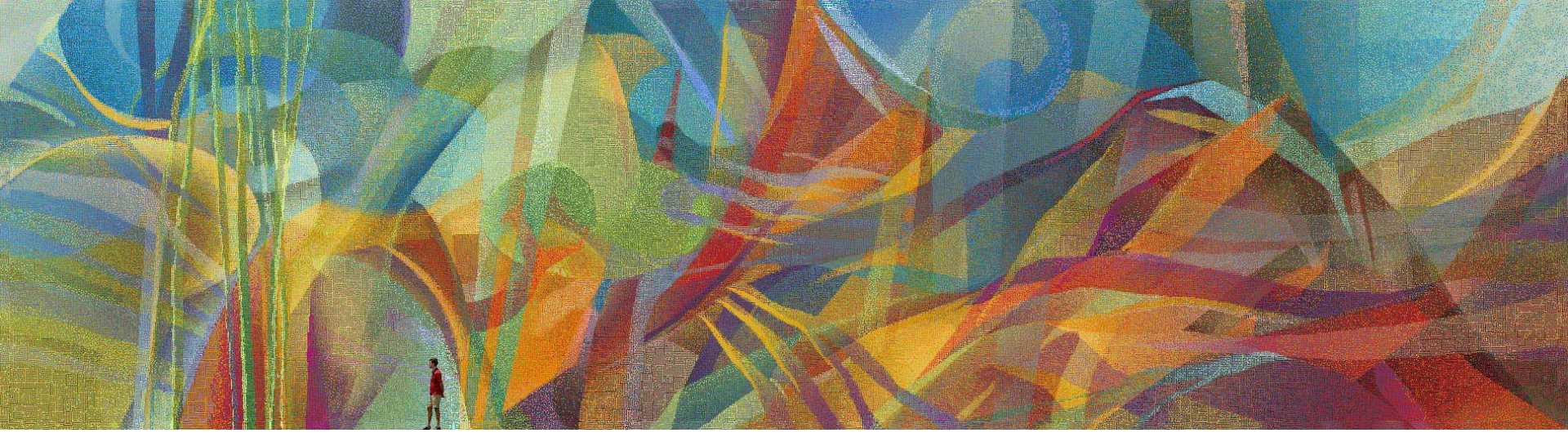


Computer-assisted Semantic Annotation of Scientific Life Works

Professor Edward Feigenbaum
Dr. Thomas Gruber
Will Snow

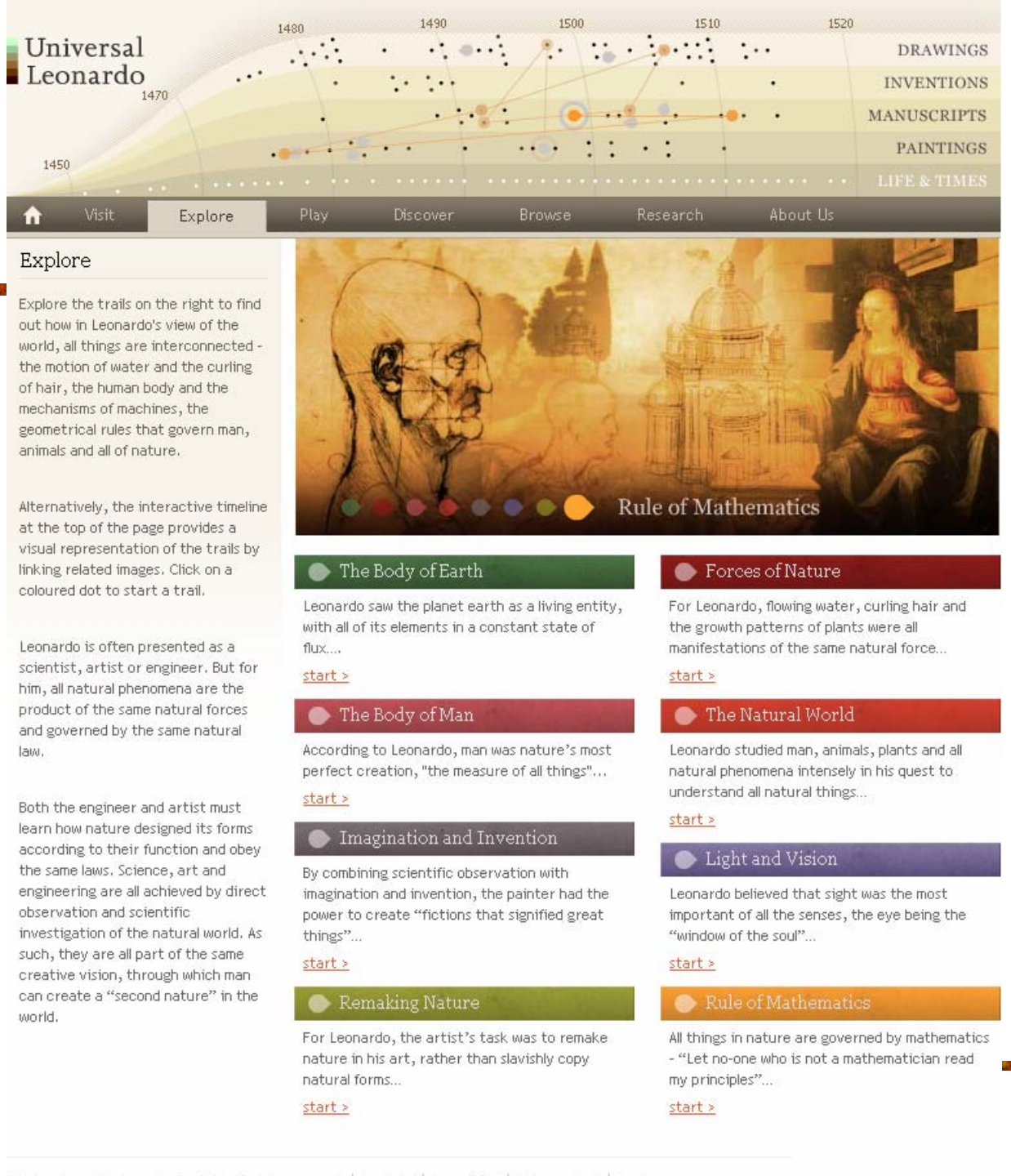
the lineup

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

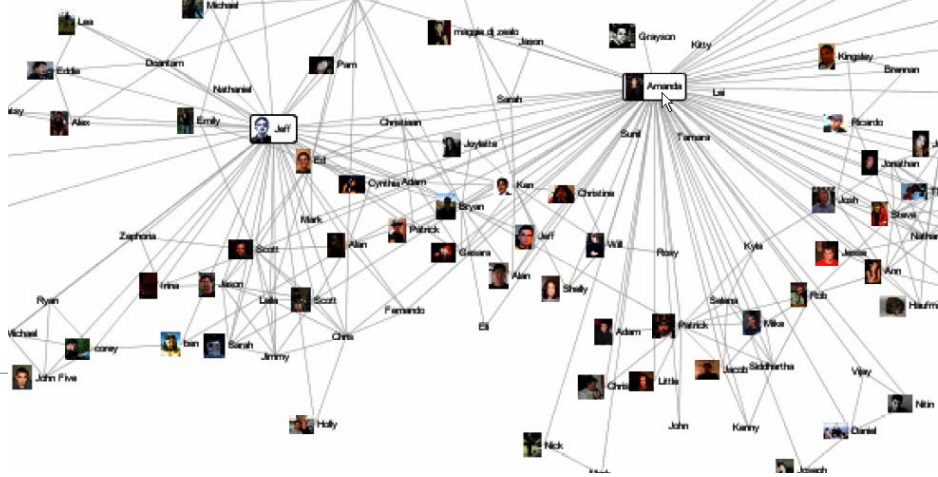
Imagine

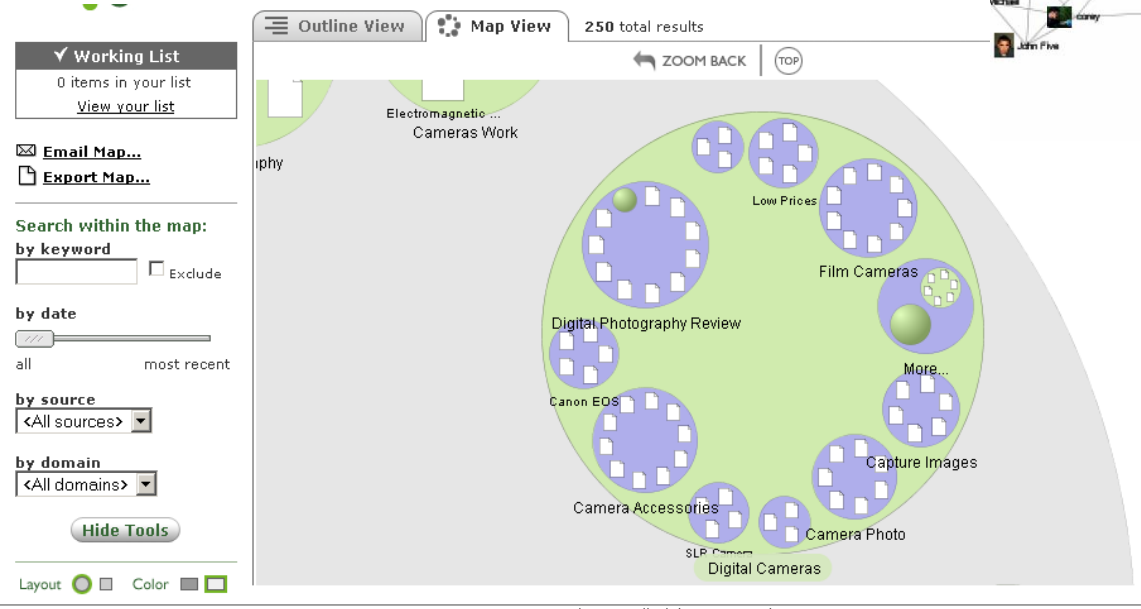
This is an incredible learning resource.

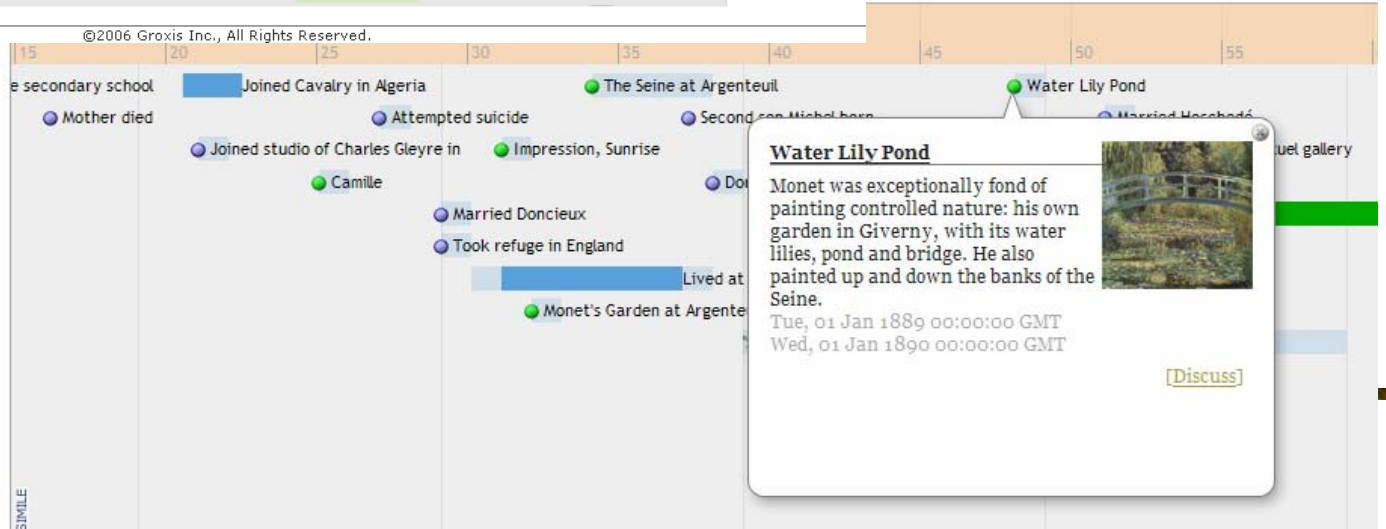
Why isn't there one of these for every member of the scientific community?



Have knowledge, can fly.



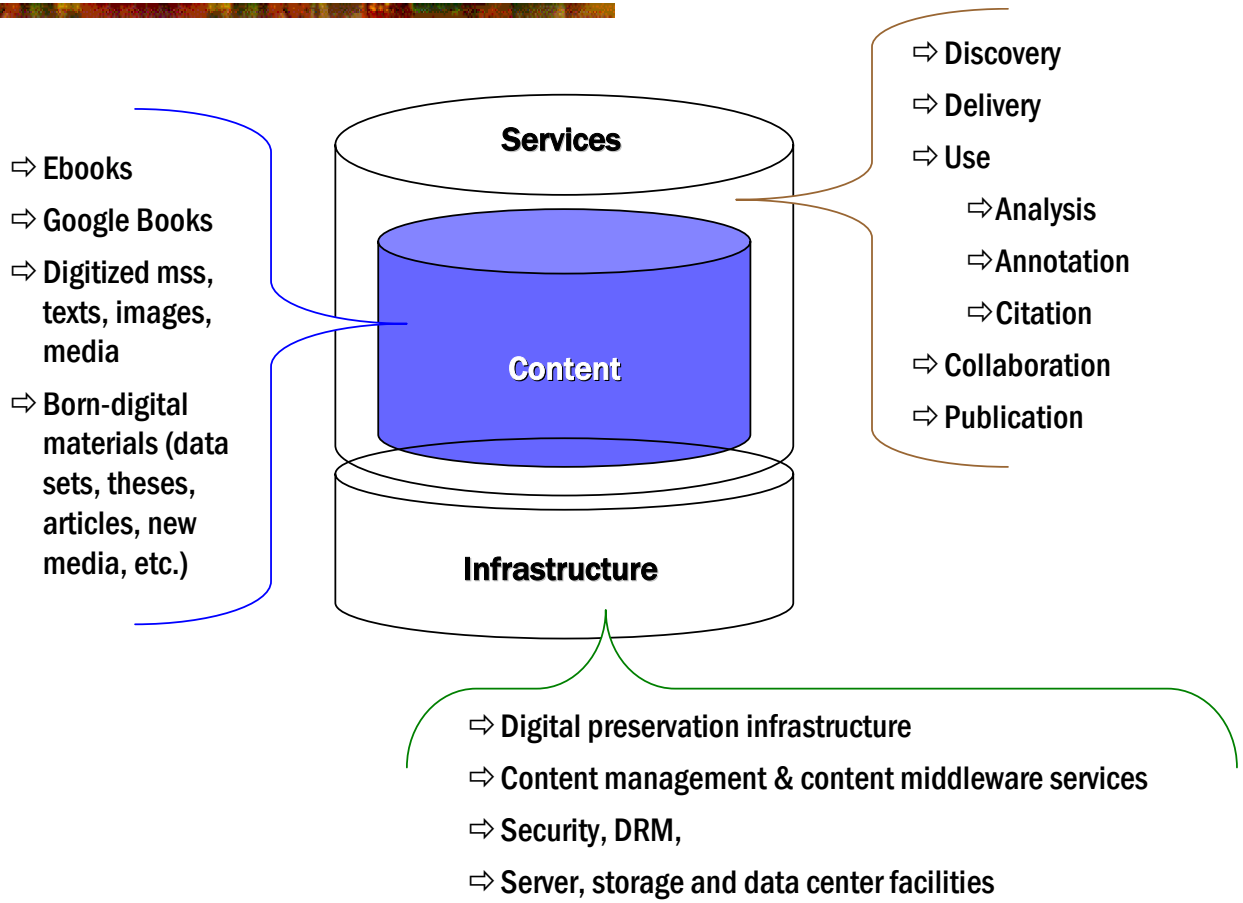




Whose work? Just look around.

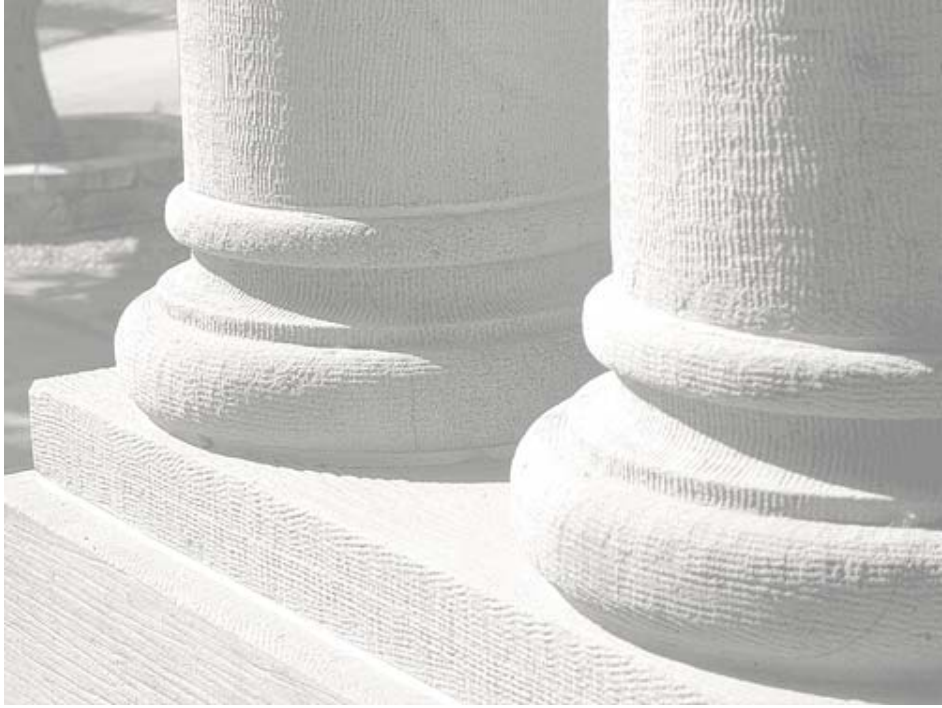
- CS
  - Ed Feigenbaum, John McCarthy, Don Knuth
- Chemistry
  - Carl Djerassi
- EE
  - Andrew Grove, John Hennessy
- Local and Historical Context
  - Stuart Brand, John Warnock, David Kennedy, Jim Adams, Roger Kornberg …

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

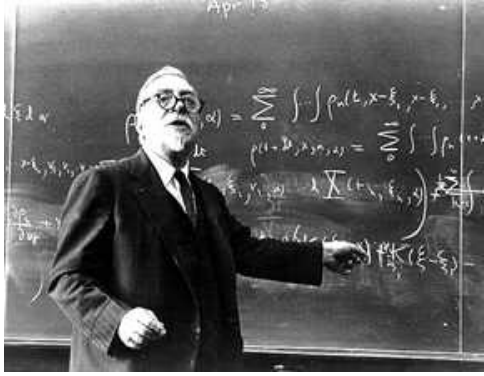
or, why this is an opportunity to make a difference in the history of science and technology

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

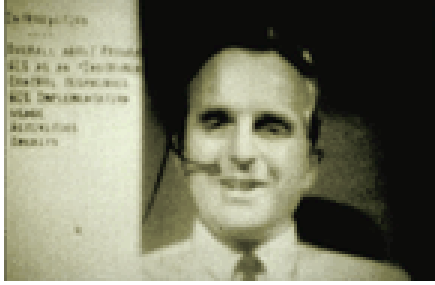


- Goal: improve human performance
- Technology:
  - *control systems with humans in the loop*

- *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

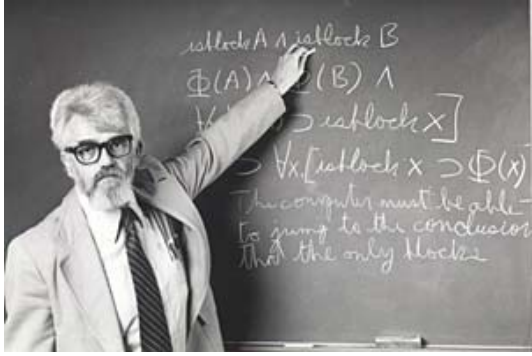
Augmentation (1960-70s) Doug Engelbart



- 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 easer 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*

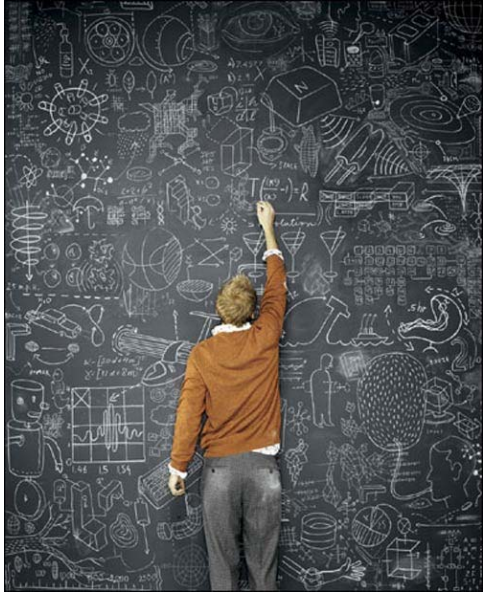
- *Insight:*
  - *Automation of intelligence is limited by the bottleneck of acquiring **human knowledge**.*

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



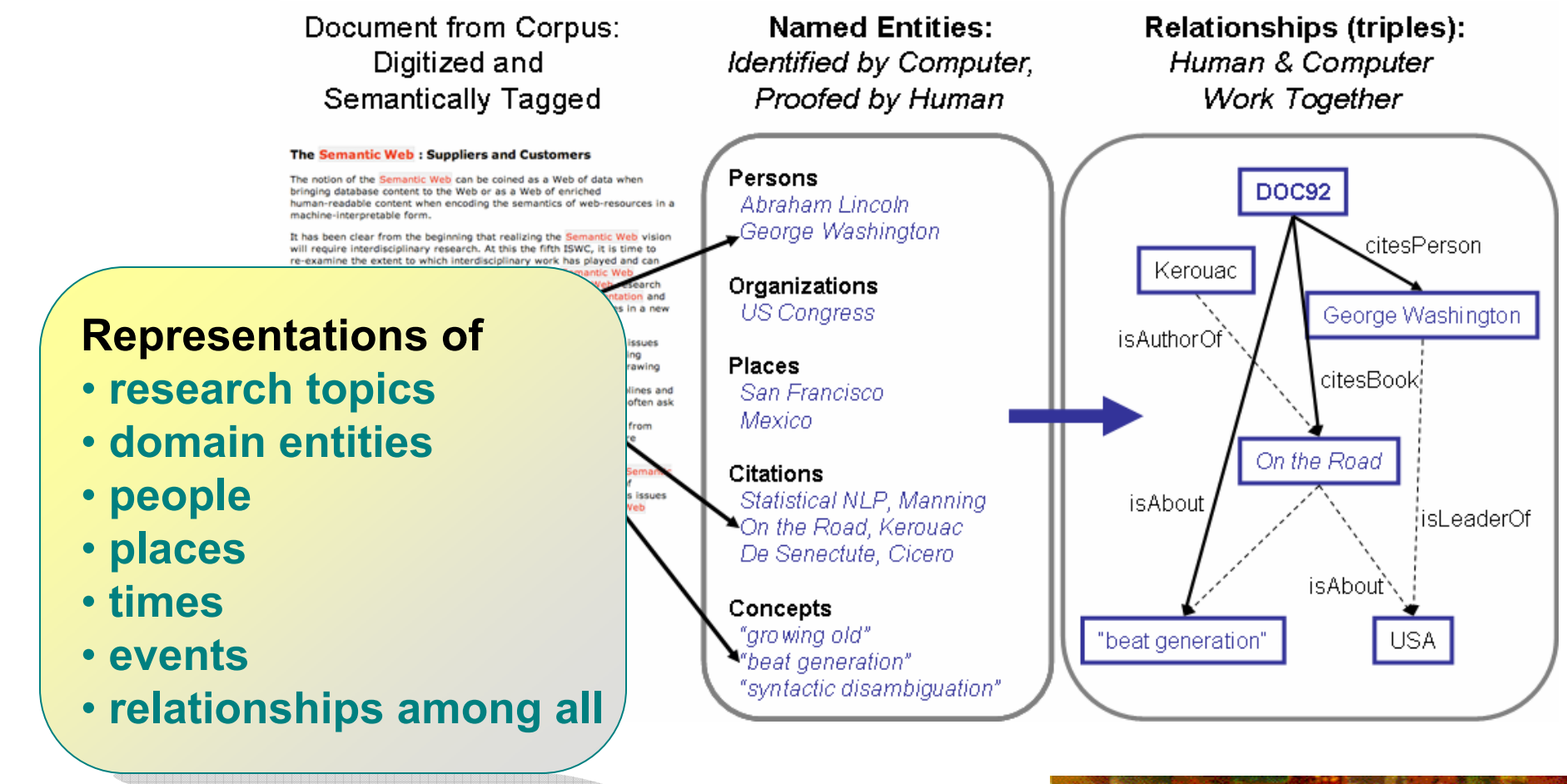
- How to get the *best* of our collective knowledge?
  - *What is best done by machines?*
  - *What is best done by humans?*
  - *What is the best application of **both**?*

Image: New York Times Photograph from the Year in Ideas by Zachary Scott December 12, 2004.

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



Representations of

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

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

doodle from http://accuracyandaesthetics.com

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