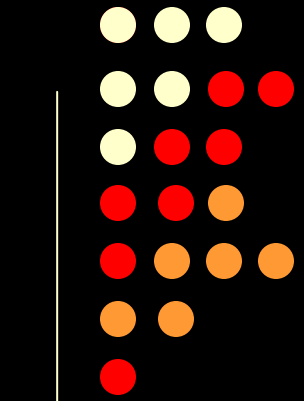
OFFICE OF NAVAL RESEARCH (ONR)

CORNELL UNIVERSITY DEPARTMENT OF COMPUTER SCIENCE

CORNELL

An Interactive Digital Library of Algorithmic Knowledge



and other Information Science projects at Cornell

April 14, 2003 / UC Berkeley

Stuart Allen
Mark Bickford
Robert Constable
Jason Hickey (Cal Tech)
Vitali Khailine (Wyoming)
Richard Eaton
Christoph Kreitz
Lori Lorigo
Jim Caldwell (Wyoming)

Topics

- The Office of Naval Research (ONR) Digital Library Project
  - Goals, Design, Implementation
- Other Information Science Projects
  - National Science Digital Library, NSDL
  - Ranking Function
  - Modeling Bursts of Activity on the Web

ONR Digital Library Project

" … to create a digital library of algorithms and constructive mathematics useable for program and software construction."

CORNELL

ONR Digital Library Project

" … to create a digital library of algorithms and constructive mathematics useable for program and software construction."

Goals

- Semantics-based interactive infrastructure
- Create, collect and organize formal content

CORNELL

Meaning of Formal

"… having a precise meaning or objective criteria of correctness … ideally computer verifiable, based on syntactic form."

Formalization I

Scheduling Application Formalization

Specify components and behavior
Use mathematical definitions and theorems

Reveals assumptions, increases assurance, and optimizes task assignments.

Doing A then B then A is the same as doing A.


Formalization II

- Halting problem, Alan Turing - informal

  Does a given algorithm ever halt when executed on given arguments?

  – no method to solve it for all inputs

Formalization II

- Halting problem, Alan Turing - informal

  Does a given algorithm ever halt when executed on given arguments?

  - no method to solve it for all inputs

- Formalization
  - Create formulas to express halting
  - Mechanically prove no method exists for all inputs

$\neg\exists h{:}\bar{A} \rightarrow \text{Bool}.\ \forall n{:}\bar{A}.\ h(n) = \text{true iff } n \text{ halts}$ ✓

CORNELL

ONR Digital Library Project

" … to create a digital library of algorithms and constructive mathematics useable for program and software construction."

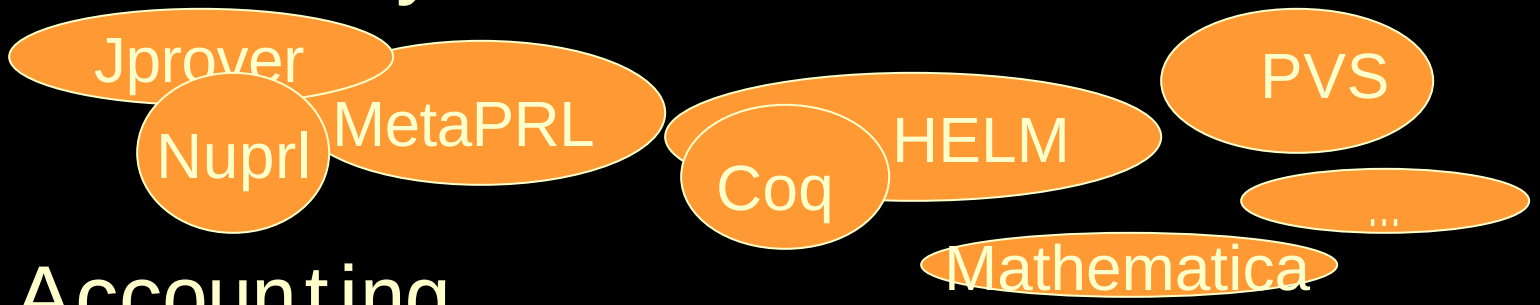
Benefits

- Accelerate discovery
- Access to formal content
- Resource for reliable software and education

CORNELL

Strategy for Meeting ONR Goals

- Community of contributors



- Accounting
  - differences and limitations amongst systems, different meanings

- Semantics-based library services

Challenges and Problems

- *Small* and disconnected community
- Formal proving is still hard work
  - Large expansion in size
  - Shallow resource base
  - Demanding skill set

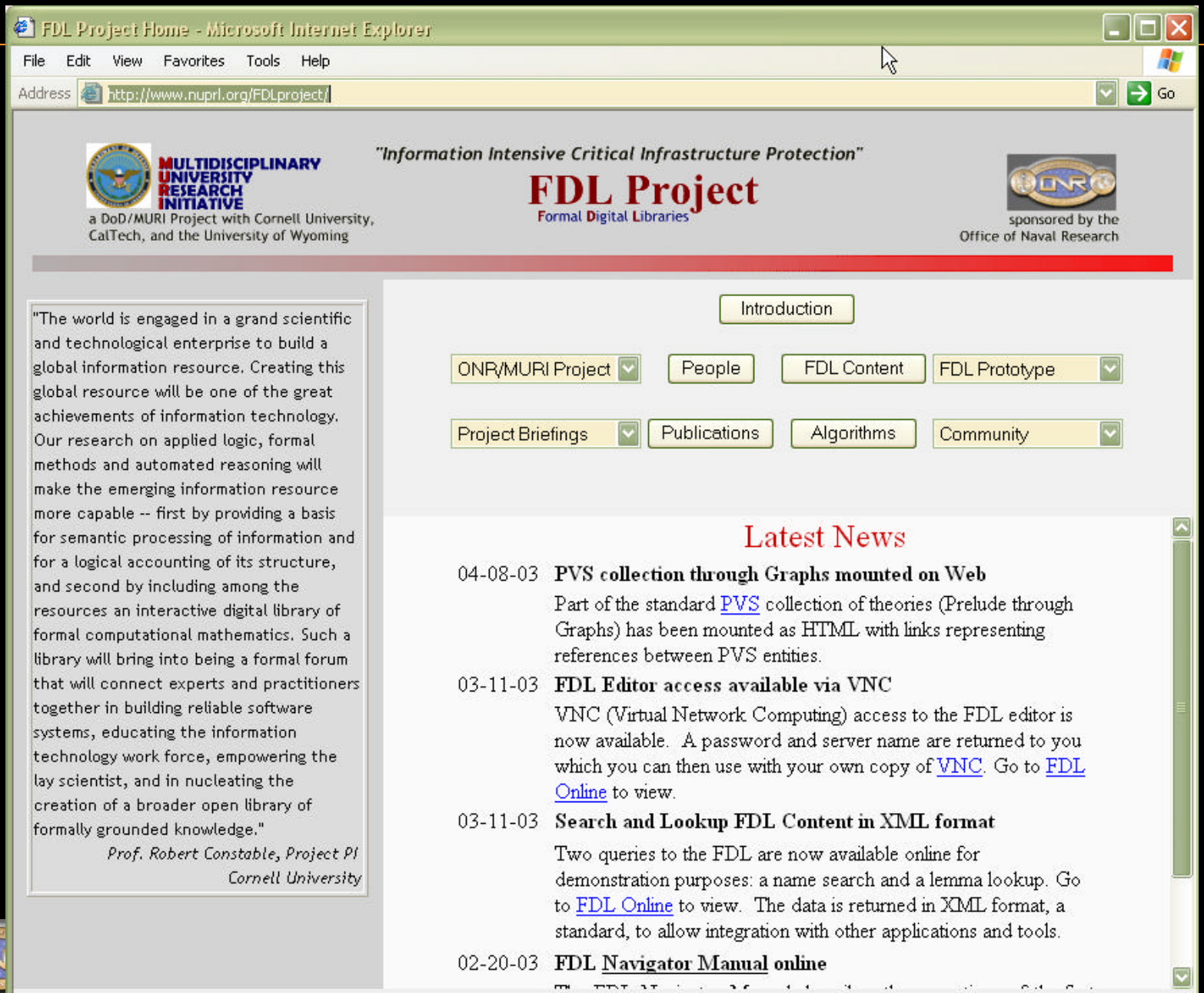
Challenge Citation

"PVS is a large and complex system and it takes a long while to learn to use it effectively. You should be prepared to invest six months to become a moderately skilled user (less if you already know other verification systems, more if you need to learn logic or unlearn Z). "

- PVS website at SRI

CORNELL

Formal Digital Library (FDL)



CORNELL

Objects

FORMAL

rules	definitions	algorithms, code
conjectures	theorems	specifications
inferences	certificates	proofs, partial proofs

CORNELL

Objects

FORMAL	rules	definitions	algorithms, code
	conjectures	theorems	specifications
	inferences	certificates	proofs, partial proofs
INFORMAL	displays	articles	documentation

CORNELL

Objects

FORMAL

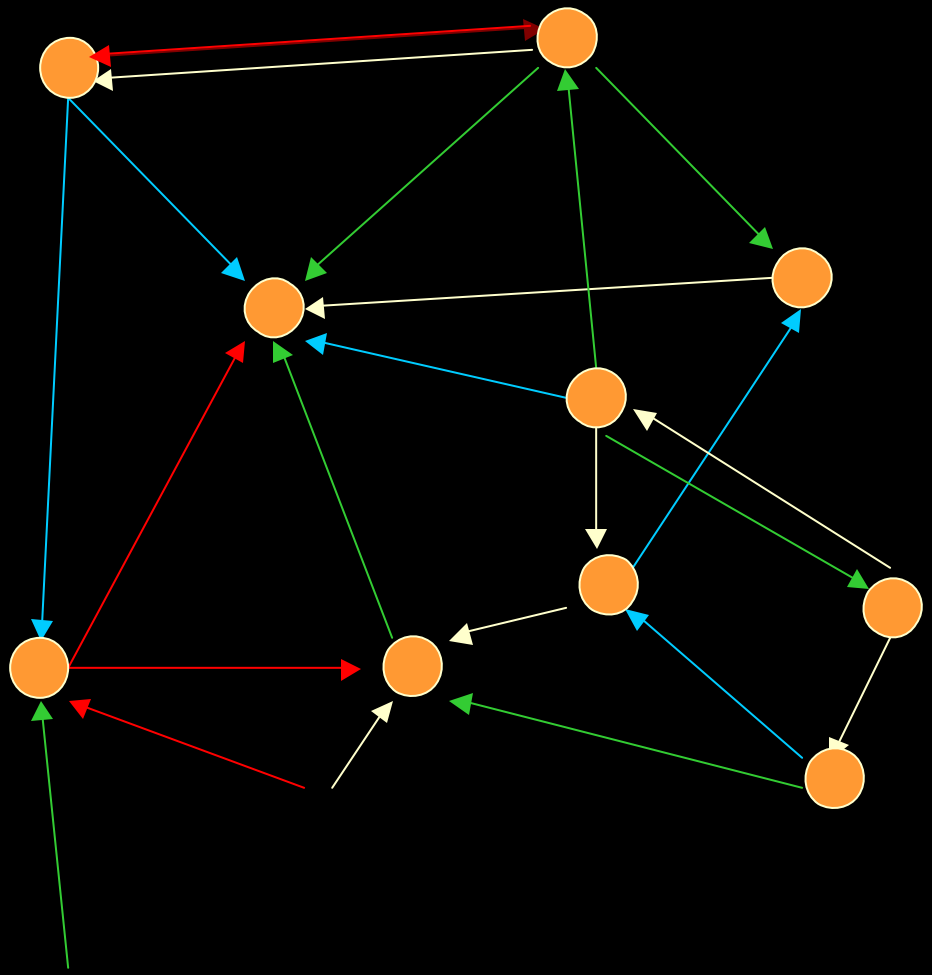
rules	definitions	algorithms, code
conjectures	theorems	specifications
inferences	certificates	proofs, partial proofs

INFORMAL

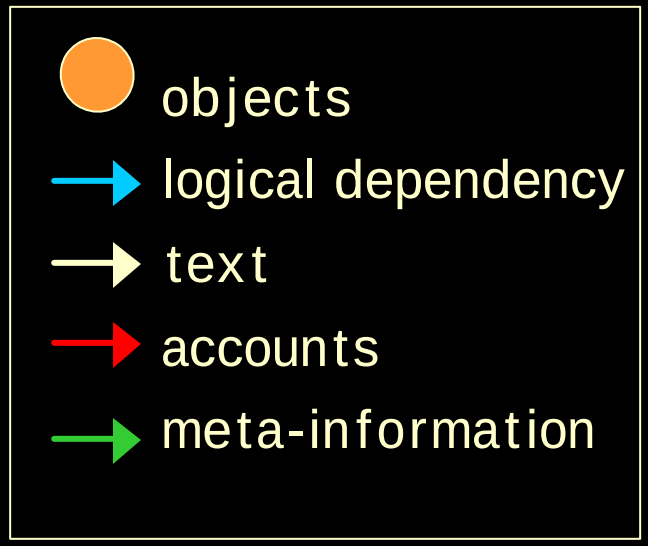
displays	articles	documentation

CORNELL

Information Graph of the FDL



Objects have interdependencies and informative links



Proof Example

proof

```
⦿ *- PRF : pos_length @edd.lori_10_13 @nuprl4.cs.corne
* top
∀A:𝕌. ∀l:A List.  ((¬(l = [])) ⇒ (||l|| ≥ 1 ))

BY aux_auto(RepeatMFor 2 (D 0) THENM D 2)

* 1

1. A : 𝕌
⊢ (¬([] = [])) ⇒ (||[]|| ≥ 1 )

BY auto(D 0 THENM D (-1))

* 2

1. A : 𝕌
2. u : A
3. v : A List
⊢ (¬([u / v] = [])) ⇒ (||[u / v]|| ≥ 1 )

BY AbReduce 0 THEN aux_auto(D 0)

* 2 1

4. ¬([u / v] = [])
⊢ (||v|| + 1) ≥ 1

BY InstLemma `non_neg_length` [⌜A⌝;⌜v⌝] THEN Auto
```

article

accounting

theorem

$$\forall A:\mathbb{U}.\ \forall l:A\ \mathrm{List}.\ \ (||l|| \geq 0\ )$$

Accounting Archive

FDL performs archival functions…

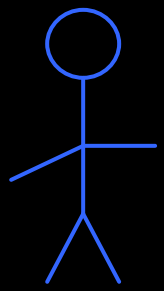
- Automath system Auto QE checked the following formalization of Landau's Grundlagen (February 22, 1974)
- EQP proved the Robbins Conjecture, an open problem for mathematicians since 1933 (October 10, 1996)
- JProver proves "the Butler did not do it" in the this logic-based "who-done-it" problem (October 13, 2000)
- CLI presents a formal model of the Java Virtual Machine (May 13, 1997)
- ACL2 verifies Y2K conversion rules (March 21, 2000)

Basic FDL Operations

- search by name
- lookup object by id
- create an object (empty)
- save
- undo
- “activate” object
- “deactivate” object

Translation Operations

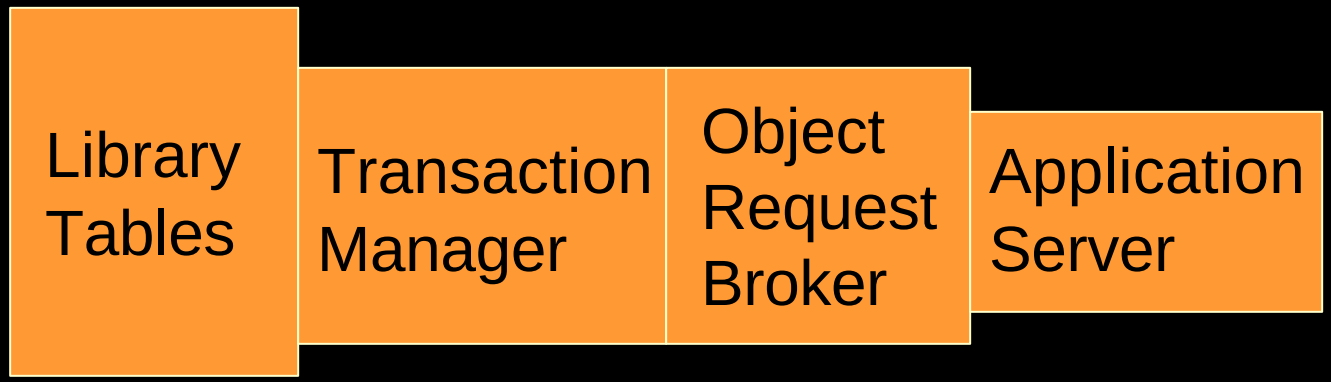
FDL performs translations on collections, or subgraphs.



Give me the smallest *closed* set with respect to logical references of objects that contains the proof of Total Order in Distributed System, D.

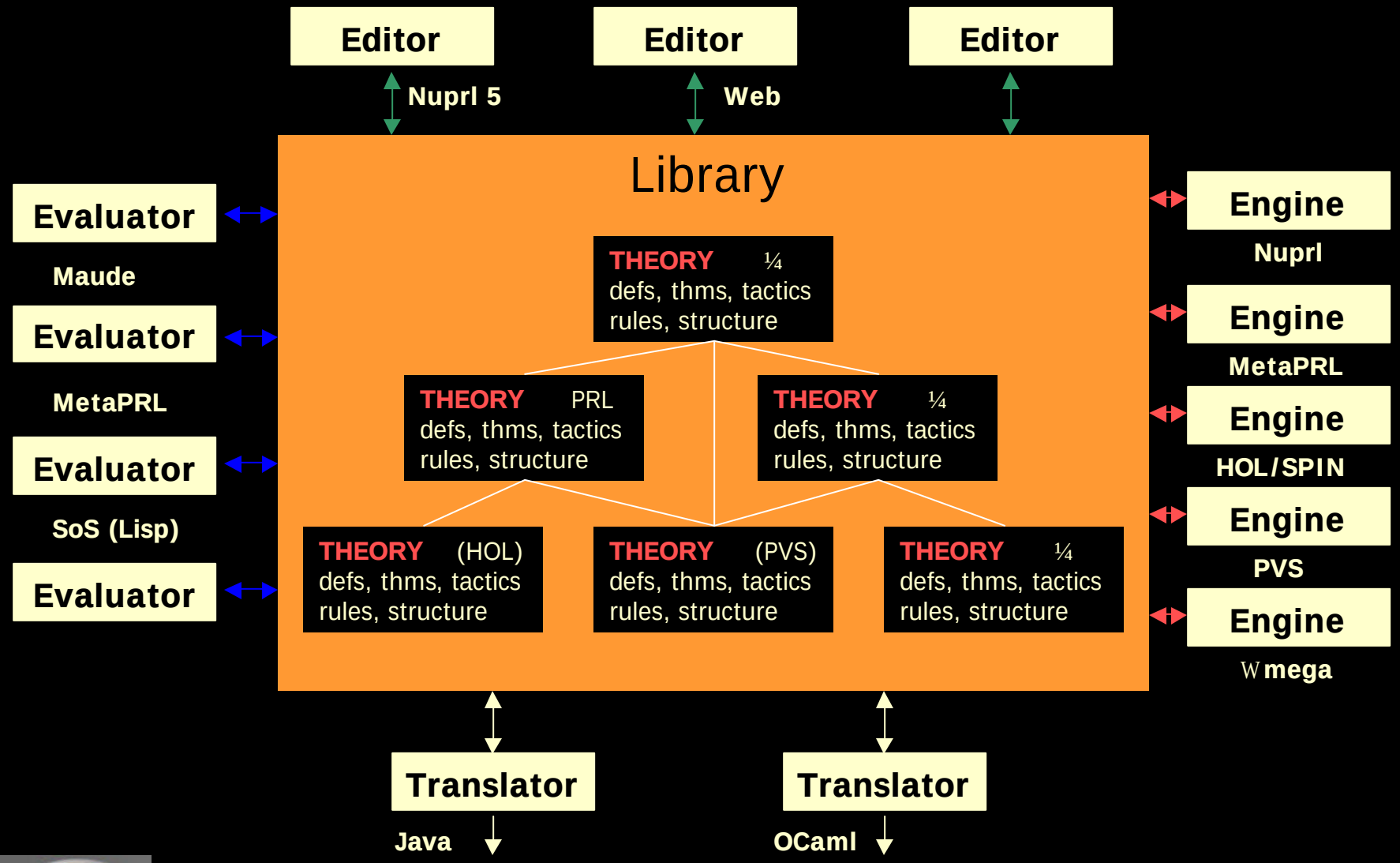
Give me the largest *closed* set with respect to textual links that contain information about prime numbers.

Implementation



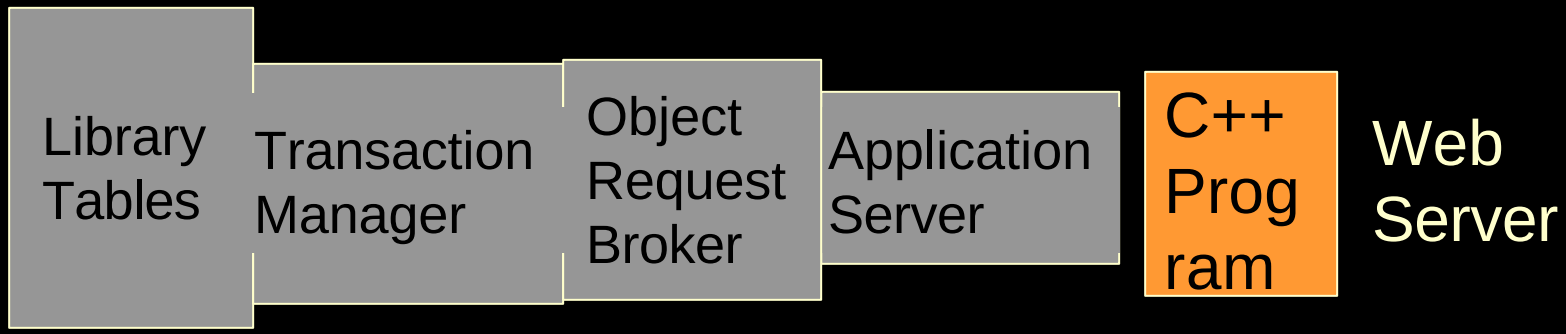
- LISP/ML based system
  - 6,000 functions
  - 62,000 lines code
  - 22,000 lines comments
- Adapted from Logical Programming Environment and Nuprl

Architecture



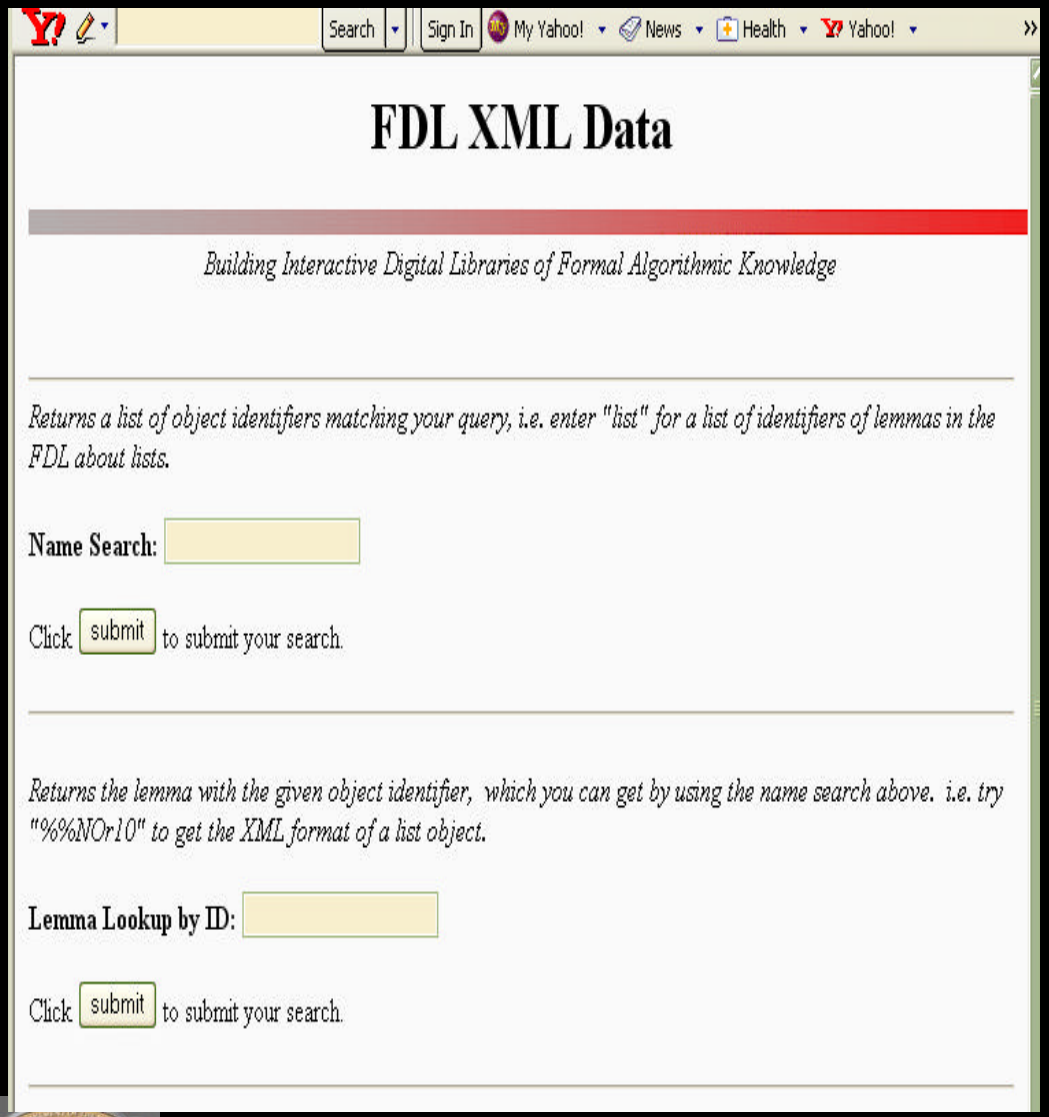
FDL Online Demo

- search by name
- lookup object by id



- vnc into editor

FDL Online Demo




```
<fdl:term>
<fdl:param_list>
  <fdl:param>
    <fdl:paramtype>t</fdl:paramtype>
    <fdl:paramvalue>implies</fdl: …>
  </fdl:param>
</fdl:param_list>
<fdl:term_list>
  <fdl:term>
    <fdl:param_list>
      <fdl:param>
        <fdl:paramtype>t</fdl:paramt
        <fdl:paramvalue>equal</fdl:
      </fdl:param>
    </fdl:param_list>
        …
```

Current Contents

- Nuprl and PVS Content
  - 422 theories (collections)
  - 1699 definitions
  - 7351 theorems
  - 3867 proofs
- MetaPRL (Cal Tech)
- In Progress – import many more in xml and mathml formats

CORNELL

Is Data Format Sufficient?

- Term: Operator (Op) x Term List
- XML (DTD for HELM)
- OMDoc
- MathBus
- Ascii (compressed and uncompressed)

Op = Opname{params}

op(Term List)
add{}(x;y)
x=natural{1}()
y=mult(z;r)

Contribution Methods

System	Method
Jprover	MathBus
C++ Program	Ascii uncompressed
MetaPRL	MathBus
OMDoc	XSLT
PVS	LISP structures
Larch from ATC	In progress

CORNELL

Related Work

- HELM (Hypertextual Electronic Library of Mathematics) - http://helm.cs.unibo.it/
- MathWeb - http://www.mathweb.org
- Mathematical Knowledge Management (MKM) - http://imps.mcmaster.ca/na-mkm-2002/presentation
- Wolfram Research MathWorld - http://mathworld.wolfram.com/
- Living Book - http://www.slicing.de/books

CORNELL

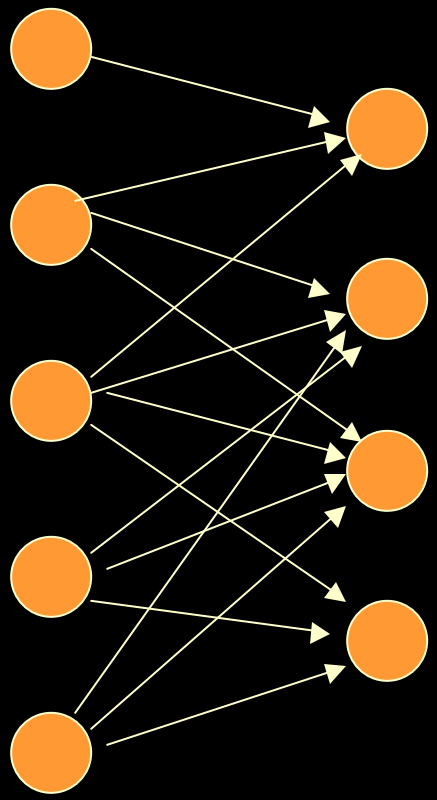
Questions

- What communities to target?
- How to most effectively search?
- Data format justification?

Questions

- What communities to target? Education, Code
- How to most effectively search? dependency
- Data format justification? MathML future?
- What usability studies and evaluations can we perform?
- Can we learn from semantic knowledge?

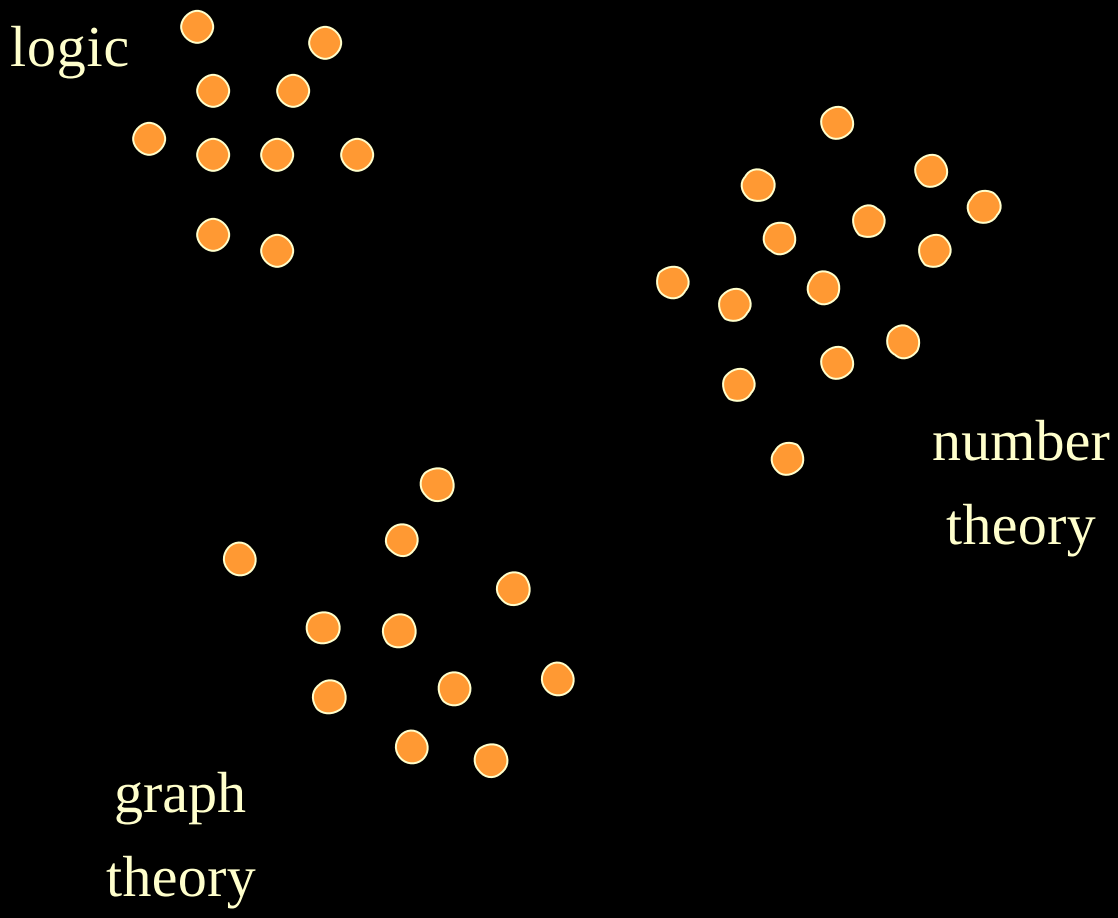
Learning from the Information Graph



hubs

authorities

Classifying by Eigenvectors



Topics

- The Office of Naval Research (ONR) Digital Library Project
  - Goals, Design, Implementation
- Other Information Science Projects
  - National Science Digital Library, NSDL
  - Ranking Function
  - Modeling Bursts of Activity on the Web

NSF NSDL

digital library: "A managed environment of multimedia materials in digital form, designed for the benefit of its user population, *structured* to facilitate access to its contents, and equipped with aids to navigate the global network ... with users and holdings totally distributed, but *managed as a coherent whole*."

--Mel Collier, *International Symposium on Research,Development, and Practice in Digital Libraries 1997*

NSF NSDL

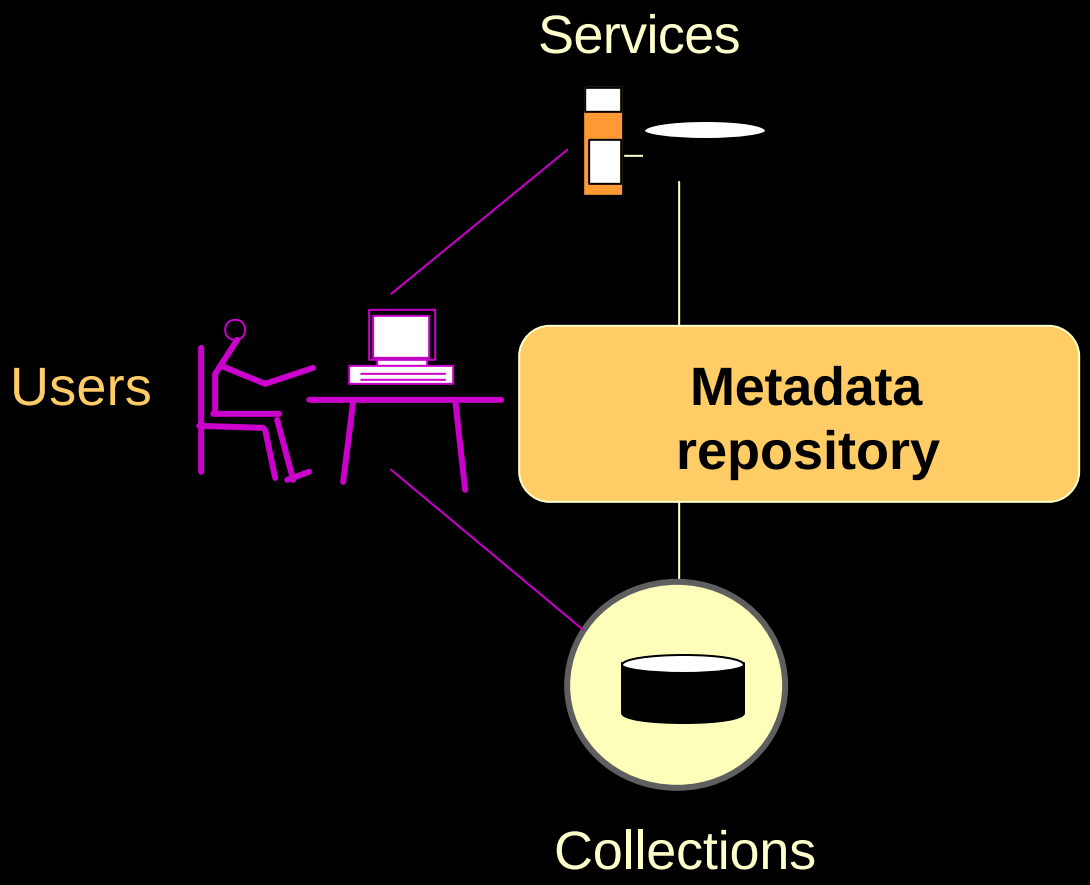
digital library: "A managed environment of multimedia materials in digital form, designed for the benefit of its user population, *structured* to facilitate access to its contents, and equipped with aids to navigate the global network ... with users and holdings totally distributed, but *managed as a coherent whole*."

--Mel Collier, *International Symposium on Research,Development, and Practice in Digital Libraries 1997*

Bill Arms (director), Naomi Dushay, Carl Lagoze, Sandy Payette, Carol Terrizi, Vicky Weissman, <more>

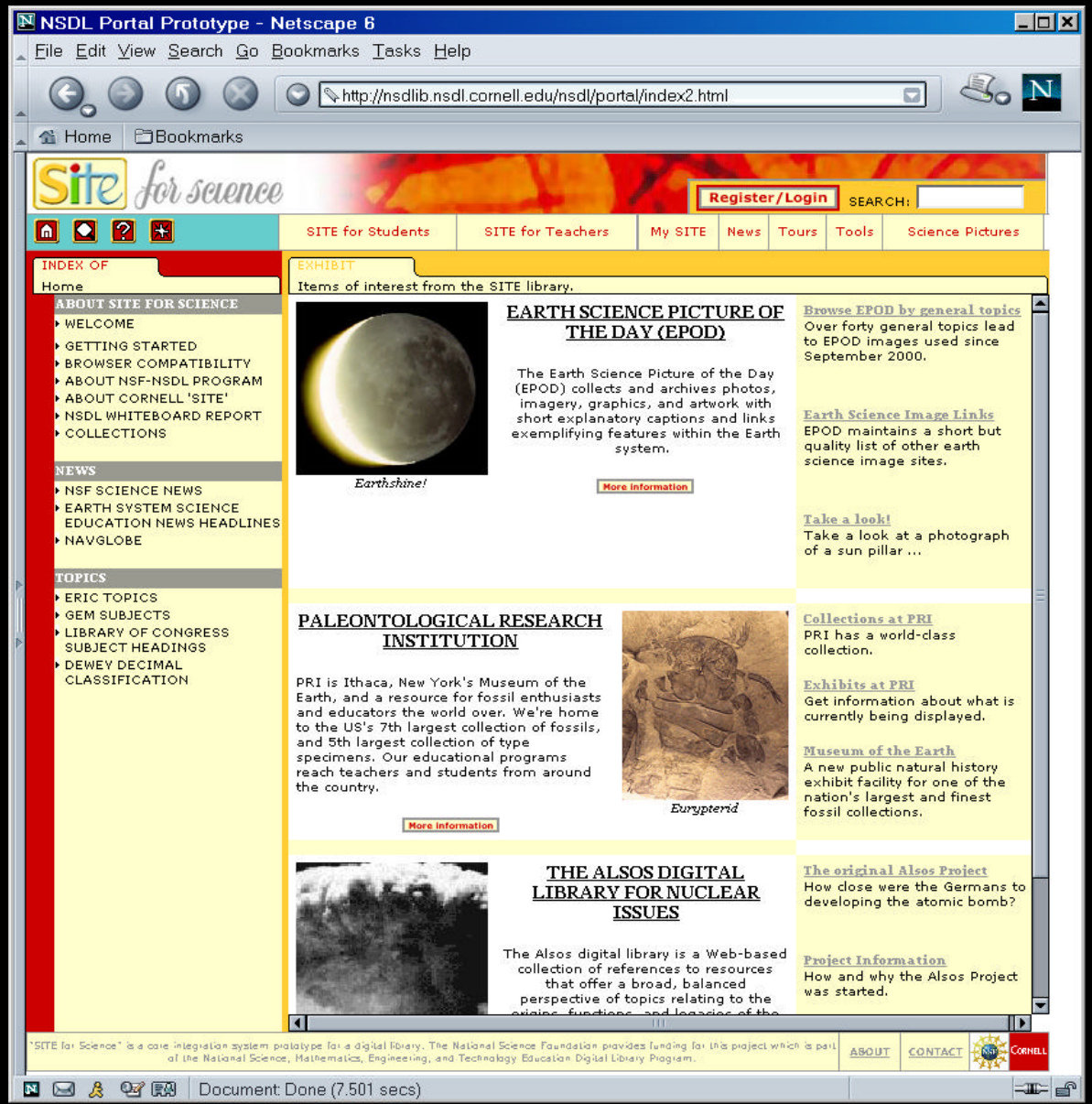
Metadata Repository

The Open Archives Initiative Protocol for Metadata Harvesting: Los Alamos, Nasa, and Cornell



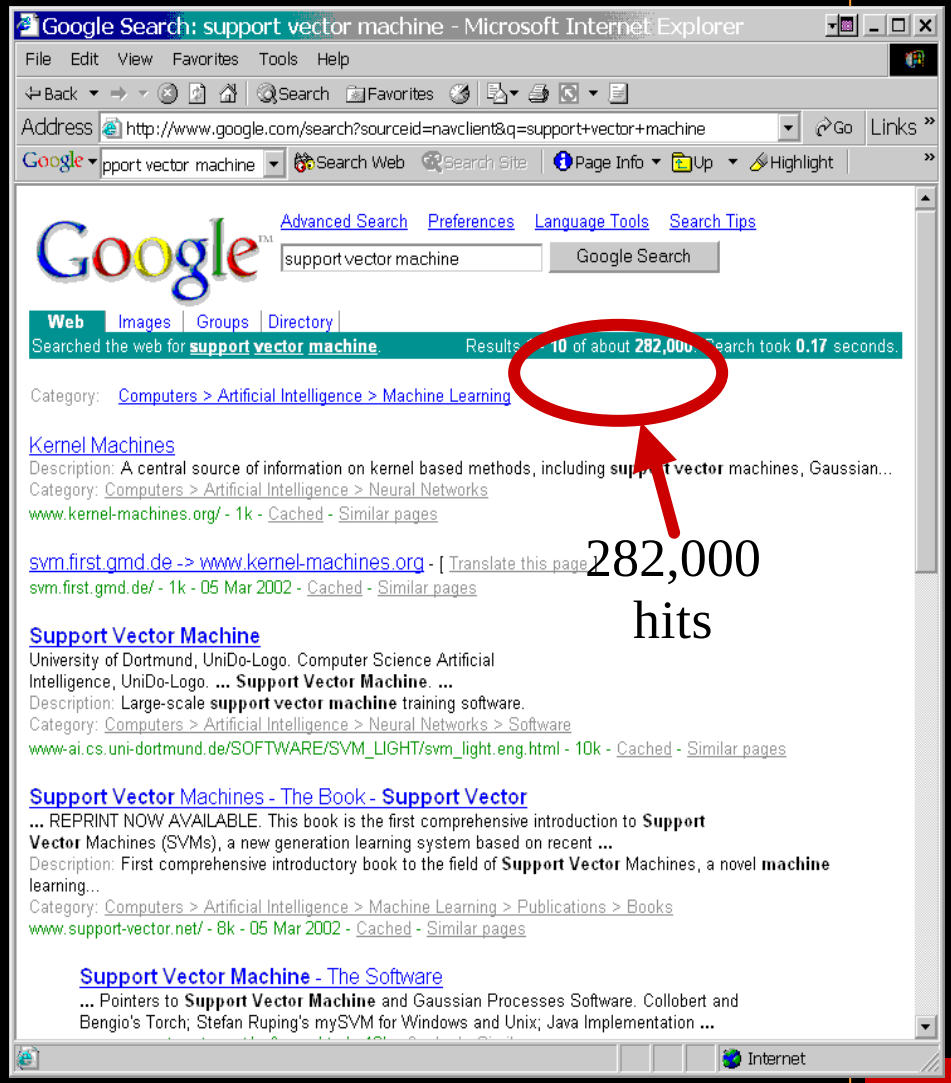
http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

Cornell's NSDL Portal



Using Machine Learning in Ranking

- T. Joachims
  - Support Vector Machines
- Classification
  - "A is good", "B is bad"
  - Often not appropriate
- Ranking
  - "A is better than B"
  - Information Retrieval
- Example
  - Improve Search Engine using Machine Learning



Experiment

- Training- 260 queries from 20 users
- Learning- ranking function, *r*
- Comparison- Google, MSN, TopRank
- Results- *r* shows improvement with a 95% significance using two-tailed binomial sign test
  - Google: in 13/88 queries users clicked on more Google links, and in 29, more from *r*

CORNELL

Modeling Bursty Streams, J. Kleinberg

- What are Bursts of Activity? – sharp rises in frequency of certain features of the stream

Email Messages	Conference Titles	Current Affairs	Web Usage Data

- Bursts are modeled as state transitions in a weighted automaton model

ACM SIGKDD '02

CORNELL

US Presidential State of the Union Address (1790 to 2002)

Word	Interval of burst
gentlemen	1790 - 1800
whilst	1857 - 1860
emancipation	1862 - 1864
coinage	1877 - 1886
depression	1930 - 1937
war	1942 - 1945
peacetime	1945 - 1947
inflation	1971 - 1980
women	1981 - 1984
college	1995 –
communities	1995 -
america	1996 –

CORNELL

Computing and Information Science

Cornell University

CIS

Cornell

"studies digital information in its human and social context."
Dean Robert Constable

- *Information systems*
- *Social systems*
- *Human centered systems*

Research Programs
Undergraduate Education
Courses
Concentration/Minor / Major
Graduate Education
Faculty and Researchers
Information Science Building
IS Seminar Series

INFORMATION SCIENCE MAJOR JUST APPROVED

Cornell

Reference Cites

- FDL www.nuprl.org/FDLproject
- NSDL www.nsdl.org/
- Rank www.cs.cornell.edu/People/tj/
- Bursts www.cs.cornell.edu/home/kleinber
- Info Science www.cis.cornell.edu/

Thank you!

lolorigo@cs.cornell.edu