


Sentence Annotation

- Similar to Question Analysis

- Annotate relevant sentences using linguistic analysis
  - Named-entity recognition
  - Syntactic parsing
  - Semantic role labeling

Architecture



Answer Extraction

- Extract candidate answers based on various information:
  - Question
    - Question Analysis: patterns
    - Question Analysis: syntactic parser
    - Question Analysis: semantic roles
    - Question Classification: question type
  - Sentence
    - Sentence Annotation: all annotated data

Answer Extraction

- Use extracted patterns
- Example:
  - Question: “In what country was Albert Einstein born?”

- Question Pattern: In what country was X born?
- Answer Pattern: X was born in Y.

- Example (Pattern):
    - S1: “Albert Einstein was born in 14 March 1879.”
    - S2: “Albert Einstein was born in Germany.”
    - S3: “Albert Einstein was born in a Jewish family.”

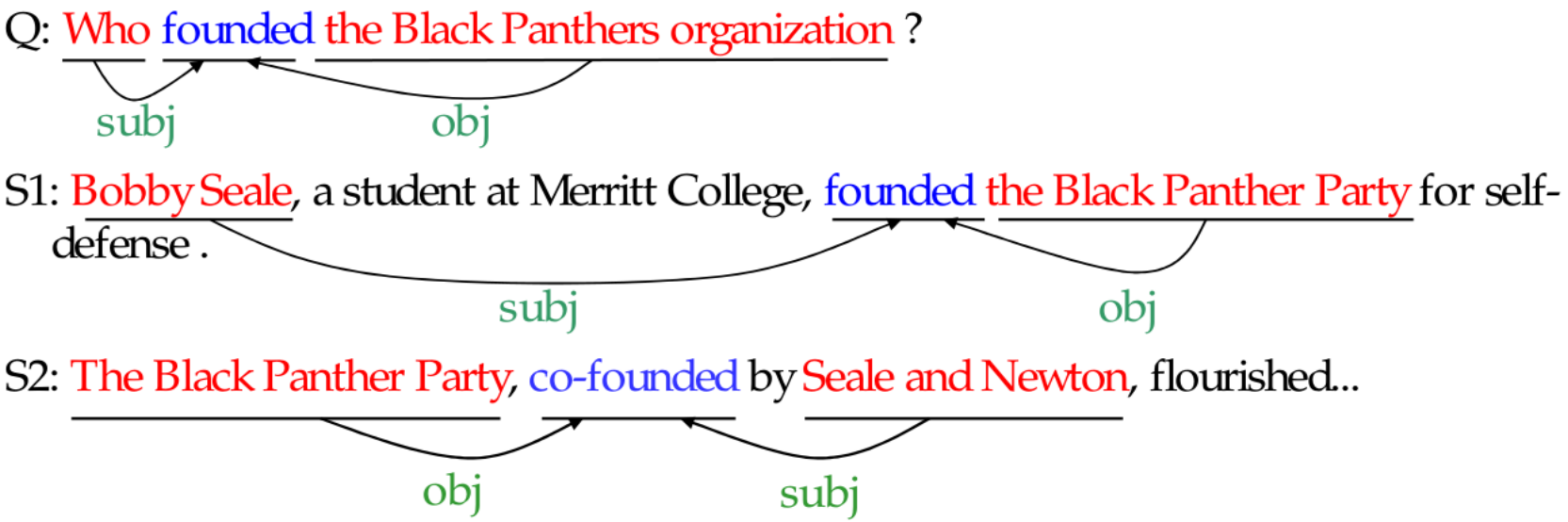
Answer Extraction

- Use question type and entity type
- Example:
  - Question: “In what country was Albert Einstein born?”
  - Question Type: LOCATION - Country

- Example (NER):
  - S1: “Albert Einstein was born in 14 March 1879.”
  - S2: “Albert Einstein was born in Germany.”
  - S3: “Albert Einstein was born in a Jewish family.”

Answer Extraction

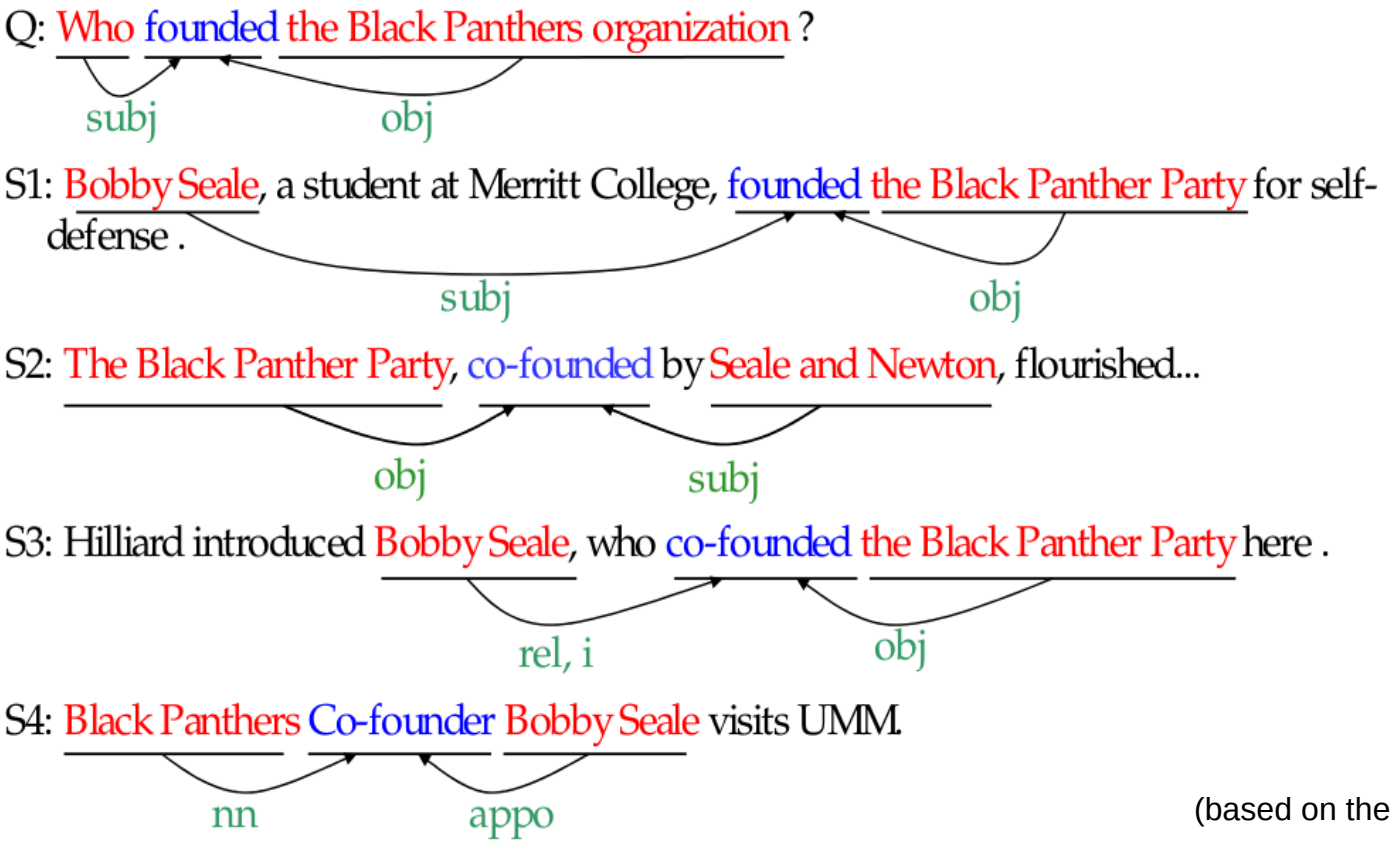
- Use syntactic parsing
  - Different wordings possible, but similar syntactic structure



(based on the work by Dan Shen)

Answer Extraction

- Use syntactic parsing
  - Many syntactic variations → need robust matching approach

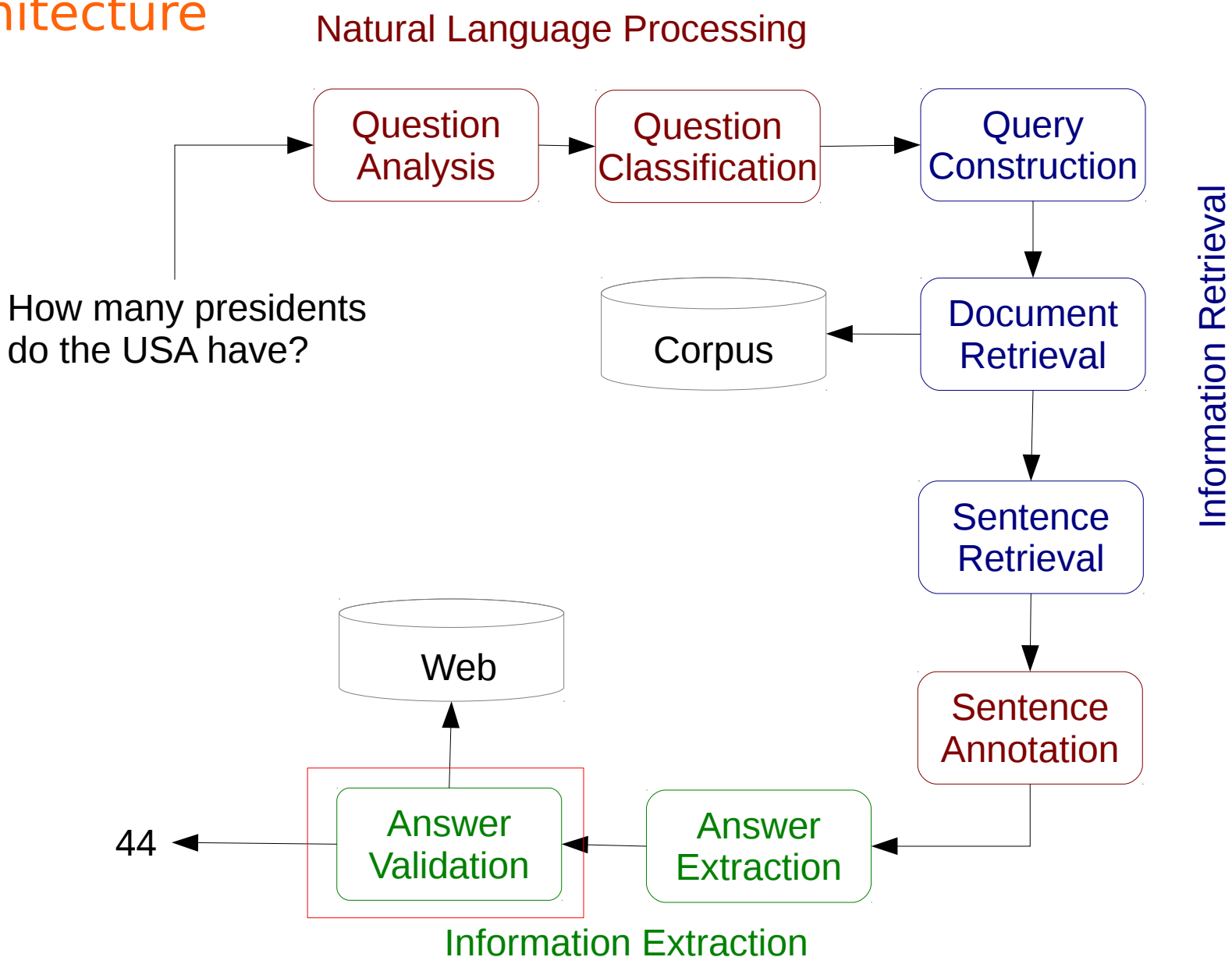


(based on the work by Dan Shen)

Answer Extraction

- Use semantic roles
- Example:
  - “Who$_{[Buyer]}$ purchased YouTube?$_{[Goods]}$”
- Example:
  - “ In 2006, YouTube$_{[Goods]}$ was purchased by Google$_{[Buyer]}$ for $1.65 billion.”

Architecture



Answer Validation

- Use the Web as a knowledge resource for validating answers

- Required steps
  - Query creation
  - Answer rating

Answer Validation

- Query creation
  - Combine the answer with a subset of the question keywords
  - Choose different combinations of subsets
    - Bag-of-Word
    - Noun phrase chunks
    - Declarative form

Answer Validation

- Example:
  - Question: “In what country was Albert Einstein born?”
  - Answer Candidate: Germany

- Queries:
  - Bag-of-Word:
    - Albert Einstein born Germany
  - Noun-Phrase-Chunks:
    - “Albert Einstein” born Germany
  - Declarative-Form:
    - “Albert Einstein born Germany”

Answer Validation

- Answer rating
  - Evaluate the query in a search engine
  - Analyze the result of the search engine
    - Count the results
    - Parse the result snippets
  - Other possibilities:
    - Use knowledge bases to find relations between the question keywords and the answer

Further Reading

- Book „Speech and Language Processing“
  - Chapters 23.1, 23.2