Building *Watson*

An overview of the DeepQA Project

David Ferrucci

Watson Principal Investigator and IBM Fellow

DeepQA Team @ IBM Research

© 2010 IBM Corporation

A Grand Challenge Opportunity

- **Drive Important Scientific Advances**
  - Envision new ways for computers to impact society & science
- **Be Relevant to IBM Customers**
  - Enable better, faster decision making over unstructured and structured content
  - *Business Intelligence, Knowledge Discovery and Management, Government, Compliance, Publishing, Legal, Healthcare, Product Support, etc.*
- **Capture the Broader Imagination**
  - The Next *Deep Blue*

© 2010 IBM Corporation

Want to Play Chess or Just Chat?

- Chess
  - A finite, mathematically well-defined search space
  - Limited number of moves and states
  - All the symbols are completely grounded in the mathematical rules of the game

- Human Language
  - Words by themselves have no meaning
  - Only grounded in **human cognition**
  - Words navigate, align and communicate an infinite space of intended meaning
  - Computers can **not** ground words to human experiences to derive meaning

© 2010 IBM Corporation

Easy Questions?

=(LN(12,546,798 *π)) ^ 3 / 34,600.47 = **0.155**

Select *Payment* where *Owner*="David Jones" and *Type(Product)*="Laptop",

Owner	Serial Number
David Jones	45322190-AK

Serial Number	Type	Invoice #
45322190-AK	LapTop	INV10895

Invoice #	Vendor	Payment
INV10895	MyBuy	$104.56

David Jones
↓↓↓↓↓↓↓↓↓↓
David Jones

=

Dave Jones
↓↓↓↓ ↓↓↓↓↓
David Jones

≠

© 2010 IBM Corporation

Hard Questions?

Computer programs are natively **explicit, fast** and **exacting** in their calculation over numbers and symbols….But **Natural Language** is implicit, highly contextual, ambiguous and often imprecise.

Person	Birth Place
A. Einstein	ULM

Structured

Unstructured

- Where was X born?

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.

Person	Organization
J. Welch	GE

- X ran this?

If leadership is an art then surely Jack Welch has proved himself a master painter during his tenure at GE.

© 2010 IBM Corporation

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.

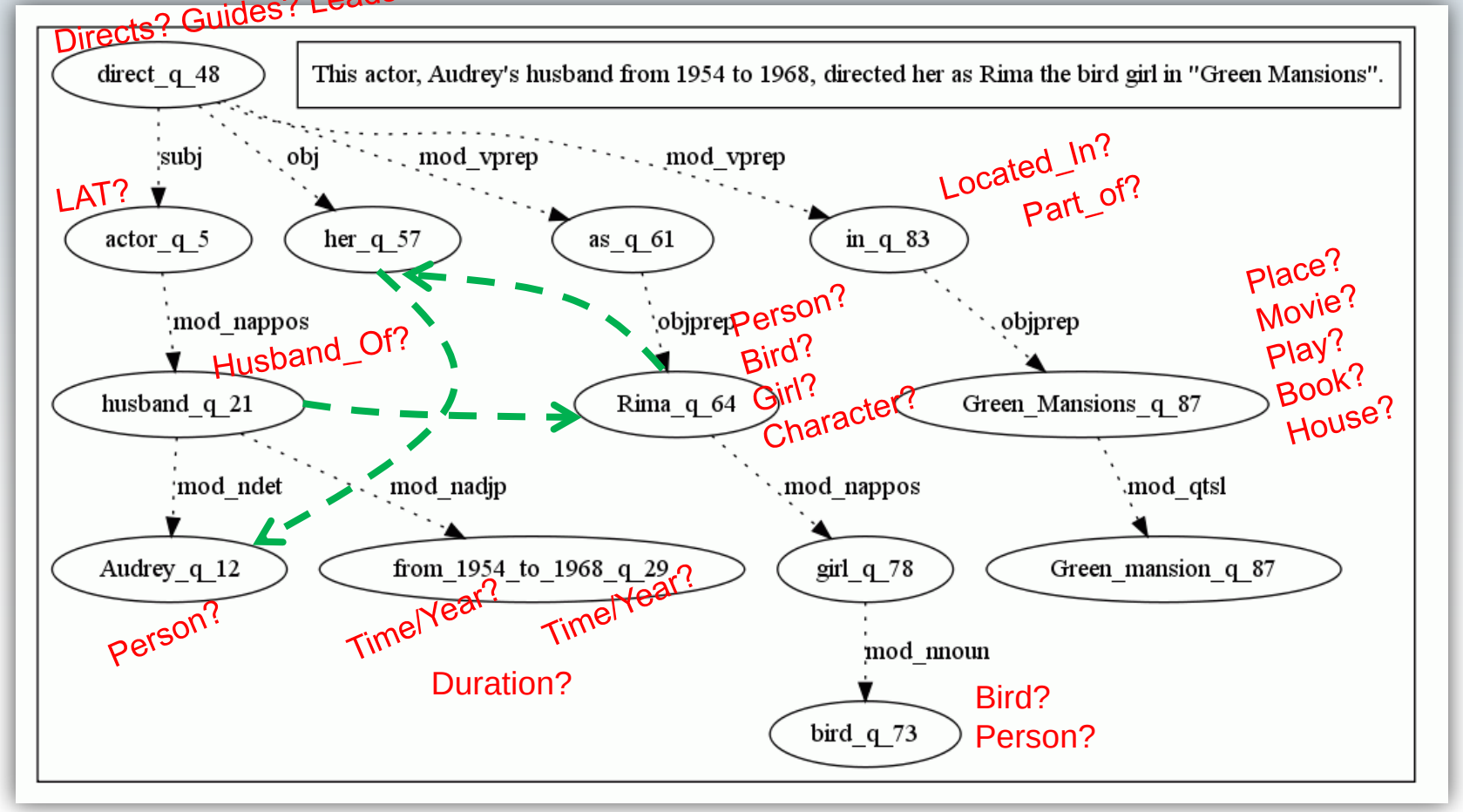
$600
In cell division, mitosis splits the nucleus & cytokinesis splits this liquid *cushioning* the nucleus

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

© 2010 IBM Corporation

The Possibilities Multiply

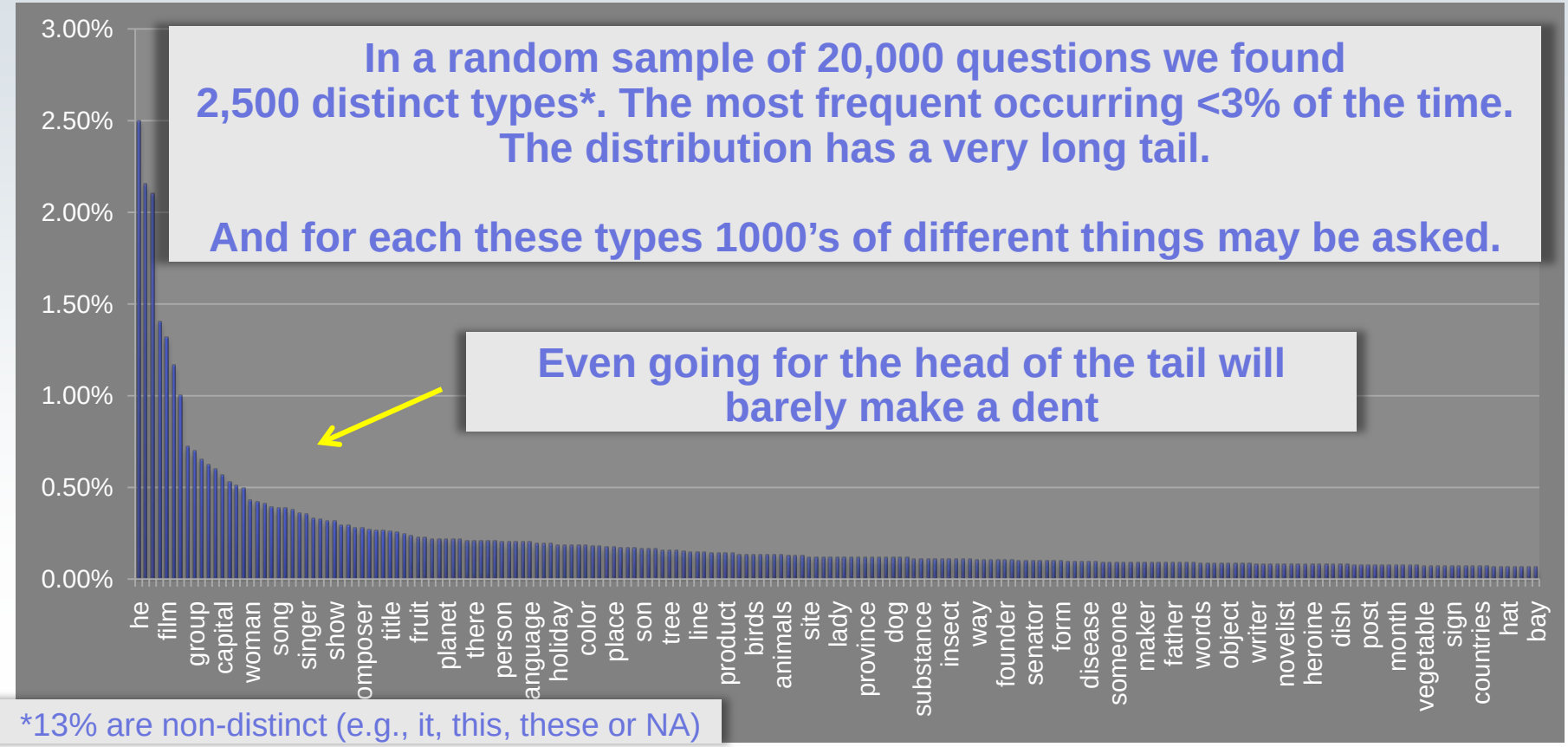
This actor, Audrey's husband from 1954 to 1968, directed her as Rima the bird girl in "Green Mansions"



© 2010 IBM Corporation

Broad Domain

We do NOT attempt to anticipate all questions and build specialized databases.

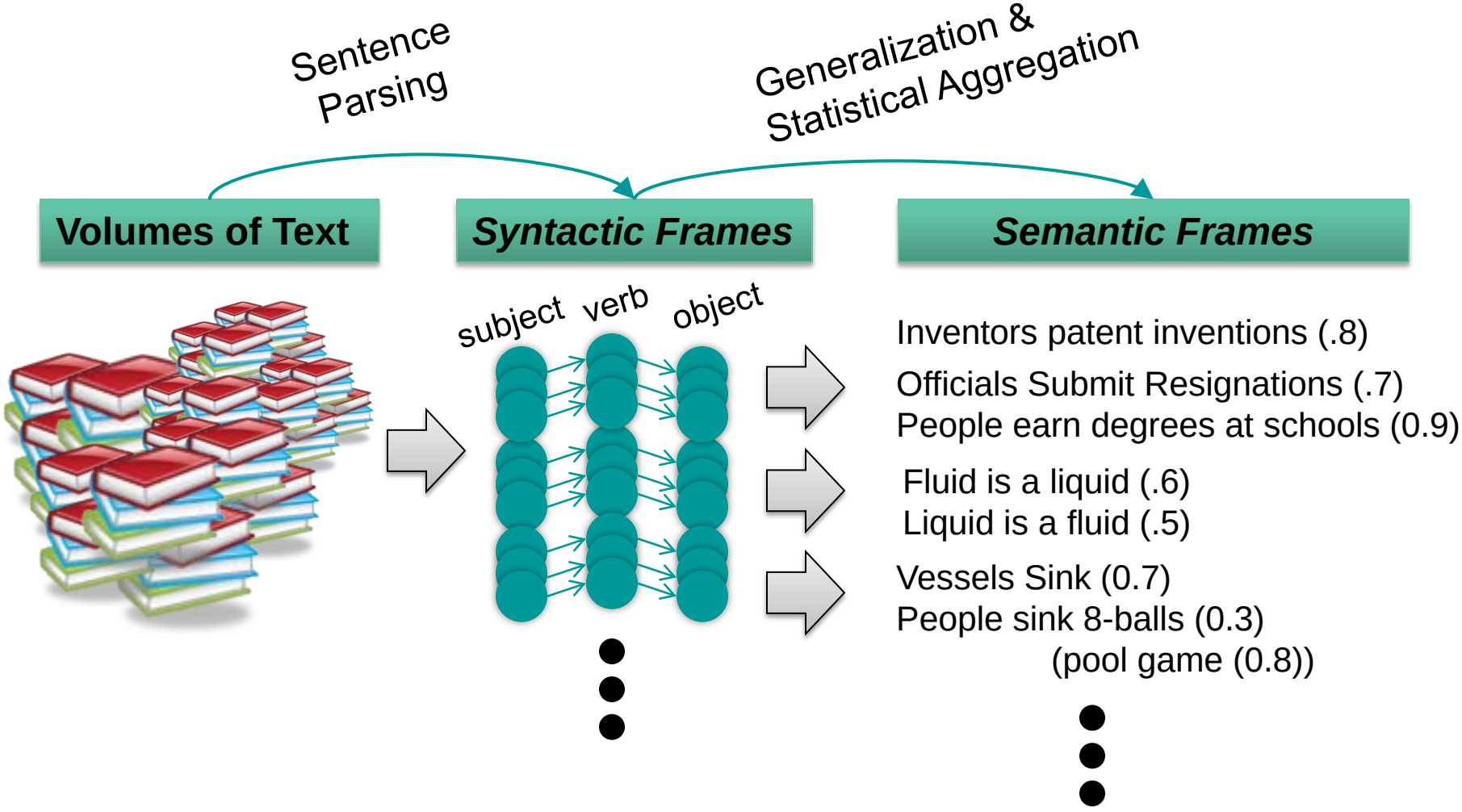


*13% are non-distinct (e.g., it, this, these or NA)

Our Focus is on reusable NLP technology for analyzing volumes of *as-is* text. Structured sources (DBs and KBs) are used to help interpret the text.

© 2010 IBM Corporation

Inducing Frames



© 2010 IBM Corporation

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

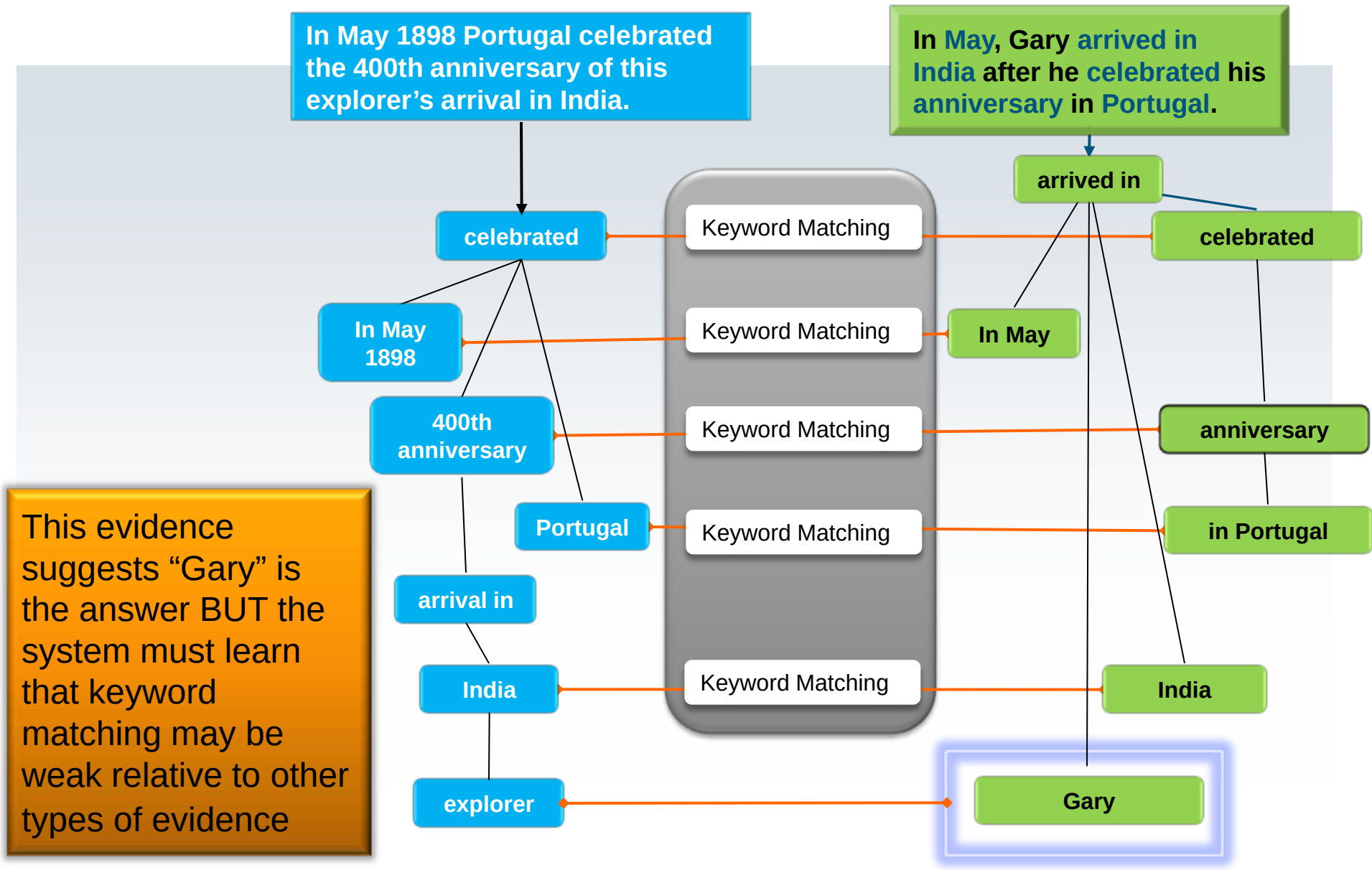
"Cytoplasm is a fluid *surrounding* the nucleus…"

Wordnet → Is_a(Fluid, Liquid) → ?

Learned → Is_a(Fluid, Liquid) → yes.

© 2010 IBM Corporation

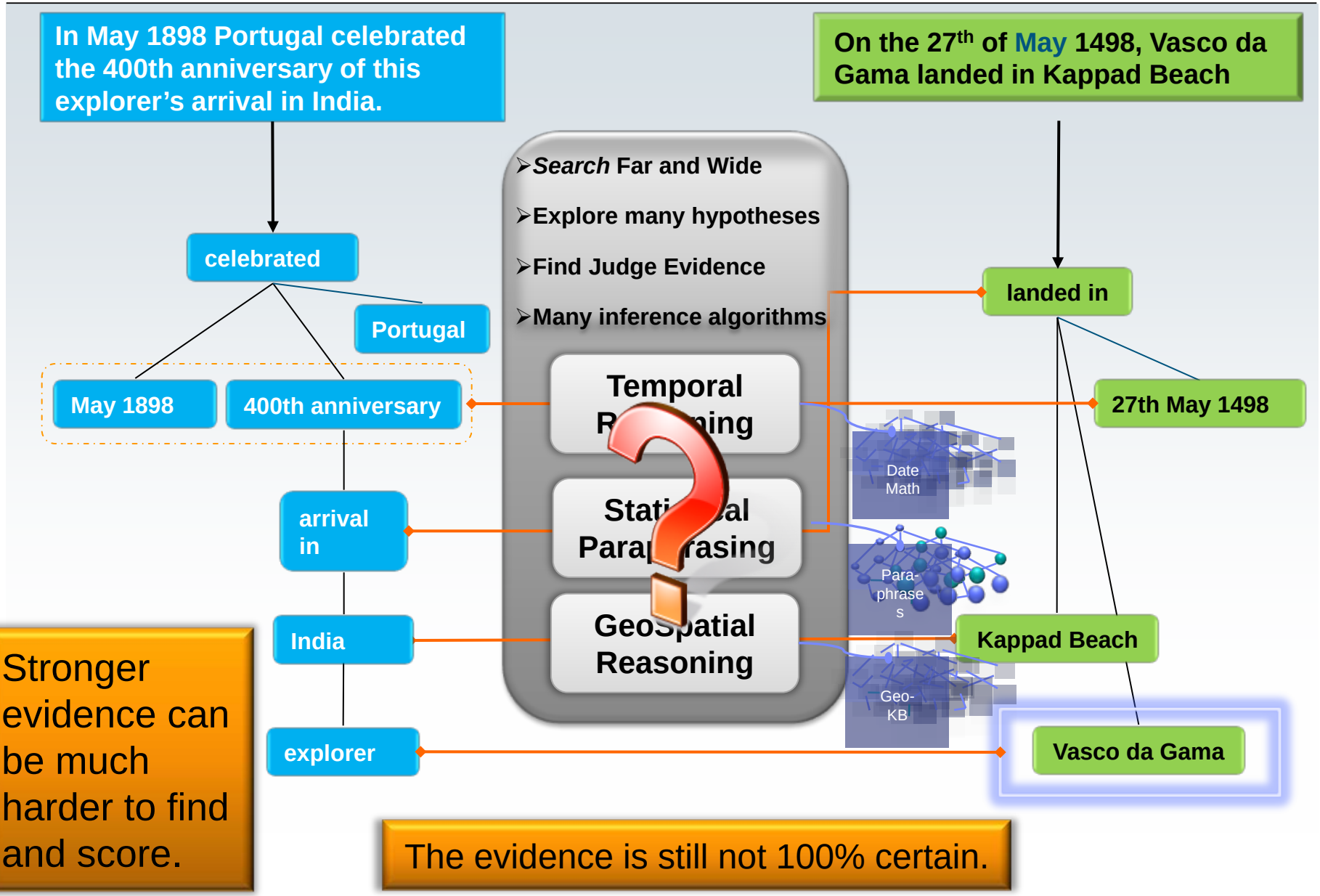
Keyword Evidence



This evidence suggests "Gary" is the answer BUT the system must learn that keyword matching may be weak relative to other types of evidence

© 2010 IBM Corporation

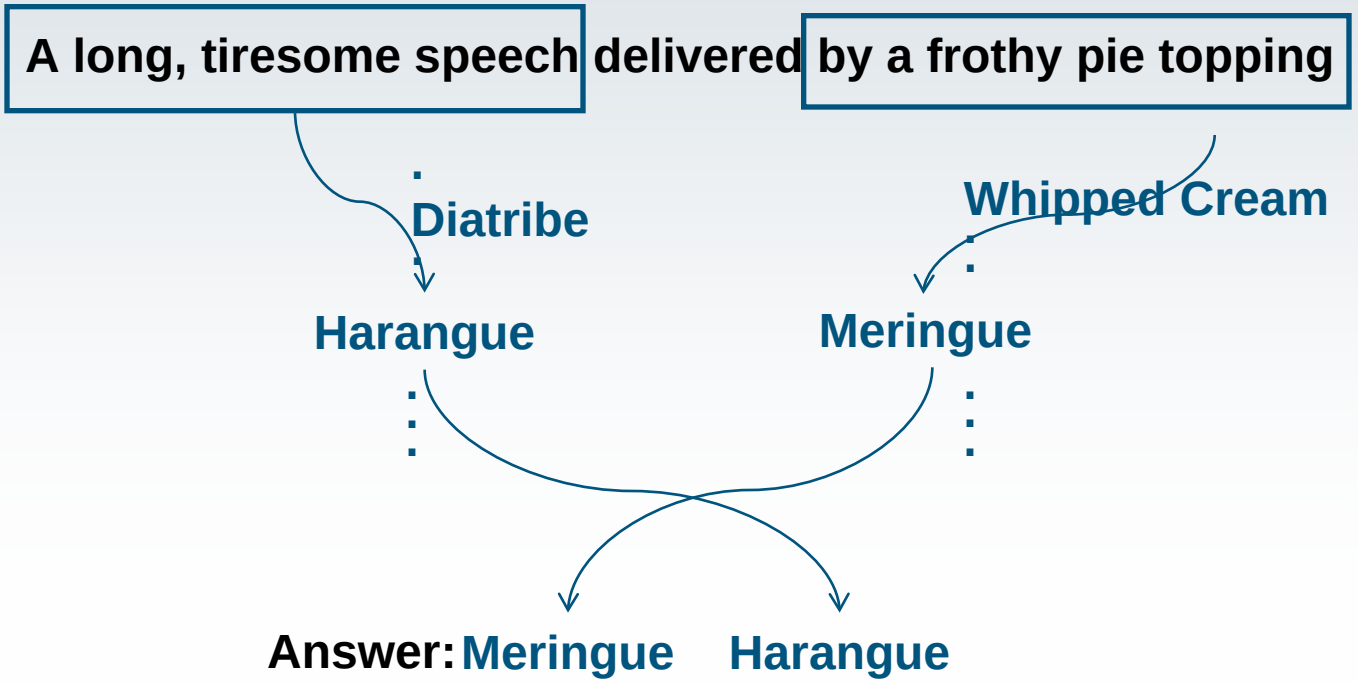
Deeper Evidence



© 2010 IBM Corporation

Not ***Just*** for Fun

Some Questions require Decomposition and Synthesis

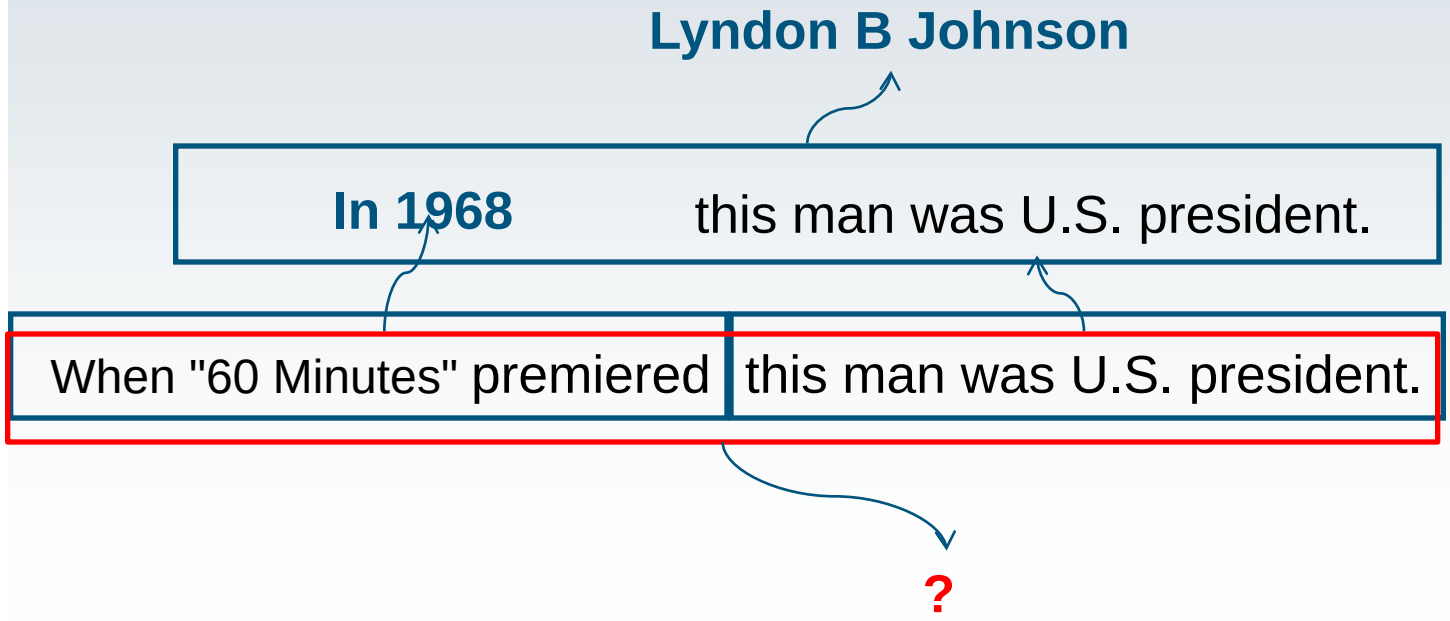


Category: Edible Rhyme Time

© 2010 IBM Corporation

Divide and Conquer
(Typical in Final Jeopardy!)

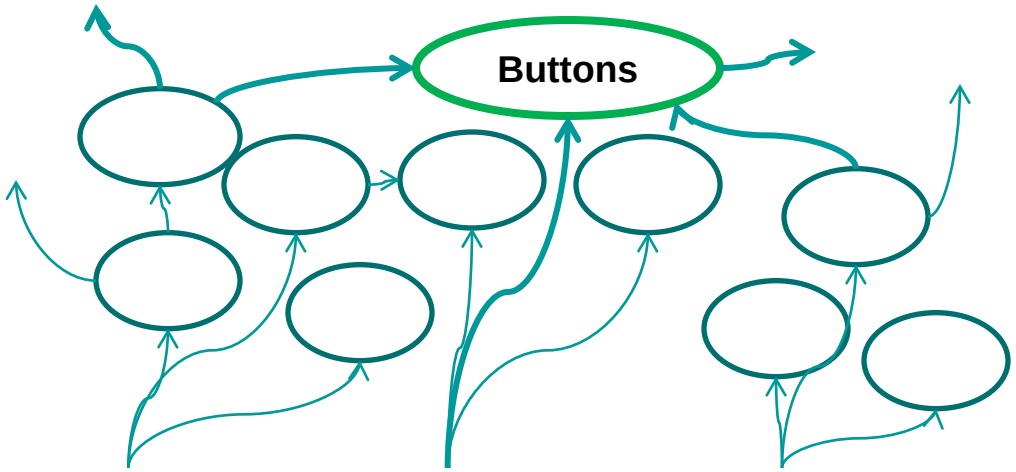
Must identify and solve sub-questions from different sources to answer the top level question

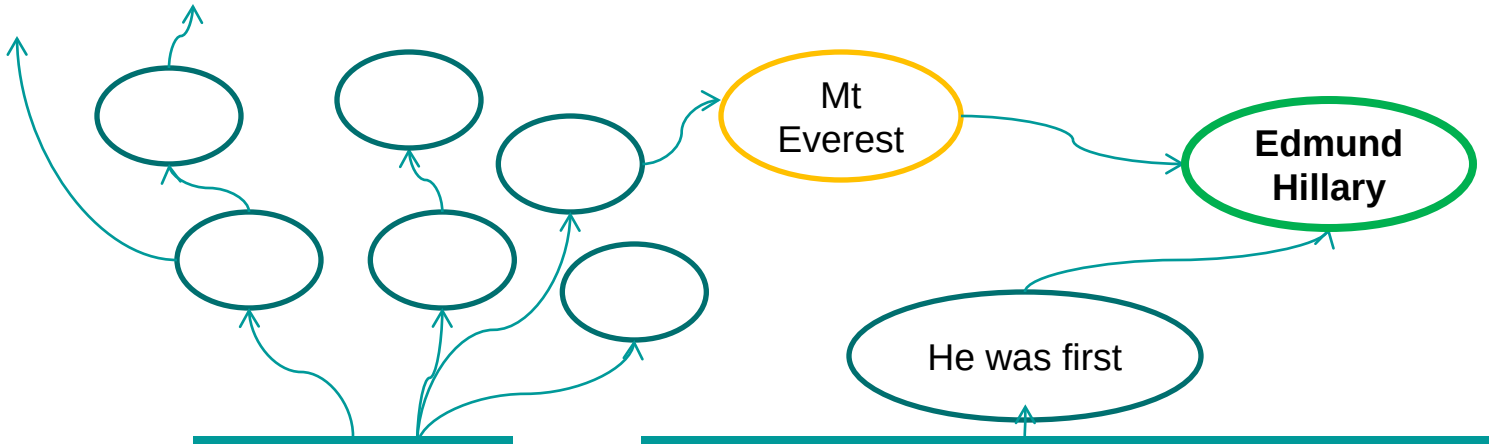


The DeepQA architecture attempts different *decompositions* and recursively applies the QA algorithms

© 2010 IBM Corporation

The Missing Link





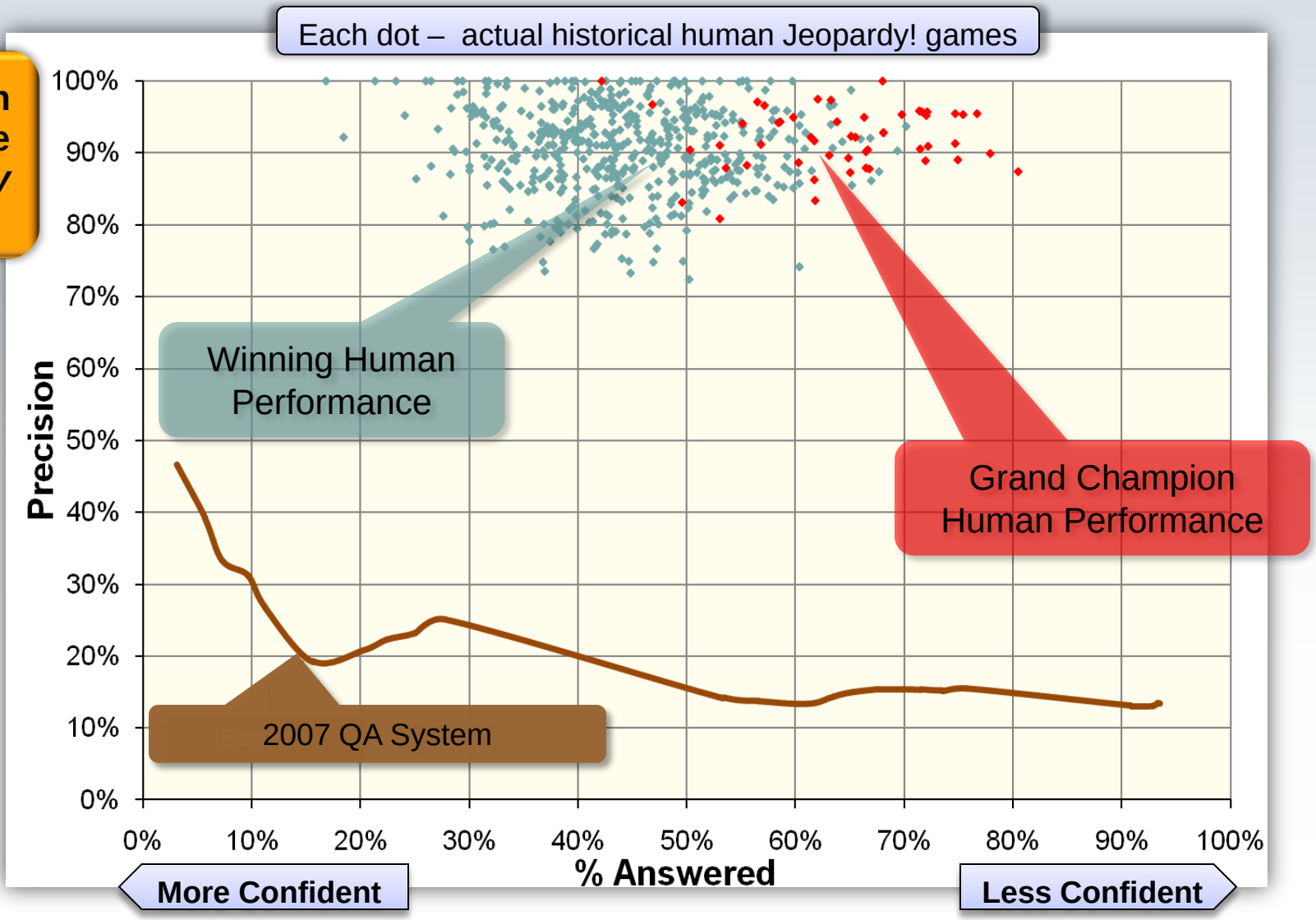
On hearing of the discovery of George Mallory's body, he told reporters he still thinks he was first.

© 2010 IBM Corporation

What It Takes to compete against Top Human Jeopardy! Players

*Our Analysis Reveals the **Winner's Cloud***

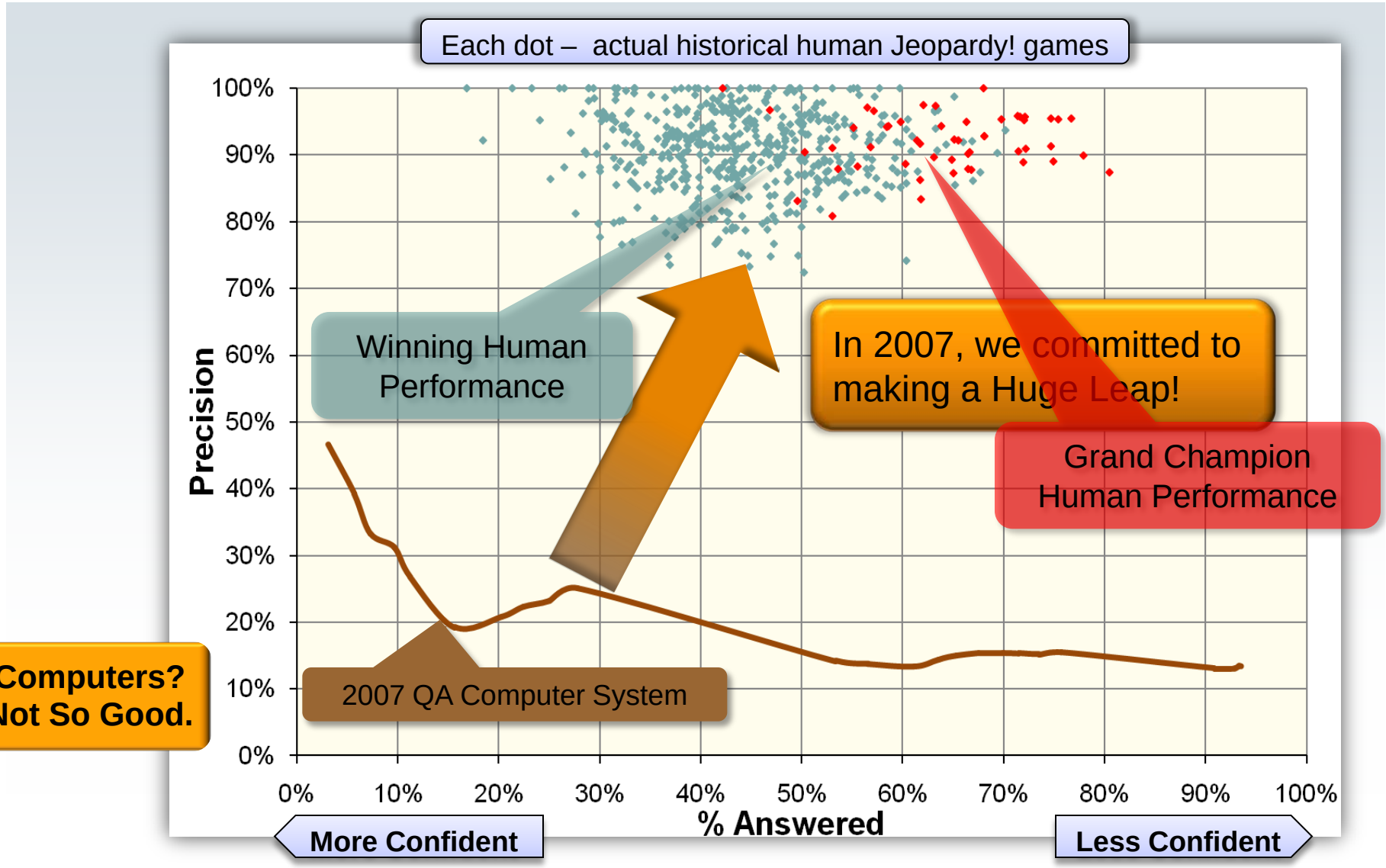
Top human players are *remarkably* good.



© 2010 IBM Corporation

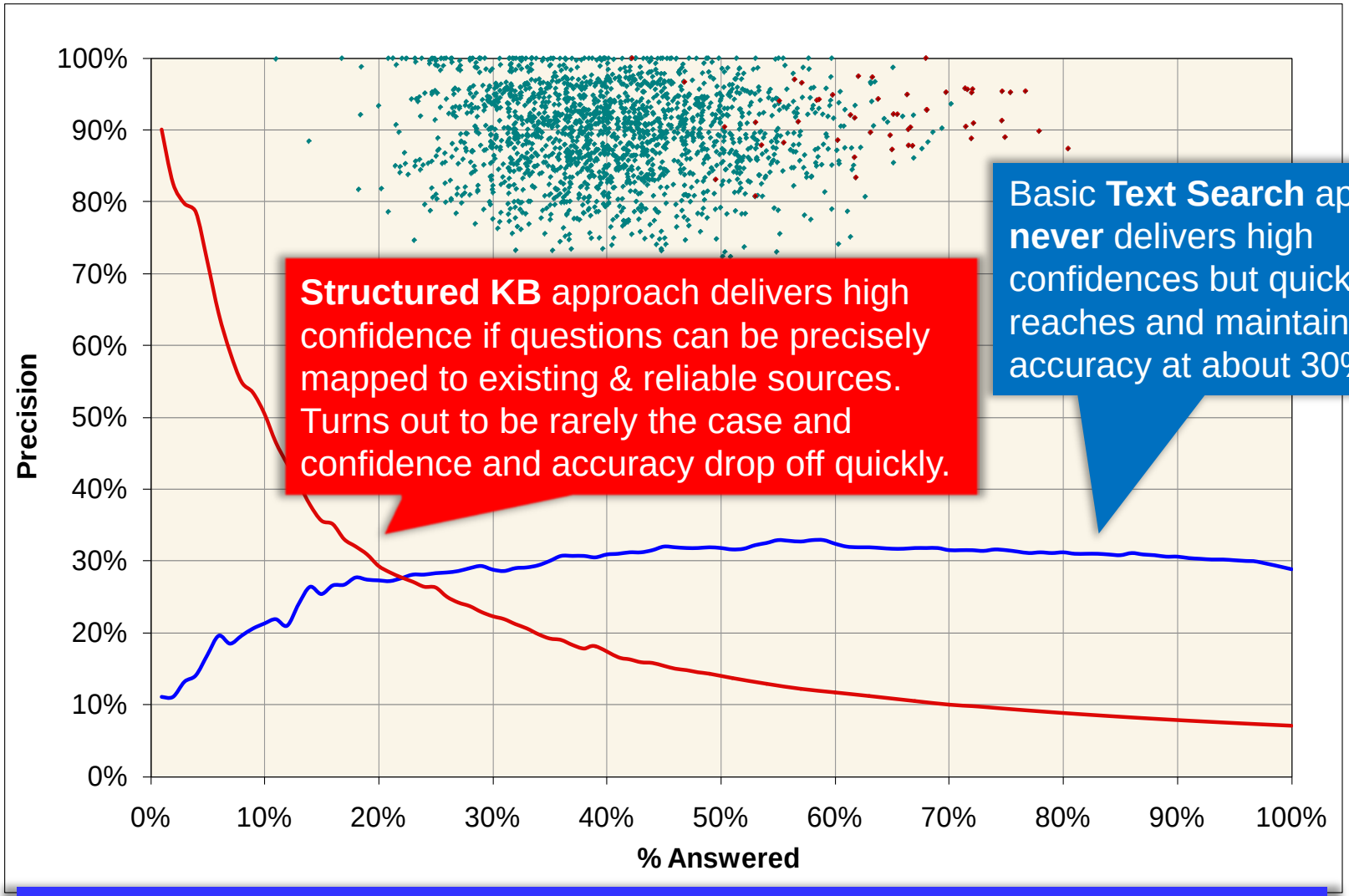
What It Takes to compete against Top Human Jeopardy! Players

*Our Analysis Reveals the **Winner's Cloud***



© 2010 IBM Corporation

Early Experiments on Focused Content



Must combine deep and shallow semantic analysis over structured & unstructured content to drive up precision , recall and confidence.

© 2010 IBM Corporation

A Few Guiding Principles

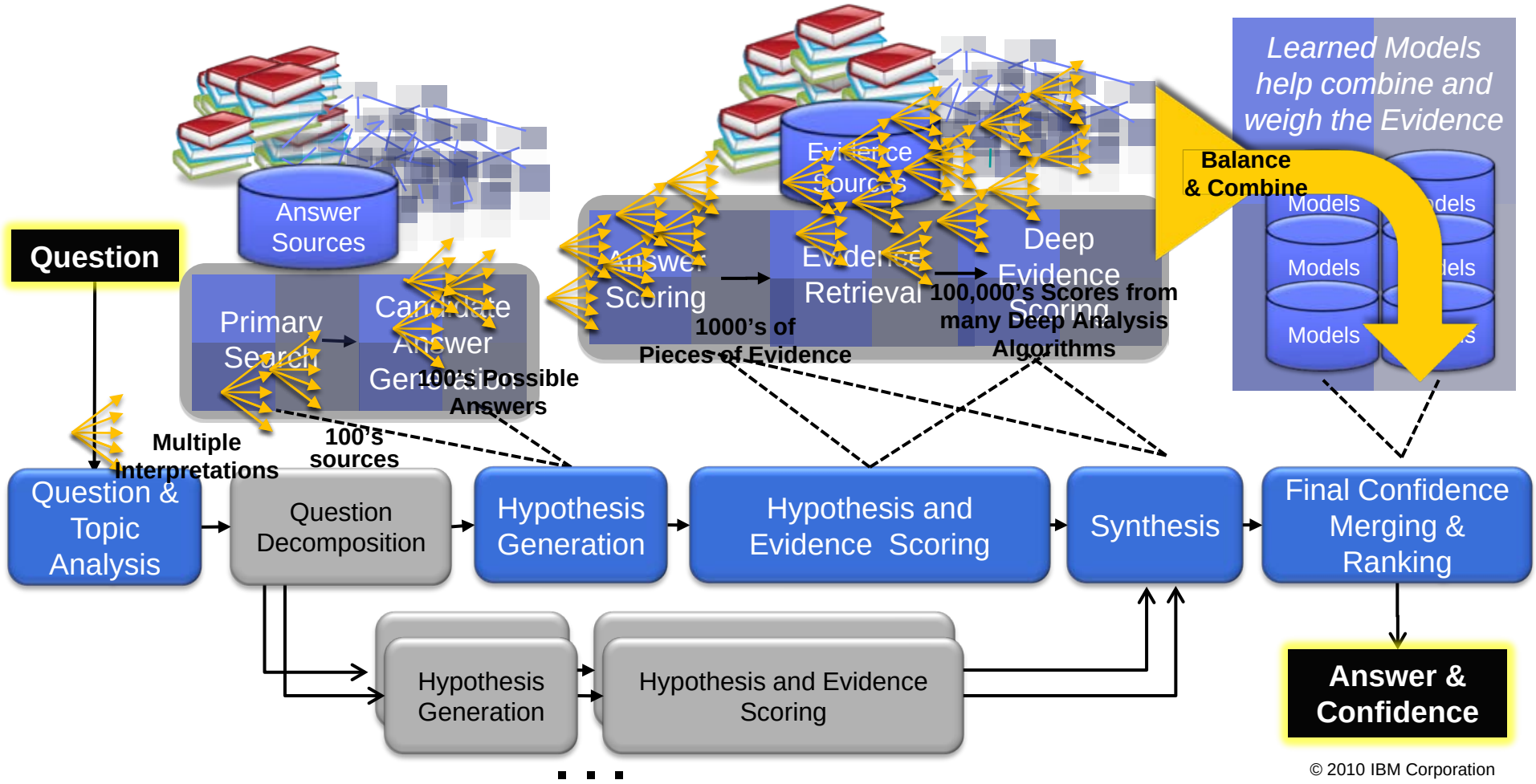
- Specific Large Hand-Crafted Models Won't Cut It
  - *Too Slow, Too Narrow, Too brittle, Too Bias*
  - *Need to acquire and analyze information from **As-Is Knowledge sources***
- Intelligence from many diverse methods
  - ***Many diverse algorithms must be combined.** No single one expected to solve the whole problem. Each addressing different weaknesses.*
  - *Relative impact of many overlapping methods must be **learned***
- *Massive Parallelism is a Key Enabler*
  - *Pursue many competing independent hypotheses over large data*
  - *Efficiency will demand simultaneous threads of evidence evaluation*

© 2010 IBM Corporation

DeepQA: The Technology Behind *Watson*

Massively Parallel Probabilistic Evidence-Based Architecture

*Generates and scores many hypotheses using a combination of 100's of **Natural Language Processing**, **Information Retrieval**, **Machine Learning** and **Reasoning Algorithms.** These gather, evaluate, weigh and balance different types of **evidence** to deliver the answer with the best support it can find.*



© 2010 IBM Corporation

A METHODOLOGY FOR RAPID ADVANCEMENT

© 2010 IBM Corporation

Methodology

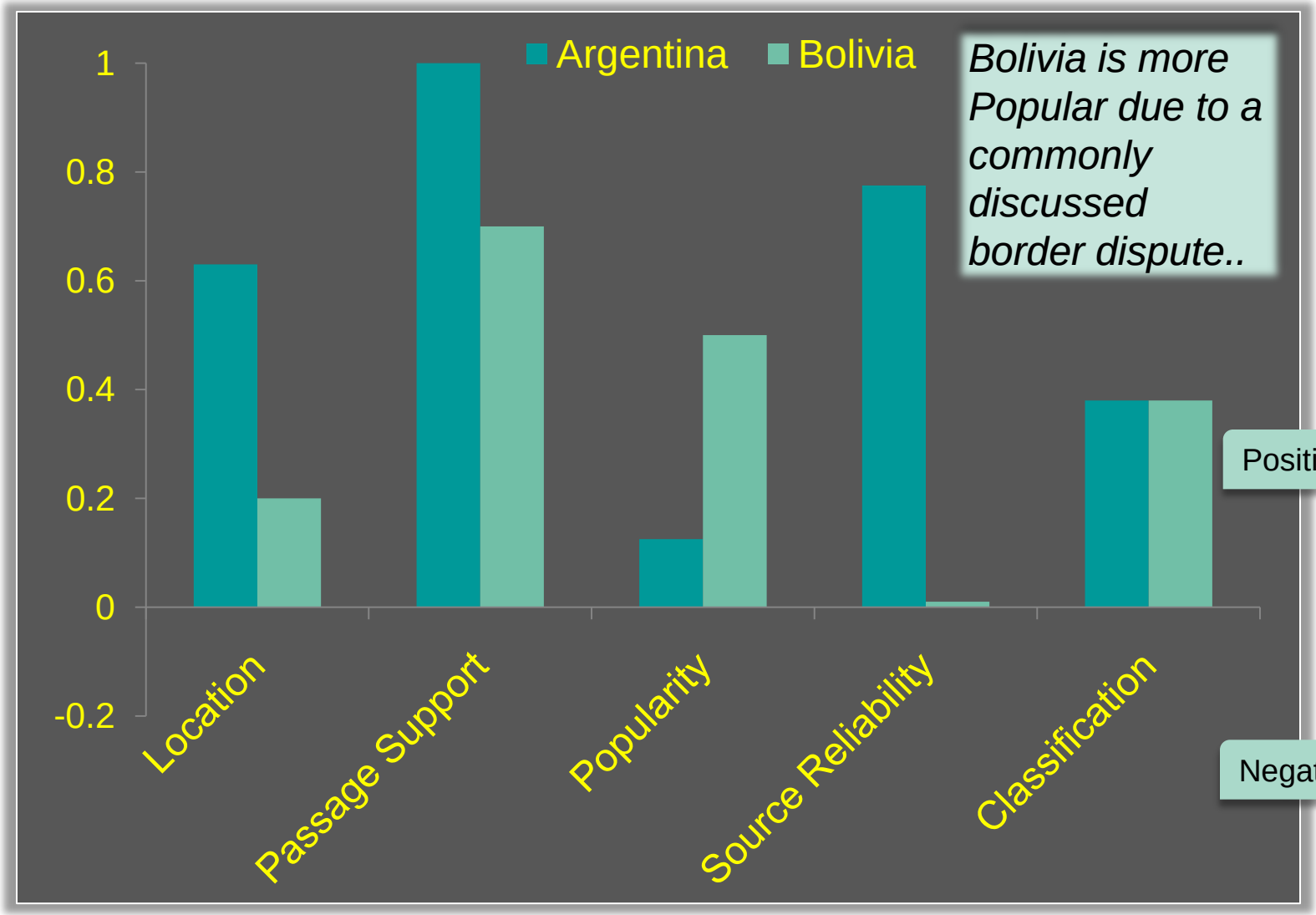
- Goal-Oriented System-Level Metrics and Longer-Term Incentives
  - Headroom Analysis: Estimate potential impact on System-Level metrics before “go”.
  - End-to-end evaluation vs. Isolated Component Metrics
  - 3-5 Year Big Team Result First vs. Incremental, Isolated Publications
  - Agile – Shift resources to project getting results
- Extreme Collaboration
  - Implemented “One Room” to optimize team work and communication
  - Immediate access to the right “expert”, spontaneous discussions
  - Real-time Documentation of ideas/status on team wiki
- Diverse Skills ALL IN One Room
  - Natural Language Processing, Information Retrieval, Knowledge Representation and Reasoning, Machine Learning, Computational Linguistics, Intelligent Systems Architecture, Software Engineering.
- Disciplined Engineering and Evaluation
  - Bi-weekly End-to-End Integration Runs & Evaluations (Large Compute Resources)
  - >10 GBs of error analysis output made accessible via Web-Based Tool
  - Feature Analysis Tools For Deeper Error Analysis

>8000 Documented experiments performed in 4 years

© 2010 IBM Corporation

Grouping Features to produce ***Evidence Profiles***

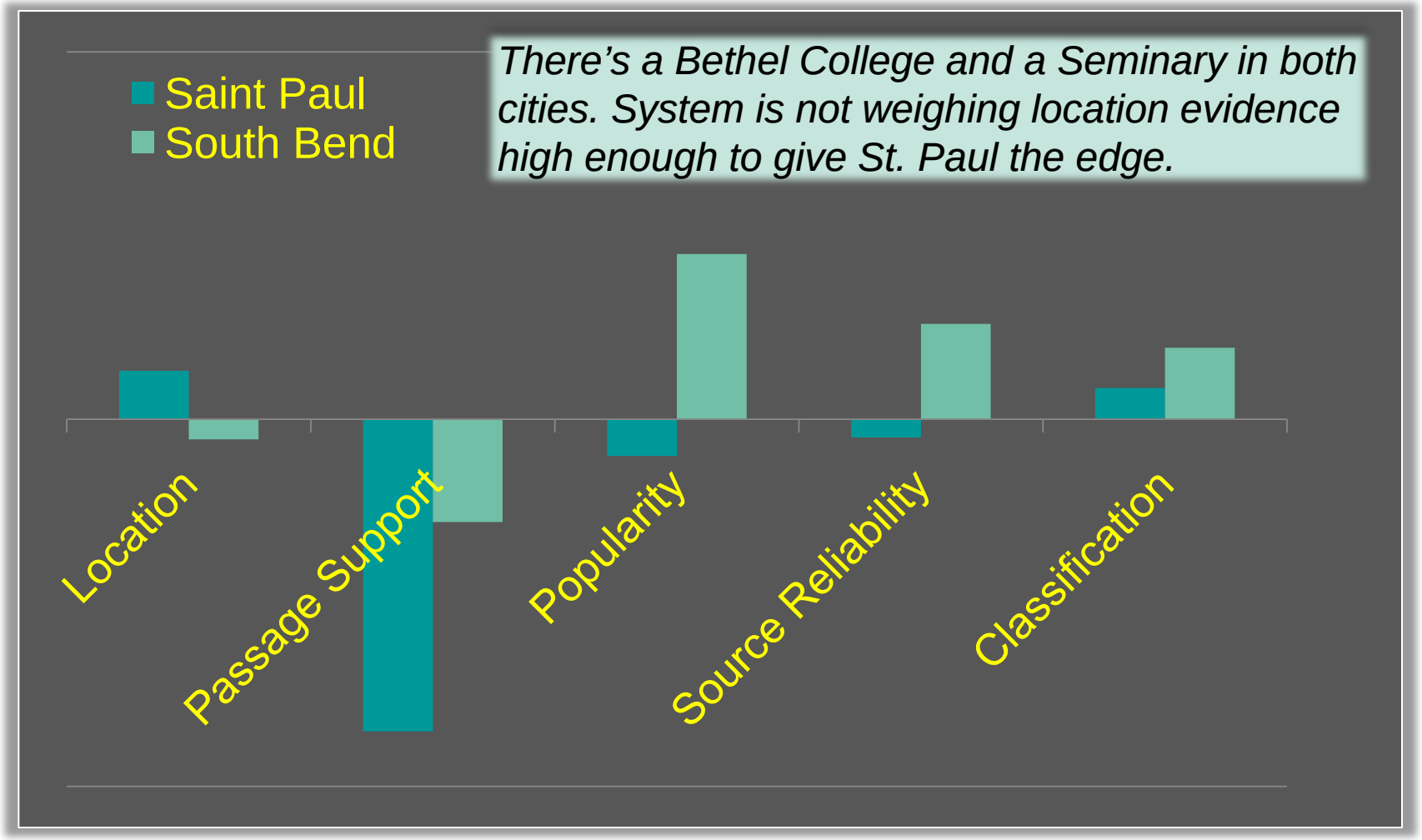
Clue: Chile shares its longest land border with this country.



© 2010 IBM Corporation

Evidence: Time, Popularity, Source, Classification etc.

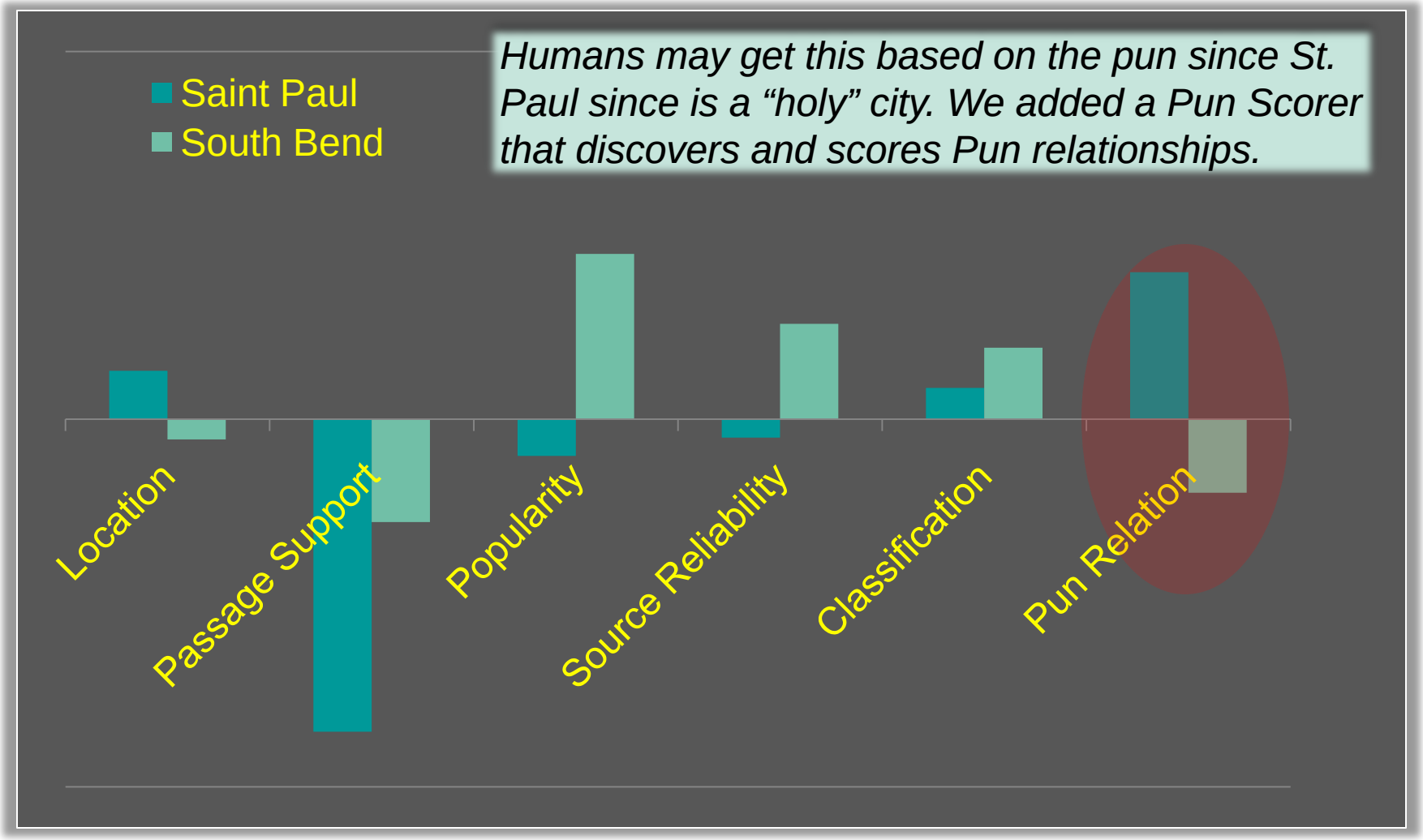
Clue: You'll find Bethel College and a Seminary in this "holy" Minnesota city.



© 2010 IBM Corporation

Evidence: Puns

Clue: You'll find Bethel College and a Seminary in this "holy" Minnesota city.



© 2010 IBM Corporation

In-Category Learning

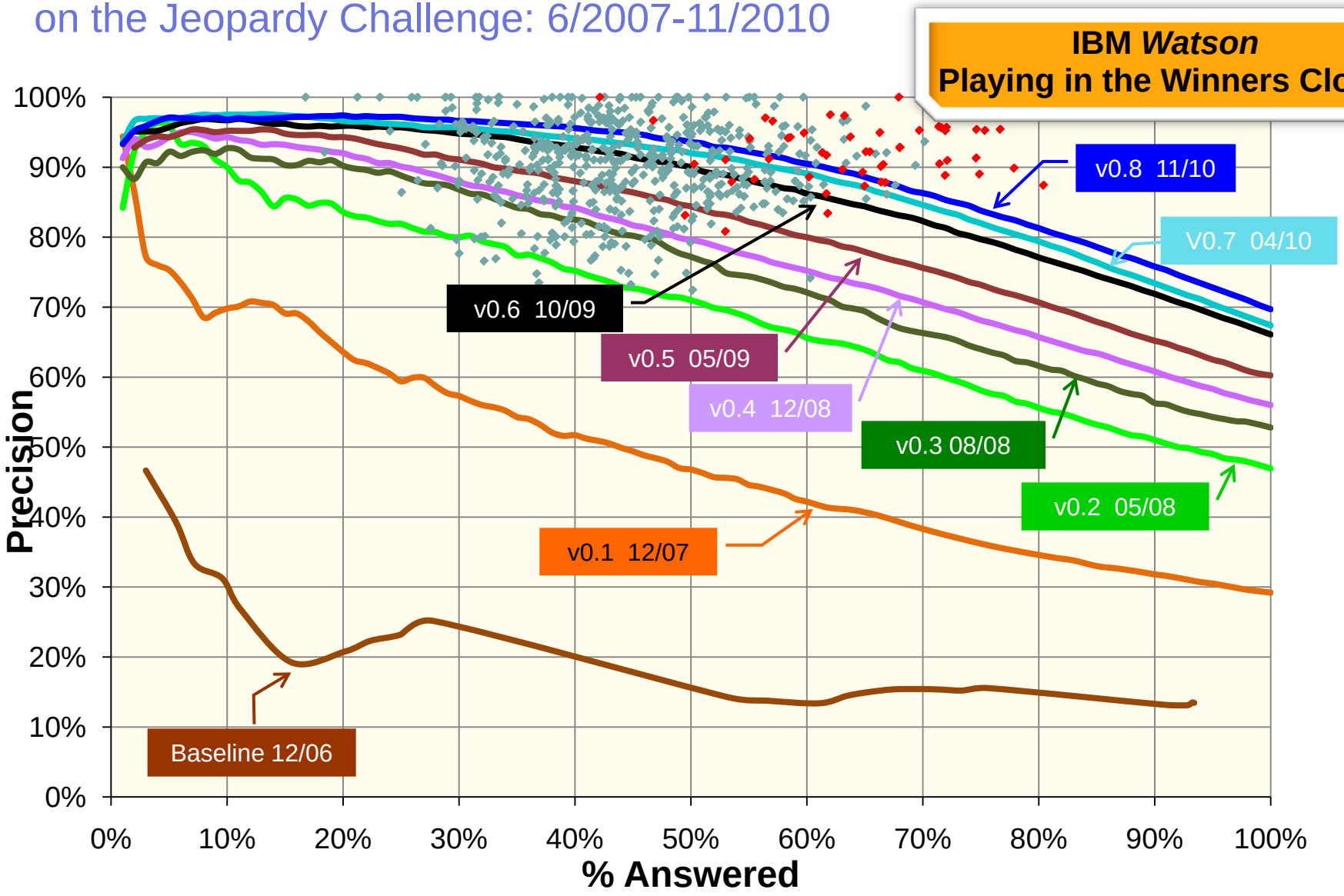
CELEBRATIONS OF THE MONTH

- What the Jeopardy! Clue is asking for is NOT always obvious
- Watson tries infer the type of thing being asked for from the previous answers.
- In this example after seeing correct answers, Watson starts to dynamically learn, using Bayesian inference, the probability that the answer type is a "**month**".

Clue	Type	Watson's Answer	Correct Answer
D-DAY ANNIVERSARY & MAGNA CARTA DAY	day	Runnymede	**June**
NATIONAL PHILANTHROPY DAY & ALL SOULS' DAY	day	Day of the Dead	November
NATIONAL TEACHER DAY & KENTUCKY DERBY DAY	Day/month(.2)	Churchill Downs	May
ADMINISTRATIVE PROFESSIONALS DAY & NATIONAL CPAS GOOF-OFF DAY	day / month(.6)	April	April
NATIONAL MAGIC DAY & NEVADA ADMISSION DAY	day / month(.8)	October	October

© 2010 IBM Corporation

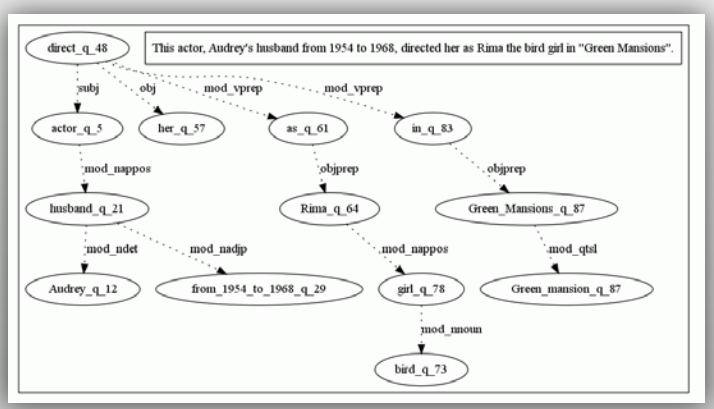
DeepQA: Incremental Progress in Answering Precision on the Jeopardy Challenge: 6/2007-11/2010



© 2010 IBM Corporation

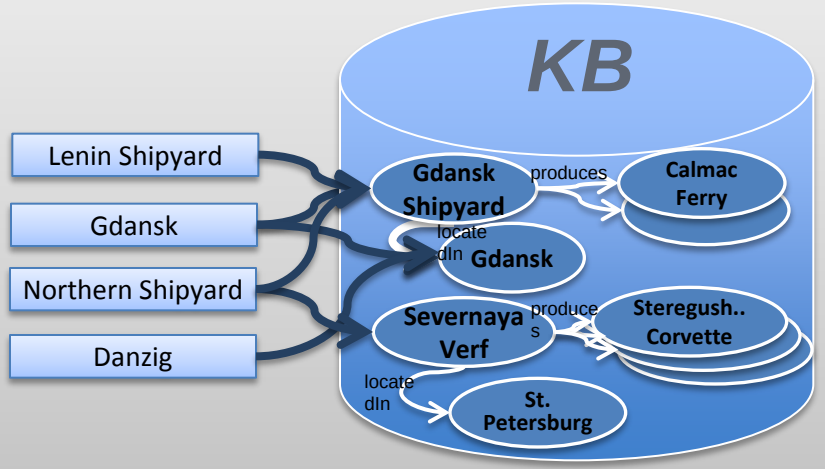
NLP Components Technology Performance

Question Processing



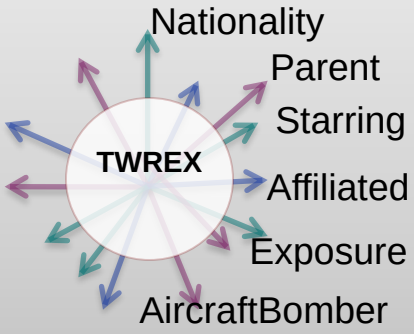
Dependency parse, Focus/LAT detection: 6000 rules, Decomp.
Eval on Jeopardy!: Parser**: 92.4% acc**, Stat LAT detection: 96.8%

KAFE: Knowledge From Extracted Content



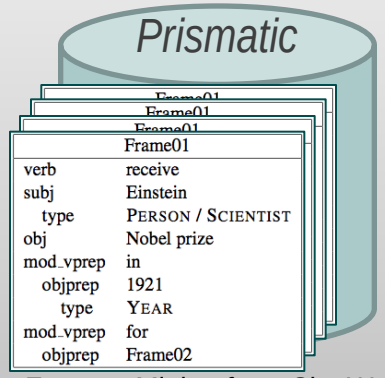
Entity Disambiguation, Entity Typing, Type Disambiguation
Eval on **Wikipedia Disambig Task, F-score 92.5**

Relation Extraction



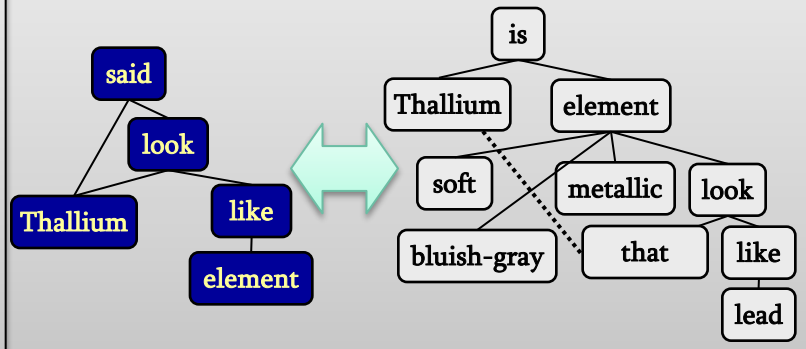
Semantic Relation Repository 7000 rels
Eval on ACE**: F-score 73.2 (**leading score)

Linguistic Frame Extraction



>1 Billlon Frames. Mining from ClueWeb:
SVO/isa/etc. cuts,
Intensional/Extensional representation

Passage Matching Ensemble



Synonymy, Temporal/Geographic Reasoning, Linguistic Axioms
Eval on RTE 2010 **Text Entailment:** F-score 48.8 (leading score)

© 2010 IBM Corporation

One Jeopardy! question can take **2 hours on a single 2.6Ghz Core** Optimized & Scaled out on 2880-Core IBM Power750's using UIMA-AS, *Watson* is answering in 2-6 seconds.

Question

Multiple Interpretations

100s sources

100s Possible Answers

1000's of Pieces of Evidence

100,000's scores from many simultaneous Text Analysis Algorithms

Question & Topic Analysis → Question Decomposition → Hypothesis Generation → Hypothesis and Evidence Scoring → Synthesis → Final Confidence Merging & Ranking → **Answer & Confidence**

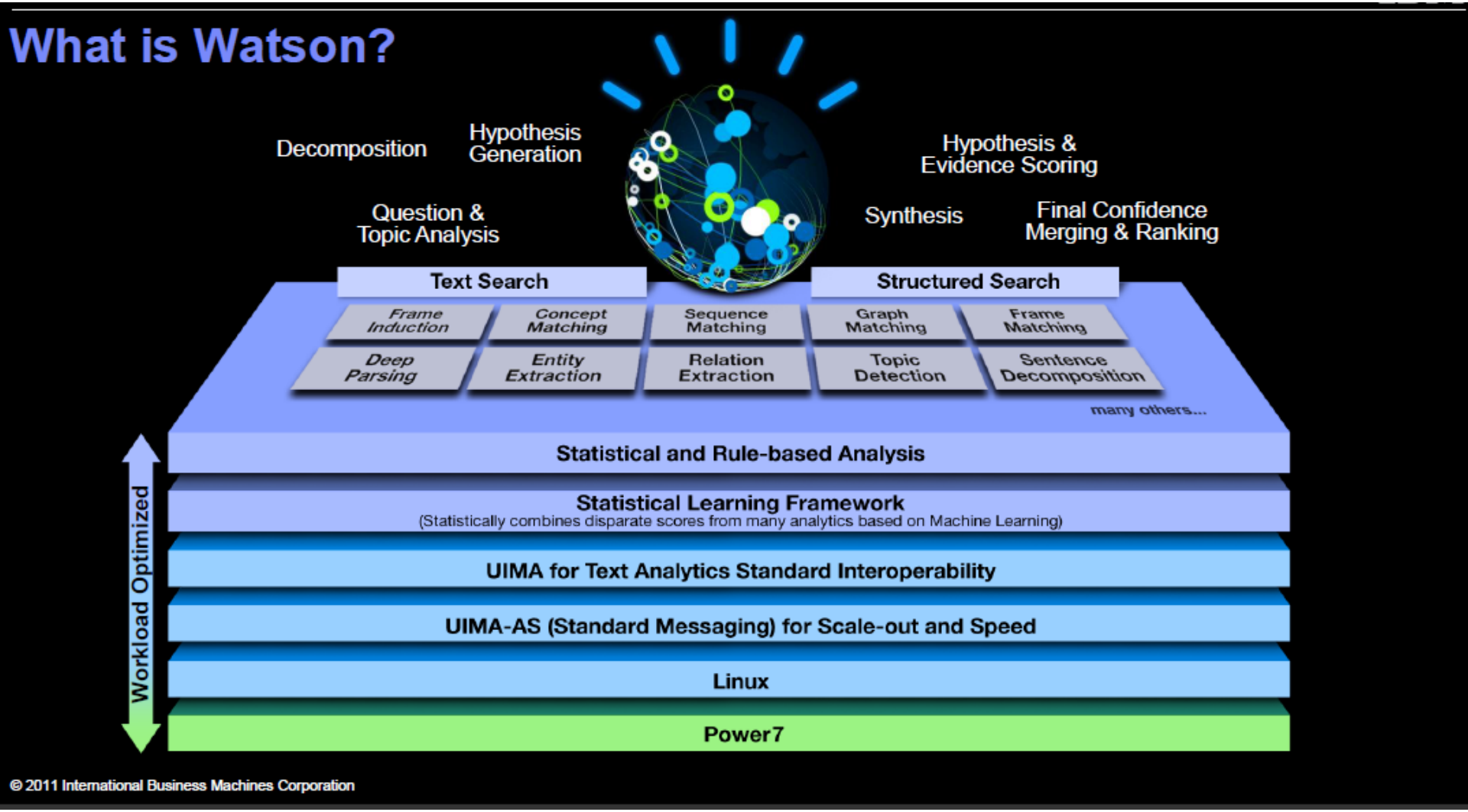
Question Decomposition → Hypothesis Generation → Hypothesis and Evidence Scoring → Synthesis

built on UIMA for interoperability

built on UIMA-AS for scale-out and speed

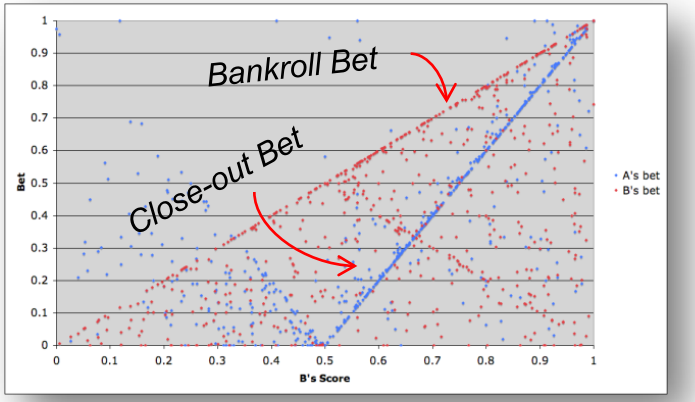
© 2010 IBM Corporation

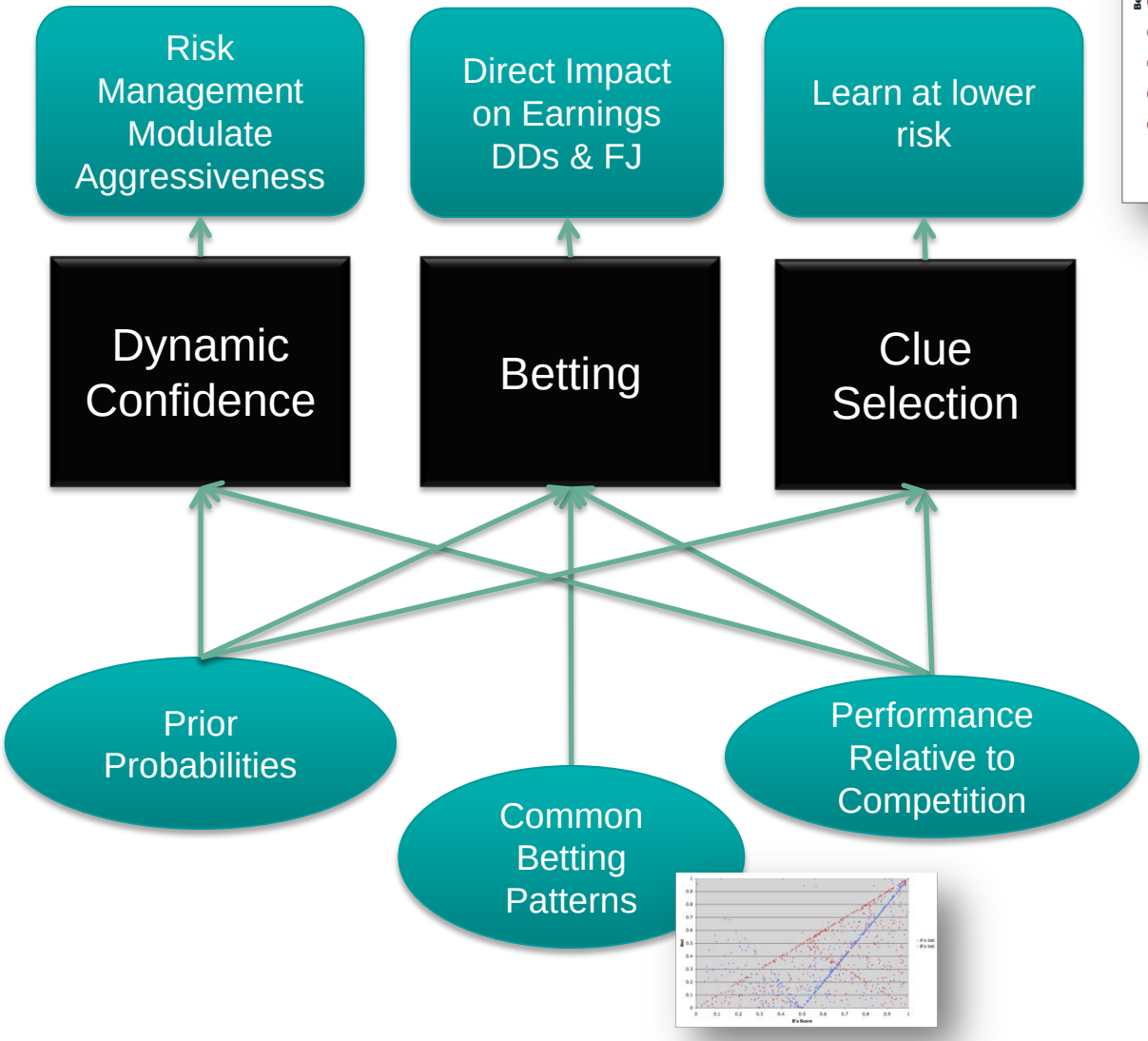
Run-Time Stack



© 2010 IBM Corporation

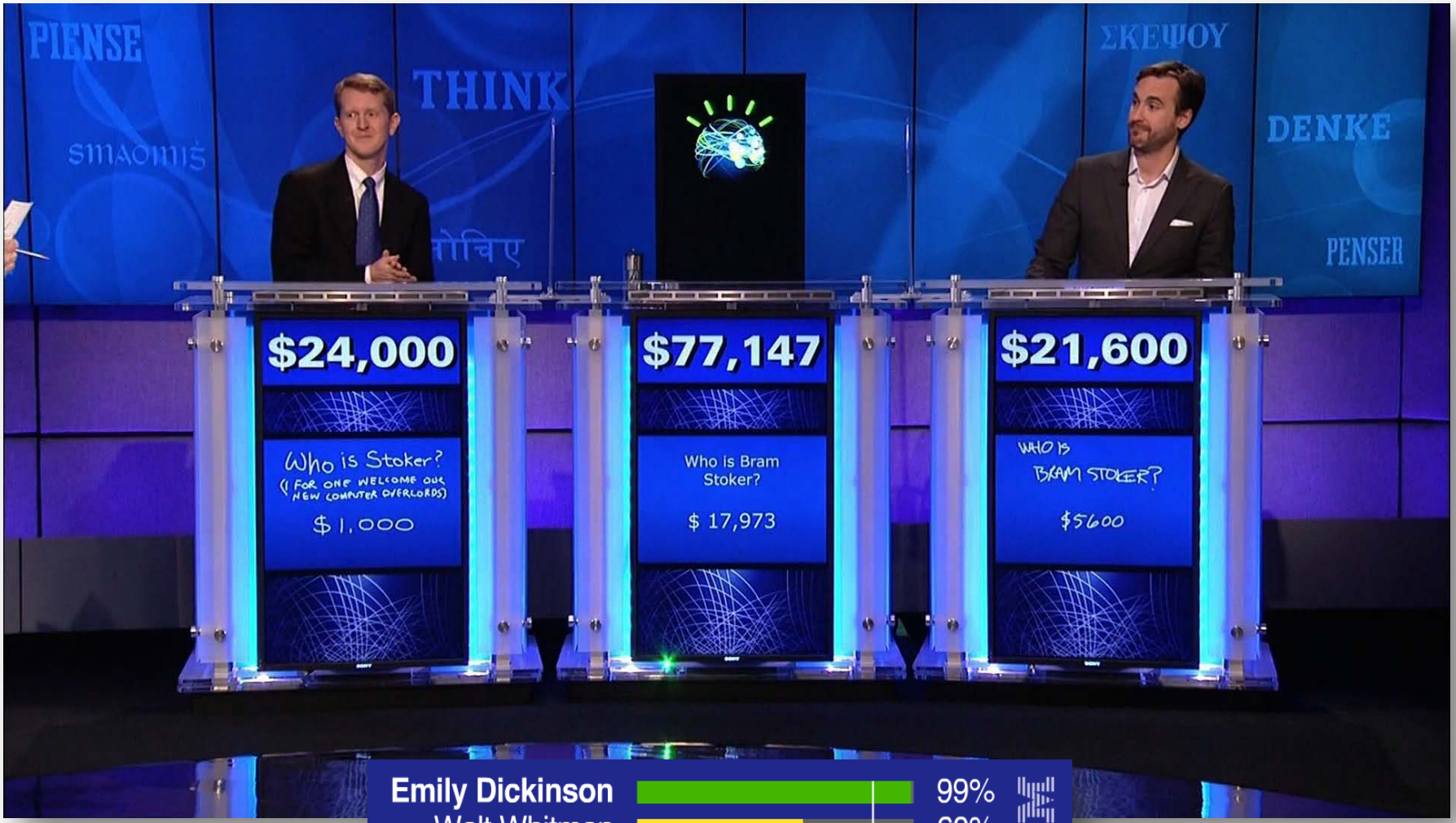
Game Strategy
Managing the Luck of the Draw





© 2010 IBM Corporation

With Precision, Accurate Confidence and Speed, the rest was History



© 2010 IBM Corporation

Potential Business Applications

Healthcare / Life Sciences: Diagnostic Assistance, Evidenced-Based, Collaborative Medicine

Tech Support: Help-desk, Contact Centers

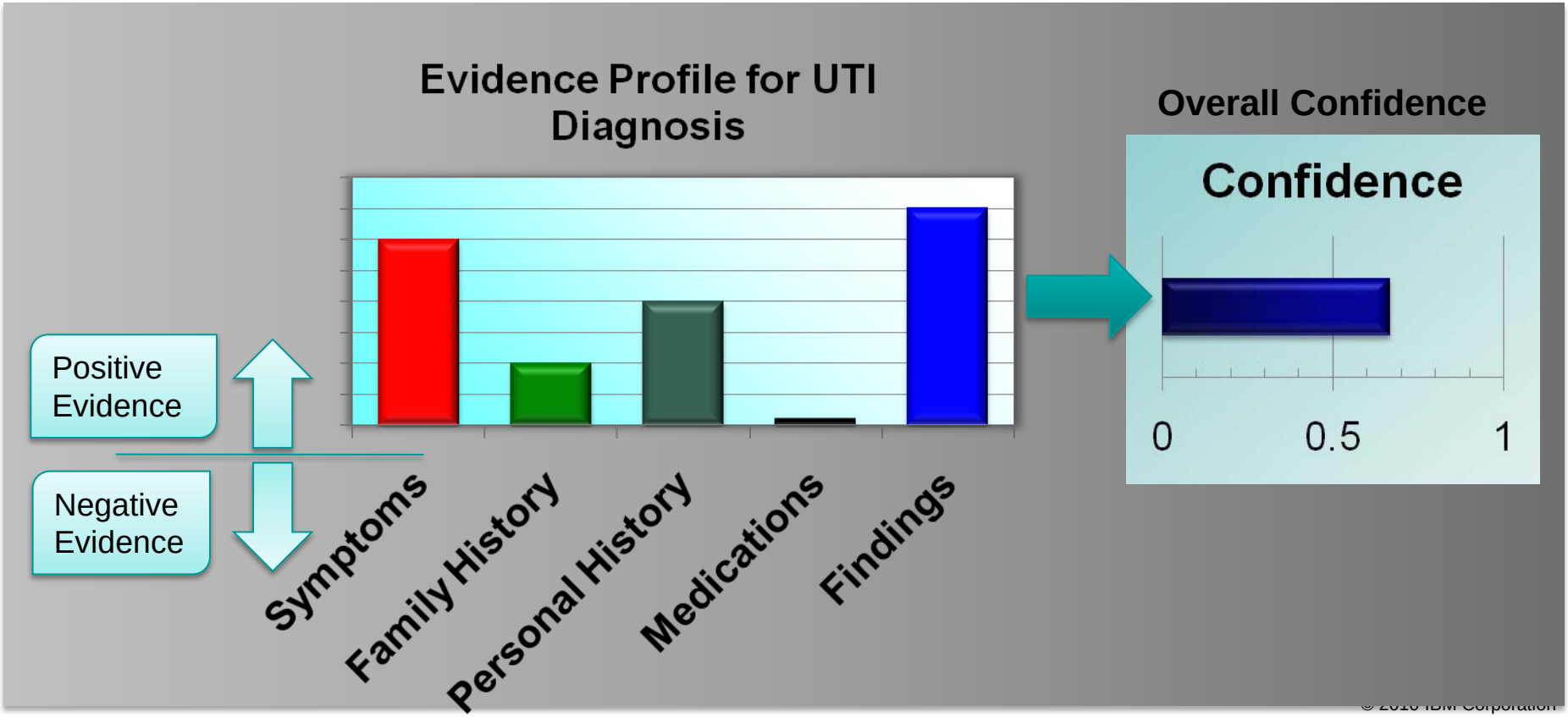
Enterprise Knowledge Management and Business Intelligence

Government: Improved Information Sharing and Security

© 2010 IBM Corporation

Evidence Profiles from disparate data is a powerful idea

- Each dimension contributes to supporting or refuting hypotheses based on
  - **Strength of evidence**
  - **Importance of dimension for diagnosis** (learned from training data)
- Evidence dimensions are combined to produce an overall confidence



© 2010 IBM Corporation

DeepQA in Continuous Evidence-Based Diagnostic Analysis

Considers and synthesizes a broad range of evidence improving quality, reducing cost

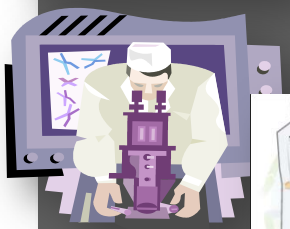
Symptoms

Family History

Patient History

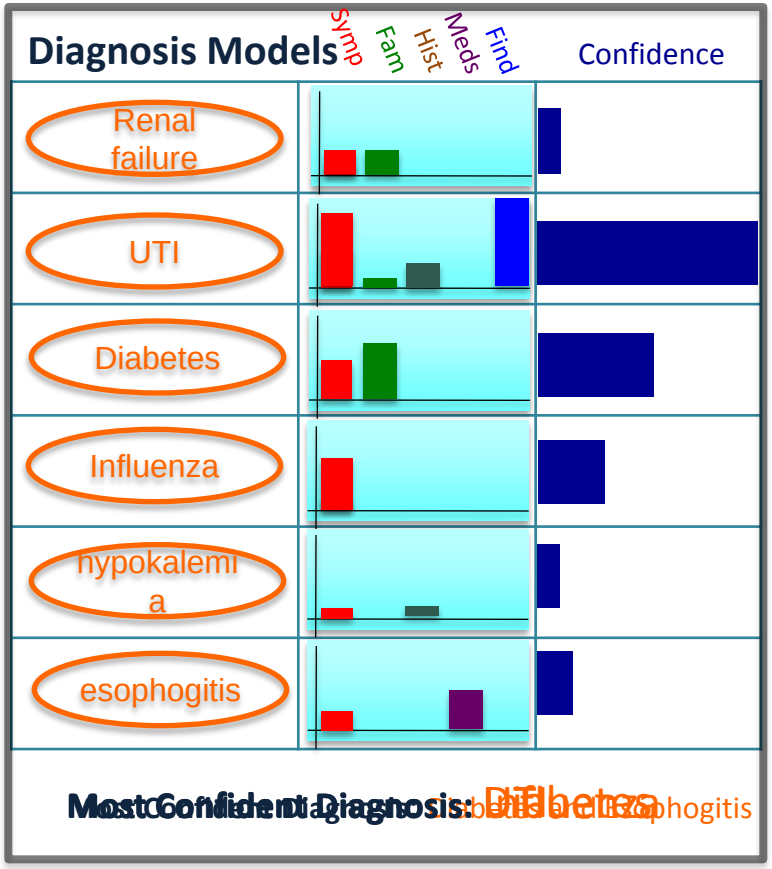
Medications

Tests/Findings



Notes/Hypotheses

Huge Volumes of Texts, Journals, References, DBs etc.



© 2010 IBM Corporation

The Core Technical Team*

Researchers and Engineers in **NLP, ML, IR, KR&R and CL** at IBM Labs and a growing number of universities

PI: David Ferrucci

Systems & Speed

Eric Brown
Jerry Cwiklik
Pablo Duboue
Eddie Epstein
Tong Fin
Dan Gruhl
Bhavani Iyer
Adam Lally
Burn Lewis
Marshall Schor

Core Algorithms

Eric Brown	Radu Florian	Dafna Sheinwald
Sugato Bagchi	David Gondek	Siddarth Patwardhan
Bran Boguraev	Aditya Kalyanpur	Kohichi Takeda
David Carmel	Hiroshi Kanayama	Yue Pan
Art Ciccolo	Adam Lally	John Prager
Jennifer Chu-Carroll	Tony Levas	Chris Welty
Bonaventura Coppola	Michael McCord	Wlodek Zadrozny
James Fan	Bill Murdock	Lei Zhang
David Ferrucci	Yuan Ni	Chang Wang
Achille Fokoue	Zhao Ming Qiu	

Strategy

David Gondek
Jon Lenchner
Gerry Tesauro
James Fan
John Prager

Speech

Andy Aaron
Raul Fernandez
Miroslav Novak
Andrew Rosenberg
Roberto Sicconi

University Collaborations & Students

Eric Nyberg (CMU)	James Allen (UMASS)	Andy Schlaikjer (CMU)
Nico Schlaefer (CMU)	Ed Hovy (USC)	Saurav Sahay (GT)
Manas Pathak (CMU)	Bruce Porter (UT)	Rutu Mulkar-Mehta (USC)
Hideki Shima (CMU)	Pallika Kanani (UMASS)	Doo Soon Kim (UT)
Chang Wang (UMASS)	Boris Katz (MIT)	
Barbara Cutler (RPI)	**Alessandro Moschitti (Trento)**	

Data Annotation

Karen Ingraffea
Matt Mulholland

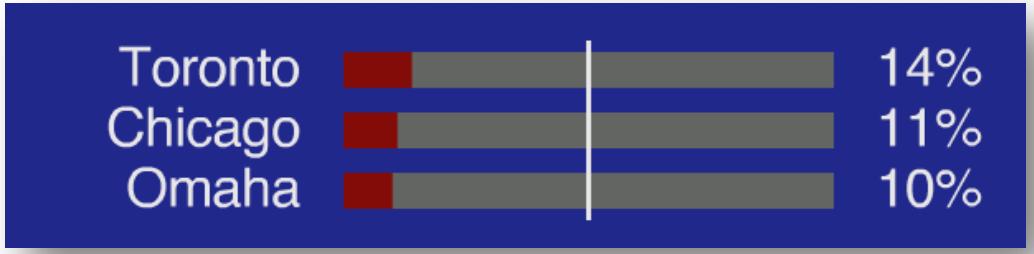
The broader team that contributes to delivering Watson for the “Stage”, to compete in Jeopardy Games is growing and changing rapidly.

Our Apologies if this is not up to date.

*NOT full-time Equivalents. Names listed if contributed some time to that part of project.

© 2010 IBM Corporation

TORONTO?



© 2010 IBM Corporation

Categories are not as simple as the seem

Watson uses statistical machine learning to discover that **Jeopardy! categories are only weak indicators of the answer type**.

U.S. CITIES	Country Clubs	Authors
St. Petersburg is home to Florida's annual tournament in this game popular on shipdecks **(Shuffleboard)**	From India, the shashpar was a multi-bladed version of this spiked club **(a mace)**	Archibald MacLeish? based his verse play "J.B." on this book of the Bible **(Job)**
Rochester, New York grew because of its location on this **(the Erie Canal)**	A French riot policeman may wield this, simply the French word for "stick“ **(a baton)**	In 1928 Elie Wiesel was born in Sighet, a Transylvanian village in this country **(Romania)**

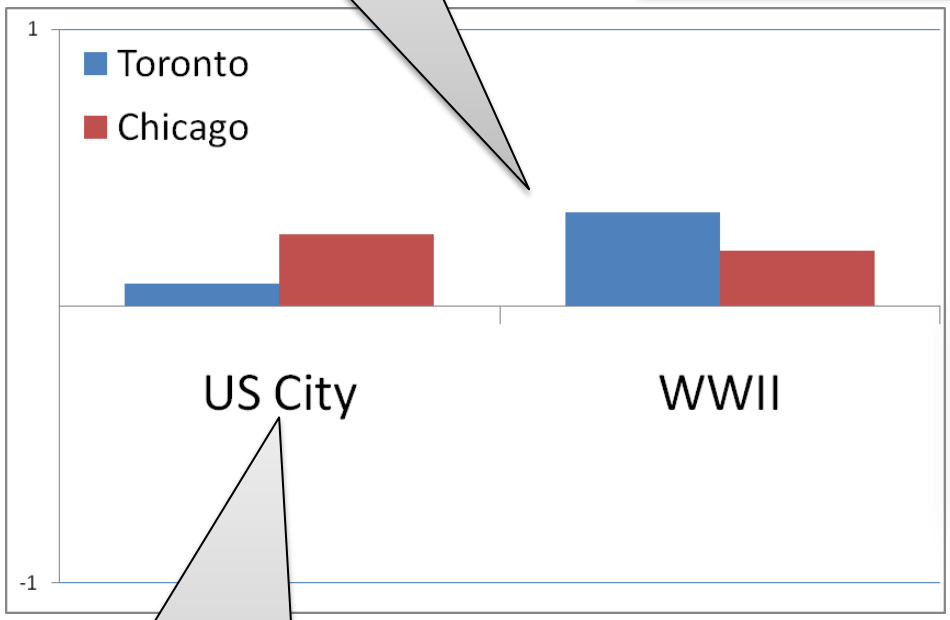
© 2010 IBM Corporation

Toronto vs. Chicago

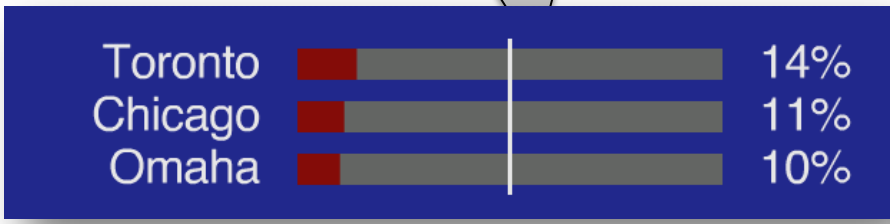
US CITIES

Its largest airport is named for a World War II hero; its second largest, for a World War II battle

Low because of weak evidence in content



Overall confidence was below threshold for both answers



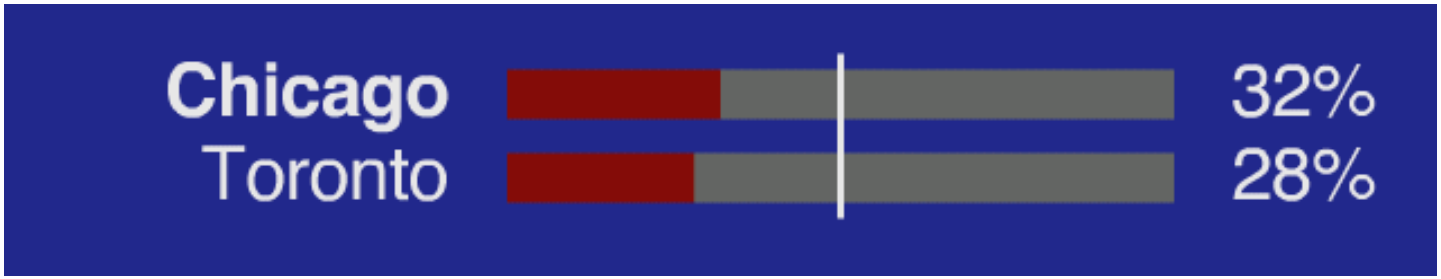
Low because being a **US City** is not a strong requirement simply based on Jeopardy! category

© 2010 IBM Corporation

When being a **US City** is a more trusted requirement, **Chicago** is in first position but still weak, because Watson is not convinced about the WWII evidence it could find.

US CITIES

This **US City's** largest airport is named for a World War II hero; its second largest, for a World War II battle



© 2010 IBM Corporation

THANK YOU

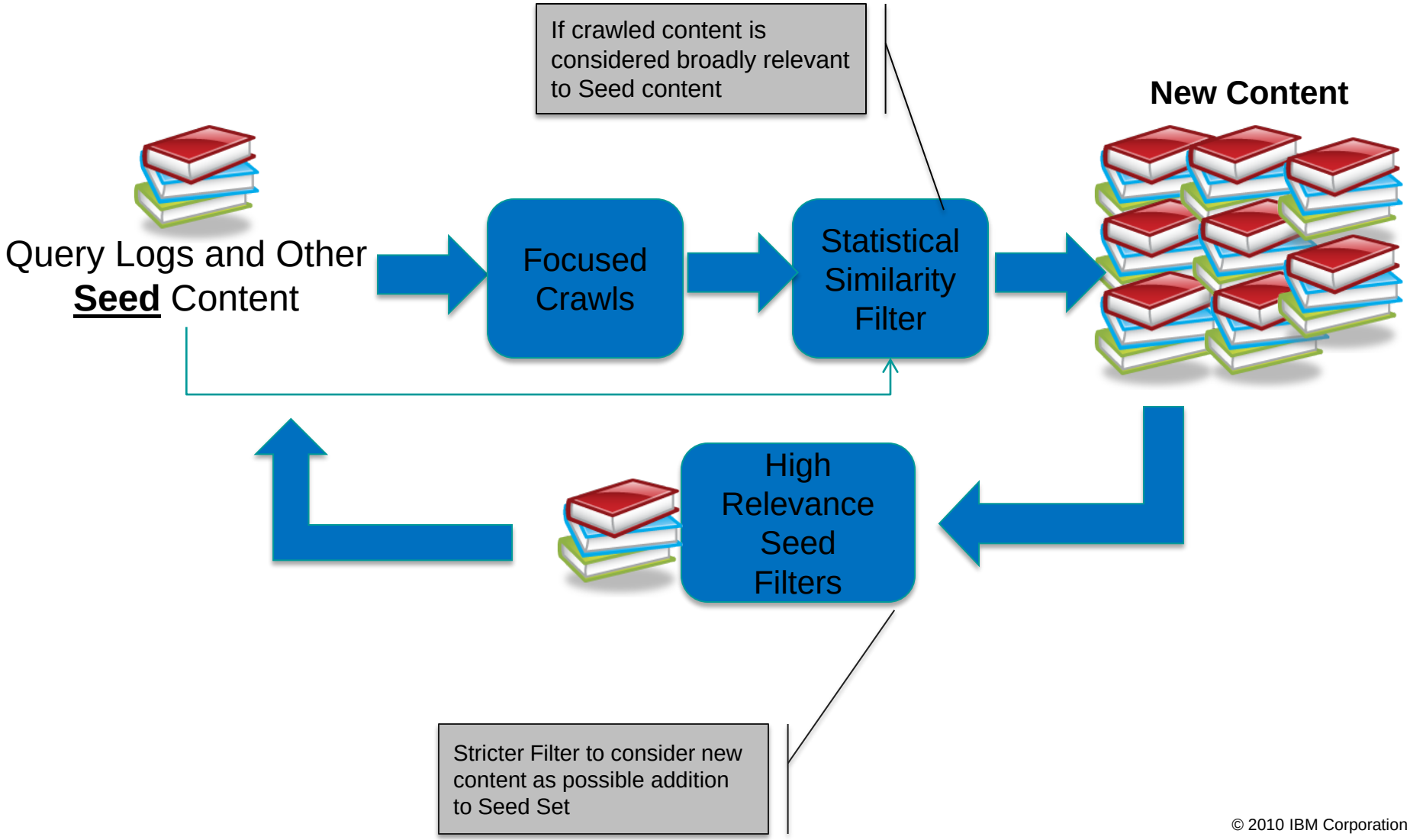
© 2010 IBM Corporation

BACKUPS

© 2010 IBM Corporation

Semi-Automatic Content Expansion

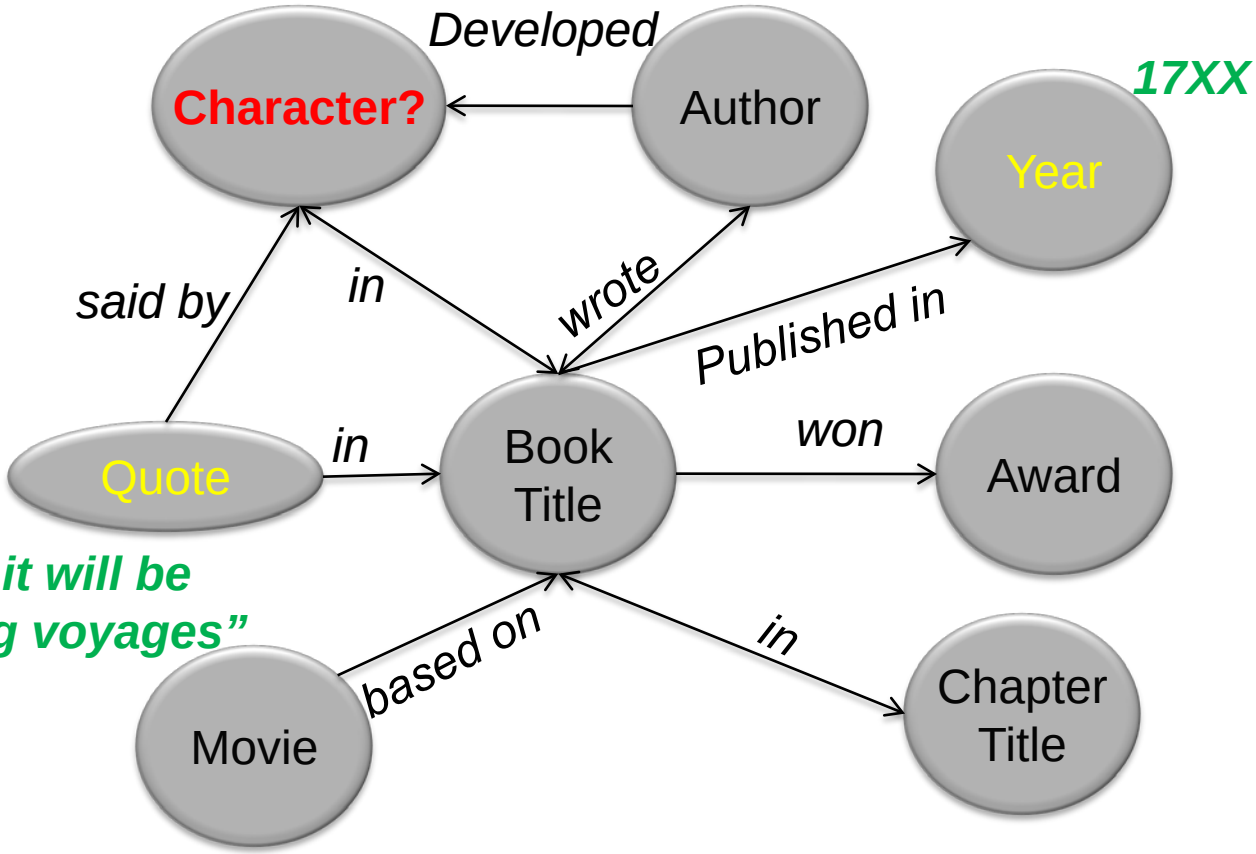
Collaborative work with CMU



© 2010 IBM Corporation

Using Frames to Find Alternatives Paths

This 18th century character studied medicine, "knowing it would be useful in long voyages"



- No *said* relation indicating who uttered the **quote** BUT…
- Can discover the quote is contained in a specific **book**, however
- The **frame** tells us how to relate a **character** to a **book** to a **year** …
- Filling out more of the frame can build confidence in selecting a character as the right answer

© 2010 IBM Corporation