

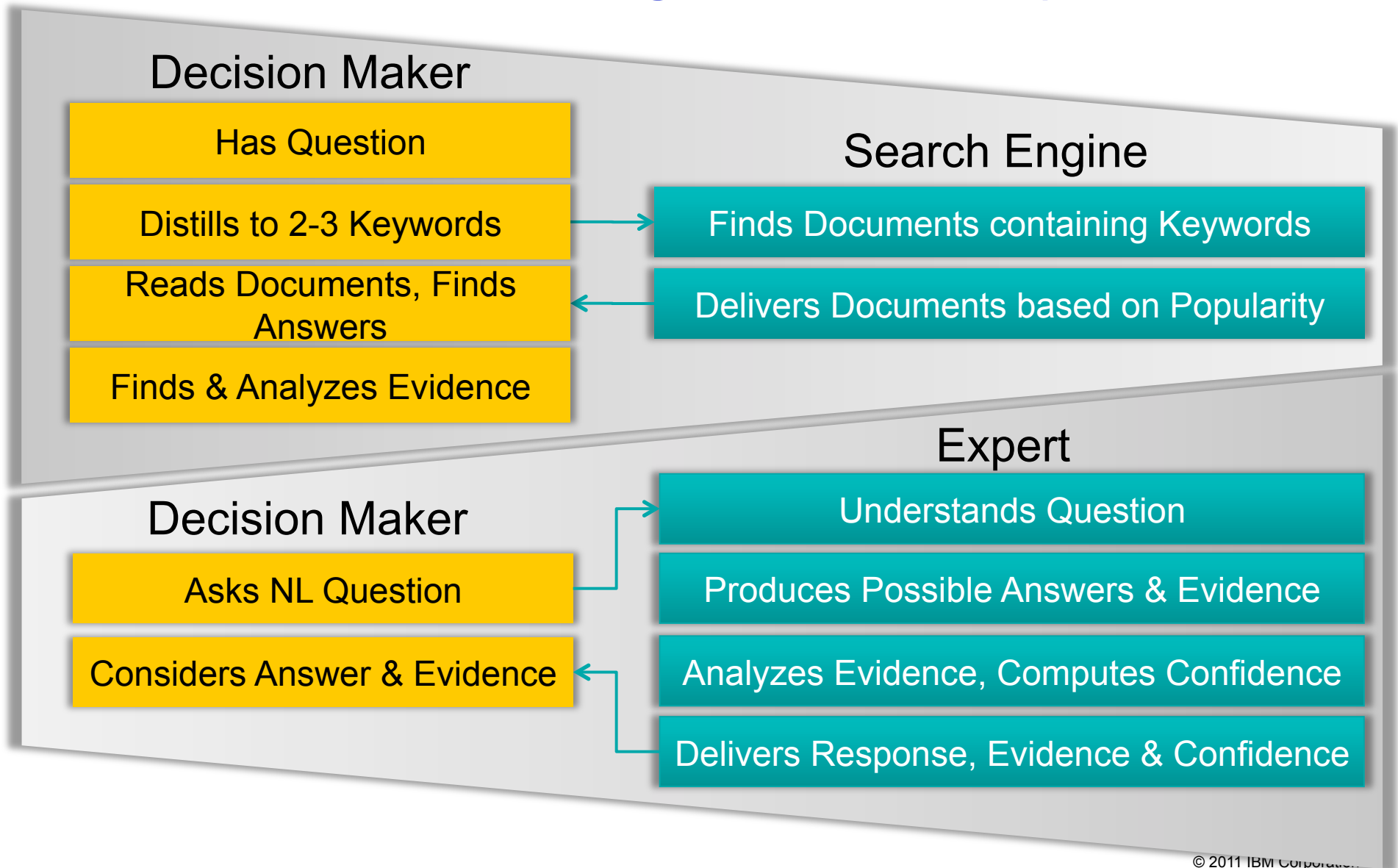
What is Watson – *An Overview*

Michael Perrone
Mgr., Multicore Computing
IBM TJ Watson Research Center

With special thanks to the
IBM Research DeepQA Team!



Informed Decision Making: Search vs. Expert Q&A



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



Easy Questions?

$$(\ln(12,546,798 * \pi))^2 / 34,567.46 = 0.00885$$

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

Owner	Serial Number
David Jones	45322190-AK

Invoice #	Vendor	Payment
INV10895	MyBuy	\$104.56

Serial Number	Type	Invoice #
45322190-AK	LapTop	INV10895

David Jones
 ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
 David Jones

=

Dave Jones
 ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
 David Jones

≠

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?

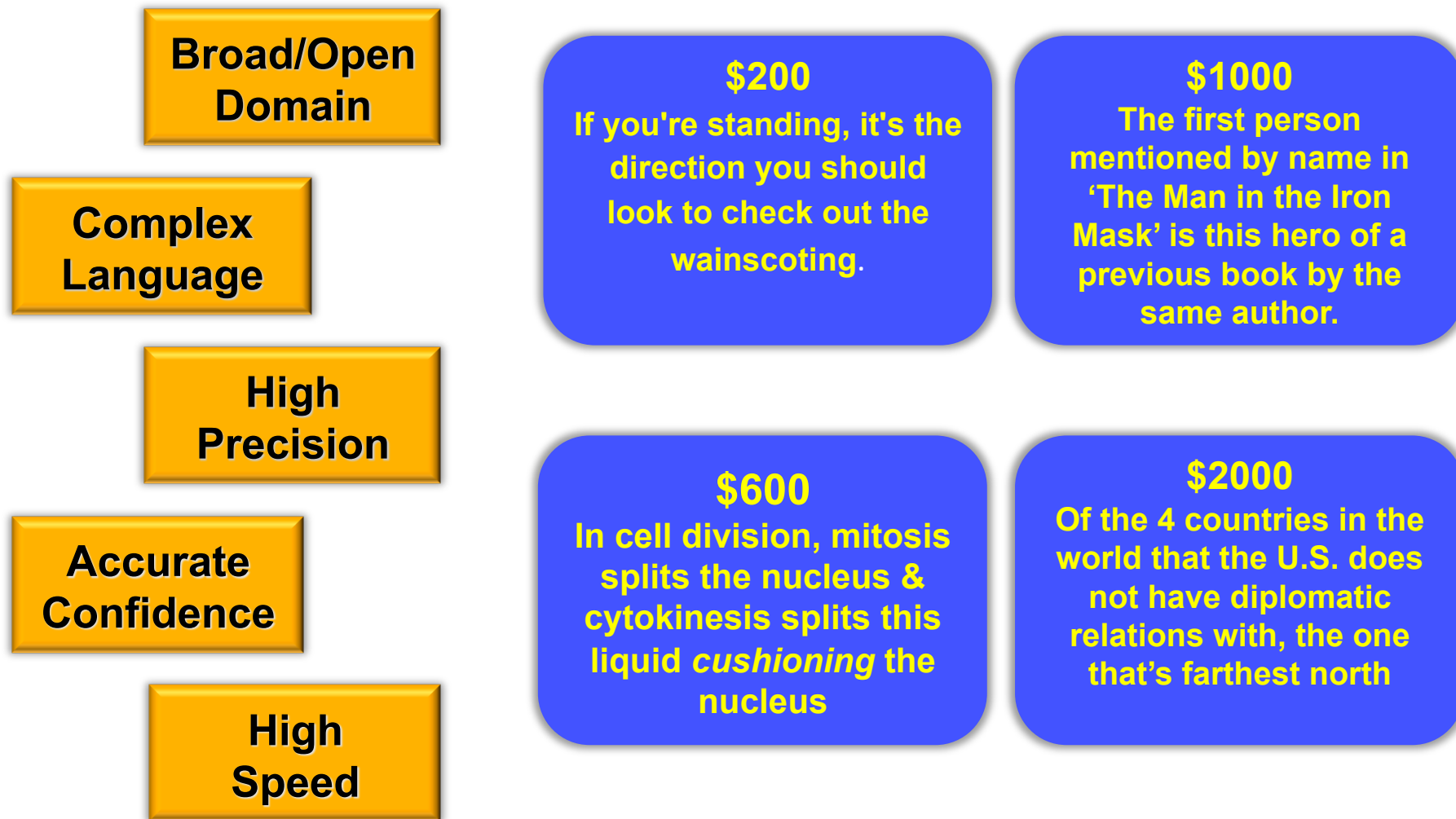
One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein's birthplace.

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.

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



Basic Game Play

Technology	Clue	Technology	Clue	Technology	Clue	Technology	Clue	Technology	Clue	Before and After
\$200	\$200									\$200
\$400	\$400									\$400
\$600	\$600									\$600
\$800	\$800									\$800
\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000	\$1000

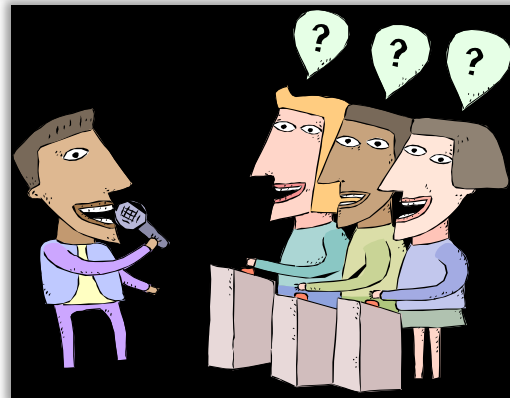
TECHNOLOGY

ALL POLICEMEN CAN THANK STEPHANIE KWOLEK FOR HER INVENTION OF THIS POLYMER FIBER, 5 TIMES TOUGHER THAN STEEL

6 Categories

5 Levels of Difficulty

- ❑ 1 of 3 Players Selects a Clue
- ❑ Host reads Clue out loud
- ❑ All Players compete to answer
- ❑ 1st to buzz-in gets to answer

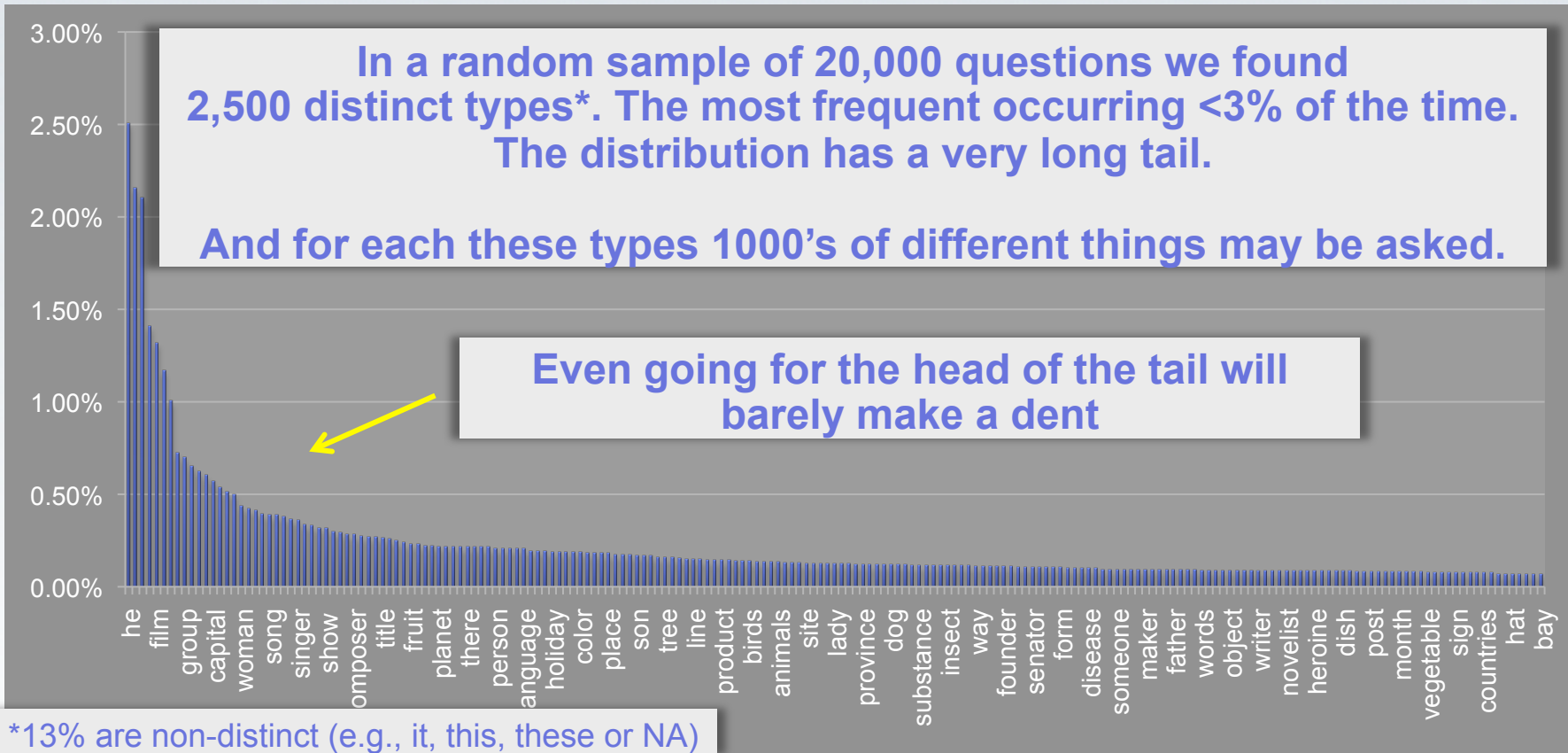


- ❑ **IF correct**
 - earns \$ value
 - selects Next Clue
- ❑ **IF wrong**
 - **loses** \$ value
 - other players buzz again (rebounds)

- ❑ Two Rounds Per Game + Final Question
- ❑ ONE Daily Double in First Round, TWO in 2nd Round

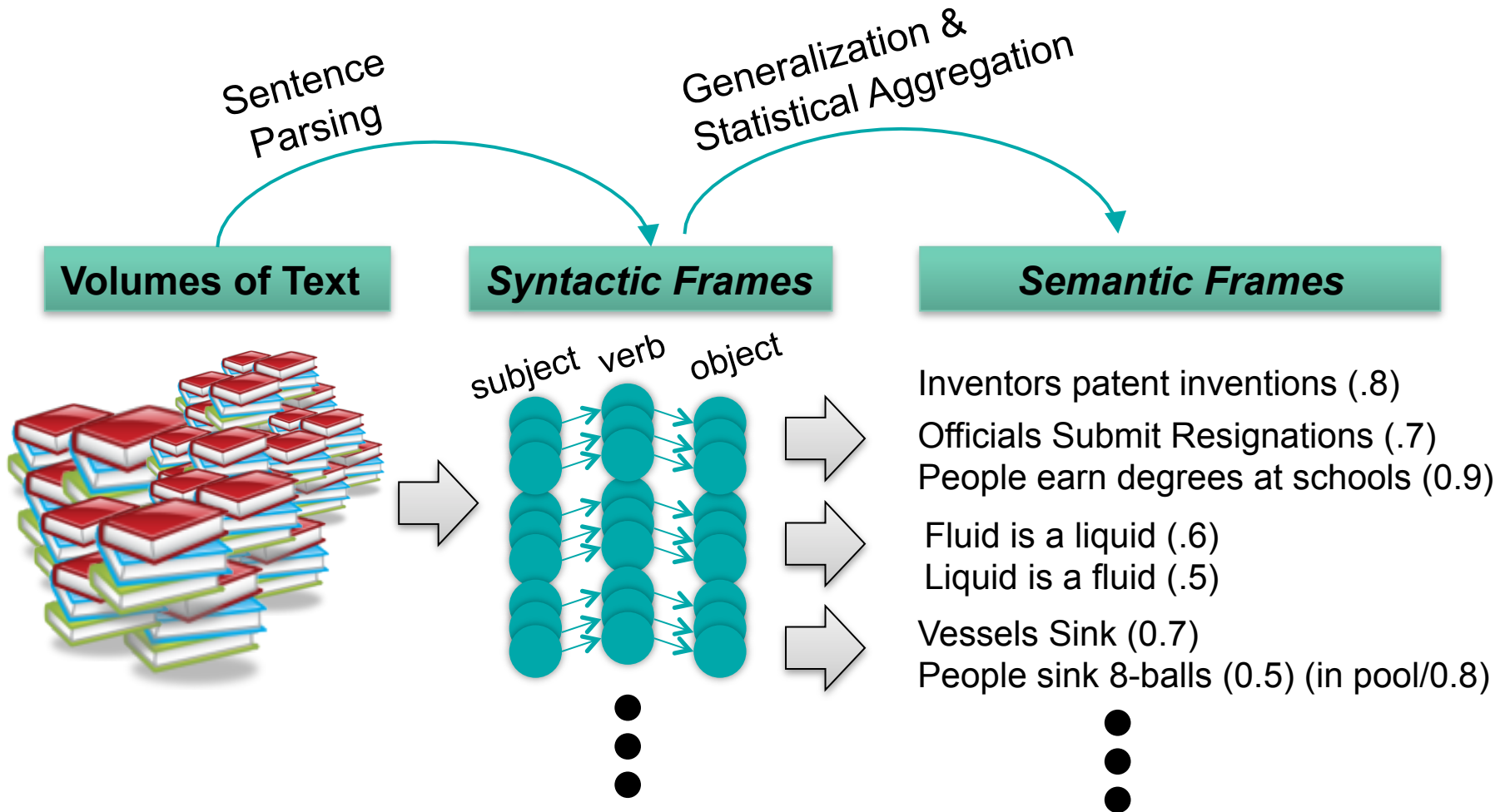
Broad Domain

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



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.

Automatic Learning From "Reading"



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) → ?

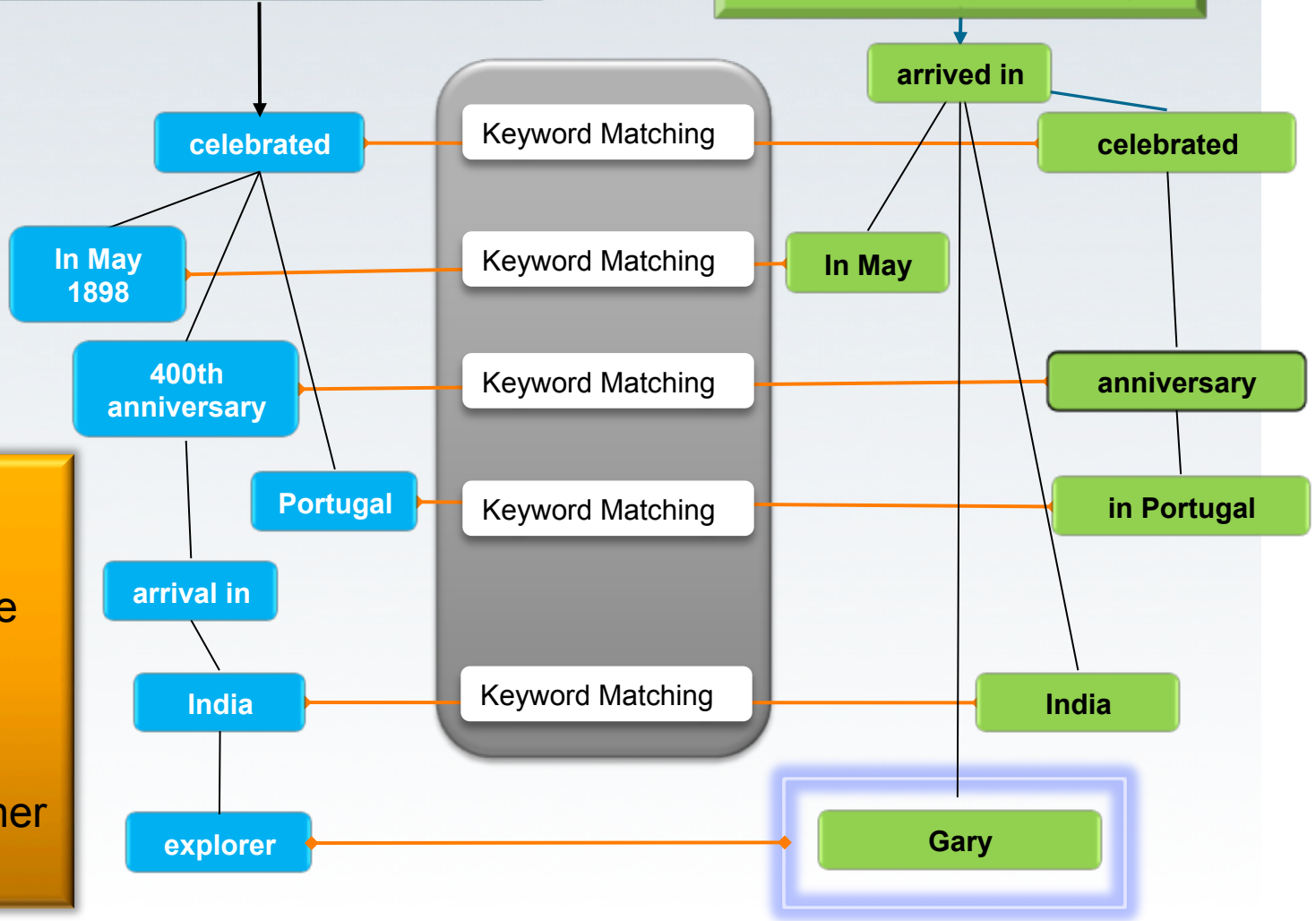
Learned → Is_a(Fluid, Liquid) → yes.

Keyword Evidence



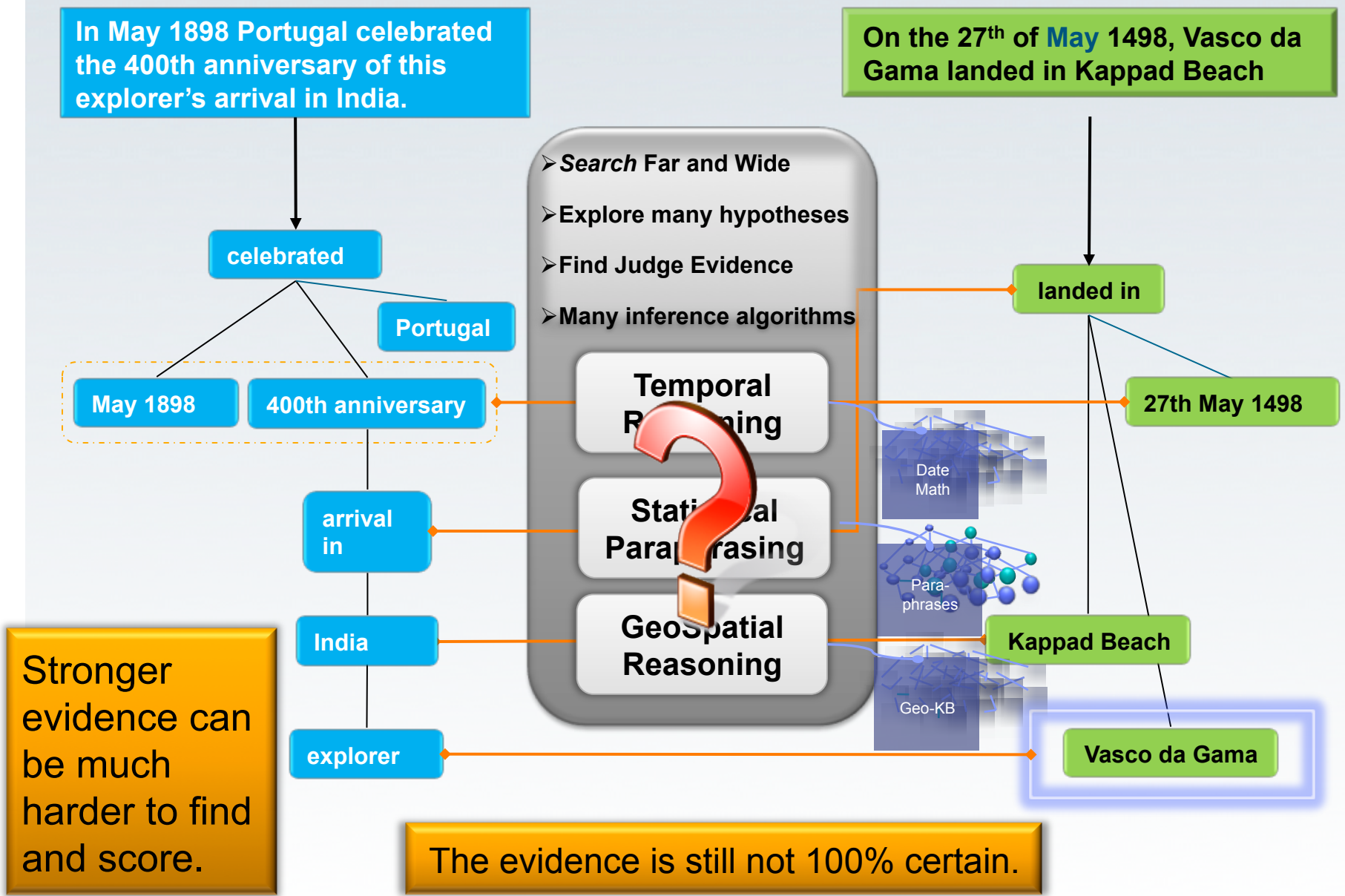
In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India.

In May, Gary arrived in India after he celebrated his anniversary in Portugal.



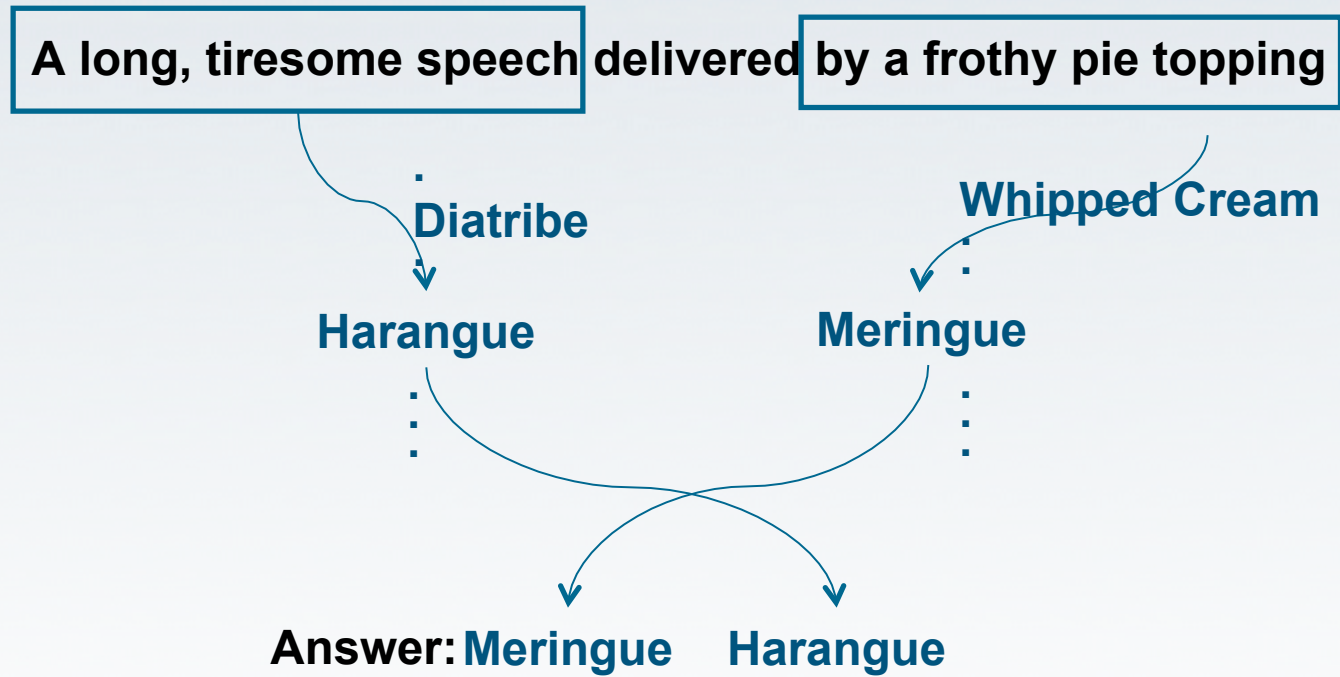
This evidence suggests "Gary" is the answer BUT the system must learn that keyword matching may be weak relative to other types of evidence

Deeper Evidence



Not *Just* for Fun

Some Questions require
Decomposition and Synthesis



Category: Edible Rhyme Time

Divide and Conquer (Typical in Final Jeopardy!)

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

Lyndon B Johnson

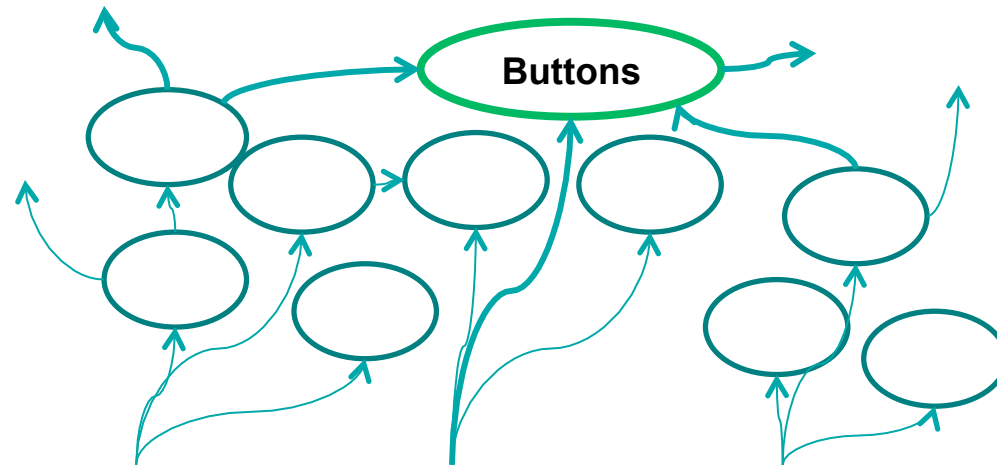
In 1968 this man was U.S. president.

When "60 Minutes" premiered this man was U.S. president.

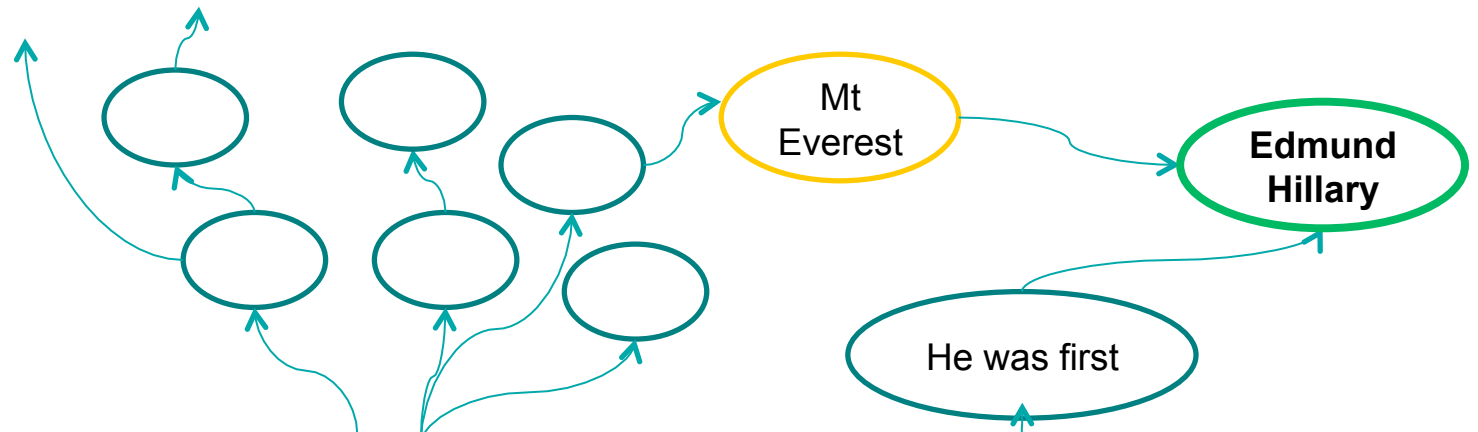
?

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

The Missing Link

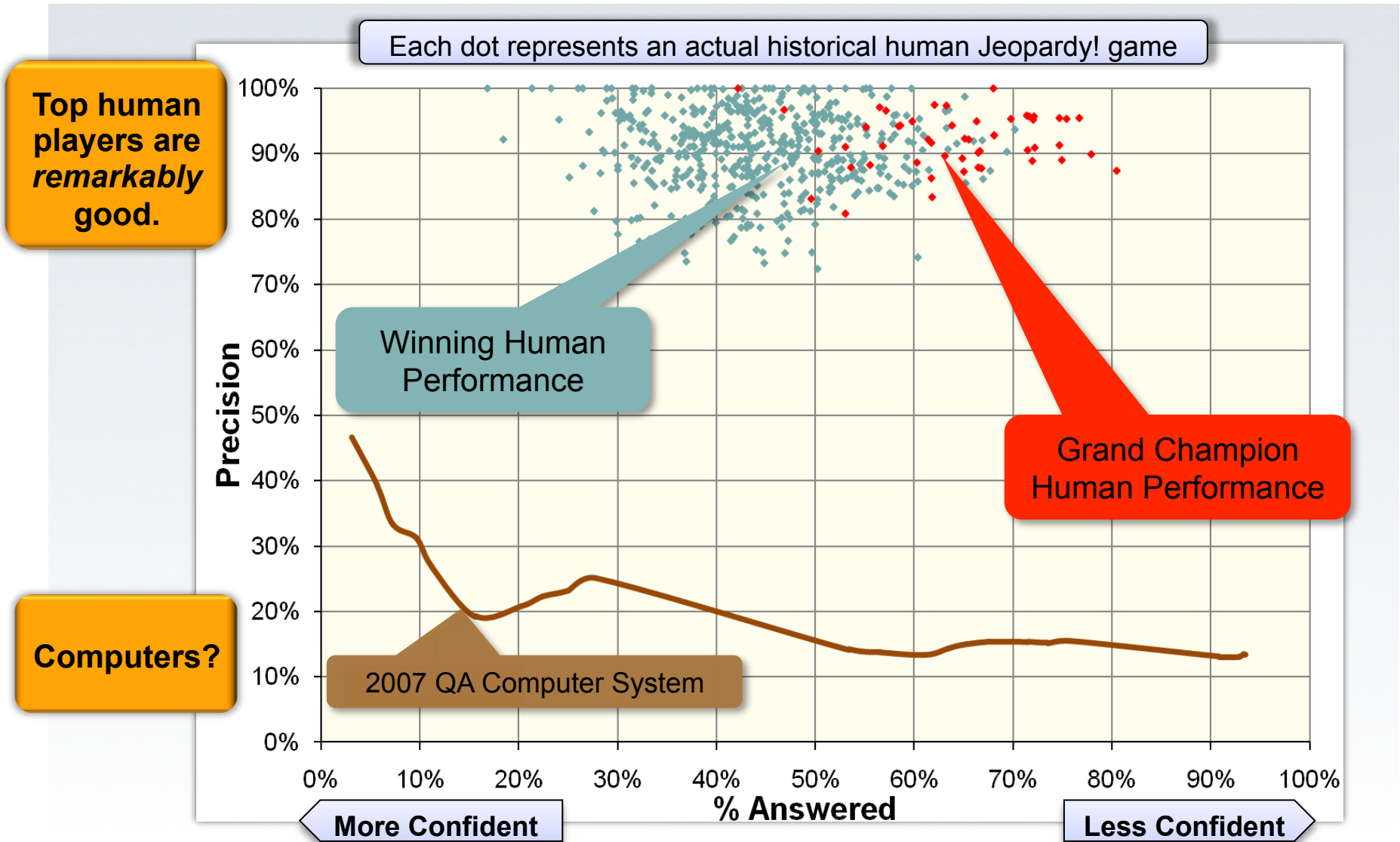


Shirts, TV remote controls, Telephones



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

The Best Human Performance: *Our Analysis Reveals the Winner's Cloud*



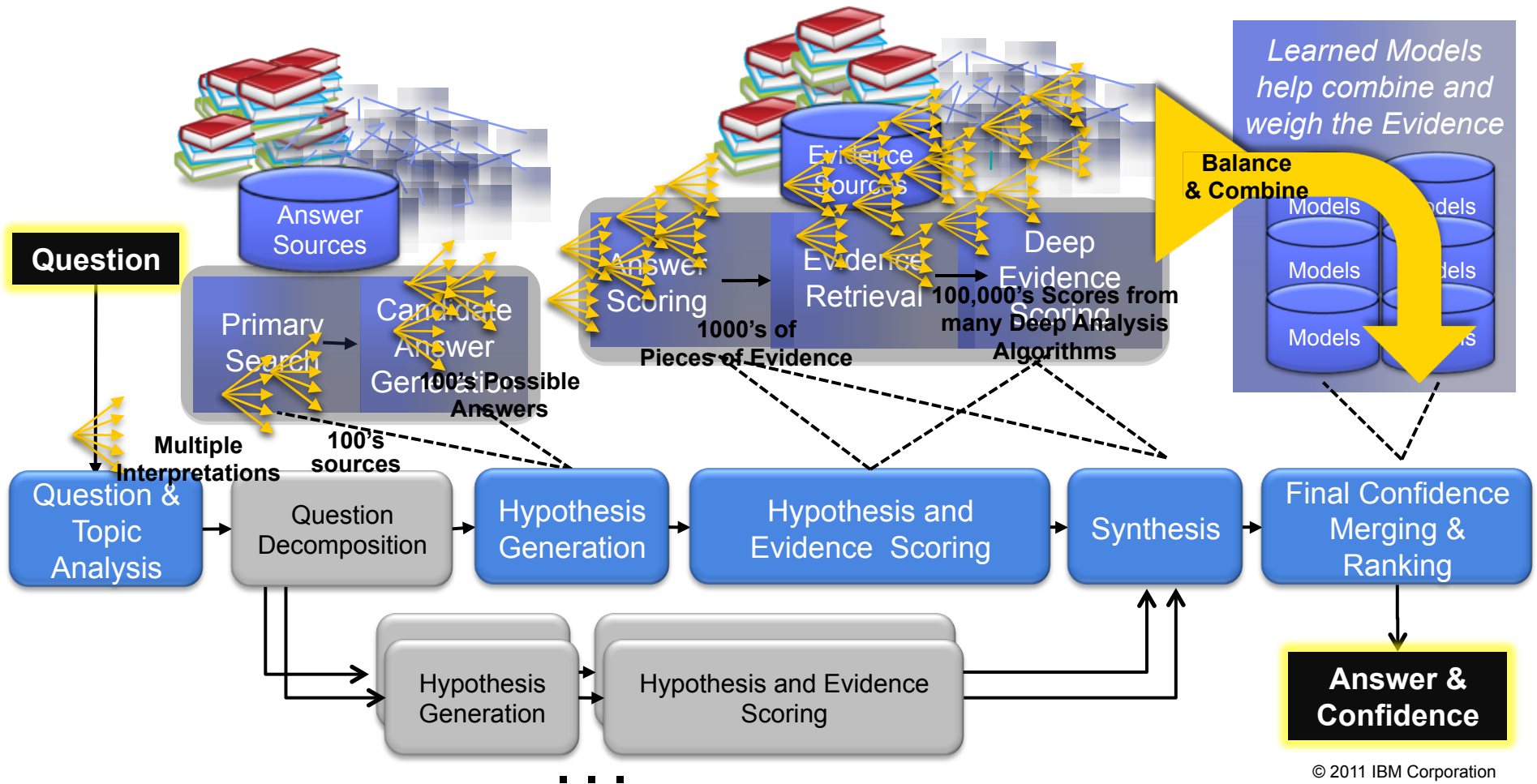
DeepQA: The Technology Behind Watson



Massively Parallel Probabilistic Evidence-Based Architecture

Generates and scores many hypotheses using a combination of 1000's **Natural Language Processing, Information Retrieval, Machine Learning and Reasoning Algorithms.**

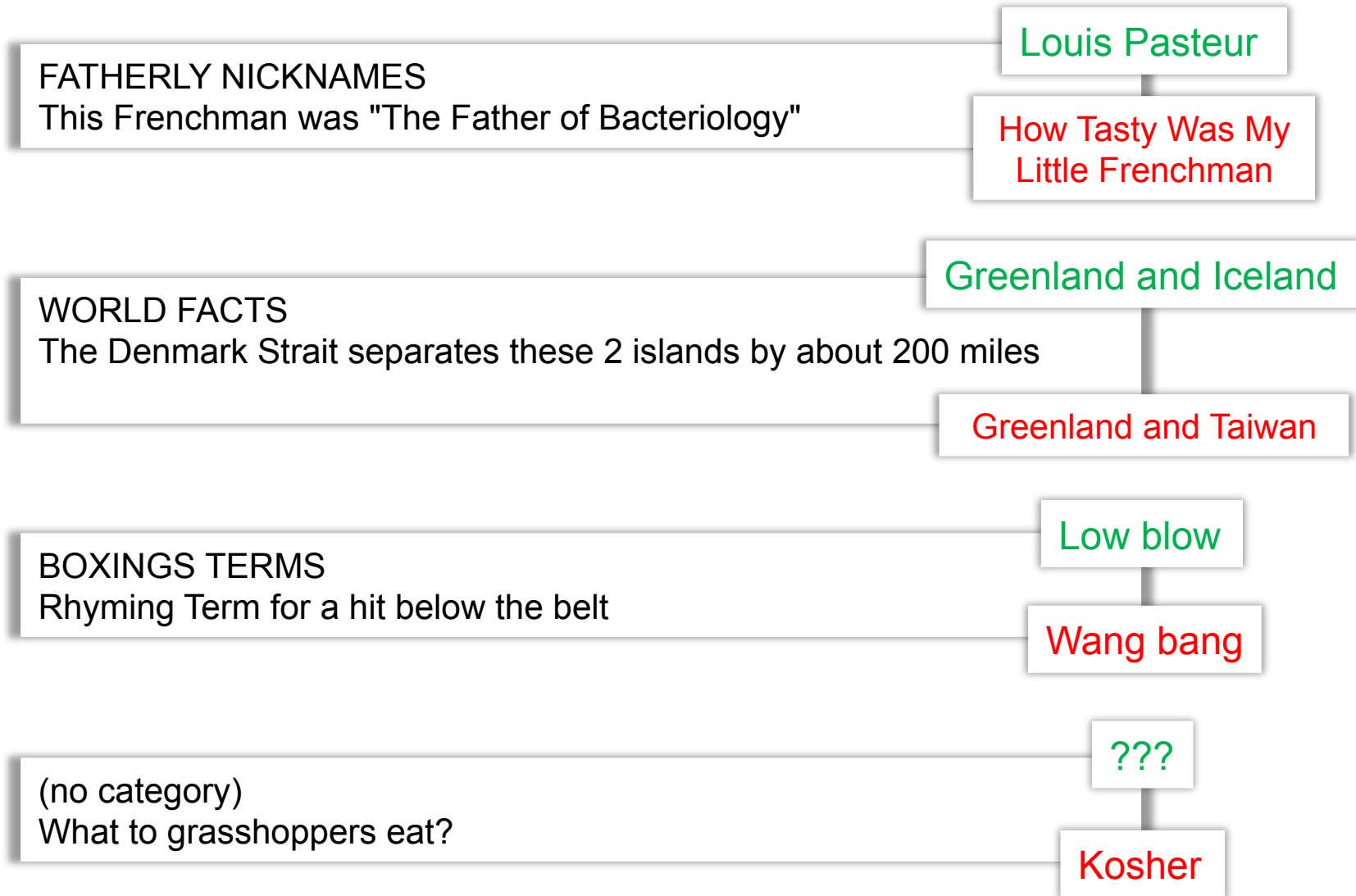
These gather, evaluate, weigh and balance different types of **evidence** to deliver the answer with the best support it can find.



Some Growing Pains (early answers)

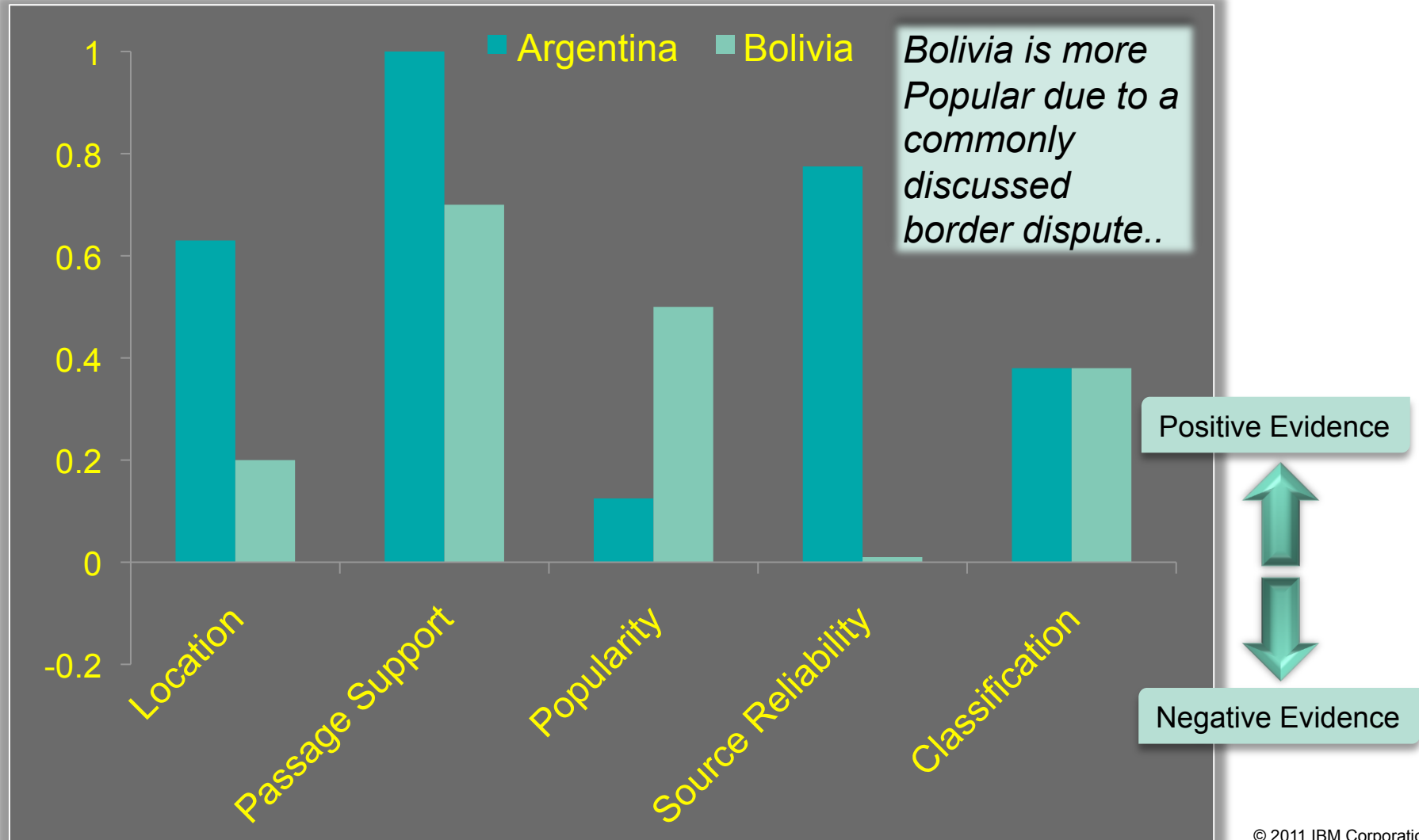


Some Growing Pains (early answers)



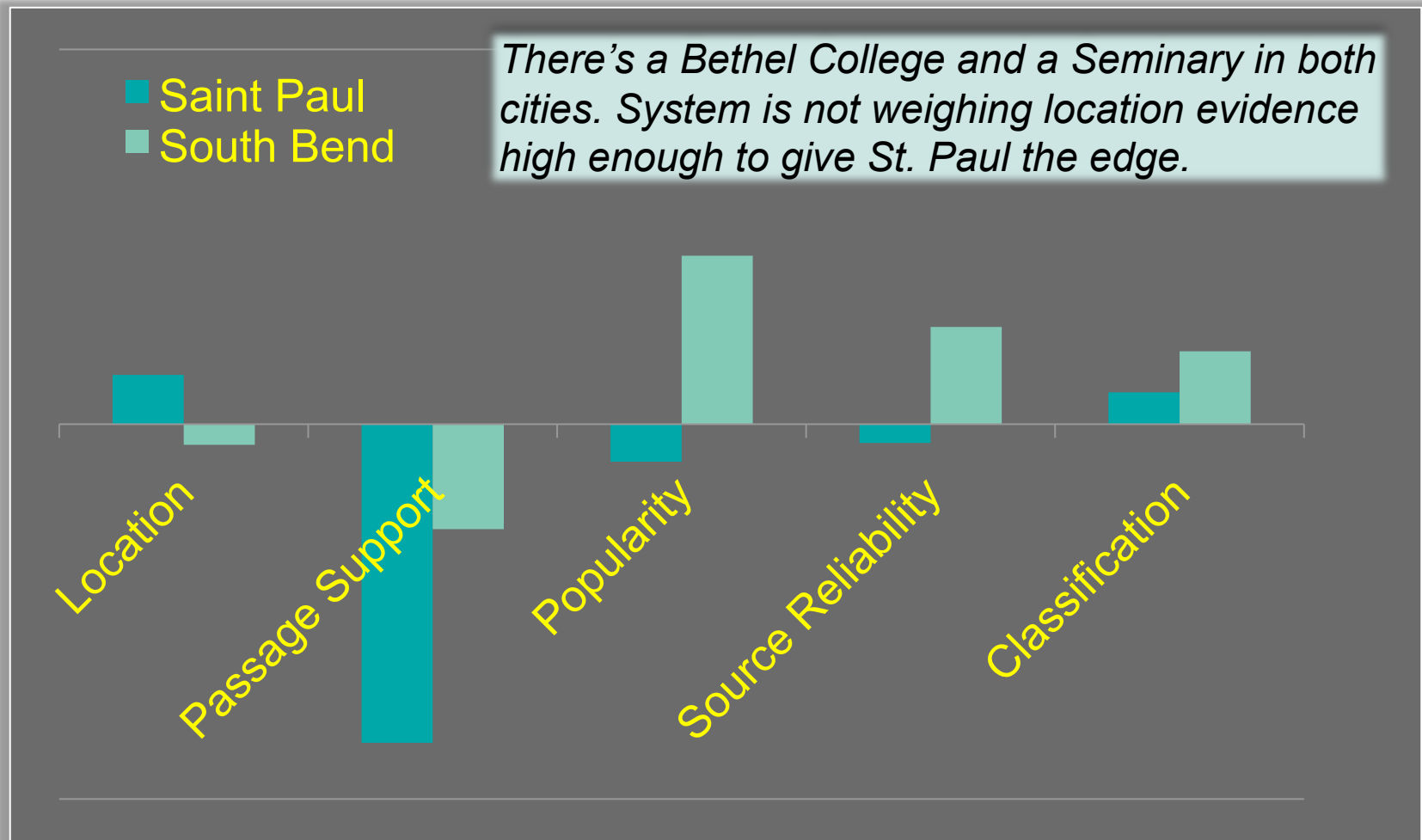
Grouping Features to produce *Evidence Profiles*

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



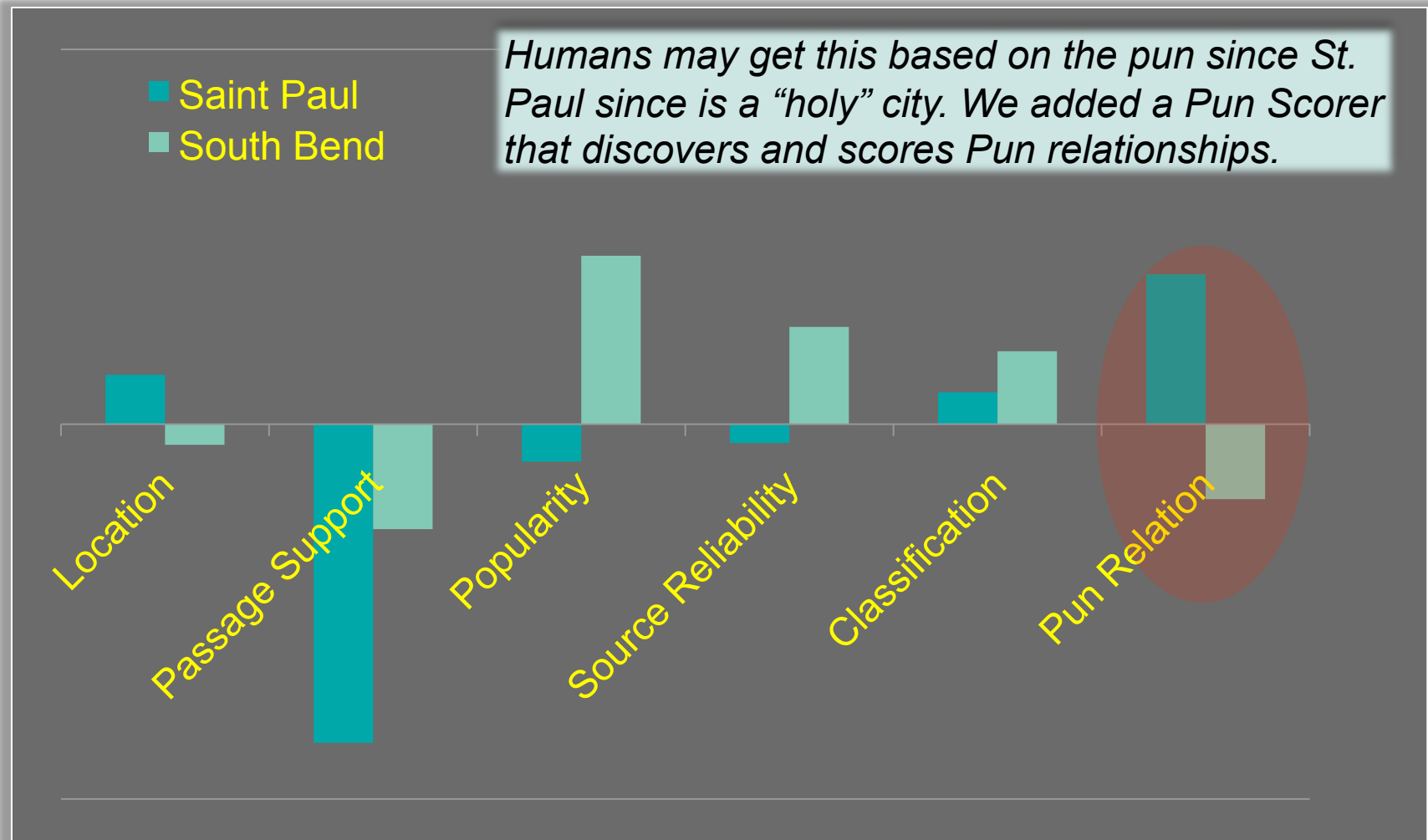
Evidence: Time, Popularity, Source, Classification etc.

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



Evidence: Puns

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



Categories are not as simple as they 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)

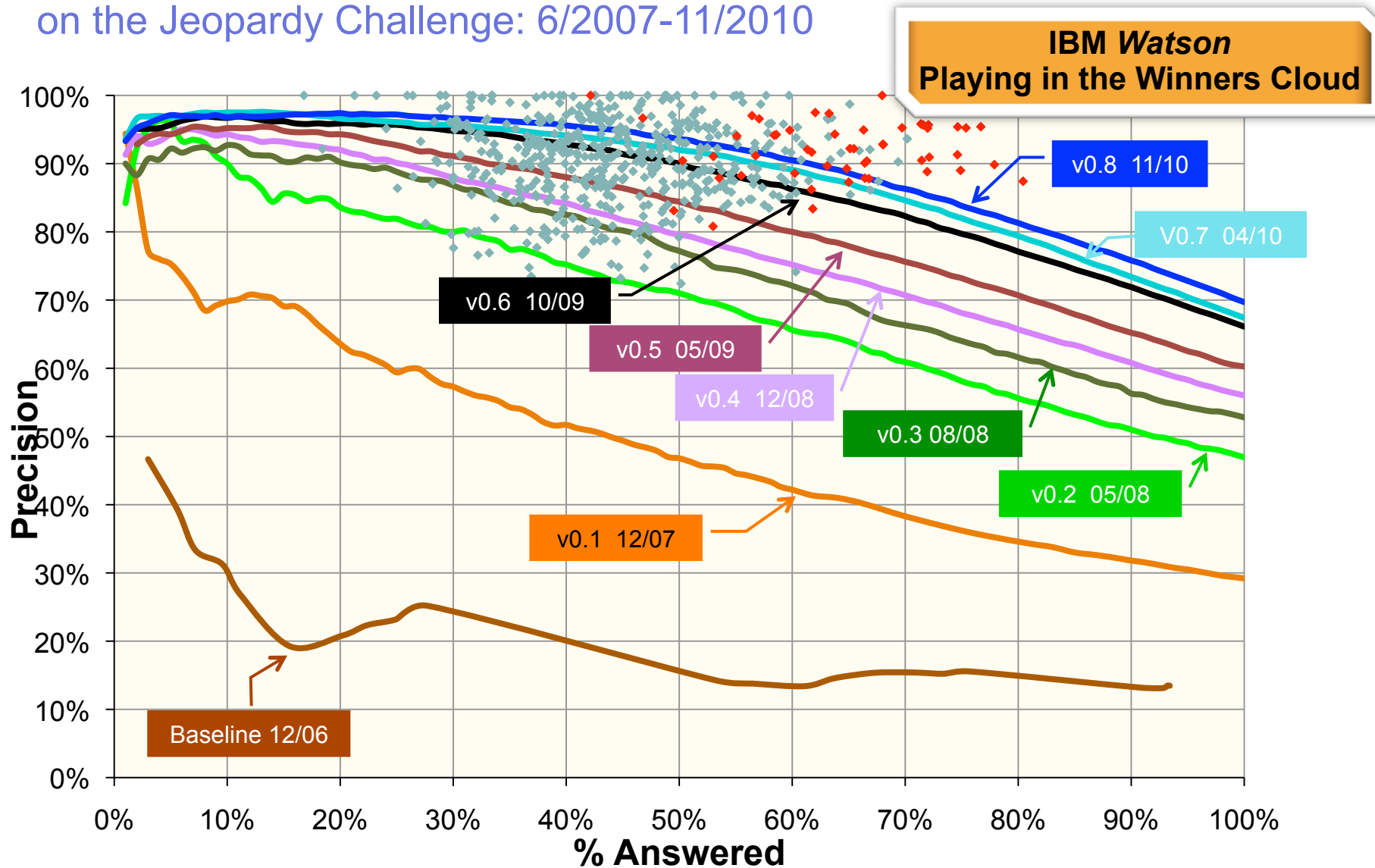
In-Category Learning

CELEBRATIONS OF THE MONTH

- What the Jeopardy! Clue is asking for is NOT always obvious
- Watson can try to infer the type of thing being asked for from the previous answers.
- In this example after seeing 2 correct answers Watson starts to dynamically learn a confidence that the question is asking for something that it can classify as a “**month**”.

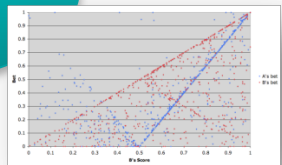
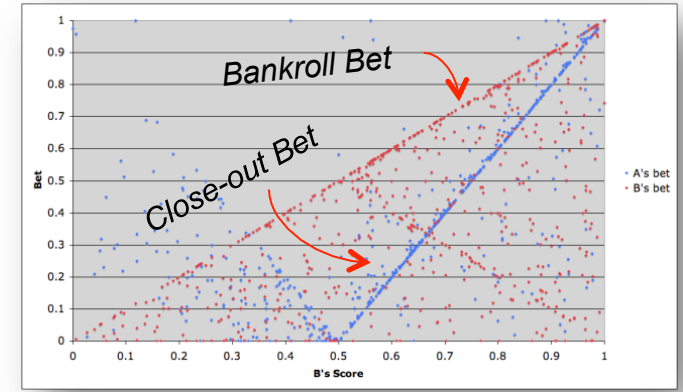
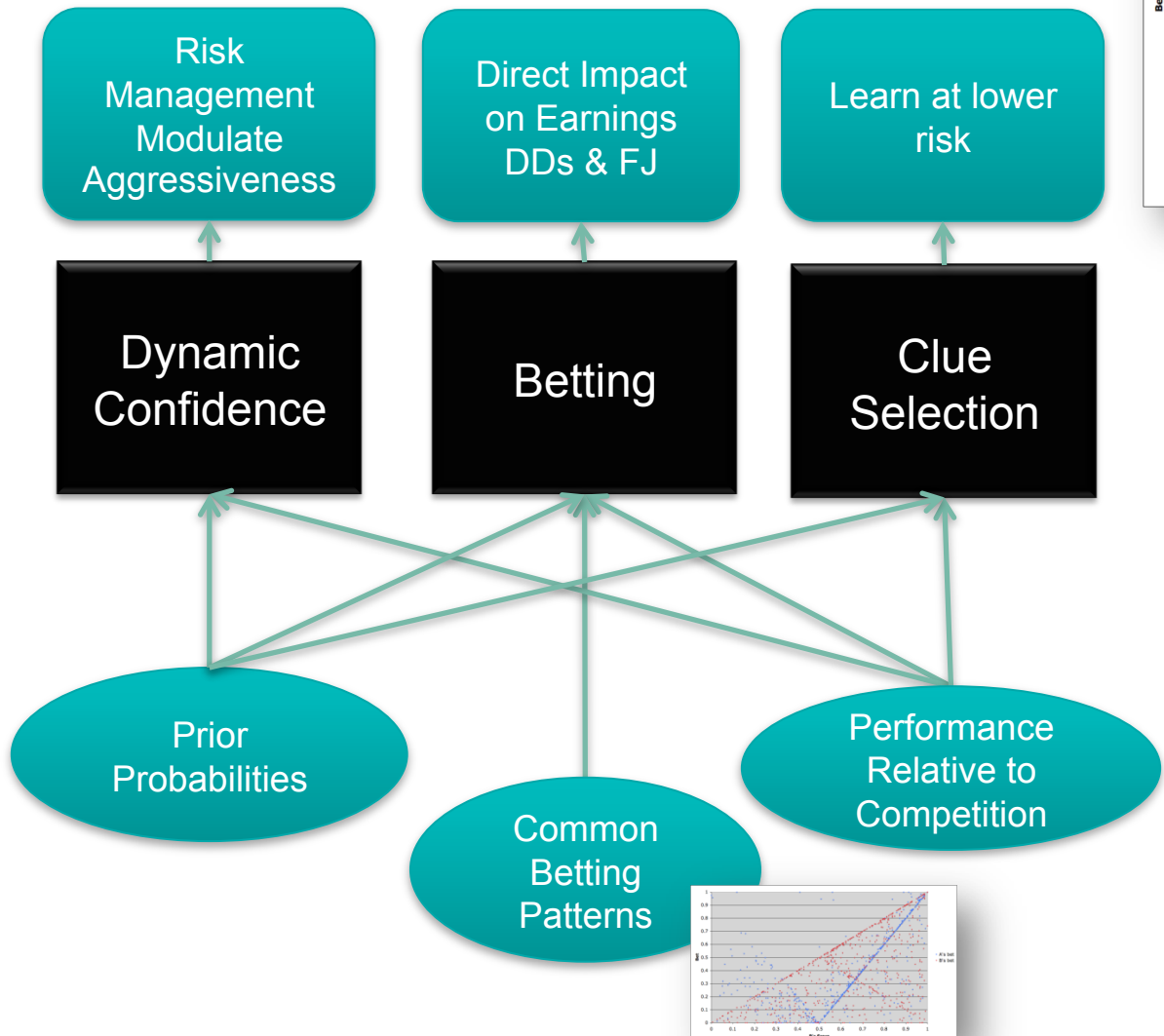
Clue	Type	Watson's Answer	Correct Answer
D-DAY ANNIVERSARY & MAGNA CARTA DAY	day	Runnymede	June
NATIONAL PHILANTHROPY DAY & ALL SOULS' DAY	day	Day of the Dead	November
NATIONAL TEACHER DAY & KENTUCKY DERBY DAY	Day/month(.2)	Churchill Downs	May
ADMINISTRATIVE PROFESSIONALS DAY & NATIONAL CPAS GOOF-OFF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

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

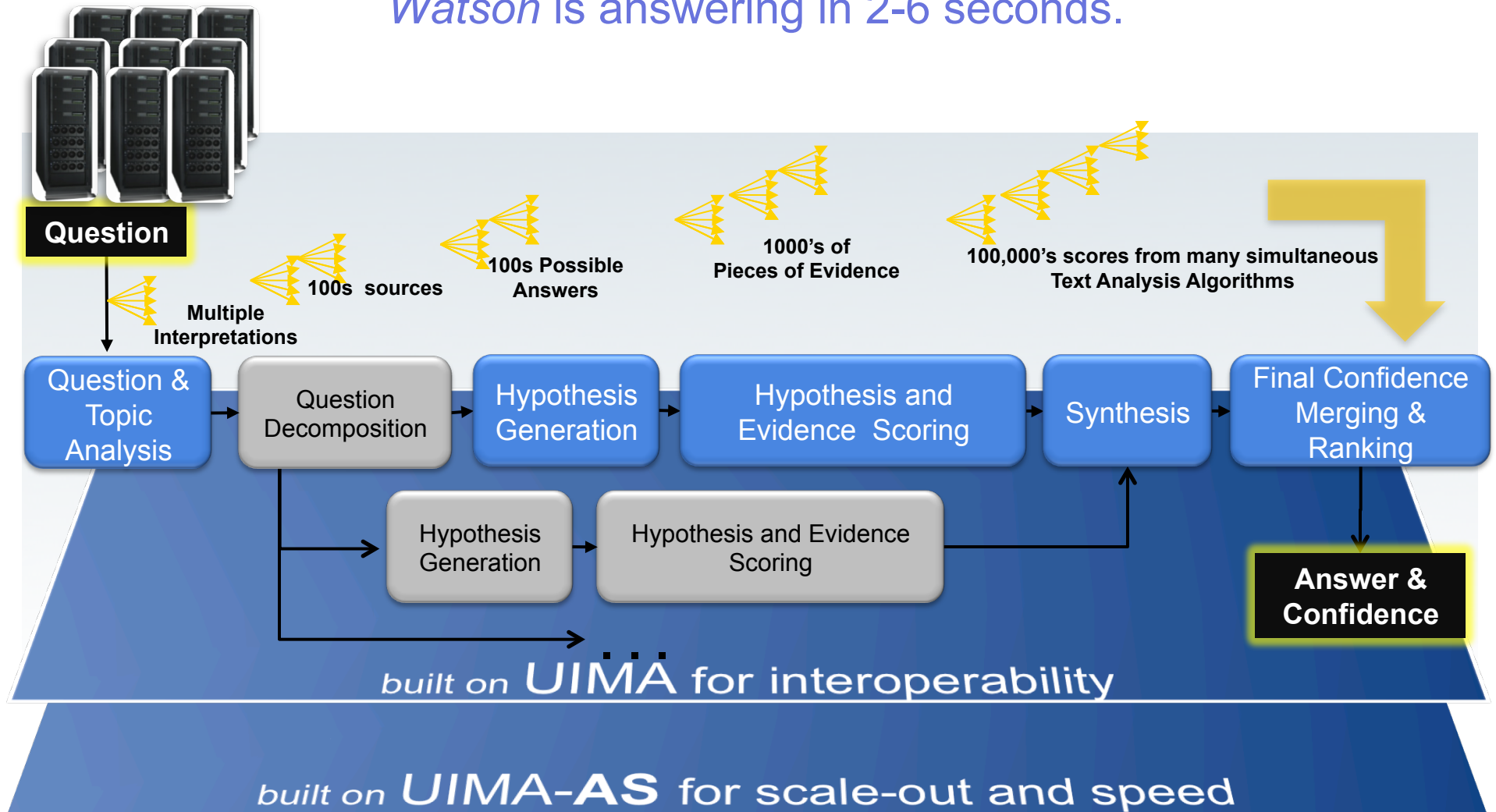


Game Strategy

Managing the Luck of the Draw

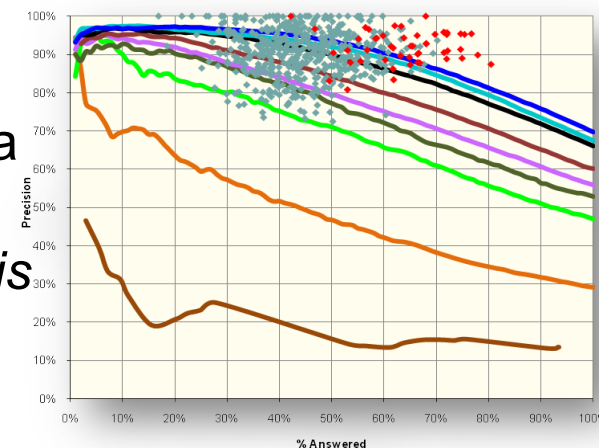
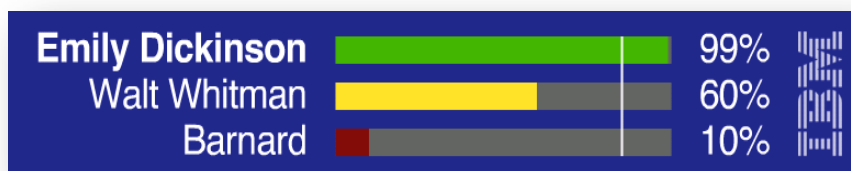


One Jeopardy! question can take **2 hours on a single 2.6Ghz Core**
Optimized & Scaled out on 2,880-Core Power750 using UIMA-AS,
Watson is answering in 2-6 seconds.



Precision, Confidence & Speed

- Deep Analytics** – Combining many analytics in a novel architecture, we achieved very high levels of *Precision* and *Confidence* over a huge variety of *as-is* content.



- Speed** – By optimizing Watson's computation for Jeopardy! on over **2,800 POWER7** processing cores we went **from 2 hours per question on a single CPU to an average of just 3 seconds.**



- Results** – in 55 real-time sparring games against former **Tournament of Champion Players last year**, Watson put on a very competitive performance in all games -- placing 1st in 71% of the them!

Next Steps

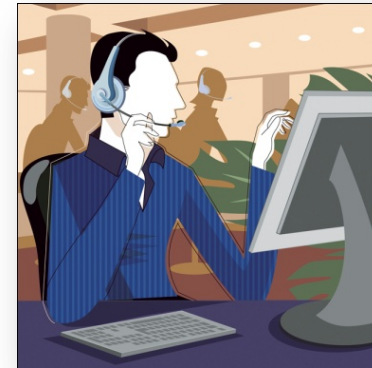
- Demonstrate ***Watson's Champion-Level*** Performance
- Broader Scientific Research: Open-Advancement of QA
 - Publish Detailed Experimental Results – What worked, What failed and why
 - Facilitate Broader Collaborative Research with Universities
 - Advance a common architecture and platform for intelligent QA systems
- Business Applications
 - Work with clients to understand and demonstrate how to impact real-world Business Applications with DeepQA technology

Potential Business Applications



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

Tech Support: Help-desk, Contact Centers



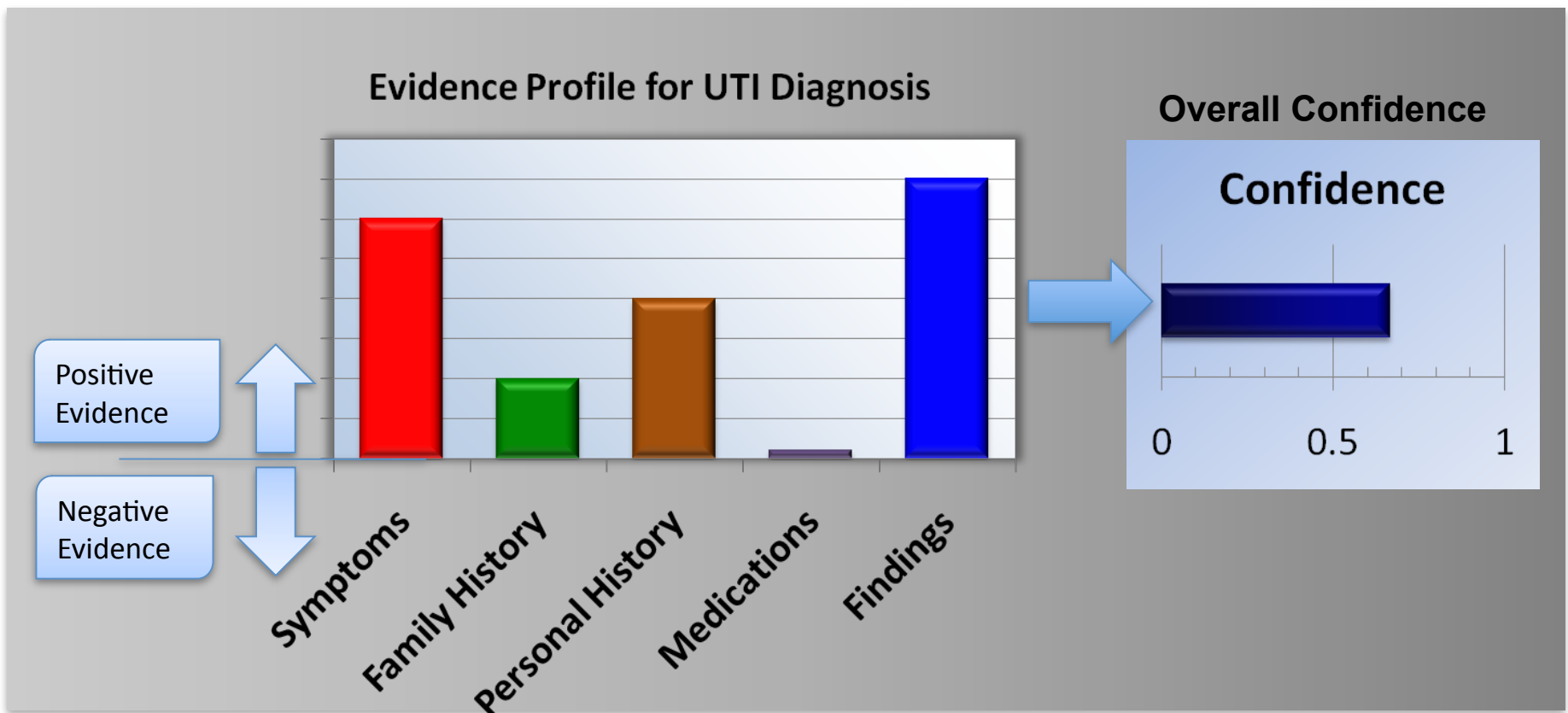
Enterprise Knowledge Management and Business Intelligence

Government: Improved Information Sharing and Security



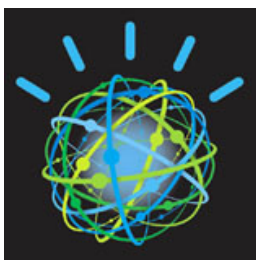
Evidence Profiles from disparate data is a powerful idea

- 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



Watson and Power Systems

- **Watson represents the future of systems design**, where workload optimized systems are tailored to fit the requirements of a new era of smarter solutions.
- Watson is a workload optimized system designed for complex analytics, made possible by integrating massively parallel POWER7 processors, Linux and DeepQA software to answer Jeopardy! questions in under three seconds.
- Building Watson on commercially available POWER7 systems and Linux ensures the acceleration of businesses adopting workload optimized systems in industries where knowledge acquisition and analytics are important.
 - Rice University uses the Power 755 with Linux for the massive analytics processing behind its cancer research program.
 - GHY International, a provider of U.S. and Canadian customs brokerage services, uses the Power 750 with AIX, IBM i and Linux to deploy new business services in as little as five minutes.



Power 750



Watson Workload Optimized System (Power 750)

- 90 x IBM Power 750¹ servers
- 2880 POWER7 cores
- POWER7 3.55 GHz chip
- 500 GB per sec on-chip bandwidth
- 10 Gb Ethernet network
- 15 Terabytes of memory
- 20 Terabytes of disk, clustered
- Can operate at 80 Teraflops
- Runs IBM DeepQA software
- Scales out with and searches vast amounts of unstructured information with UIMA & Hadoop open source components
- SUSE Linux provides a cost-effective open platform which is performance-optimized to exploit POWER 7 systems
- 10 racks include servers, networking, shared disk system, cluster controllers



¹ Note that the Power 750 featuring POWER7 is a commercially available server that runs AIX, IBM i and Linux and has been in market since Feb 2010

In summary, Watson is good but...

- Watson

- 10 compute racks
- 80kW of power
- 20 tons of cooling

- Human

- 1 brain - fits in a shoe box
- Can run on a tuna-fish sandwich
- Can be cooled with a hand-held paper fan.

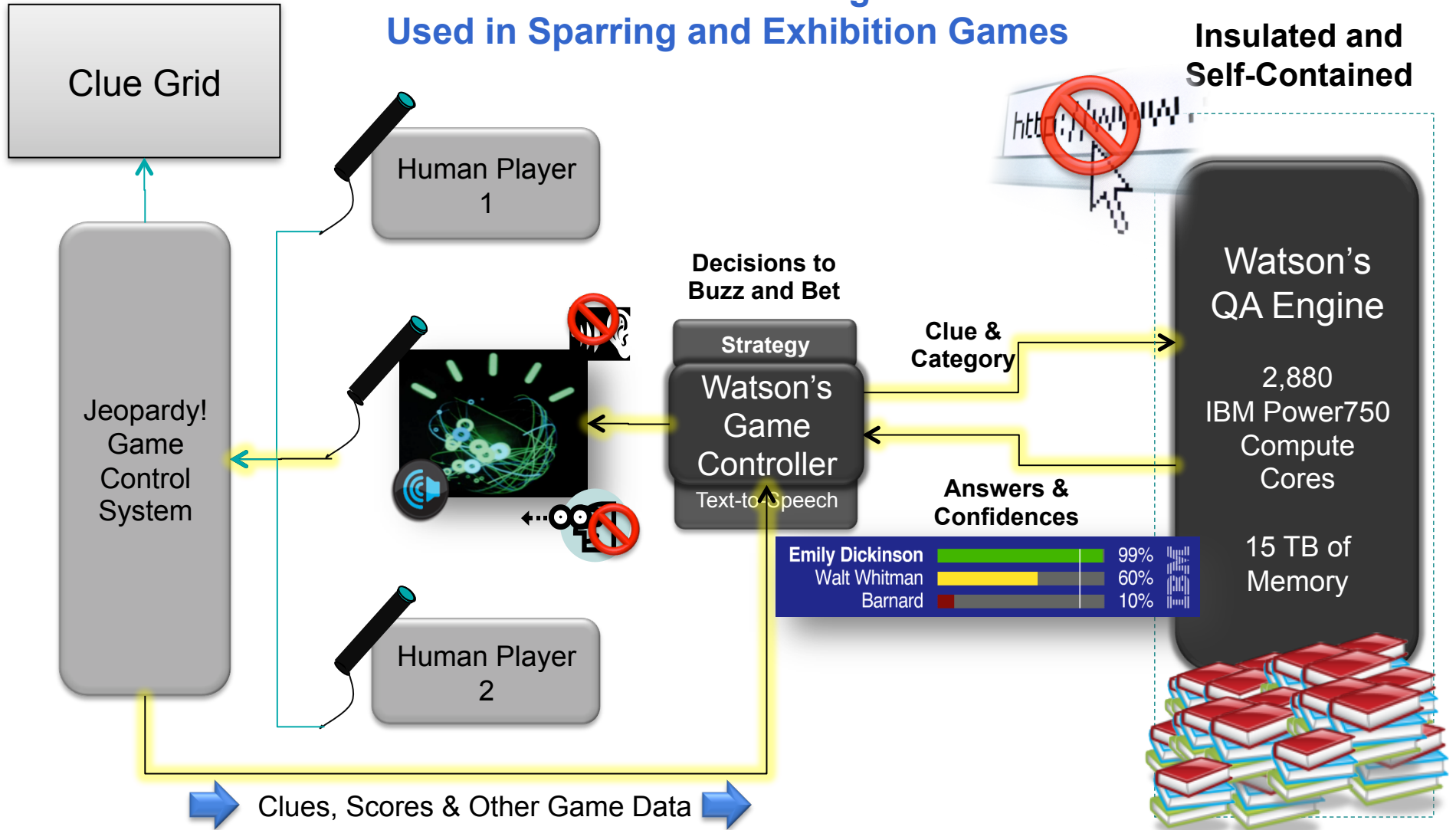
More info:

http://w3.ibm.com/ibm/resource/res_watson_grand_challenge.html

Questions?

BACK UPS

Real-Time Game Configuration Used in Sparring and Exhibition Games



Analysis of natural language content equivalent to 1 Million Books

Watson Does and Does NOT

Does

- Speaks Clue Selections
- Speaks Responses
- Physically Presses the Button
- Gets the clue electronically when you see it

Does NOT

- Hear
 - Can answer with same wrong answer
- See
- Process Audi/Visual Clues
 - These are excluded from the contest
- Buzz *Perfectly*
 - Watson is quick but not the quickest
 - Variable time to compute confidences and answers
 - Confidence-Weighed Buzzing
 - Not always fast or confident enough

ABSTRACT

The TV quiz show Jeopardy! is famous for providing contestants answers to which they must supply the correct questions in order to win. Contestants must be fast, have an almost encyclopedic knowledge of the world, and they must be able to handle clues that are vague, involve double-meanings and frequently rely on puns. Early this year, Jeopardy! aired a match involving the two all-time most successful Jeopardy! contestants and Watson, a artificial intelligence system designed by IBM. Watson won the Jeopardy! match by a wide margin and in doing so, brought the leading edge of computer technology a little closer to human abilities. This presentation will describe the supercomputer implementation of Watson use for the Jeopardy! match and the challenges overcome to create a computer capable of accurately answering open-ended natural language questions in real-time - typically in under 3 seconds.