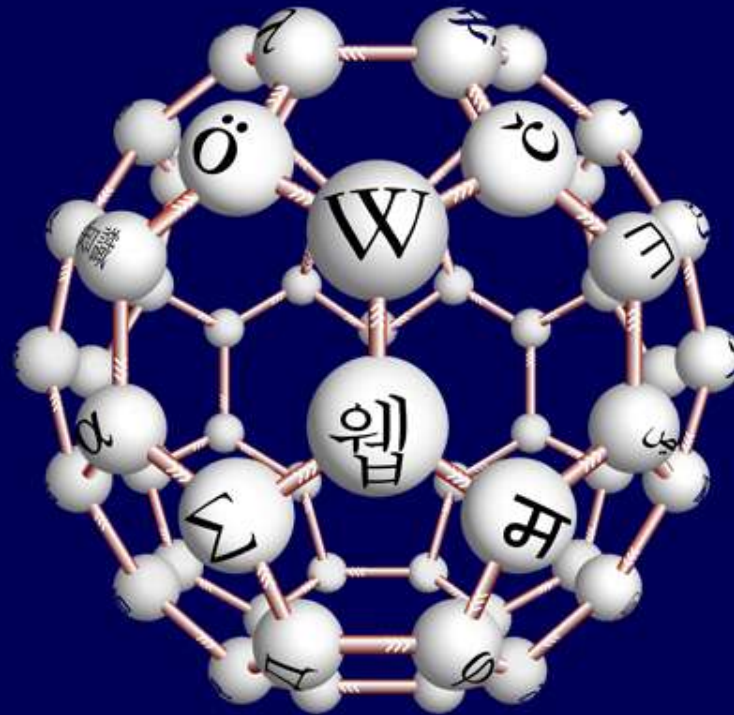


# کارگاه آموزشی وب معنایی



زمستان ۸۸

<http://wtlab.um.ac.ir>



Session 2: Applications

# Information Retrieval

Majid Sazvar

[sazvar@stu-mail.um.ac.ir](mailto:sazvar@stu-mail.um.ac.ir)

WTLAB

Ferdowsi University of Mashhad

2009

# Outline



- What is Information Retrieval?
- Retrieval Process
- Some Problems and Solutions
- Semantic Retrieval Process
- Semantic Search
- Question Answering

# What is Information Retrieval?



Information retrieval (IR) is **finding material** (usually documents) of an **unstructured nature** (usually text) that satisfies an **information need** from within **large collections** (usually stored on computers).

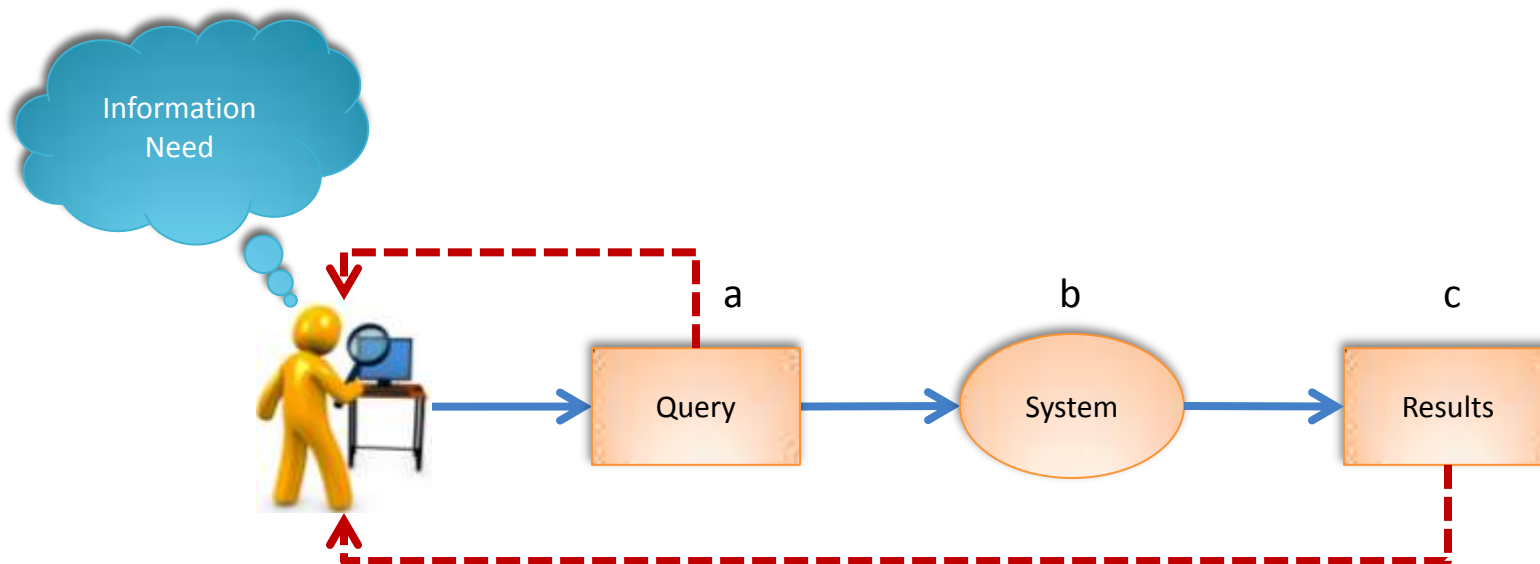
Information retrieval (IR) deals with the **representation, storage, organization** of, and **access** to information items.

Web search engines are the most visible IR applications.

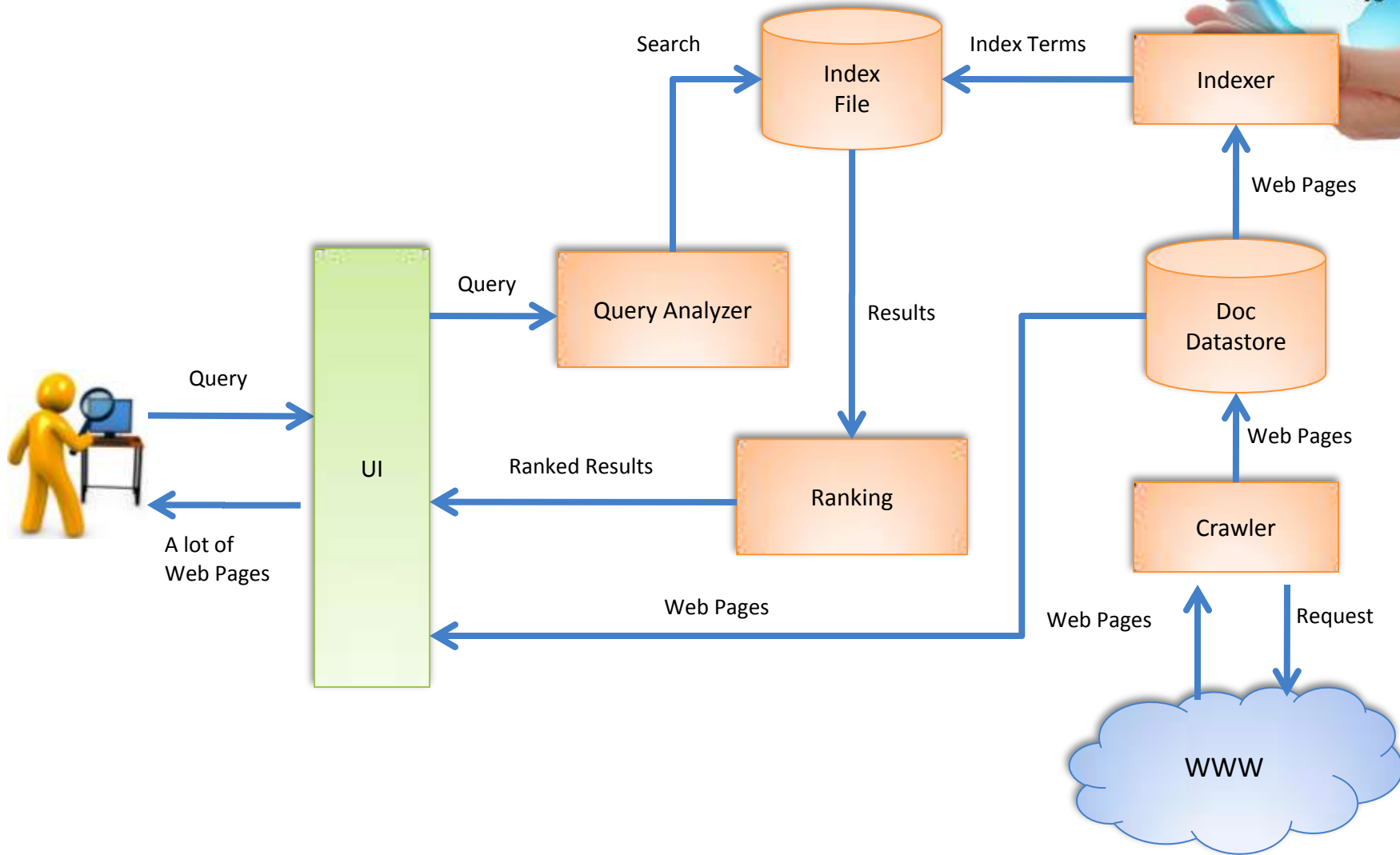
# Retrieval Process



- a) query construction
- b) search algorithm of the system
- c) presentation of the results



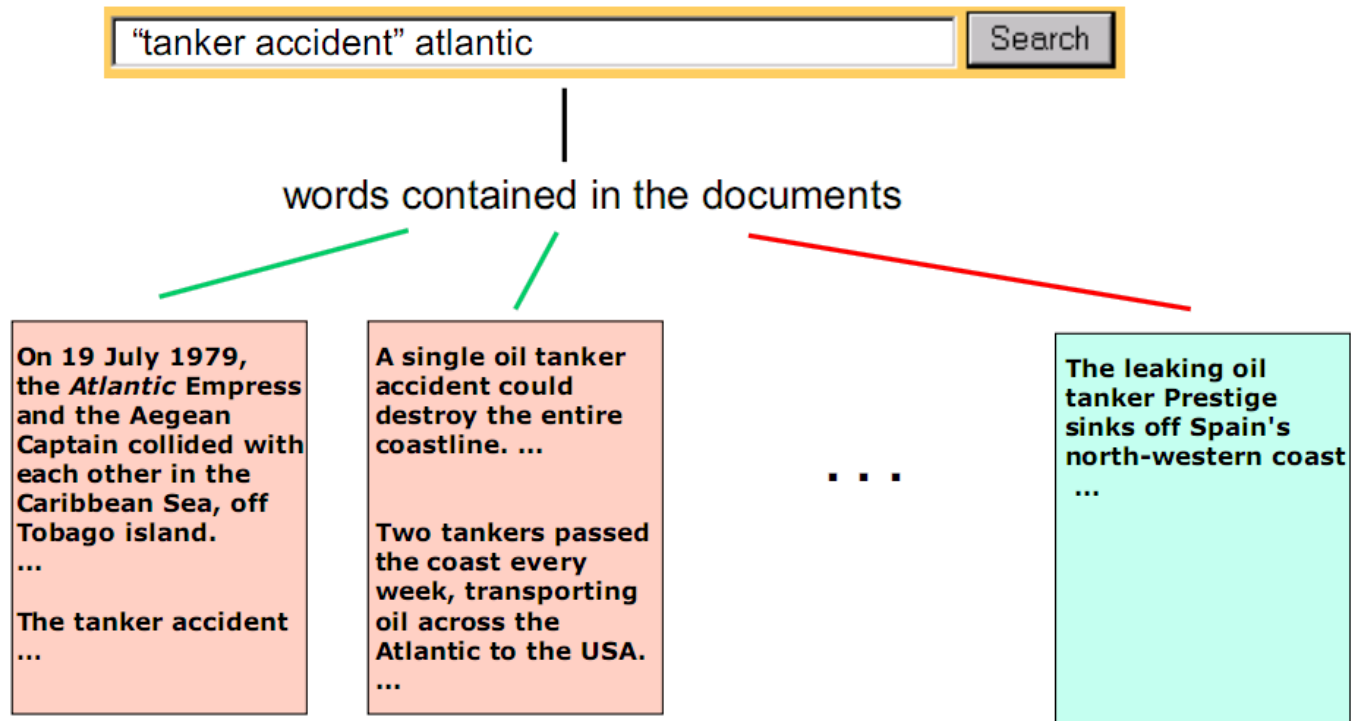
# Retrieval Process (2)



# Some Problems



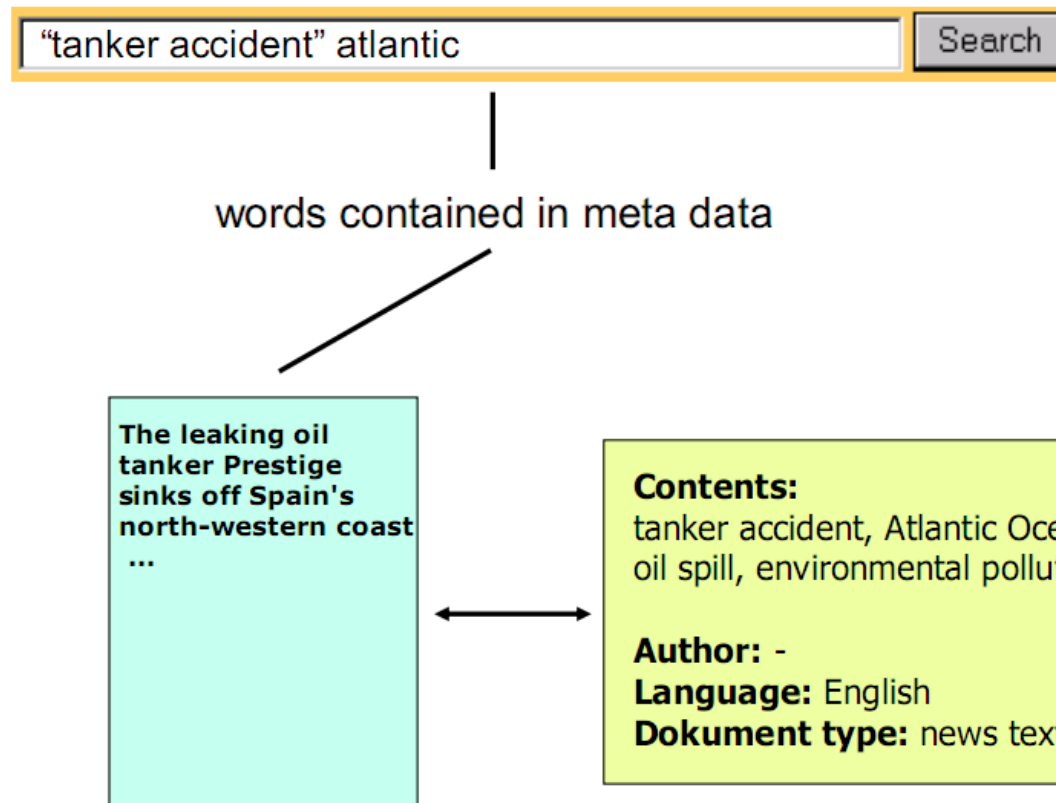
- Information search with search engines



# Some Problems (2)



- Using meta data to describe document contents

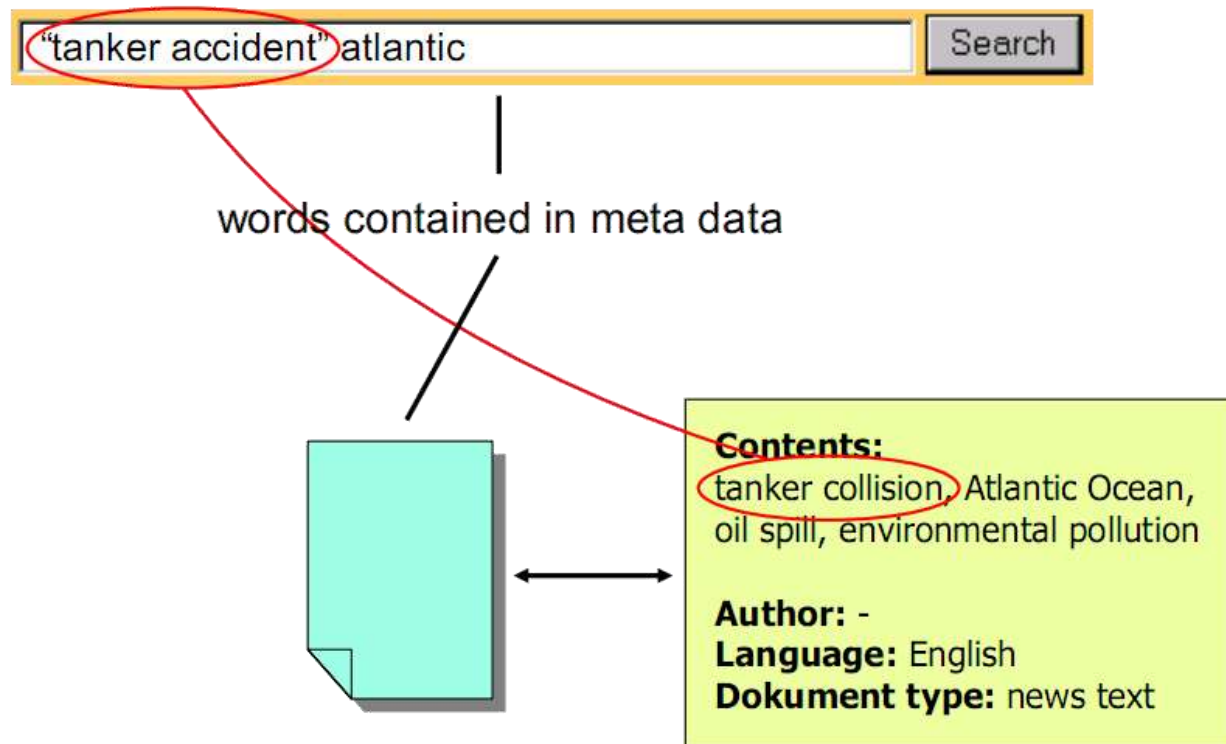




# Some Problems (3)



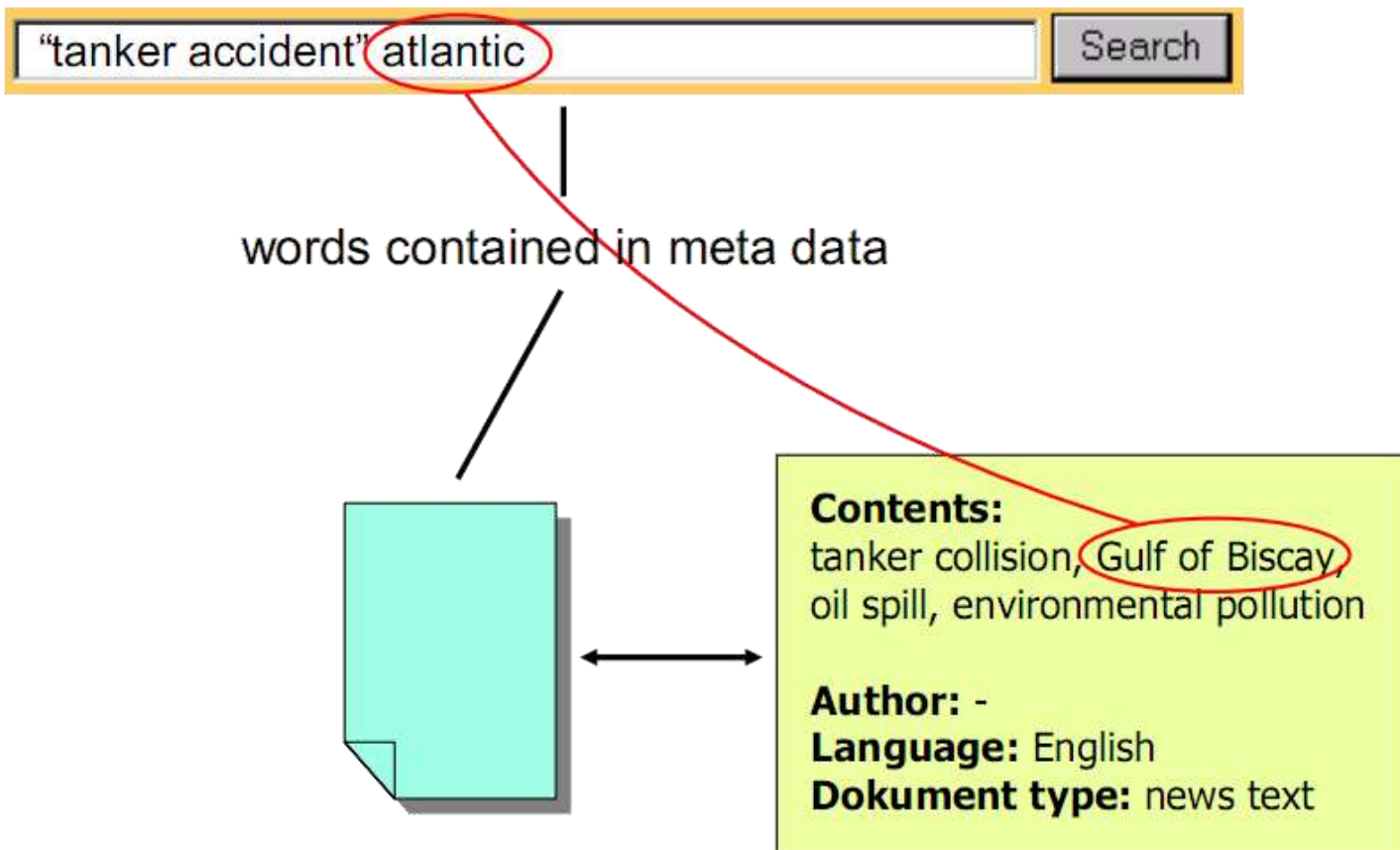
- Different terms in meta data and query formulation



# Some Problems (4)



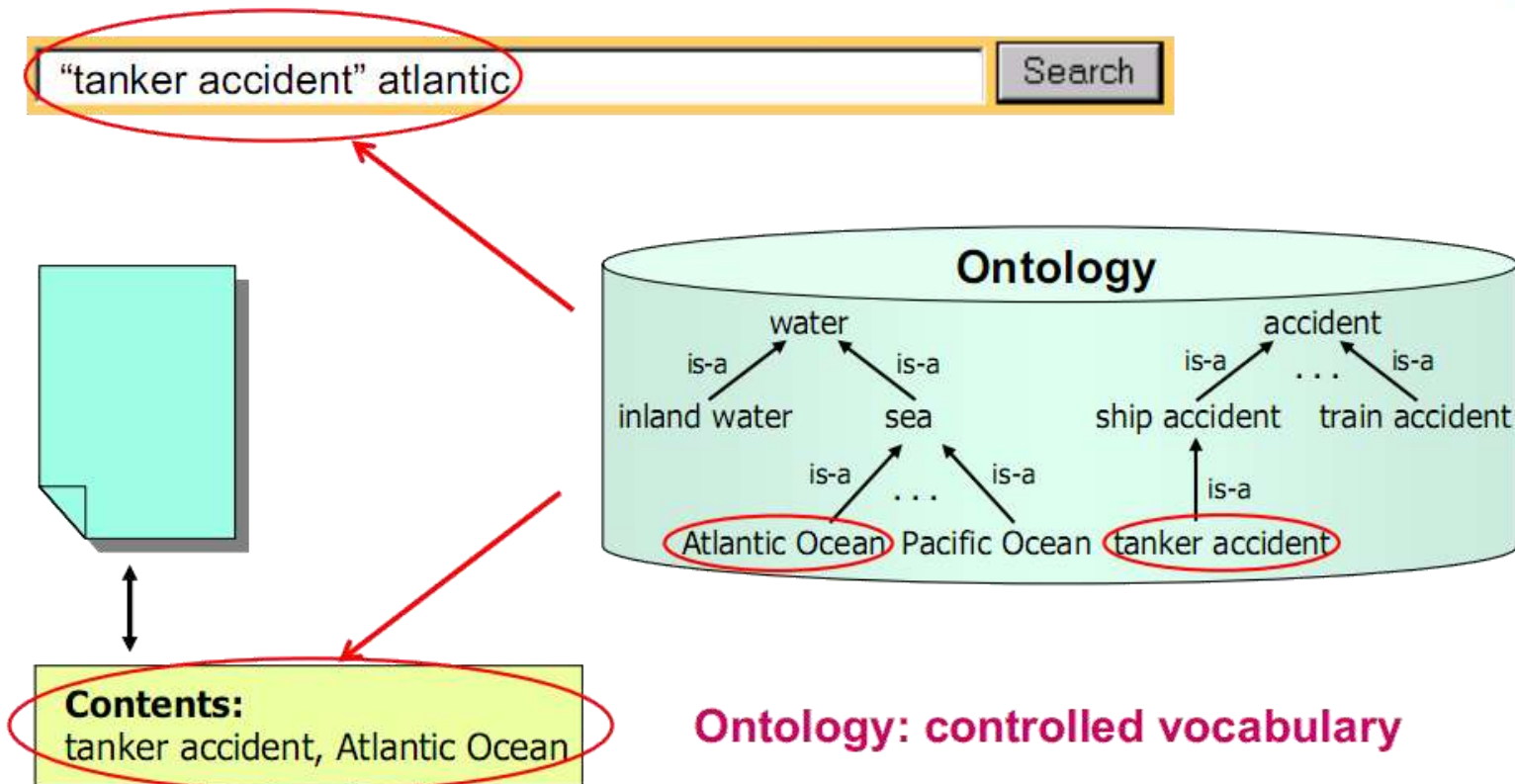
- Different vocabulary for indexing and querying



# Some Problems (5)



- Controlled vocabulary for indexing and querying

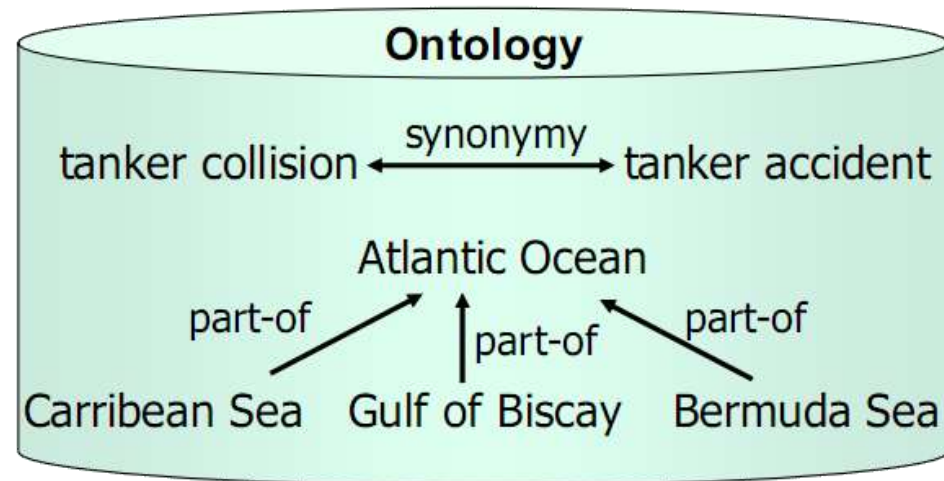


# Some Problems (6)



- Using background knowledge to extend query

**automatic  
query extension**

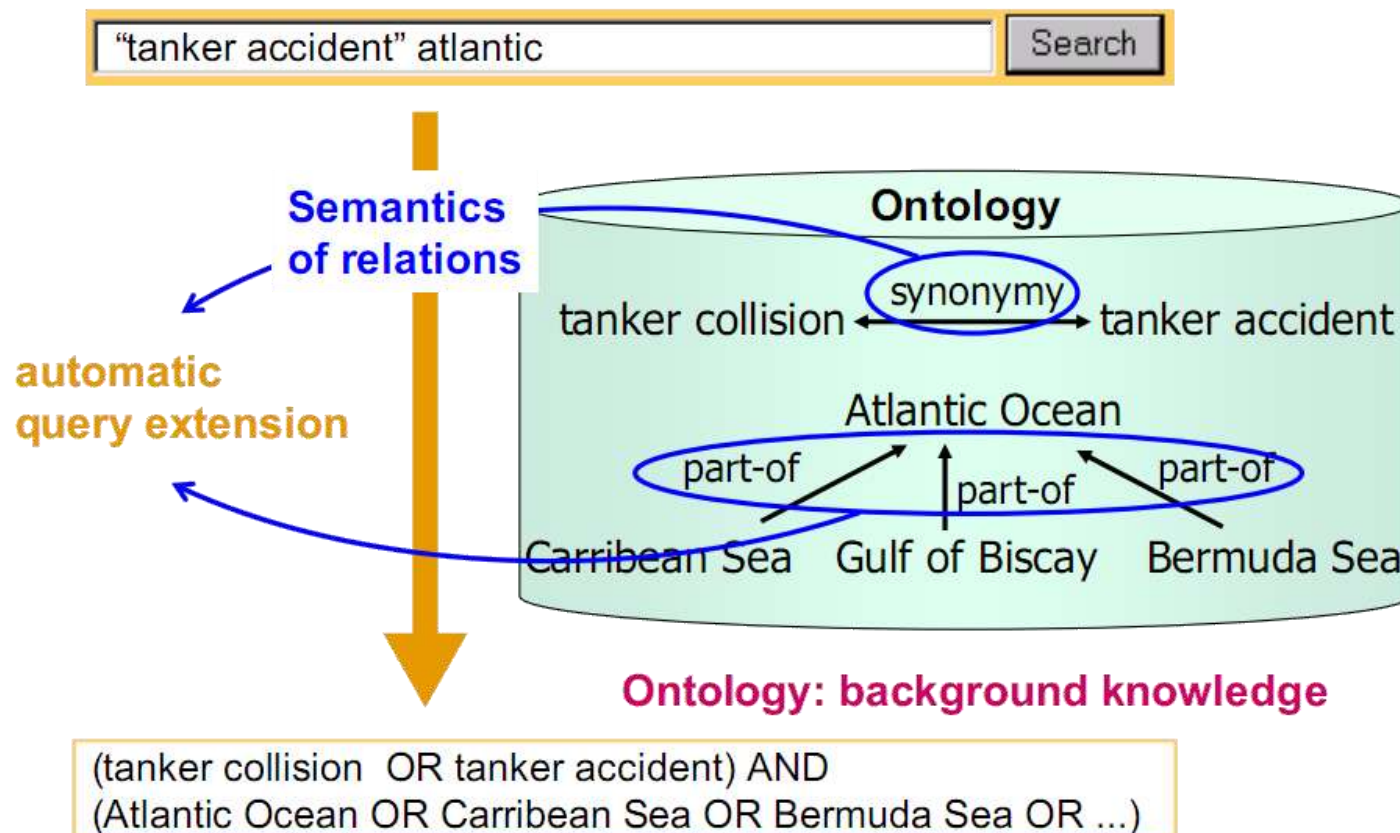


**Ontology: background knowledge**

# Some Problems (7)



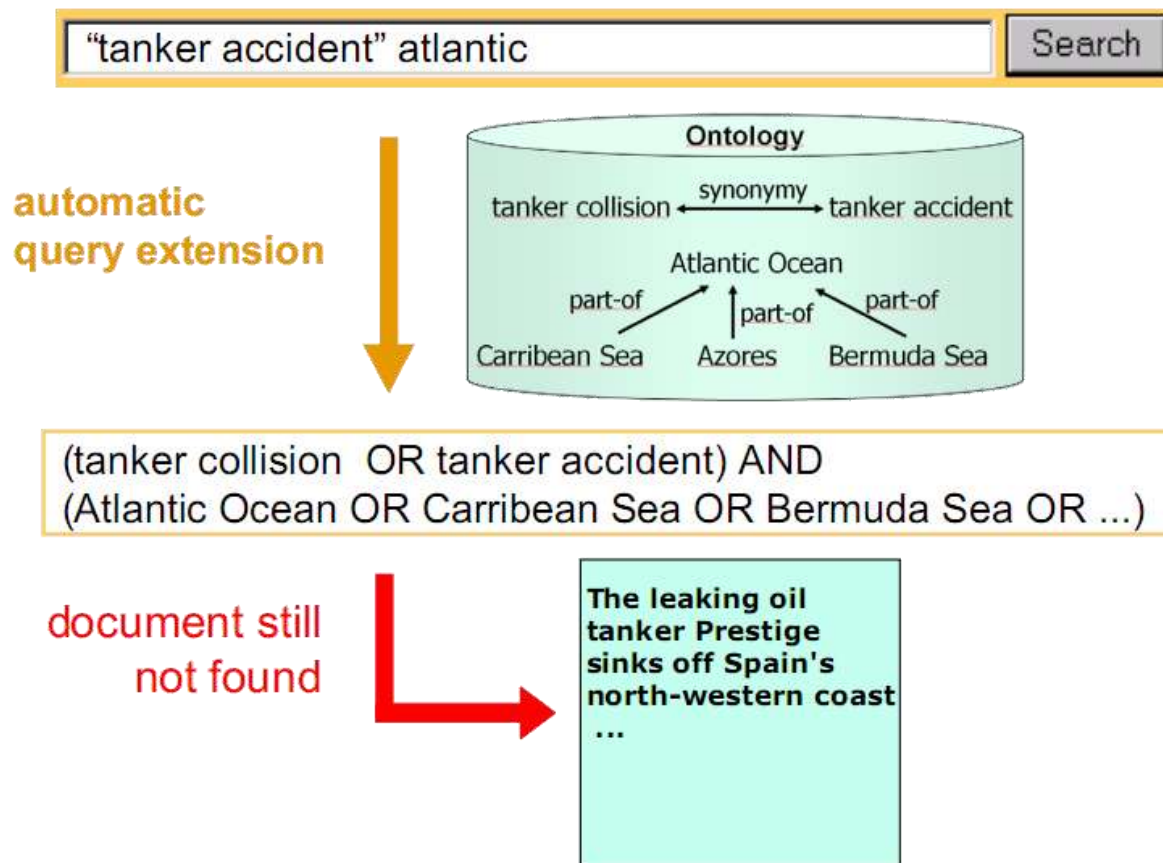
- Using background knowledge to extend query



# Some Problems (8)



- Required background knowledge can be complex



# Therefore, what is needed?

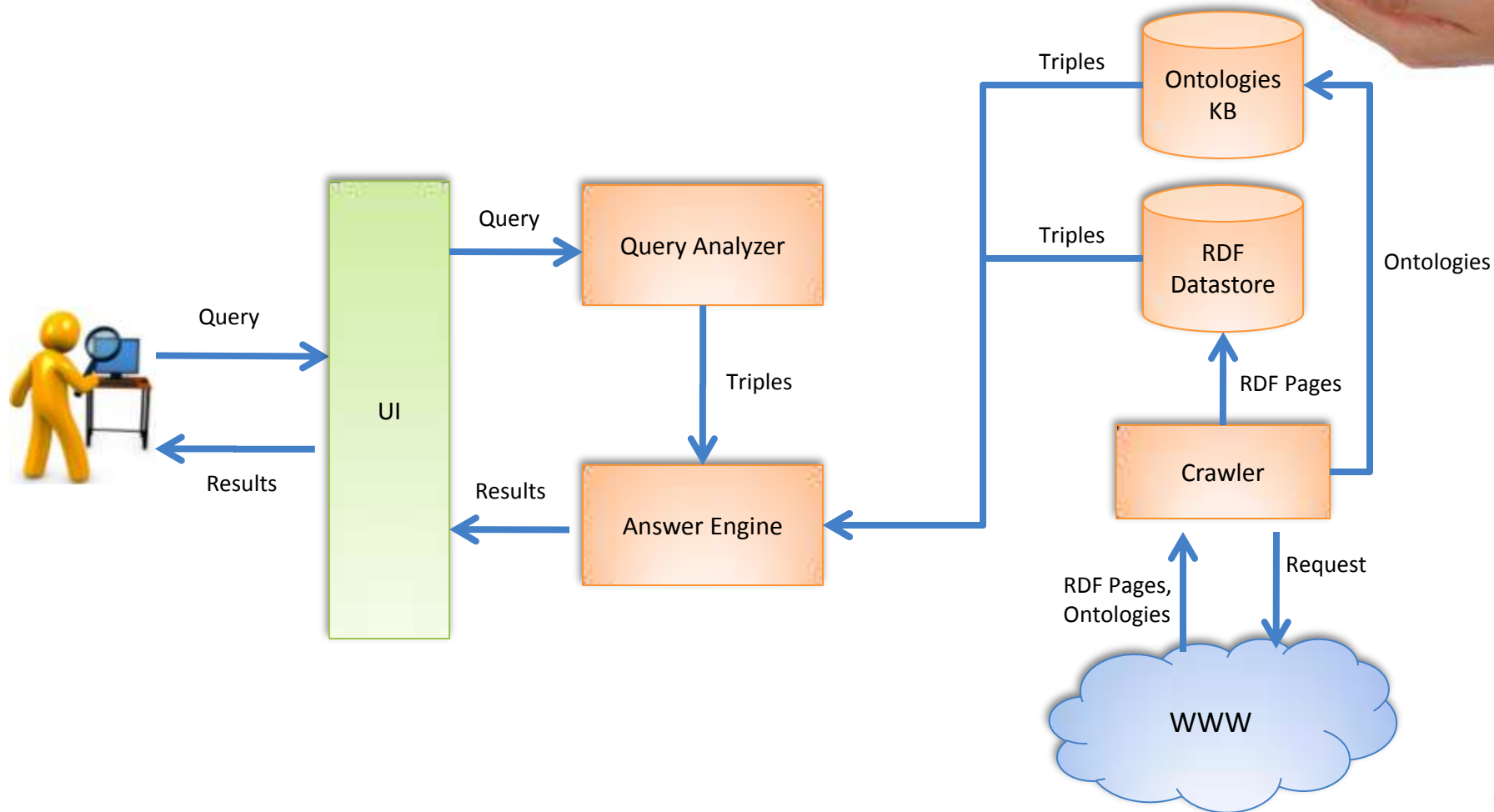


- Assign meta data to information objects
- Content description with **RDF** concepts and relations between them

- Provision of background knowledge
- Provision of the **RDF Schema, OWL, Rules** for query extension, ontology integration, etc.



# Semantic Retrieval Process





# Semantic Search



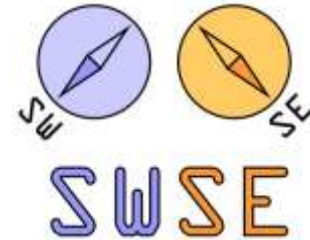
- Application of semantic technologies such as ontology and resource description languages to IR tasks is typically referred to as **Semantic Search**.

# Types of Semantic Search Engines



- Human-Oriented
  - E.g. Falcon, SWSE
- Application-Oriented
  - E.g. Swoogle, Watson

Falcons



Swoogle  
semantic web search

watson  
exploring the semantic web

# Some examples



Falcons

Object [Concept](#) [Document](#)


Objects 1 - 10 of 65,847 for your search **iran** (0.08 seconds)

## Iran - Concept

- label: **Iran**
- type: Concept

<http://www4.wiwiss.fu-berlin.de/bookmashup/subject/Iran>

## Iran - State, State, Democracy

- label: **Iran**
- comment: De Islamitische Republiek **Iran** (Perzisch: جمهوری اسلامی ایران) (Getranslitereerd: Djoemhūri-ye Eslāmi-ye Īrān) of **Iran** (ایران) is een land in het Midden-Oosten."
- sameAs: <http://www4.wiwiss.fu-berlin.de/factbook/resource/Iran>
- image: 
- type: Democracy

<http://dbpedia.org/resource/Iran>

## Counties in Iran - Concept

- label: Counties in **Iran**
- preferred label: Counties in **Iran**
- has broader: Subdivisions of **Iran**
- type: Concept
- has broader: Counties by country

[http://dbpedia.org/resource/Category:Counties\\_in\\_Iran](http://dbpedia.org/resource/Category:Counties_in_Iran)

## History of Iran - Concept

- label: History of **Iran**
- preferred label: History of **Iran**
- has broader: **Iran**
- type: Concept

### Type

#### Any type

- Abstraction
- Agent
- Concept
- Country
- Group
- Items
- Noun Synset
- Ontology
- Organization
- Person
- Physical Entity
- Social Entity
- Social Group
- Spatial Thing
- Subject

# Some examples (2)



RelFinder

between examples

(1) Albert Einstein

(2) Kurt Gödel

add clear Find Relations

Filter by:

length class link conn...

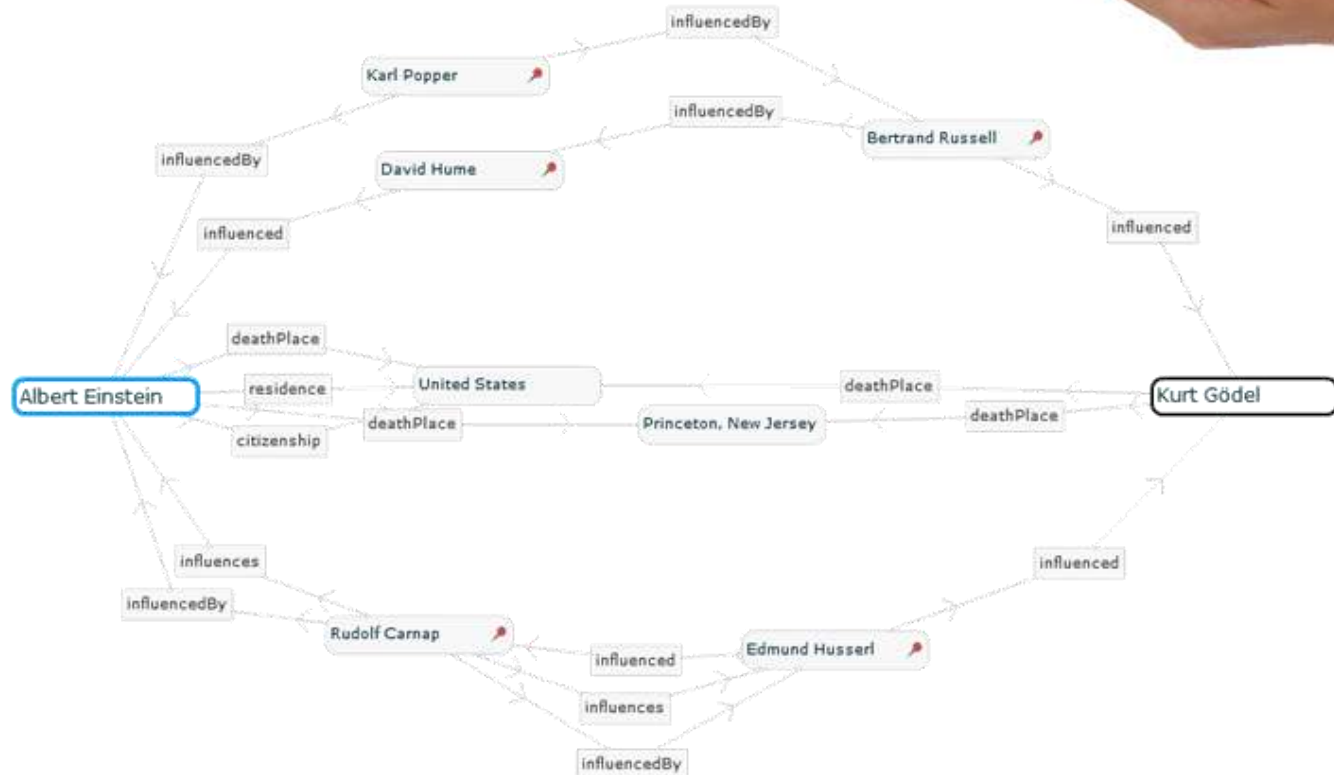
number of objects	num	vi
1	4/4	☺
2	8/9	☺

**Albert Einstein** en

More Infos: [dbpedia.org](http://dbpedia.org)  
[en.wikipedia.org](http://en.wikipedia.org)



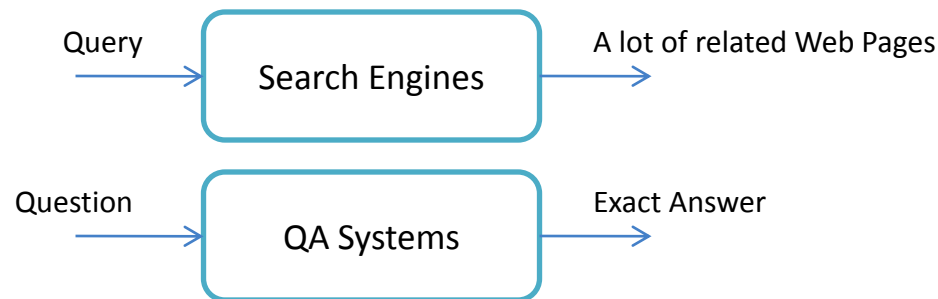
Albert Einstein (14 March 1879 – 18 April 1955)



# Question Answering



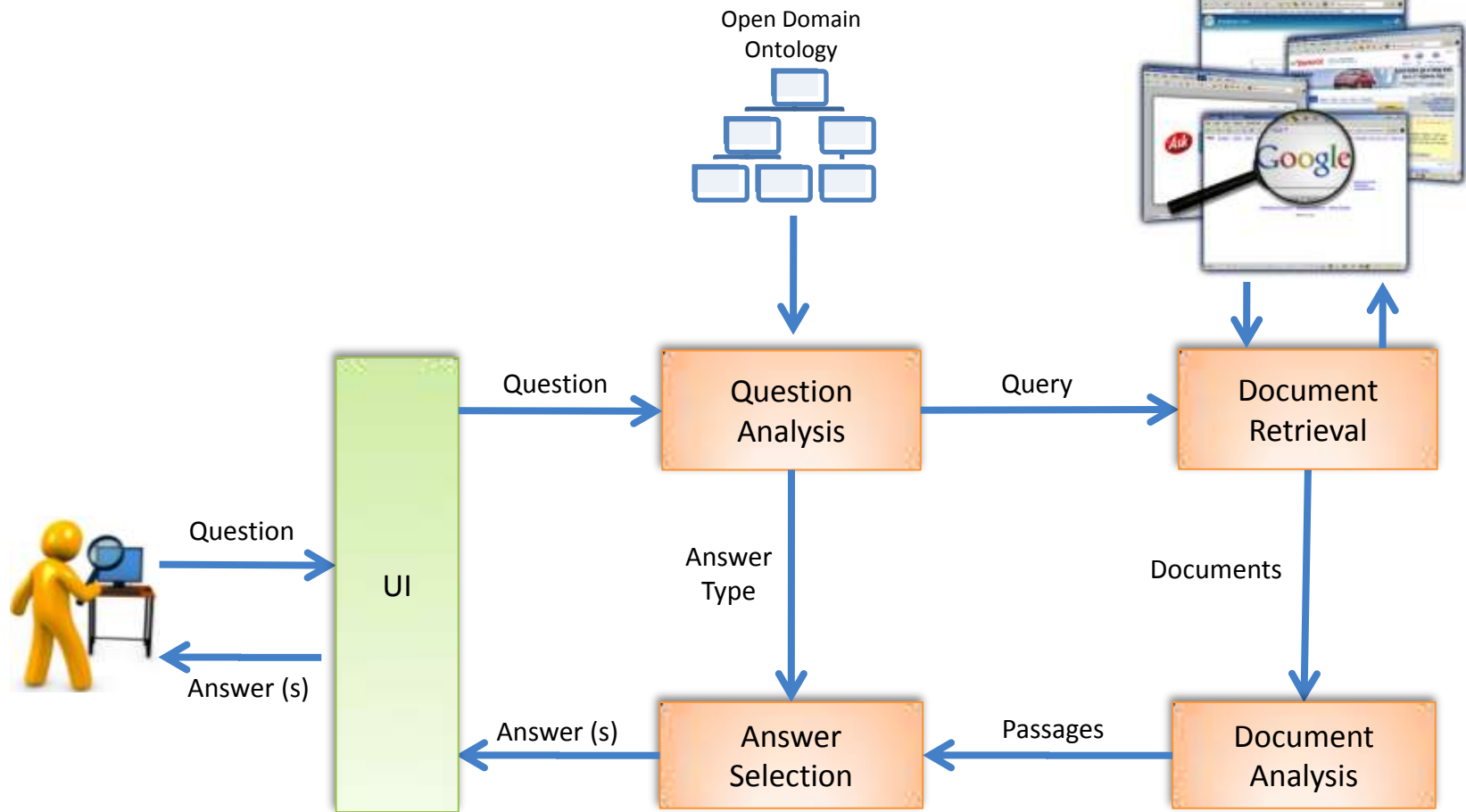
- Question answering (QA) systems **take users' natural language questions** and automatically locate answers from large collections of documents.
- Two types of QA systems
  - Closed-Domain (or restricted domain) Question Answering
  - Open-Domain Question Answering



# Question Answering (2)



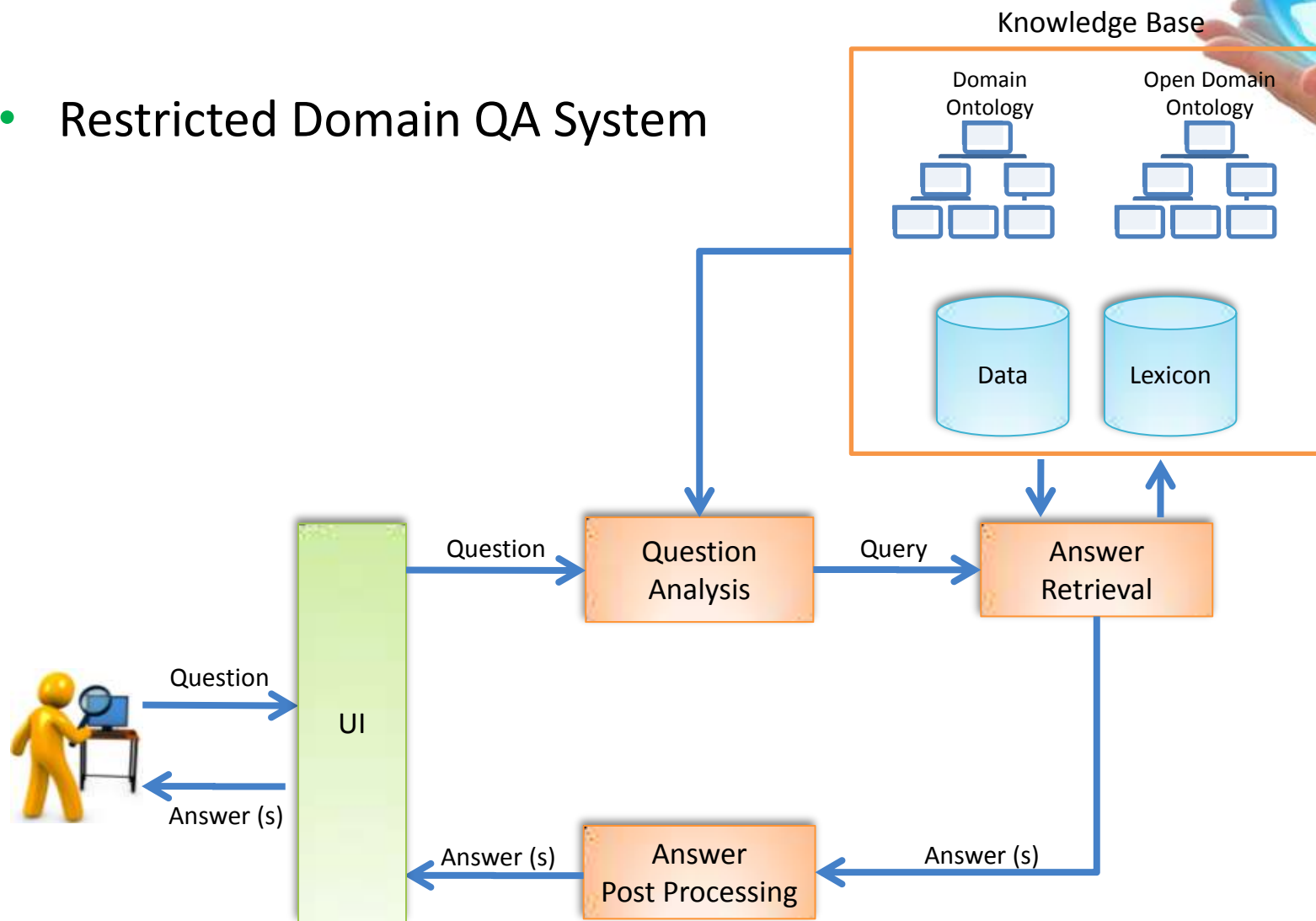
- Open Domain QA System



# Question Answering (2)



- Restricted Domain QA System





?