Question Answering

Lê Thanh Hương
School of Information and Communication Technology

Question Answering

- An idea originating from the IR community
- IR: find *relevant documents*, but we want *answers* from textbases
- QA: give short answer, perhaps supported by evidence

Sample TREC questions

- Who is the author of the book “The Iron Lady: A Biography of Margaret Thatcher”?
- What was the monetary value of the Nobel Peace Prize in 1989?
- What does the Peugeot company manufacture?
- How much did Mercury spend on advertising in 1993?
- Why did David Koresh ask the FBI for a word processor?

People want to ask questions

- Examples from AltaVista query log (late 1990s)
  - Who invented surf music?
  - How to make stink bombs
  - Which english translation of the bible is used in official catholic liturgies?
- Examples from Excite query log (12/1999)
  - How can i find someone in Texas
  - Where can i find information on puritan religion?
  - What vacuum cleaner does Consumers Guide recommend

A Brief (Academic) History

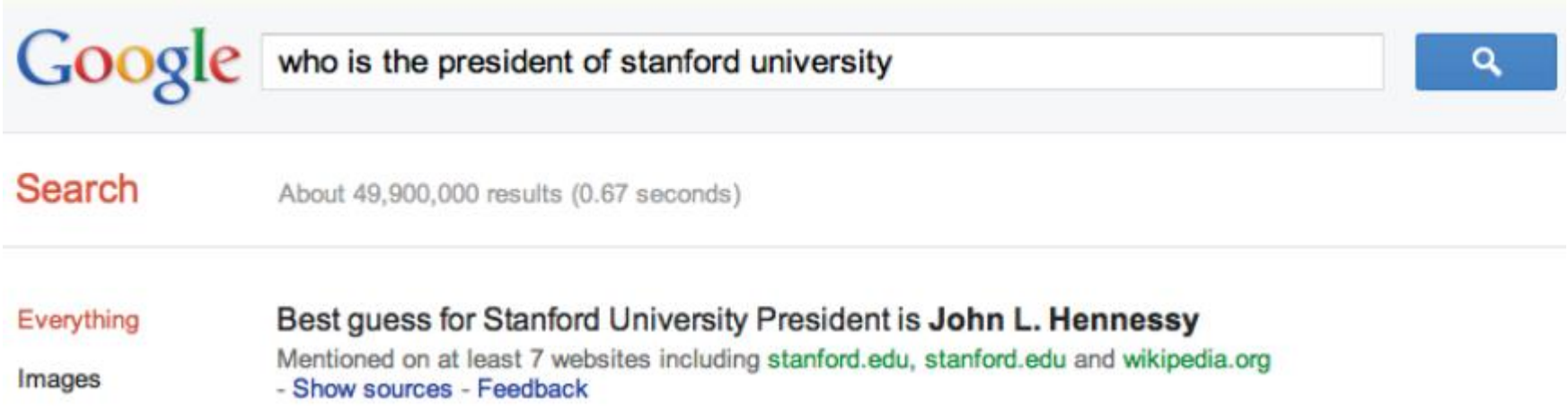
- Question answering is not a new research area
- Question answering systems can be found in many areas of NLP research, including:
  - Natural language database systems
    - A lot of early NLP work on these: e.g., LUNAR system
    - There's still Microsoft English Query
  - Spoken dialog systems
    - Currently very active and commercially relevant

A Brief (Academic) History

- Focusing on open-domain QA is new focus
  - MURAX (Kupiec 1993): Encyclopedia answers
  - Hirschman: Reading comprehension tests
  - TREC QA competition: 1999–
- But not really new either: Simmons et al. 1965
  - Take an encyclopedia and load it onto a computer.
  - Take a question and parse it into a logical form
  - Perform simple information retrieval to get relevant texts, parse those into a logical form, match and rank
    - What do worms eat? Worms eat ???
      - Candidates
        - Worms eat grass
        - Grass is eaten by worms
        - Birds eat worms

Online QA Examples

- LCC: http://www.languagecomputer.com/demos/question_answering/index.html
- AnswerBus is an open-domain question answering system: www.answerbus.com
- EasyAsk, AnswerLogic, AnswerFriend, Start, Quasm, Mulder, Webclopedia, TextMap, etc.
- Google



AskJeeves

- …is most hyped example of QA
- …does pattern matching to match your question to their own knowledge base of questions
  - If that works, you get the human-curated answers to that known question
  - If that fails, return regular web search
- A potentially interested middle ground, but a weak shadow of real QA

TextMap - The Entity Search Engine - Newspaper Analy...

File Edit View Favorites Tools Help

Back Search Favorites

Address http://www.textmap.com/cgi-bin/qa_main.cgi?question=In+wh Go Links

TEXTMAP
THE ENTITY SEARCH ENGINE

Monitoring the World So You Don't Have To ...

ENTITIES SOU

Search!

TextMap : TextMed : Textblg : TextBiz : Make homepage! : Link to us : Help?

Question Answering

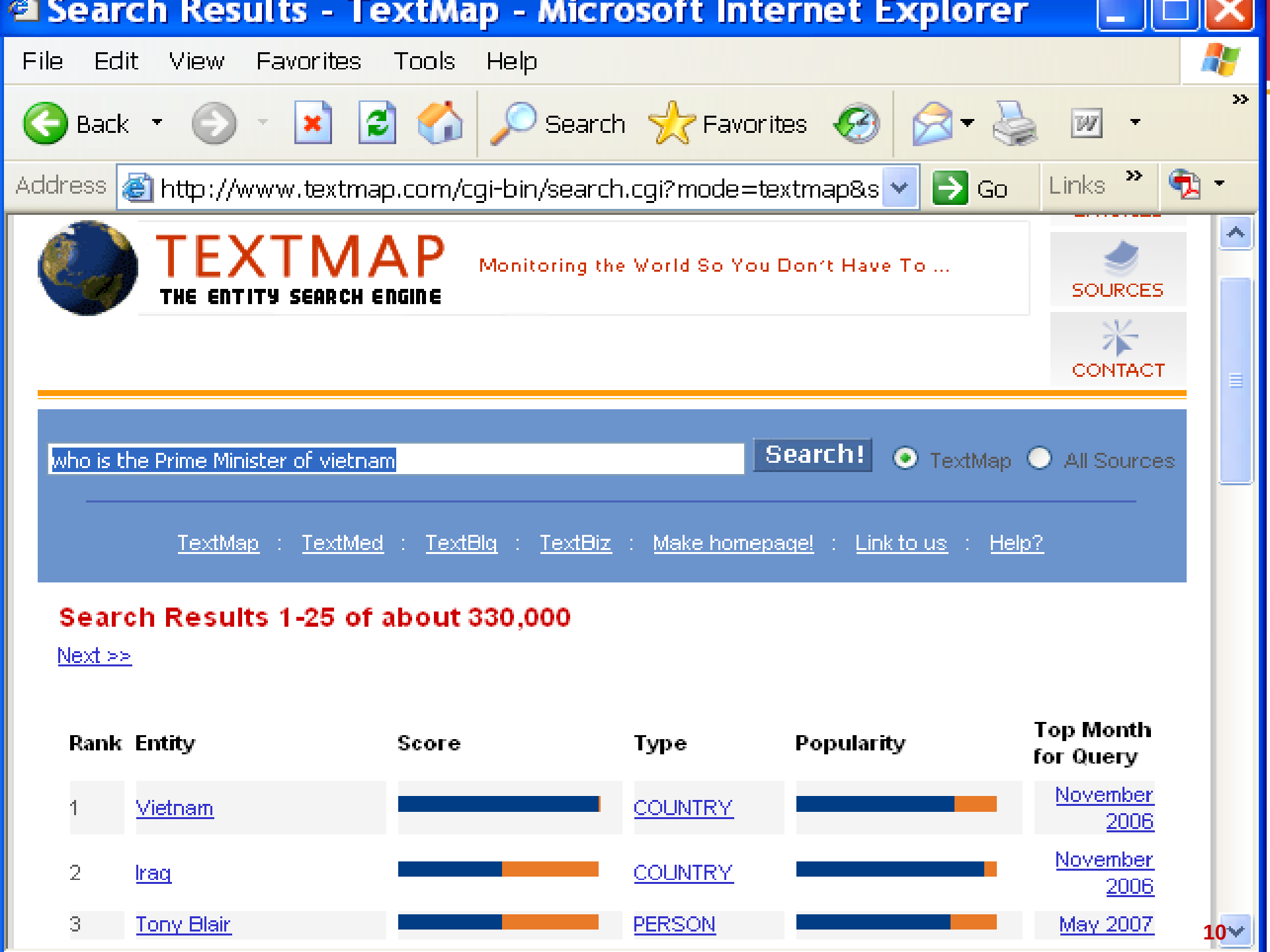
Wednesda

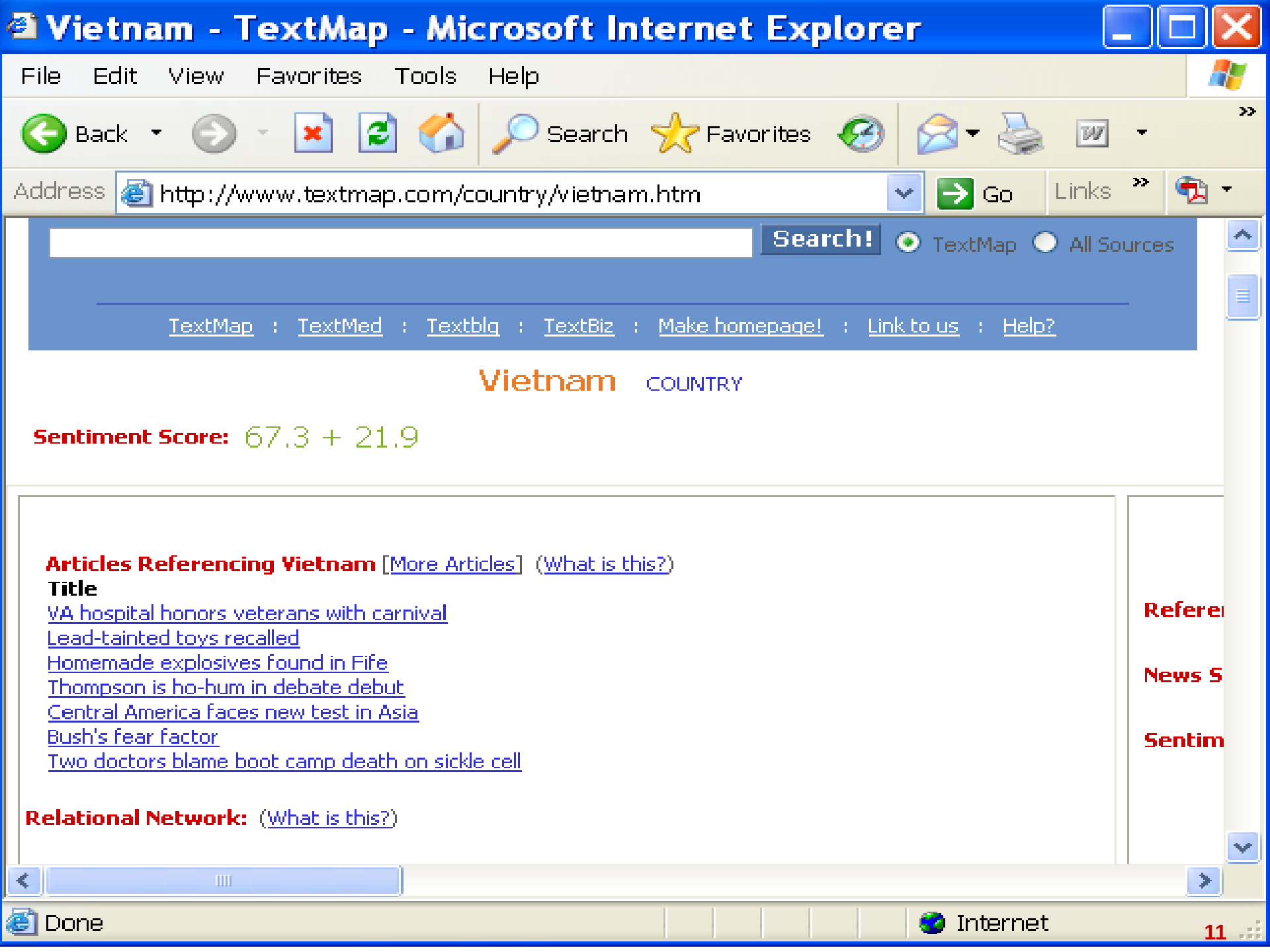
in what year did John Lennon die?

Answer: 1980

[The Beatles Anthology 02/28/2006 wiki]

Internet


Search Results - TextMap - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.textmap.com/cgi-bin/search.cgi?mode=textmap&s
Go Links
TEXTMAP
THE ENTITY SEARCH ENGINE
Monitoring the World So You Don't Have To ...
SOURCES
CONTACT
who is the Prime Minister of vietnam
Search!
TextMap
All Sources
TextMap : TextMed : TextBlg : TextBiz : Make homepage! : Link to us : Help?
Search Results 1-25 of about 330,000
Next >>
Rank Entity Score Type Popularity Top Month for Query
1 Vietnam COUNTRY November 2006
2 Iraq COUNTRY November 2006
3 Tony Blair PERSON May 2007


Vietnam - TextMap - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.textmap.com/country/vietnam.htm
Go Links
Search! TextMap All Sources
TextMap : TextMed : Textblg : TextBiz : Make homepage! : Link to us : Help?
Vietnam COUNTRY
Sentiment Score: 67.3 + 21.9
Articles Referencing Vietnam [More Articles] (What is this?)
Title
VA hospital honors veterans with carnival
Lead-tainted toys recalled
Homemade explosives found in Fife
Thompson is ho-hum in debate debut
Central America faces new test in Asia
Bush's fear factor
Two doctors blame boot camp death on sickle cell
Relational Network: (What is this?)
Done
Internet
11

The TREC Document Collection

- …uses news articles from the following sources:
  - AP newswire, 1998-2000
  - New York Times, 1998-2000
  - Xinhua News Agency newswire, 1996-2000
- 1,033,461 documents – 3GB of text
- This is a lot of text to process entirely using advanced NLP techniques → information retrieval is done first

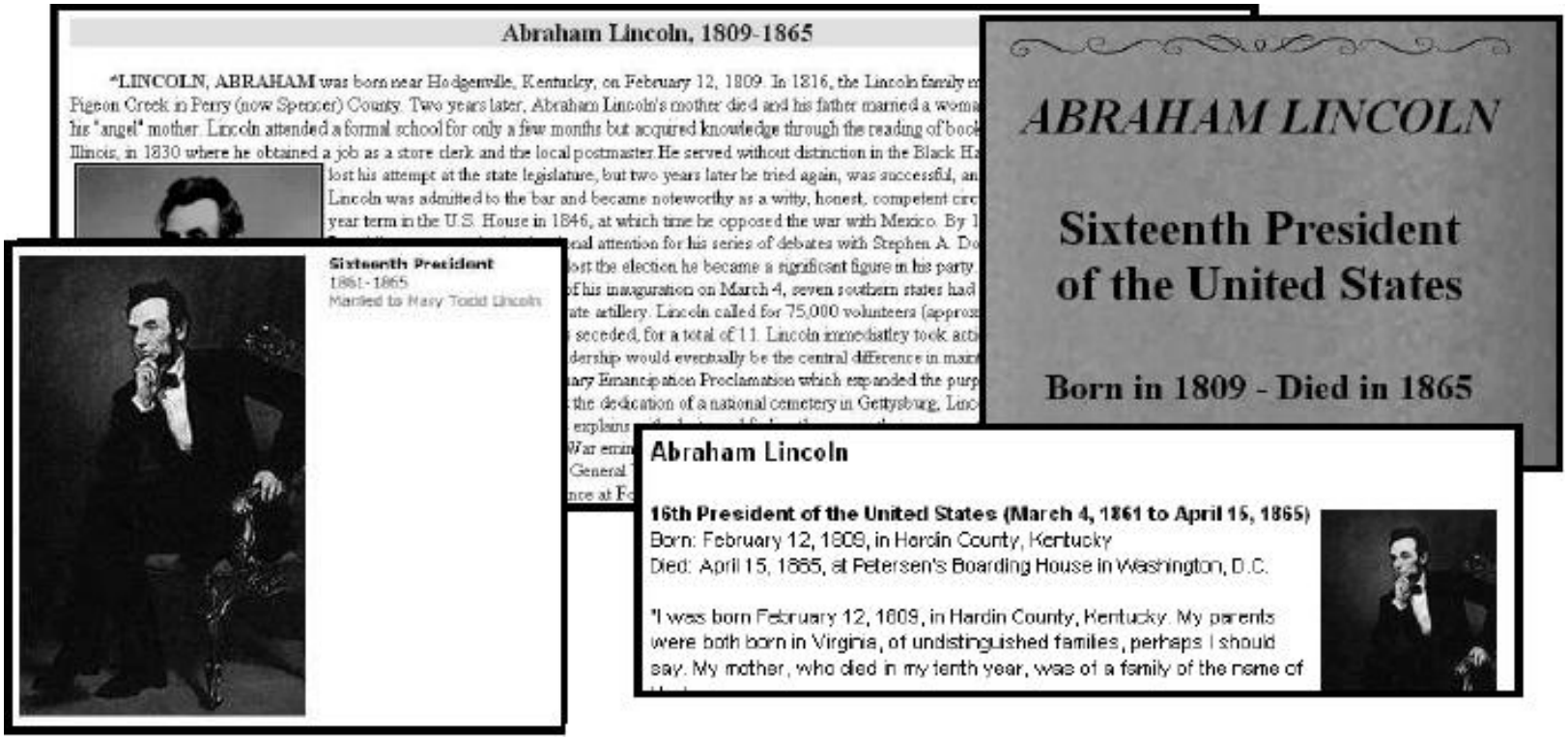
Top Performing Systems

- …can answer ~70% of the questions

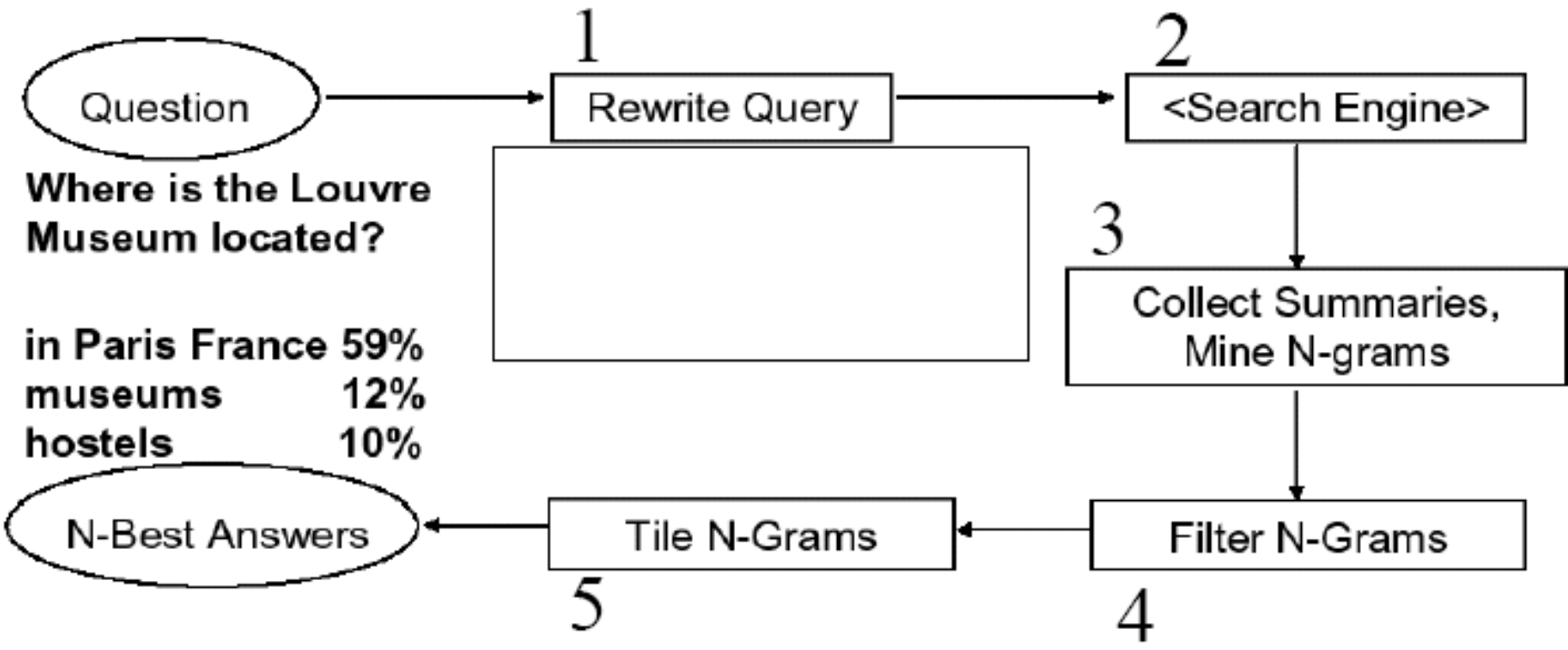
- Approaches:
  - Knowledge-rich approaches, using many NLP techniques (Harabagiu, Moldovan et al.-SMU/UTD/LCC)
  - AskMRS: shallow approach
  - Middle ground use large collection of surface matching patterns (ISI)

AskMRS: shallow approach

- In what year did Abraham Lincoln die?
- Ignore hard documents and find easy ones



AskMSR: Details



Step 1: Rewrite queries

- Intuition: The user's question is often syntactically quite close to sentences that contain the answer
  - Where is the Louvre Museum located?
  - The Louvre Museum is located in *Paris*
  - Who created the character of Scroogle?
  - *Charles Dickens* created the character of Scrooge.

Query rewriting

- Classify question into 7 categories
  - Who is/was/are/were…?
  - When is/did/will/are/were…?
  - Where is/are/were…?

a) Category-specific transformation rules
  - E.g., For Where question, move “is” to all possible locations
  - Where is the Louvre Museum located?

→ is the Louvre Museum located?
→ the is Louvre Museum located?
→ the Louvre is Museum located?
→ the Louvre Museum is located?
→ the Louvre Museum located is?

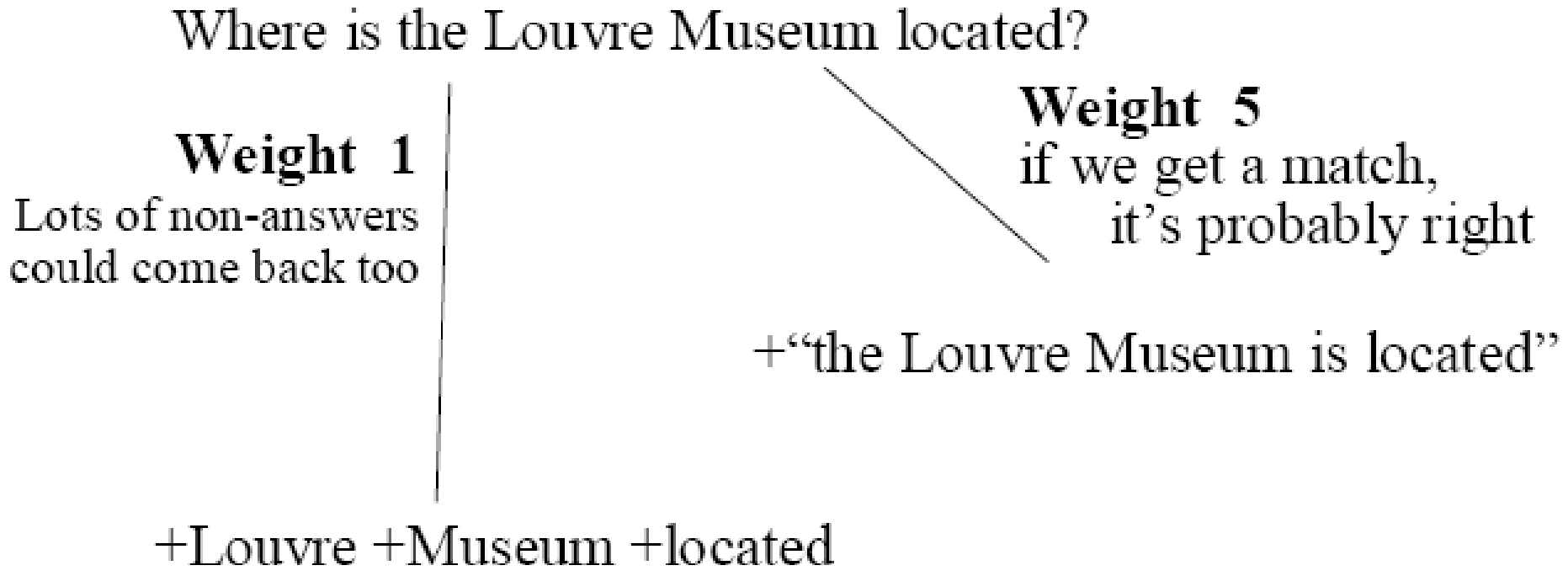
b) Expected answer “Datatype” (eg, Date, Person, Location,…)

→ When was the French Revolution? → DATE

- Hand-crafted classification/rewrite/datatype rules (Could they be automatically learned?)

Query Rewriting - weights

- Some query rewrites are more reliable than others



Step 2: Query search engine

- Send all rewrites to a Web search engine
- Retrieve top N answers (100?)
- Rely just on search engine's words/phrases, not the full text of the actual document

Step 3: Mining N-Grams

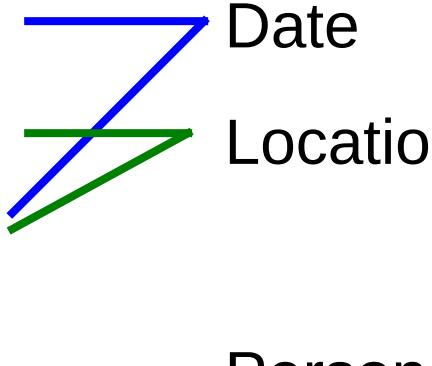
- Unigram, bigram, trigram, …, N-gram: list of N adjacent term in a sequence
  - Eg. “Web Question Answering: Is More Always Better”
  - Unigram: Web, Question, Answering, Is, More, Always, Better
  - Bigram: Web Question, Question Answering, Answering Is, Is More, More Always, Always Better
  - Trigram: …

Mining N-grams

- Simple: Enumerate all N-grams (N=1,2,3…) in all retrieved phrases
  - Use hash table and other tools to make this efficient
- Weight of an n-gram: occurrence count
  - Eg, “Who created the character of Scrooge?”
  - Dickens – 117
  - Christmas Carol – 78
  - Charles Dickens – 75
  - Disney – 72
  - Carl Banks – 54
  - A Christmas – 41
  - Christmas Carol - 45

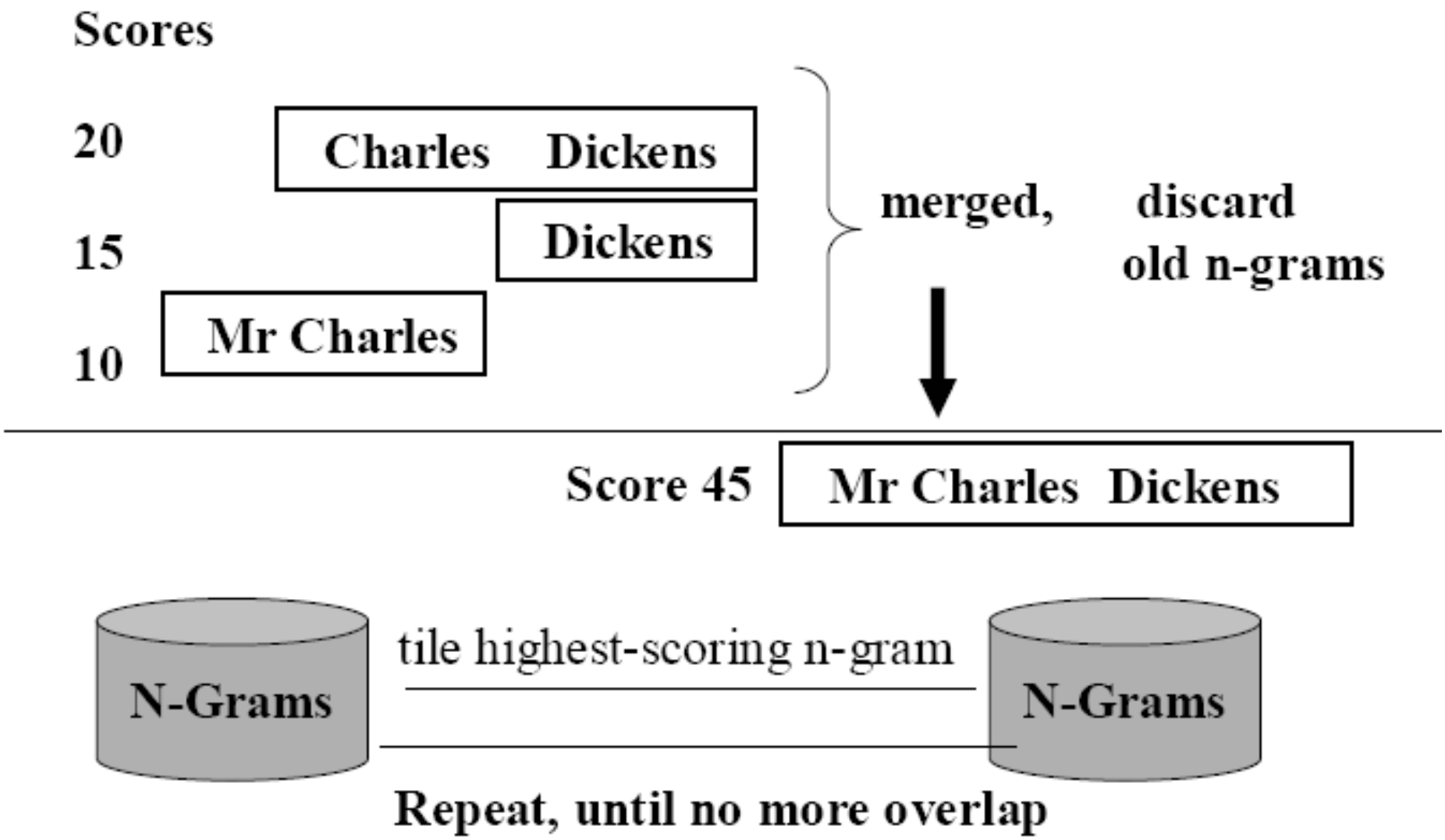
Step 4: Filtering N-Grams

- Each question type is associated with one or more "data-type filters" = regular expression
- When…
- Where…
- What…
- Who…



- Boost score of n-grams that do match regexp
- Lower score of n-grams that don't match regexp

Step 5: Tiling the Answers



Results

- Standard TREC contest test-bed:
  - ~1M documents; 900 questions
  - Technique doesn’t do well (but rank in top 9/30 participants!)
- Limitation:
  - Works best only for fact-based questions
  - Limited range of
    - Question categories
    - Answer data types
    - Query rewriting rules

Surface matching patterns (Ravichandran and Hovy, ISI)

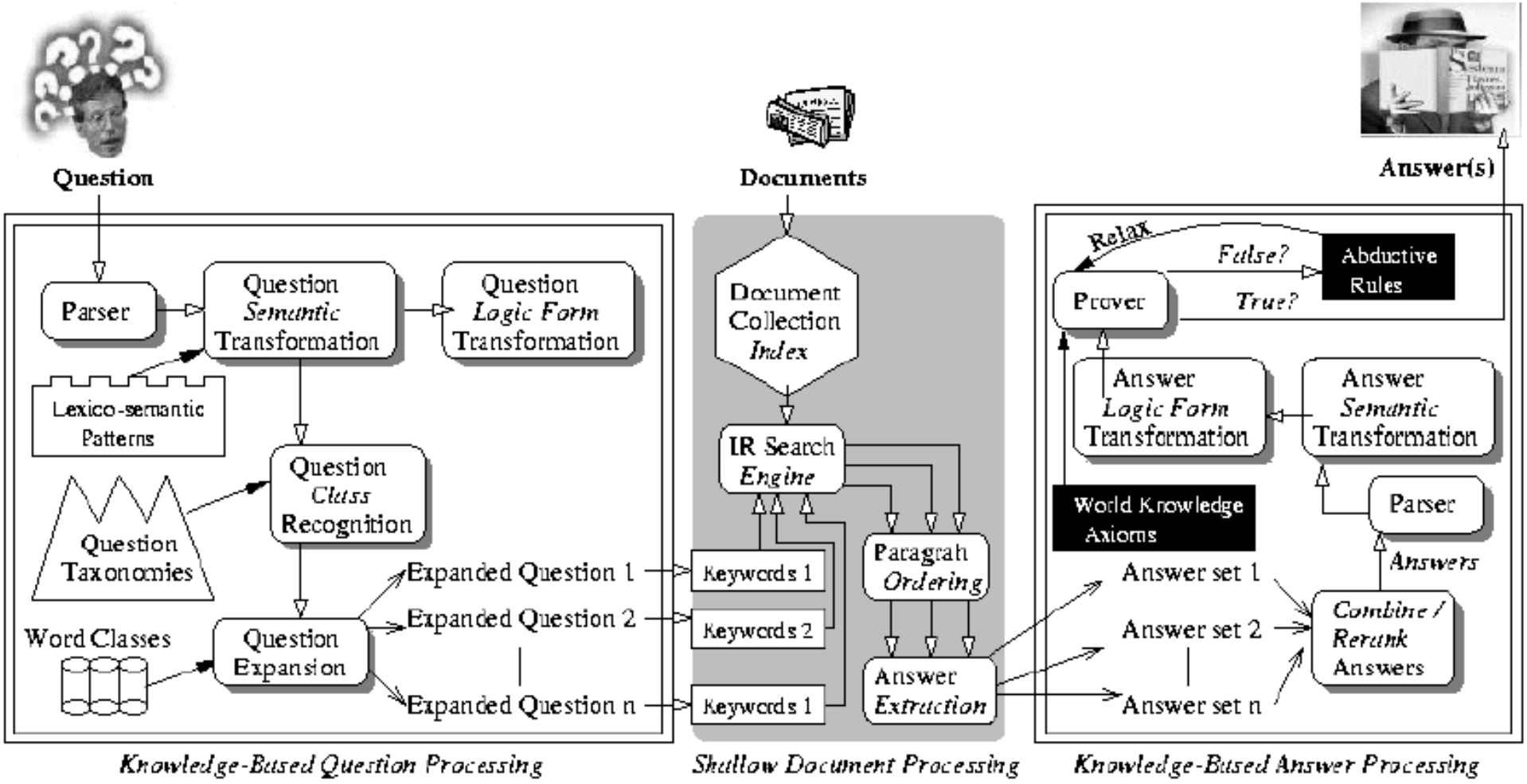
- When was X born?
  - Mozart was born in 1756
  - Gandhi (1869—1948)
- <NAME> was born in <BIRTHDATE>
- <NAME> (<BIRTHDATE>-
- Use a Q-A pair to query a search engine
- Extract patterns and compute their accuracy

Example: INVENTOR

- <ANSWER> invents <NAME>
- the <NAME> was invented by <ANSWER>
- <ANSWER> invented the <NAME> in
- <ANSWER>’s invention of the <NAME>
- …

- Many of these patterns have high accuracy
  - But still some mistakes

Full NLP QA (LCC: Harabagiu,Moldovan et al.)

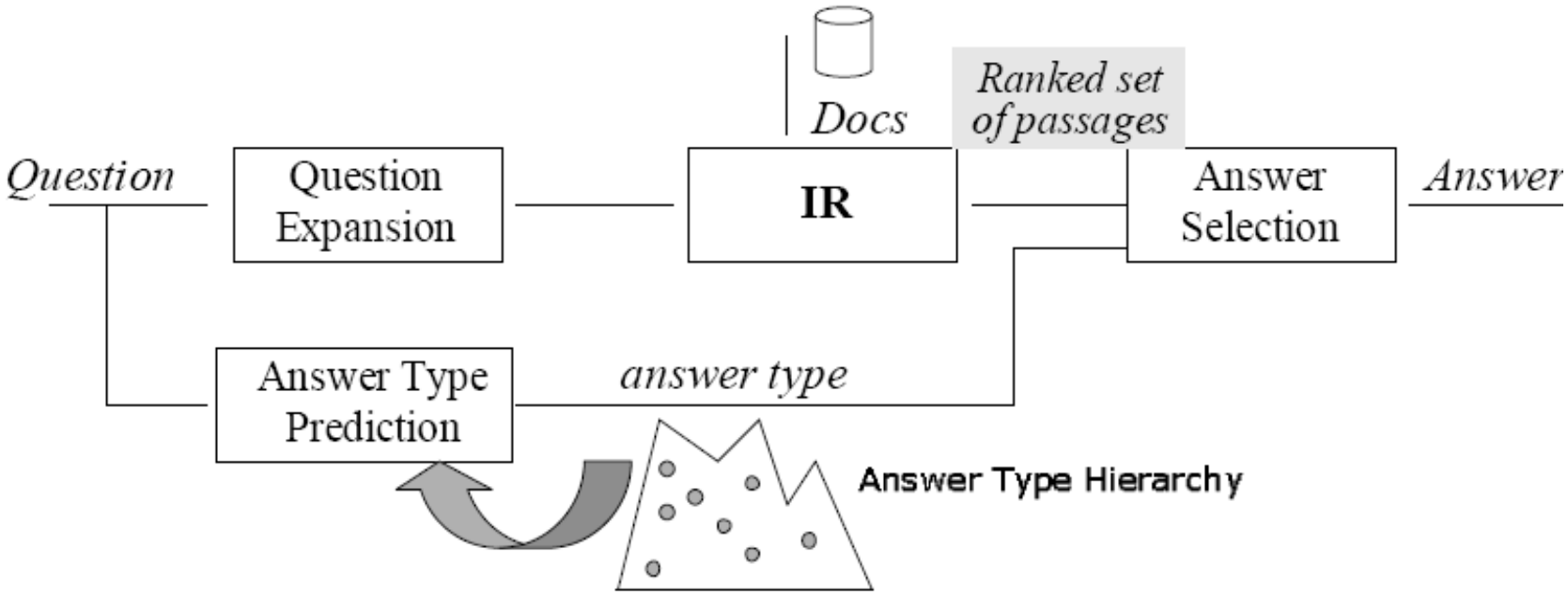


Value from sophisticated NLP (Pasca & Harabagiu, 2001)

- Good IR is needed: SMART paragraph retrieval
- Large taxonomy of question types and expected answer types is crucial
- Statistical parser is used to parse questions and relevant text for answers, and to build KB
- Query expansion loops (morphological, lexical synonyms, and semantic relations) important
- Answer ranking by simply ML method

Answer types in State-of-the-art QA systems

- Answer type:
  - Labels questions with answer type based on a taxonomy
  - Classifies questions (eg., by using a maximum entropy model)



Answer Types

- “Who” questions can have organizations as answers
  - Who sells the most hybrid cars?
- “Which” questions can have people as answers
  - Which president went to war with Mexico?

Keyword Selection Algorithm

Select all…

- Non-stopwords in quotations
- NNP words in recognized named entities
- Complex nominals with their adjectival modifiers
- Other complex nominal
- Nouns with adjectival modifiers
- Other nouns
- Verbs
- The answer type word

Passage Extraction Loop

- Passage Extraction Component
  - Extracts passages that contain all selected keywords
  - Passage size/start position dynamic
- Passage quality and keyword adjustment
  - 1st iteration: use the first 6 keyword selection heuristics
  - If #passages < θ → query is too strict →drop a keyword
  - If #passages > θ → query is too relaxed →add a keyword

Passage Scoring

Involve 3 scores:

- #words from the question that are recognized in the same sequence in the window
- #words that separate the most distant keywords in the window
- #unmatched keywords in the window

Rank candidate answers in the retrieved passages

- Name the first private citizen to fly in space

- Answer type: Person
- Text passage:

  "Among them was Christa McAuliffe, the first private citizen to fly in space. Karen Ailen, best known for her starring role in "Raiders of the Lost Ark", plays McAuliffe. Brian Kerwin is featured as shuttle pilot Mike Smith…"

- Best candidate answer: Christa McAuliffe

Name Entity Recognition

- Current QA is determined by the recognition of name entities

QUANTITY	55	ORGANIZATION	15	PRICE	3
NUMBER	45	AUTHORED WORK	11	SCIENCE NAME	2
DATE	35	PRODUCT	11	ACRONYM	1
PERSON	31	CONTINENT	5	ADDRESS	1
COUNTRY	21	PROVINCE	5	ALPHABET	1
OTHER LOCATIONS	19	QUOTE	5	URI	1
CITY	19	UNIVERSITY	3		

- Precision of recognition
- Coverage of name classes
- Mapping into concept hierarchies
- Participation into semantic relations (eg, predicate-argument structures or frame semantics)

Semantics and Reasoning for QA: Predicate-argument structure

- *When was Microsoft established?*

  Microsoft plans to establish manufacturing partnerships in Brazil and Mexico in May.

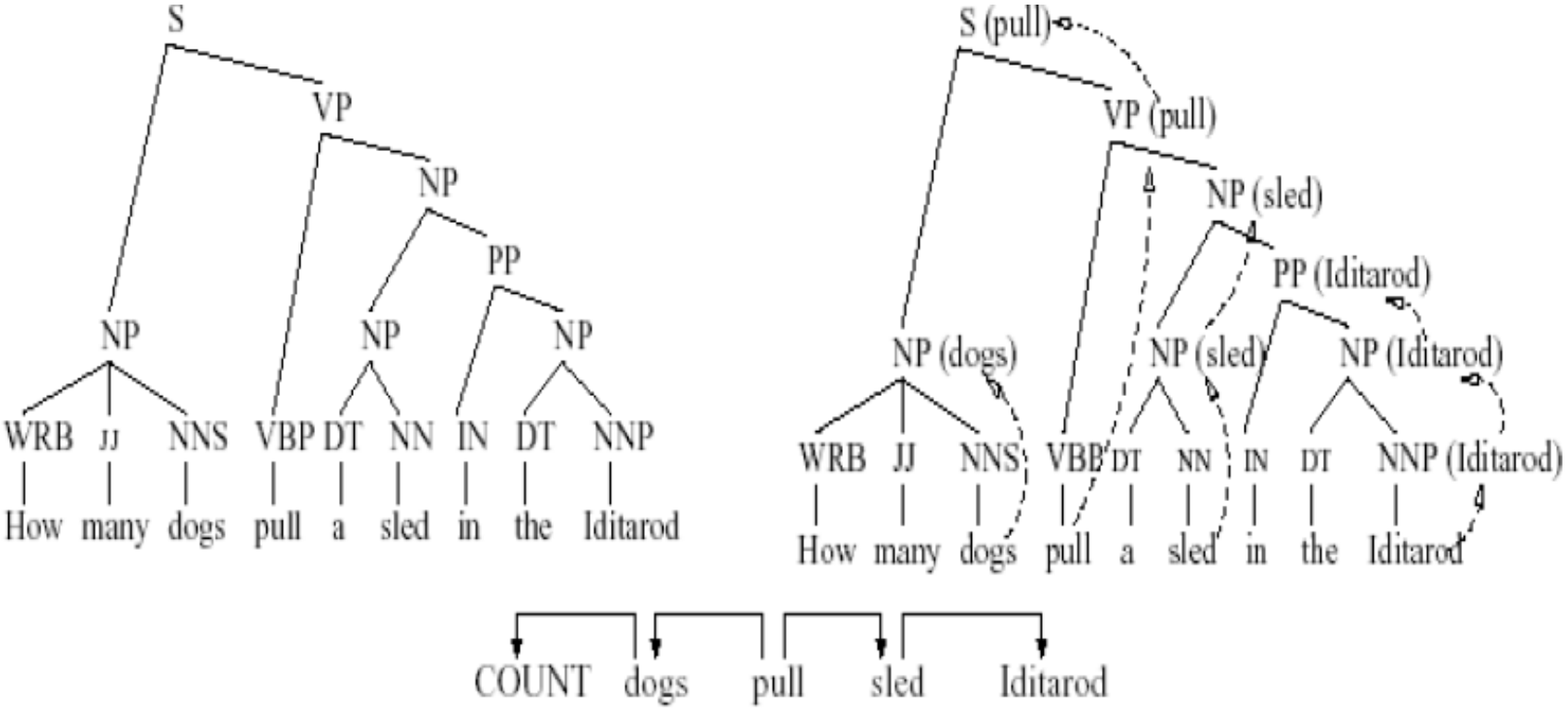
- Need to be able to detect sentences in which ‘Microsoft’ is object of ‘establish” or close synonym.
- Matching sentence:

  Microsoft Corp was founded in the US in 1975, incorporated in 1981, and established in the UK in 1982.

- Require analysis of sentence syntax/ semantics

Semantics and Reasoning for QA: Syntax to Logical Forms

- Syntactic analysis plus semantic → logical form
- Mapping of question and potential answer LFs to find the best match



Inference

- System attempts inference to justify an answer (often following lexical chains)
- Their inference is a middle ground between logic and pattern matching
- But very effective: 30% improvement

- Q: When was the internal combustion engine invented?
- A: The first internal-combustion engine was built in 1867.
- Invent → create_mentally → create → build

QA Example

- How hot does the inside of an active volcano get?
- Get(TEMPERATURE, inside(volcano(active))))
- “lava fragments belched out of the mountain were as hot as 300 degrees Fahrenheit”
- Fragments(lava, TEMPERATURE(degrees(300)), belched(out, mountain))
  - Volcano ISA mountain
  - Lava ISPARTOF volcano
  - Lava inside volcano
  - Fragments of lava HAVEPROPERTIESOF lava

→ use Wordnet

Answer Validation motivates the Robust Textual Inference Task

- The task: Can systems correctly perform ‘local textual inferences’ [individual inference steps]?
  - On the assumption that some piece of text (T) is true, does this imply the truth of some other hypothesis text (H)?
    - *Sydney was the host city of the 2000 Olympics →*
    - *The Olympics have been held in Sydney TRUE*
- The format could be used for evaluating extended inferential chains or knowledge
  - But, in practice, fairly direct stuff

The textual inference task

- Does text *T justify an inference to hypothesis H?*
  - Emphasis on variability of linguistic expression
- Robust, accurate textual inference would enable:
  - Semantic search: H: *lobbyists attempting to bribe U.S. legislators*
  - T: *The A.P. named two more senators who received contributions engineered by lobbyist Jack Abramoff in return for political favors.*
  - Question answering: H: *Who bought J.D. Edwards?*
  - T: *Thanks to its recent acquisition of J.D. Edwards, Oracle will soon be able…*
  - Customer email response
  - Relation extraction (database building)
  - Document summarization

Natural Examples: Reading Comprehension

- (CNN Student News) -- *January 24, 2006*
- Answer the following questions about today's featured news stories. Write your answers in the space provided.
- 1. Where is the country of Somalia located? What ocean borders this country?
- 2. Why did crew members from the USS Winston S.Churchill recently stop a small vessel off the coast of Somalia? What action did the crew of the Churchill take?

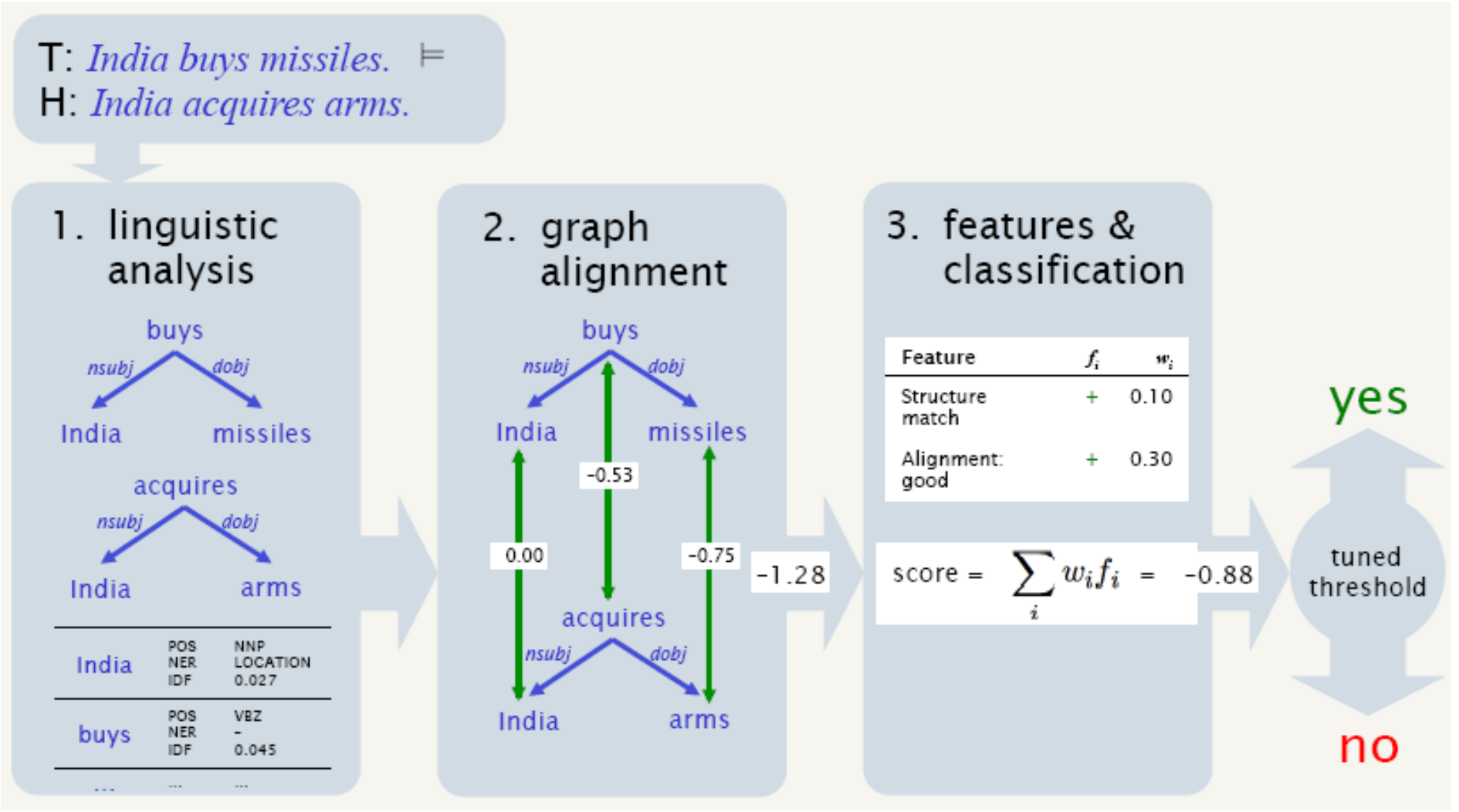
Verification of terms [Dan Roth]

- Non-disclosure Agreement

- WHEREAS Recipient is desirous of obtaining said confidential information for purposes of evaluation thereof and as a basis for further discussions with Owner regarding assistance with development of the confidential information for the benefit of Owner or for the mutual benefit of Owner and Recipient; THEREFORE, Recipient hereby agrees to receive the information in confidence and to treat it as confidential for all purposes. Recipient will not divulge or use in any manner any of said confidential information unless by written consent from Owner, and Recipient will use at least the same efforts it regularly employs for its own confidential information to avoid disclosure to others. Provided, however, that this obligation to treat information confidentially will not apply to any information already in Recipient‘s possession or to any information that is generally available to the public or becomes generally available through no act or influence of Recipient. Recipient will inform Owner of the public nature or Recipient's possession of the information without delay after Owner's disclosure thereof or will be stopped from asserting such as defense to remedy under this agreement. Each party acknowledges that all of the disclosing party's Confidential Information is owned solely by the disclosing party (or its licensors and/or other vendors) and that the unauthorized disclosure or use of such Confidential Information would cause irreparable harm and significant injury, the degree of which may be difficult to ascertain. Accordingly, each party agrees that the disclosing party will have the right to obtain an immediate injunction enjoining any breach of this Agreement, as well as the right to pursue any and all other rights and remedies available at law or in equity for such a breach. Recipient will exercise its best efforts to conduct its evaluation within a reasonable time after Owner's disclosure and will provide Owner with its assessment thereof without delay. Recipient will return all information, including all copies thereof, to Owner upon request. This agreement shall remain in effect for ten years after the date of it's execution, and it shall be construed under the laws of the State of Texas.

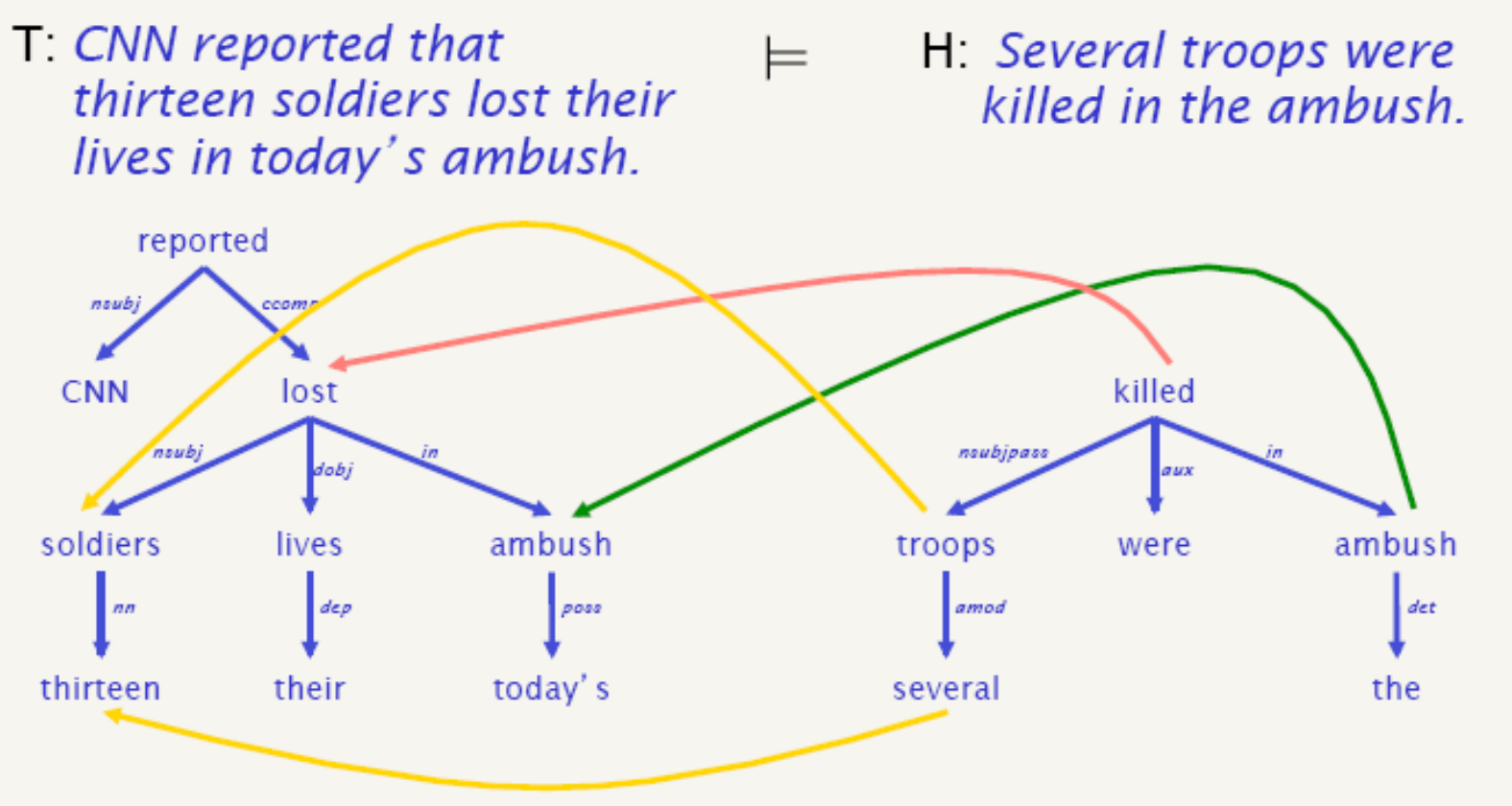
- Conditions I care about:
  - All information discussed is freely shareable unless other party indicates in advance that it is confidential
  - TRUE? FALSE?

Stanford system three-stage architecture [MacCartney et al. 2006]

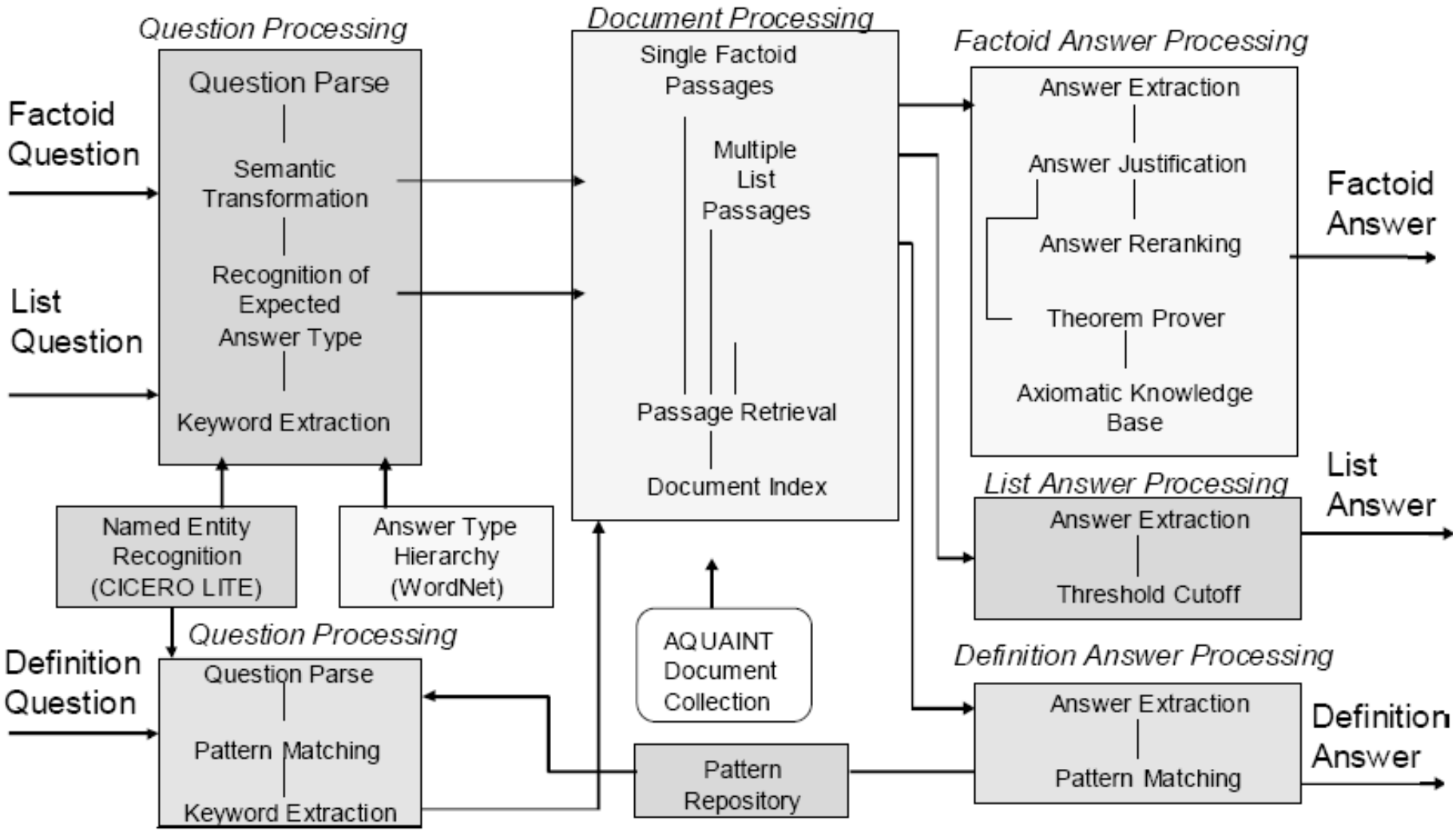


Textual inference as graph alignment
[Haghighi et al. 05, de Salvo Braz et al. 05]

- Find least cost alignment of H to part of T, using locally decomposable cost model (lexical and structural costs)
- Assumption: good alignment ⇒ valid inference



Architecture of LCC's QA system



Question types

- Factoid queries: WH questions like when, who, where.
- Yes/ No queries: Is Berlin capital of Germany?
- Definition queries: what is leukemia?
- Cause/consequence queries: How, Why, What. what are the consequences of the Iraq war ?
- Procedural queries: which are the steps for getting a Master degree?
- Comparative queries: what are the differences between the model A and B?
- Queries with examples: list of hard disks similar to hard disk X.
- Queries about opinion: What is the opinion of the majority of Americans about the Iraq war?

Data types

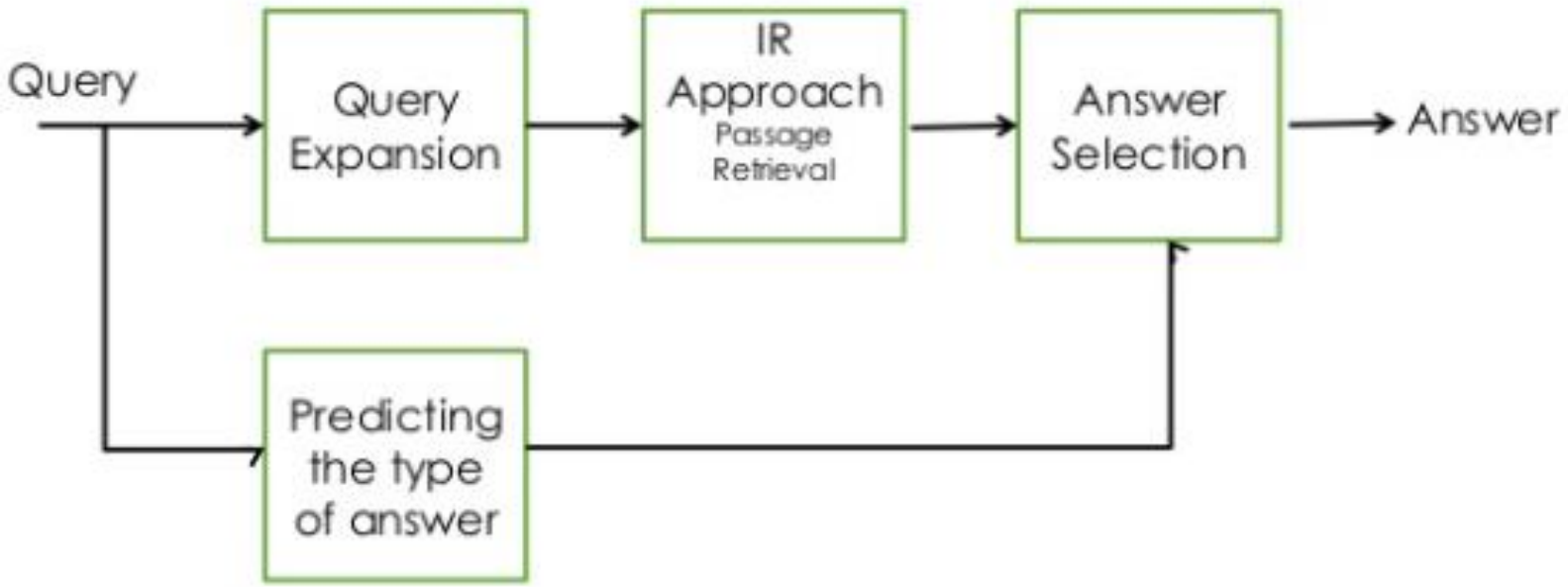
- Structured data (relational data bases, RDF knowledge bases).
- Semi-structured data (XML databases)
- Free text
- Multimodal data: image, voice, video

Approaches

- Open-domain: domain independent QA systems can answer any query from any corpus
  - + covers wide range of queries
  - - low accuracy
- Closed-domain: domain specific QA systems are limited to specific domains
  - + High accuracy
  - - limited coverage over the possible queries - Needs domain expert

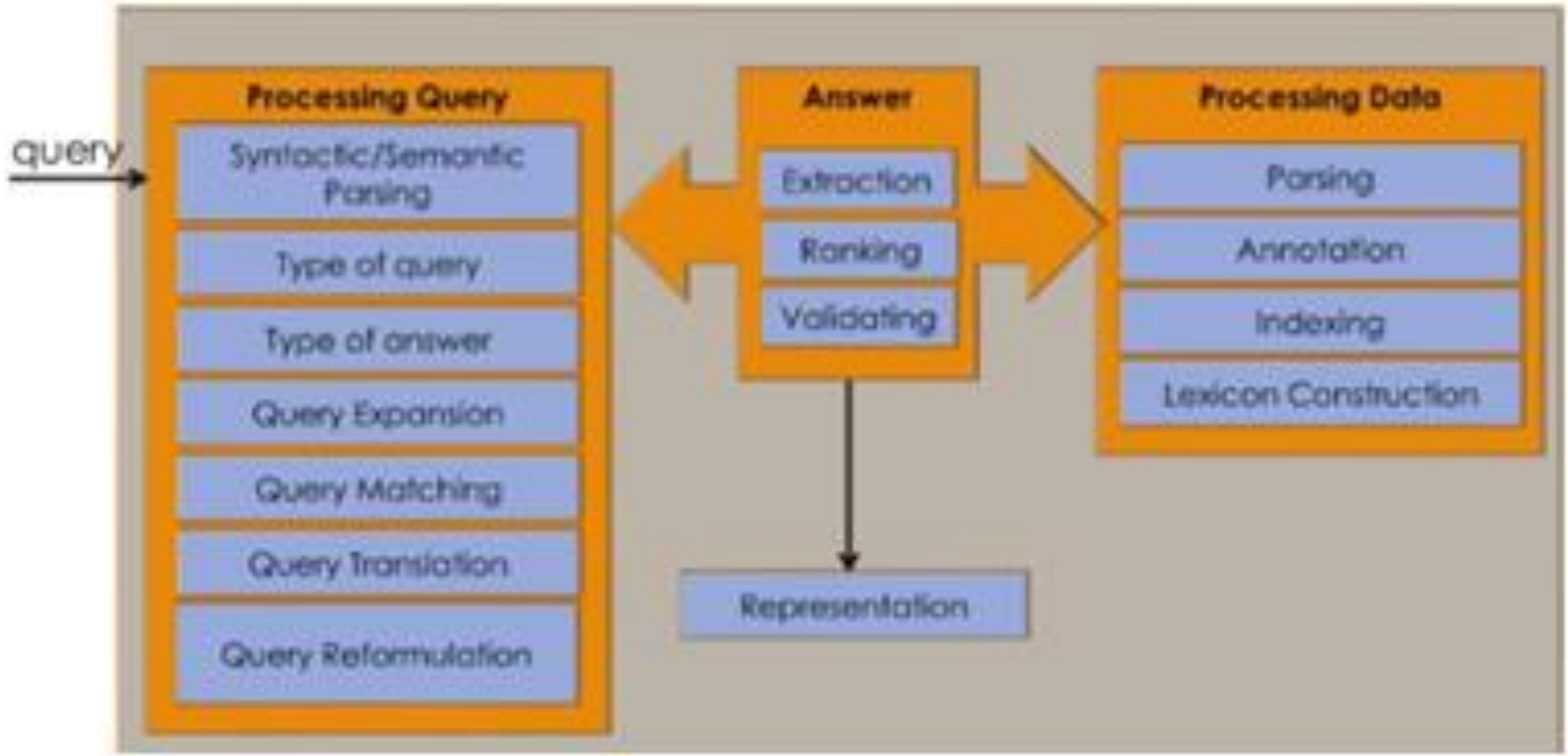
Evaluation

- Text Retrieval Conference (Trec), 1999-2002, with QA track
- Typical framework of QA systems

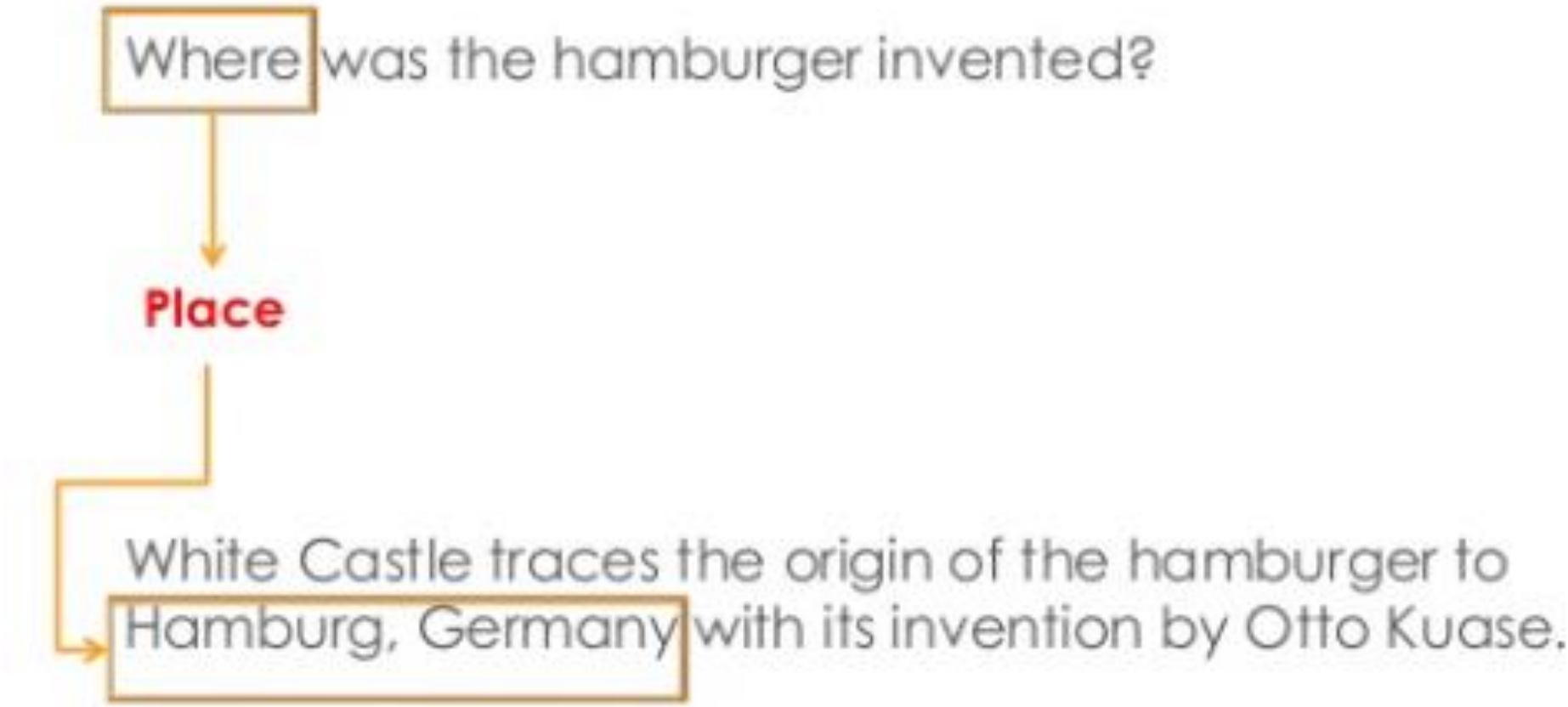


Evaluation

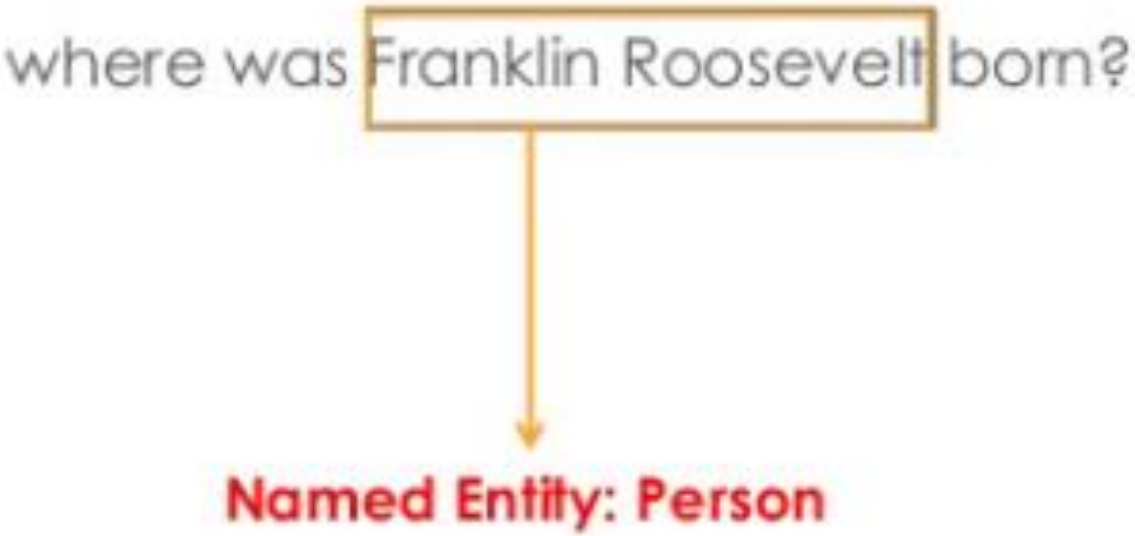
- Cross-Language Evaluation Forum (CLEF), from 2003 for European languages, for single and cross-language contexts
- Basic architecture



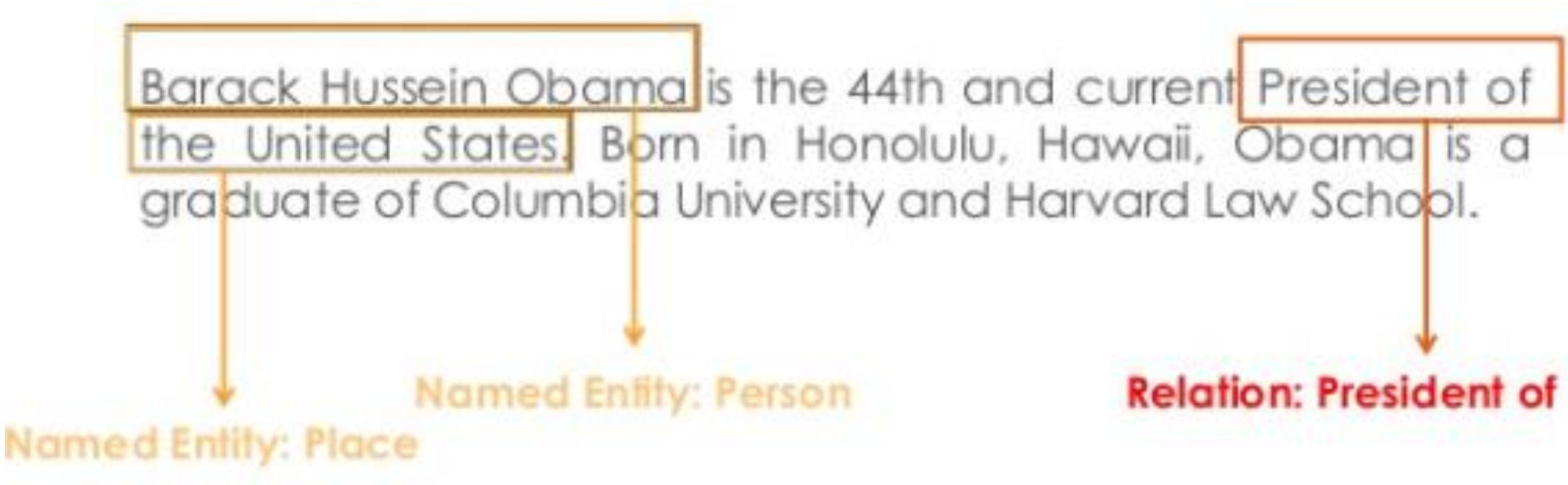
Question types



NER in QA



Relation extraction



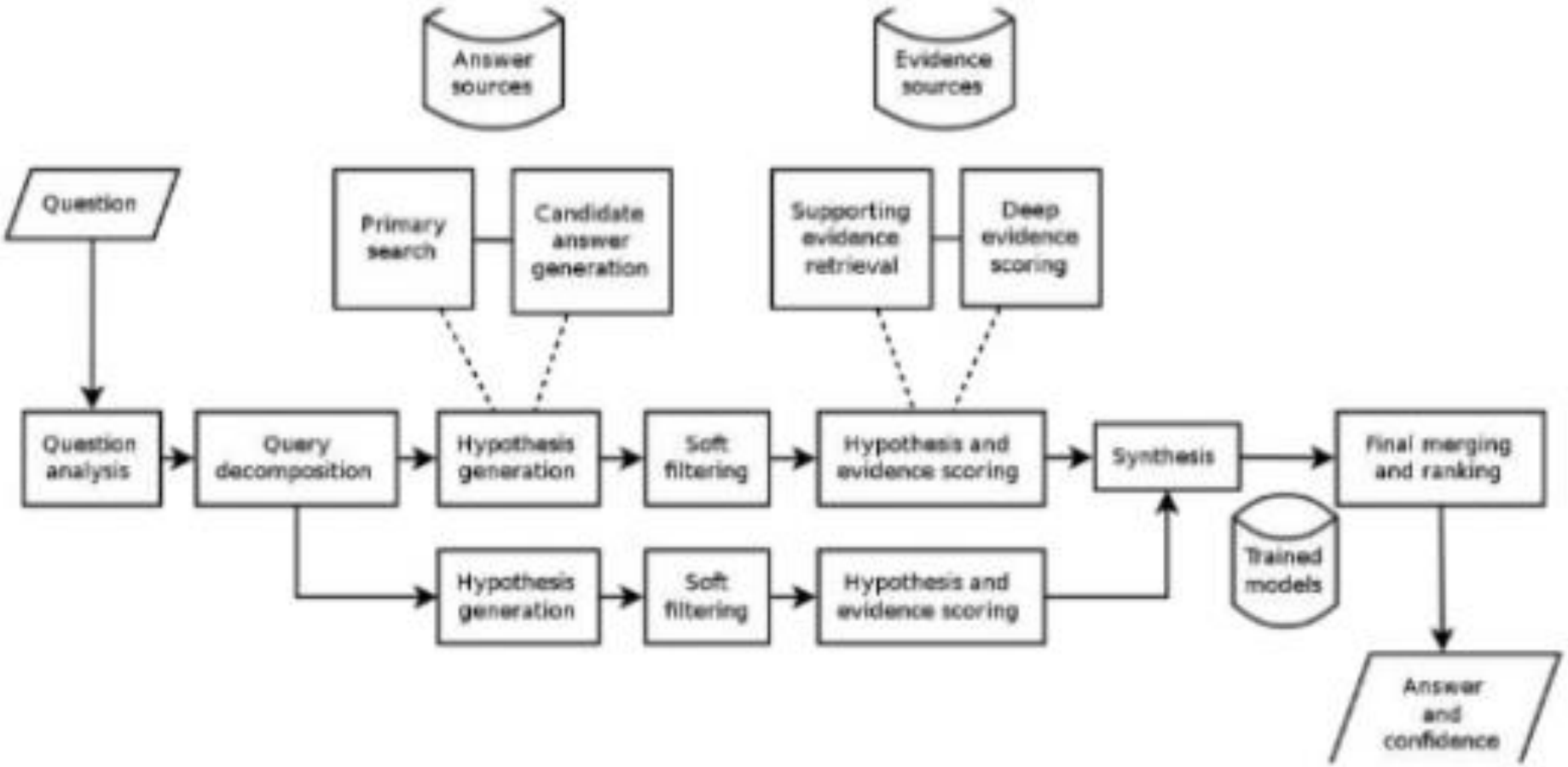
Watson project

- Watson is a computer which is capable of answering question issued in natural language.
- Questions come from quiz show called Jeopardy.
- The software of this project is called DeepQA project.
- In 2011, Watson won the former winners of quiz show Jeopardy

Watson – description

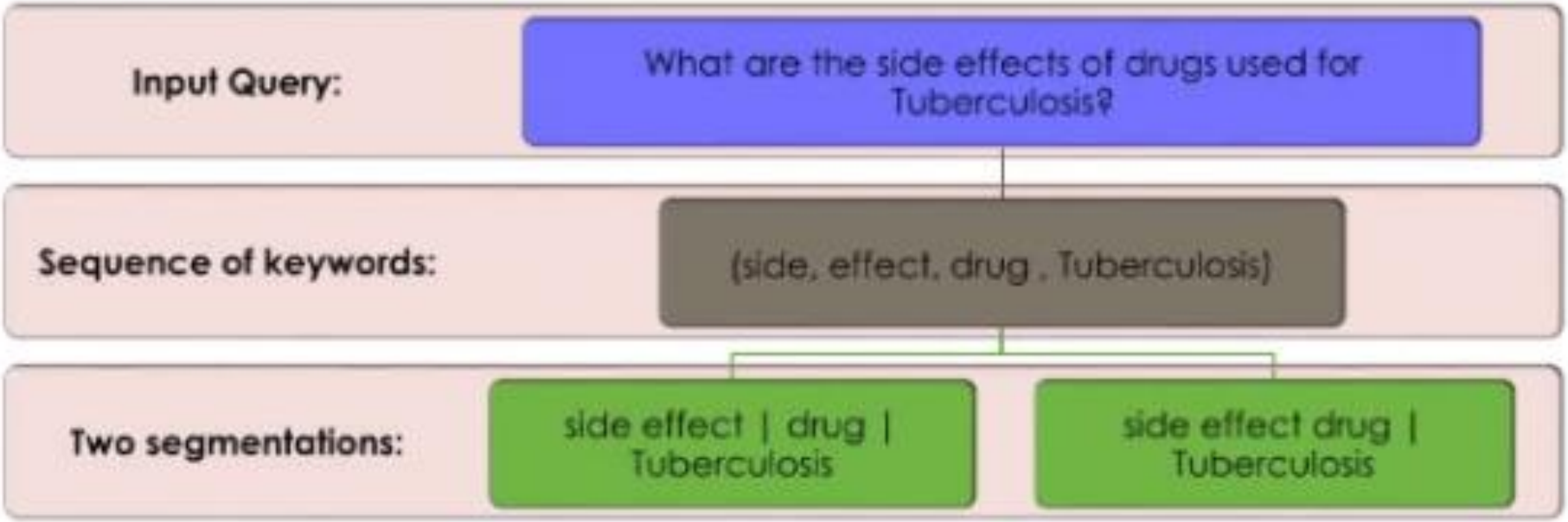
- Hardware: Watson system has 2,880 POWER processor threads and has 16 terabytes of RAM.
- Data: encyclopedias, dictionaries, thesauri, newswire articles, and literary works. Watson also used databases, taxonomies, and ontologies such as DBPedia, WordNet, and Yago were used.
- Software: DeepQA software and the Apache UIMA framework. The system was written in various languages, including Java, C++, and Prolog, and runs on the SUSE Linux Enterprise Server 11 operating system using Apache Hadoop framework to provide distributed computing.

DeepQA architecture



Question analysis

- Define most important phrases in each question to generate queries by keywords



Ambiguity resolution

- Resource disambiguation is the process of recognizing the suitable resources in the underlying knowledge base.