Watson: Jeopardy! and Beyond

Murthy Devarakonda, Ph.D.
Research Scientist and Team Manager
Watson Technologies
IBM Research

© 2013 IBM Corporation
The juice of this bog fruit is sometimes used to treat urinary tract infections.

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

Most modern supercomputers run on this operating system developed by Linus Torvalds.

In cell division, mitosis splits the nucleus & cytokinesis splits this liquid cushioning the nucleus.

Grace Murray Hopper is credited with applying this 3-letter term to a mysterious computer problem.
LINCOLN BLOGS
TREASURY SECY. CHASE JUST SUBMITTED THIS TO ME FOR THE THIRD TIME — GUESS WHAT, PAL. THIS TIME I'M ACCEPTING IT

RESIGNATION?

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

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

Diverse and Extensible Evidence Scoring

[0.58 0.5 -1.3 ... 0.97]
[0.71 1 13.4 ... 0.60]
[0.42 0 2.0 ... 0.90]
[0.84 0.5 10.6 ... 0.88]
[0.33 0 6.3 ... 0.83]
[0.21 1 11.1 ... 0.92]
[0.91 0 -8.2 ... 0.31]
[0.91 0 -1.7 ... -0.20]

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

Merging & Ranking Based on Statistical Machine Learning

© 2013 IBM Corporation
DeepQA: The architecture underlying Inside Watson

Using a highly parallelized architecture generates many hypotheses, collects a wide range of evidence and balances the combined confidences of over 100 different analytics that analyze the evidence form different dimensions.
Watson Learns Relationships by “Reading”

Volumes of Text → Syntactic Frames → Semantic Frames

- Sentence Parsing
- Generalization & Statistical Aggregation

Syntactic Frames:
- subject
- verb
- object

Semantic Frames:
- Inventors patent inventions (0.8)
- Officials Submit Resignations (0.7)
- People earn degrees at schools (0.9)
- Fluid is a liquid (0.6)
- Liquid is a fluid (0.5)
- Vessels Sink (0.7)
- People sink 8-balls (0.5) (in pool/0.8)
In cell division, mitosis splits the nucleus & cytokinesis splits this **liquid cushioning** the nucleus.

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”.

- **Organelle**
- **Vacuole**
- **Cytoplasm**
- **Plasma**
- **Mitochondria**
- **Blood** …

\[
\text{Is(“Cytoplasm”, “liquid”) } = 0.2 \\
\text{Is(“organelle”, “liquid”) } = 0.1 \\
\text{Is(“vacuole”, “liquid”) } = 0.2 \\
\text{Is(“plasma”, “liquid”) } = 0.7
\]

“Cytoplasm is a **fluid surrounding** the nucleus…”

Wordnet → Is_a(Fluid, Liquid) → ?

Learned → Is_a(Fluid, Liquid) → yes.
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.

Evidence suggests “Gary” is the answer BUT the system must learn that keyword matching may be weak relative to other types of evidence.
In May 1898 Portugal celebrated the 400th anniversary of this explorer’s arrival in India.

On the 27th of May 1498, Vasco da Gama landed in Kappad Beach.

The evidence is still not 100% certain.
With Precision, Accurate Confidence and Speed, the rest was History
Watson is ushering in a new era of computing . . .

. . . enabling new opportunities and outcomes
Watson takes on medical applications...
Started on American College of Physician’s Doctor’s Dilemma questions such as: *Diagnosis for fever, adenopathy, and arthralgia following drug exposure*

Correct Response: Lemierre’s Syndrome

3+ years of progress
Teaching Watson Medical Concepts

Watson Uses US National Library of Medicine’s UMLS (Unified Medical Language System) for Medical Vocabulary
Answering some questions requires finding the missing link…
On hearing of the discovery of George Mallory's body, he told reporters he still thinks he was first.
A 63-year-old patient is sent to the neurologist with a clinical picture of resting tremor that began 2 years ago. At first it was only on the left hand but now it compromises the whole arm. At physical exam, the patient has an unexpressive face and difficulty in walking, and a continuous movement of the tip of the first digit over the tip of the second digit of the left hand is seen at rest.

What part of his nervous system is most likely affected?

Choices:
A. Cerebellum
B. Pons
C. Substantia Nigra
D. Caudate Nucleus
A 63-year-old patient is sent to the neurologist with a clinical picture of resting tremor that began 2 years ago.

At physical exam, the patient has an unexpressive face and difficulty in walking, and a continuous movement of the tip of the first digit over the tip of the second digit of the left hand is seen at rest.

Multi-Step Inferencing
A More Detailed View of the Analysis
Thank You