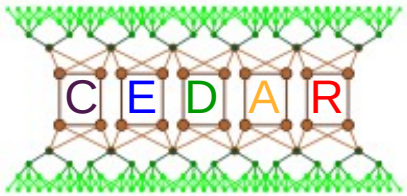
Is It Possible to Make the Semantic Web a Reality?

Hassan Aït-Kaci

ANR Chair of Excellence

Université Claude Bernard Lyon 1

Constraint Event-Driven Automated Reasoning Project



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 rate at which it accumulates. **Implicit in this data hides a *wealth* of information**—literally!

A recent article in the **Wired magazine** illustrates this by reporting the noticeable prediction success of a small data analysis company called **Recorded Future** whose main office is located in Gothenburg, Sweden.

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***.

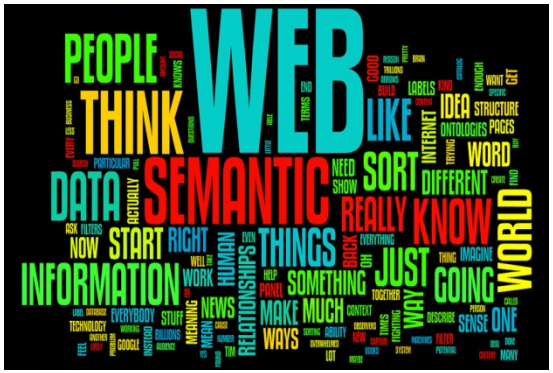
Wherein Lies the Knowledge?

Yet, as blatantly 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: *how to extract and use knowledge hidden but implicit in public data.* And we're talking about **Big Data**!

The “Semantic Web”?

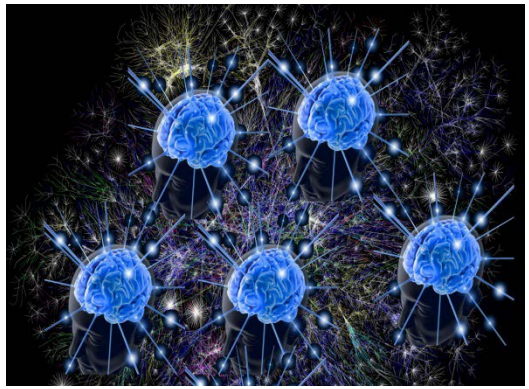


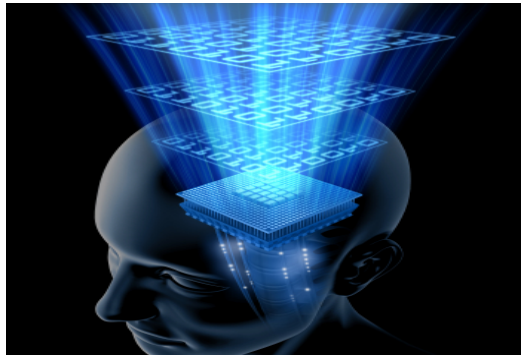
It has been now over a decade that the **Semantic Web** has been heralded as the means to infuse **meaning** into the World-Wide Web.

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 knowledge** of **massively interconnected standardized information**.

Semantic Web Challenges





- Reasoning with interconnected information
- Automate its knowledge structuring (standard?)
- Automate its reasoning power
- Need to agree on (a) standard(s)

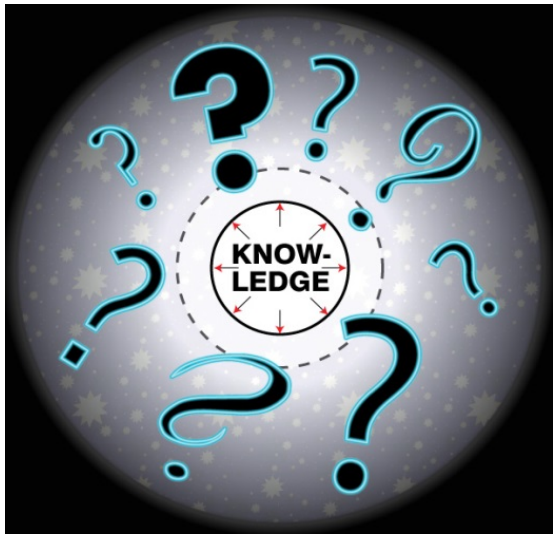
Standards—KIF

In AI, **KIF** is not a narcotic but it means:

Knowledge **I**nterchange **F**ormat

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.



Standards—RIF

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

Rule Interchange Format

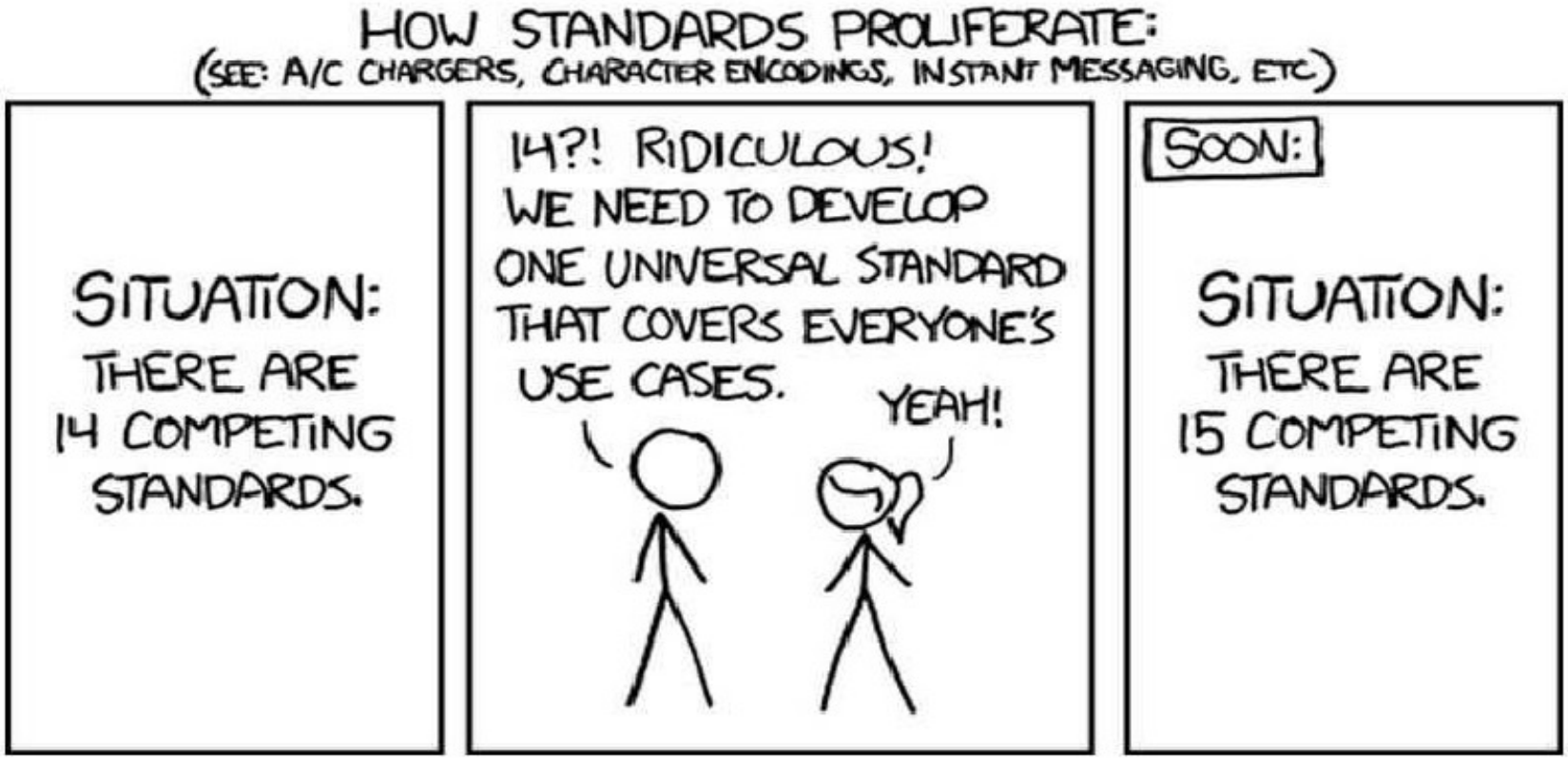
http://www.w3.org/standards/techs/rif

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

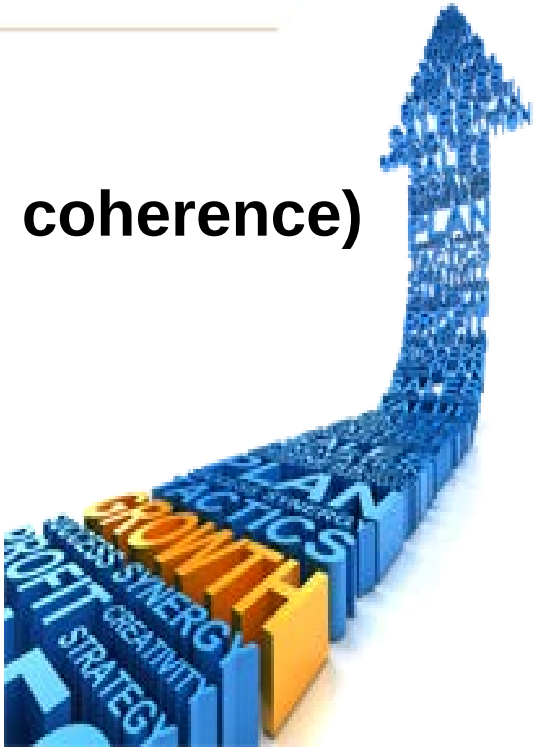
Standards galore ... but:

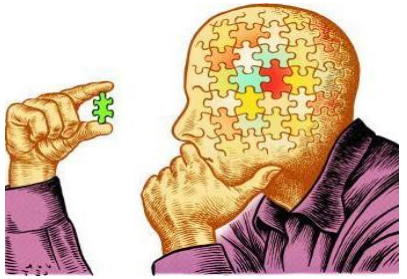
How many are **really used**? ... beyond trivial use cases.

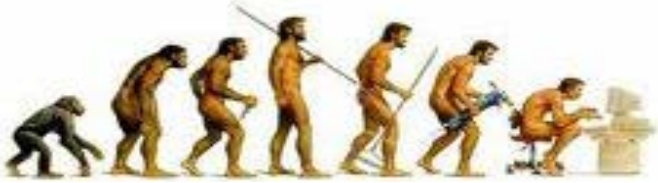


Semantic Web Reasoning Challenges

- Scalability
- Distribution (incrementality, data diffusion and coherence)
- Structural reasoning
- Temporal reasoning
- Approximate reasoning
- Learning—Abductive and inductive reasoning
- Big Linked Data = “Blinked” Data?
- Knowledge evolution management
- ...

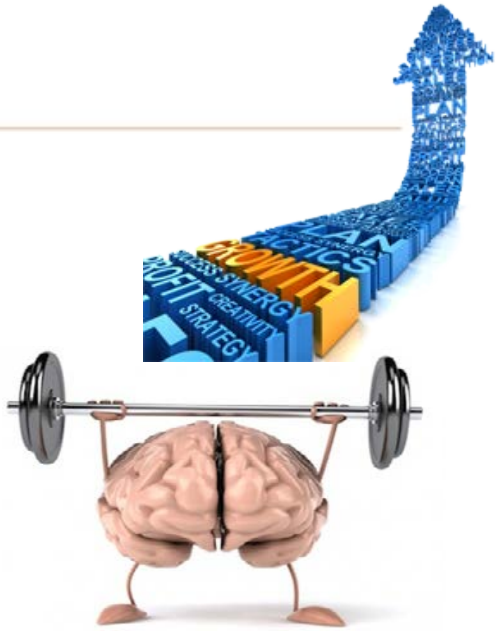






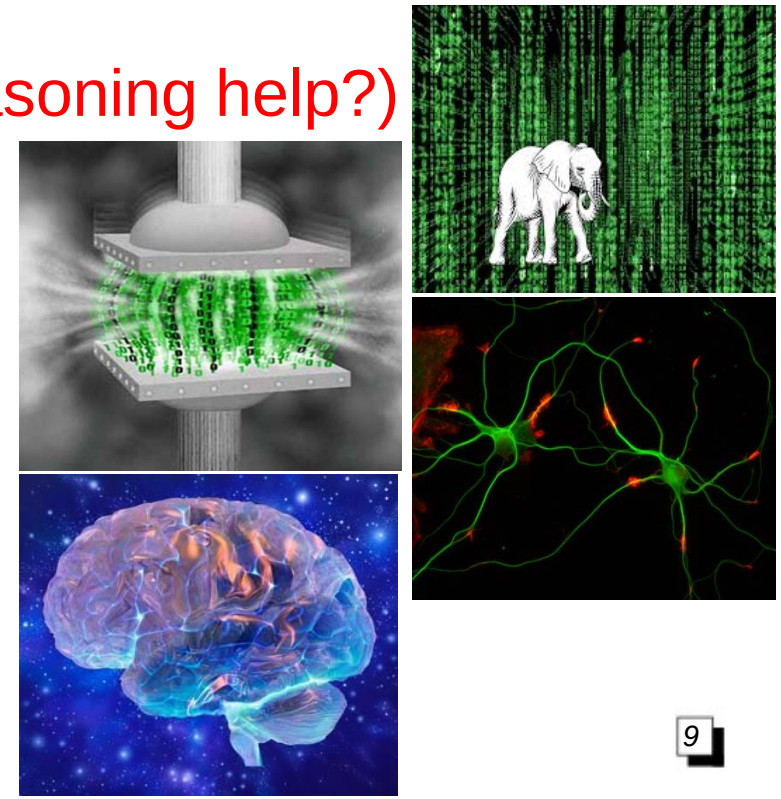
Semantic Web Challenges—Scalability

Scalability



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)
  - “Blinked Data?” (huge brain)



Semantic Web Challenges—Distribution

Distribution (incrementality, data diffusion and coherence)

Triplestores in the Cloud

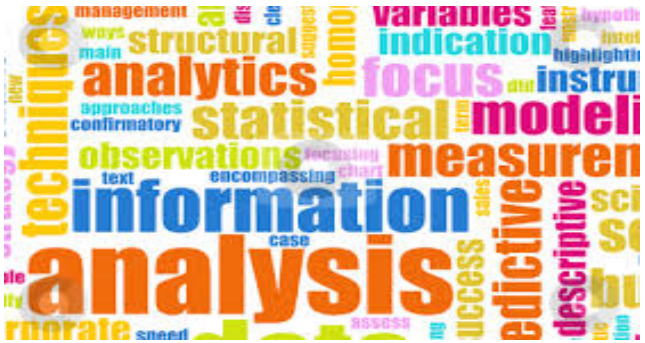
- **Performance**
  - Tbox reasoning (“ontological data” schema?)
  - Abox querying (SPARQL *vs.* NoSQL triple-as-relation)
- **Data handling**
  - Big Data (Relational/Semi-structured)
  - Linked Data (RDF Triples)
  - “Blinked Data?” (interconnected massive triplestores)

Semantic Web Challenges—Structural reasoning

Structural reasoning

- Efficient knowledge processing

- Default tolerance (detail abstraction)



- Semantic context

Temporal reasoning

- Event processing

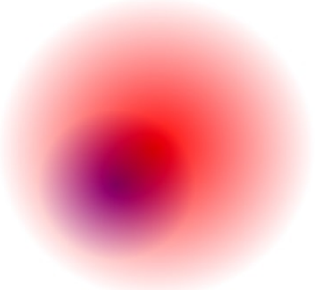
- Time-relative logic

- Time-sensitive knowledge

Approximate reasoning

- Probabilistic logic (Bayesian, Markovian)

- Fuzzy set logic



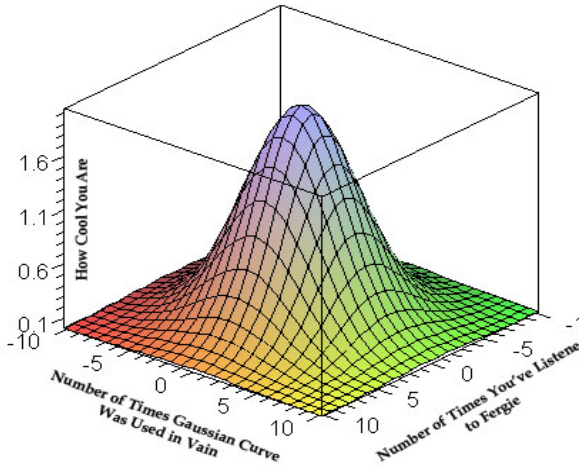
- Rough set logic

Semantic Web Challenges—Learning

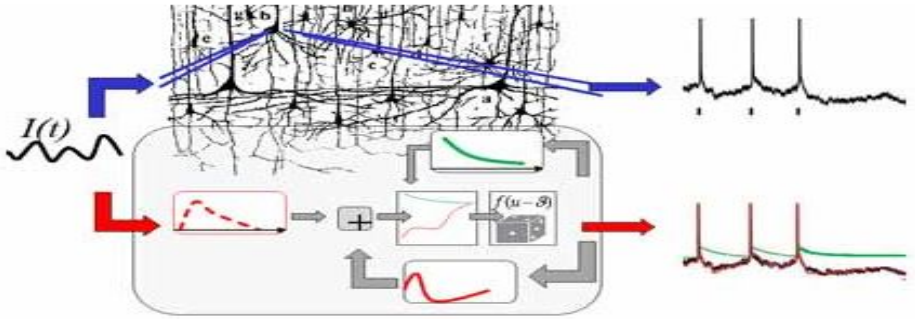
Learning—Abductive and inductive reasoning

- Structural learning

- Statistical learning



- Combinations



Semantic Web Challenges—Linked Data

Linked data

- interconnection management

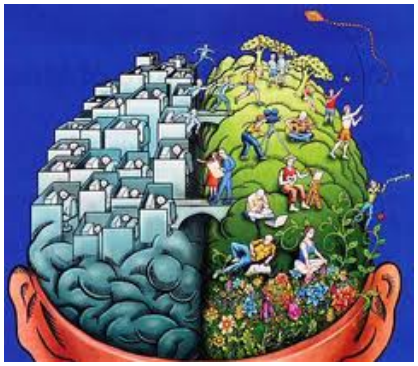
- “Blinked Data”

- Combinations

Semantic Web Challenges—Knowledge evolution

Knowledge evolution management

- Coherence maintenance



- Provenance and trustability

- Context management

CEDAR—Constraint Event-Driven Automated Reasoning

Véritable mémoire du temps,

le cèdre de l’Atlas nous raconte l’Histoire ...

OMAR MHIRIT & MOHAMED BENZYANE

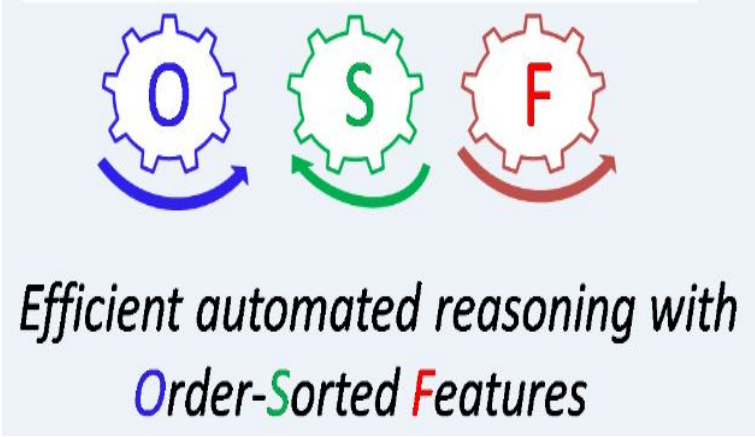
Le Cèdre de l’Atlas : Mémoire du Temps

http://books.google.fr/books?id=6wFPkWJ0PTEC

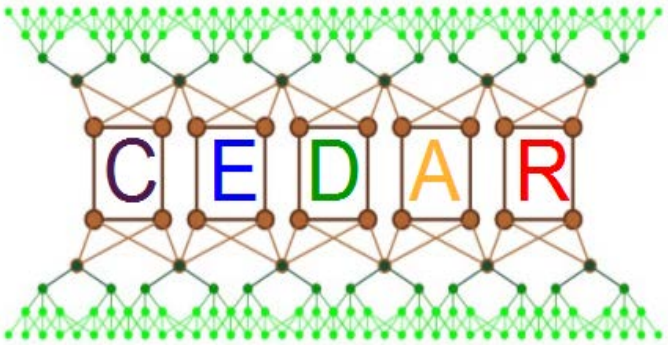
CEDAR—Constraint Event-Driven Automated Reasoning

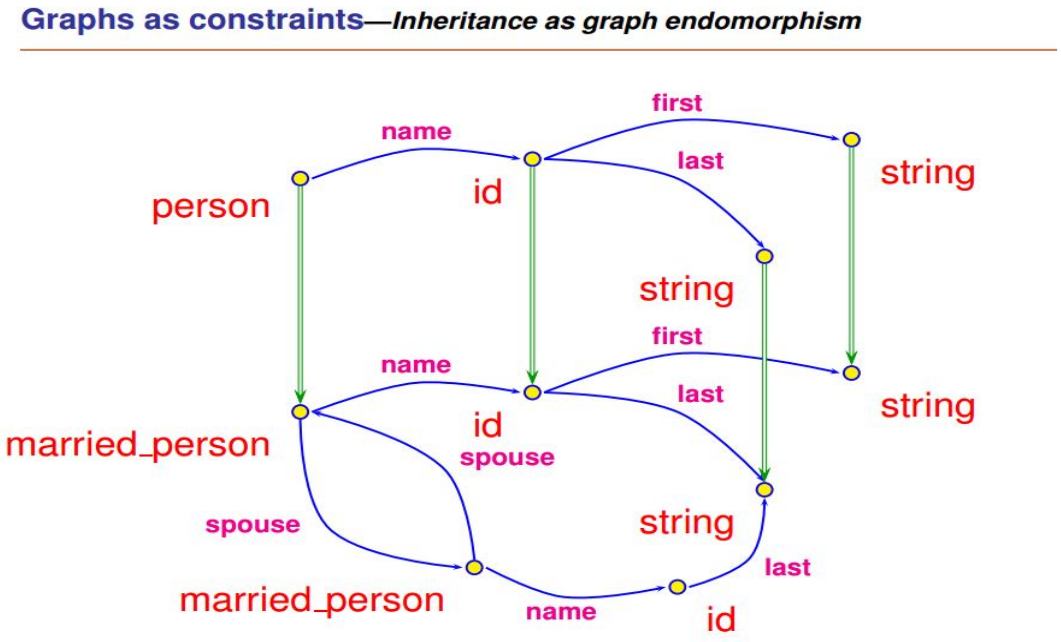
Owls break easily!

Is there a remedy?



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

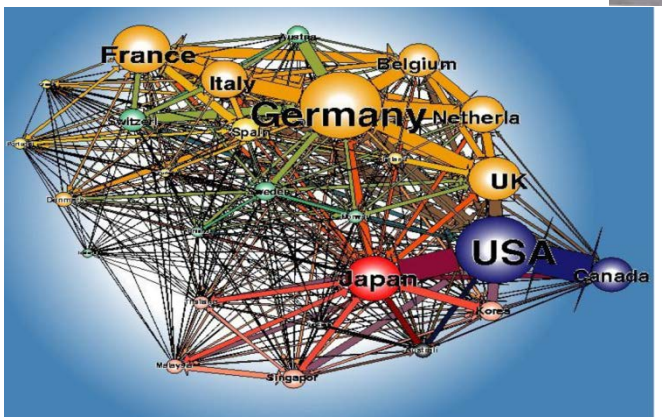




CEDAR—Scalability and Distribution

The CEDAR project addresses mainly two concerns:

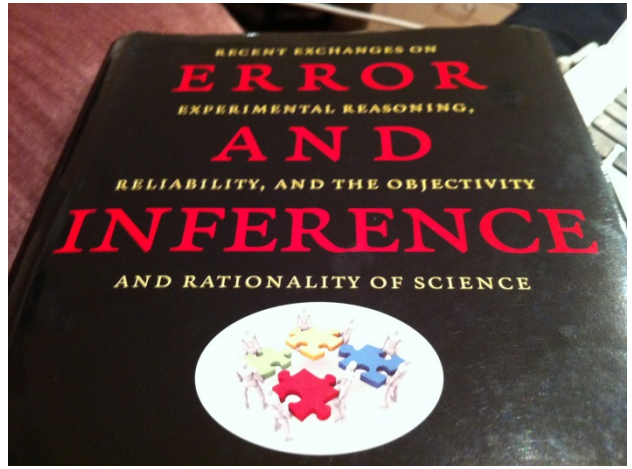
- **Scalability of ontological reasoning**
- Management and access of **distributed ontological knowledge** and “Blinked Data”



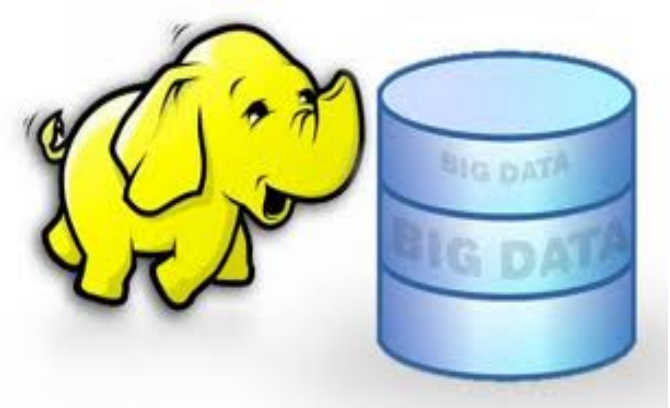
CEDAR—Scalability and Distribution

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 and "Blinked Data"

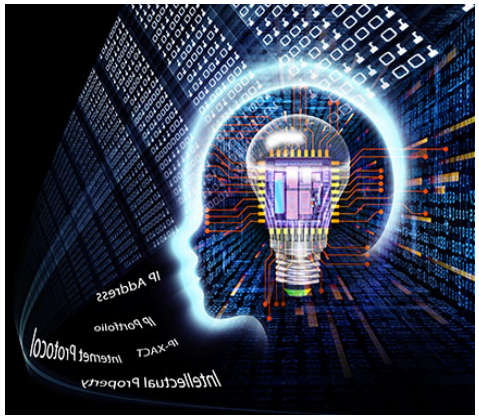


Semantic Web = World-Wide Brain?

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 such a hope could 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.”

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 it must be **scalable**, **incremental**, capable of dealing with **approximate** data (fuzzy, probabilistic, incomplete) in **real time**, and manage information of **enormous size** and diversity **distributed** all over the Internet.

Semantic Web—Where we may be tomorrow?

So how may we expect the Semantic Web to turn into a reality?

- We have surveyed a few **challenges** and **potentials** faced by the **W3C** to make the Semantic Web a reality.
- Such a large effort is bound to produce unforeseen serendipitous offshoots in the same manner as the "MoonTechnology" of the 60's did pursuing **JFK**'s otherworldly dream of human moon settling.

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

Thank You For Your Attention !

For more information:

hassan.ait-kaci@univ-lyon1.fr

http://cedar.liris.cnrs.fr