

Course: Common Sense Reasoning

12. Using Natural Language to Access to Data

Martin Molina



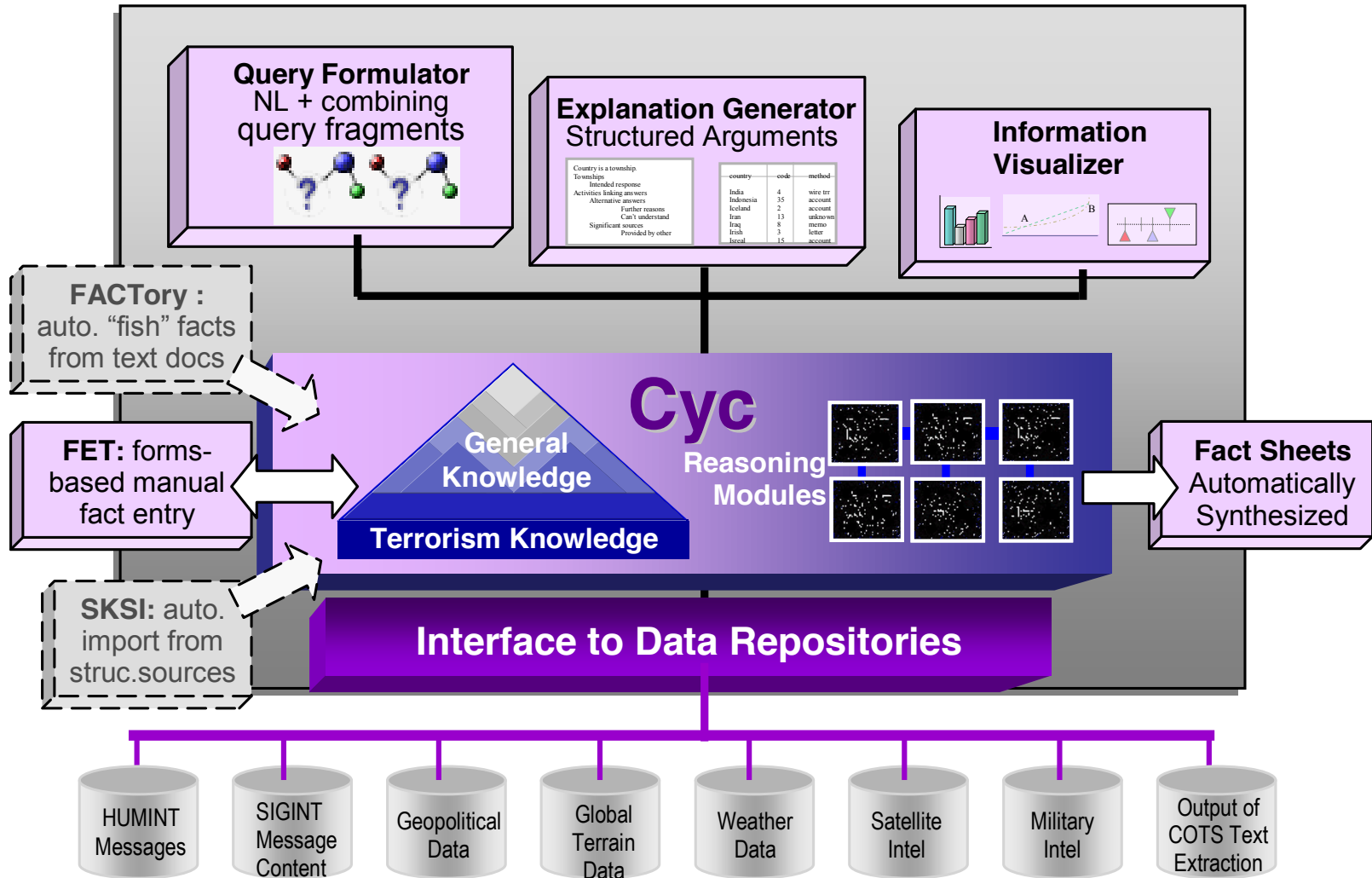
Common sense knowledge can facilitate natural language access to data

- Questions is ordinary language
- Using multiple databases
- Common sense can be used for taxonomic reasoning

[Condoravdi, et al. 2015]

[Strzalkowski, Harabugiu, 2006]

Cyc Analytic Environment (CAE) is a Cyc application in the domain of terrorism



CAE facilitates access to data about terrorism

- CAE uses a large amount of terrorism data:
 - Over 300,000 facts about 2,700 terrorists, 800 terrorists groups, 14,000 attacks.
- CAE accepts diverse types of questions:
 - When and where of an attack
 - Weapons and tactics used
 - Individuals and groups
 - Affected people (killed, captured, etc.)

Example of query and answer

Query:

What terrorist attacks that occurred later than the year 2000 targeted government building in capital cities of greater than 1,000,000 residents?

Answer:

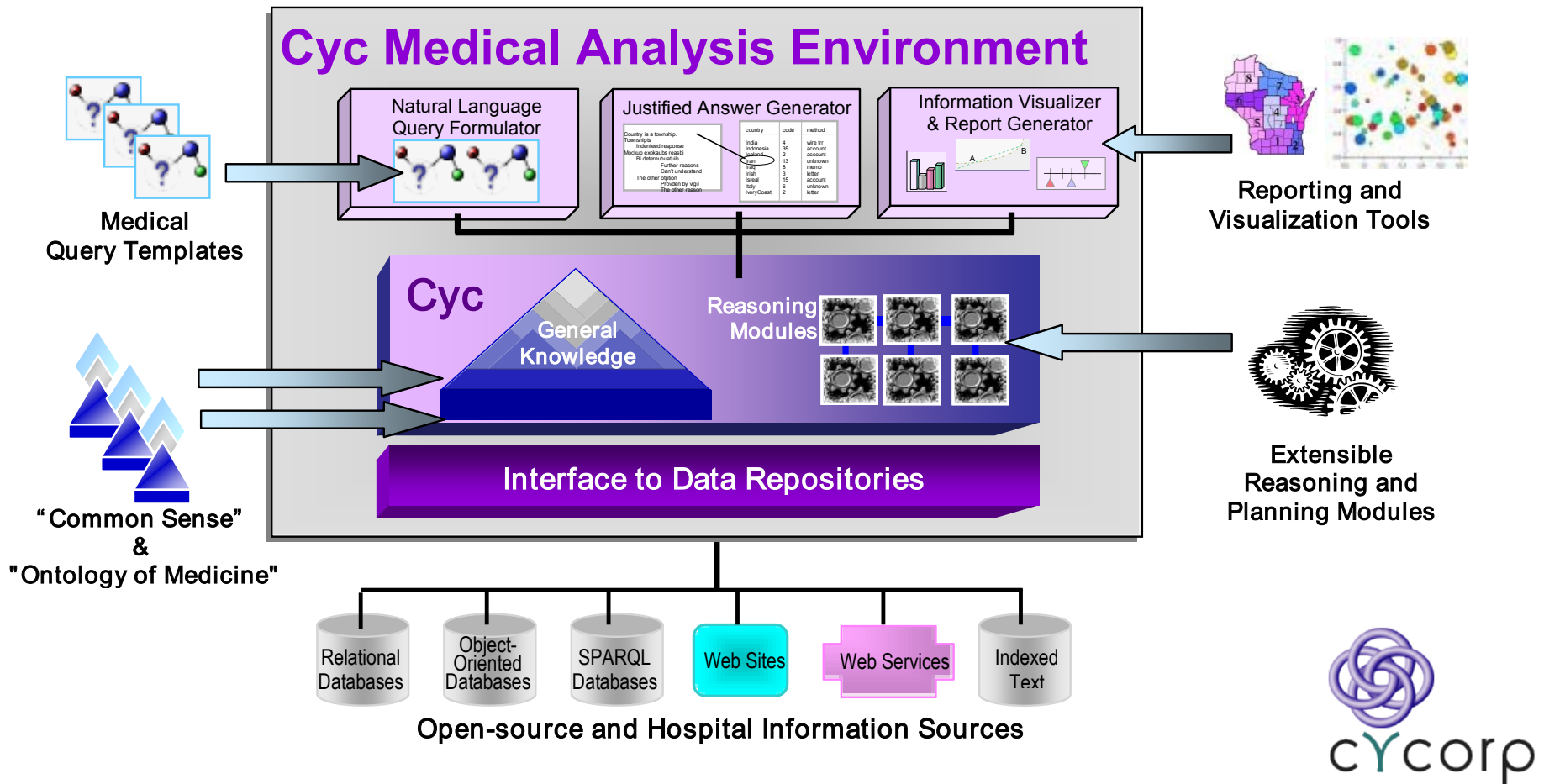
- Setting: Ankara, Turkey
- Population: 1,877,755
- Date: March 26, 2001
- Government building: Iran embassy in Ankara

Explanation:

- Iran embassy in Ankara is an embassy
- An embassy is a type of government building
- Ankara is the capital city of Turkey
- ...

In the explanation, the sentence “an embassy is a type of government building” describes common sense knowledge used to generate the answer

Medical Analysis Environment (MAE) is a Cyc application in the domain of medicine



Example questions in MAE

- Which patients had a heart attack less than 2 weeks prior to a coronary artery bypass graft (CABG) between 2000 and 2005?
- What were the post-operative morbidity and mortality rates for patients who had septicemia or bacteremia within one month following a pericardial window?
- What percentage of aortic valve replacement (AVR) patients in 2006 received a premarket aortic prosthesis?
- What is the percentage of patients that had a mitral valve prosthesis that was manufactured by an EU company implanted between 2000 and 2005 and later explanted?
- For patients over 65 years old, what was the average time between successive stent surgeries?

Watson is a general question-answering system



IBM Watson

- Watson is capable of answering open domain questions written in natural language
- Watson won the first prize of the American quiz TV show "Jeopardy" competing against human opponents in 2011

[Ferrucci et al. 2010]

Watson answers questions over a broad domain of topics

Clue:

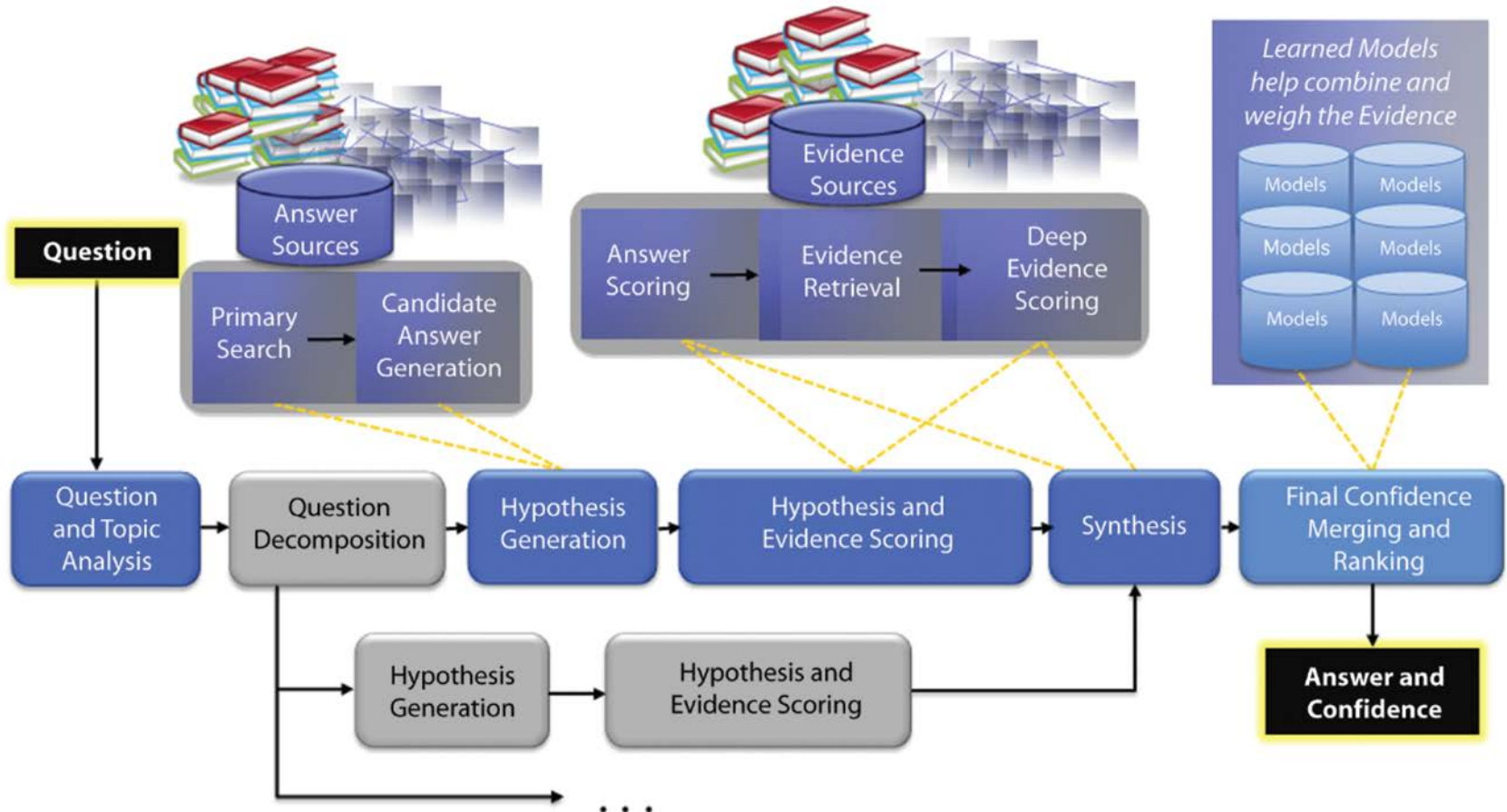
“This drug has been shown to relieve the symptoms of ADD with relatively few side effects”

Answer in ≤ 5 seconds:

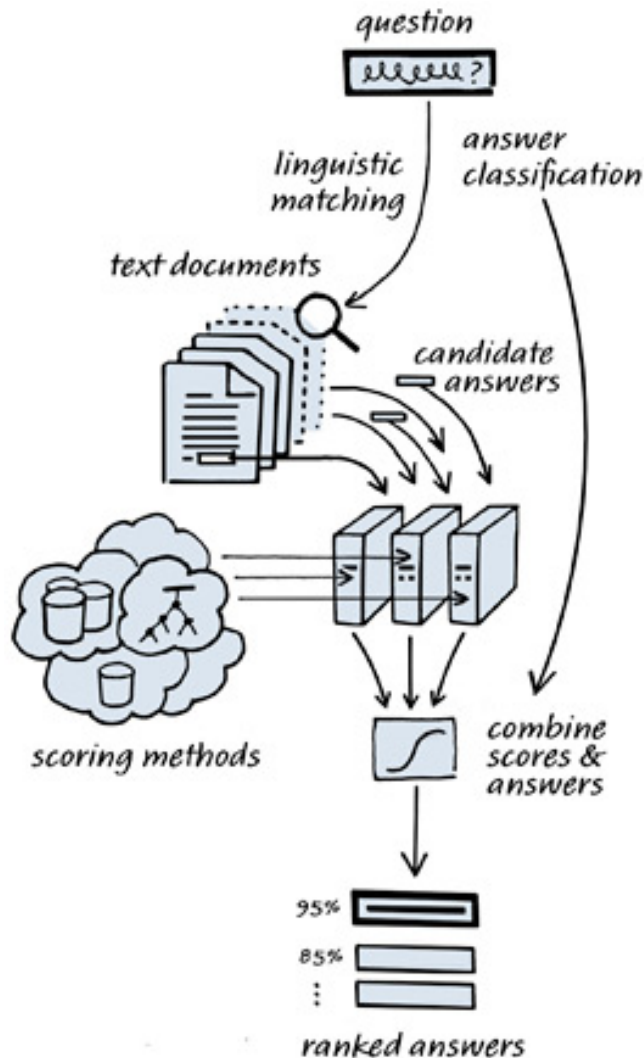
“Ritalin”

NOTE: According to the rules of the TV show, the answer is actually given as a question, i.e., “What is Ritalin?”

Watson uses information acquired from documents from the web, encyclopedias, dictionaries, etc.



Watson first generates candidate answers and then combines and scores them to generate the final answer

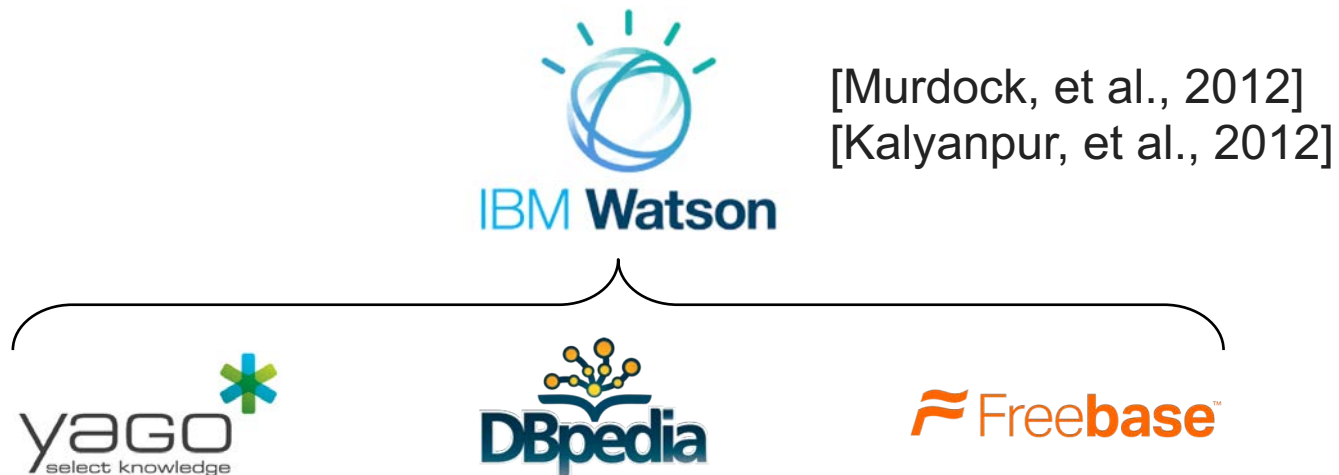


Watson uses multiple processes running in parallel with specialized tasks

Source of figure:
<http://blog.stephenwolfram.com/2011/01/jeopardy-ibm-and-wolframalpha/>

Watson uses large scale data bases for taxonomic reasoning

Watson uses Yago, DBpedia and Freebase for taxonomic reasoning (e.g., typing coercion) including subsumption, disjointness, etc.,



Course “Common sense reasoning”.
© 2019 Martin Molina

This work is licensed under Creative Commons license CC BY-NC-SA 4.0:
<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>



Work citation in APA style:

Molina, M. (2019). Common sense reasoning [Lecture slides]. OpenCourseWare, Universidad Politécnica de Madrid. Retrieved from <http://ocw.upm.es/course>