



Computer-assisted Semantic Annotation of Scientific Life Works

Professor Edward Feigenbaum
Dr. Thomas Gruber
Will Snow

the lineup

- **Imagine what** Prof. Edward Feigenbaum
 - **Consider why** Dr. Thomas Gruber
 - **Discuss** Will Snow
-

Imagine

This is an incredible learning resource.

Why isn't there one of these for every member of the scientific community?

Universal Leonardo

1450 1470 1480 1490 1500 1510 1520

DRAWINGS
INVENTIONS
MANUSCRIPTS
PAINTINGS
LIFE & TIMES

Visit Explore Play Discover Browse Research About Us

Explore

Explore the trails on the right to find out how in Leonardo's view of the world, all things are interconnected - the motion of water and the curling of hair, the human body and the mechanisms of machines, the geometrical rules that govern man, animals and all of nature.

Alternatively, the interactive timeline at the top of the page provides a visual representation of the trails by linking related images. Click on a coloured dot to start a trail.

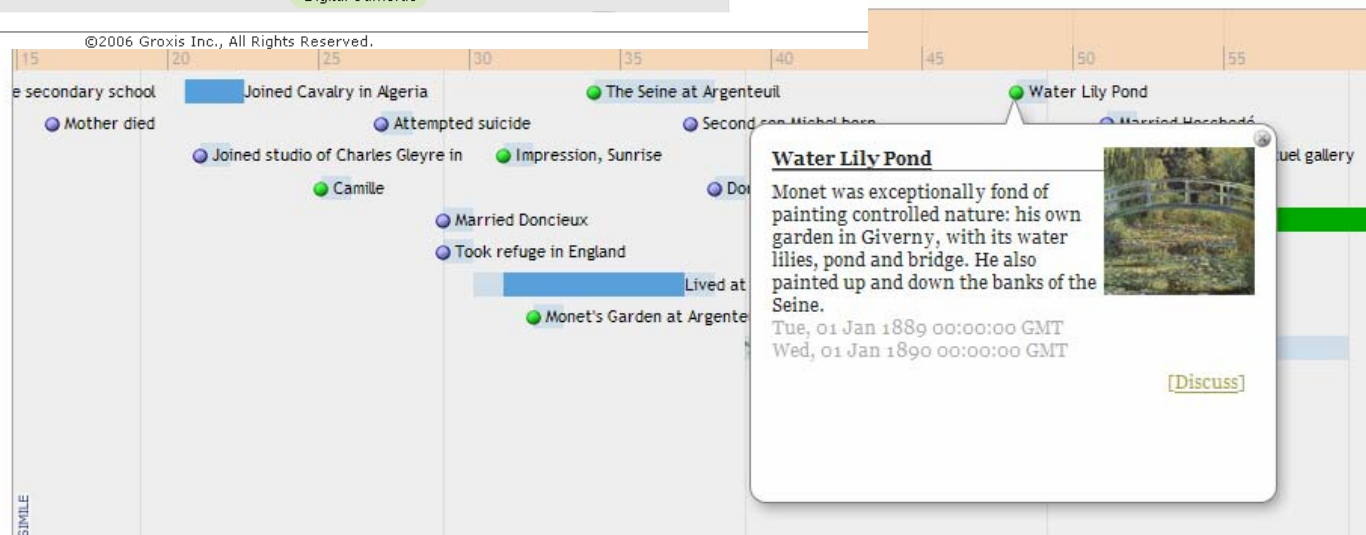
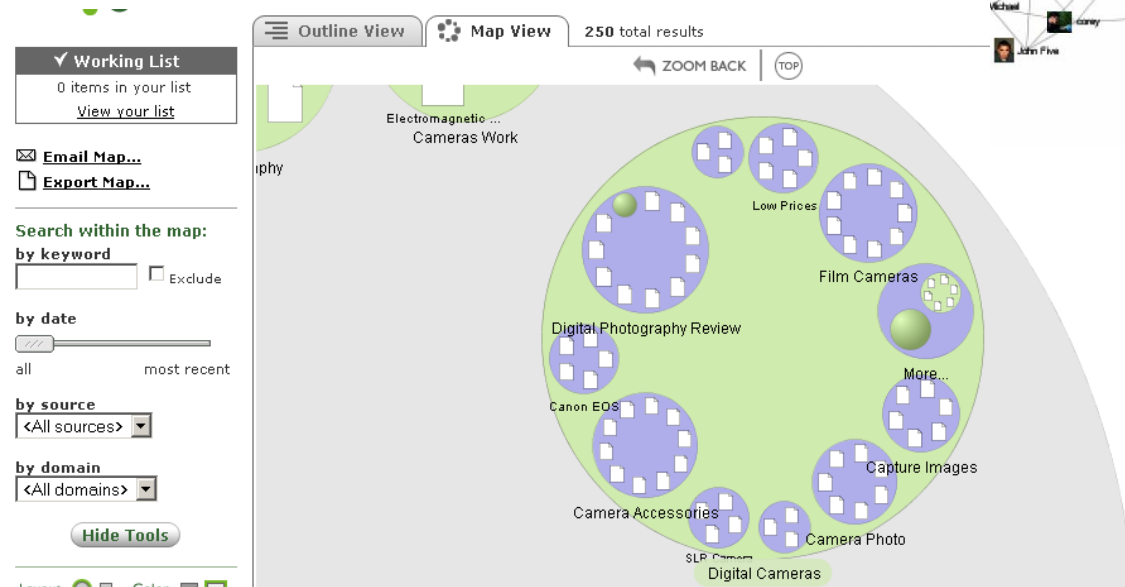
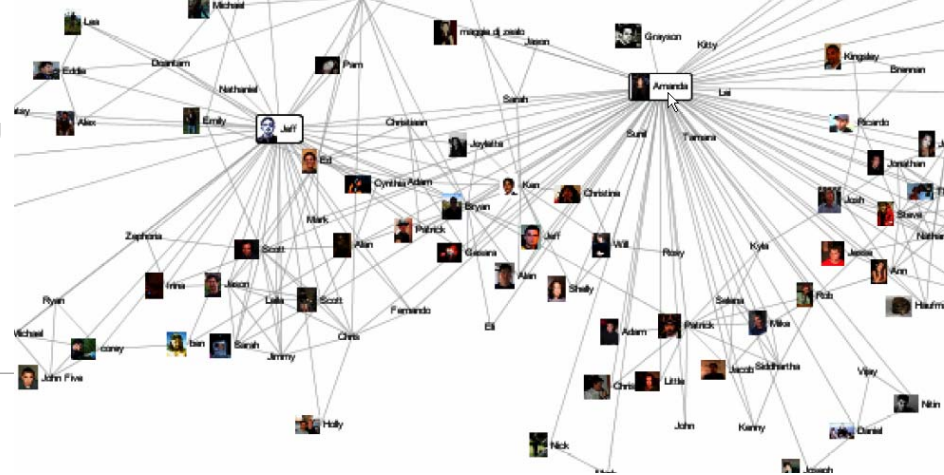
Leonardo is often presented as a scientist, artist or engineer. But for him, all natural phenomena are the product of the same natural forces and governed by the same natural law.

Both the engineer and artist must learn how nature designed its forms according to their function and obey the same laws. Science, art and engineering are all achieved by direct observation and scientific investigation of the natural world. As such, they are all part of the same creative vision, through which man can create a "second nature" in the world.

Rule of Mathematics

- The Body of Earth**
Leonardo saw the planet earth as a living entity, with all of its elements in a constant state of flux...
[start >](#)
- Forces of Nature**
For Leonardo, flowing water, curling hair and the growth patterns of plants were all manifestations of the same natural force...
[start >](#)
- The Body of Man**
According to Leonardo, man was nature's most perfect creation, "the measure of all things"...
[start >](#)
- The Natural World**
Leonardo studied man, animals, plants and all natural phenomena intensely in his quest to understand all natural things...
[start >](#)
- Imagination and Invention**
By combining scientific observation with imagination and invention, the painter had the power to create "fictions that signified great things"...
[start >](#)
- Light and Vision**
Leonardo believed that sight was the most important of all the senses, the eye being the "window of the soul"...
[start >](#)
- Remaking Nature**
For Leonardo, the artist's task was to remake nature in his art, rather than slavishly copy natural forms...
[start >](#)
- Rule of Mathematics**
All things in nature are governed by mathematics - "Let no-one who is not a mathematician read my principles"...
[start >](#)

Have knowledge, can fly.



Whose work? Just look around.

■ CS

- Ed Feigenbaum, John McCarthy, Don Knuth

■ Chemistry

- Carl Djerassi

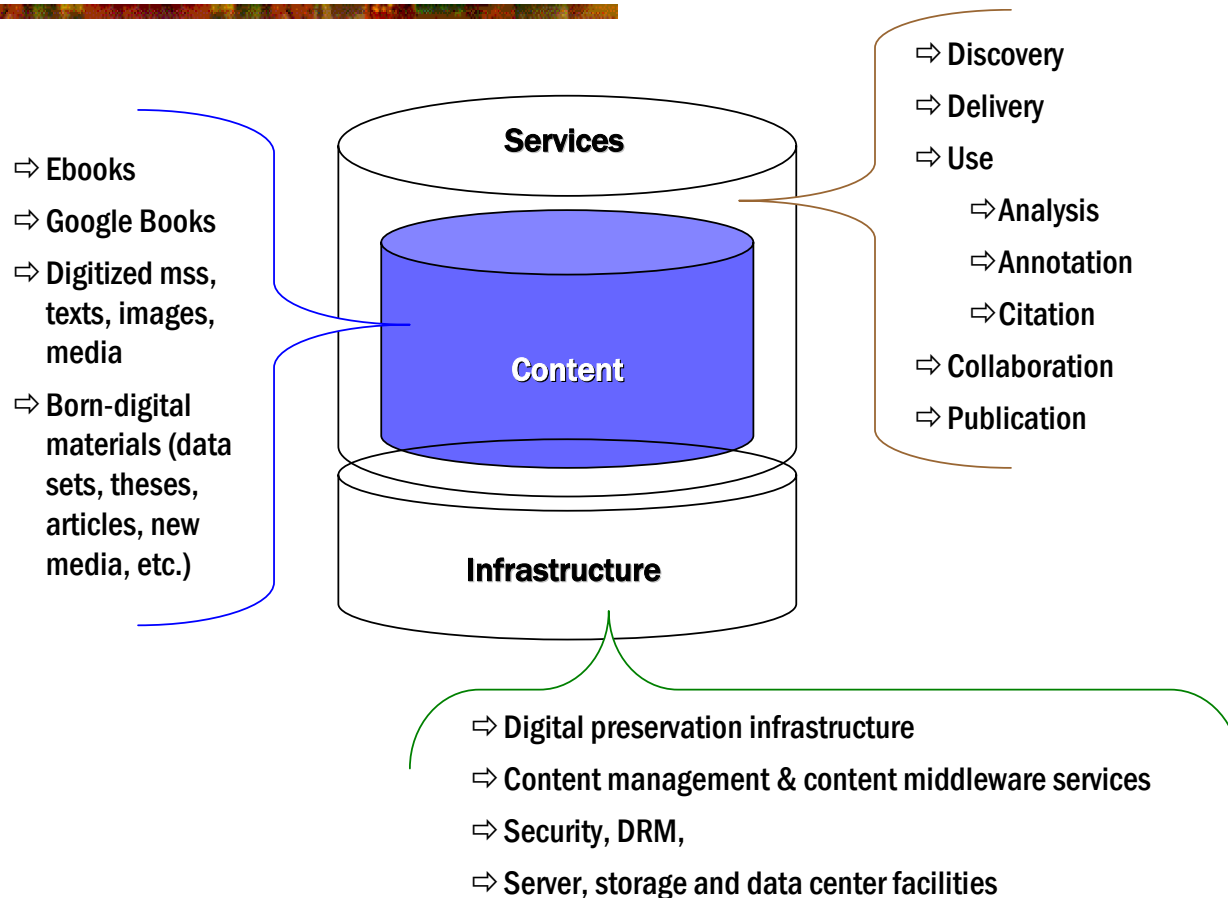
■ EE

- Andrew Grove, John Hennessy

■ Local and Historical Context

- Stuart Brand, John Warnock, David Kennedy, Jim Adams, Roger Kornberg ...
-

Where? How about Stanford's Digital Library

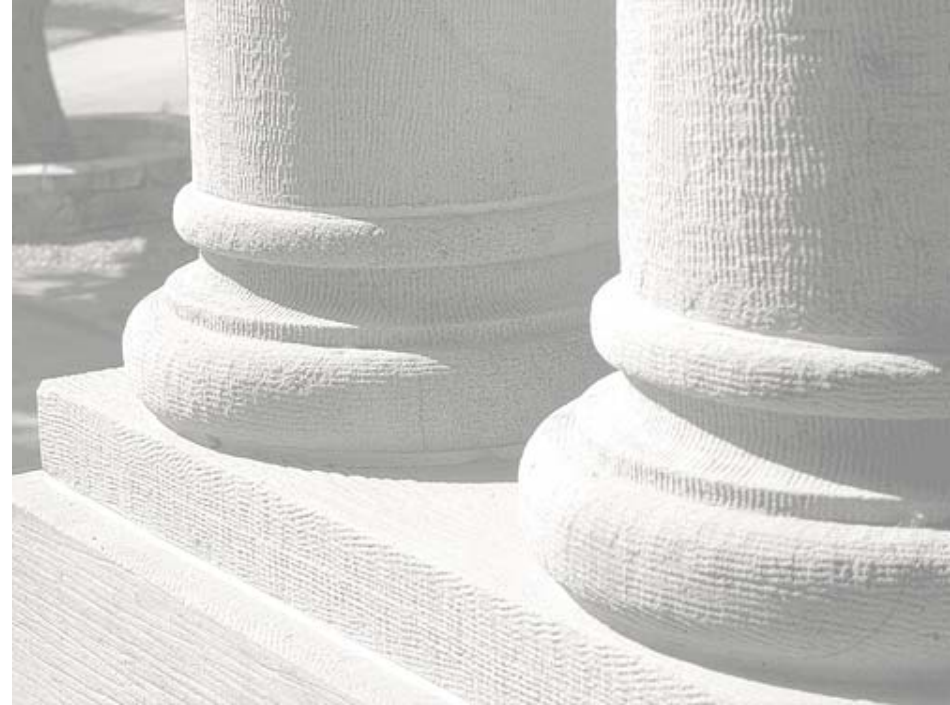


You are invited.

- Research Assistants wanted
 - PhD Students
 - *can lead to thesis*
 - MS students
 - *great prep for Web 3.0 companies*
-

The Research

or, why this is an opportunity to
make a difference in the history
of science and technology

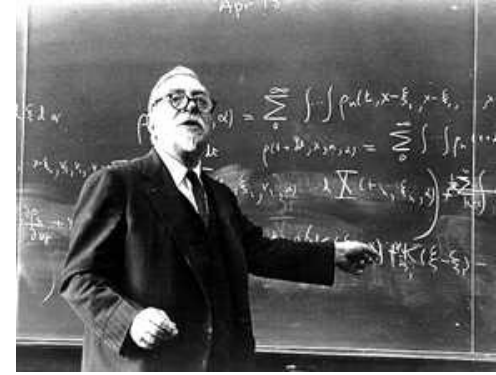


Milestones in the History of Technology for Intelligence

- 1940s Cybernetics – Norbert Wiener
 - 1960s Augmentation – Doug Engelbart
 - 1970s Early AI – McCarthy, ...
 - 1980s Knowledge Systems – Feigenbaum, ...
 - 2000s Collective Intelligence – WWW, Google, Wikipedia, ...
-
- *What do they have in common?*
-

Cybernetics (1940s-50s)

Norbert Wiener

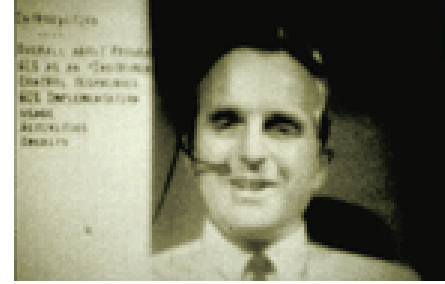


- Goal: improve human performance
- Technology:
 - *control systems with humans in the loop*
- *Insight:*
 - Intelligence comes from “**Man-machine synergy**”
(and you can’t just throw algorithms at the data)

“Many people suppose that [computing machines] are replacements for intelligence and have cut down the need for original thought. ... This is not the case. If simple devices need simple thought to get the most out of them, complicated devices need a vastly reinforced level of thought. ... Moreover this work cannot be put off until the machines have already processed their data. It is very rare, and to say the least, by no means normal, that data that has been thoughtlessly selected can be organized by an afterthought so as to produce significant results.” - N.W. 1948

Augmentation (1960-70s)

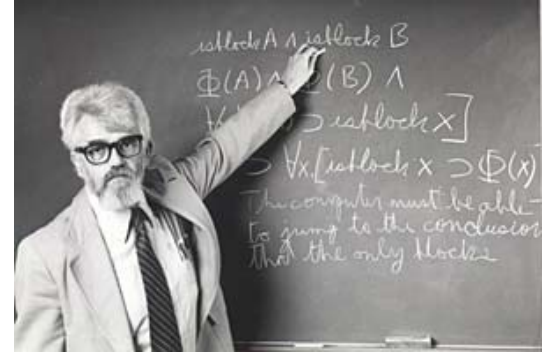
Doug Engelbart



- Goal: “improve collective IQ”
- Technology
 - *net-based collaboration*
 - *high bandwidth human interfaces*
 - *external, collective memories using semistructured representation*
- *Insight:*
 - *Technology can **augment** collective human knowledge, not replace it.*

Early AI (1960-80s)

John McCarthy et al.



- Goal: give machines human-level common sense
- Technology
 - *time sharing – machines should be flexible like people*
 - *high level languages – machines should be easier to teach*
 - *formal representations of knowledge and inference engines*
- *Insight:*
 - *Automating intelligence can be achieved by **representing knowledge** in machine-understandable forms.*

Expert Systems (70-80s)

Ed Feigenbaum et al.



- Goal: Expert-level competence
 - *in diagnosis, classification, configuration, etc.*
 - Technology
 - *mechanical inference on formally represented knowledge*
 - *Insight:*
 - *Automation of intelligence is limited by the bottleneck of acquiring **human knowledge**.*
-

Collective Intelligence

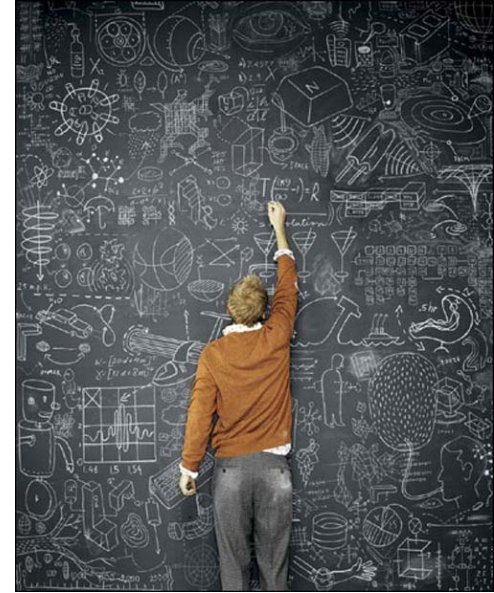
WWW, Google et al.



- Goal: access to “all the world’s knowledge”
 - Technology
 - *Google’s breakthrough:*
harnessing human input to bias inductive inference
 - *Insight:*
 - *Scalable computations on vast sources of **human knowledge**, with **human** feedback and interpretation.*
-

The Challenge for Today

- How to get the *best* of our collective knowledge?
 - *What is best done by machines?*
 - *What is best done by humans?*
 - *What is the best application of **both**?*



Opportunity to meet the challenge

■ Focused domain:

- *digitized life-work collections of scientific careers*

■ Powerful technology:

- *semantic information processing algorithms*

■ Potential for impact:

- *bring the legacy of eminent scientists to the world, now and for the future*
-

Semantic Annotation of Life-Work Collections

Document from Corpus:
Digitized and
Semantically Tagged

The Semantic Web : Suppliers and Customers

The notion of the **Semantic Web** can be coined as a Web of data when bringing database content to the Web or as a Web of enriched human-readable content when encoding the semantics of web-resources in a machine-interpretable form.

It has been clear from the beginning that realizing the **Semantic Web** vision will require interdisciplinary research. At this the fifth ISWC, it is time to re-examine the extent to which interdisciplinary work has played and can

semantic Web
research
citation and
in a new

issues
ing
rawing

lines and
often ask

from
e

Semantic
f
s issues
Web

Representations of

- research topics
- domain entities
- people
- places
- times
- events
- relationships among all

Named Entities:
*Identified by Computer,
Proofed by Human*

Persons

Abraham Lincoln
George Washington

Organizations

US Congress

Places

San Francisco
Mexico

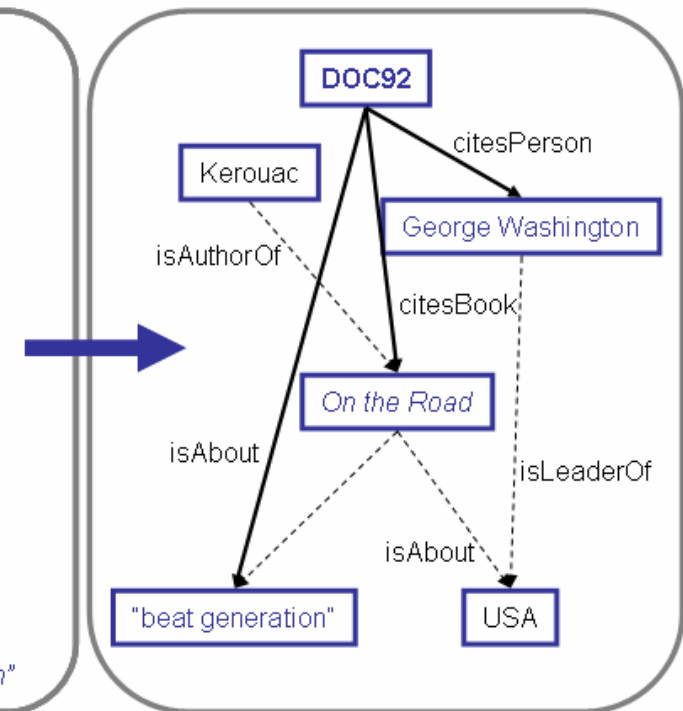
Citations

Statistical NLP, Manning
On the Road, Kerouac
De Senectute, Cicero

Concepts

"growing old"
"beat generation"
"syntactic disambiguation"

Relationships (triples):
*Human & Computer
Work Together*



What to do with the knowledge?

- Trace the history of an important idea
 - DIY “James Burke Connections”
 - State of the art prior art
 - Computational autobiography
 - How do paradigm shifts work?
 - Where is the next New, New Thing? 😊
-

The Scientific and Engineering Questions

- What are the limits of automation?
 - *and how to evaluate beyond precision/recall*
 - What tasks can be done by humans?
 - *and how to make them efficient*
 - What are the 10x breakthroughs
 - *new kinds of knowledge to be revealed*
 - *new kinds of inferences to be discovered*
 - *revolutionize research into scientific life works*
-

Potential Generalizations

- enable semantic annotation of
 - *your life work as you create it*
 - *history as it is made*
 - create the public toolkit for “semantic OCR”
 - *transform the work of archivists, historians of science*
-

Engage.

- opportunities to participate
Feigenbaum@cs.stanford.edu
- copy of this talk with links
tomgruber.org/writing



Further Readings

- *Cybernetics: or Control and Communication in the Animal and the Machine* by Norbert Wiener, 1948.
 - *A Conceptual Framework for the Augmentation of Man's Intellect* by Doug Engelbart, 1962. <http://www.bootstrap.org/>
 - *Netizens: On the History and Impact of Usenet and the Internet (Perspectives)*, by Michael Hauben, Ronda Hauben, Thomas Truscott, 1997. http://www.firstmonday.org/issues/issue3_8/chapter6/index.html
 - *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*, by Fred Turner, 2006. <http://fredturner.stanford.edu/>
 - *Collective Knowledge Systems: Where the Social Web meets the Semantic Web*, by Tom Gruber. *Journal of Web Semantics*, 2007. <http://tomgruber.org/writing/collective-knowledge-systems.htm>
 - *Whither Academic Information Services in the Perfect Storm of the Early 21st-century?* by Mike Keller. http://www-sul.stanford.edu/staff/pubs/keller_biconf06_finalpaper.pdf
-