Engineering Challenges in Building Watson

A talk on DeepQA and the Jeopardy! Challenge

Chris Welty
DeepQA Team @ IBM Research

© 2009 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)
Barbara Cutler (RPI)	Boris Katz (MIT)	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.

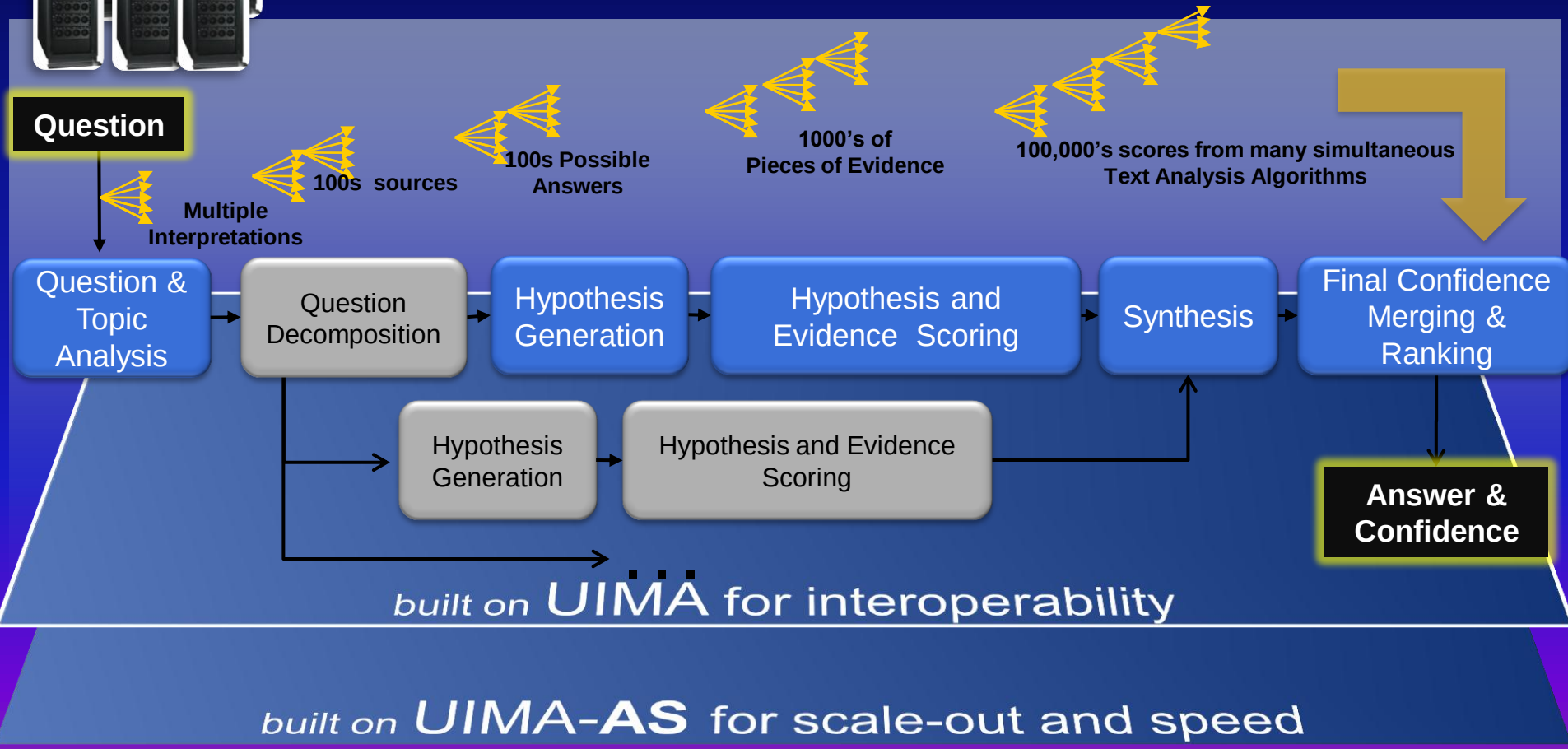
*NOT full-time Equivalents. Names listed if contributed some time to that part of project.

Premise:
80% of Human Knowledge *is in natural language text*

Research Mission:
How can we exploit this knowledge to solve problems?

© 2009 IBM Corporation

One Jeopardy! question can take **2 hours on a single 2.6Ghz Core** Optimized & Scaled out on 3000-Core IBM HPC using UIMA-AS, *Watson* is answering in ~2 seconds.



© 2010 IBM Corporation

Watson – a Workload Optimized System

- 90 x IBM Power 750[1] 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
- Linux provides a scalable, open platform, optimized to exploit POWER7 performance
- 10 racks include servers, networking, shared disk system, cluster controllers

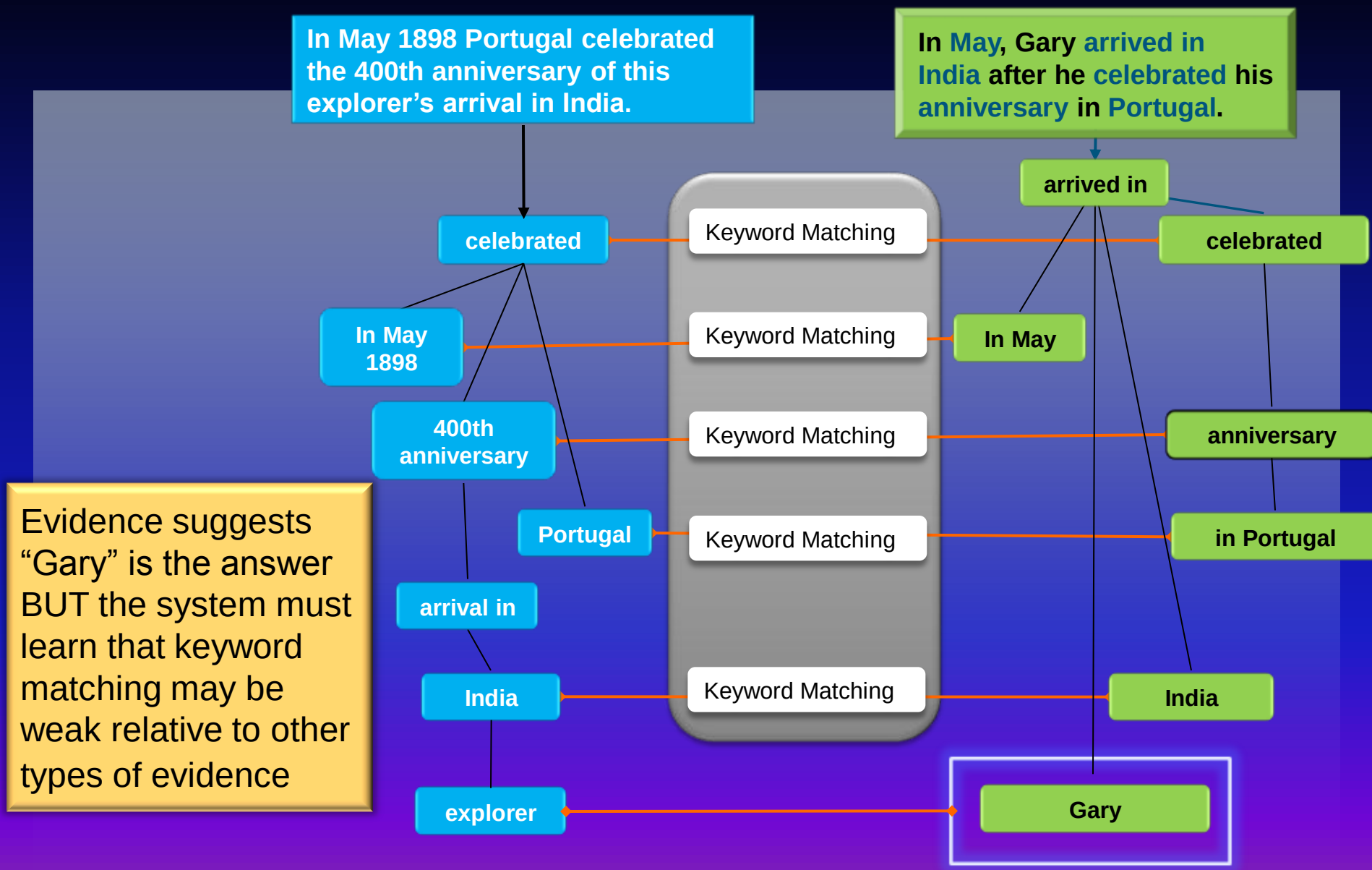
Linux

POWER7™ BUILT ON Power™

[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

© 2010 IBM Corporation

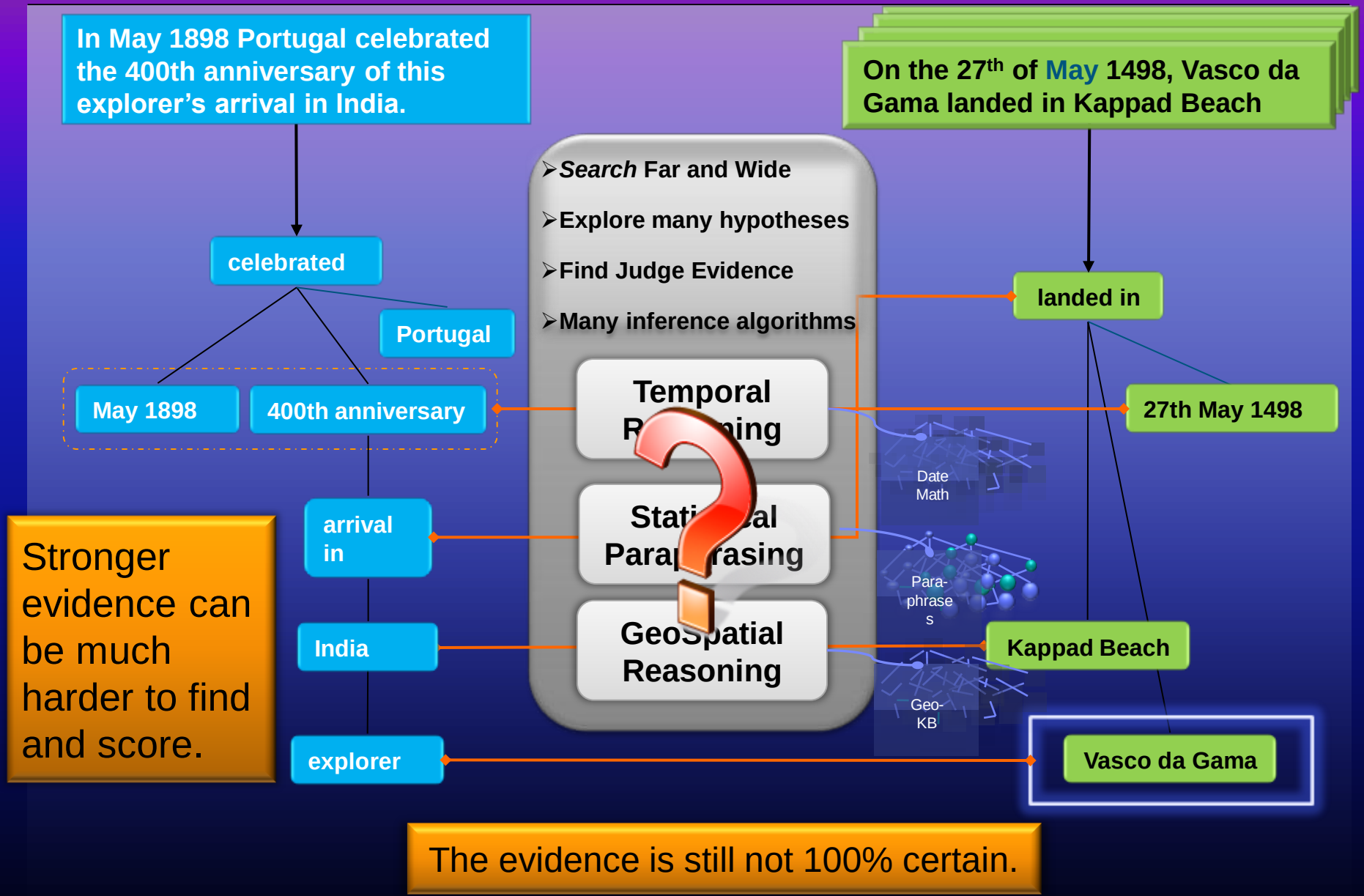
Different Types of Evidence: Keyword Evidence



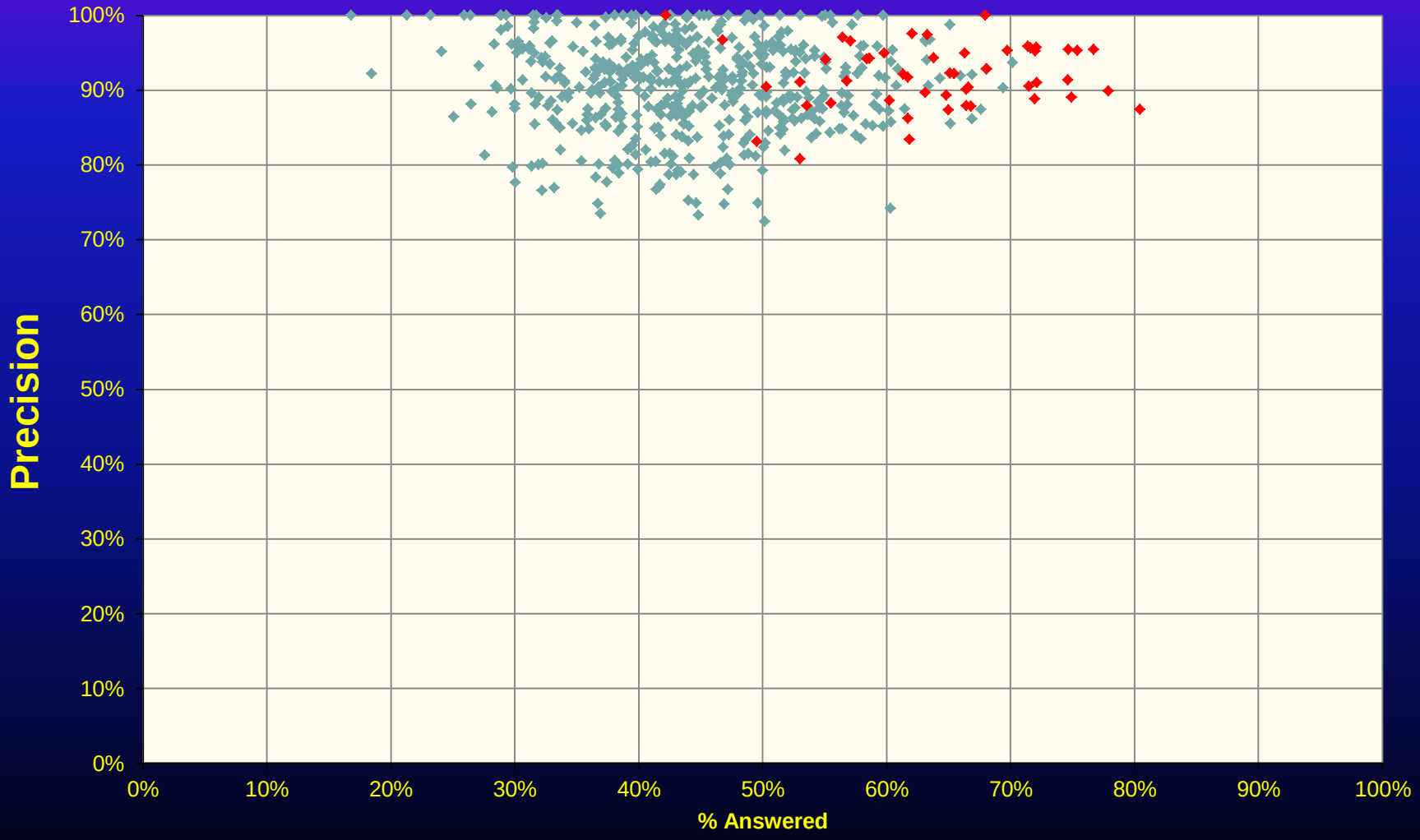
Evidence suggests "Gary" is the answer BUT the system must learn that keyword matching may be weak relative to other types of evidence

© 2010 IBM Corporation

Different Types of Evidence: Hybridization

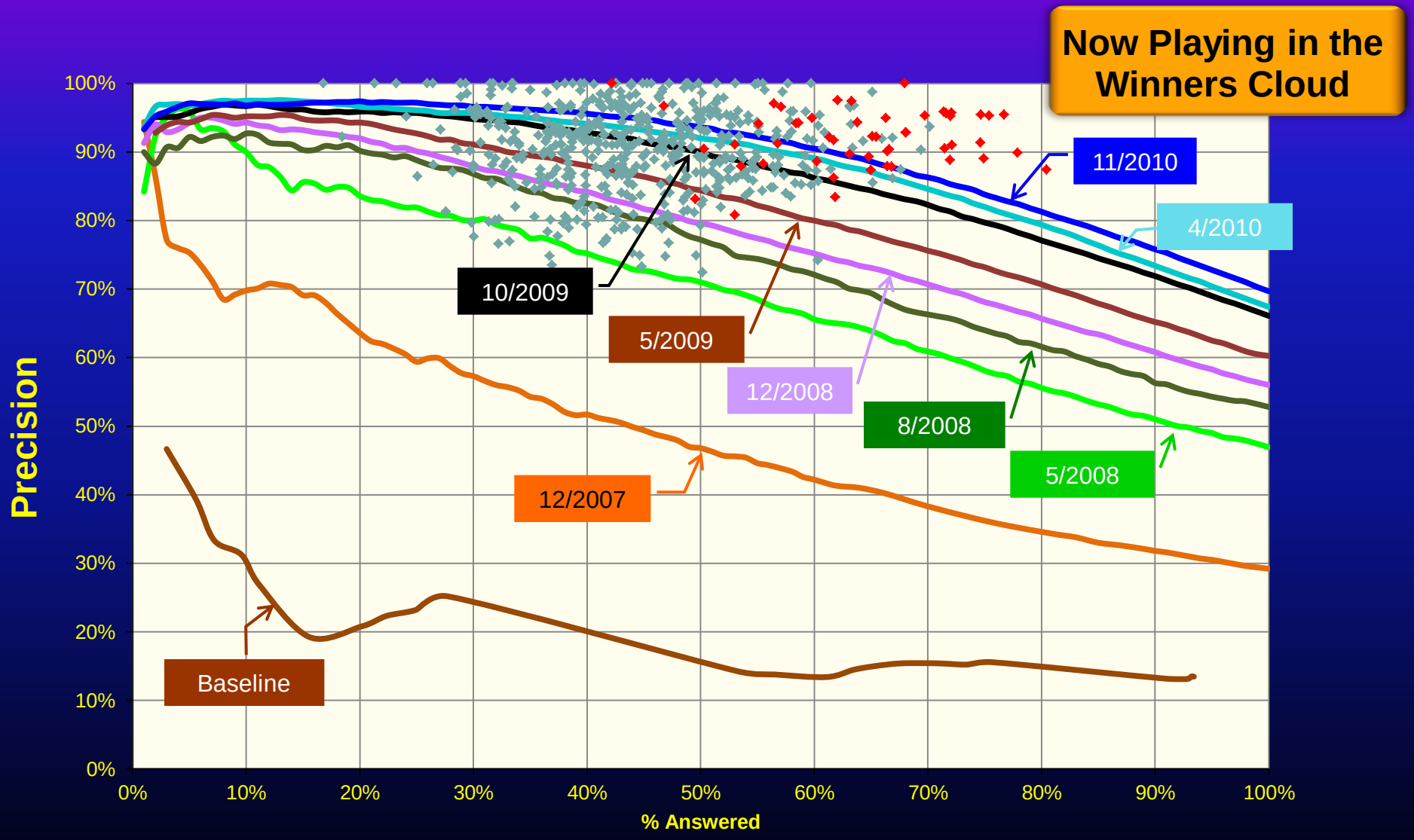


Cognitive Computing: Performance against a metric



© 2010 IBM Corporation

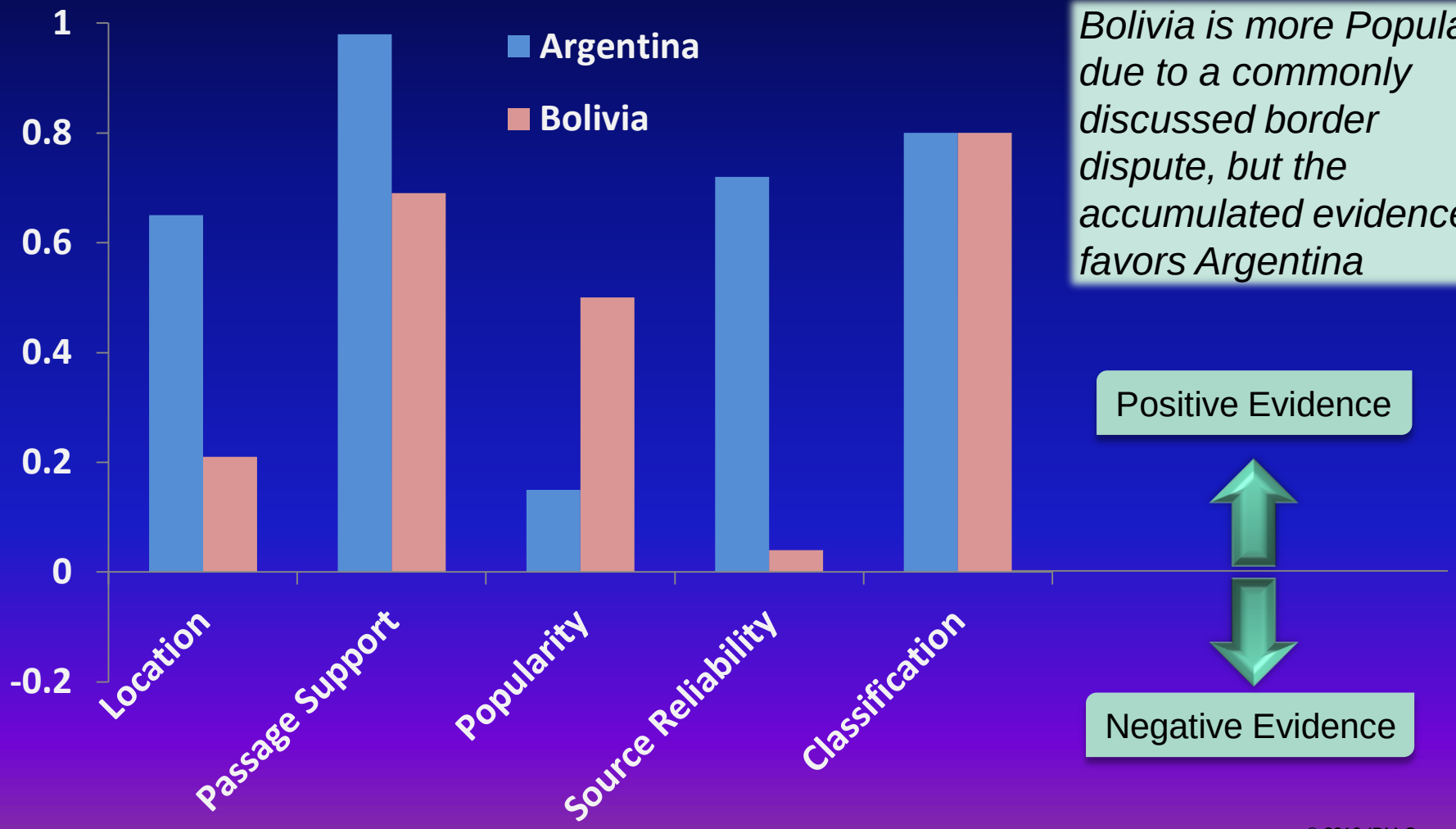
Deep QA: Incremental Progress 6/2007-11/2010



TECHNOLOGY
BEFORE & AFTER
LATIN AMERICA
$200
$400
$600
CHILE SHARES ITS LONGEST LAND BORDER WITH THIS COUNTRY
$200
$400
$600
ARGENTINA
$1000
$1000
$1000
$1000
$1000
$1000

Grouping features to produce Evidence Profiles

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

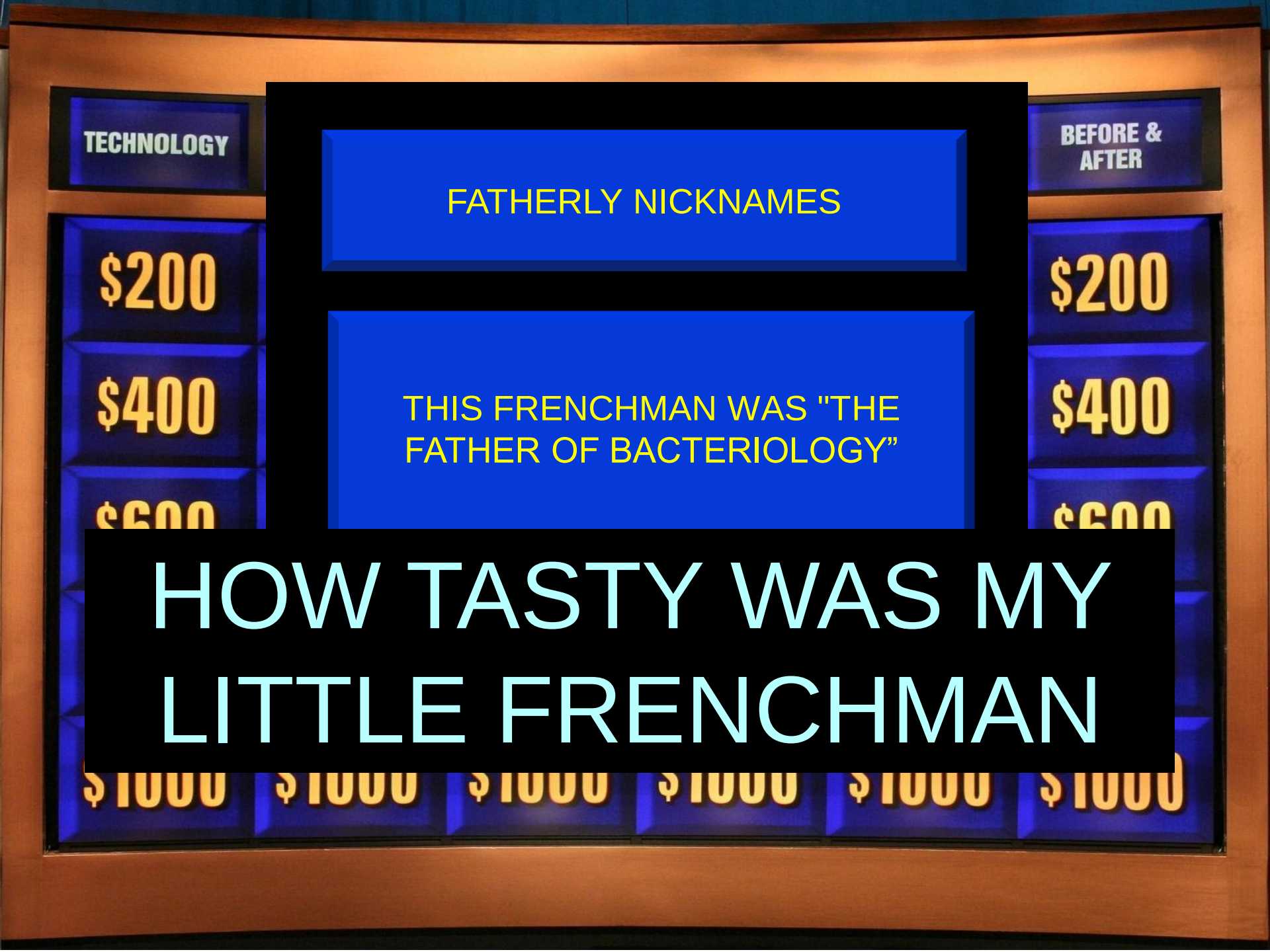


Bolivia is more Popular due to a commonly discussed border dispute, but the accumulated evidence favors Argentina

Positive Evidence

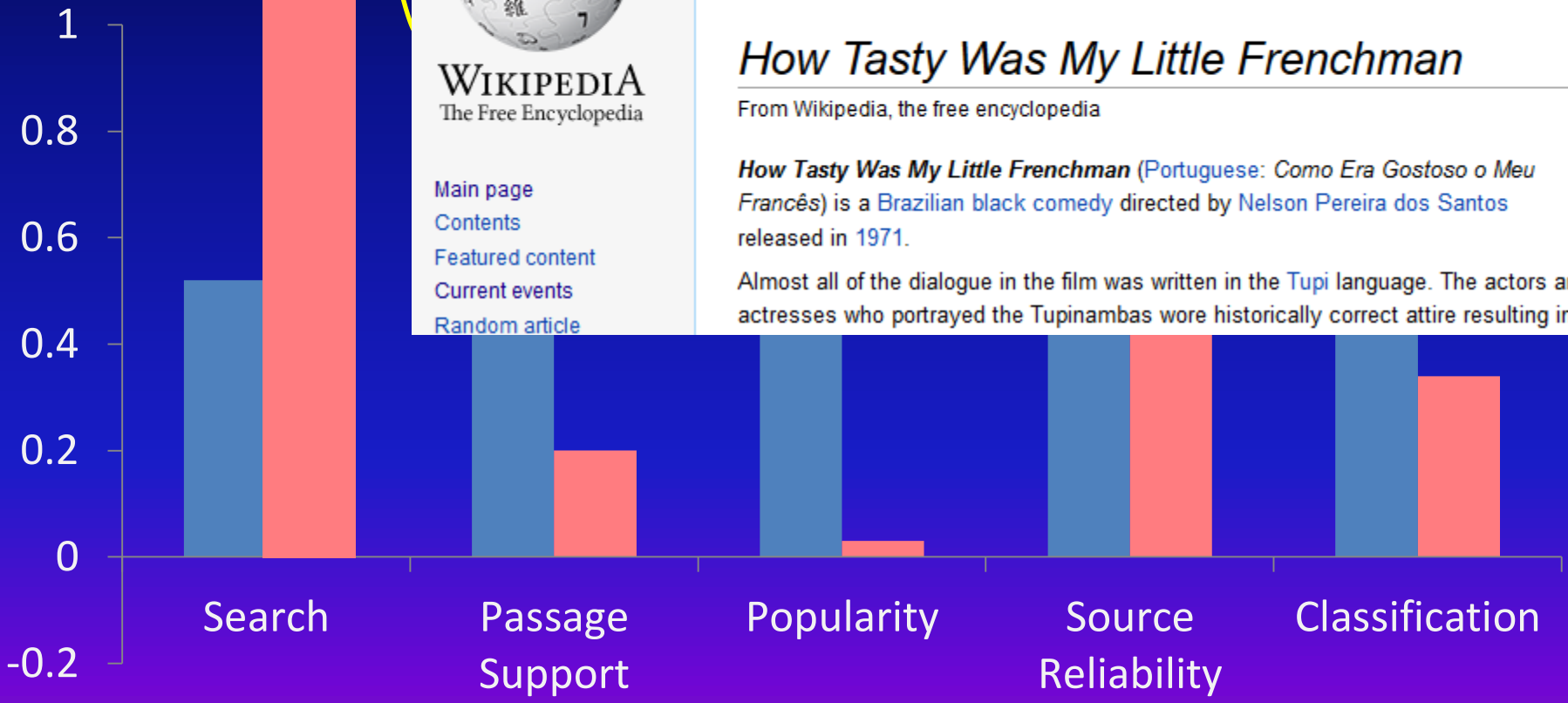
Negative Evidence

© 2010 IBM Corporation


TECHNOLOGY
BEFORE & AFTER
FATHERLY NICKNAMES
$200
$400
$200
$400
THIS FRENCHMAN WAS "THE FATHER OF BACTERIOLOGY”
HOW TASTY WAS MY LITTLE FRENCHMAN

Error Analysis: Values out of range

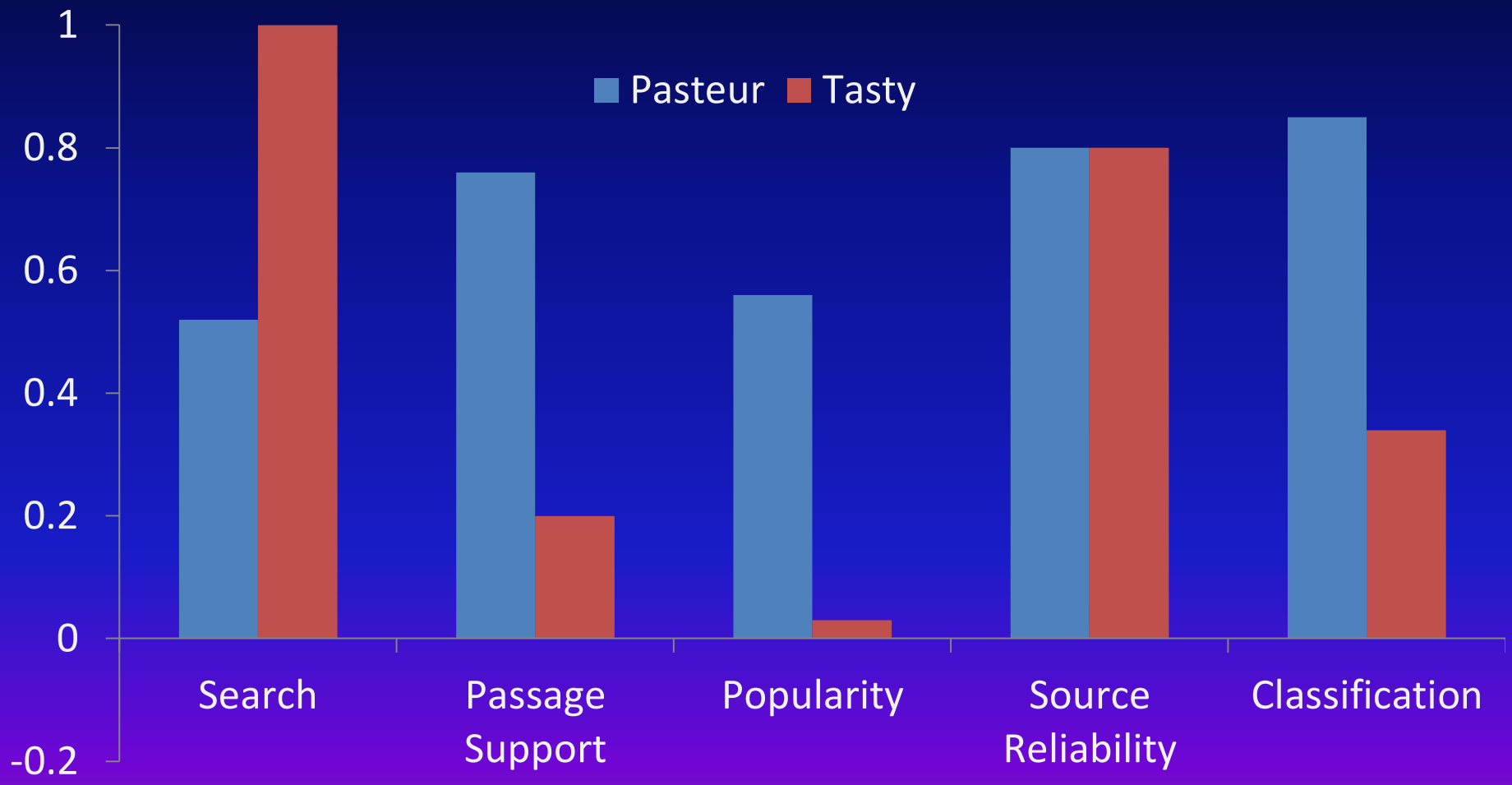
Clue: This Frenchman was "The Father of Bacteriology".



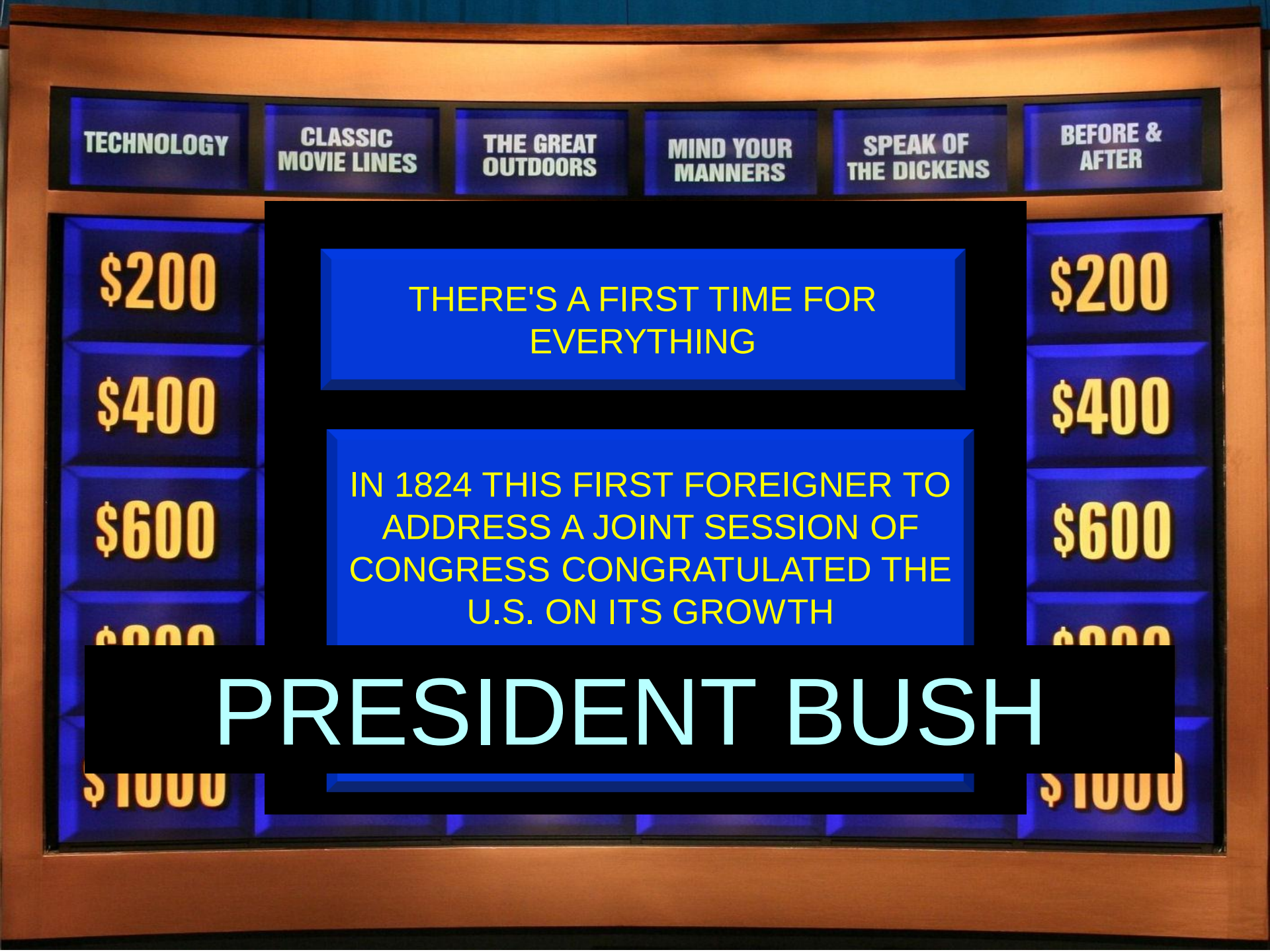
© 2010 IBM Corporation

Solution: Cap values to max seen in training

Clue: This Frenchman was "The Father of Bacteriology".

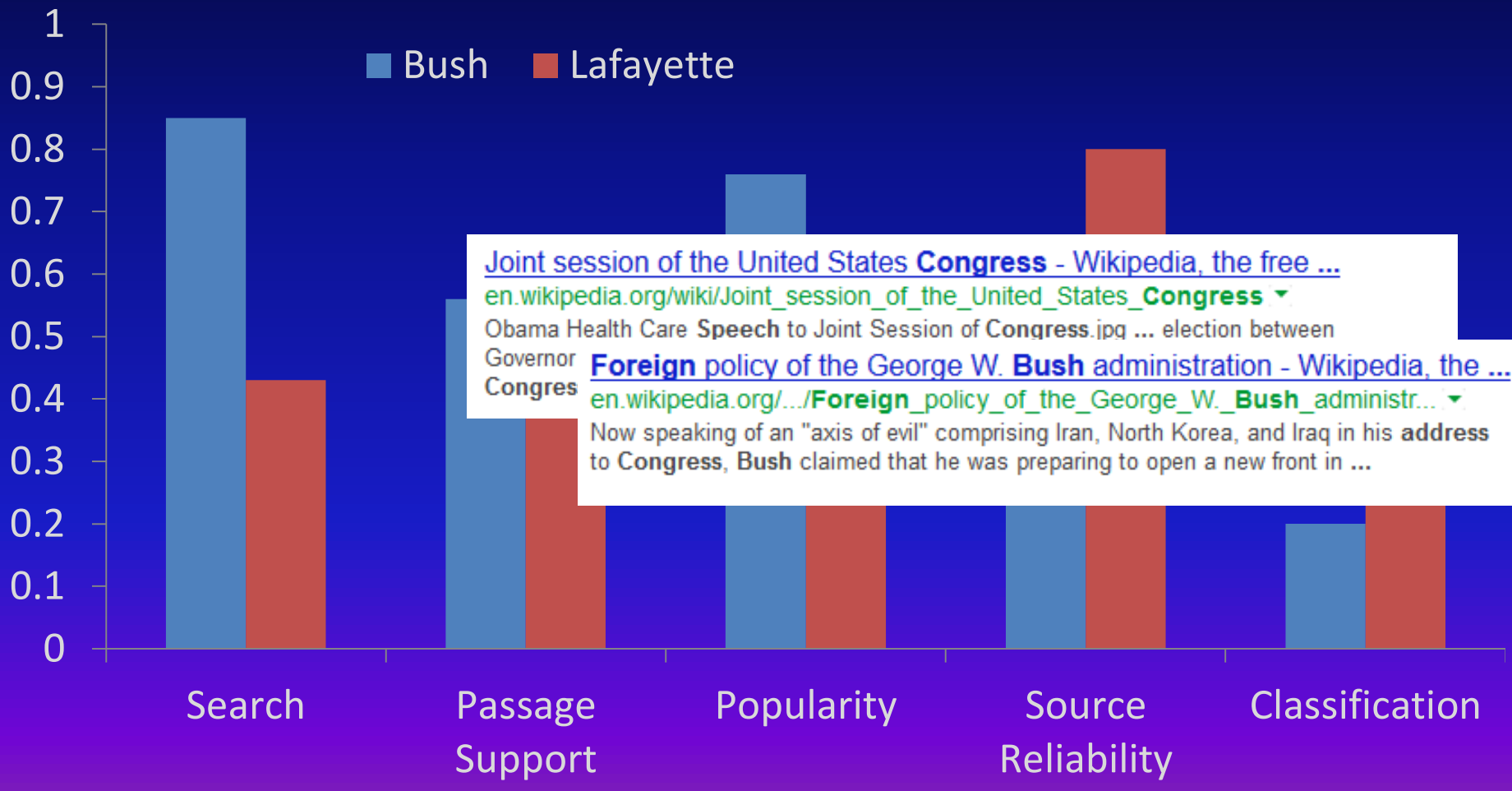


© 2010 IBM Corporation


TECHNOLOGY
CLASSIC MOVIE LINES
THE GREAT OUTDOORS
MIND YOUR MANNERS
SPEAK OF THE DICKENS
BEFORE & AFTER
$200
$400
$600
$1000
THERE'S A FIRST TIME FOR EVERYTHING
IN 1824 THIS FIRST FOREIGNER TO ADDRESS A JOINT SESSION OF CONGRESS CONGRATULATED THE U.S. ON ITS GROWTH
PRESIDENT BUSH
$200
$400
$600
$1000

Error Analysis: Temporal Relevance Missing

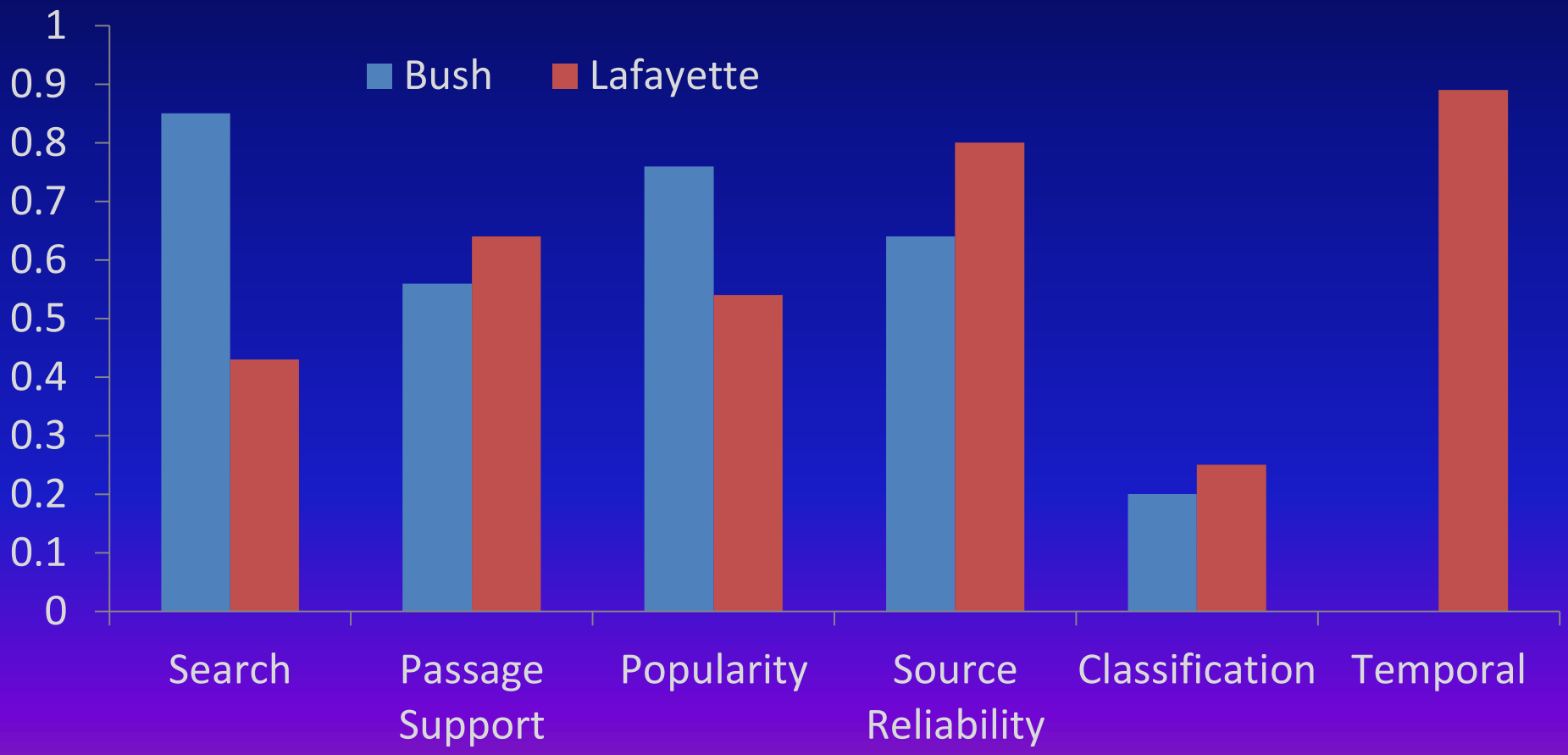
Clue: In 1824 this first foreigner to address a joint session of Congress congratulated the U.S. on its growth



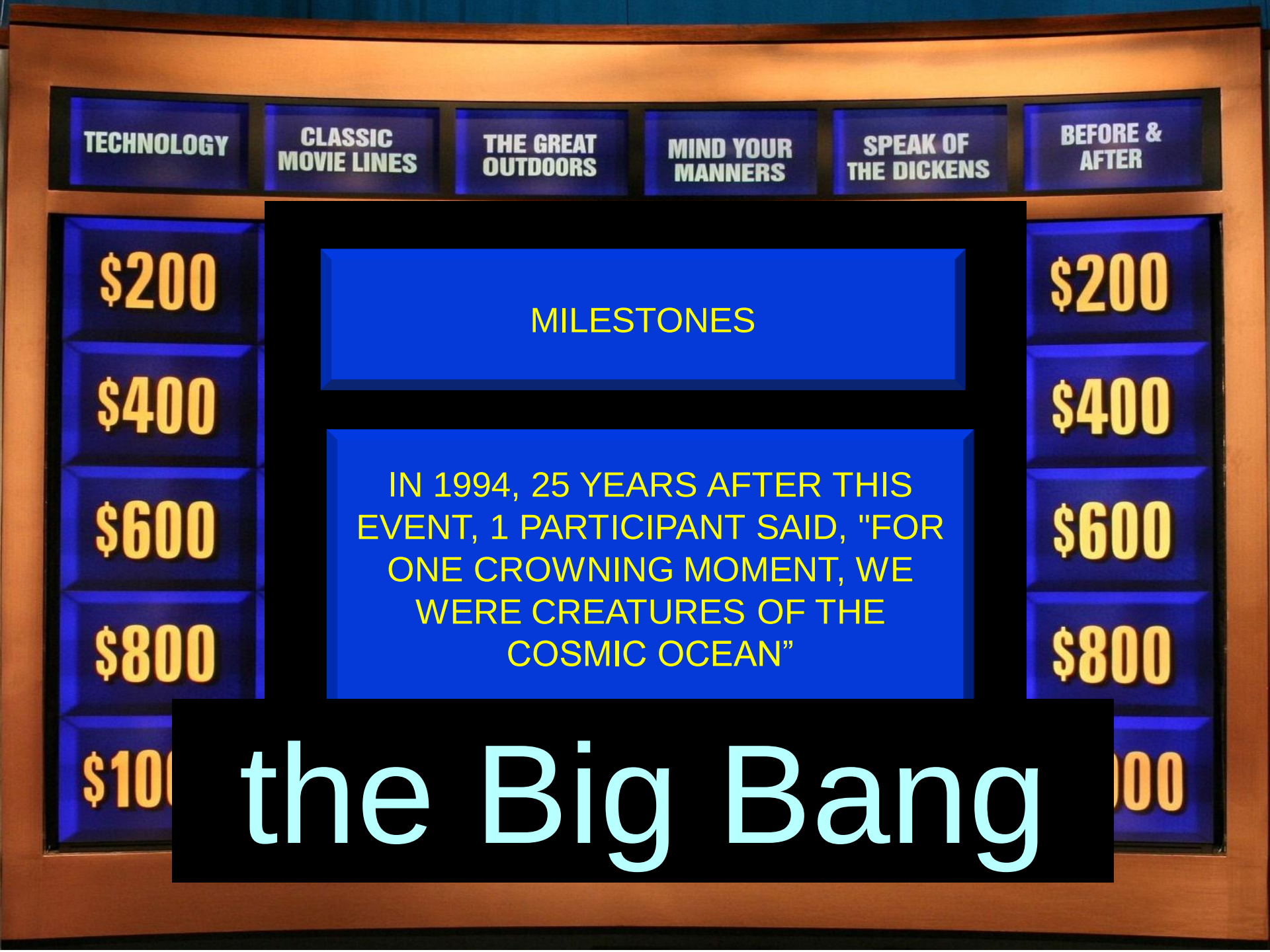
© 2010 IBM Corporation

Solution: Temporal Scoring

Clue: In 1824 this first foreigner to address a joint session of Congress congratulated the U.S. on its growth

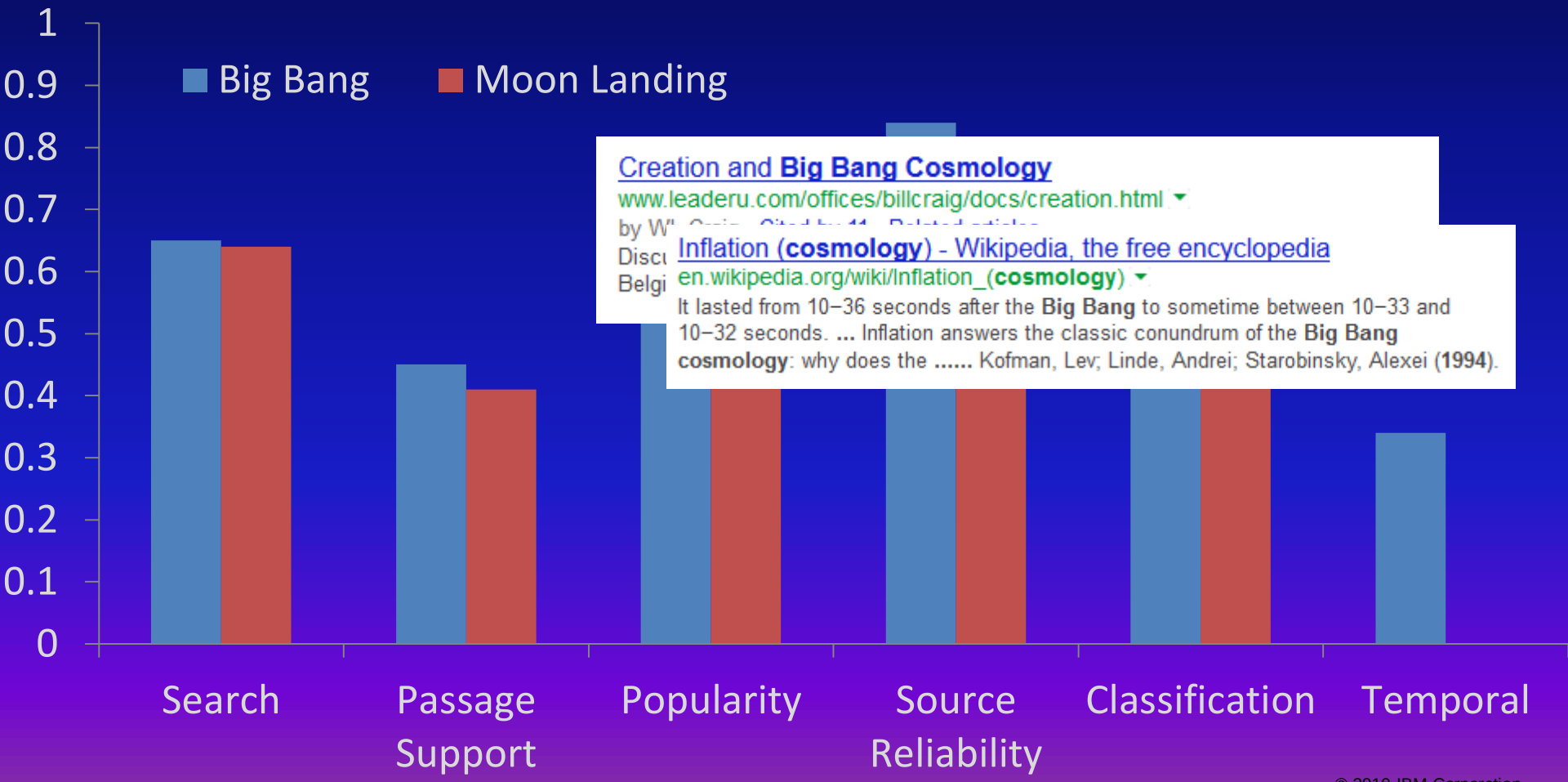


© 2010 IBM Corporation


TECHNOLOGY
CLASSIC MOVIE LINES
THE GREAT OUTDOORS
MIND YOUR MANNERS
SPEAK OF THE DICKENS
BEFORE & AFTER
$200
$400
$600
$800
$200
$400
$600
$800
MILESTONES
IN 1994, 25 YEARS AFTER THIS EVENT, 1 PARTICIPANT SAID, "FOR ONE CROWNING MOMENT, WE WERE CREATURES OF THE COSMIC OCEAN"
the Big Bang

Error Analysis: Relative Date in Clue

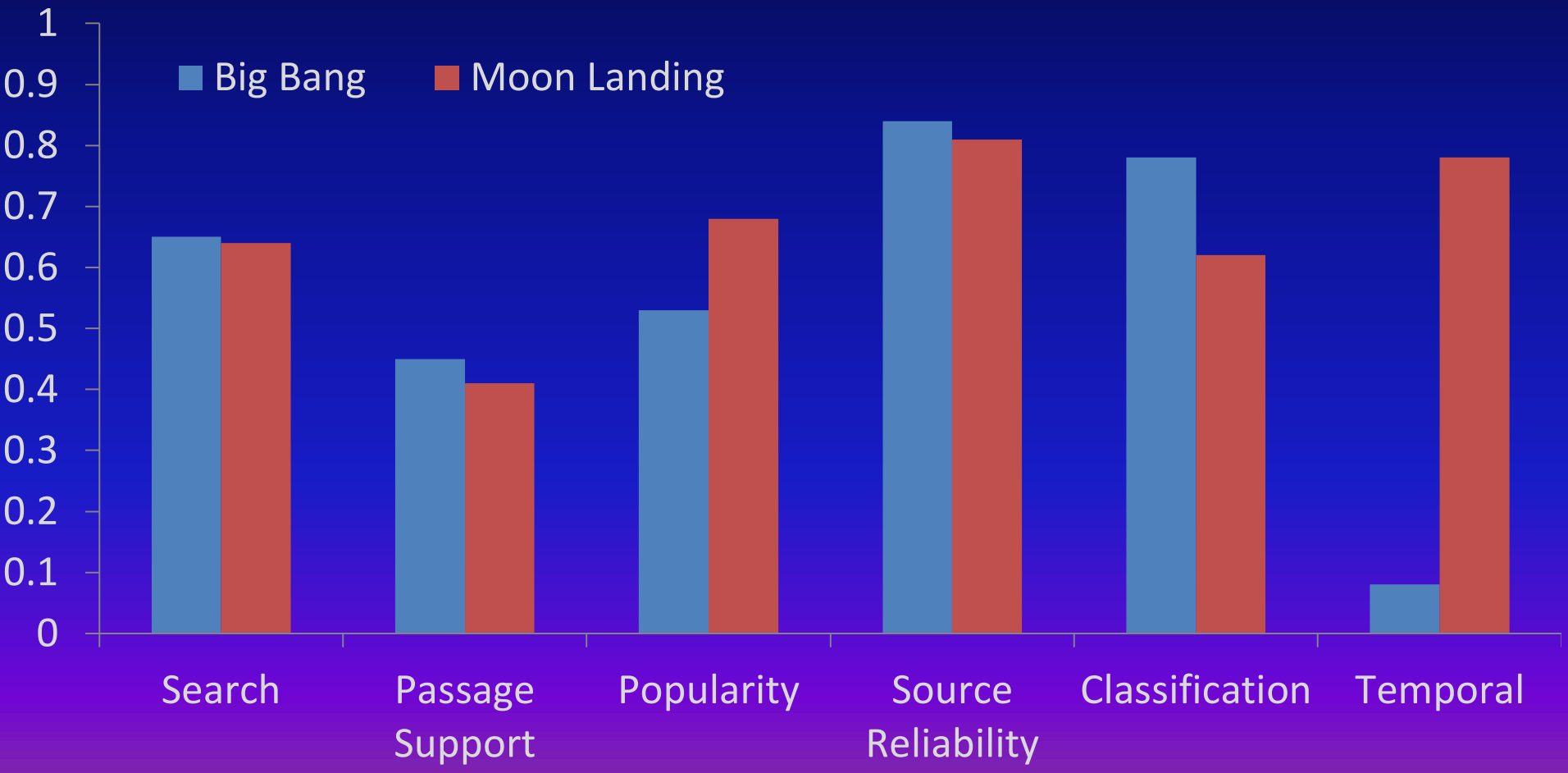
Clue: In 1994, 25 years after this event, 1 participant said, "For one crowning moment, we were creatures of the cosmic ocean"



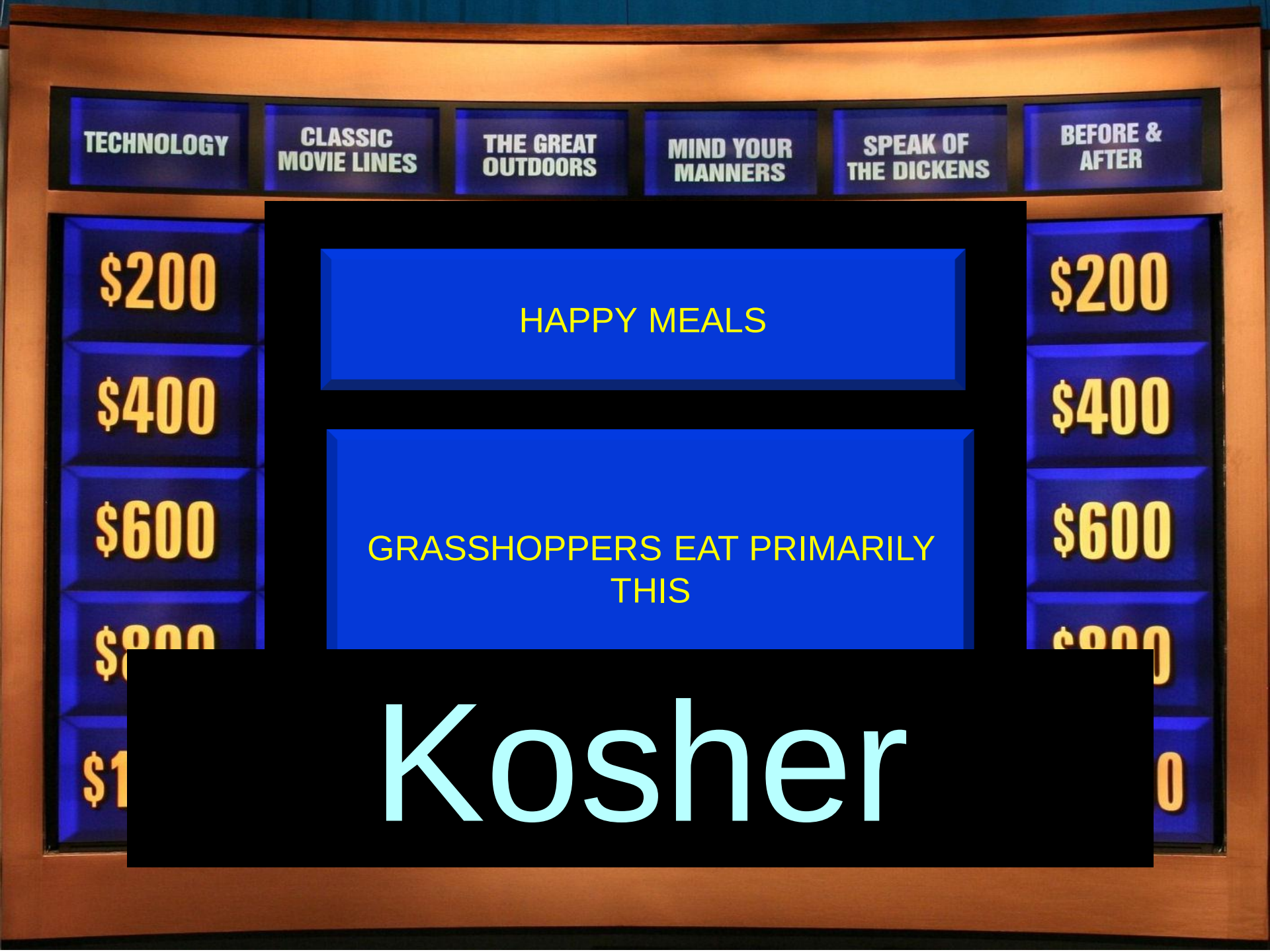
© 2010 IBM Corporation

Solution: Relative Date Recognition

Clue: In 1994, 25 years after this event, 1 participant said, "For one crowning moment, we were creatures of the cosmic ocean"

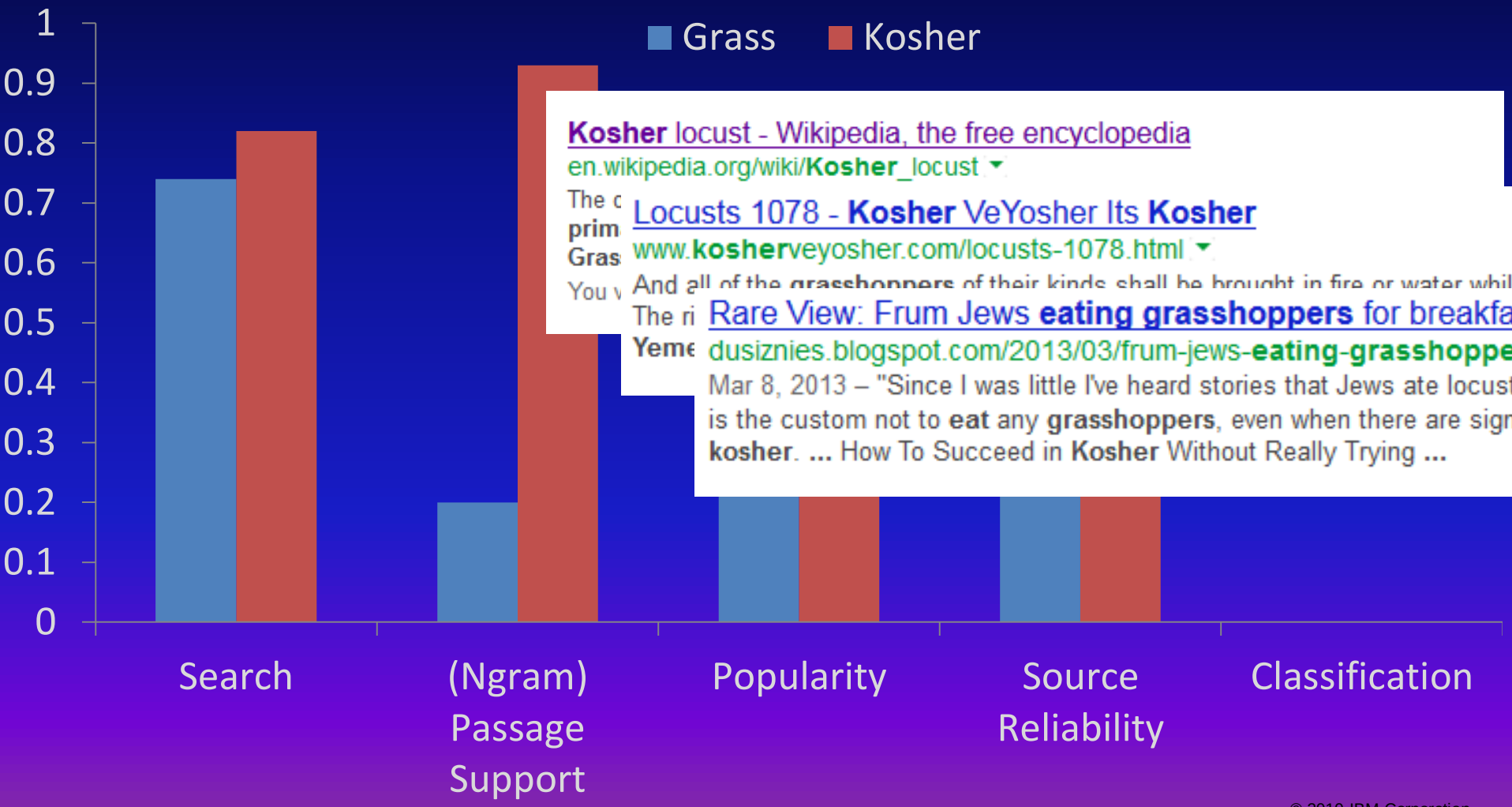


© 2010 IBM Corporation


TECHNOLOGY
CLASSIC MOVIE LINES
THE GREAT OUTDOORS
MIND YOUR MANNERS
SPEAK OF THE DICKENS
BEFORE & AFTER
$200
$400
$600
$200
$400
$600
HAPPY MEALS
GRASSHOPPERS EAT PRIMARILY THIS
Kosher

Error Analysis: shallow passage matching too strong

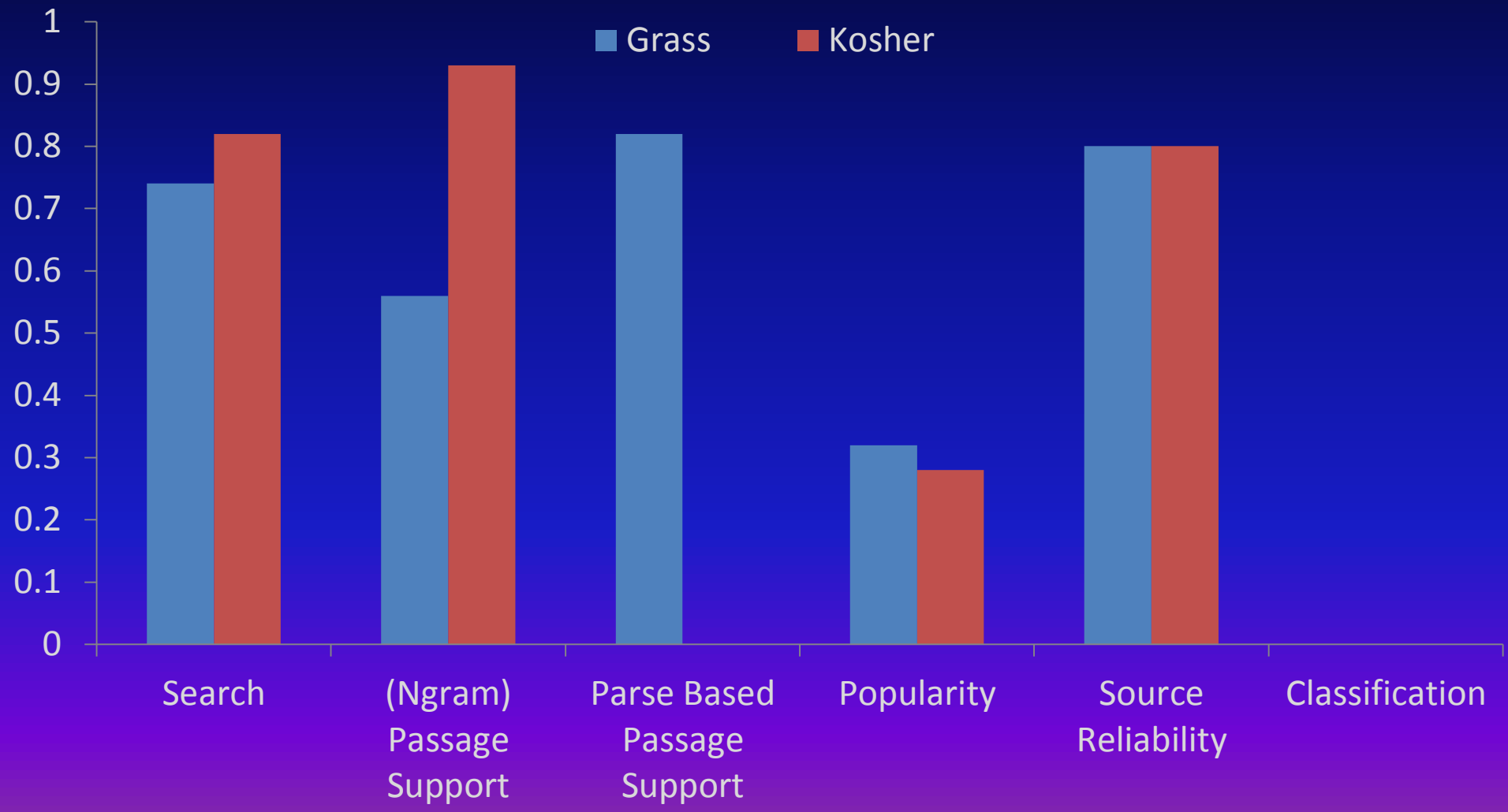
Clue: Grasshoppers eat primarily this.



© 2010 IBM Corporation

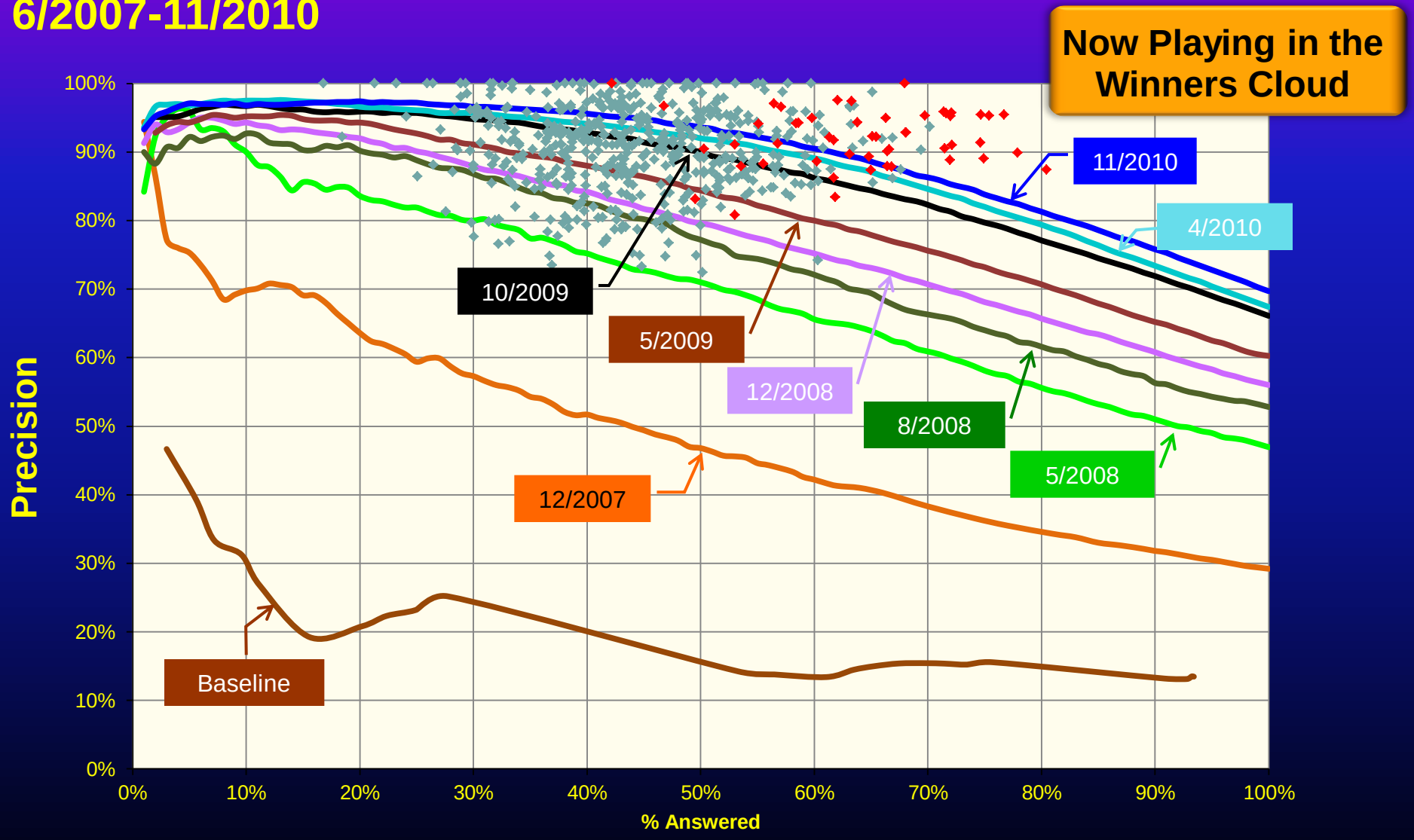
Solution: Parse-based passage matching

Clue: Grasshoppers eat primarily this.



© 2010 IBM Corporation

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



…and we won

© 2010 IBM Corporation

THANK YOU

© 2010 IBM Corporation