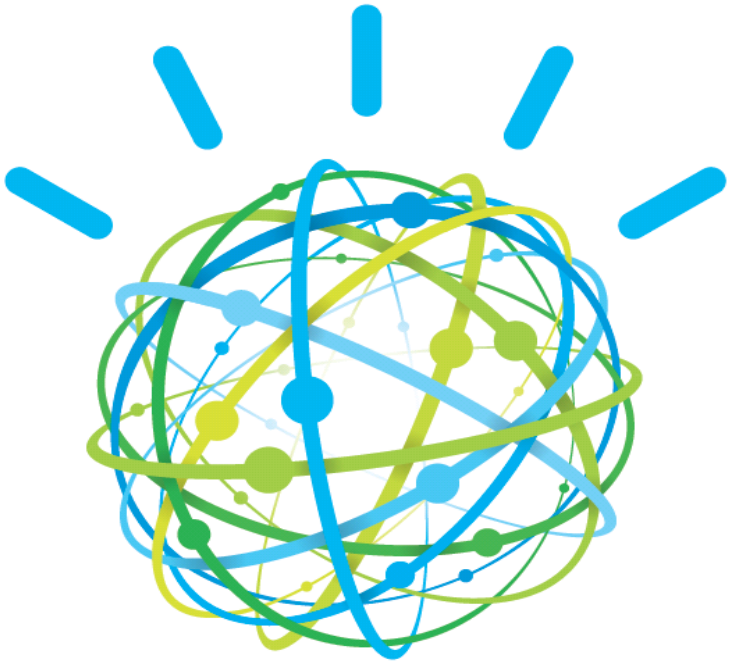
After Watson



Wlodek Zadrozny

UNCC Computer Science Dept.

Watson Research Team,
IBM (Jan 2008– Dec 2012).

“Watson became possibly the first nonhuman millionaire by besting its human competition”

Final Score:	$ 24,000	**$ 77,147**	$ 21,600

PLAN

1. Setting the stage: what NLP can/can't do
2. Classical semantics in a caricature
3. How Watson
   A. does not answer a question
   B. answers a question
4. Watson innovations in semantics
5. Discussion

A few meta-points

Personal view -- I'm not representing IBM's view of Watson

I believe there's a logic in Watson way's of doing things, and hope to start a discussion by abstracting a few important features of Watson

It should be connected to work already done (hope for your suggestions)

Successes of language processing technologies

1. Gist translation (e.g. Google translation)
2. Information extraction into (e.g. financial) databases
3. Essay grading
4. Stock market predictions based on sentiment (anxiety) in the blogs
5. U.S. Social Security applications approval
6. Email generation in Obama campaign
7. Mental conditions diagnosis

Return on investment > 10x, 100x?

However …

Natural language processing remains an AI-complete and unsolved problem (along with computer vision)

For example… the Turing test

(…) chatbots competing in the final round for the Loebner Prize (…)

The contest implements the Turing Test by having judges sit at computer terminals and decide whether a human or chatbot is talking.

Judges were able to determine chatbots were talking after only three or four lines of chat, sometimes less, *and the chatbots often delivered irrelevant, off-the-wall responses.*

New Scientist (10/20/11)

2012 update: "(…) the (best) program did not fool any judges (…)"

What makes NLP an AI-complete problem?

- Language (meaning) is ambiguous and contextual
- Language (use) requires reasoning beyond current capabilities of machines
- Language understanding requires understanding of the difference btw. reality and fiction
- Understanding might require having a body (symbol grounding)
- Understanding might only be possible within a human body and brain. [very controversial]

Text in classical compositional semantics (very rough sketch)

- Parse the text, create a semantic representation (compositionally) for each sentence
- Create a model of the text by summing all relations
- Use an equivalence relation on discourse entities (coreference) to create a better model

Text in classical compositional semantics (very rough sketch)

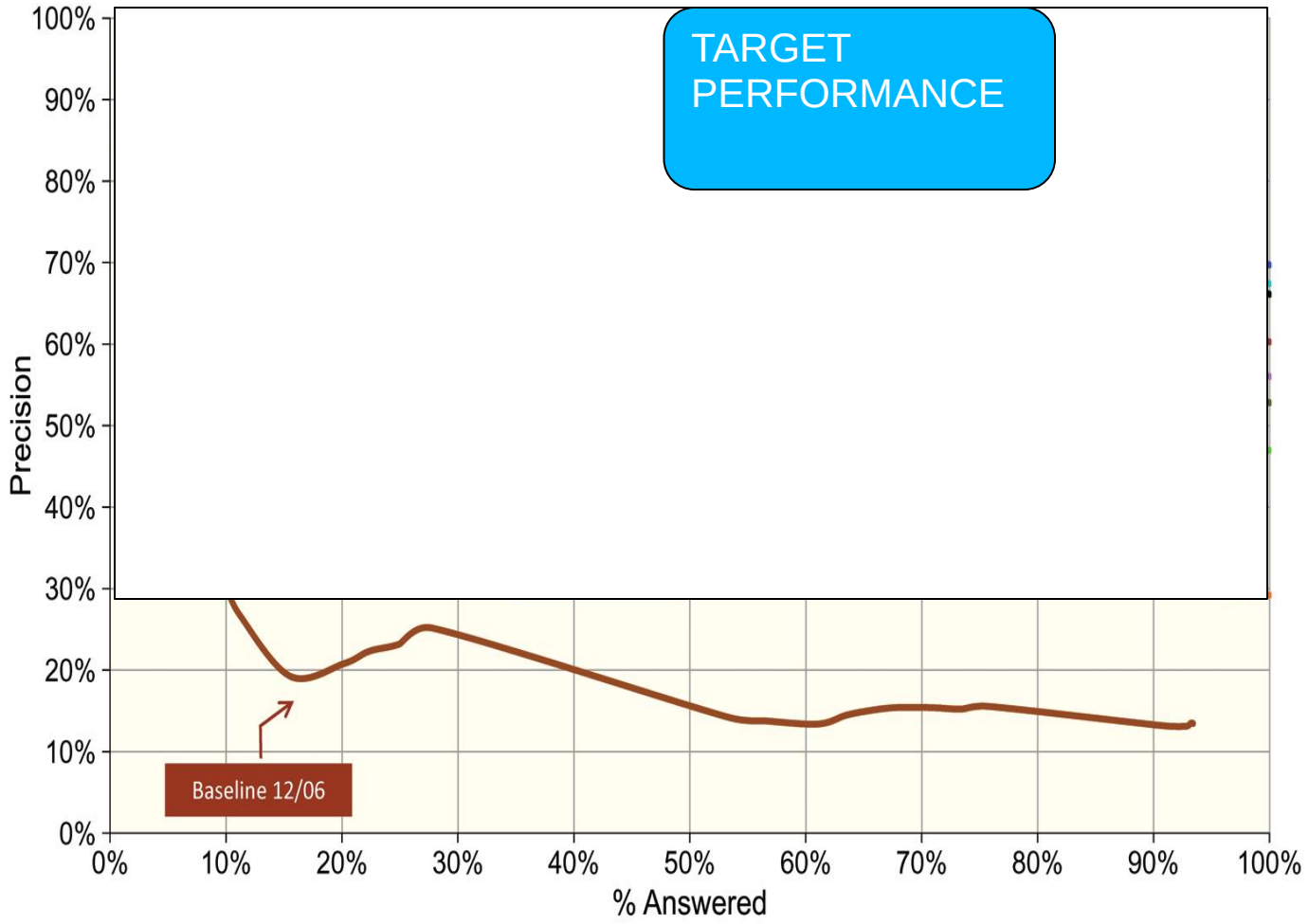
- Parse the text, create a semantic representation (compositionally) for each sentence
- Create a model of the text by summing all relations
- Use an equivalence relation on discourse entities (coreference) to create a better model

Then *question answering* can be done via unification

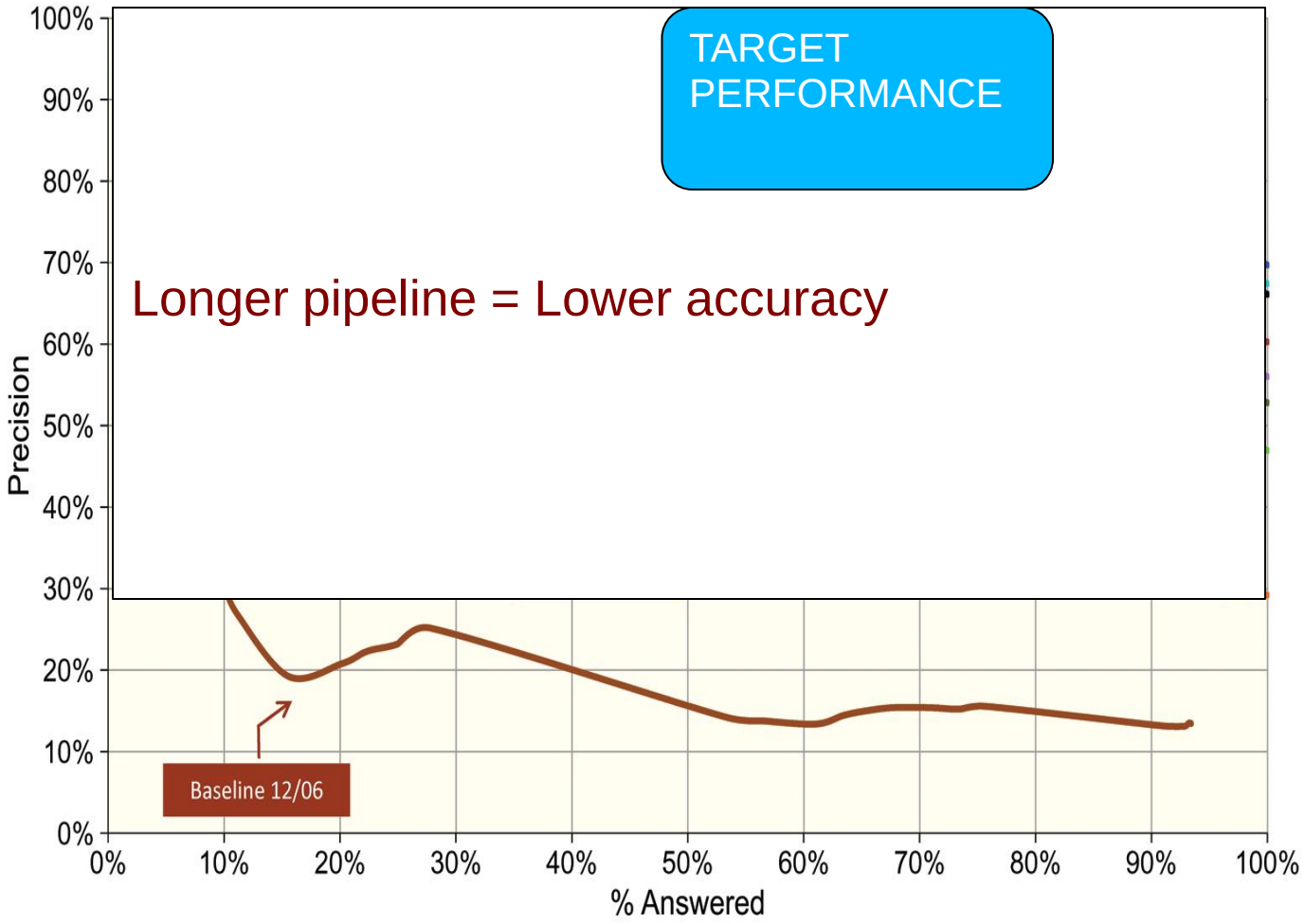
Question answering before Watson

- Question classification: *who, what, when*, *how big,* …
- Search for the right passages
- Interpretation (shallow and partial compositional semantics)
- Matching relations (“unification”) to produce an answer
- Estimating the likelihood of the answer being correct

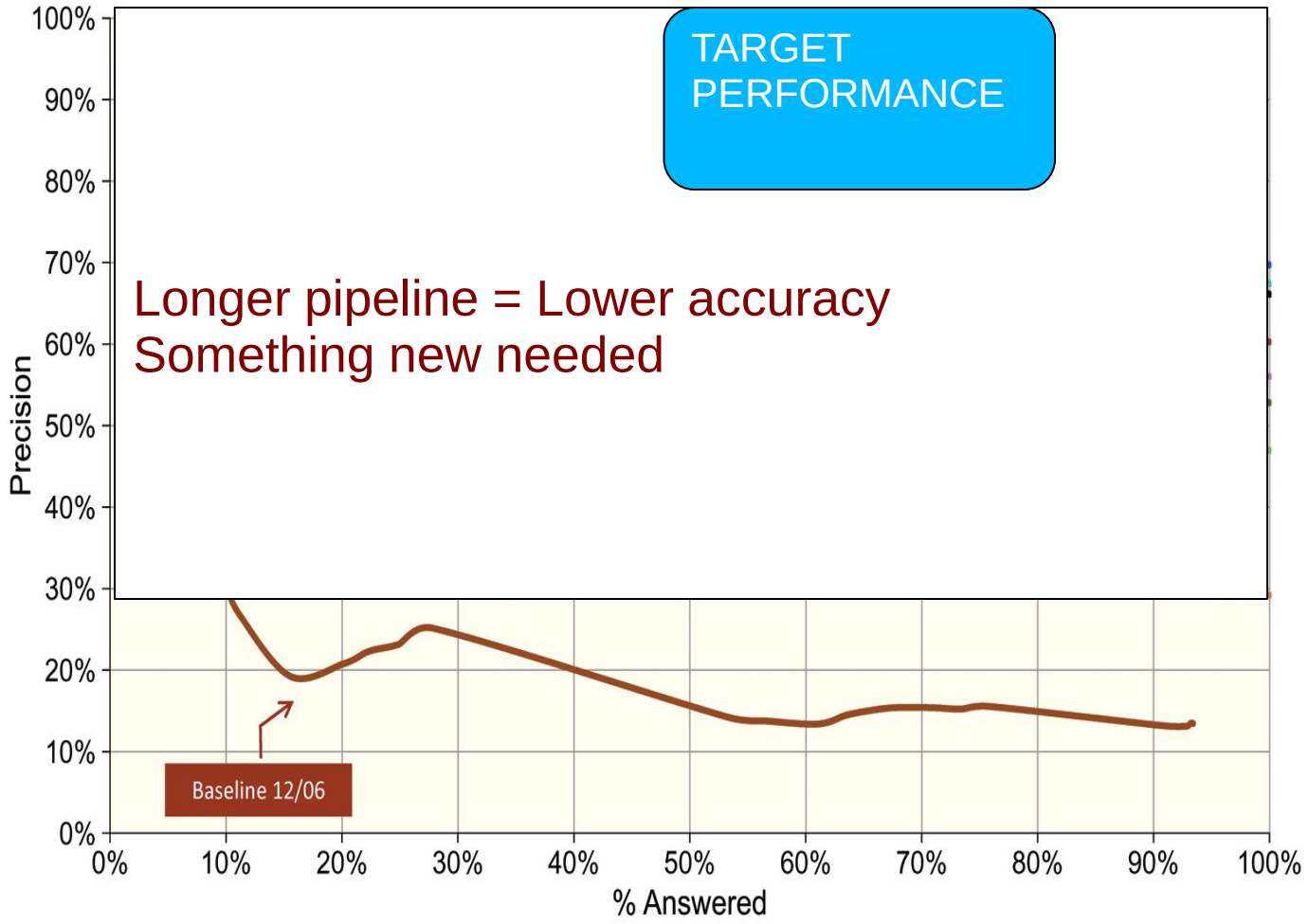
The classical pipeline didn’t work



The classical pipeline didn’t work



The classical pipeline didn’t work



Question Answering in Watson

- Why is open domain question answering difficult? – The language problem and the knowledge problem.
- Why are Jeopardy questions difficult?
- How does Watson answer a question?

Where was X born? Is it a hard Question?

DB

Person	Birth Place
Einstein	ULM

Information extraction: Albert Einstein was born in Ulm, in the Kingdom of Württemberg in the German Empire on 14 March 1879.

Supporting evidence for Watson: *One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein´s birthplace.*

The Jeopardy! Challenge: *A compelling and notable way to **drive** and **measure** the technology **of automatic Question Answering** along 5 Key Dimensions*

Broad/Open Domain

Complex Language

High Precision

Accurate Confidence

High Speed

$200
If you're standing, it's the direction you should look to check out the wainscoting.

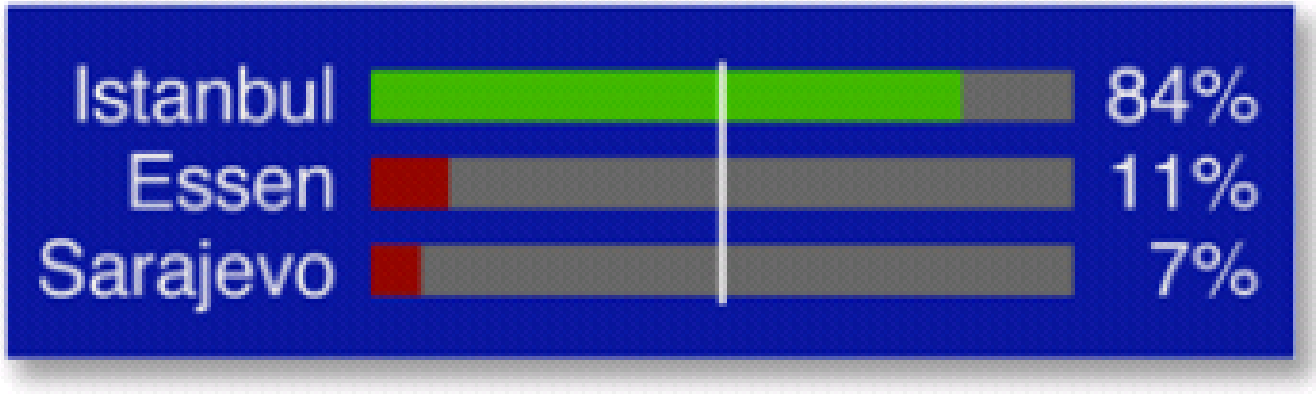
$1000
The first person mentioned by name in 'The Man in the Iron Mask' is this hero of a previous book by the same author.

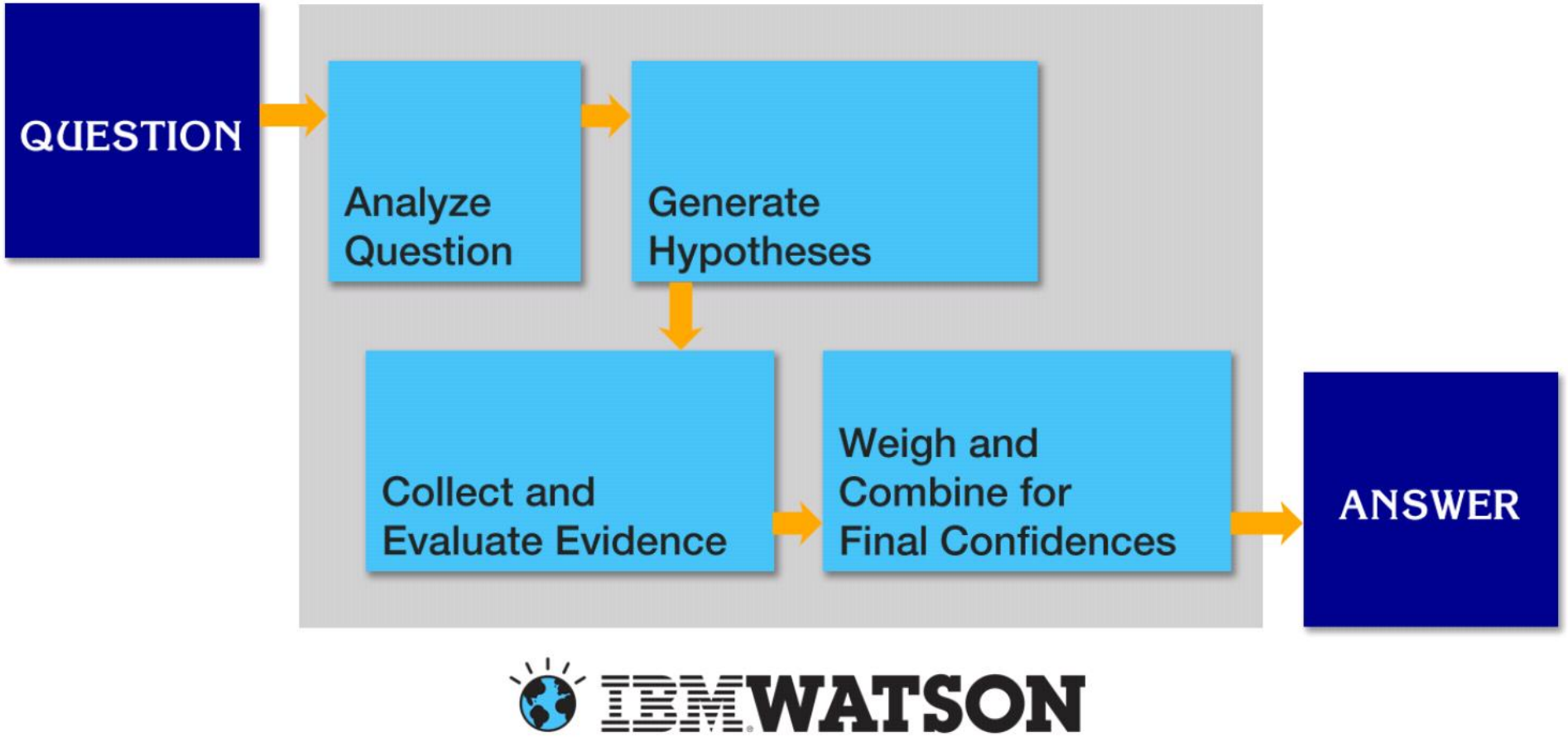
$600
In cell division, mitosis splits the nucleus & cytokinesis splits this liquid *cushioning* the nucleus

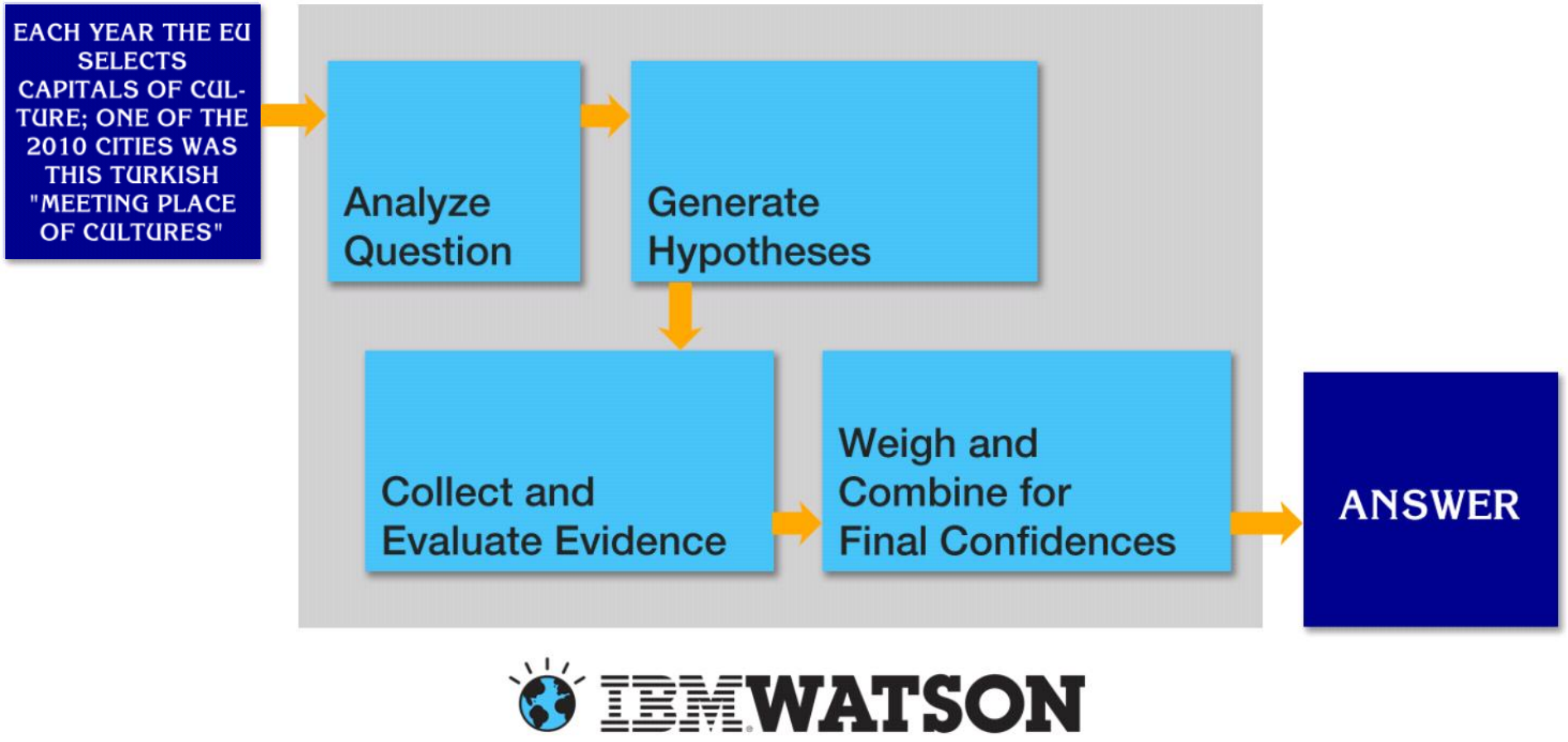
$2000
Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north

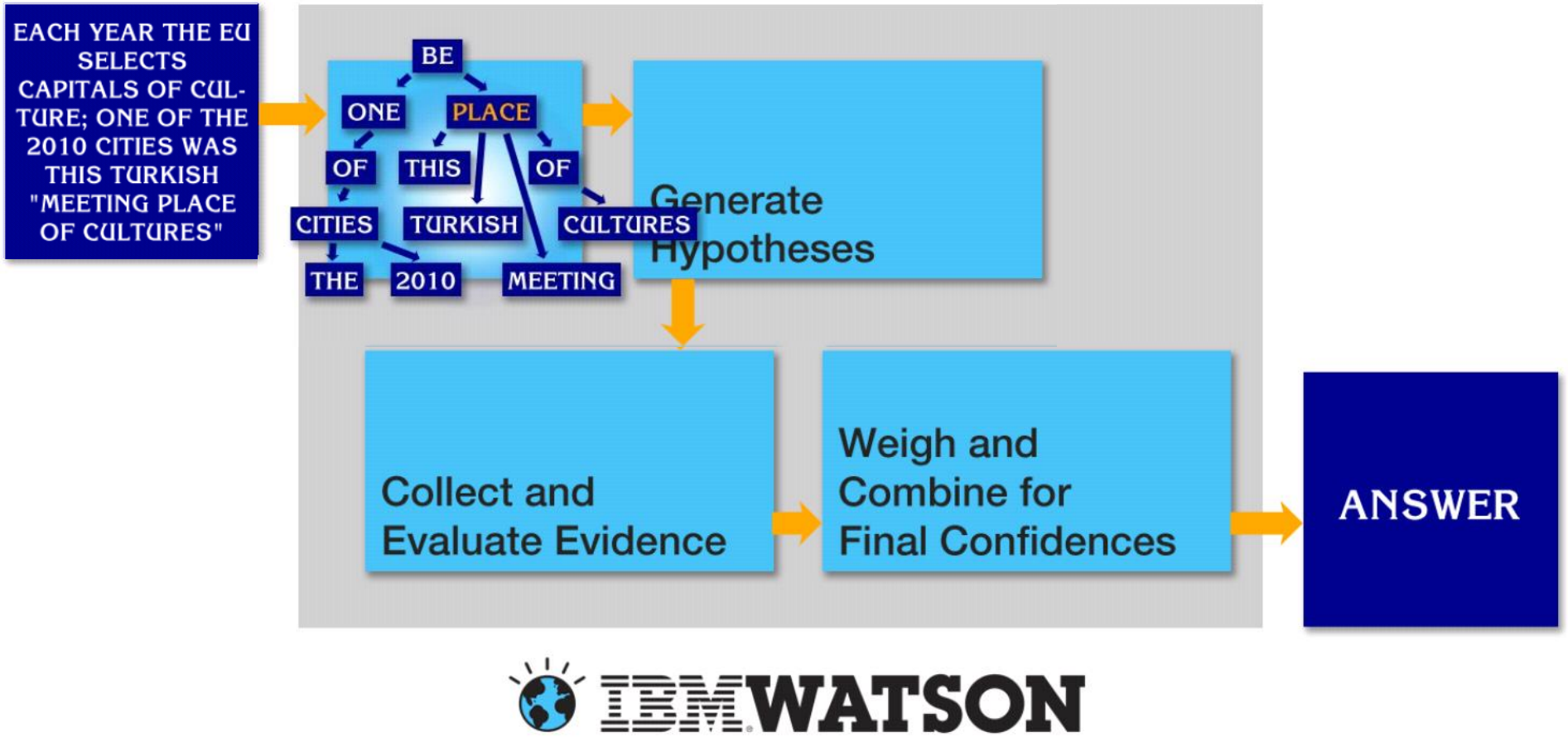
EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"

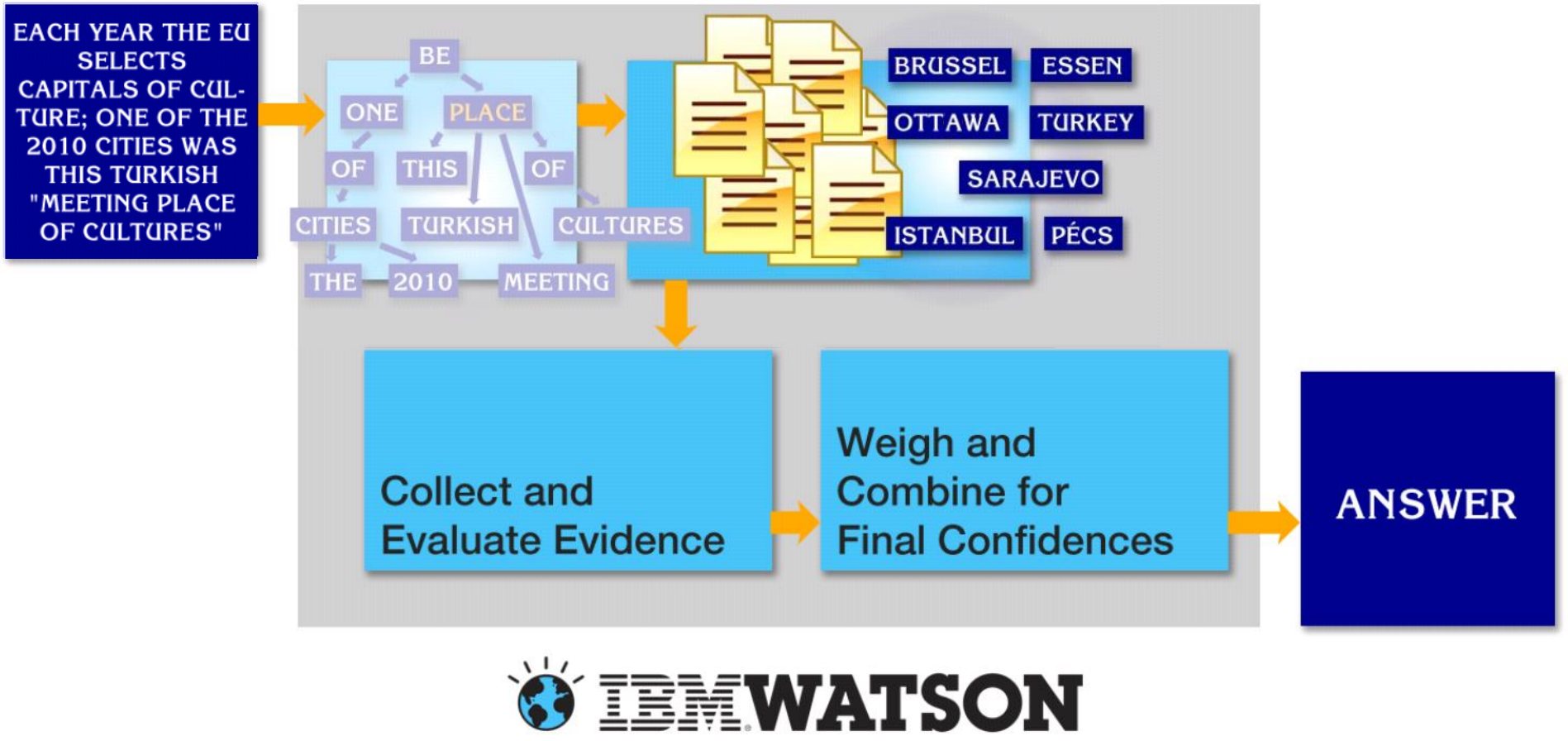
WHAT IS ISTANBUL? ✓



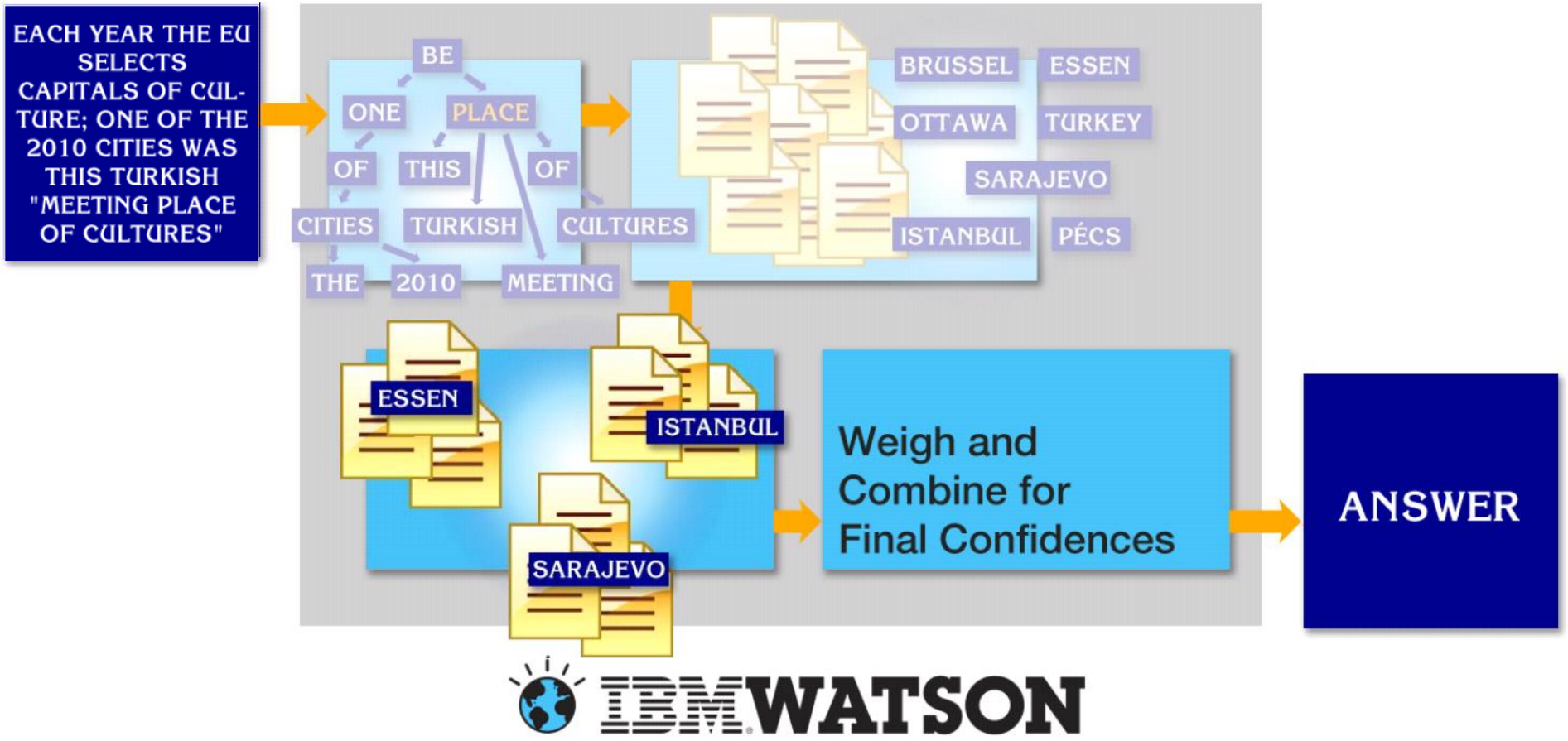

QUESTION
Analyze Question
Generate Hypotheses
Collect and Evaluate Evidence
Weigh and Combine for Final Confidences
ANSWER
IBM WATSON

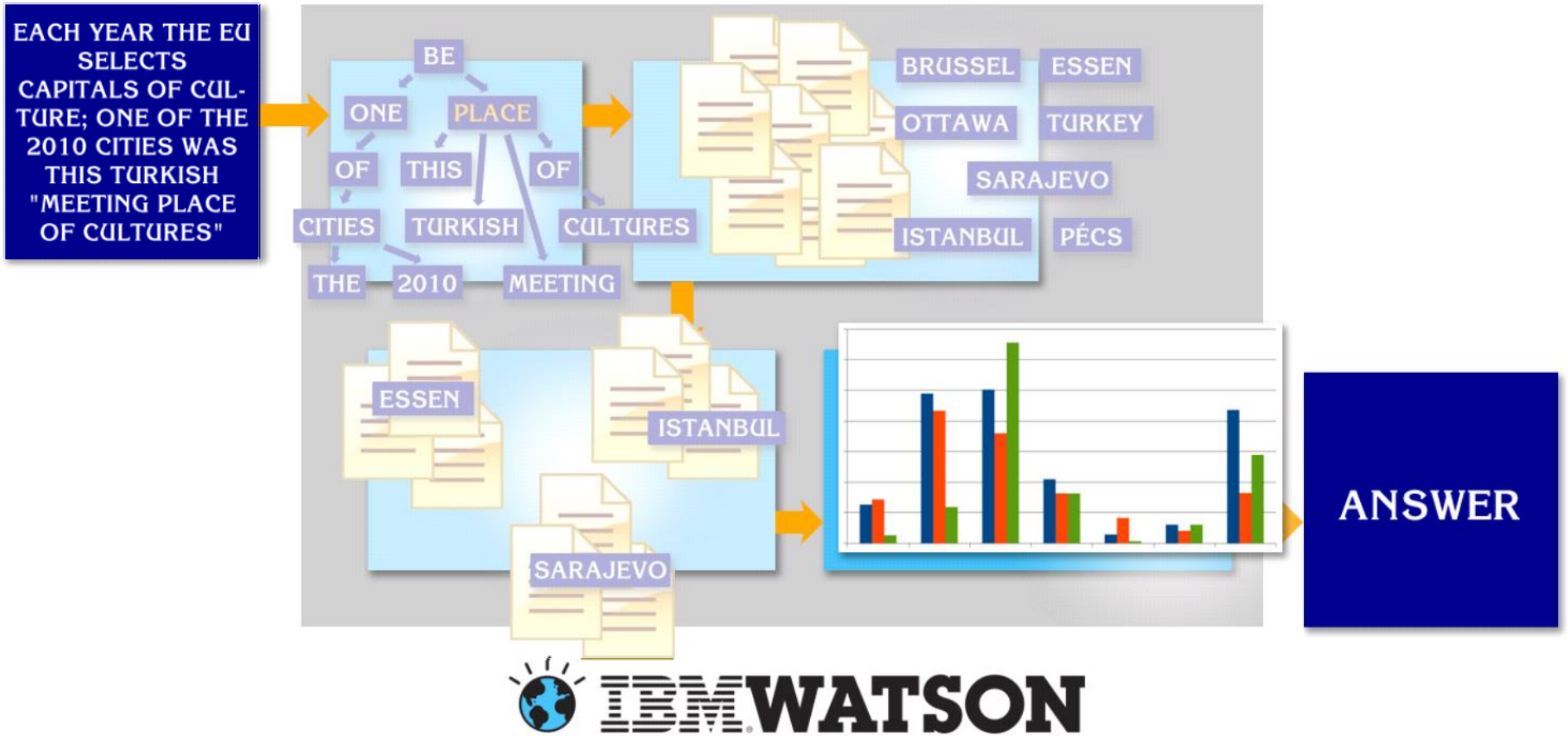

EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
Analyze Question
Generate Hypotheses
Collect and Evaluate Evidence
Weigh and Combine for Final Confidences
ANSWER
IBM WATSON


EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
BE
ONE
PLACE
OF
THIS
OF
CITIES
TURKISH
CULTURES
THE
2010
MEETING
Generate Hypotheses
Collect and Evaluate Evidence
Weigh and Combine for Final Confidences
ANSWER
IBM WATSON

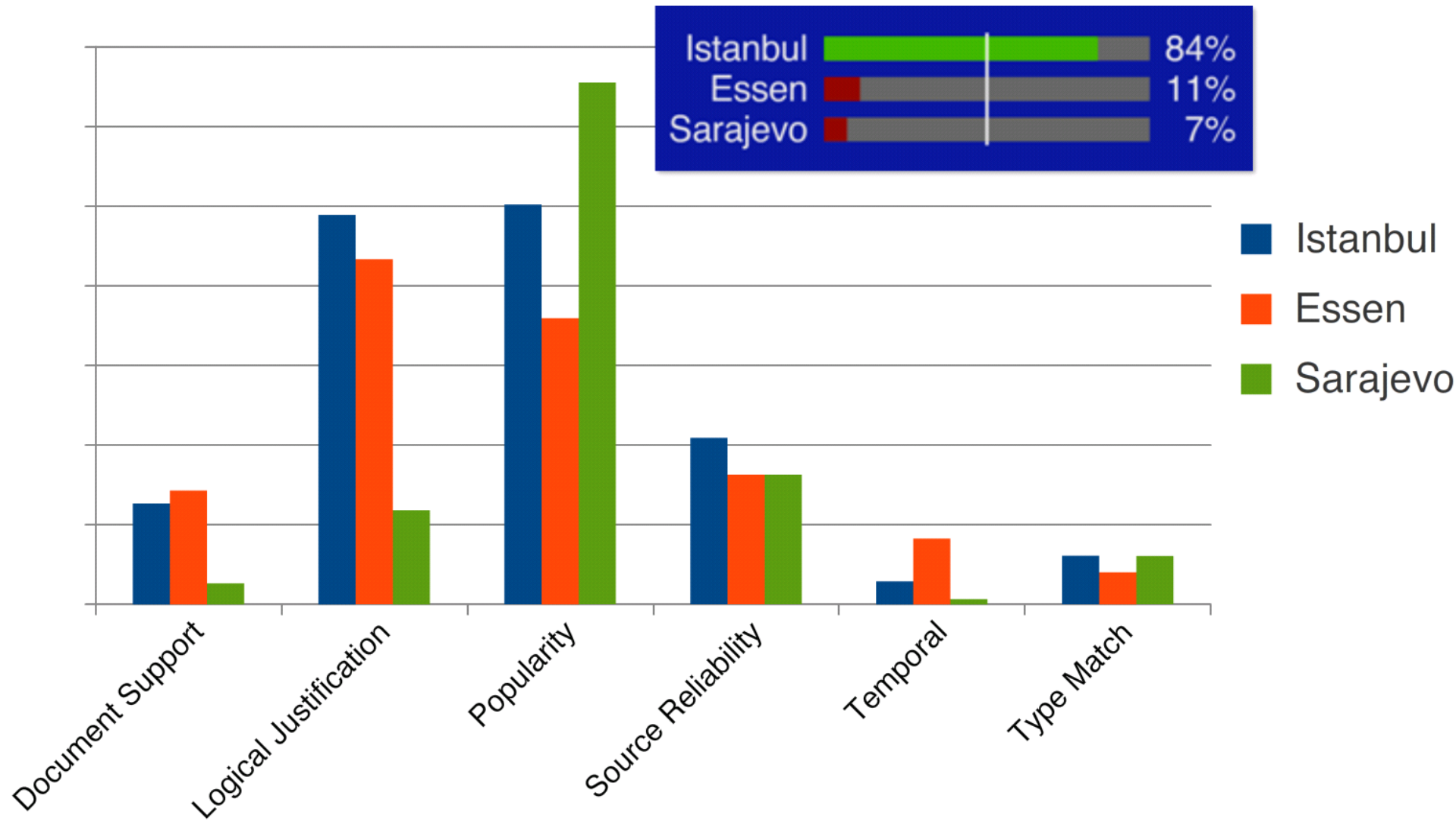

EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
BE
ONE
PLACE
OF
THIS
OF
CITIES
TURKISH
CULTURES
THE
2010
MEETING
BRUSSEL
ESSEN
OTTAWA
TURKEY
SARAJEVO
ISTANBUL
PÉCS
Collect and Evaluate Evidence
Weigh and Combine for Final Confidences
ANSWER

IBM WATSON

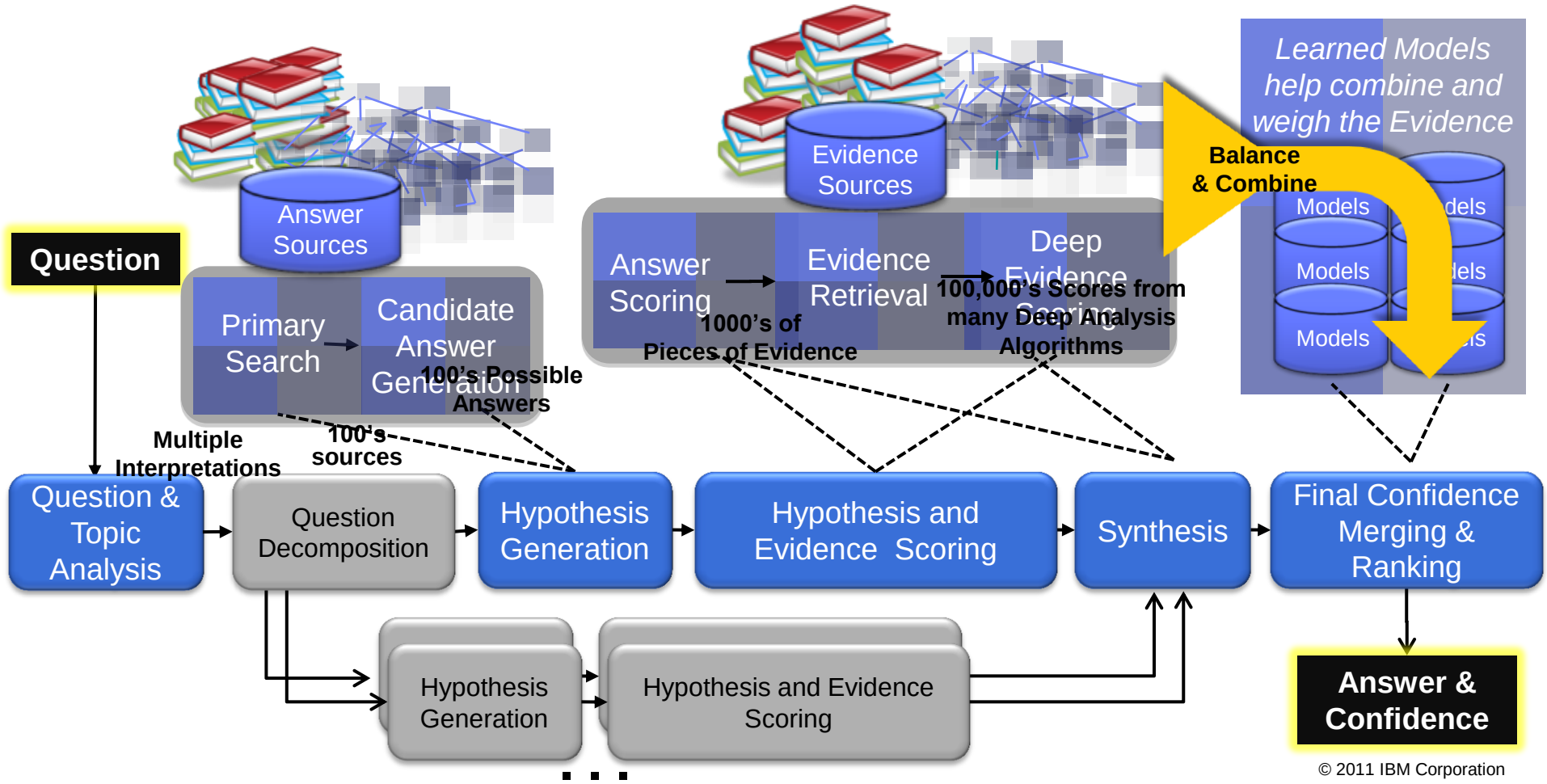

EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
BE
ONE
PLACE
OF
THIS
OF
CITIES
TURKISH
CULTURES
THE
2010
MEETING
BRUSSEL
ESSEN
OTTAWA
TURKEY
SARAJEVO
ISTANBUL
PÉCS
ESSEN
ISTANBUL
SARAJEVO
Weigh and Combine for Final Confidences
ANSWER
IBM WATSON


EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
BE
ONE
PLACE
OF
THIS
OF
CITIES
TURKISH
CULTURES
THE
2010
MEETING
BRUSSEL
ESSEN
OTTAWA
TURKEY
SARAJEVO
ISTANBUL
PÉCS
ESSEN
ISTANBUL
SARAJEVO
ANSWER
IBM WATSON


EACH YEAR THE EU SELECTS CAPITALS OF CULTURE; ONE OF THE 2010 CITIES WAS THIS TURKISH "MEETING PLACE OF CULTURES"
BE
ONE
PLACE
OF
THIS
OF
CITIES
TURKISH
CULTURES
THE
2010
MEETING
BRUSSEL
ESSEN
OTTAWA
TURKEY
SARAJEVO
ISTANBUL
PÉCS
ESSEN
ISTANBUL
SARAJEVO
WHAT IS ISTANBUL?
IBM WATSON
Istanbul 84%
Essen 11%
Sarajevo 7%


Istanbul 84%
Essen 11%
Sarajevo 7%
Istanbul
Essen
Sarajevo
Document Support
Logical Justification
Popularity
Source Reliability
Temporal
Type Match

Watson Architecture



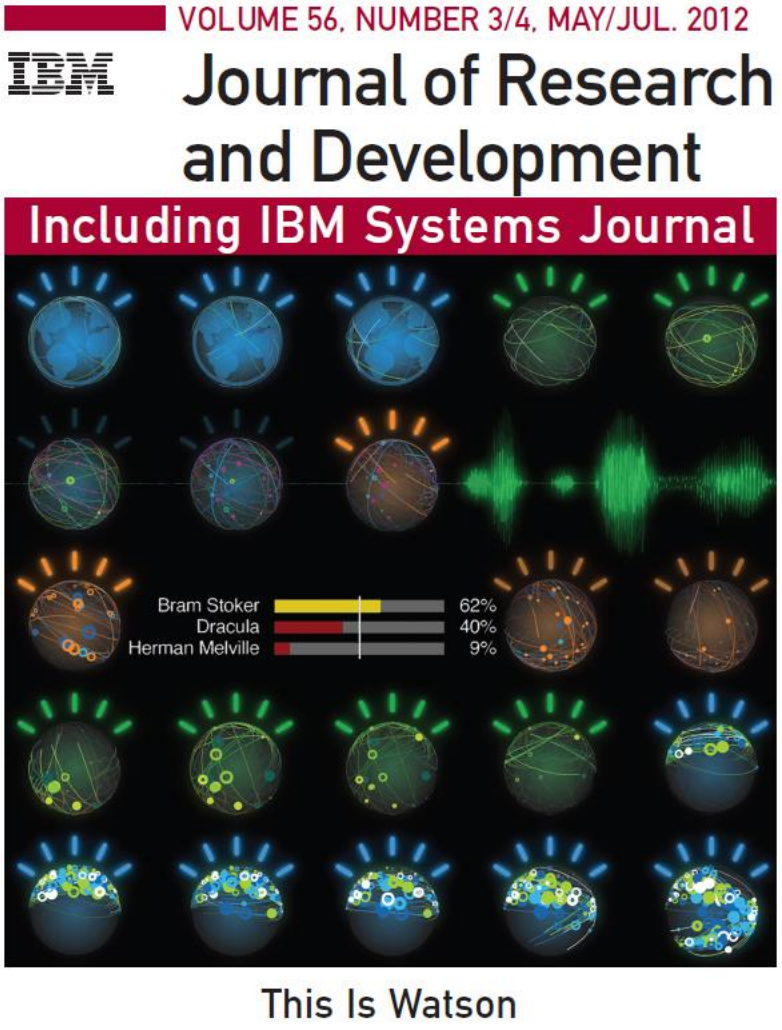
Watson's conceptual innovations

- Dialogue-like view of meaning: *generate candidate interpretations, find supporting evidence, and evaluate it*
- Aboutness of knowledge – *source engineering* to convert knowledge into a referent-oriented format
- Contextual meaning derivation – *deferred type evaluation* to make type matching based on all available evidence

Why these three?

- Empirically valuable and formally unexplored
- I don’t know a logic that would support all of them
- Aspects seem appear in works of other researchers

Is it all? No, we can extend the list to 50+ innovations



This Is Watson

patft.uspto.gov

50+ patent applications

Several patents already granted

Dialogue-like view of language meaning

Dialogue-like view meaning (and not a static, text centric view). The meaning – typically, *a correct referent-- will emerge from evaluation of evidence* in the task of answering a question.

- The answer is a "paraphrase" of the question.
  Multiple meanings are pursued in parallel
- The evidence is knowledge needed to understand the question.
  Multiple collections of data support different interpretations
- Evaluation is the process of choosing the best interpretation.

Process: Generate and evaluate

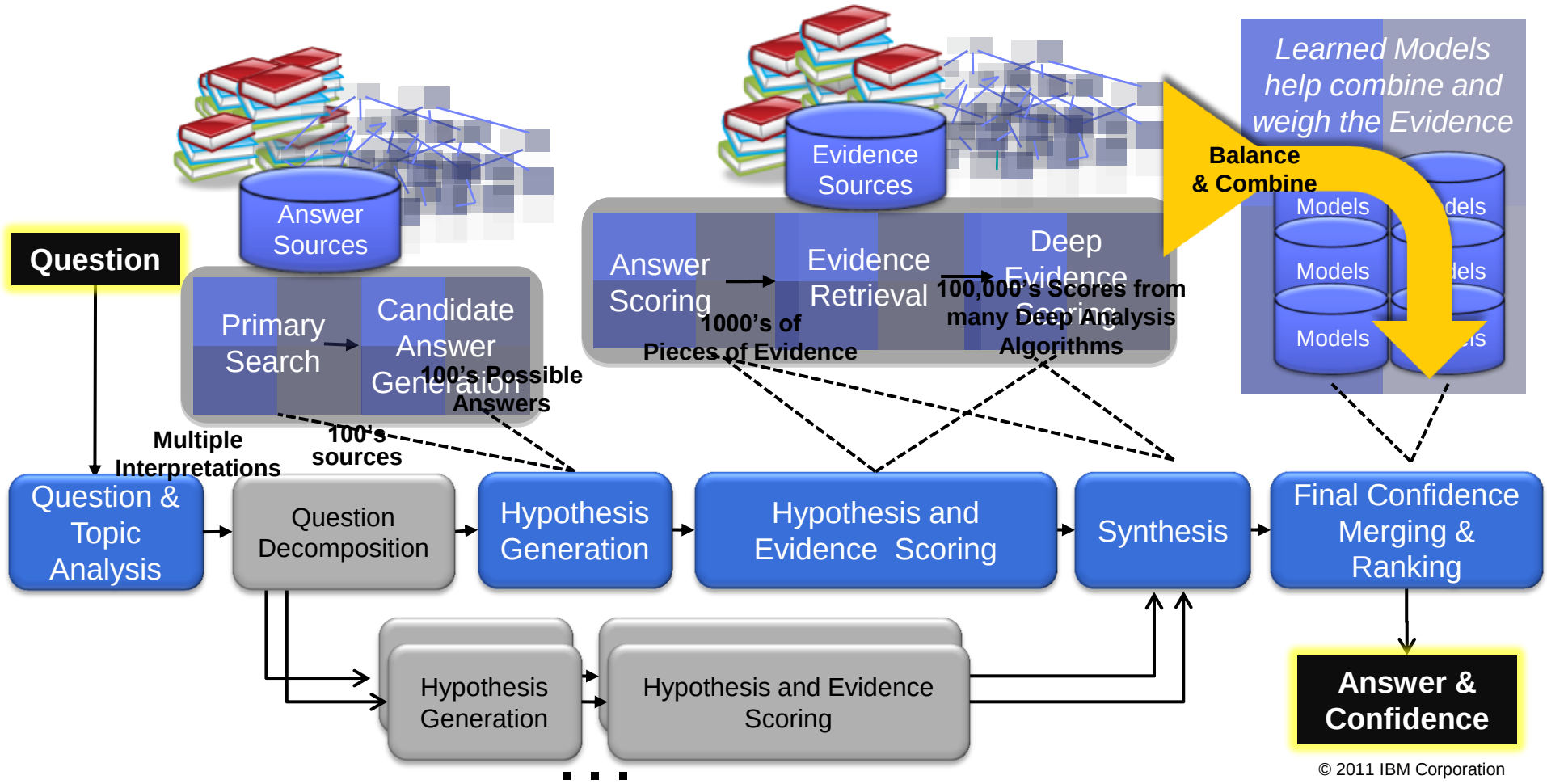
Generate candidate answers and check them one by one to choose the best.

Generate: Use document referents *(e.g. Wikipedia titles) as candidate answers. Use search to find the most relevant documents.

Evaluate: Automatically score each candidate answer along 400+ dimensions. Use machine learning to find how to best combine these partial scores.

*Note: Every document is about something

Watson Architecture



Example Question

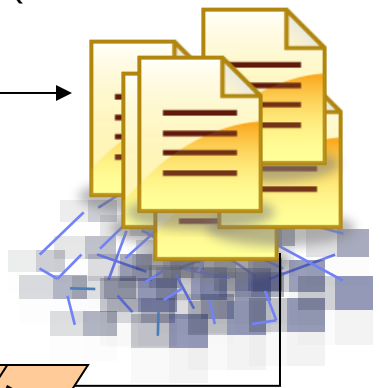
In 1894 C.W. Post created his warm cereal drink Postum in this Michigan city

Question Analysis

Keywords: 1894, C.W. Post, created …
Lexical AnswerType: (Michigan city)
Date(1894)
Relations:
Create(Post, cereal drink)
…

Primary Search

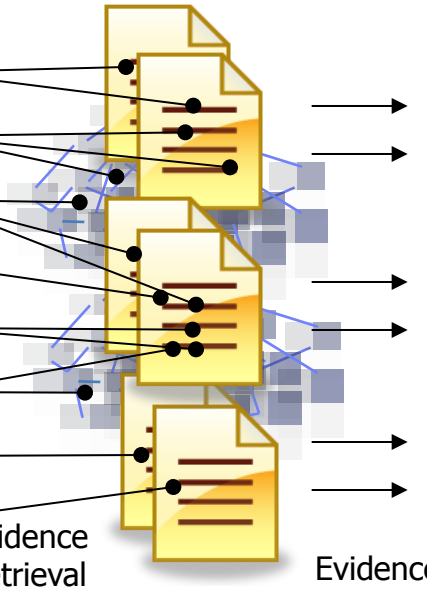
Related Content (Structured & Unstructured)



Candidate Answer Generation

General Foods
1895
Post Foods
Battle Creek
Grand Rapids
…
…
…

Evidence Retrieval



Evidence Scoring

Lexical | Taxonomic | Spatial | … | Temporal

```
[0.58  0  -1.3 … 0.97]
[0.71  1  13.4 … 0.72]
[0.12  0   2.0 … 0.40]
[0.84  1  10.6 … 0.21]
[0.33  0   6.3 … 0.83]
[0.21  1  11.1 … 0.92]
[0.91  0  -8.2 … 0.61]
[0.91  0  -1.7 … 0.60]
```

Merging & Ranking

1) Battle Creek (0.85)
2) Post Foods (0.20)
3) 1895 (0.05)

Watson's conceptual innovations

- Dialogue-like view of meaning: *generate candidate interpretations, find supporting evidence, and evaluate it*
- Aboutness of knowledge – *source engineering* to convert knowledge into a title-oriented format
- Contextual meaning derivation – *deferred type evaluation* to make type matching based on all available evidence

Aboutness of knowledge

<DOC>

<DOCNO>Shakespeare211</DOCNO>

<TITLE>Ophelia</TITLE>

<TEXT>But, good my brother,
Do not, as some ungracious pastors do,
Show me the steep and thorny way to heaven,
Whiles, like a puff' d and reckless libertine,
Himself the primrose path of dalliance treads,
And reaks not his own rede.</TEXT>

</DOC>

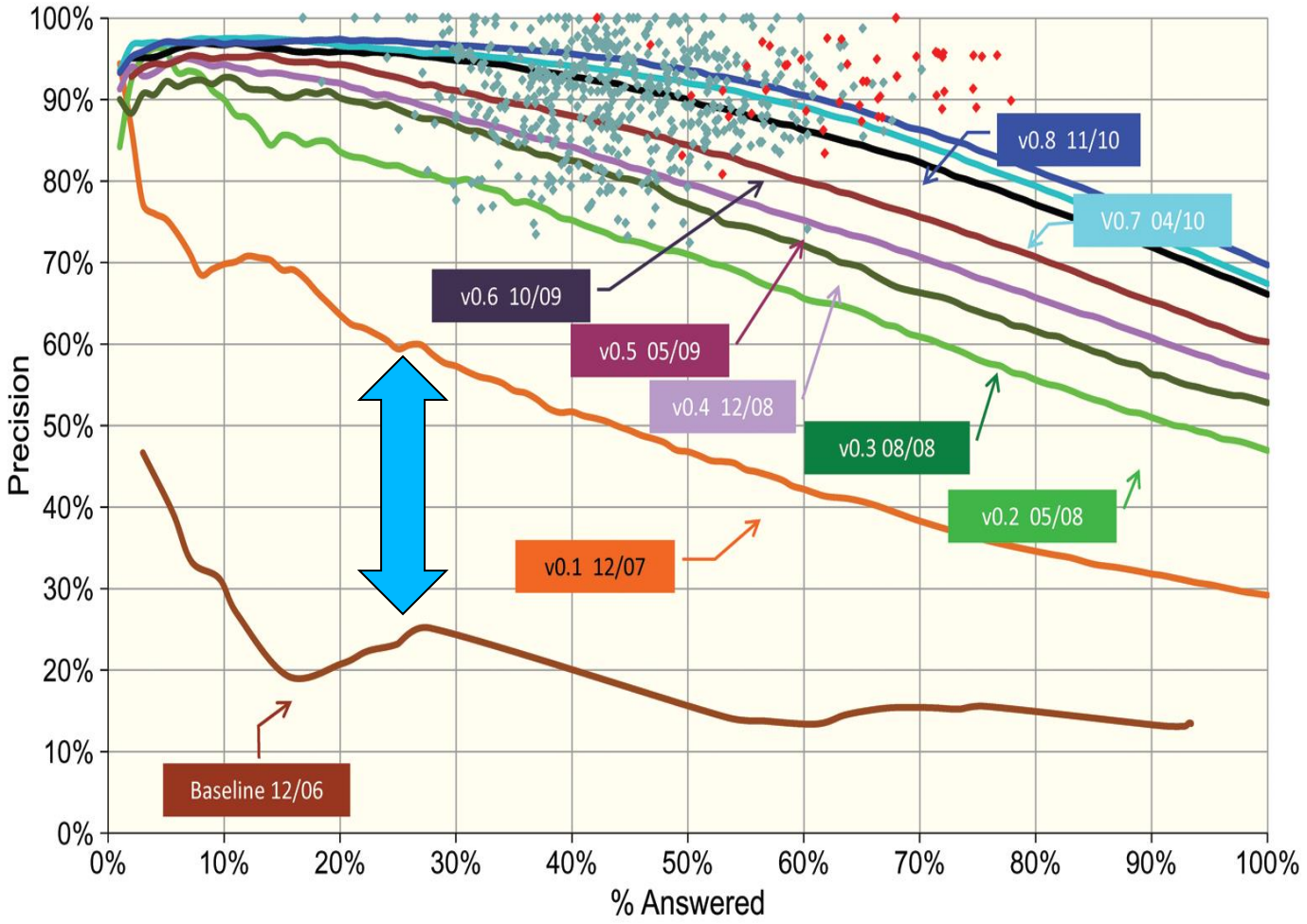
Aboutness of knowledge

- A document is about something
- The same text might be about multiple entities (but not about all the entities in the text). E.g. Ophelia, Shakespeare, Hamlet.
- Some sources are already title oriented (encyclopedias, dictionaries), but needed to be cleaned (to improve candidate generation performance). Other needed to be transformed into the title-oriented format. E.g. “important books”, song lyrics, poetry, …

Source selection, preparation, expansion, testing…

- Prior research showed that adding more text might be harmful, so care was needed
- “Raw” text in TREC format used in search
- Text processed syntactically and semantically used in supporting passage retrieval
- (Source processing also essential when answers are not titles. For example, in help desk support.)

Initial jump: Aboutness of knowledge



Watson's conceptual innovations

- Dialogue-like view of meaning: *generate candidate interpretations, find supporting evidence, and evaluate it*
- Aboutness of knowledge – *source engineering* to convert knowledge into a title-oriented format
- Contextual meaning derivation – *deferred type evaluation* to make type matching based on all available evidence

Deferred type evaluation

Meaning representation is an unsolved problem.

Let many scorers focus on aspects of meanings to see how well a candidate matches relations in the question.

With a slight exaggeration:

Passages have no meanings; they only provide candidate scores!

Example

"This Polish football player was recently promised to Bayern Munich"

Robert Lewandowski

Example -- Problem

Simplest evidence: Google search for

"polish football player" " promised to Bayern Munich“

yields 2 results both mention Lewandowski

However

lewandowski "polish football player“ 160K results

lewandowski "german football player“ >1M results

Problem: The evidence for the correct answer is weak.

Example -- Solution

Both types and a default co-exist

lewandowski IS-A *"german football player“*

lewandowski IS-A *"polish football player“*

■* $Polish(x) \neq German(x)$

Measurement (answering the question) produces the types, as one of the pieces of evidence.

(Do these co-exist in some kind of superposition?)

Deferred type evaluation

Compute full meaning – referent -- after all evidence is provided.

In the J! system, it meant finding the full meaning of a query in several steps:

- Searching for documents mentioning entities in the query, getting their referents
- Finding a "type" (i.e. a descriptor) of a candidate answer only when all information about all relevant entities is available.
- the "type" is not required as part of a predetermined ontology but is only a lexical/grammatical item.

Discussion

I'm not aware of any formal treatment of meaning postulating all three:

1. Interactivity or superposition of properties
2. Aboutness of all knowledge
3. Deferred type evaluation

(1) matches with physics and physiology (e.g. our vision)

(2) is philosophically plausible

(3) makes computational sense and has appeared in computer science before in different context

Can we make interesting (predictive) mathematics out of it?

Can we make interesting (predictive) mathematics out of it?

Can we connect it with other interesting and unsolved problems like granularity and transitions between discrete and continuous?

Thank you!

Impact of source engineering

	Wikipedia Only Baseline	**Source Acquisition + Transformation**	**Source Acquisition + Transformation + Expansion**
Accuracy	59.0%	66.7% (+7.7%)	70.4% (+3.7%)
Precision@70	77.0%	84.3% (+7.3%)	87.2% (+2.9%)
Candidate Binary Recall	77.1%	84.7% (+7.6%)	87.5% (+2.8%)
# of Documents	3.5 M	7.6 M	8.6 M
Corpus Size	13 GB	25 GB	59 GB