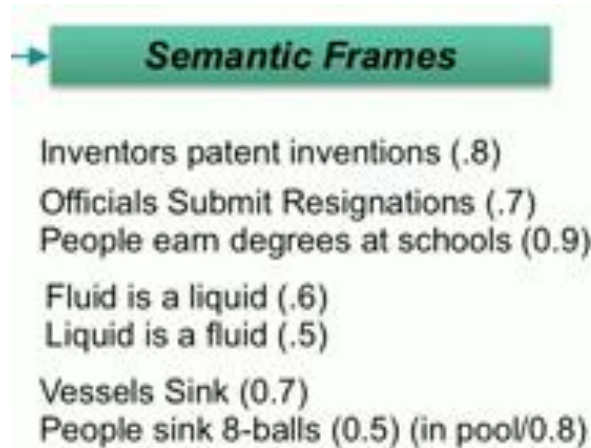




# What's Wrong with Watson?

Jeopardy Questions

# First, Semantic Frames



This is presumably state of the art – as used in Watson.

If you scan them you will see several problems – they are non-contextual – the probability that a fluid is a liquid varies in different areas – marine, aerospace, medical. The one that is particularly bad is

**Liquid is a fluid with a confidence level of 0.5**

AI is full of this stuff – nonsensical facts, from which valid results should spring

Examples from: Ferrucci's presentation

[http://www.youtube.com/watch?v=3G2H3DZ8rNc&feature=player\\_embedded](http://www.youtube.com/watch?v=3G2H3DZ8rNc&feature=player_embedded)

# Question 1

## Question

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

The question itself is confused but we will ignore that

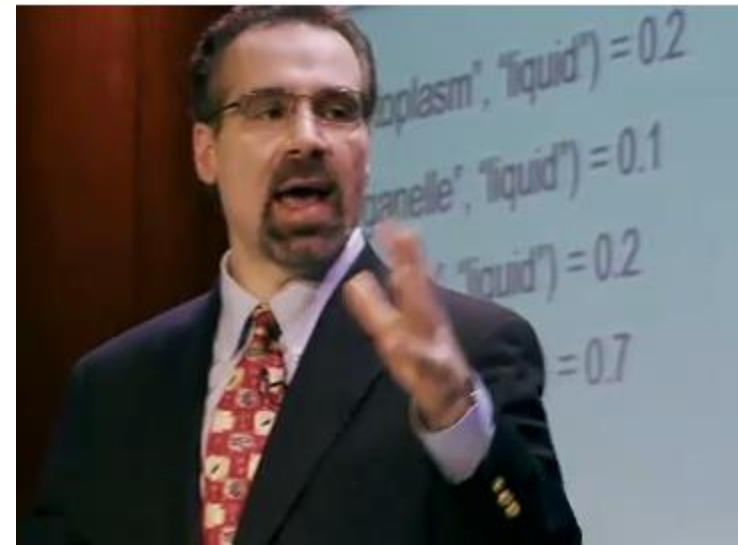
# A Possibility

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

# Some Possibilities

## Intermediate Hypotheses

- is ( "cytoplasm", "liquid" ) = 0.2
- is ( "organelle", "liquid" ) = 0.1
- is ( "vacuole", "liquid" ) = 0.2
- is ( "plasma", "liquid" ) = 0.7



# Wordnet

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

└ **Wordnet:** is\_a(Fluid, Liquid) ?/  
is ("cytoplasm", "liquid) = **0.2**

You can't rely on Wordnet, it is not very good

# What We Are Told

The IBM team provided Watson with millions of documents, including ... 200 million pages of structured and unstructured content consuming four terabytes of disk storage, **including the full text of Wikipedia.**

## Cytoplasm

---

From Wikipedia, the free encyclopedia

The **cytoplasm** is a thick liquid residing between the cell membrane

This looks pretty definitive – why wasn't it used?

Given the description of the mechanism, it should have been used, so one concludes the description of the knowledge access mechanism is incorrect

# Why Doesn't It Use What It Has?

To decide cytoplasm is a liquid with a confidence of 0.2 when it has a simple “Cytoplasm is a liquid” statement available to it is not very good.

It would appear to have accumulated a lot of garbage in its efforts to summarise

– liquid is fluid 0.5

How can you rely on a system that has been allowed to fill up with garbage?



# Question 2

This may seem obvious, but answering the wrong question is confusing

Category: US Cities

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

Answer: Toronto

The requirement of the question – US Cities – was ignored

The explanations given in the Wikipedia article -

[http://en.wikipedia.org/wiki/Watson\\_\(computer\)](http://en.wikipedia.org/wiki/Watson_(computer)) - only make the situation worse

- It couldn't parse the second part of the question – doesn't handle context or semicolons
- US City wasn't mentioned in the question – it was the category
- Cities in the US named Toronto
- An American baseball team in Toronto, which is presumably enough to make it a US City (this sounds like a wild reason – Watson would not have used such a long path)
- Didn't know Toronto was not a US City – it did know it was a city though, and it has its GPS coordinates

In other words, it had no idea what the question said, and could not ensure that the answer was consistent with the question

# Question 3

Category: Bounce

In August 1812 a US Frigate got this nickname when British cannonballs were seen to bounce off its thick oak hull

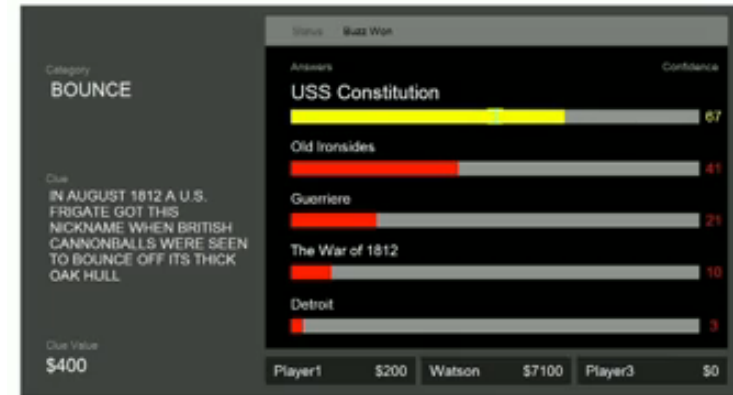
Watson answer: USS Constitution

The answer had to be a nickname, not the real name – the meaning of the question is being ignored

The nickname requirement should have made 4 out of 5 of the hypotheses fail

Guerriere was the English ship, so it should have failed completely – as not being a US Frigate – Watson doesn't do nouns, or it doesn't do adjectives

If we are talking evidence-based probabilistic Q&A, how does Detroit (the fifth ranked answer) fit into the picture? It was a long way from the action.



**Wikipedia:**

**Old Ironsides** is usually applied as a nickname, and may refer to: [USS Constitution](#), a 44-gun US Navy frigate and the oldest commissioned warship afloat

This fact was available to Watson

# But It Can Be Improved!

What is needed to improve it would destroy it

The question needs to be accurately parsed if the answer is to be reliable, but for that you need the knowledge to do so, so you can't just rush out looking for cross-connections on just the words in the question

Need to handle context – this has been the bogey of AI for 60 years

The knowledge needs to be made consistent before it is searched for a particular answer – maybe not for Jeopardy, but certainly for medical diagnosis

Hold the most precise information:

Cytoplasm is a fluid, cytoplasm is a liquid

Weapon is a gun, weapon is a Glock semiautomatic 9 mm pistol

Need to handle multiple ways of determining something – if you know the viscosity of a fluid, you can tell whether it is a liquid or a gas

# Parsing 1

Watson uses a three slot parser

So what – subject, verb, object

Let's take some advanced text, see how  
Watson would perform on it

# Parsing 2

Mary had a little lamb,  
Its fleece was white as snow,  
And everywhere that Mary went,  
That lamb was sure to go.

We run the three slot parser on it – we get

Mary had            lamb,

Mary went,

.

# Parsing 3

What went wrong?

It uses a three slot parser – it expects noun, verb, noun.

It doesn't know what to do with:

- A complex subject like “Its fleece”
- A layered verb like “is sure to go”
- A preposition like “as” as in “white as snow”
- A relative pronoun, “everywhere that”
- Anaphora, “that lamb”
- And a million other things

# Parsing 4

This leaves five alternatives:

1. The analysis of the parsing given here is wrong
2. It got all the semantic information required
3. It might read something else that will have the text arranged differently
4. We knew that already, but why disappoint people – what they don't know won't hurt them
5. We use a better parsing method, so it won't be Watson or any of its foundations

# So How Come It Worked on Jeopardy?

Slight changes to the questions, like using more prepositions (something a human wouldn't notice) would have resulted in Watson getting every question wrong.

It disassociates the question into its component parts – individual words or triples consisting of noun verb noun. Text, even the simplest – a nursery rhyme - just isn't like that.

We need to see through the marketing hype to understand how it works, so its suitability as a basis for further advancement can be assessed.