Transforming the Academy: Knowledge Formation in the Age of Digital Information

Robert L. Constable
Computing & Information Science

Cornell University

Northeastern University
College of Computer & Information Science

February 9, 2006

Introduction

Your College and our Faculty are key to the future of our universities

Why is CIS so critical?

In the age of Digital Information, the core business of universities is changing

- discovery
- knowledge dissemination
- knowledge preservation

This talk will examine these changes and their implications for the academy.

Outline

- Nature of the changes in
  - Discovery
  - Dissemination
  - Preservation
- Role of computer science - automation
- Role of information science - human access to digital information

Conclusion

Discovery in the Digital Age – three examples

- Genomics
- The National Virtual Observatory (NVO)
- Computational Science

Yet Bigger Tomatoes…

Elber/Tanksley Discovery

Chromosome 2

fw 2.1, 2.2, 2.3

TG608
TG189
CT205
TG554
TG493
TG266
TG469
TG463
CT9
TG337
TG48
TG34
TG91
TG167
TG151
TG59
TG154

stuffer

ovate

Se 2.1

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

National Virtual Observatory

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

Other Examples

- Life Sciences

  As in the tomato growth gene and cancer

- 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

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

Dissemination of Digital Information

The Web

- CS papers
- Dspace @MIT
- Wikipedia

Google

- Translation
- Images
- “Awareness”

The Cornell arXiv

- Physics
- Math
- CS

Changing Nature of Instruction

Professors are not primarily information providers but rather

- guides
- interpreters
- critics

They seek to impart knowledge based on sifting and winnowing information, by teaching how to evaluate it, and organize it.

Preserving Information

Example of the arXiv

New Dark Ages possible?

Extinction of digital machines?

Changing Nature of Scholarship

In all fields, we can link the primary data sources to publications.

This will revolutionize historical subjects.

Outline

- Nature of the changes in
  - Discovery
  - Dissemination
  - Preservation
- Role of computer science - automation
- Role of information science - human access to digital information

Conclusion

Role of Computing

automating calculation

- data structures (numerical, symbolic)
- algorithms

formalizing proof

- advanced data structures
- specialized algorithms

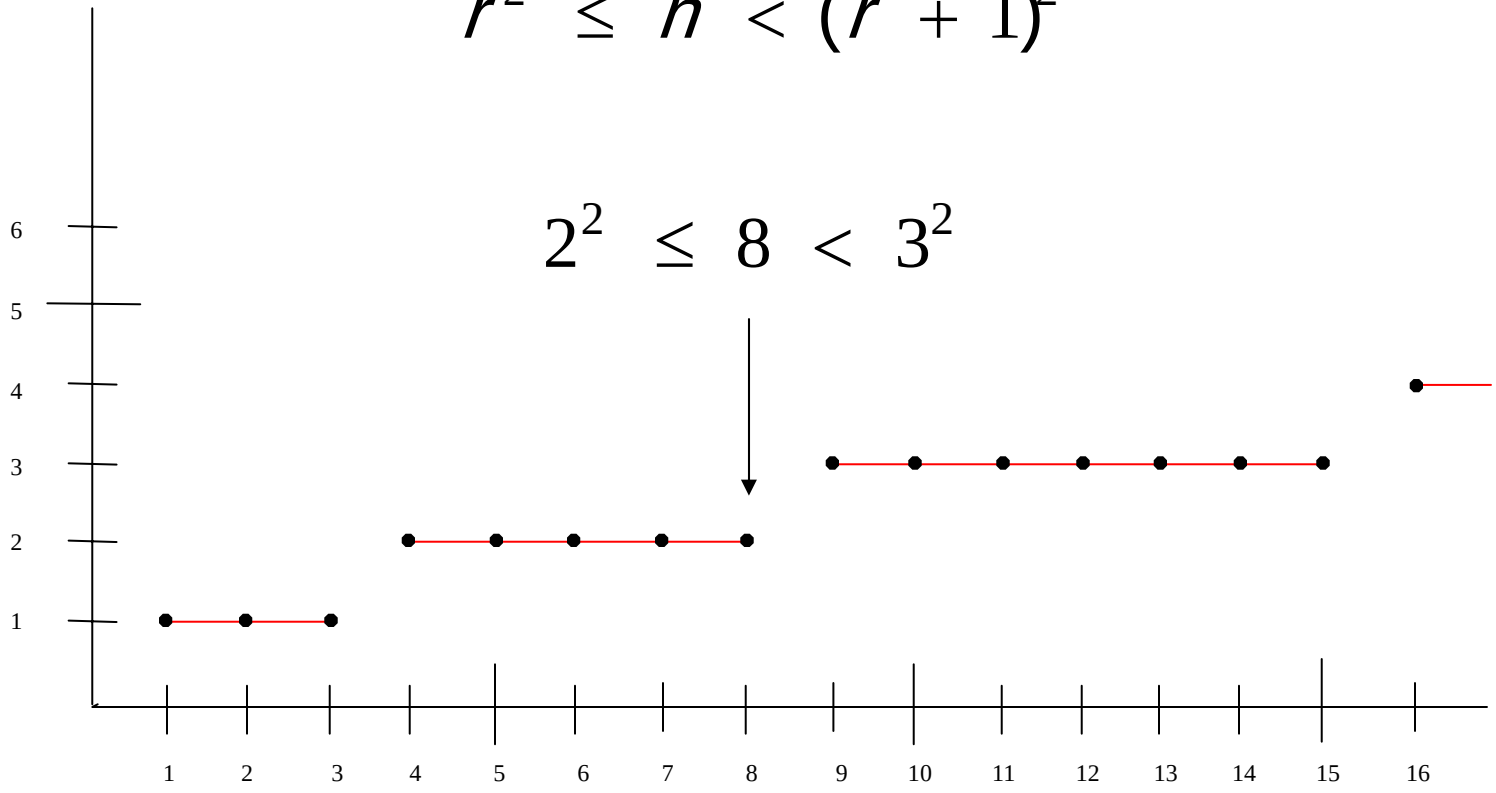
automating reasoning

- decision procedures
- tactics
- heuristics

Integer Square Root

$$r^2 \leq n < (r+1)^2$$

$$2^2 \leq 8 < 3^2$$



Proof of Root Theorem

$\forall n : \mathbb{N}.\ \exists r : \mathbb{N}.\ r^2 \le n < (r+1)^2$

BY `allR`

$n : \mathbb{N} \vdash \exists r : \mathbb{N}.\ r^2 \le n < (r+1)^2$

BY NatInd 1

. . . . .base case.....

$\vdash \exists r : \mathbb{N}.\ r^2 \le 0 < (r+1)^2$

BY `existsR` ⌈0⌉ THEN `Auto`

. . . . .induction case .....

$i : \mathbb{N}^+,\ r : \mathbb{N},\ r^2 \le i - 1 < (r+1)^2$

$\vdash \exists r : \mathbb{N}.\ r^2 \le i < (r+1)^2$

BY `Decide` $\lceil (r+1)^2 \le i \rceil$ THEN `Auto`

Proof of Root Theorem (continued)

. . . . .Case 1.....

$$i : \mathbb{N}^+,\ r : \mathbb{N},\ r^2 \le i - 1 < (r+1)^2,\ (r+1)^2 \le i$$

$$\vdash \exists r : \mathbb{N}.\ r^2 \le i < (r+1)^2$$

BY `existsR` $\lceil \underline{r} + 1 \rceil$ THEN `Auto'`

. . . . .Case 2.....

$$i : \mathbb{N}^+,\ r : \mathbb{N},\ r^2 \le i - 1 < (r+1)^2,\ \neg\left((r+1)^2 \le i\right)$$

$$\vdash \exists r : \mathbb{N}.\ r^2 \le i < (r+1)^2$$

BY `existsR` $\lceil \underline{r} \rceil$ THEN `Auto`

Nuprl's Automation (Auto)

Consider what Auto' can figure out

It knows $(r + 1)^2 \leq i$ and

$(i\text{-}1) < (r + 1)^2$

Thus $i < (r + 1)^2 + 1$

Need to know $i < ((r + 1) + 1)^2$

Is $(r + 1)^2 + 1 \leq (r + 2)^2 = r^2 + 4r + 4$?

$r^2 + 2r + 2 < r^2 + 4r + 4$

$2(r + 1) < 4(r + 1)$

yes

What can provers figure out?

- Restricted arithmetic, but not all quantifier free arith.

- Peano arithmetic

- Equality reasoning (congruence closure)

- Euclidean geometry

- Fragments of algebra etc.

- Nuprl also uses fully automatic JProver

Provers can transform proofs

- Replay on altered goals
- Convert some classical proofs to constructive
- Track dependency information
- Translate to natural language
- Extract implicit computational content

The Root Program Extract

Here is the extract term for this proof in ML notation with proof terms (pf) included:

```
let rec sqrt i =
    if i = 0 then < 0, pf0 >
        else let  < r, pfi-1 >= sqrt(i - 1)
        in if  (r + 1)^2 ≤ n then < r + 1, pfi >
            else  < r, pfi' >
```

How does extraction work?

The key insight lies in changing the meaning of

$A \Rightarrow B$ and $A \vee B$.

Normally given by truth tables

A	B	A&B	AvB	A⇒ B
T	T	T	T	T
T	F	F	T	F
F	T	F	T	T
F	F	F	F	T

to assert $A \Rightarrow B$ means to provide an effective method of taking evidence for A into evidence for B.

Outline

- Nature of the changes in
  - Discovery
  - Dissemination
  - Preservation
- Role of computer science - automation
- Role of information science - human access to digital information

Conclusion

Role of digital information

building blocks of formal knowledge structures

proofs as knowledge structures

reasoning as knowing

What does Nuprl know?

Consider the square root proof.

What does Nuprl know about $\mathbb{N}$?

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

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

What does it mean to know Mathematics?

The logicians best answer:

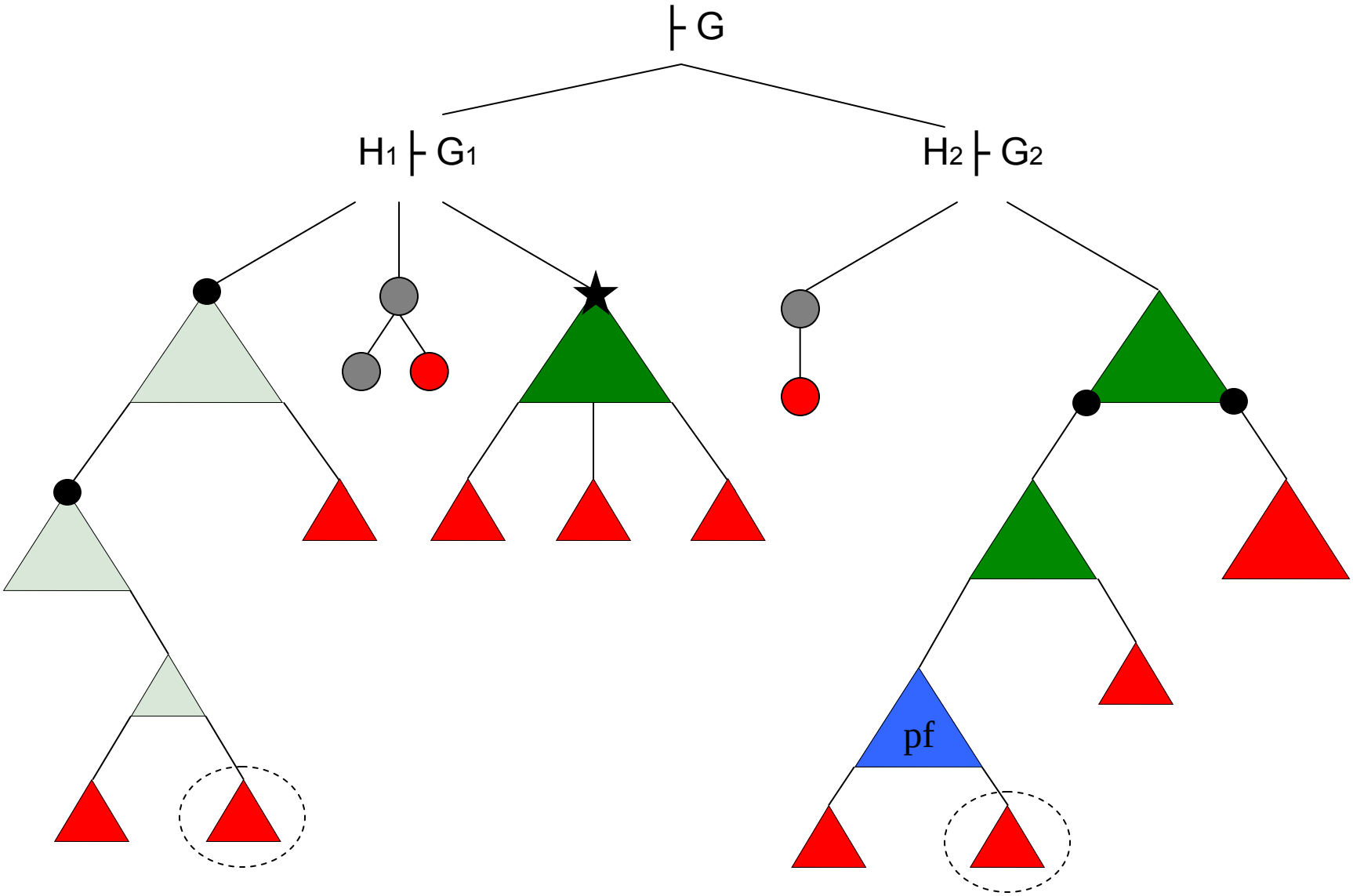
to know is to prove

Machines rely on complete formal proofs.

Humans rely on intuitive proofs.

How do intuitive proofs relate to formal proofs?

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

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.

Extension to distributed protocols

Mutual exclusion protocols – Paxos

Security protocols – APSS

Commitment protocols

Consensus protocols

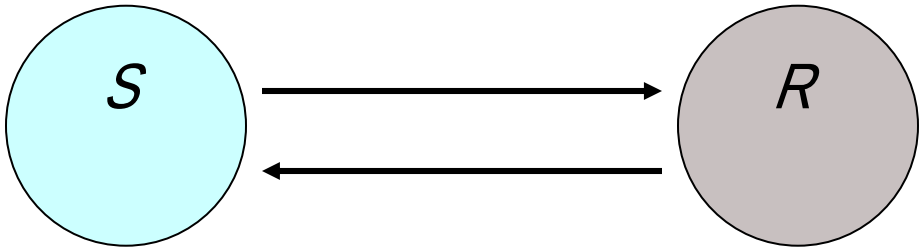
Snap shots

Adaptation

Fault Tolerance

Distributed Task Specification

Here is two phased-handshake in our Logic of Events



$$E_p = \{e : E \mid loc(e) = p\}$$

$$Snd_{p,l} = \{e : E_p \mid sends(e, l) \neq nul\}$$

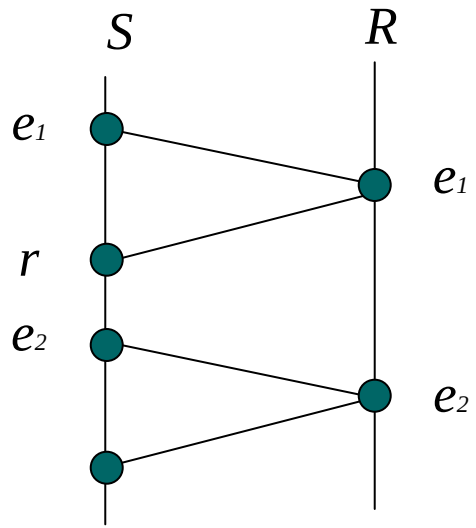
$$Rcv_{p,l} = \{e : E_p \mid kind(e)\ is\ receive\ on\ l\}$$

Deriving the Two-Phase Handshake

We illustrate this process by deriving a protocol for the two-phase handshake from a proof that its specification is realizable.

(1) $\forall e_1, e_2 : \mathsf{Snd}_{S,l_1}.\quad \exists r : \mathsf{Rcv}_{S,l_2}.\quad (e_1 < e_2 \Rightarrow e_1 < r < e_2)$

(2) $\forall e_1, e_2 : \mathsf{Snd}_{R,l_2}.\quad \exists r_1, r_2 : \mathsf{Rcv}_{R,l_1}.\quad (e_1 < e_2 \Rightarrow r_1 \leq e_1 < r_2 \leq e_2)$



Two-Phase Handshake Theorem

Theorem:

$\forall e_1, e_2 : Snd_s . e_1 < e_2 \Rightarrow \exists r : Rcv_s . e_1 < r < e_2$

What are the consequences of $e_1 < e_2$?

S sent two messages

Can we infer further consequences?

Relate send events to knowledge, create a boolean variable *rdy*

- Require *rdy* to be true when a send event occurs
- Require *rdy* to be false after a send event e

Two-Phase Handshake

Theorem:

$\forall e_1, e_2 : \mathrm{Snd}_s . e_1 < e_2 \Rightarrow \exists r : \mathrm{Rcv}_s . e_1 < r < e_2$

How to establish this by reasoning?

Suppose $\mathbf{e}_1 < \mathbf{e}_2$ are sends on link ℓ from S,

Then by $\mathbf{L}_1$ (**rdy** after $\mathbf{e}_1$) = false.

Since $\mathbf{e}_2$ is a send, **rdy** must be true at $\mathbf{e}_2$ by $\mathbf{L}_2$.

Therefore some event $\mathbf{e}'$ before $\mathbf{e}_2$ and after $\mathbf{e}_1$

must set **rdy** to true.

By $\mathbf{L}_3$ the event $\mathbf{e}'$ must be received from **R** on link ℓ'.

Let **r** be this $\mathbf{e}'$.

If we have the lemmas, then the theorem is true.

L1 & L2 & L3 ⇒ Theorem

Realizable Lemmas

L1. If S sends on link ℓ then it waits

$\forall$e: $\mathsf{Snd}_{s,\ell}$(rdy after e) = false realized by **R_1**

L2. S sends on link ℓ only when rdy = true

$\forall$e: $\mathsf{Snd}_{s,\ell}$(rdy when e) = true realized by **R_2**

L3. After S on link ℓ sends it is ready only after an acknowledgement on link ℓ'

$\forall$e: $\mathsf{Snd}_{s,\ell}$ e changes rdy to true $\Rightarrow$

e is a receive from R on link ℓ' realized by **R_3**.

If the lemmas are realizable, then so is the theorem,

by **R_1 + R_2 + R_3**

Handshake Message Automation

action local (a) sends on ℓ_1 < **tag, v** >

only [a] sends on ℓ_1

state **rdy** : $\mathbb{B}$; (**initially** **rdy** = true)

precondition **a** is **rdy** = true

effect **local** (**a**) **rdy** := false

action **rcv**$_{\ell_2}$ < **ack** > : **Atom**

effect **rdy** := true

only [**local** (**a**), **rcv**$_{\ell_2}$ <**ack**>] affect **rdy**

This message automaton realizes the theorem.

Handshake Message Automation

action local (a) sends on ℓ_1 < **tag, v** >
only [a] sends on ℓ_1
state **rdy** : $\mathbb{B}$; (**initially** **rdy** = true)
precondition **a** is **rdy** = true

R_2

effect **local** (**a**) **rdy** := false

R_1

action **rcv**$_{\ell_2}$ < **ack** > : **Atom**
effect **rdy** := true
only [**local** (**a**), **rcv**$_{\ell_2}$ <**ack**>] affect **rdy**

R_3

This message automaton realizes the theorem.

The Future of Automated Reasoning

larger data bases – from 50K theorems to 500K theorems: many more “I know that” hits in proofs.

mega-tactics – big steps very large subtrees

decision procedures (tree automata, etc.)

transcribing proof sketchs

learning prover behavior

deployment on more critical problems

Provers Will Become Indispensible

Provers already help up solve complex mathematical problems. They help us know since proof is a process of knowing.

See Imre Lakatos Proofs & Refutations

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

Impact of CIS – continued

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.

Students will be increasingly computer savy and demand to know how computing 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.

Impact of CIS – continued

To do this requires:

- a broad CIS corriculum
- enough CIS faculty
- interdisciplinary flexibility

These forces will become increasingly clear as we advance deeper into the Age of Digital Information.

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.