

# Some Personal Thoughts on Semantic Web and “Non-symbolic” AI

Ora Lassila, Ph.D

Research Fellow, Nokia Research Center  
Cambridge, Massachusetts, USA

October 2008

URSW 2008 workshop keynote address



Nokia Research Center

 ISWC 2008

# Garbage In, Garbage Out...



**COMPUTERWORLD**  
Business Intelligence

Print Article Close Window

## Greenspan, Cox tell Congress that bad data hurt Wall Street's computer models

Patrick Thibodeau

October 23, 2008 ([Computerworld](#)) WASHINGTON — Using nature to describe a man-made financial disaster, [Alan Greenspan](#), former chairman of the [Federal Reserve](#), today called the [financial-market meltdown](#) a "once in a century tsunami" and explained to a [U.S. House](#) committee what he thought went wrong. And insufficient data was one of the causes he pointed to.

Greenspan has long praised computer technology as a tool that can be used to limit risks in financial markets. For instance, in 2005, he credited improved computing power and risk-scoring models with making it possible for lenders to extend credit to subprime mortgage borrowers.

But at a [hearing](#) held today by the House Committee on Oversight and Government Reform, Greenspan acknowledged that the data fed into financial systems was often a case of garbage in, garbage out.



# It is all about data

(I will explain...)



Nokia Research Center

 ISWC 2008

# Original Semantic Web Vision

[Berners-Lee, Hendler & Lassila 2001]

- (Advertised as) the next generation of the World Wide Web
- Make Web content amenable to automated processing
  - interpretation by machines
  - (most content out there is for human consumption)



- Implies the use of
  - artificial intelligence
  - agents

# Original Semantic Web Vision

[Berners-Lee, Hendler & Lassila 2001]

- (Advertised as) the next generation of the World Wide Web
  - Make Web content amenable to automated processing
    - interpretation by machines
    - (most content out there is for human consumption)
- ↓
- Implies the use of
    - artificial intelligence
    - agents

- Next generation of personal computing
  - Computers working on behalf of users
    - (current usage: as tools)
    - more autonomy, handling unanticipated situations
- ↓
- Implies the use of
    - artificial intelligence
    - agents



# Original Vision Deconstructed

- Heavily predicated on
  - multi-agent technologies
  - ontologies + associated reasoning
  - availability of data
- General idea is that agents
  - access data
  - use ontologies to interpret data
  - draw conclusions (to the benefit of human users)



# Original Vision Was Criticized as “Science Fiction”

- Where does all the data come from?
- What makes us think agents will get access to all the data?
- What would make organizations provide data in “semantic” form?
- How can we get any agreement on what ontologies to use for data?
- ...



# Original Vision Was Criticized as “Science Fiction”

- Where does all the data come from?
- What makes us think agents will get access to all the data?
- What would make organizations provide data in “semantic” form?
- How can we get any agreement on what ontologies to use for data?
- ...

**It is all about data**





# The “AI Winter” Revisited

- 1980s “AI boom” ended in the early 1990s in a severe “hangover”
  - many promises of “intelligent” software could not be fulfilled
  - expectations vastly exceeded practical results
- For the past decade, we have seen a revival of interest in AI
  - (the “AI Spring”)



# The “AI Winter” Revisited

- “New”, successful AI often “non-symbolic”
  - fuzzy logic, neural networks, machine learning, data mining, ...
  - note: there is a lot more data available now (thanks to the Web)
- Classical, “symbolic” AI mostly seen as unrealistic and idealistic
  - aspirations of completeness, consistency, ...
- New methods work in practice
  - (without us necessarily understanding why...)

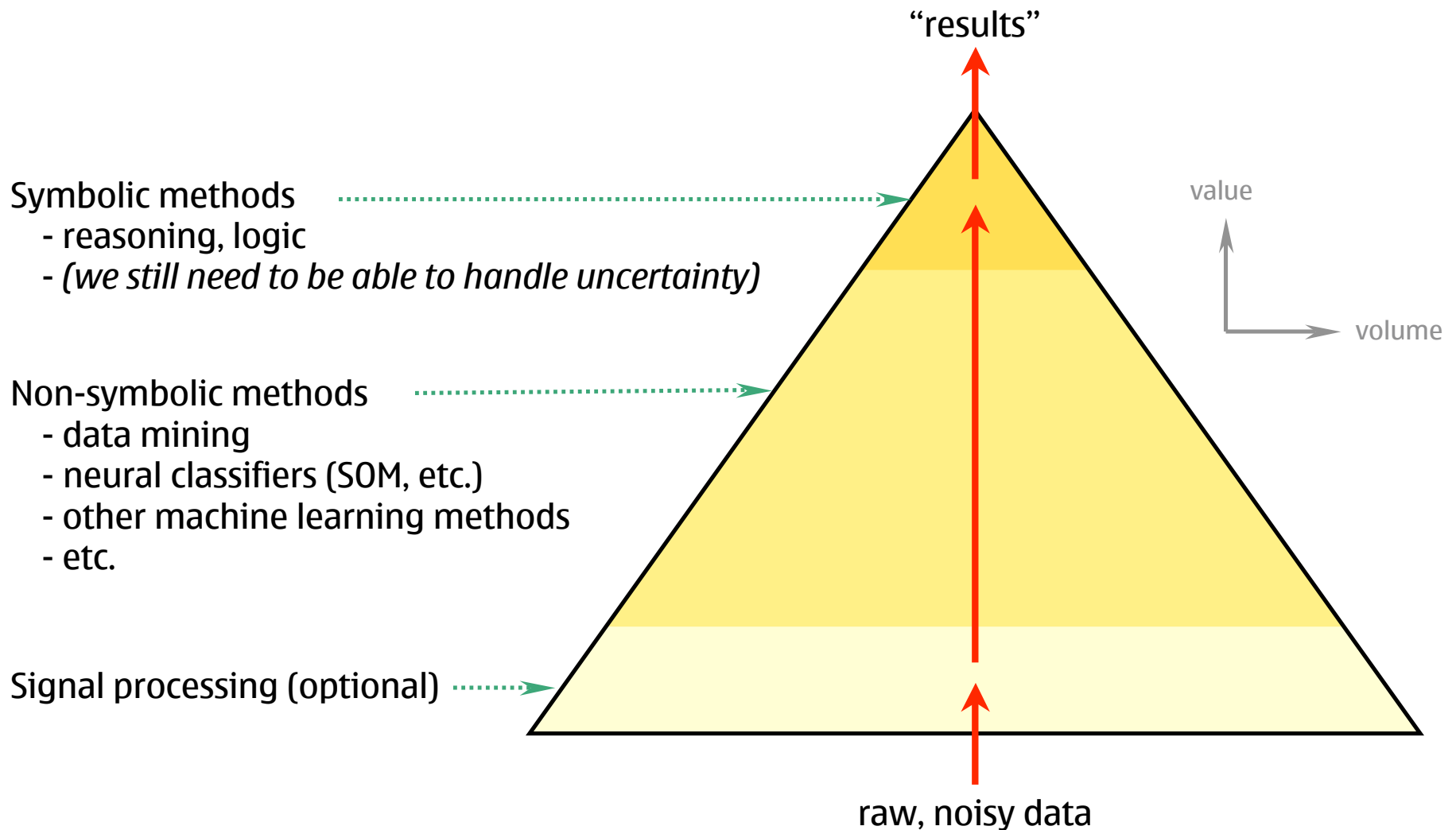


# “Diluting” the Semantic Web Vision

- Various versions of the Semantic Web vision have appeared
  - “Giant Global Graph”, Linking Open Data, “Data Web”
  - “lowercase” semantic web, microformats
- Trying to remove the AI component from the vision
  - (it seems that this only postpones the inevitable...)
- Emphasis is on data
  - how to link data sets
  - uniform representation
- Issues with entity-resolution, object identification remain

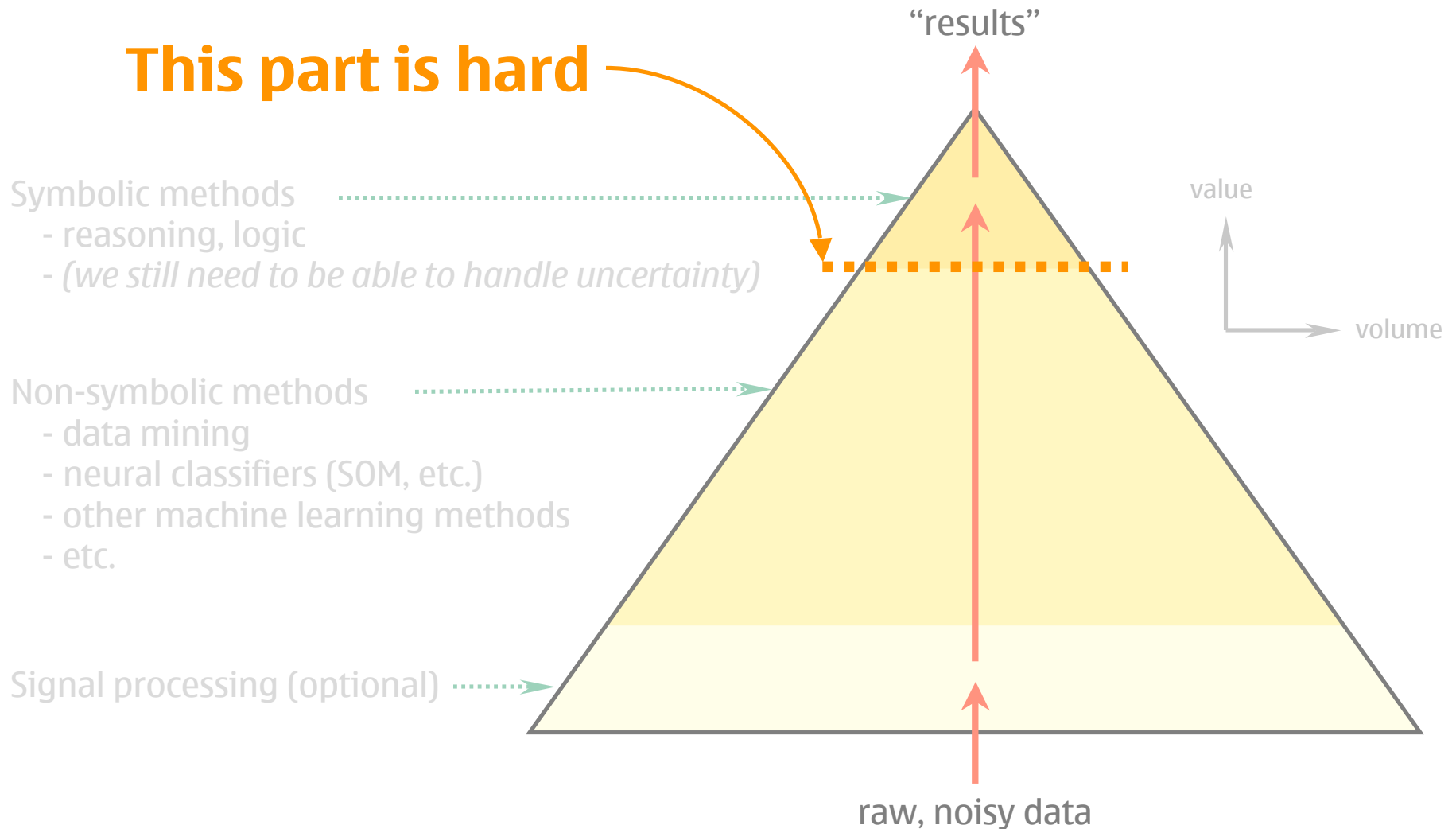


# “Data Value Chain” (abstract, conceptual view)



# “Data Value Chain” (abstract, conceptual view)

**This part is hard**



# My Group's Experiments: Context-Awareness

- Context derivation [Lassila & Khushraj 2004, 2005]
  1. data “clean-up”
  2. DL reasoning + rules
    - some colleagues of mine believe they can do it all using just non-symbolic methods [e.g., Flanagan, Mäntyjärvi & Himberg 2002]
- User modeling with activity capture [van Kleek & Shrobe 2007]
- Auditory context capture [Perttunen, van Kleek, Lassila & Riekkilä 2008]



# My Group's Experiments: PIM

- Capturing user notes as structured, actionable data
  - “natural” language → RDF [van Kleek, Bernstein, Karger, schraefel 2007]
- Music metadata applications
  - noisy data, lots of entity resolution issues [Khushraj, in progress]
- Virtual personal assistant
  - speech and dialogue -based user interface [Adler et al, in progress]



# Common Denominators in My Group's Experiments

- Semantic Web data models, logic-based reasoning
- Data must first be “cleaned-up” using non-symbolic or heuristic methods
- Problems: uncertainty, unreliability





# Conclusions

- I believe the original Semantic Web vision is valid and worth pursuing
  - it describes the future of personal computing (not the future of the Web)
  - it implies a fundamental change in how we use information technology
- The vision is predicated on pervasive availability of data
  - real-world data is noisy (and must be cleaned up)
  - (business and social issues remain, in addition to technical issues...)
- Non-symbolic AI methods have enjoyed great success lately
  - let's use them to make better quality data available
- Even after clean-up, issues remain with data
  - uncertainty, unreliability, ...
  - (dealing with these is an integral part of the success of the Semantic Web)
  - note that “diluted” versions of the vision (e.g., linked data) have the same problems...



# Questions? Comments?

- <mailto:ora.lassila@nokia.com>