The Semantic Web
Challenges and Potentials

ICWR 2018

Special Session on the Future of the Web

Tehran, Iran

April 26, 2018

Hassan Aït-Kaci

HAK Language Technologies

Wherein Lies the Knowledge?

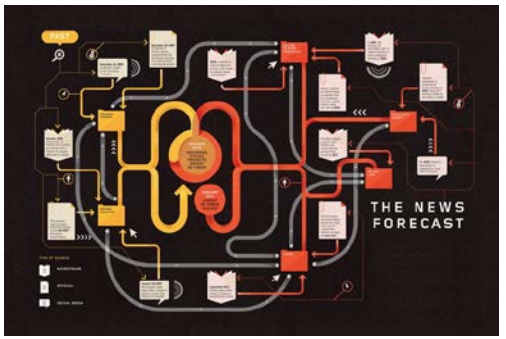
The next wave of information processing must adapt to a **radical change of reality**—namely, the **enormous quantity of available data** and the supernova **rate at which it accumulates**.

Implicit in this data hides

a wealth of information

—literally!

An article in a 2011 issue of the **Wired magazine** illustrates this with the **remarkable prediction success** of a ***small* data analysis company** in Gothenburg, Sweden, called **Recorded Future**.



Wherein Lies the Knowledge?

Such is this company's rate of success in predicting world's events and situations before anyone else, that **most major world players** (including **Google** and the **CIA**!) **line up as its customers**.

How they do it is their trade secret, of course—but, put simply, ***they find all they need in publicly available data***.

Wherein Lies the Knowledge?

Yet, as successful as his company may be, Recorded Future's co-founder and CEO **Christopher Ahlberg** makes the following statement:

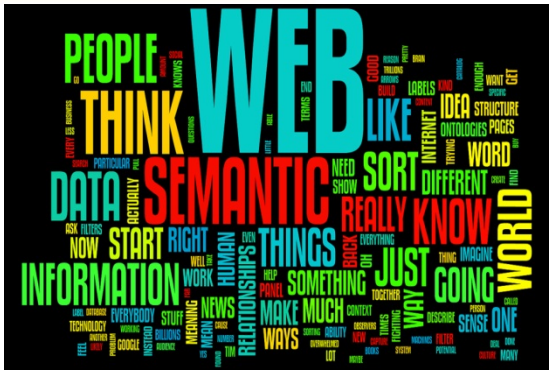
*"... to develop **a tool that could create predictions for any input**, from finance to terrorism, **would be much harder**. [One] would not only have to index the internet, but also **understand and interpret it**."*

—Christopher Ahlberg as quoted by Tom Cheshire in *Wired*–November 10, 2011

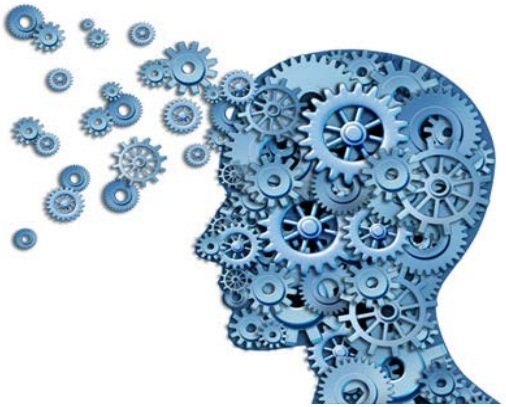
Indeed, Recorded Future's boon may only be the ***tip of an iceberg***. So the challenge is: ***extract*** *and* ***use implicit knowledge*** *hidden in* ***public data***.

And we're talking about ***Big Data!***

Understanding the Web — the "*Semantic*" Web?



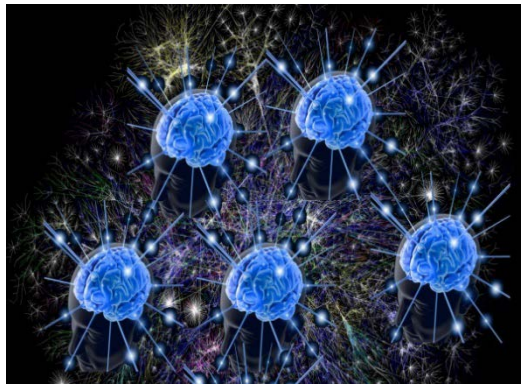
The **Semantic Web** has been heralded for two decades now as the means to infuse **meaning** into the World-Wide Web.

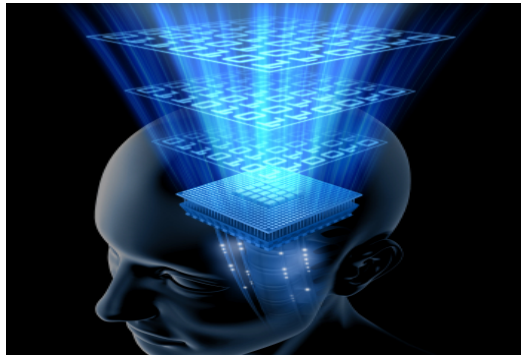


Subject of controversy, this **ambitious objective** has been disputed re. **what is actually *meant* by "*meaning*."**

Many see this as a truly achievable potential made possible by the **sublimation into AI knowledge** of **massively interconnected standardized data**.

Semantic Web Challenges





- Distributed **information interconnection**—*whose?*
- **Knowledge representation** language(s) — *which?*
- **Automated reasoning** power(s) — *what logics?*
- Need to agree on (a) **standard**(s) — *how many?*

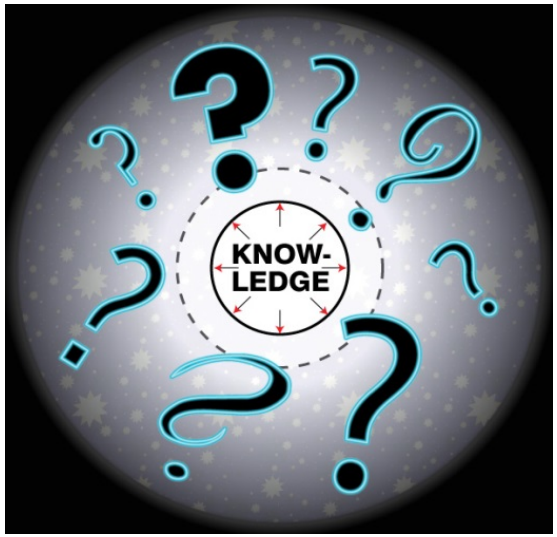
KIF Standard…?

In AI, **KIF** is not a narcotic; but it is…

the… **Knowledge Interchange Format**

http://www-ksl.stanford.edu/knowledge-sharing/kif/

A **LISP**-like language and S-expression structure language proposed to describe many (all?) **knowledge representation formalisms** so they each provide their own standardized form to one another.

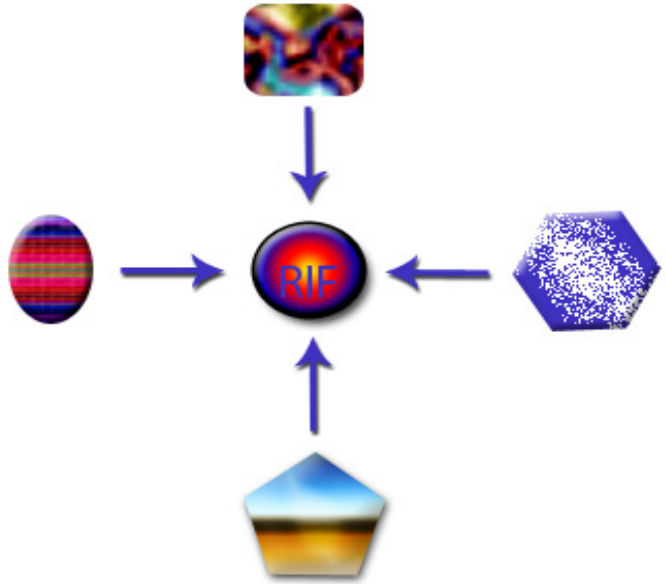


RIF Standard…?

In AI, the **RIF** is not a mountain range in northern Morocco…

It is the… **Rule Interchange Format**
http://www.w3.org/standards/techs/rif

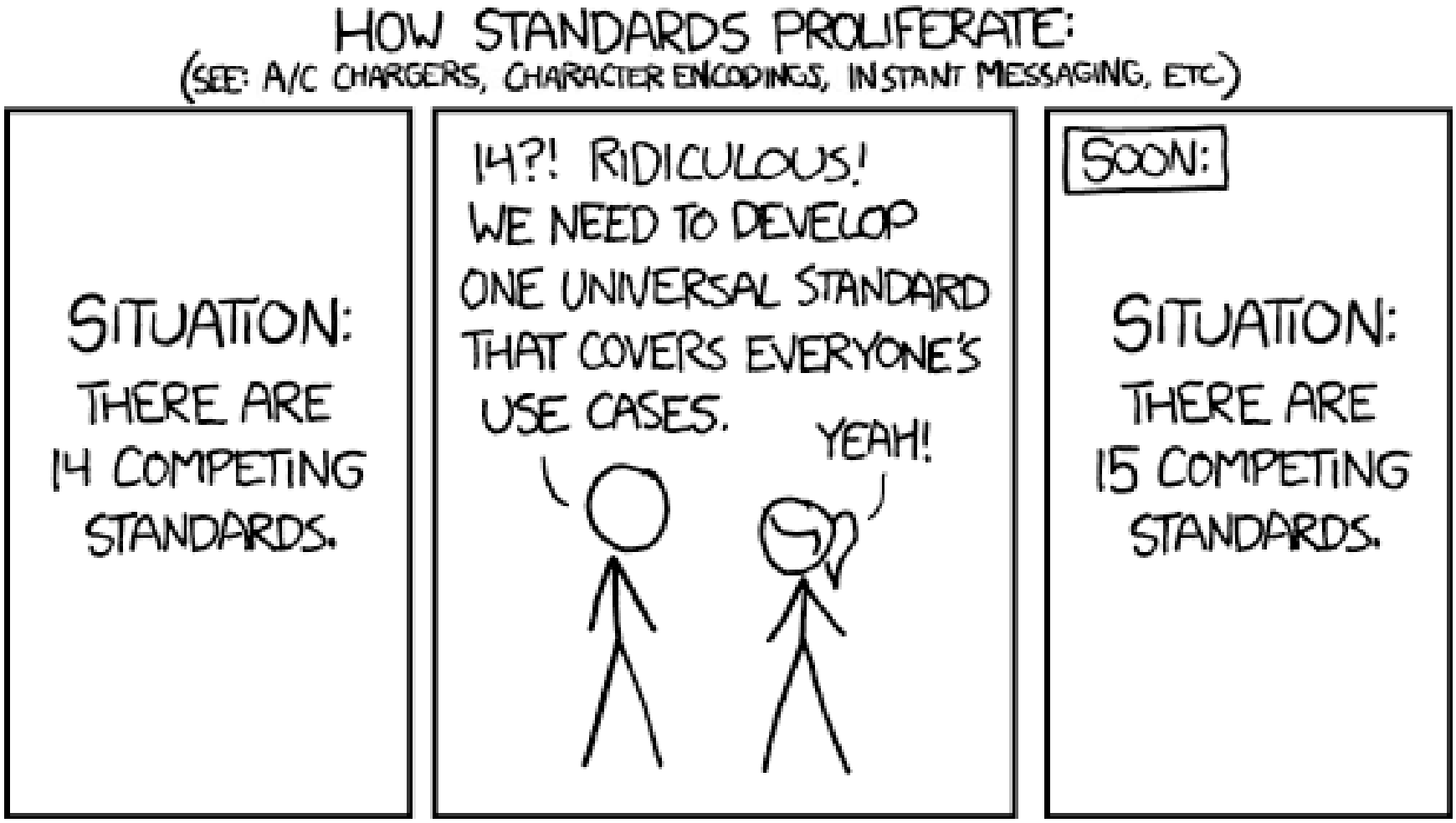
An **XML** standard language (using its own meta-syntax and structure) proposed to describe many (all?) rule formalisms so they each provide their own standardized form to one another.



Semantic Web Challenges

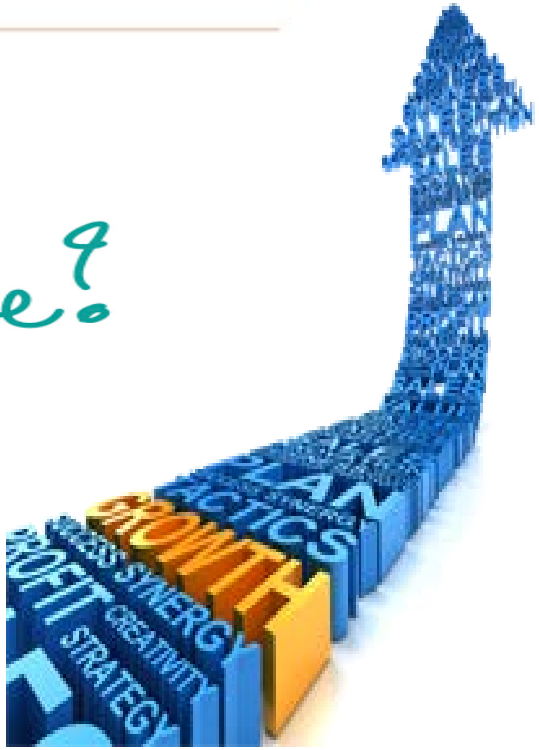
Standards galore ... but:

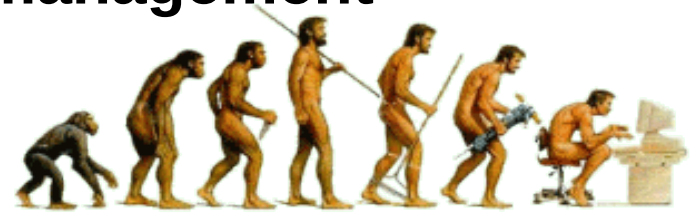
How many are **really used**? ... beyond trivial use cases.

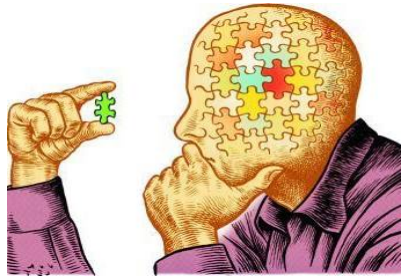


Semantic Web Reasoning Challenges

- Scalability
- Distribution
- Structural reasoning
- Temporal reasoning
- Approximate reasoning
- Learning
- Big Linked Data = “Blinked” Data?
- Knowledge evolution management
- Ethics
- ...

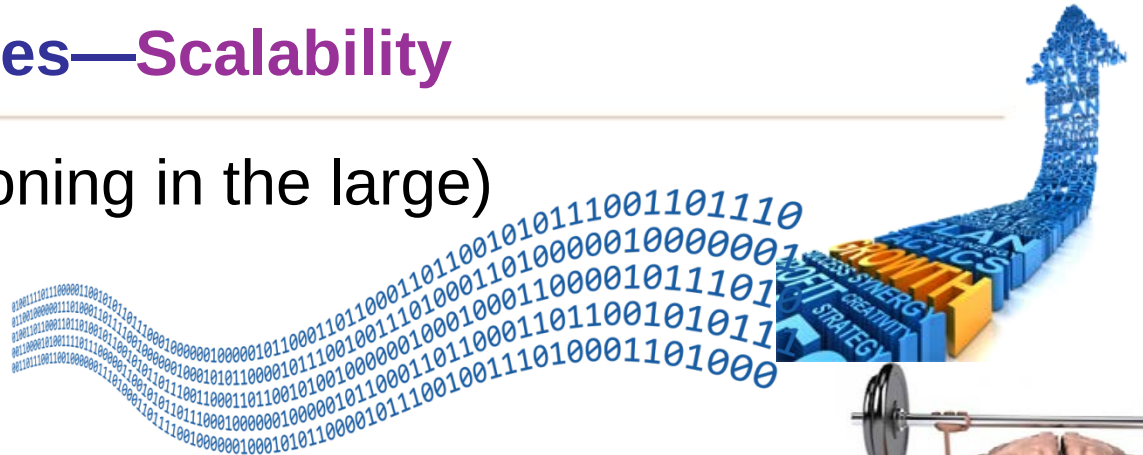






Semantic Web Challenges—Scalability

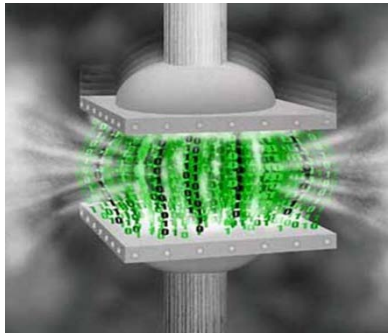
Scalability (affects reasoning in the large)

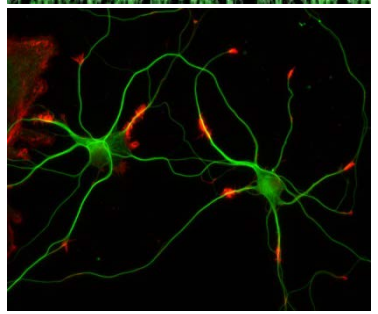


- **Performance**
  - Tbox reasoning (“ontological” reasoning)
  - Abox querying (where does the reasoning help?)

- **Data handling**
  - Big Data (synopsize the essence)
  - Linked Data (synaptic reasoning)
  - Big Linked Data (huge brain)

— “*Blinked*” Data?





Semantic Web Challenges—Distribution

Distribution (incremental data diffusion and coherence)

Triplestores in the Cloud

- **Performance**
  - Tbox reasoning (“ontological data” schema?)
  - Abox querying (SPARQL *vs.* NoSQL triple-as-relation)

- **Data handling**
  - Big Data (Relational/Semi-structured)
  - Linked Data (RDF Triples)
  - “Blinked Data?” (interconnected massive triplestores)

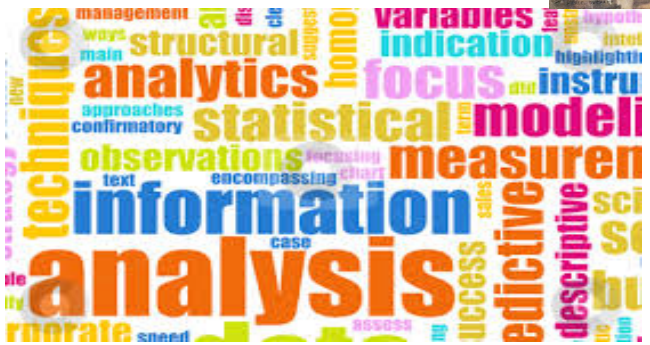
Semantic Web Challenges—Structural reasoning

Structural reasoning (deriving facts from facts)

- **Efficient knowledge processing**

- **Default tolerance (detail abstraction)**

- **Semantic context**



Semantic Web Challenges—Temporal reasoning

Temporal reasoning (taking time into account)

- **Event processing**

- **Time-relative logic**

- **Time-sensitive knowledge**

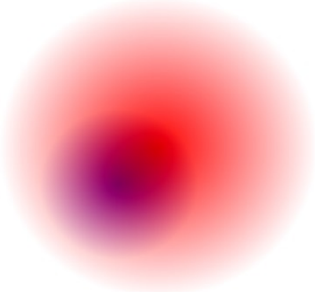
Semantic Web Challenges—Approximate reasoning

Approximate reasoning

(deriving partial knowledge from partial facts)

- **Probabilistic logic (Bayesian, Markovian)**

- **Fuzzy set logic**



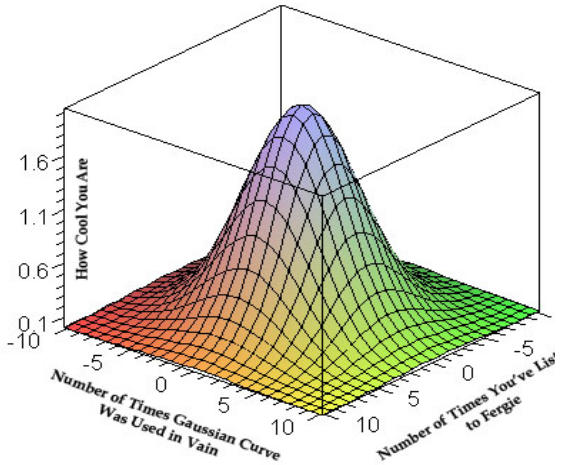
- **Rough set logic**

Semantic Web Challenges—Learning

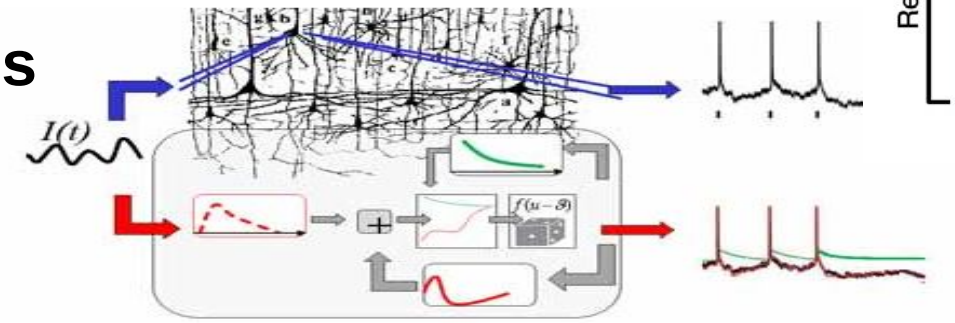
Learning—Abductive and inductive reasoning

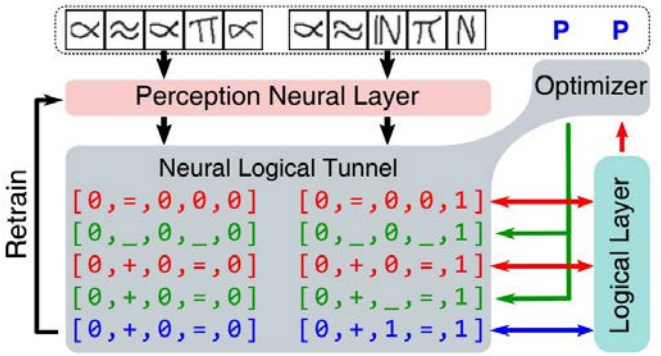
- **Structural learning**

- **Statistical learning**



- **Combinations**





Semantic Web Challenges—Linked Data

Linked data (interconnected computing)

- **Interconnectivity management**

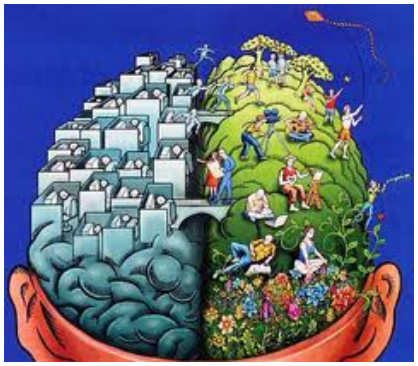
- **"Blinked Data"**

- **The Internet of Everything**

Semantic Web Challenges—Knowledge evolution

Knowledge evolution management
(take into account that things change)

- **Coherence maintenance**



- **Provenance and trustability**

- **Context management**

Semantic Web—Ethical challenges

Fundamental concern:

how may such power be made to **respect ethics?**

- **Who** can **control web technology** and **how**?

- Can **industrial profit** outweigh **social rights**?

- **International governments** and all **public sectors must catch up with this technology**.

Semantic Web = Worldwide Evolving Brain?

The essential argument is that it is expected that **standardized knowledge** can somehow **arise** and be **used** in the form of **ontologies** from **massively interconnected information**.

Such is the potential for **Linked Data**, for example.

Even if this can be achieved, yet another **challenge** for such **knowledge**, however it may be represented, is to be effectively, let alone **efficiently**, processed to **provide intelligence**.

The key is that, **whatever the standards may be**, one cannot escape the need for **formal encoding** of such knowledge to lend itself to **inference of implicit networked knowledge**, beyond the classical **processing of explicit silo-ed data**.

6th Generation Computing?

Hence, this all smells, tastes, and looks again like a **“been there, done that!”**; *viz.*, the promises of the **5th Generation Project** of the 80’s.

In fact, the SW’s objective is **much more challenging today** taking into account the **exponential explosion of data** and the inescapable **need for scalable processing**.

In addition, **cloud networking** and the ubiquitous **distribution of information** has made this task even more daunting.

“I think you should be more explicit here in step two.”

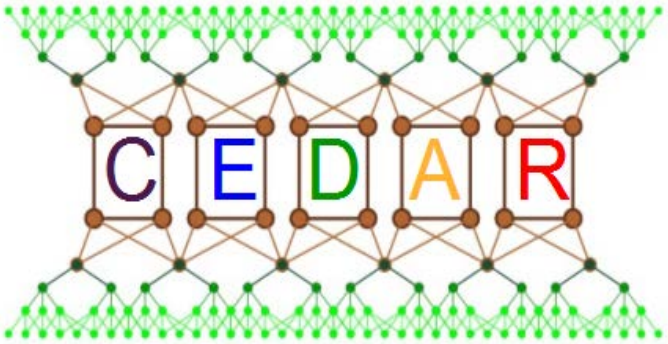
CEDAR—Constraint Event-Driven Automated Reasoning

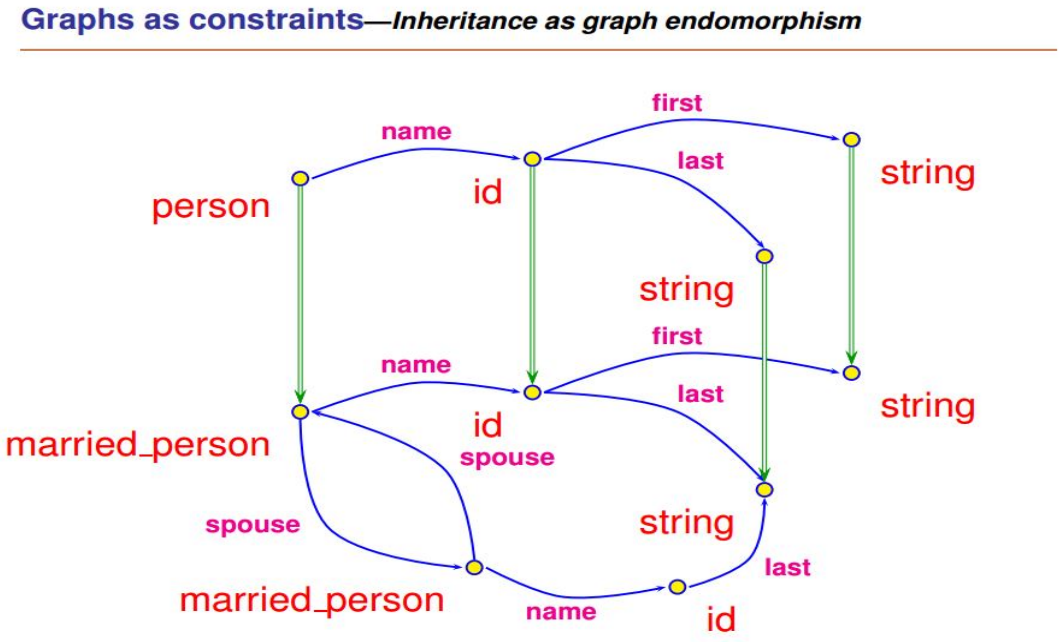
Owls break easily!

Is there a remedy?

Efficient automated reasoning with Order-Sorted Features

ANR funded chair of excellence – Jan. 2013 ➡ Jan. 2015



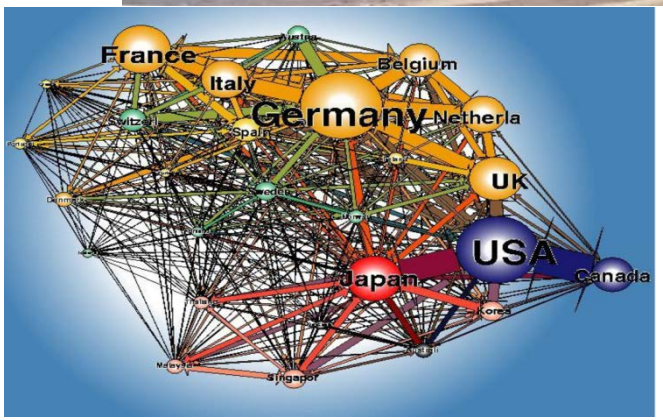


CEDAR—Scalability and Distribution

The CEDAR project addressed mainly two concerns:

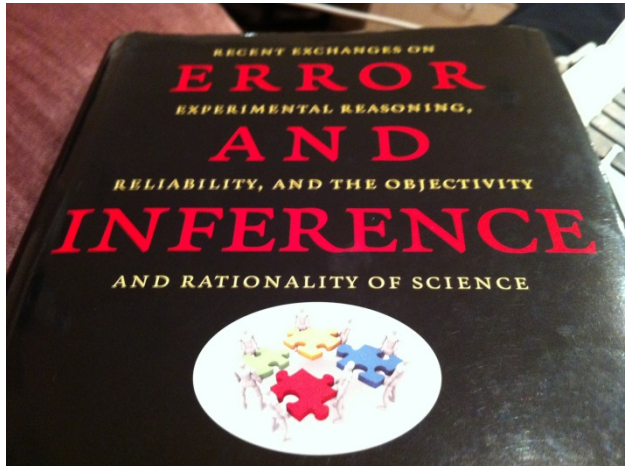
- **Scalability of ontological reasoning**

- Management and access of **distributed ontological knowledge** and “Blinked Data”

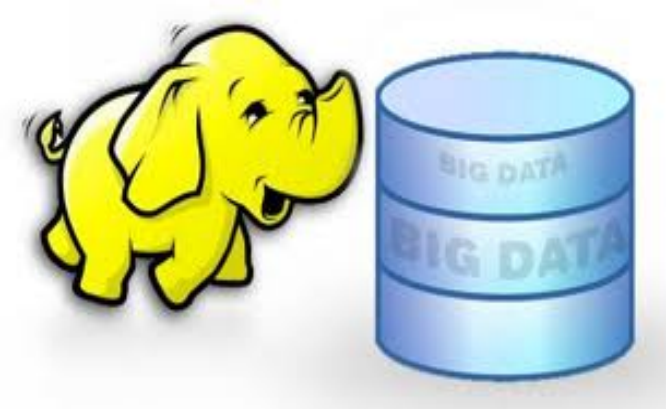


CEDAR—Scalability and Distribution

The CEDAR project's approach:



- experiment with **existing systems** ***vs.*** **our** own **reasoning technology**



- experiment with **Hadoop-style** architecture for **concurrent processing** **of** **distributed** knowledge as **"Blinked Data"**

Semantic Web—Where are we today?

If one must be critical:

- **W3C SW standards** have not really been tested
- **Viable alternatives** have not really been considered

However, all SW formalisms must **imperatively** take into account the **formidable challenges** described above.

Namely:

any **knowledge representation** and **efficient inference** based on them must be **scalable**, **incremental**, capable of dealing with **approximate** information (*probabilistic, fuzzy, paraconsistent, …*) in **real time**, and manage data of **enormous size** and **diversity** that is **distributed** all over the Internet.

Semantic Web—Where we may be tomorrow?

We have surveyed a few **challenges** and **potentials** faced by the **W3C** to make the **Semantic Web** a **reality**.

- So may we expect the W3C **to give meaning to the Web**?

- Such a large effort is bound to produce unexpected **serendipitous offshoots** — the objective itself is irrelevant even if not achievable: *"Moon Technology" was a benefit even if no one is there* (yet?) …

- For it to do so, it must **adapt to** any **unexpected** reshaping of the **(computing) world**, taking every opportunity to make what is possible become **real**.

Thank You For Your Attention !

Hassan Aït-Kaci

hak@acm.org