What Deans of Informatics Should Tell Their University Presidents

Robert L. Constable
Dean of the Faculty of

Computing & Information Science Cornell University

euroTICS 2006
ETH, Zurich
Monday October 16, 2006

Introduction

This talk is about how my senior colleagues and I create support for computing and information science (informatics) among US policy makers, university presidents, legislators, and funding agency heads.

Strategies and Facts

There are various strategies for getting the attention of policy makers and motivating them:

- Excitement of *opportunity*
- Fear of *failure*
- Responsibility to *save the world*
- Pride of the *accomplishment*

All strategies depend on basic *facts* about our field.

Basic Facts about Informatics

(1) The ideas, methods, and discoveries of informatics (Computing and Information Science) are changing the way we work, learn, discover, play, heal, manage resources,…

Computers and digital information change whatever they touch.

Basic Facts about Informatics

(2) The impact of (1) on the academy will be large, it changes how we *create*, *preserve* and *disseminate knowledge* – that's university business, and it increases demand for CIS in other academic disciplines.

(3) The western economies are *knowledge based* and increasingly dependent on the IT sector.

5

How we are responding in North America?

- Creating colleges (faculties) of ***CIS***
- Broadening the ***I-Schools***
- Broadening the Computing Research Association (***CRA***), now includes IT Deans Group
- Changing Computer Science education
- Exploring **cross-cutting academic structures** – multidisciplinary and interdisciplinary

6

Plan of the Talk

1. Illustrate the Facts by Examples
2. Examine College Structures
3. Historical Perspective and Conclusion

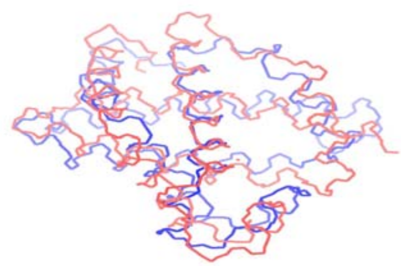
Relevance of CIS ideas and methods by examples

- Computational Biology
- Astronomy
- Digital Age Mathematics

Structures Are Evolutionary Templates

High degree of structural similarity is often observed in proteins with diverse sequences and in different species (below noise level – 15 percent sequence identity).

Oxygen Transport Proteins



Leghemoglobin in Plants

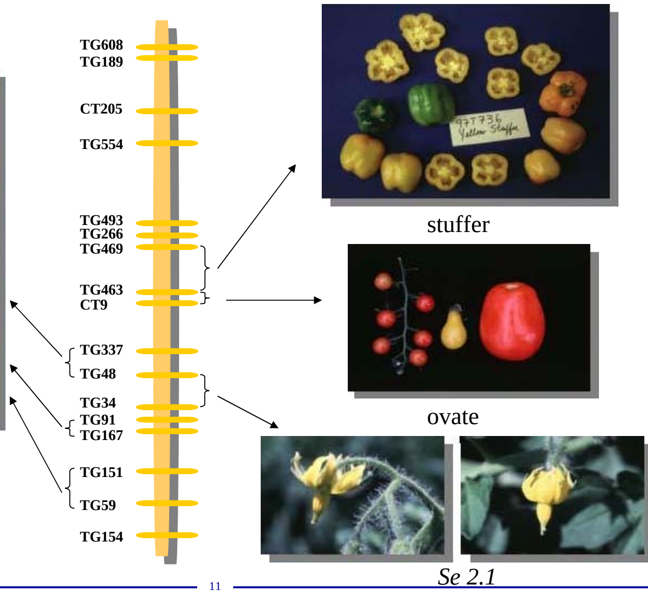
Yet Bigger Tomatoes…

10

Elber/Tanksley Discovery

Chromosome 2

fw 2.1, 2.2, 2.3

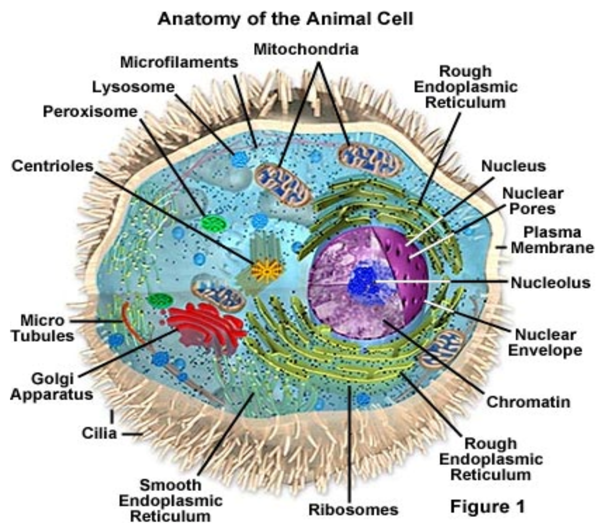


Elber/Tanksley Discovery - continued

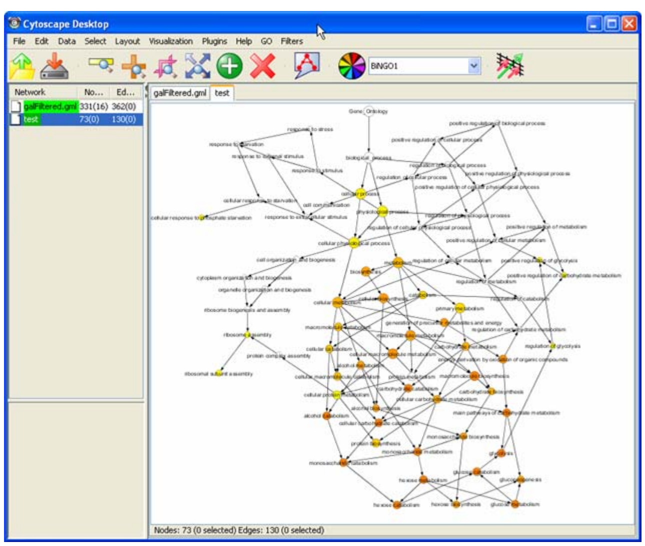
Human Ras p21

- Molecular switch based on GTP hydrolysis
- Cellular growth control and cancer
- Ras oncogene: single point mutations at positions Gly12 or Gly61

simulate complex systems as cells



biological metabolic network



But so far success was very limited…

National Virtual Observatory

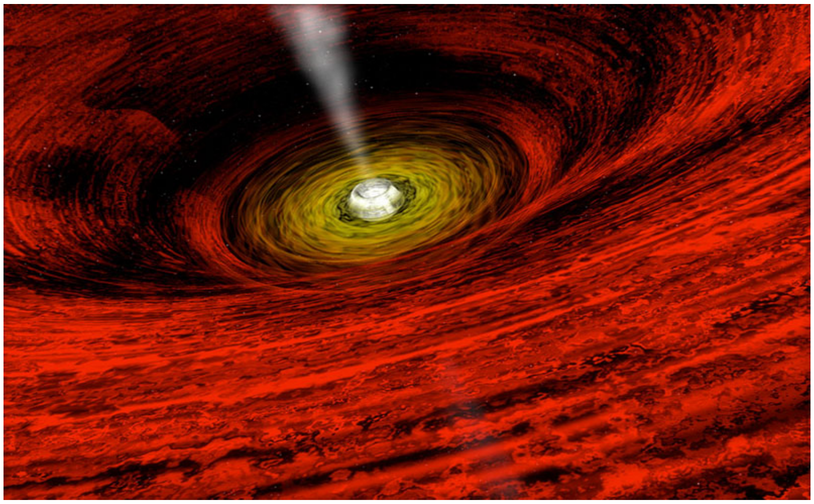
The PC is a telescope for viewing “digital stars”.

Changing the face of astronomy

The astronomer Alexander Szalay has said the work of computer scientists on the NVO has “changed astronomy as we know it”.

Machine learning applied to large databases has led to new discoveries, e.g. new exotic sources,
identification of unidentified sources.

And we can even extrapolate to more complex exotic systems



17

Consider these famous problems…

- The Poincare Conjecture
- The Four Color Theorem
- The Kepler Conjecture

Computers and the Web have fundamentally changed how they are being solved.

18

Digital Age Mathematics – The Poincaré Conjecture

In 2006 the International Mathematical Union (IMU) tried to award Gregory Perelman of St. Petersburg its highest honor, the Fields Medal, for solving the Poincaré Conjecture, one of the seven Millenium problems. He would not accept.

"If the proof is correct, no other recognition is needed."

Digital Age Mathematics – The Poincaré Conjecture Continued

On November 11, 2002 Perelman posted a proof of the Poincaré Conjecture on the Cornell arXiv, Paul Ginsparg's digital library of "e-prints." This posting stimulated the math community to "fill in the details."

(Paul Ginsparg is an Information Science professor in CIS, and Perelman's proof builds on the work of William Thurston, a Field's Medalist who has a joint CIS appointment with Math.)

The Poincaré Conjecture - Controversy

Fields Medalist Shing-Tung Yau said in 2006,

"In Perelman's work, spectacular as it is, many key ideas of the proofs are sketched or outlined, and complete details are often missing."

The idea of a proof is central to modern mathematics. They have strict forms, like a sonnet. It can now be measured against a new standard, the complete formal proof – an idea from Hilbert made precise and implemented by computer scientists.

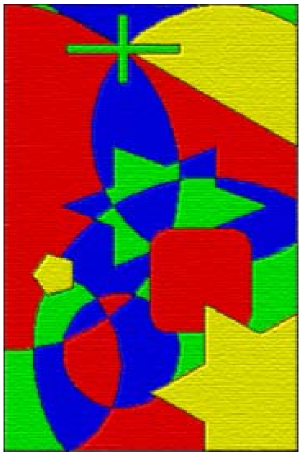
Digital Age Mathematics

One of the most profound contributions of computer science to intellectual history is the demonstration that computers can implement many high level mental functions.

(The converse is also profound, the discovery that our mundane mental functions are extremely difficult to automate.)

The Four Color Theorem 1976

In 1976 computers helped Appel and Haken prove the 1852 four color conjecture – that any planar map can be colored using four colors so that no two adjacent regions have the same color.



Concerns about the Appel/Haken Proof

The programs used to show that the 1,476 reducible maps could be four colored were not proved to be correct, and ran for hundreds of hours.

Formal Proof of the Four Color Theorem

In 2004, Georges Gonthier at MSR used the Coq theorem prover, with help from Benjamin Werner, to give a definitive computer checked proof of the four color theorem.

25

The Nature of Formal Proofs

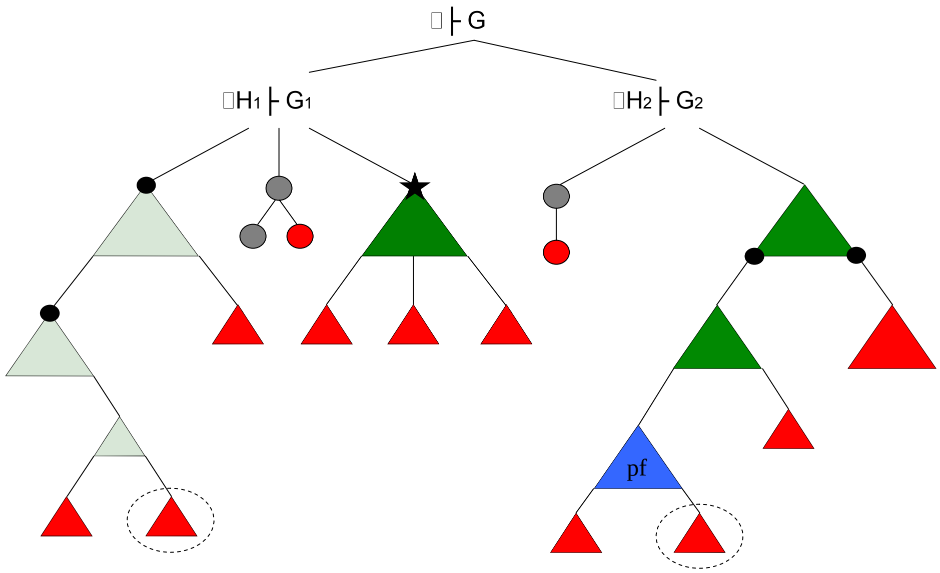
Formal proofs are elements of a tree-like data structure whose nodes are called **sequents**. They have the form

$$H_1, \ldots, H_n \vdash G$$

Where the are propositions called the **hypotheses** and G is the **goal**. H_1

26

A Picture of Proof Structure



Analyzing Proof Structure

- key insights ●
- clever step ★
- filled in by machine
  - humans ignore
  - humans need
- experts need ★
- routine
- learners need
- obvious
- trivial
- well known
- minor variant of pf

The Kepler Conjecture

In 1998 Thomas Hales used computers to "solve" Kepler's conjecture from 1611.

The most dense packing of spheres is as grocers do it.

Since the proof could not be confirmed by the usual social process, Hales turned to computer science and formal proof (using HOL-Light). He relied on a major discovery from CS.

Other Examples

- Social Sciences

  There are laws of social networks, e.g., six degrees of separation

- Humanities

  Assembling the map of the city of Rome, circa 210 A.D.

- Business

  The World is Flat by T. Friedman

Summary – Lessons from Examples

- In all of these examples, computer science is an essential partner in the research. In the case of the life sciences, the assembly of the human genome was a 50/50 effort.

  Computer science ideas, methods, and discoveries are essential.

- All of these examples have opened exciting new areas of research in informatics.

- These interactions with CIS are accelerating

31

Plan of the Talk

1. Illustrate the Facts by Examples
2. **Examine College Structures**
3. Historical Perspective and Conclusion

32

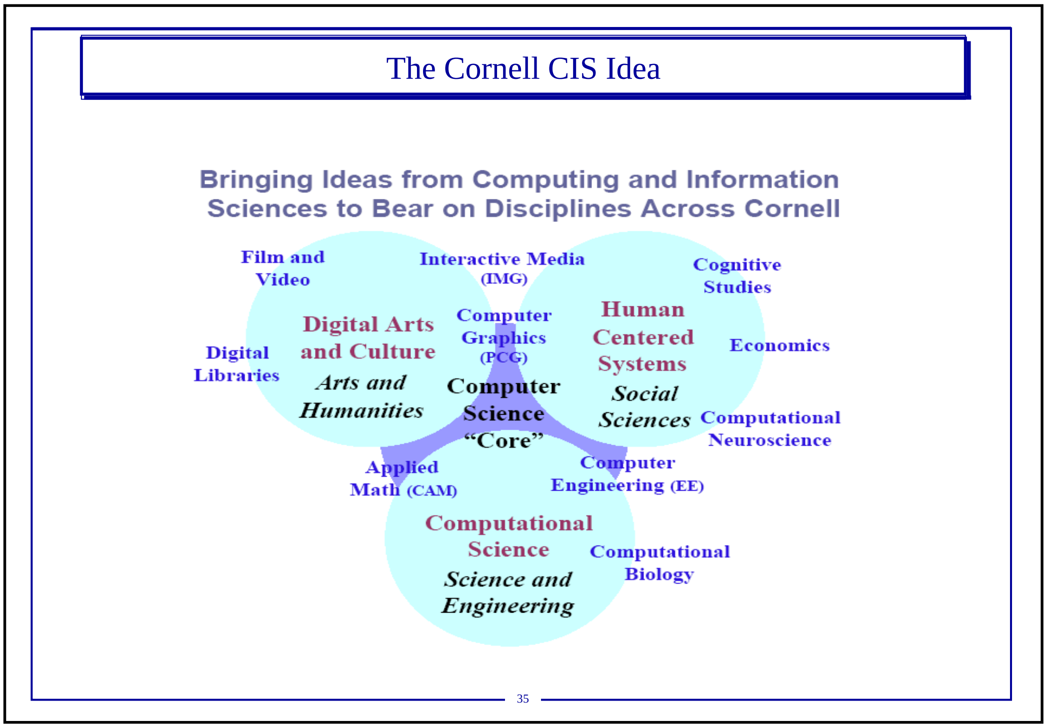
College Structures

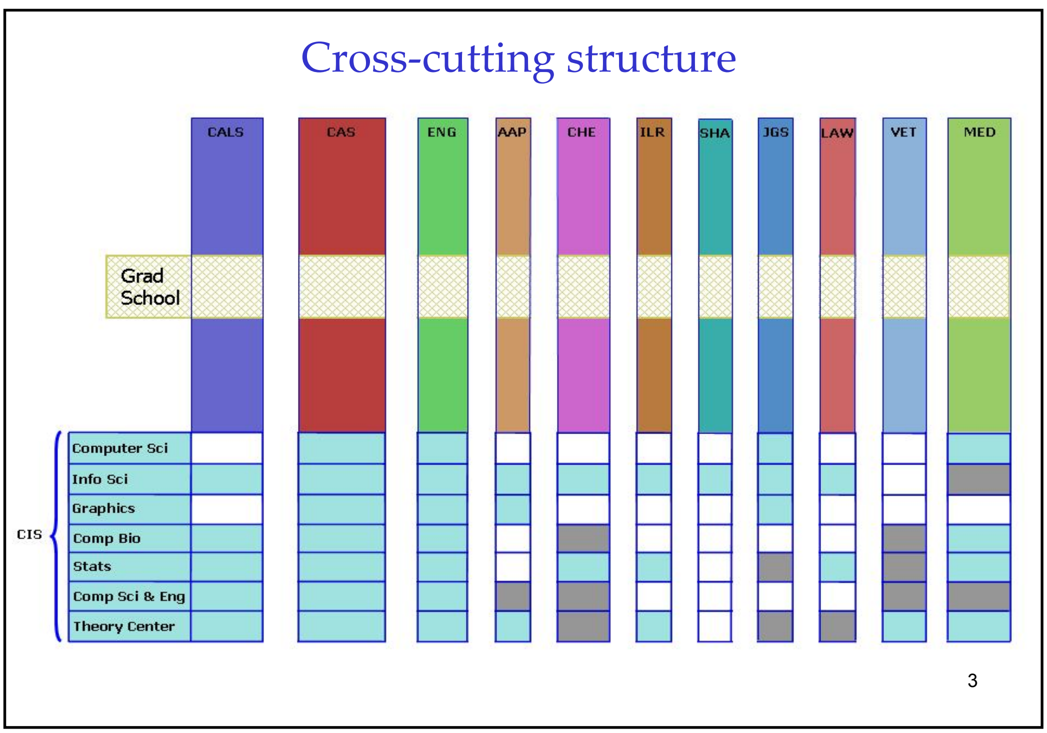
There are over a dozen colleges of CIS and over 20 I-schools in North America. Here are three of the top CIS colleges:

- Cornell
- CMU
- Georgia Tech

Comparing Colleges

Cornell	**CMU**	**Georgia**
- Computer Science	- Computer Science	- Computing Science & Systems
- Statistics	- Machine Learning	
- Information Science	- HCI Institute Language Technologies	- Interactive & Intelligent Computing
- Computational Biology	- Robotics	
- Computational Science & Engineering	- Software Research	- CSE
- (Digital Arts)	- Entertainment Technologies	


The Cornell CIS Idea
Bringing Ideas from Computing and Information Sciences to Bear on Disciplines Across Cornell
Film and Video
Interactive Media (IMG)
Cognitive Studies
Digital Libraries
Digital Arts and Culture
Arts and Humanities
Computer Graphics (PCG)
Computer Science "Core"
Human Centered Systems
Social Sciences
Economics
Computational Neuroscience
Applied Math (CAM)
Computer Engineering (EE)
Computational Science
Science and Engineering
Computational Biology
35


Cross-cutting structure
CALS
CAS
ENG
AAP
CHE
ILR
SHA
JGS
LAW
VET
MED
Grad School
CIS
Computer Sci
Info Sci
Graphics
Comp Bio
Stats
Comp Sci & Eng
Theory Center
3

Impact of CIS

CIS will impact every discipline because it goes to the core of what they do. Perhaps 5% to 7% of the faculty in most disciplines will want to be connected as well to a center of CIS research and teaching.

5% to 7% of faculty
(at Cornell 80 to 110)

37

Impact of CIS – continued

Students will be increasingly computer savy and demand to know how computational thinking applies to their interests.

The economy will need more knowledge workers.

Taken together, this means having many more faculty trained in CIS and connected to it academically as well as intellectually.

38

Plan of the Talk

1. Illustrate the Facts by Examples

2. Examine College Structures

3. Historical Perspective and Conclusion

Historical perspective

The Industrial Revolution (IR1) is about: extending muscle power (mass, energy, force, power, space, and time)

The Information Revolution (IR2) is about: extending brains (information, intelligent processes, computation, complexity, and networks)

IR1 created colleges of engineering, shaping the physical sciences.

IR2 is creating colleges of computing, shaping the information sciences.

Conclusion: Only the Beginning

We are in early stages of the Information Revolution.

Combining digital information with digital computation is an explosive mix. We will see the birth of machines that know and reason, that are continuously interactive and autonomous.

It will be more clear that CIS is about modeling information processes and automating intellectual processes.

THE

END

Changing University Research

Partnership with CIS are key, unit must be large enough to help many interested departments

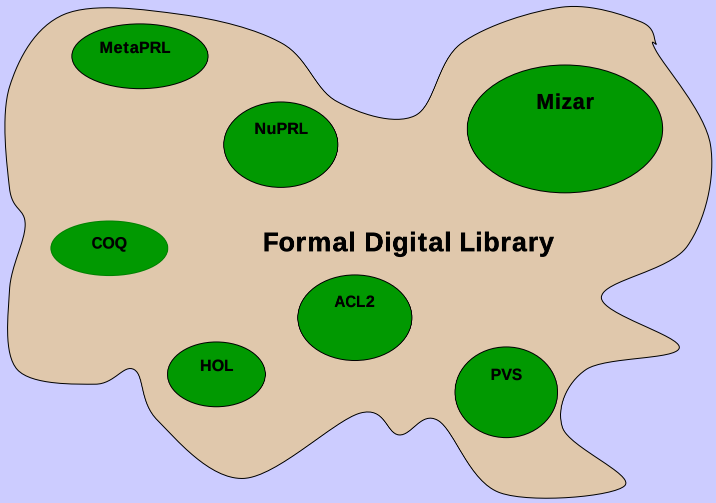
5% to 7% of faculty (at Cornell 80 to 110)

Not for applications, for joint discovery

General Goal

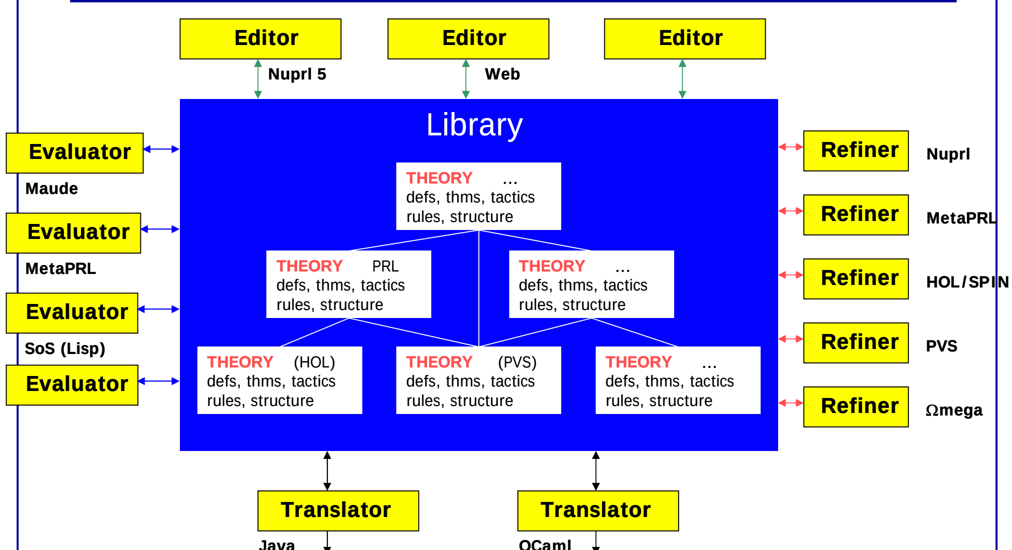
Investigate the scientific and social potential of an unusual information resource – approximately 50,000 formal theorems and proofs created by interactive theorem provers

This resource represents a frontier of a century and a half effort to perfect the notion of a mathematical proof



The FDL joins all the small islands into a single big island that can accommodate much more than the sum of all individual islands, and allows people to stand on the ground in between.

Formal Digital Library



What does Coq know?

What does Coq know about □ ?

- calculation
- symbolic calculation
- induction
- 5K basic facts

What if Coq knew 50K basic facts? 500K? 5M?

What does it mean to know Mathematics?

The logicians best answer:

to know is to prove

Machines rely on complete formal proofs.

Humans rely on intuitive proofs.

How do intuitive proofs relate to formal proofs?

How Provers Are Used?

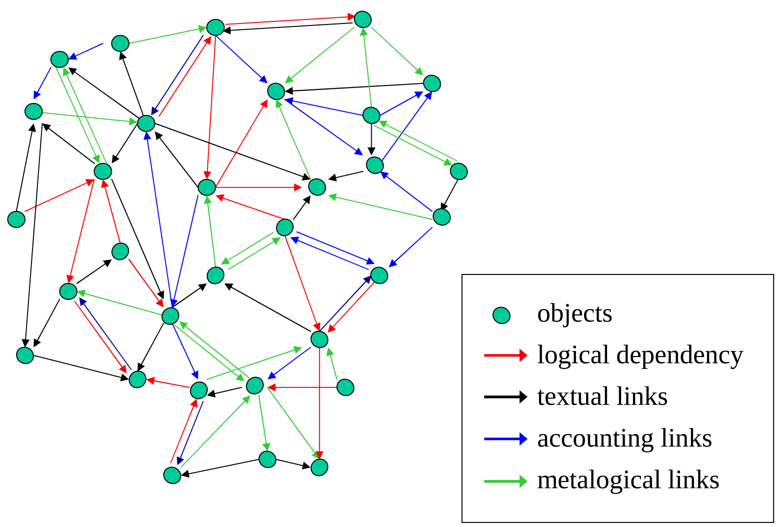
American provers must follow economic needs

In Europe they can create formal mathematics

- The Fundamental Theorem of Algebra
- The Prime Number Theorem
- Grobner Basis Theorem
- The Four Color Theorem

In the US we verify protocols and algorithms.

Information Graph of the FDL



Conclusion: Only the Beginning

We are in early stages of the Information Revolution.

Combining digital information with digital computation is an explosive mix. We will see the birth of machines that know and reason, that are continuously interactive and autonomous.

Conclusion: Impact of CIS

What is the impact of CIS on universities?

Is the impact of the Information Revolution greater than that of Industrial Revolution?

What was that impact?

- Colleges of Eng were created
  - Physics - AEP
  - Chem - CHE
  - - Civil
  - - Mechanical
- New kinds of students, larger demand for math, physics (among the largest departments)
- Closer ties to industry