


Lecture 19

Question Answering



CS 6320

Outline

- Define the QA problem
- QA Architecture
- QA Performance
- How to extract an answer
- Logic proving an answer

Problem definition

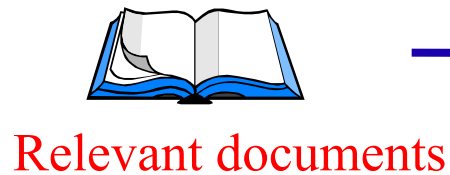
- Open-Domain QA
 - Find answers to open-domain natural language questions by searching a large collection of documents.
Documents: texts, web pages, digital libraries, maps, databases, images, etc.
- Canned-QA and Frequently Asked Questions
 - Map new questions into predetermined questions for which answers exist.

In this lecture we focus on open-domain QA.

QA is not only:



IE:



Who did:
What:
to Whom:
When:
Where:

IE + IR

Template

Simple example

- Easy

- Q: What is the fastest car in the world?

- Correct answer:

- .., the Jaguar XJ220 is the dearest (Pounds 415,000), **fastest** (217 mph / 350 kmh) and most sought-after **car** in the **world**.

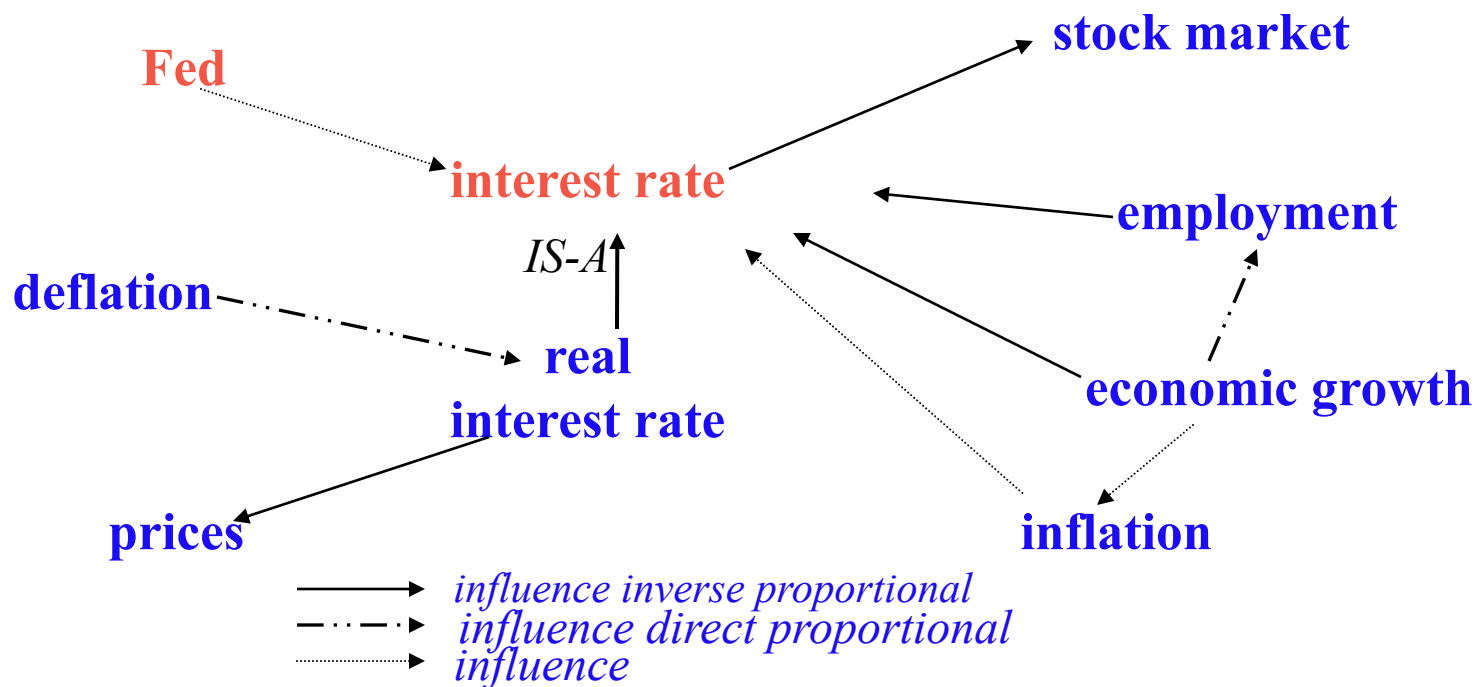
- Wrong answer:

- .. will stretch Volkswagen's lead in the **world's fastest**-growing vehicle market. Demand is expected to soar in the next few years as more Chinese are able to afford their own **cars**.

Complex example

- Hard
 - Q: Will the Fed change interest rate at their next meeting?

Needs knowledge acquisition and much more.



A taxonomy of QA systems *Moldovan et al., ACL 2002*

Class	Type	Example
1	Factual	Q33: What is the largest city in Germany? A: ... Berlin, the largest city in Germany ...
2	Simple - reasoning	Q198: How did Socrates die? A: ... Socrates poisoned himself...
3	Fusion - list	What are the arguments for and against prayer in school? Answer across several texts
4	Interactive - context	Clarification questions
5	Speculative	Q: Should the Fed raise interest rates at their next meeting? Answer provides analogies to past actions

Enabling Technologies and Applications

Enabling Technologies

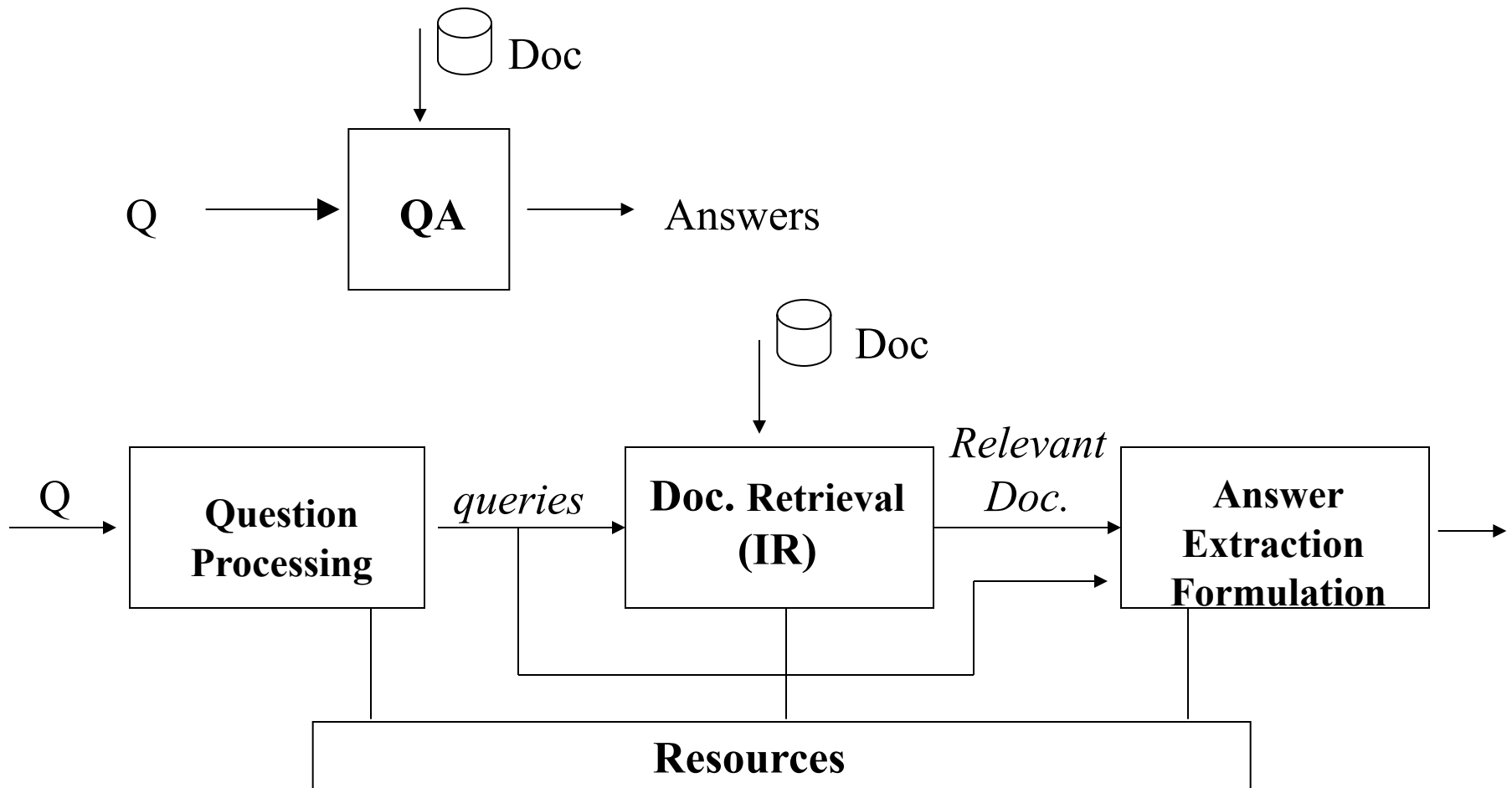
- POS Tagger
- Parser
- WSD
- Named Entity
- Inference engine
- WordNet
- Knowledge Acquisition
- Knowledge Classification
- Language generation
- Database retrieval

QA

Applications

- ◆ Smart agents
- ◆ Situation Management
- ◆ E-Commerce
- ◆ Summarization
- ◆ Tutoring / Learning
- ◆ Personal Assistant in business, private
- ◆ On-line Documentation
- ◆ On-line Troubleshooting
- ◆ Semantic Web
- ◆ Customer Relation Mgmt

A Generic QA Architecture



Question Processing

1. Determines the type of question
2. Determines the type of answer expected
3. Determine question focus
4. Translates the question into queries for the search engine

Question type

- Question types:
 - Definition question
 - Math question
 - List question
 - When, where, how long, how much,
 - What questions (are more difficult)
- Build a question classifier:
 - Supervised learning
 - Rule-based
 - Neural networks

Answer Type

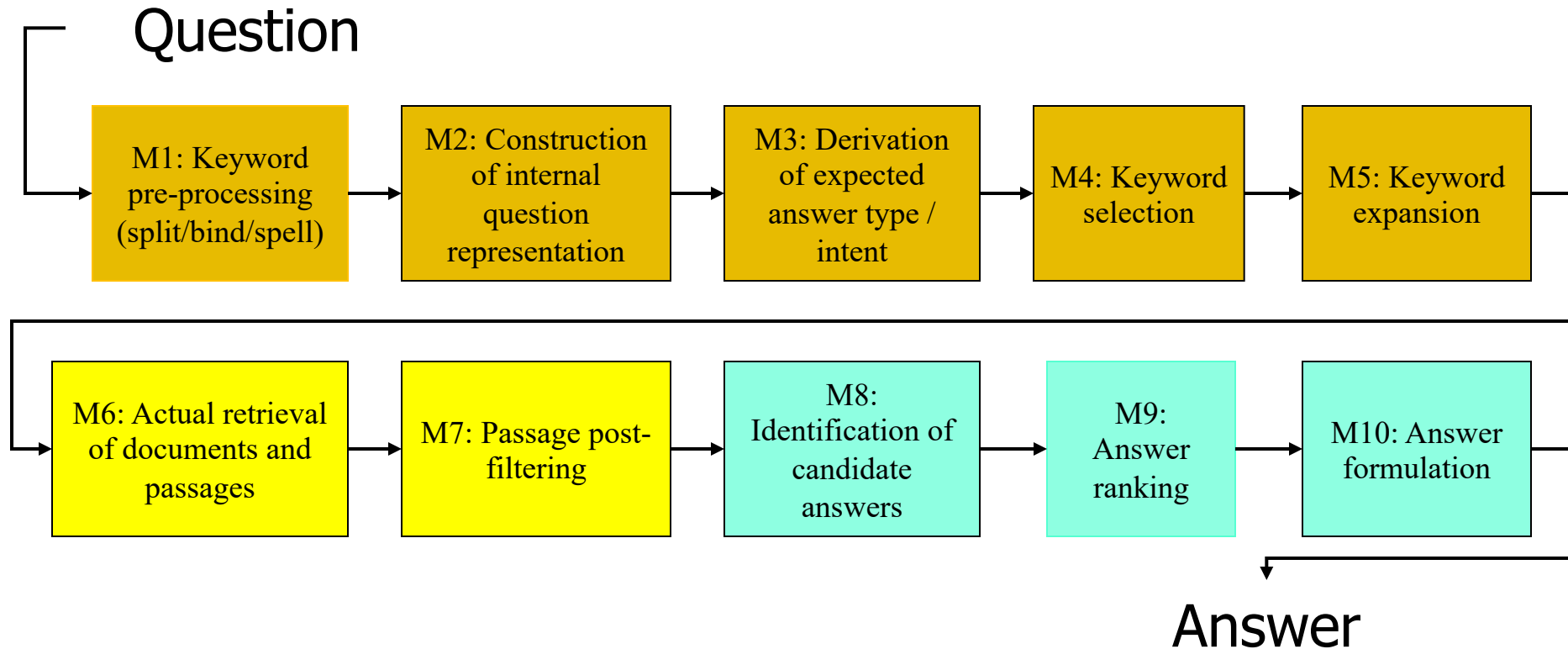
- **Answer types:** Any Name Entity can be an answer type
Ex: person, location, time, occupation, money, vehicles, etc
- Build a **taxonomy** of answer types, and link it to question types

Question Focus

- Question focus is the entity central to the question, what the quest is all about
- Q: What's the height of Mount Everest ?
- A: Mount Everest's height is 29,029 ft.

Architecture modules

Serial System Architecture



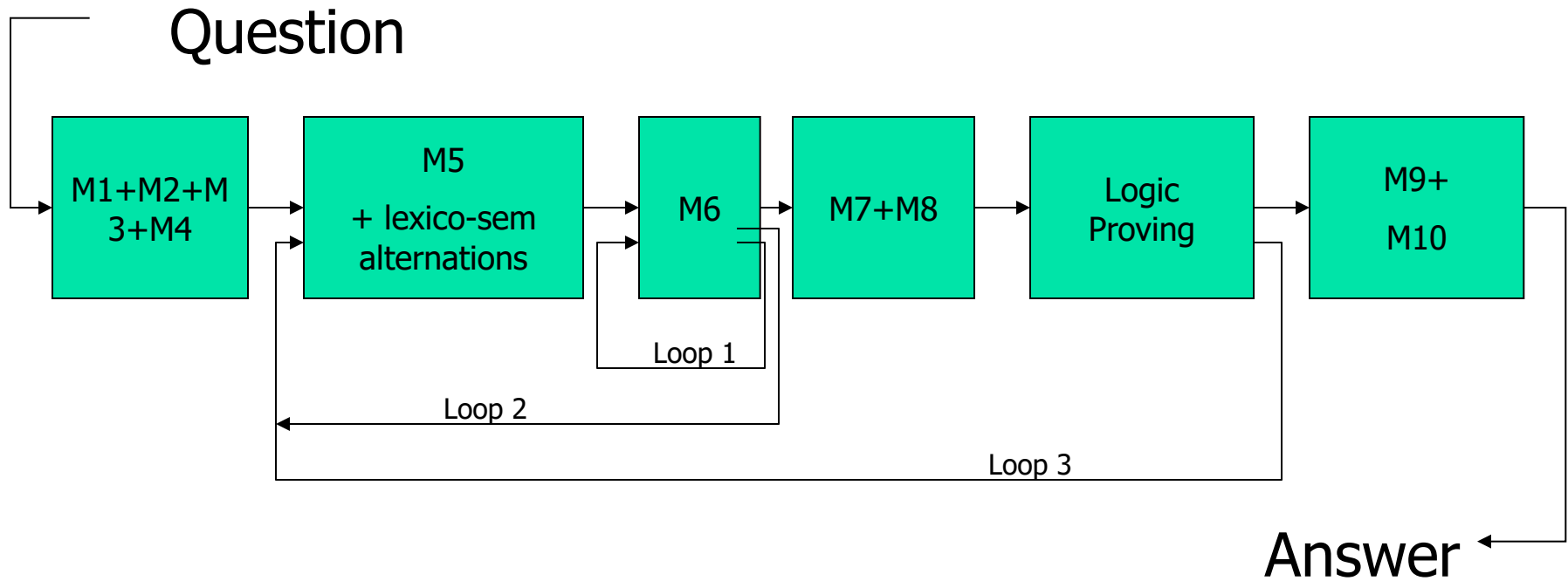
Performance Analysis

Distribution of Errors

Module	Module definition	Errors (%)
M1	Keyword pre-processing (split/bind/spell check)	1.9
M2	Construction of internal question representation	5.2
M3	Derivation of expected answer type	36.4
M4	Keyword selection (incorrectly added or excluded)	8.9
M5	Keyword expansion desirable but missing	25.7
M6	Actual retrieval (limit on passage number or size)	1.6
M7	Passage post-filtering (incorrectly discarded)	1.6
M8	Identification of candidate answers	8.0
M9	Answer ranking	6.3
M10	Answer formulation	4.4

Performance Analysis

Architecture with Feedbacks



Performance Analysis

Impact of System Parameters

Feedback added	Precision (MRR)	Incremental enhancement
none	0.421=b	0%
Passage retrieval (loop 1)	0.468=b1	b+11%
Lexico-semantic (loop 2)	0.542=b2	b1+15%
Proving (loop 3)	0.572=b3	b2+5%

Keywords Alternations

- **Morphological alternations**

- Q209 *Who invented the paper clip?*
- Query: [paper AND clip AND (invent OR inventor)]

- ◆ **Lexical alternations**

- Q206 *How far is the moon?*
- Query: [distance AND moon]

- ◆ **Semantic alternations**

- Q258 *Where do lobsters like to leave?*
- Query: [lobster AND (like OR prefer)]

Question reformulations

- Q397 When was the Brandenburg Gate in Berlin built?
- Q814 When was Berlin's Brandenburg gate erected?

- Q411 What tourist attractions are there in Reims?
- Q711 What are the names of the tourist attractions in Reims?
- Q712 What do most tourists visit in Reims?
- Q713 What attracts tourists in Reims?
- Q714 What are tourist attractions in Reims?
- Q715 What could I see in Reims?
- Q716 What is worth seeing in Reims?
- Q717 What can one see in Reims?

Question reformulations (cont'd)

- Question similarity algorithm:
 - Q: $W_1 W_2 \dots W_n$
 - Q': $W_1' W_2' \dots W_m'$
- Lexical relations between W_i and W_j' based on:
 - Common morphological root
 - ♦ E.g. *owns, owner*
 - ♦ Q416 *Who owns CNN?*
 - ♦ Q741 *Who is the owner of CNN?*
 - WordNet synonyms
 - ♦ E.g. *gestation and pregnancy*
 - ♦ Q763 *How long is human gestation?*
 - ♦ Q765 *A normal pregnancy lasts how many months?*

Question reformulations (cont'd)

- WordNet hypernyms
 - *E.g. erect and build*
 - Q814 *When was Berlin's Brandenburg gate erected?*
 - Q397 *When was the Brandenburg Gate in Berlin build?*

TREC Examples

■ Question

What causes the greenhouse effect?

■ Paragraph

“The layer of the atmosphere that produces the greenhouse effect has undergone concentration of gases of 30 percent in the last 100 years, which might raise the temperature of the Earth” said the environmentalist in an interview with the RPC Radio.

■ Answer

The layer of the atmosphere

■ Question

What is the largest volcano in Europe?

■ Paragraph

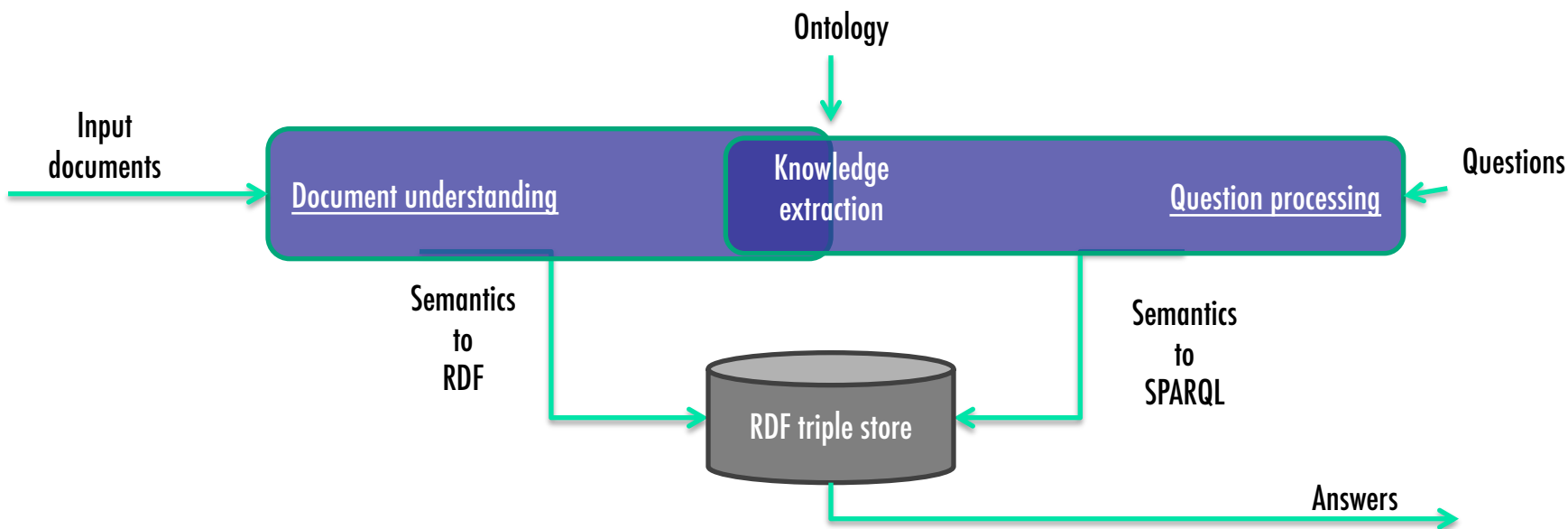
The crater, which burst Thursday, was formed in 1971 and is the youngest on Mount Etna. The 3.5-kilometer-high Etna, the largest and most active volcano in Europe, has erupted frequently since the beginning of this year and Thursday's eruption was the 59th, a record high in 30 years

■ Answer

Etna

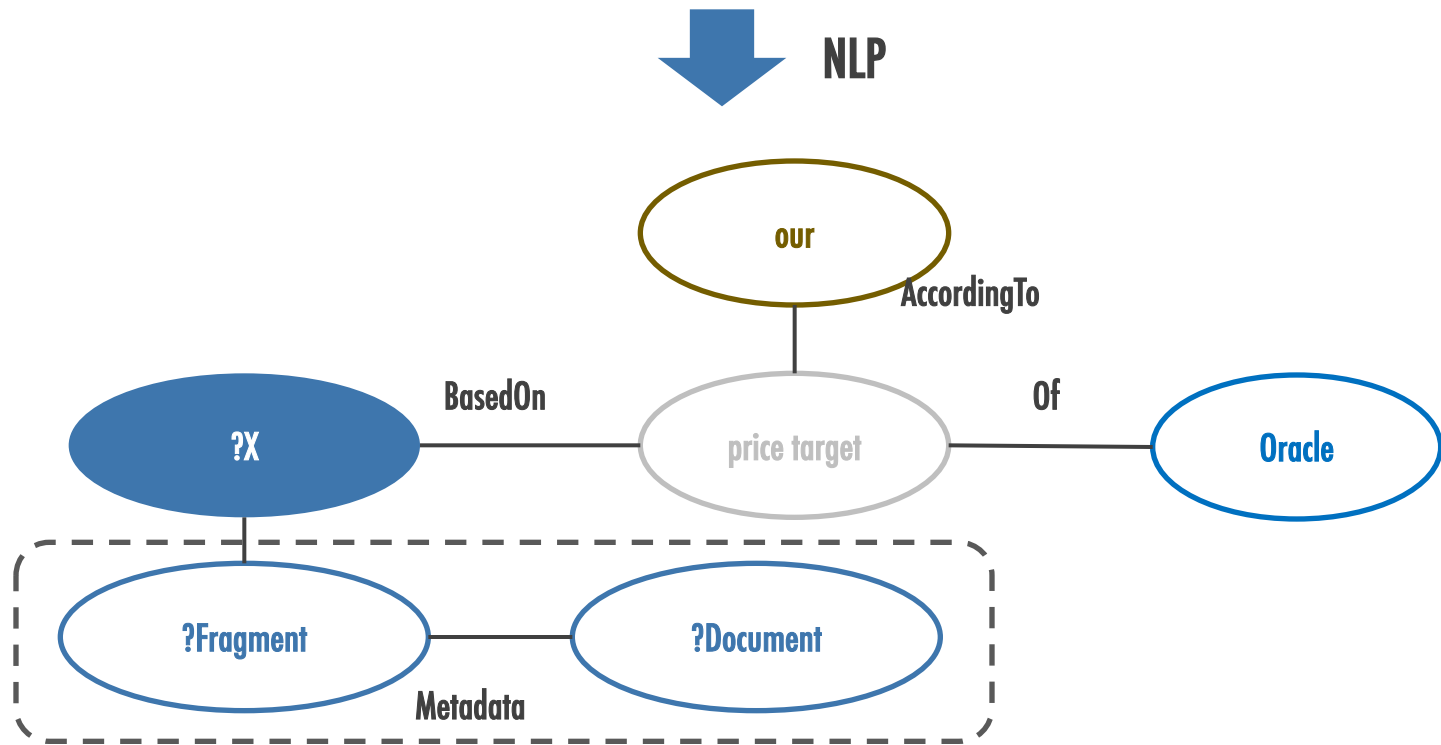
Knowledge Graph QA

- **Ontology-driven customized knowledge extraction**
 1. Indexing of document knowledge
 2. Processing of natural language questions to create queries



Text to SPARQL

What supports **our** **Oracle** price target?



LFT Definitions

Predicate – for every noun, verb, adjective and adverb.

Verb predicates:

action/state/event – predicate(e , x_1 , x_2)

e – eventuality

x_1 – syntactic subject

x_2 – syntactic direct object

Example:

{supporter, protagonist, champion, admirer, booster}
(a person who backs a politician)

Fix – slot allocation

Person: $n(x_1)$ & back: $v(e_1, x_1, x_2)$ & politician: $n(x_2)$

LFT Example 1

Q: Who shot Billy the Kid?

P1: The scene called for Phillips' character to be saved from a lynching when Billy the Kid (Emilio Estevez) shot the rope in half just as he was about to be hanged.

P2: In 1881, outlaw William H. Bonney Jr., alias Billy the Kid, was shot and killed by Sheriff Pat Garrett in Fort Summer, N.M.

LFT:

Q : PERSON(x_1) & shoot(e_1, x_1, x_2) & Billy _ the _ Kid(x_2)

P1 : Billy _ the _ Kid(x_1^l) & shoot(e_1^l, x_1^l, x_2^l) & rope(x_2^l)

P2 : Billy _ the _ Kid(x_1^l) & shoot(e_1^l, x_2^l, x_1^l) & Sheriff _ Pat _ Garret(x_2^l)

Logic proof for P2 succeeds and agent of shooting Sheriff_Pat_Garret unifies with PERSON

$e1 = e1'$

$x1 = x2'$

$x2 = x1'$

LFT Example 2

Q045: When did Lucelly Garcia, former ambassador of Columbia to Honduras die?

A: "Several gunmen on a highway leading to the Columbian city of Ibaque murdered Colombian ambassador to Honduras Lucelly Garcia today."

WordNet axioms:

$Colombian(x_1) \leftrightarrow of(x_1, x_2) \& Colombia(x_2)$
 $kill(e, x_1, x_2) \leftrightarrow cause(e_1, x_1, x_2) \& die(e_2, x_2)$
 $kill(e, x_1, x_2, x_3) \leftrightarrow put(e, x_1, x_2, x_3) \& to(e, x_3) \& death(x_3)$
 $murder(e_1, x_1, x_2) \leftrightarrow kill(e_1, x_1, x_2) \& intentionally(e) \& with(e, x_3) \& premeditation(x_3)$

Logical Proof

Step 0:

QLF: $Lucelly_Garcia(x_1) \& former(x_1) \& ambassador(x_1) \& of(x_1, x_2) \& Colombia(x_2) \& Honduras(x_3) \& die(e_1, x_2) \& TIME - STAMP(e_1)$

ALF: $gunman(x_2^l) \& murder(e_1^l, x_2^l, x_1^l) \& Colombian(x_1^l) \& ambassador(x_1^l) \& to(x_1^l, x_3^l) \& Honduras(x_3^l) \& Lucelly_Garcia(x_1^l) \& TIME - STAMP(e_1^l)$

QLF: $of(x_1^l, x_2) \& Colombia(x_2) \& die(e_1, x_1^l)$

Step 1:

ALF: $gunman(x_2^l) \& murder(e_1^l, x_2^l, x_1^l) \& of(x_1^l, x_7^l) \& Colombia(x_7^l) \& ambassador(x_1^l) \& to(x_1^l, x_3^l) \& Honduras(x_3^l) \& Lucelly_Garcia(x_1^l) \& TIME - STAMP(e_1^l)$

Logical Proof (cont'd)

Step 2:

QLF: $die(e_1, x_1^l)$

ALF: $gunman(x_2^l) \& kill(e_1^l, x_2^l, x_1^l) \& intentionally(e_1^l) \& with(e_1^l, x_8^l) \& premeditation(x_8^l) \& of(x_1^l, x_7^l) \& Colombia(x_7^l) \& ambassador(x_1^l) \& to(x_1^l, x_3^l) \& Honduras(x_3^l) \& Lucelly_Garcia(x_1^l) \& TIME - STAMP(e_1^l)$

QLF: $die(e_1, x_1^l)$

Step 3:

ALF: $gunman(x_2^l) \& cause(e_2^l, x_2^l, x_3^l) \& die(e_3^l, x_1^l) \& intentionally(e_1^l) \& with(e_1^l, x_8^l) \& premeditation(x_8^l) \& of(x_1^l, x_7^l) \& Colombia(x_7^l) \& ambassador(x_1^l) \& to(x_1^l, x_3^l) \& Honduras(x_3^l) \& Lucelly_Garcia(x_1^l) \& TIME - STAMP(e_1^l)$

$e_1 = e'$ **Answer:** Today (the date of the article)