

Semantic Knowledge for Commodity Computing

Evelyne Viegas
Director – Semantic Computing
Microsoft Research

Semantics as the study of Meaning

- Data semantics – extract and map from structured and semi-structured sources into ontologies
- Lexical semantics – identify/learn concepts, roles from sentences (e.g. Powerset; MindNet)
- Statistical semantics – discover meaning from patterns of use (e.g. concept similarity)
- Computational semantics – automate the process of constructing and reasoning with meaning representations
- Semantic web – linked data via URI, common graph structure with RDF, inferences via ontologies and OWL
- Formal semantics – in linguistics? in logic?

Semantics is hard

- At a Santa Fe (NM) petrol station
 - “We will sell gasoline to anyone in a glass container”
- On a long-established NM dry cleaning store
 - “Thirty eight years on the same spot”
- In the offices of a loan company
 - “Ask about our plans for owning your home”
- “Laptops under \$1000 and with 2 or more GB of RAM”
- “Are there any friends of a friend of mine at the next conference?”
- “I’m really hungry”

Vision – Enable *Next Generation Experiences* by working with academia, stakeholders from industry, government, and consumers/innovators to make sense of data

**DATA > INFORMATION > KNOWLEDGE >
INTELLIGENCE**

Data/Information

- To help explore the data value chain, Microsoft's collaborations provide access to data that enables:
 - Innovation – By having access to real world data, researchers can unveil new analysis or research directions based on shared assets and explore new questions
 - Science – By allowing wider use of data, repeatability of experiments can be performed and data misrepresentations or faulty results avoided
 - Training – real-world large-scale data is a powerful tool for training the next generation of data analysts and researchers
- Cloud-based services: Web Language and Query Language Models
 - Used to research topics such as human speech, spelling, information extraction, learning, and machine translation.

Data-Driven Research

Researchers need **access** to large scale real world data, and infrastructure to drive innovation, enable science (repeatability)

- [Search Summit 2007](#) new asks:
 - Need more data, larger scale;
 - Need to follow a user (privacy!)
- [Beyond Search – Semantic Computing and Internet Economics 2009](#)
new asks:
 - Need data access (as opposed to data release);
 - Compute power
- [Language Model Services](#) (Web Ngram and Query Language Models)
 - Access to large scale real world data via cloud-based Services

Multi-word Tag Cloud from Government Dataset Titles

Single Tag Cloud



Multi Tag Cloud



Ref: Dr. Li Ding, Rensselaer Polytechnic Institute

Next Generation Experiences

- Shift from users understanding computers to computers understanding users
 - Data → Information → **Knowledge** → **Intelligence**
- Data is a 1st -class citizen under different multimedia encoding: text, speech, gesture, image, pen, sensor , touch, video
 - Semantics is emerging as a unifying paradigm
 - Context, *beliefs* need to be taken into account

Knowledge/Intelligence

- From *data services* to *knowledge services*
 - Future of information processing and discovery at web-scale?
 - Role of collective intelligence?
- Transforming semantic knowledge into commodity computing
 - Opportunities
 - Challenges

Semantic Knowledge for Commodity Computing: Myth or Reality?

- *Ontologies for Real World*—[Deborah McGuinness](#), Rensselaer Polytechnic Institute
- *A Probabilistic Knowledgebase for Text Understanding*—[Haixun Wang](#), Microsoft Research
- *Data Challenges in the Real World*—[Shoshanna Budzianowski](#), Microsoft

Semantic Knowledge for Commodity Computing: Myth or Reality?

- *Text and Context: Using Context to Better Understand Searcher's Intentions*—[Susan Dumais](#), Microsoft Research
- *Learning to Map Sentences to Meaning*—[Luke Zettlemoyer](#), University of Washington
- *Citizen Sensing—Opportunities and Challenges in Mining Social Signals and Perceptions*—[Amit Sheth](#), Wright University
- *Semantics and Knowledge-Enabled Experiences for Search and Discovery*—[M. Srikanth](#), Microsoft

Thanks!

- Questions?
- More info: <http://research.microsoft.com/~evelynev>
- Contact: evelynev@microsoft.com



Microsoft Research
FacultySummit

© 2011 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.

Microsoft Research

Faculty Summit



FUTURE WORLD

2011 ← 2031