Az IBM Watson mögötti MI

Mátyás-Barta Csongor

2015. január 9.

Áttekintés

Automatic Open-Domain Question Answering

Megadva egy *széles ismerettartományra* vonakozó *természetes nyelvű kérdést*, szolgáltasson:

- pontos válaszokat
- pontos válasz bizonyosságot
- helyes magyarázatot
- gyorsan

Jeopardy

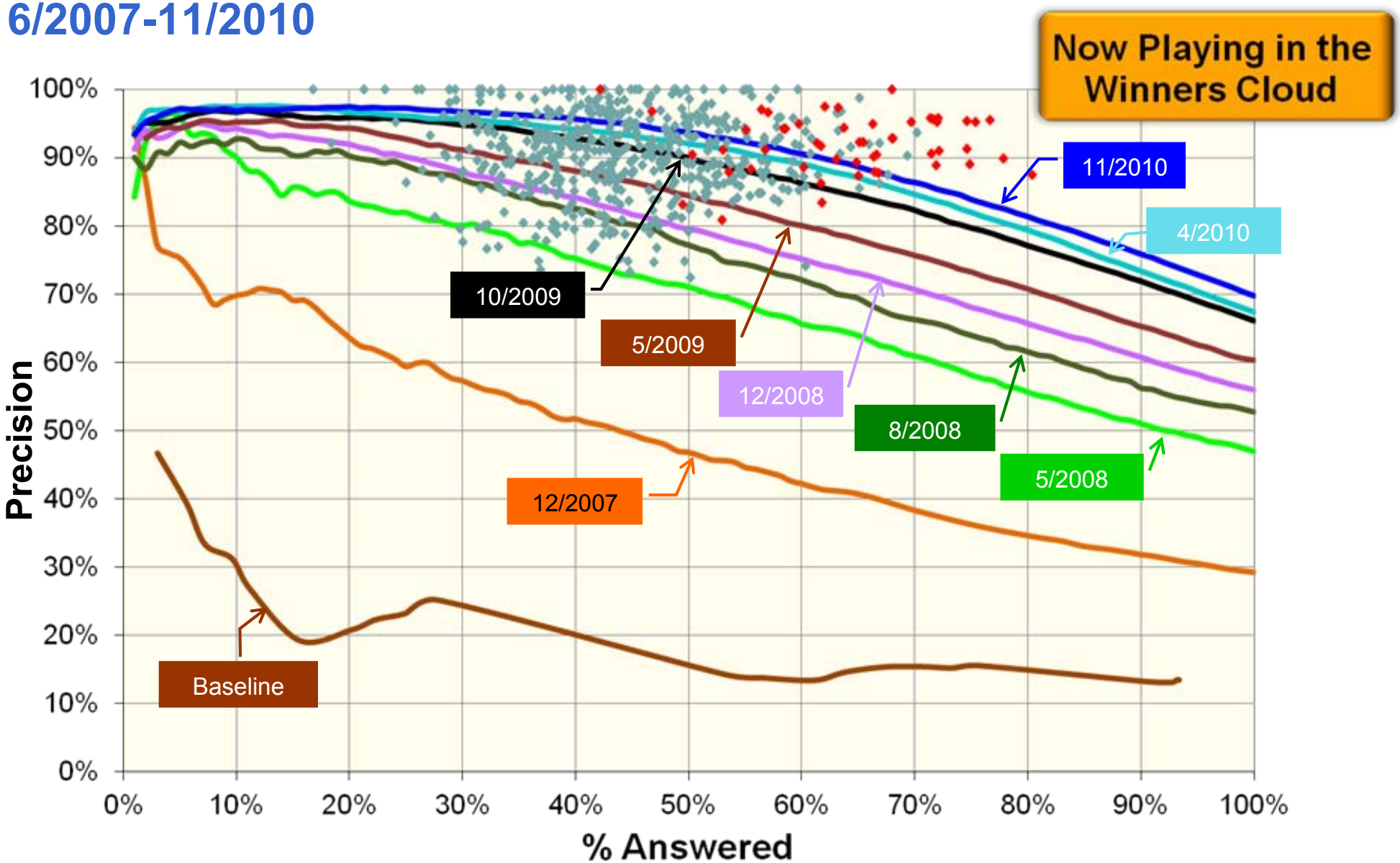
- széles ismerettartományra vonatkozó kérdések
- több játékos
- számít a gyorsaság
- számít a pontosság

Jeopardy

- *Category:* Diplomatic Relations
- *Clue:* Of the four countries in the world that the United States does not have diplomatic relations with, the one that's farthest north.
- *Inner subclue:* The four countries in the world that the United States does not have diplomatic relations with (Bhutan, Cuba, Iran, North Korea).
- *Outer subclue:* Of Bhutan, Cuba, Iran, and North Korea, the one that's farthest north.
- *Answer:* North Korea

- enumerate zero or more decomposition hypotheses for each question as possible interpretations.

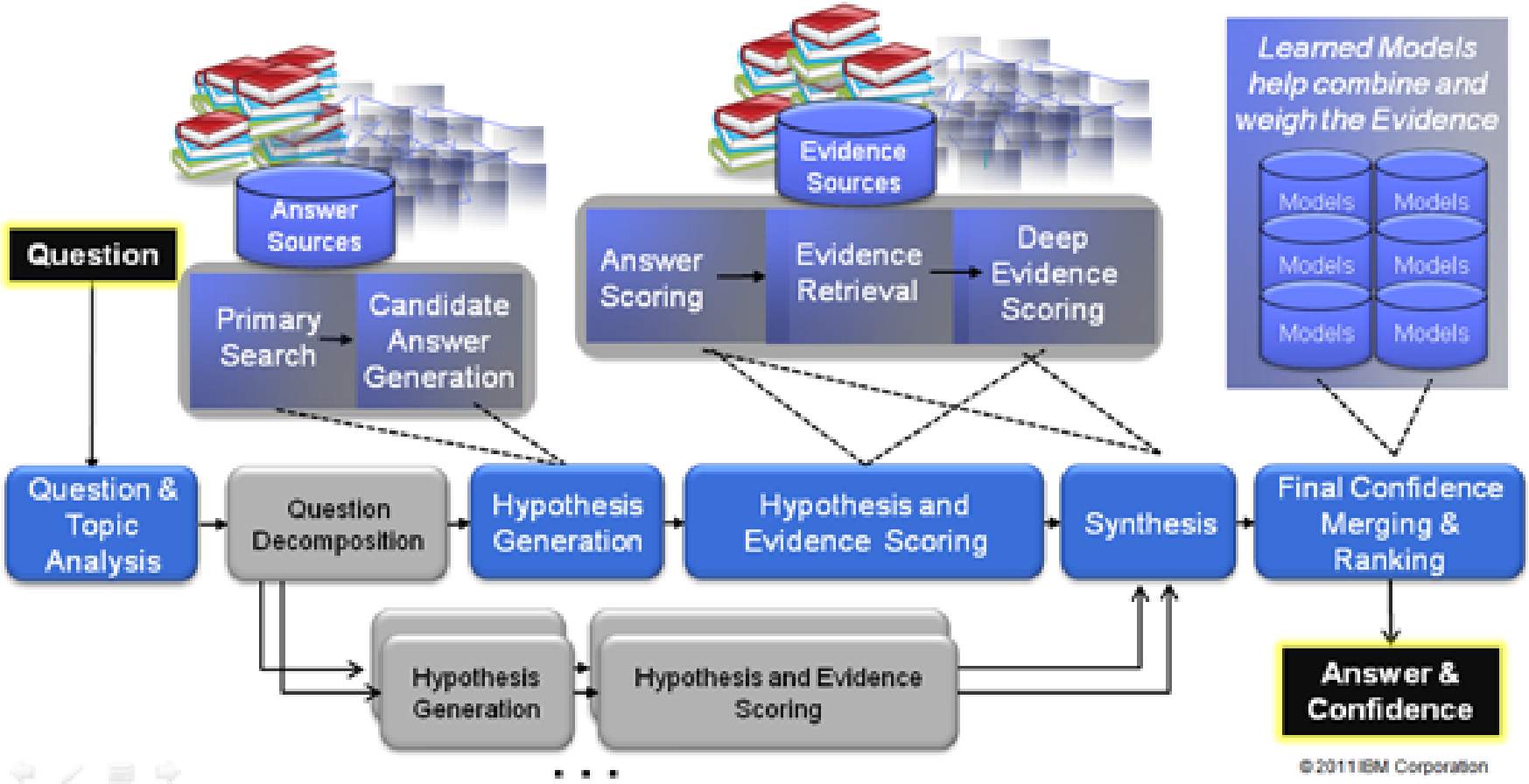
Deep QA: Incremental Progress in Precision and Confidence 6/2007-11/2010



DeepQA

- *Massive parallelism*: Exploit massive parallelism in the consideration of multiple interpretations and hypotheses.
- *Many experts*: Facilitate the integration, application, and contextual evaluation of a wide range of loosely coupled probabilistic question and content analytics.
- *Pervasive confidence estimation*: No component commits to an answer; all components produce features and associated confidences, scoring different question and content interpretations. An underlying confidence-processing substrate learns how to stack and combine the scores.
- *Integrate shallow and deep knowledge*: Balance the use of strict semantics and shallow semantics, leveraging many loosely formed ontologies.

Architektúra - DeepQA



DeepQA - Content Acquisition

- offline
- question type (manual) and domain(automatic) analysis
- corpus: encyclopedias, dictionaries, thesauri, newswire articles, literary works
- corpus expansion process:
  - identify seed documents and retrieve related documents from the web
  - extract self-contained text nuggets from the related web documents
  - score the nuggets based on whether they are informative with respect to the original seed document
  - merge the most informative nuggets into the expanded corpus
- other kinds of semistructured and structured content: dbPedia, WordNet, and the Yago ontology

Automatic Learning for “Reading”

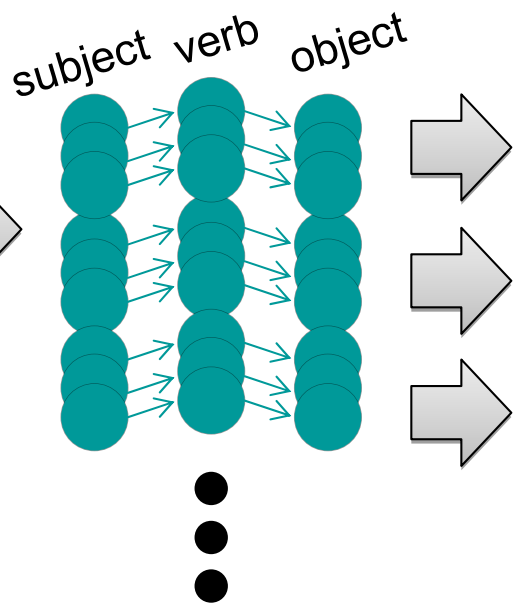
Sentence Parsing

Generalization & Statistical Aggregation

Volumes of Text

Syntactic Frames

Semantic Frames



Inventors patent inventions (.8)

Officials Submit Resignations (.7)
People earn degrees at schools (0.9)

Fluid is a liquid (.6)
Liquid is a fluid (.5)

Vessels Sink (0.7)
People sink 8-balls (0.5) (in pool/0.8)

DeepQA - Question Analysis

- online, during the contest
- mixture of experts: shallow parses, deep parses, logical forms, semantic role labels, coreference ...
- **question classification**: puzzle, math question, definition question; puns, constraints, subclues
- **focus and LAT detection**
- **relation detection**: "They're the two states you could be reentering if you're crossing Florida's northern border," - borders(Florida,?x,north)
- **decomposition**: rule based deep-parsing and statistical classification - subquestions

Hypothesis Generation and Soft Filtering

- **primary search**: find as much as possible answer bearing content; Multiple text search engine, knowledge based search using SPARQL, triple store queries
- **candidate answer generation**: specific to each search result type; Recall over precision is prefered. Several hundred candidate answers generated.

Soft filtering

- lightweight analysis, scoring of candidates, with threshold
- for example the likelihood of a candidate answer being an instance of the LAT
- lets through 100 candidates

Hypothesis and Evidence Scoring

- **evidence retrieval**, for example passage search: returns passages that contain the candidate answer used in the context of the original question terms
- **scoring**: multiple scoring algos, determine the degree of certainty that the evidence supports the candidate answers; multiple scorers, common format, allows fast expansion and tunning. Scoring for example based on source reliability, geospatial location, temporal relationships, taxonomic classification, aliases, logical form alignment etc.
- Evidence profile for each candidate answer: contains all scores for the given answer

Evaluating Possibilities and Their Evidence

In cell division, mitosis splits the nucleus & cytokinesis splits this liquid *cushioning* the nucleus.

- *Organelle*
- *Vacuole*
- *Cytoplasm*
- *Plasma*
- *Mitochondria*
- *Blood …*

- Many candidate answers (CAs) are generated from many different searches
- Each possibility is evaluated according to different dimensions of evidence.
- Just One piece of evidence is if the CA is of the right type. In this case a "liquid".

Is("Cytoplasm", "liquid") = 0.2↑

Is("organelle", "liquid") = 0.1

Is("vacuole", "liquid") = 0.2

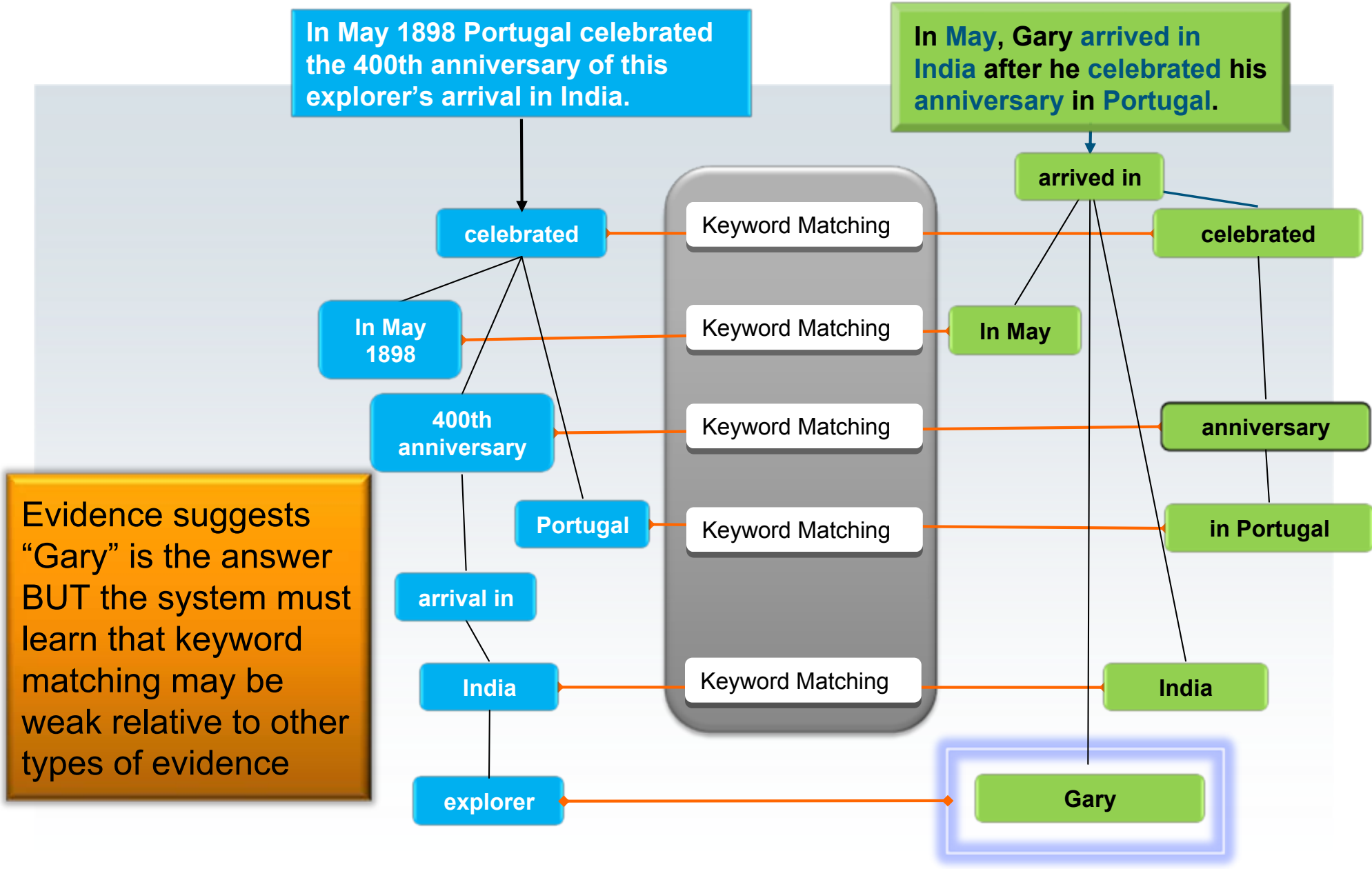
Is("plasma", "liquid") = 0.7

"Cytoplasm is a fluid *surrounding* the nucleus…"

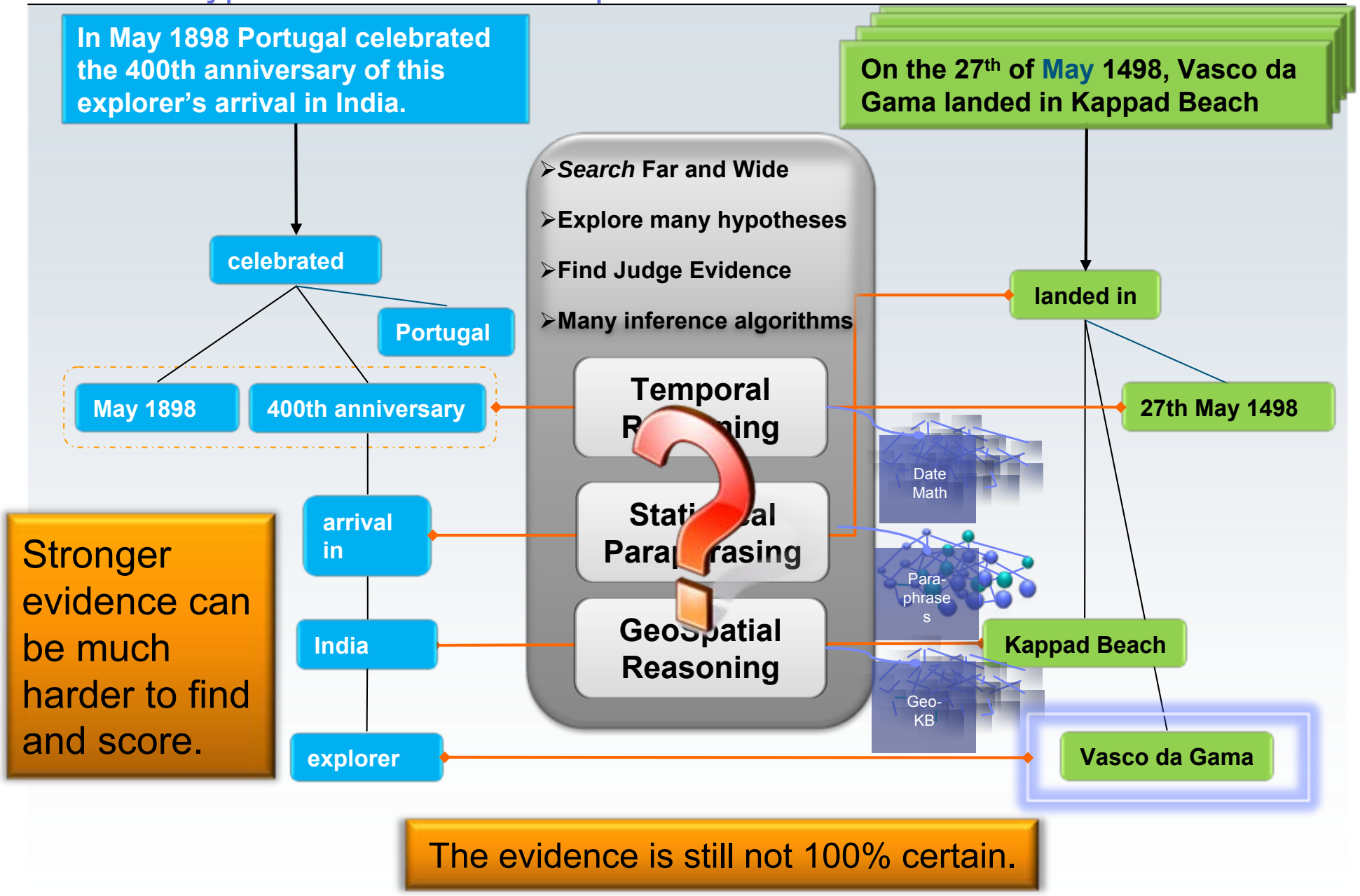
Wordnet → Is_a(Fluid, Liquid) → ?

Learned → Is_a(Fluid, Liquid) → yes.

Different Types of Evidence: Keyword Evidence



Different Types of Evidence: Deeper Evidence



Final merging and ranking

- **answer merging**: merge the equivalent and related hypotheses and their scores (Abraham Lincoln and Honest Abe)
- **ranking and confidence estimation**: machine learning aproach: different scores are considered for different types of questions

Köszönöm a figyelmet!

Használt források

- Fóliák - http://www.cs.hku.hk/watson/
- http://www.aaai.org/Magazine/Watson/watson.php