Watson Beyond Jeopardy!: Adaptation to the Medical Domain

Jennifer Chu-Carroll
IBM T.J. Watson Research Center

Research Work by the Watson Technologies Team
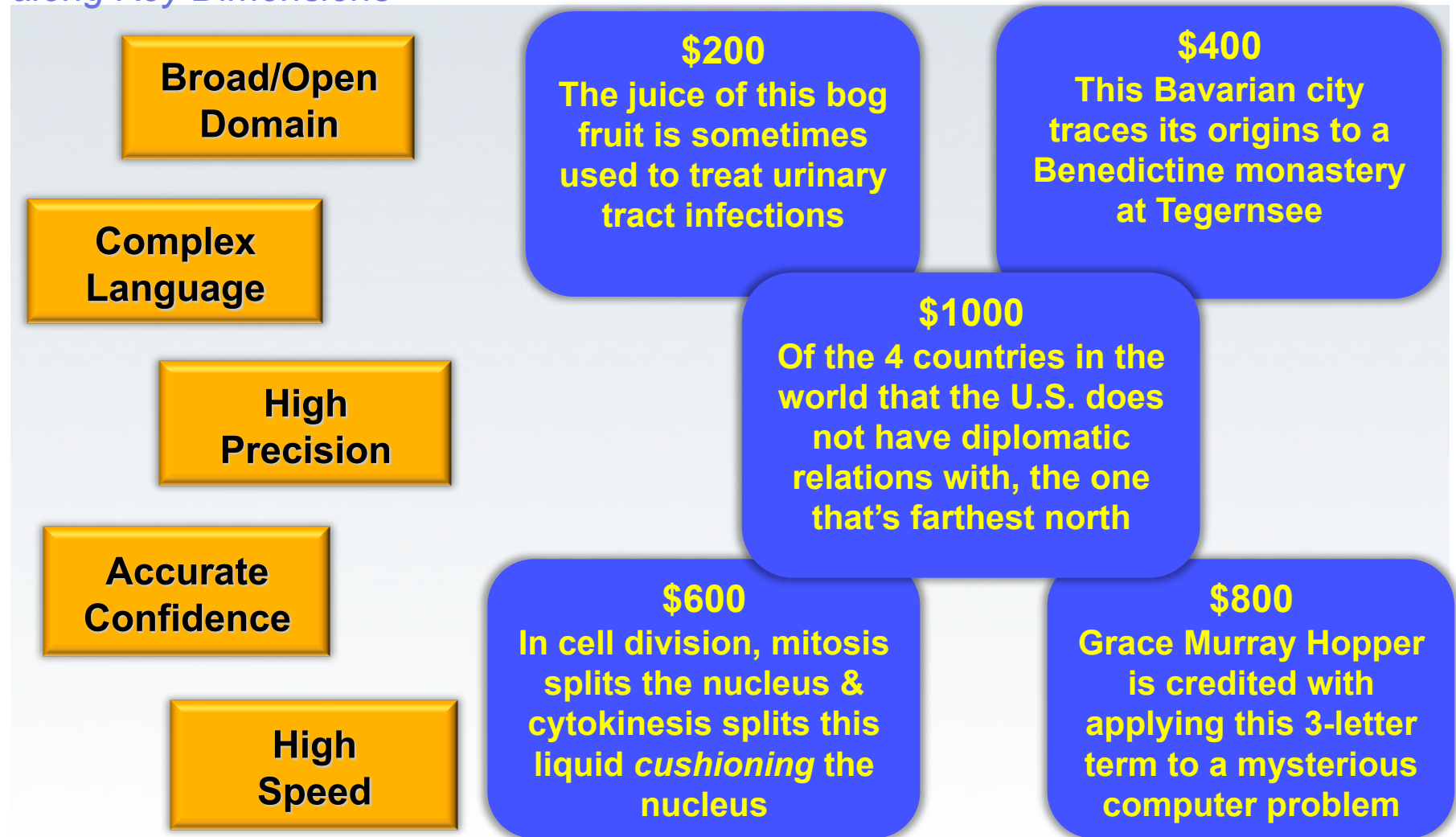
WATSON AND THE JEOPARDY! CHALLENGE

Automatic Open-Domain Question Answering
A Long-Standing Challenge in Artificial Intelligence to emulate human expertise

- Given
  - Rich **Natural Language Questions**
  - Over a **Broad Domain of Knowledge**

- Deliver
  - **Precise Answers:** Determine what is being asked & give precise response
  - **Accurate Confidences:** Determine likelihood answer is correct
  - **Consumable Justifications:** Explain why the answer is right
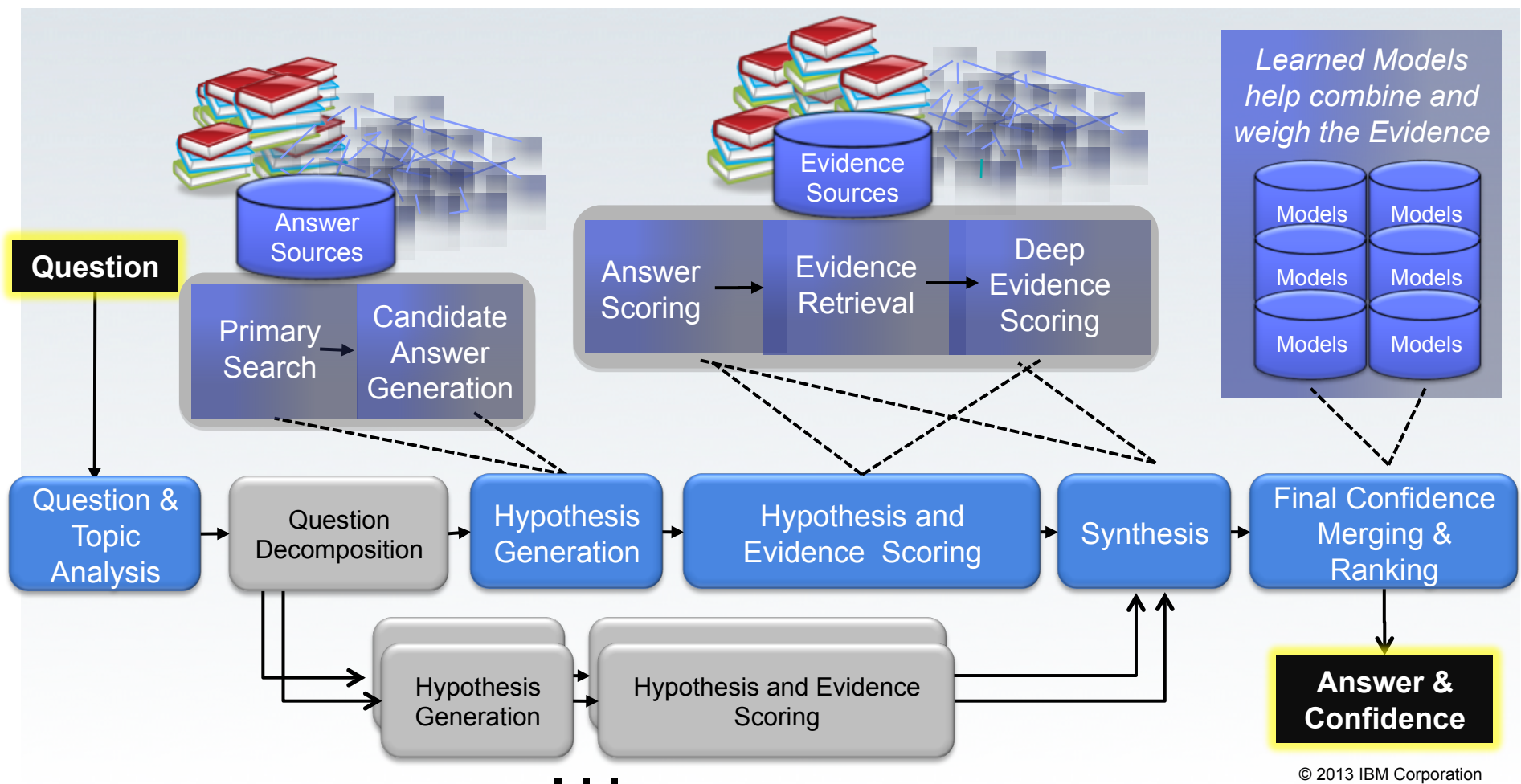 - **Fast Response Time:** Precision & Confidence in <3 seconds

The Jeopardy! Challenge
A palpable, compelling and notable way to drive the technology of Question Answering along Key Dimensions

Broad/Open Domain

Complex Language

High Precision

Accurate Confidence

High Speed

$200
The juice of this bog fruit is sometimes used to treat urinary tract infections

$400
This Bavarian city traces its origins to a Benedictine monastery at Tegernsee

$1000
Of the 4 countries in the world that the U.S. does not have diplomatic relations with, the one that's farthest north

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

$800
Grace Murray Hopper is credited with applying this 3-letter term to a mysterious computer problem

DeepQA: The architecture underlying Watson

*Generates many hypotheses, **collects a wide range of evidence** and balances the combined confidences of **over 100 different analytics that analyze the evidence from different dimensions***

Learned Models help combine and weigh the Evidence

Question

Answer Sources

Primary Search → Candidate Answer Generation

Evidence Sources

Answer Scoring → Evidence Retrieval → Deep Evidence Scoring

Models Models Models Models Models Models

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

Hypothesis Generation → Hypothesis and Evidence Scoring

Answer & Confidence

© 2013 IBM Corporation

Watson answers by finding, reading, scoring and combining evidence

IBM WATSON

IN 1698, THIS COMET DISCOVERER TOOK A SHIP CALLED THE PARAMOUR PINK ON THE FIRST PURELY SCIENTIFIC SEA VOYAGE

Question Analysis

Important Terms: 1698, comet, paramour, pink, ...
AnswerTypes: comet discoverer
Date(1698),
Took(discoverer, ship)
Called(ship, Paramour Pink)
...

Primary Search

Content (Structured & Unstructured)

Candidate Answer/Hypothesis Generation

Term Overlap | Classification | Relations | ... | Temporal

100's of NLP Scoring Algorithms

High-Speed Evidence Retrieval

Isaac Newton	[0.58 0.5 −1.3 … 0.97]
Wilhelm Tempel	[0.71 1 13.4 … 0.60]
HMS Paramour	[0.42 0 2.0 … 0.90]
Christiaan Huygens	[0.84 0.5 10.6 … 0.88]
Halley's Comet	[0.33 0 6.3 … 0.83]
Edmond Halley	[0.21 1 11.1 … 0.92]
Pink Panther	[0.91 0 −8.2 … 0.31]
Peter Sellers	[0.91 0 −1.7 … −.20]

...

1) **Edmond Halley (0.85)**
2) **Christiaan Huygens (0.20)**
3) **Peter Sellers (0.05)**
4) **...**

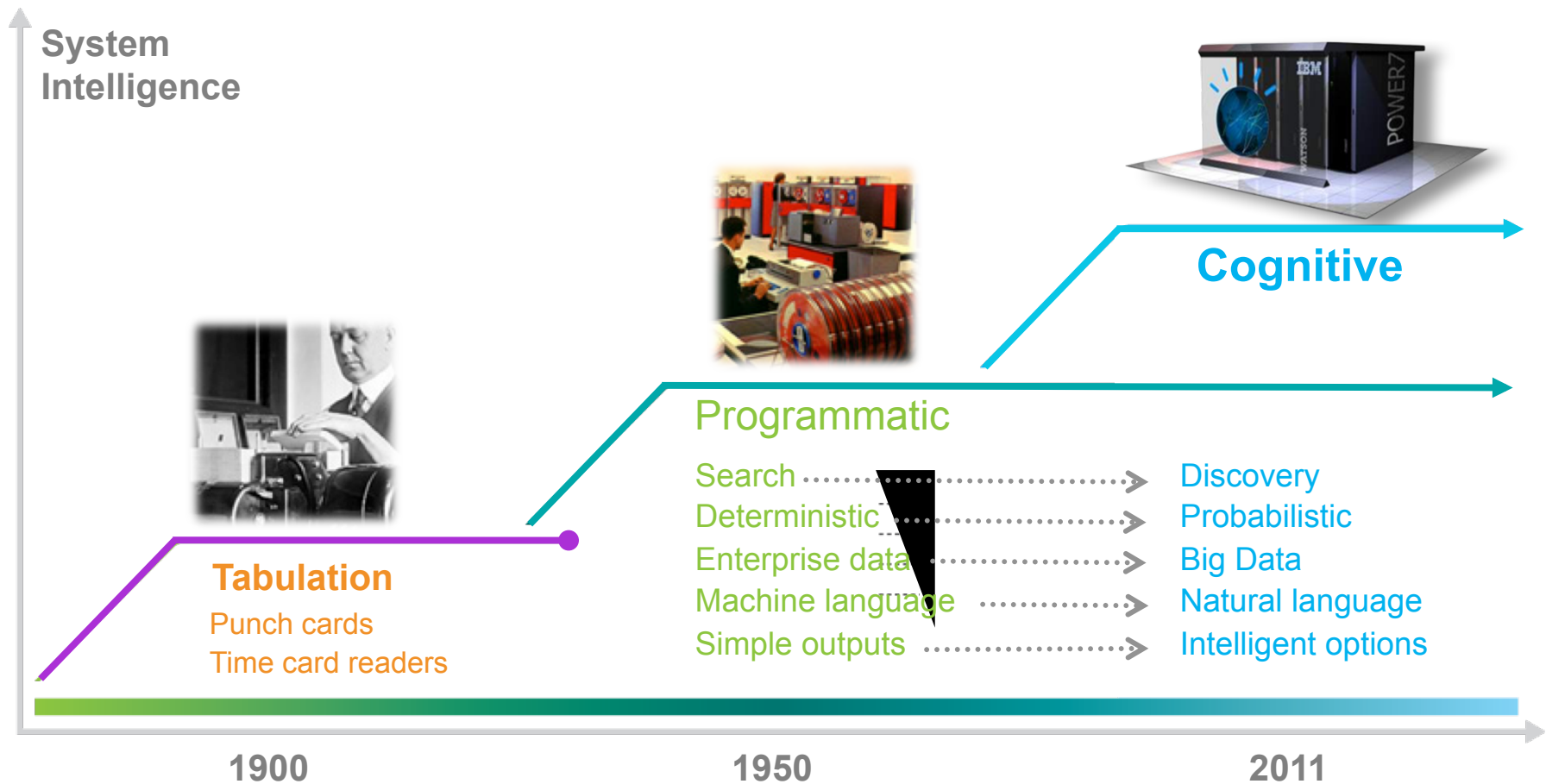**Merging & Ranking Based on Statistical Machine Learning**

Diverse and Extensible Evidence Scoring

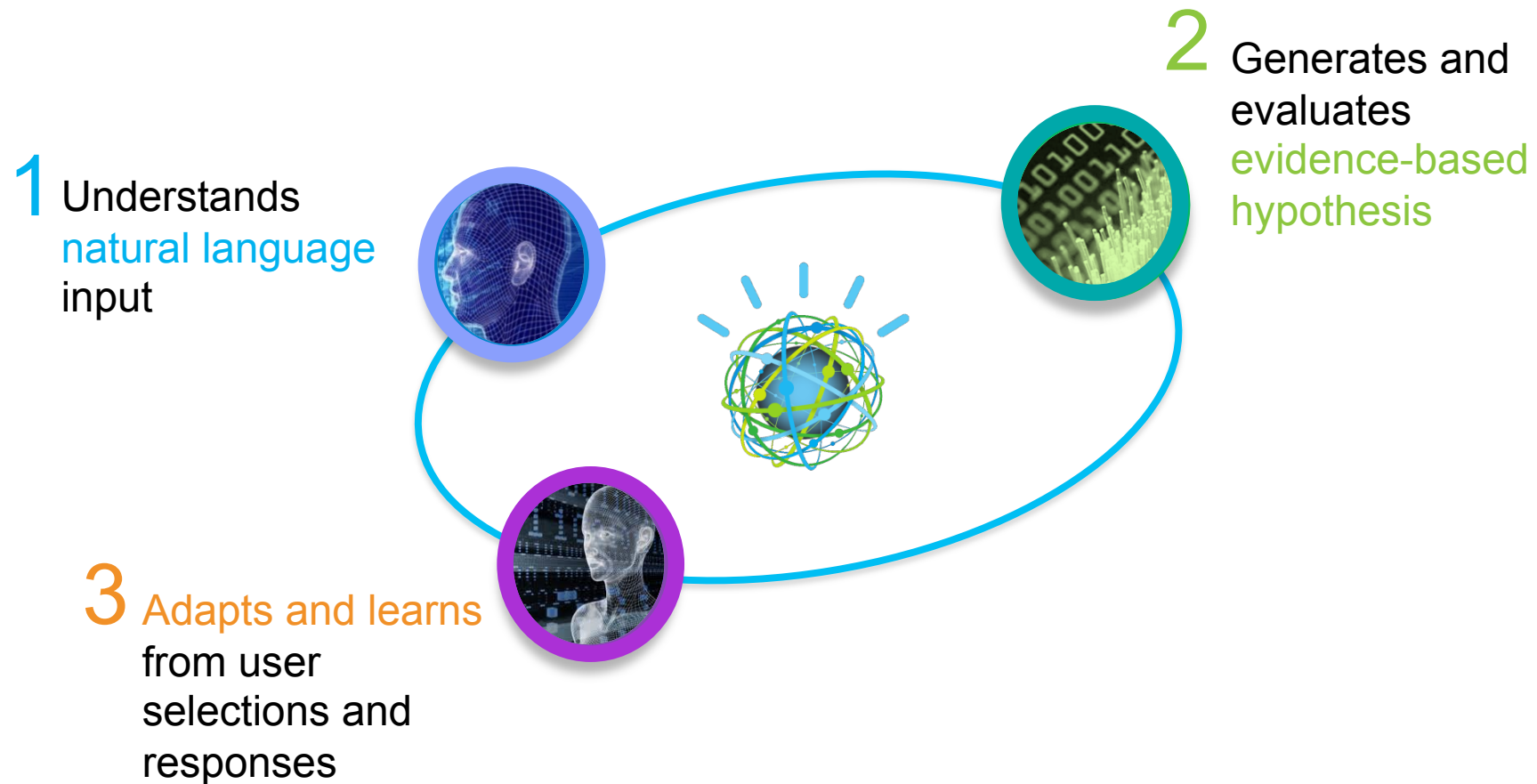**DeepQA: Incremental Progress in Answering Precision on the Jeopardy Challenge: 6/2007-11/2010**

IBM WATSON

**IBM *Watson*
Playing in the Winners Cloud**

Precision

% Answered

v0.8 11/10
V0.7 04/10
v0.6 10/09
v0.5 05/09
v0.4 12/08
v0.3 08/08
v0.2 05/08
v0.1 12/07
Baseline 12/06

© 2013 IBM Corporation

With Precision, Accurate Confidence and Speed, the rest was History
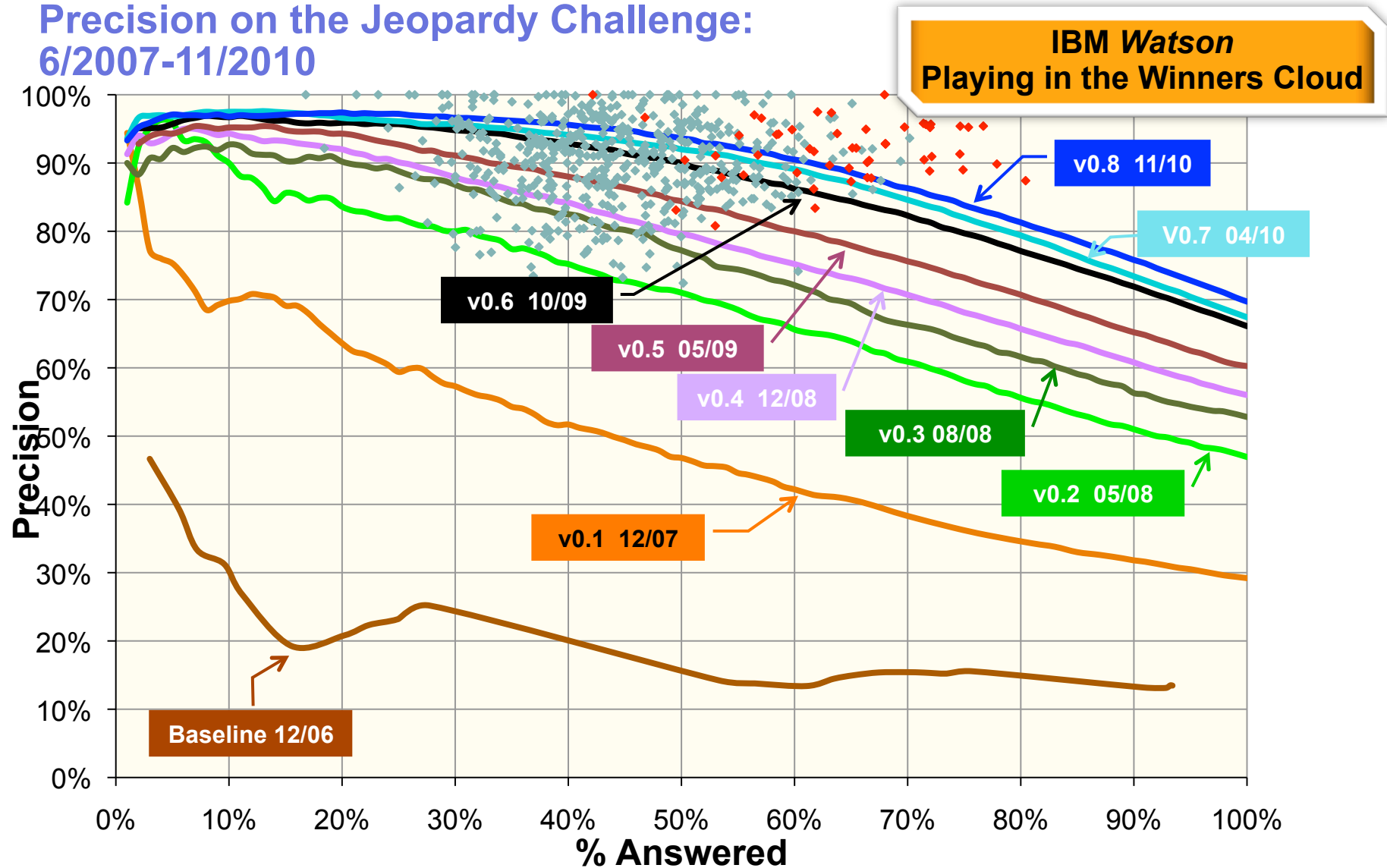
Watson: A New Era of Computing

System Intelligence

Cognitive

Programmatic

Search	·····>	Discovery
Deterministic	·····>	Probabilistic
Enterprise data	·····>	Big Data
Machine language	·····>	Natural language
Simple outputs	·····>	Intelligent options

Tabulation
Punch cards
Time card readers

1900   1950   2011

Watson Beyond Factoid Question Answering

1 Understands natural language input

2 Generates and evaluates evidence-based hypothesis

3 Adapts and learns from user selections and responses

WATSON IN HEALTHCARE

Use of Question Answering in Medical Diagnosis

After Watson's win on Jeopardy!, people assumed that anything that could be phrased as a question could be correctly answered by Watson:

Watson, "Given my medical record <insert hundreds of pages of structured and unstructured data here>, what's wrong with me?"

But that isn't what Watson was designed for:

- Watson wants a single sentence question
- Watson wants to find passages based on concepts in the question
- Watson wants to explore candidates found in relevant passages
- Watson wants to align answer-bearing passages with questions

The New Watson Challenge

We accepted the implied challenge to facilitate the reasoning process over a complex scenario:

Input

Complex natural language description of a problem

Output

Evidence-based inference chains leading to hypotheses

Our first domain of exploration is medical diagnosis because of its mature, complex and meaningful problem solving nature

Taking Watson beyond Jeopardy!

Understanding	Interacting	Explaining	Learning
Specific Questions	Question-In/Answer-Out	Precise Answers & Accurate Confidences	Batch Training Process

The type of murmur associated with this condition is harsh, systolic, and increases in intensity with Valsalva

Emily Dickinson — 99%
Walt Whitman — 60%
Barnard — 10%

From specific questions to rich, incomplete problem scenarios (e.g. EHR)

Evidence analysis and look-ahead, drive interactive dialog to refine answers and evidence

Move from quality answers to quality answers and evidence

Scale domain learning and adaptation rate and efficiency

Entire Medical Record

Input, Responses

Dialog

Refined Answers, Follow-up Questions

Answers, Corrections, Judgements

Responses, Learning Questions

Rich Problem Scenarios	Interactive Dialog Teach Watson	Comparative Evidence Profiles	Continuous Training & Learning Process

Sample Patient Scenario from US Medical Licensing Exam

A mother brings her 5-year-old son into your office. The boy has papular and pustular lesions on his face. A serous honey-colored fluid exudes from the lesions. A Gram stain of the pus reveals many neutrophils and Gram-positive cocci in chains. The organism is non-motile, catalase-negative, beta-hemolytic on blood agar, and is bacitracin sensitive. What organism is the most likely cause of the disease in this patient?
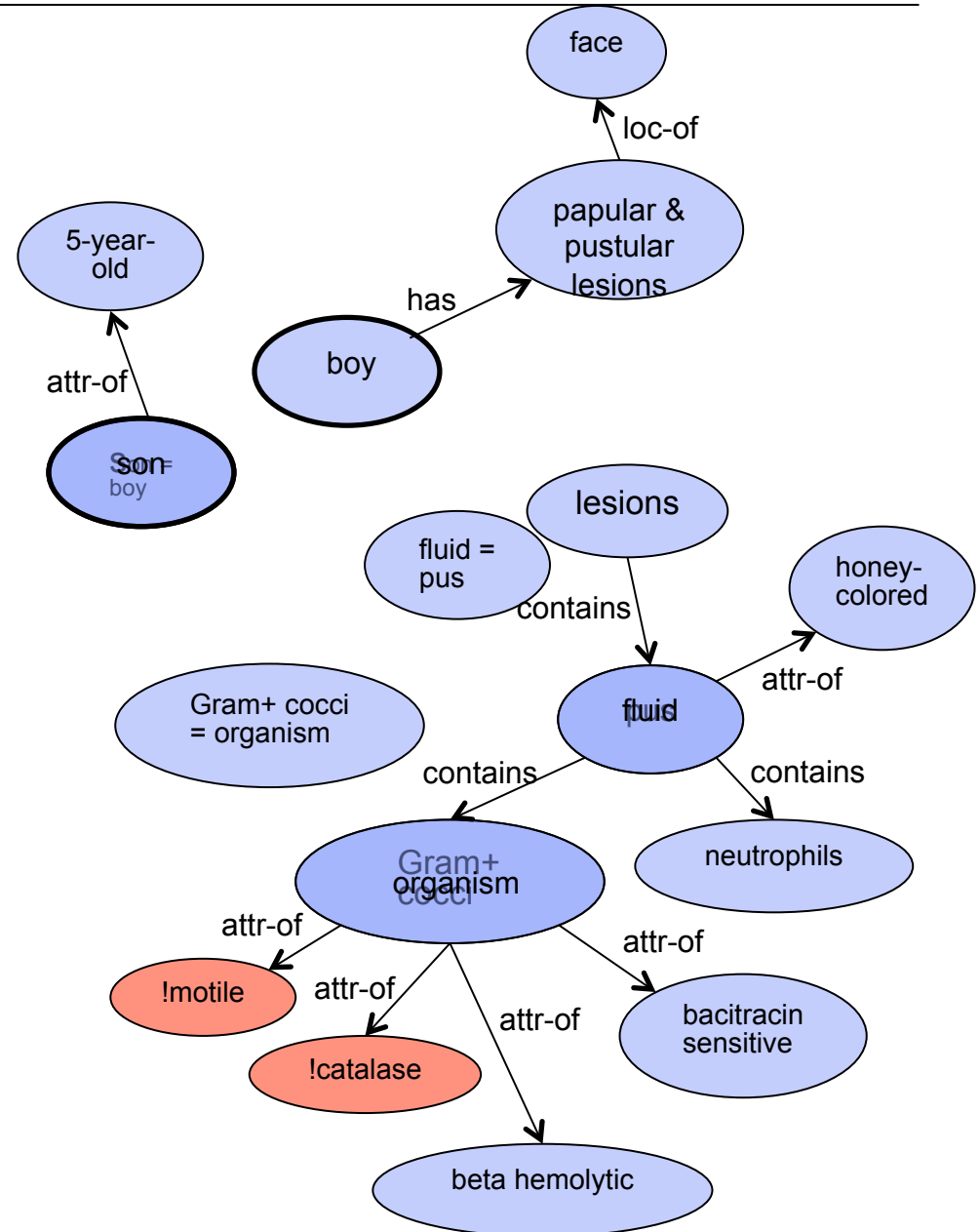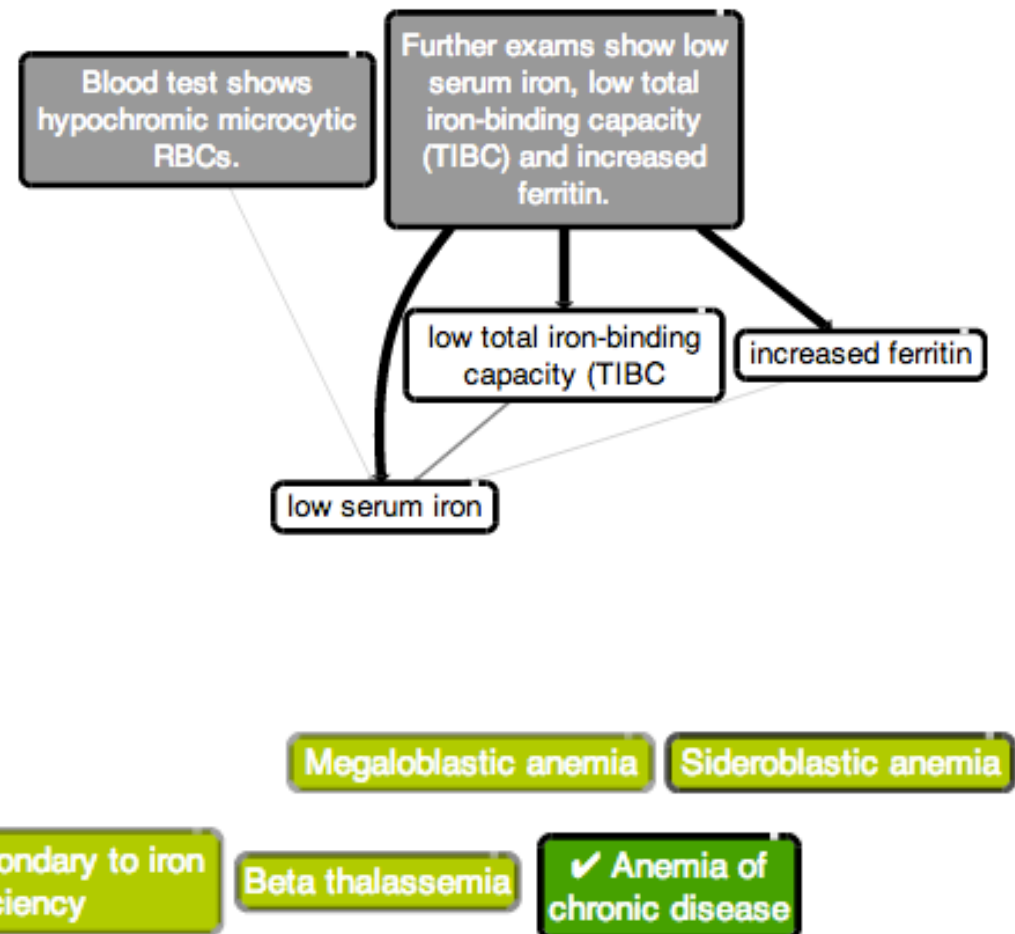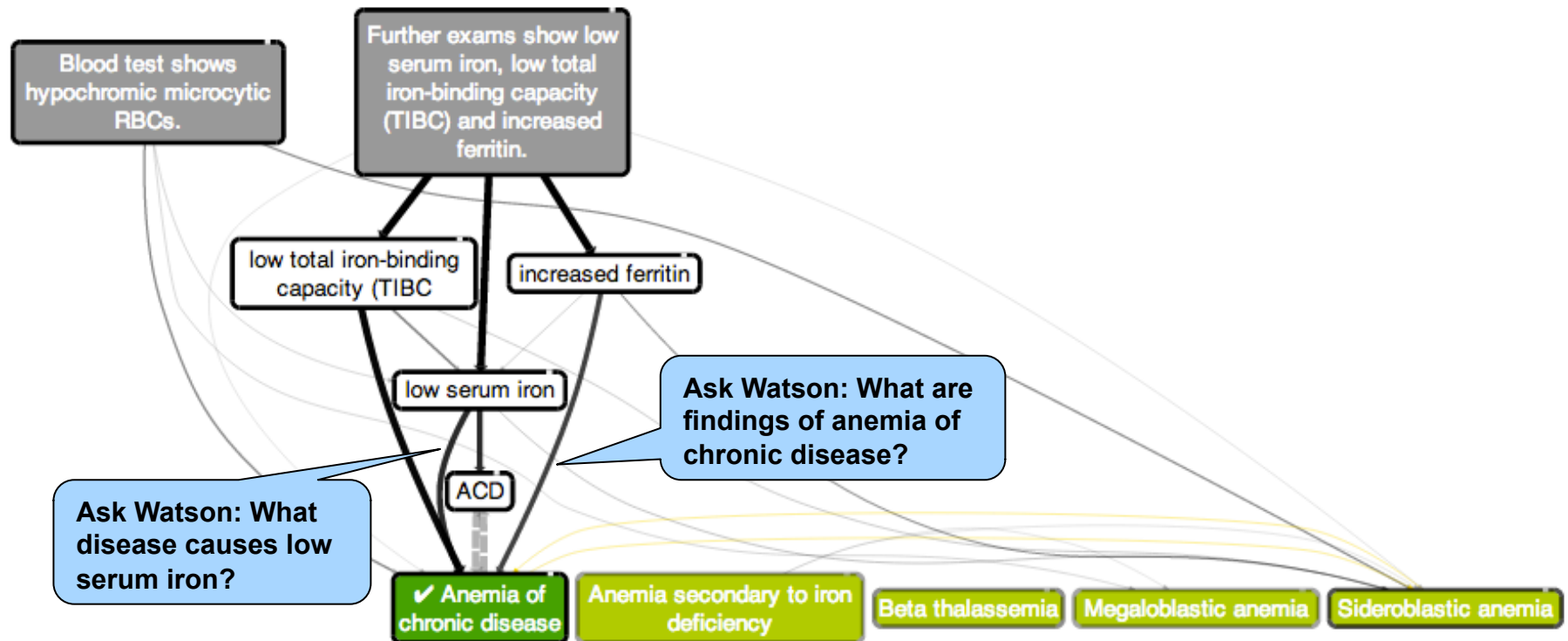
(A) Streptococcus pneumoniae
(B) Staphylococcus aureus
(C) Peptostreptococcus
(D) Streptococcus pyogenes
(E) Staphylococcus epidermidis

A 70-year-old man comes for a follow up with his cardiologist. There are no specific complaints. Findings at the physical exam are BP- 130/80 mmHg, HR- 80 beats/min, and appearance of pale mucous membranes. Lungs are clear to auscultation, and there is no edema of lower extremities. Fecal occult blood test (FOBT) was negative. Blood test shows hypochromic microcytic RBCs. Further exams show low serum iron, low total iron-binding capacity (TIBC) and increased ferritin. What is the most probable diagnosis in this patient?

(A) Anemia of chronic disease
(B) Anemia secondary to iron deficiency
(C) Beta thalassemia
(D) Megaloblastic anemia
(E) Sideroblastic anemia

➤ The answers are not one step away
➤ Finding them requires *connecting the dots*
➤ Shallow language understanding is not enough
➤ Discovering rationalized paths through the content becomes a key value

WatsonPaths: Beyond Factoid QA
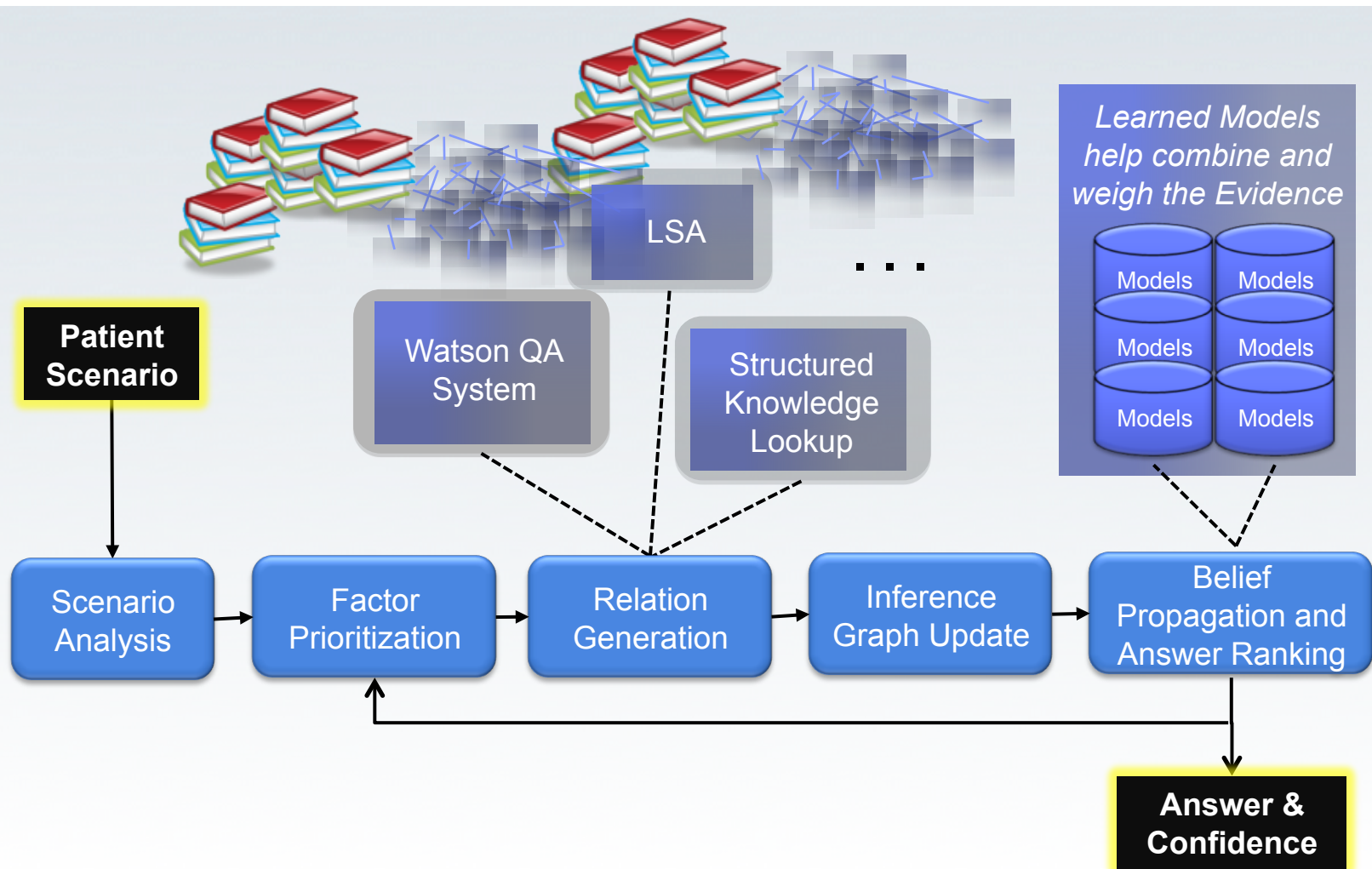
Processing Input Scenarios

A mother brings her 5-year-old son into your office. The boy has papular and pustular lesions on his face. A serous honey-colored fluid exudes from the lesions. A Gram stain of the pus reveals many neutrophils and Gram-positive cocci in chains. The organism is non-motile, catalase-negative, beta hemolytic on blood agar, and is bacitracin sensitive. What organism is the most likely cause of the disease in this patient?

- Objectives:
  1. Figure out what's wrong with the patient
  2. Identifying effective treatment/next steps

- Identify critical information from scenario for diagnosis
  - Parsing
  - Co-reference resolution
  - Negation detection
  - Clinical factor identification

Scenario Analysis Results

A mother brings her 5-year-old son into your office. The boy has papular and pustular lesions on his face. A serous honey-colored fluid exudes from the lesions. A Gram stain of the pus reveals many neutrophils and Gram-positive cocci in chains. The organism is non-motile, catalase-negative, beta hemolytic on blood agar, and is bacitracin sensitive. What organism is the most likely cause of the disease in this patient?

Clinical Factor Identification

IBM WATSON

Scenario Description → Clinical Factor Identification → Matching against Text corpus

Visualization

Sub-Question Generation

Parser | **Dictionary** | **Rules** | **Learning** | **Factors**

UMLS
· Concepts
· Semantic types

Lab Value Analysis
Negation
Syntactic &
    Semantic Patterns

Entity Detection
Relation Detection

Clinical Factor
· type
· score
Attributes
· type

"She has pain in the epigastric region and sometimes on the right side of her abdomen." → Clinical Factors

pain / finding
+**epigastric region** / anatomy

pain / finding
+**right side of the abdomen** / anatomy
+**sometimes** / temporal

A 70-year-old man comes for a follow up with his cardiologist. There are no specific complaints. Findings at the physical exam are BP- 130/80 mmHg, HR- 80 beats/min, and appearance of pale mucous membranes. Lungs are clear to auscultation, and there is no edema of lower extremities. Fecal occult blood test (FOBT) was negative. Blood test shows hypochromic microcytic RBCs. Further exams show low serum iron, low total iron-binding capacity (TIBC) and increased ferritin. What is the most probable diagnosis in this patient?

Blood test shows hypochromic microcytic RBCs.

Further exams show low serum iron, low total iron-binding capacity (TIBC) and increased ferritin.

low total iron-binding capacity (TIBC

increased ferritin

low serum iron

Megaloblastic anemia

Sideroblastic anemia

Anemia secondary to iron deficiency

Beta thalassemia

✔ Anemia of chronic disease

WatsonPaths for Medical Diagnosis (Cont'd)

2000057 What disease causes low serum iron?

Back to Questions
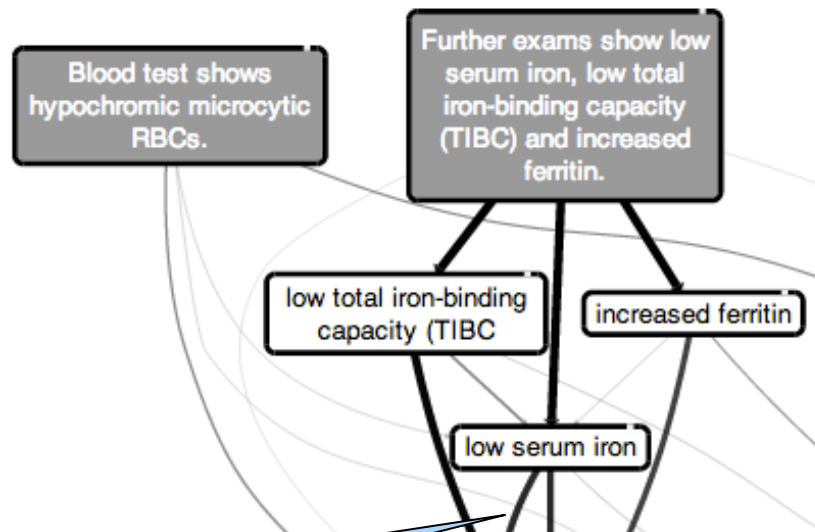
Evidence Profile

Anemia of chronic disease	59%	
Sideroblastic anemia	8%	

Diagram nodes:
- Blood test shows hypochromic microcytic RBCs.
- Further exams show low serum iron, low total iron-binding capacity (TIBC) and increased ferritin.
- low total iron-binding capacity (TIBC
- increased ferritin
- low serum iron

Bar chart categories: PASSAGE_SUPPORT, DOCUMENT_SUPPORT, TWREX

100% "Rheumatoid arthritis" *Corpus: Web Corpus Expansion*

▸ Details

In most cases, the reduced red cell mass is caused by the **anemia of chronic disease**, a normocytic-normochromic process characterized by a low concentration of serum iron, a low serum iron-binding capacity, and a normal or increased serum ferritin concentration.

65% "Iron deficiency anemia" *Corpus: Web Corpus Expansion*

▸ Details

Increased stainable iron in macrophages. **Sideroblastic anemia** is suspected in patients with microcytic anemia or a high RDW anemia, particularly with increased serum iron, serum ferritin, and transferrin saturation (see Anemias Caused by Deficient Erythropoiesis: Iron Deficiency Anemia)

A young woman presented with pneumonia, and a course of cephalosporins were given to treat this condition. She developed a non-bloody, watery diarrhea with abdominal cramps in a few days after starting the treatment and came back to the doctor's office. Her stool was sent to the laboratory and a colonoscopy performed to find out the cause of this diarrhea. Colonoscopy showed yellow-white plaques and membranes on colonic mucosa. Which of the followin is the most likely toxin, secreted from this microorganism, which is causing these colonic changes?

She developed a non-bloody, watery diarrhea with abdominal cramps in a few days after starting the treatment and came back to the doctor's office.

Colonoscopy showed yellow-white plaques and membranes on colonic mucosa.

watery diarrhea

abdominal cramps in a few days after starting the treatment
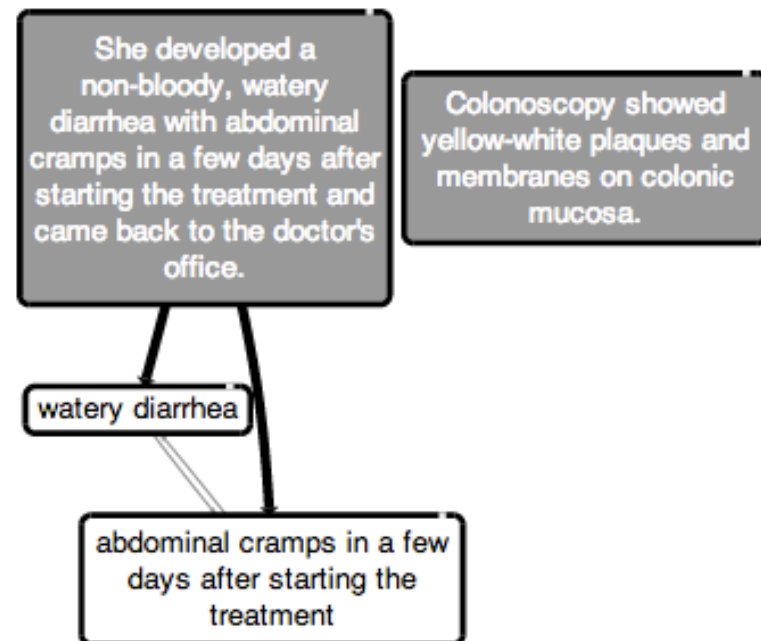
Neurotoxin   Endotoxin   Tetanospasmin   Alpha toxin   ✔ Cytotoxin B

Diagnosis First: Finding the Missing Links

Ask Watson: What causes colonoscopy to show yellow-white plaques and membranes on colonic mucosa?
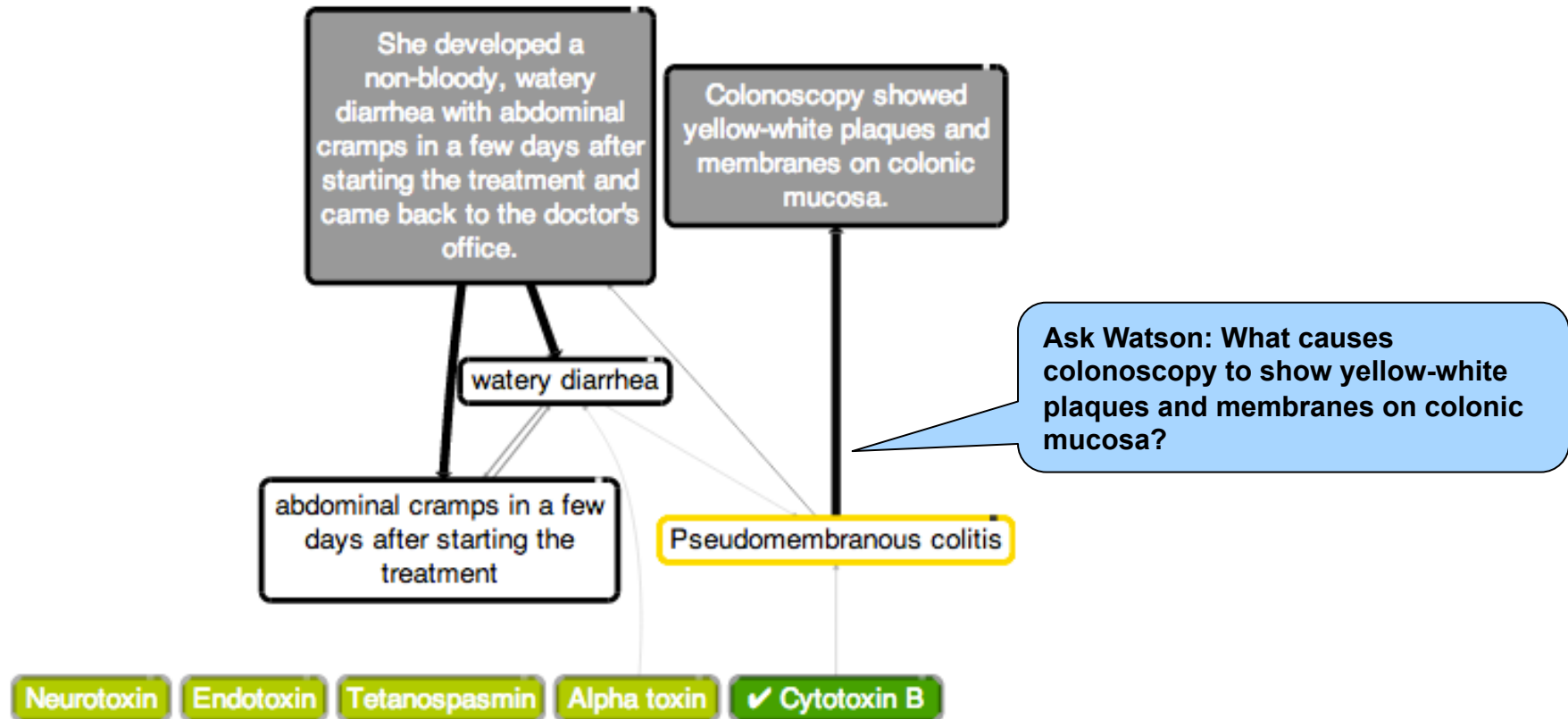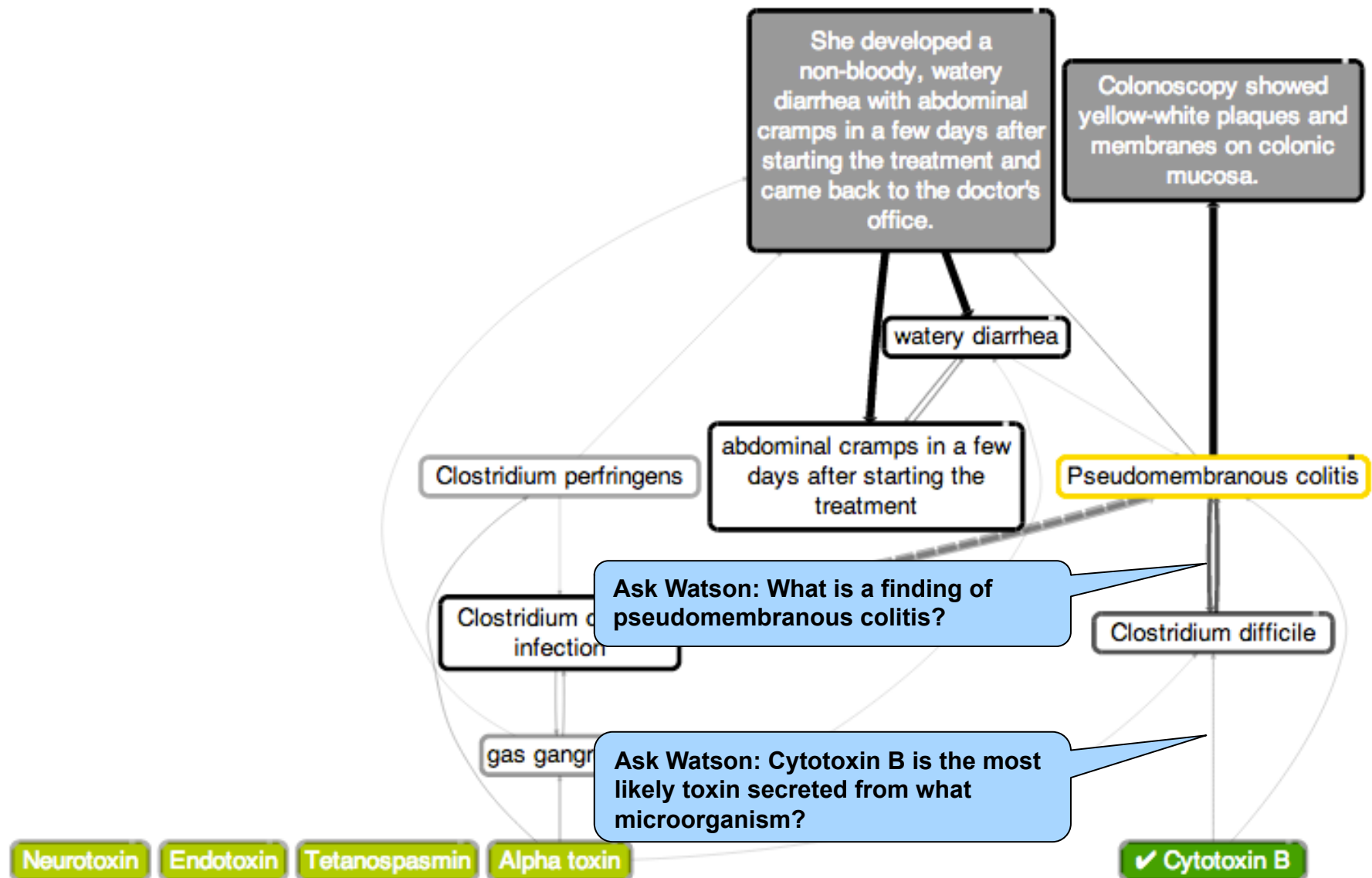
WatsonPaths: Leveraging Watson and Beyond
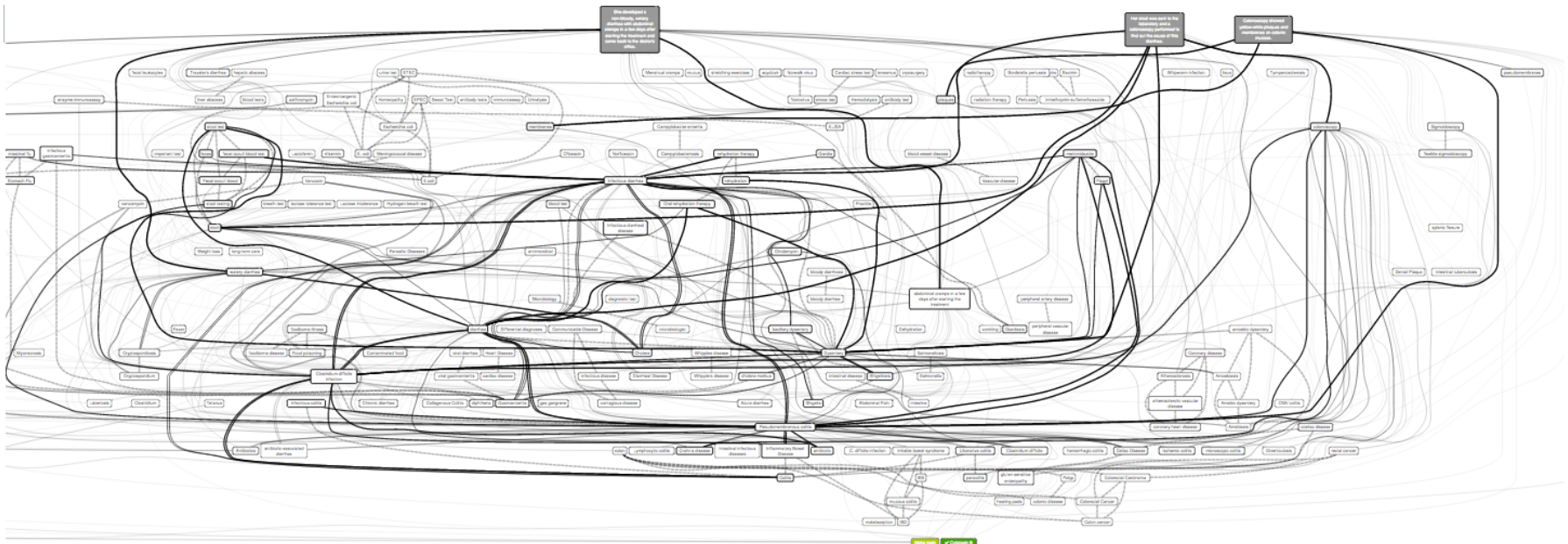
WatsonPaths builds complex inference graphs by relying on various systems (including Watson) to generate relations and confidence between nodes. With this capability WatsonPaths can answer questions that the base Watson system cannot. It provides a powerful and interactive decision support paradigm over large volumes of unstructured content.

Leveraging Existing Medical Resources
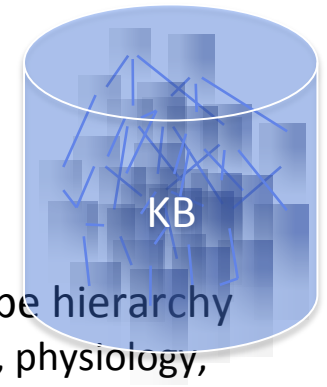
- UMLS (Unified Medical Language System) from NLM
  - ~100 sources, sort of merged
  - ~3M unique concept identifiers (not unique concepts), organized in a type hierarchy
    - activities, anatomy, chemicals/drugs, devices, disorders, genetics, organisms, physiology, procedures, …
  - ~350 relation types; ~30M unique relation instances
    - diagnoses, treats, finding_site_of, has_causative_agent, contraindicates, …

- Sample Uses of UMLS
  - In Watson QA system
    - Type Coercion: does a candidate answer match the type the question is seeking
    - Candidate generation
    - Term matching
  - In WatsonPaths
    - Clinical factor identification
    - Relation generation in inference graph
    - Term matching

Mining over medical corpus: Prismatic

As with other NSAIDs, ibuprofen may be useful in the treatment of severe orthostatic hypotension

Lasix (furosemide), a diuretic, and ibuprofen, an NSAID, can be taken together

Rule-based relation detector identifies hyponymy relations in text

Frame01	
subj	Ibuprofen
type	NSAID

Frame02	
subj	Lasix
type	diuretic

Frame03	
subj	NSAID
type	drug
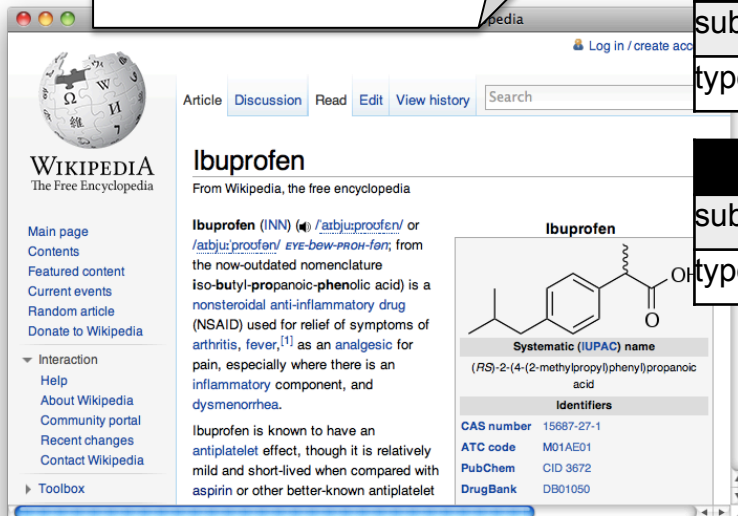
Ibuprofen isa NSAID

Lasix isa diuretic

NSAID isa drug

.
.
.

200M

Associating Symptoms with Diseases: Symptom KB

Urinary Tract Infection
The most common symptoms of a bladder infection are burning with urination, frequency of urination, an urge to urinate, without vaginal discharge or significant pain.[4] An upper urinary tract infection or pyelonephritis may additionally present with flank pain and a fever. Healthy women have an average of 5 days of symptoms.[ ] The symptoms of urinary tract infections may vary with ag
urinary system t
children, urinary
may include diar
nausea and vom
crying that cann
measures. [5] 6]

Cystitis
Symptoms:
   - **Sudden onset**
   - **Dysuria (painful urination)**
   - **Nocturia**
   - **Low back pain**
   - **pneumaturia**

Disease	Symptom	Sensitivity	Specificity
UTI	burning with urination	#	#
UTI	fever	#	#
Cystitis	dysuria	#	#
…	…	…	…

Topics Clusters : term co-occurrence in documents - LSA

exhibition

Picasso

painting

gallery

art

artist

hotel engine car driver

vehicle transport

drive

Type Clusters: terms share similar syntactic roles (Prismatic)

explorer	ibm	rome
firefox	dell	paris
safari	hp	london
opera	acer	venice
chrome	asus	milan
		florence

Complexity of Language in the Medical Domain

food would "get stuck" when she was swallowing

Causation

...can cause food to move slowly **in the esophagus**.

swallowing difficulty...

Terminology

...food gets held-up...

Normal QRS Pattern

Chronology

Abnormal QRS Complex

sudden onset of chills

Delta-Wave

cold

chills

PR Interveral

coryza

Domain

Terminology

Magnitude

Fever

High Temperature

Temperature

Fever **after** acute symptoms subside...

Chronology

---

productive cough *after nonproductive cough*

Negation

nonproductive cough

productive cough

Chronology

Abdomen Pain

Terminology

Urination Pain

Causation

pneumaturia

bubbles in the urine

Dysuria

Abdomen Pain exacerbated by exercise

---

between the upper abdomen and the back

Flank Pain

Location

Kidney Pain

Lower Back Pain

Syndrome of sore throat, fever, **sepsis** and **unilateral neck swelling**	…cause **septic thrombophlebitis** of the internal jugular vein (Lemierre syndrome). Most patients have fever, sore throat, odynophagia, and **swelling in the neck down to the hyoid bone**
Nasal mucosal atrophy and **foul-smelling crusts in the nasal passages**	Atrophic rhinitis is characterized by progressive nasal atrophy, mucosal colonization with Klebsiella ozaenae or other organisms and **foul smelling nasal discharge**
Syndrome characterized by hypokalemic metabolic alkalosis, mild hypotension, **calluses on the knuckles** and **enamel erosion**	Many individuals with bulimia have **skin abrasions on their knuckles** from inducing vomiting. The most common effect of anorexia and bulimia is **tooth enamel erosion**.
Flexing patient's right hip and knee to elicit pain is used to diagnose this condition	For example, the obturator sign is present when the **internal rotation of the thigh elicits pain** (i.e., pelvic appendicitis), and the psoas sign is present when the **extension of the right thigh elicits pain** (i.e., retroperitoneal or retrocecal appendicitis)

Paraphrases/Entailment

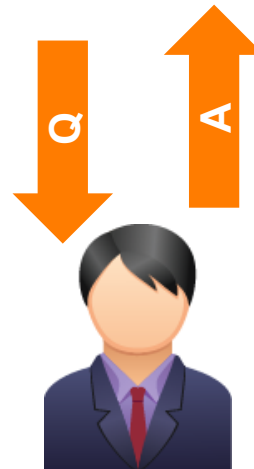| Question Text	Passage Text	Learned Axiom
Murmur associated with this condition is harsh, systolic, diamond-shaped, and increases in intensity with Valsalva	A systolic murmur that increases with the valsalva maneuver and disappears with squatting suggests hypertrophic cardiomyopathy	X suggests Y => X associated with Y
Class of drugs causing regression of polyposis in familial adenomatous polyposis	NSAIDs have been shown to induce adenoma regression in patients with familial adenomatous polyposis	X has been shown to induce Y => X causes Y
Intravenous treatment for cyanide poisoning	Antidotes for cyanide poisoning include amyl nitrate, sodium nitrate, and intravenous sodium thiosulfate.	Antidotes for X include Y => Y is treatment for X
Syndrome characterized by narrowing of the extra-hepatic bile duct from mechanical compression by a gallstone impacted in the cystic duct	Mirizzi's syndrome, a rare condition in which a gallstone impacting the cystic duct obstructs the common bile duct by edema and extrinsic compression	X obstructs Y => narrowing of Y by X
Preferred corrective treatment for acute episodes of angioedema in patients with hereditary angioedema	For acute episodes of angioedema in hereditary angioedema, adminster intravenous, purified, nanofiltered C1-INH concentrate as first-line therapy	For X, administer Y as first-line therapy => Y is preferred treatment for X

Explore the Evidence and Teach Watson

See Watson's evidence for making a connection. Judge it to improve the inference and Teach Watson in the process.
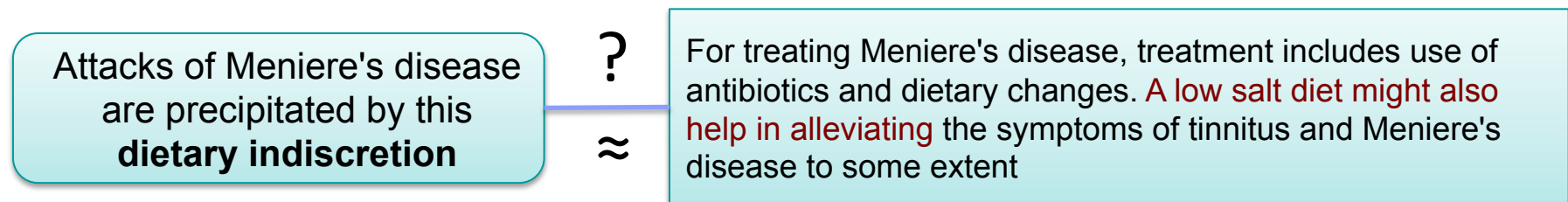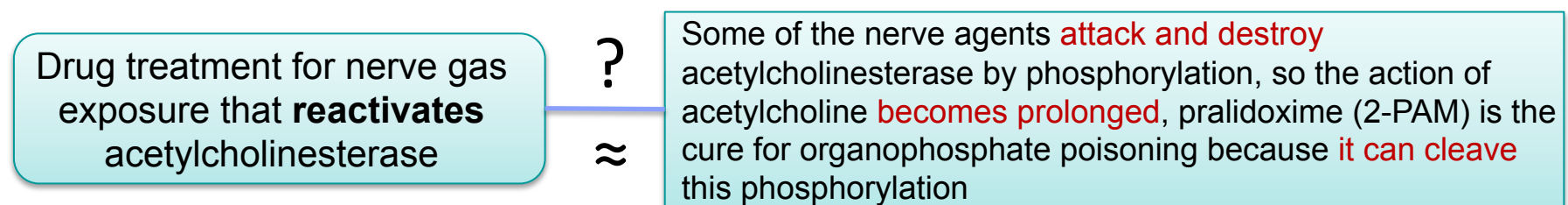
Learning Types
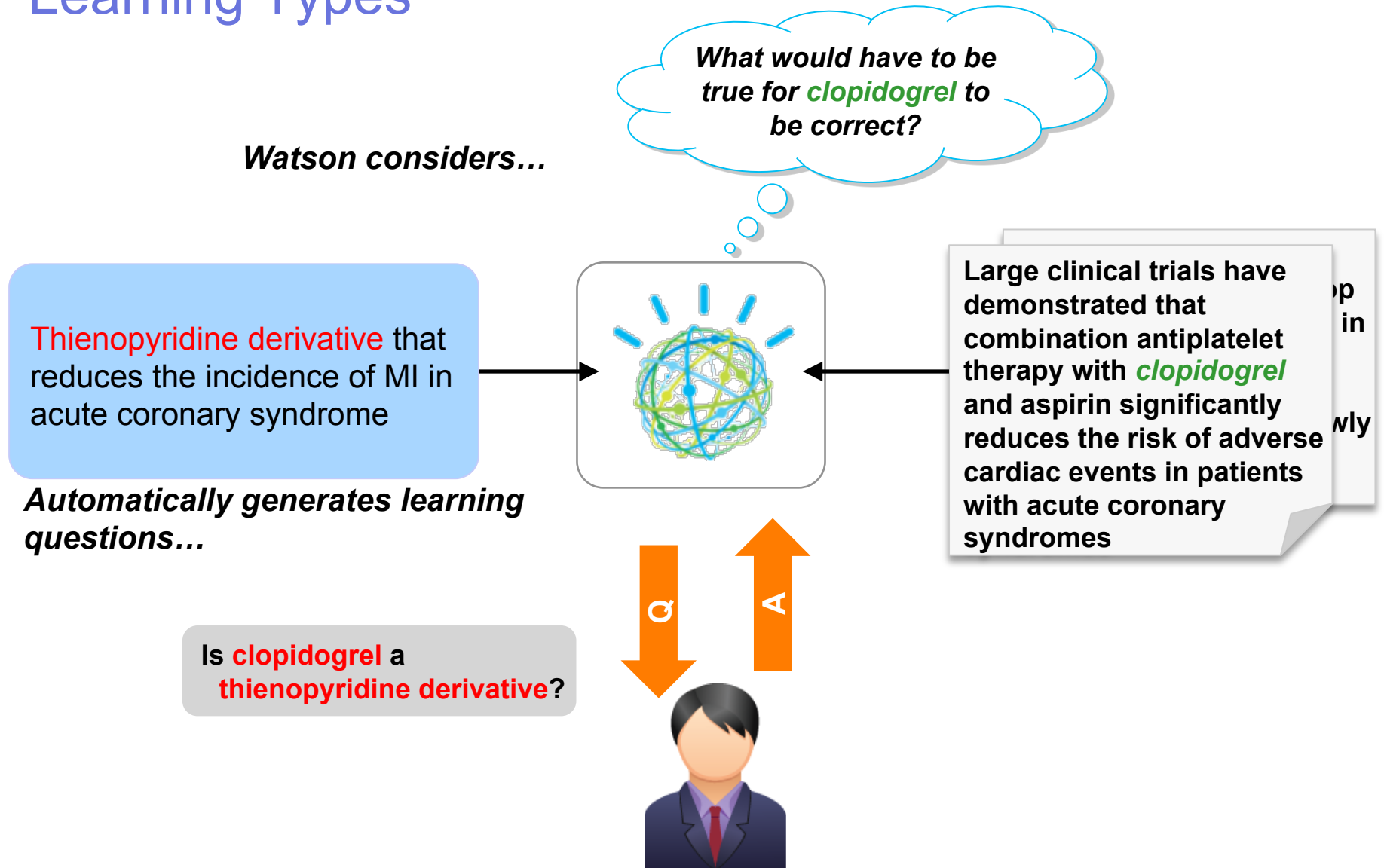
Some Language Challenges Require a Longer-Term Investment

Tackling the language/logic problem in a general and comprehensive way will take a longer-term sustained research investment. We need to provide value to the user today while the system actively learns through collaboration.

> Attacks of Meniere's disease are precipitated by this **dietary indiscretion**

?

≈

> For treating Meniere's disease, treatment includes use of antibiotics and dietary changes. A low salt diet might also help in alleviating the symptoms of tinnitus and Meniere's disease to some extent
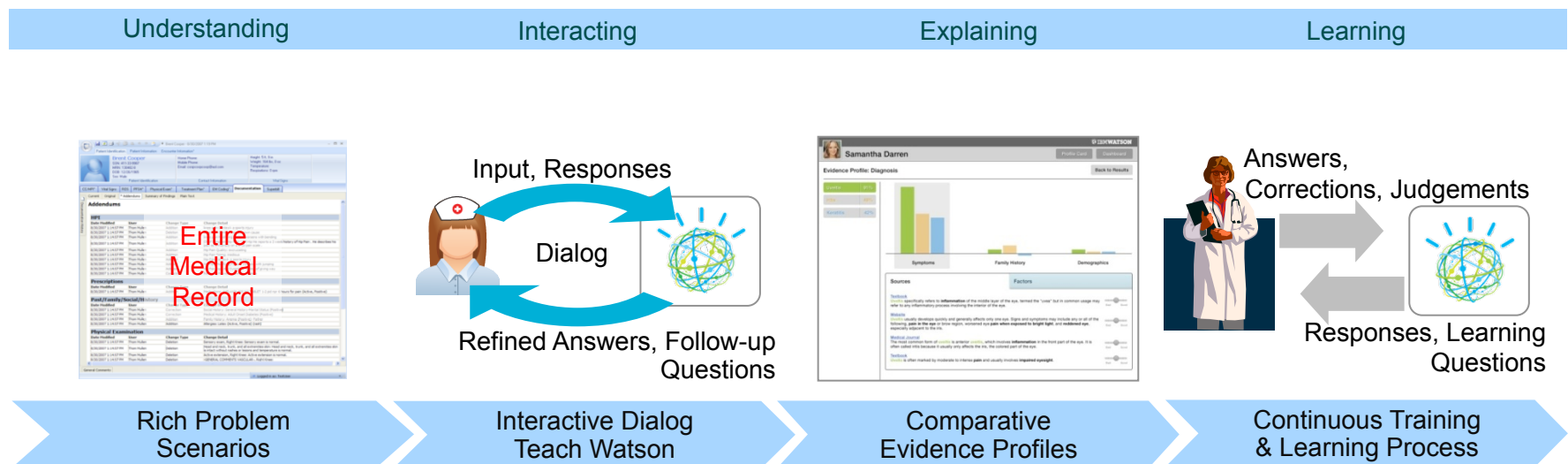
Both are referring to eating too much salt. But a **low-salt diet** is **not** a **dietary indiscretion**. However, if it alleviates a problem then perhaps the opposite (eating too much salt) is?

> Drug treatment for nerve gas exposure that **reactivates** acetylcholinesterase

?

≈

> Some of the nerve agents attack and destroy acetylcholinesterase by phosphorylation, so the action of acetylcholine becomes prolonged, pralidoxime (2-PAM) is the cure for organophosphate poisoning because it can cleave this phosphorylation

Reactivatation may be described by cleaving an action that may have caused *an attack and destruction*? If we worked hard to identify and manually model and generate such a complex and highly contextual rule – how many times would it be used? There are potentially millions of these. But where? Will they apply generally? Are they worth the investment?

Taking Watson beyond Jeopardy!: Recap

| Understanding | Interacting | Explaining | Learning |

Input, Responses

Dialog

Refined Answers, Follow-up Questions

Answers, Corrections, Judgements

Responses, Learning Questions

Entire Medical Record

| Rich Problem Scenarios | Interactive Dialog Teach Watson | Comparative Evidence Profiles | Continuous Training & Learning Process |

Additional Business Applications

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

Tech Support: Help-desk, Contact Centers
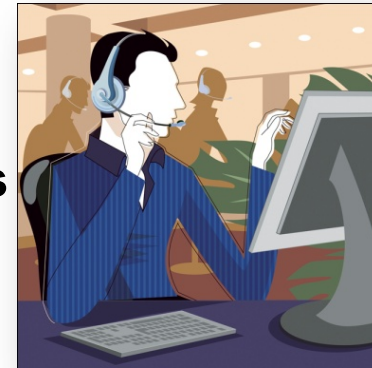
Enterprise Knowledge Management and Business Intelligence

Government: Improved Information Sharing and Education

THANK YOU