Services of the Formal Digital Library (FDL)

NA-MKM 2004

January 6, 2004

Lori Lorigo, Cornell University

Talk Outline

- FDL Guidelines
- Search Service
- Accounting Service
- Proof Sharing Scenarios

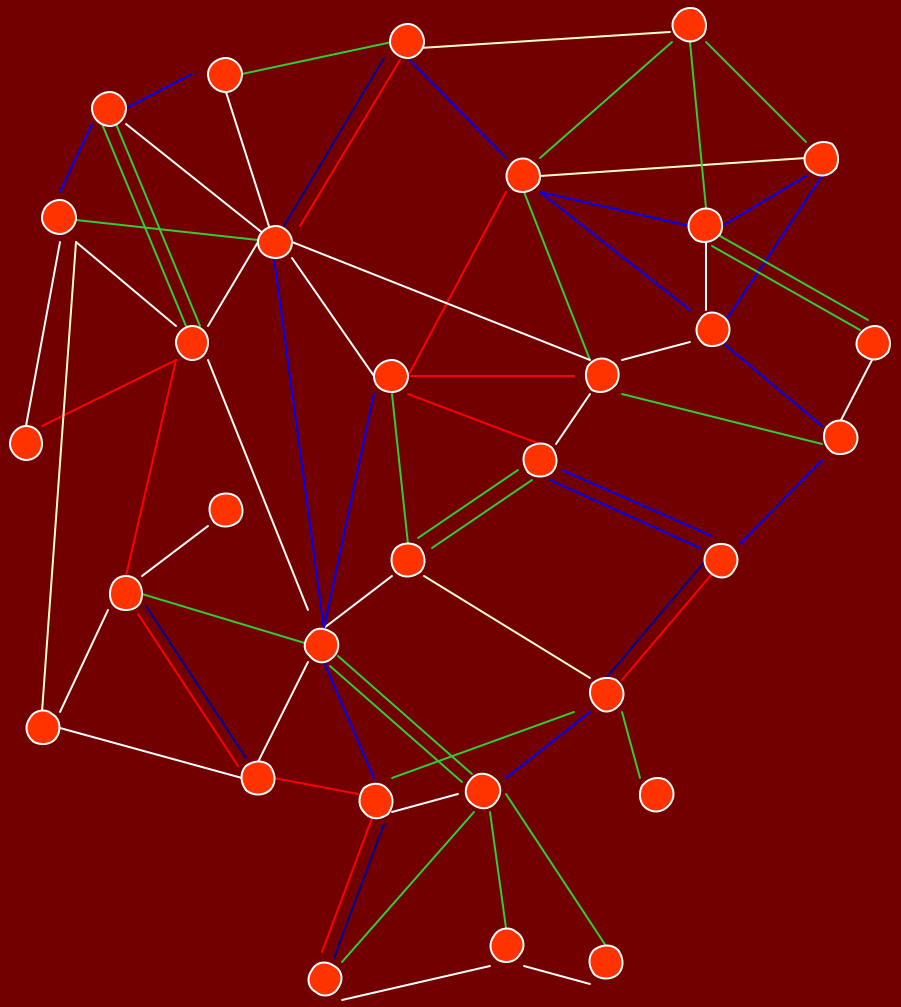
FDL Guidelines

- Offer theoretical neutrality
- Include results from all major provers
- Offer knowledge management services
- Avoid a monolithic software system

Cornell/Caltech/Wyoming MURI ONR research grant

Cornell NSF National Science Digital Library (NSDL) funding

Knowledge Network



- Objects: definitions, theorems, proofs, texts, certificates, inference steps, code

→ logical dependency

→ textual links

→ accounting links

→ metalogical links

Search and Classification

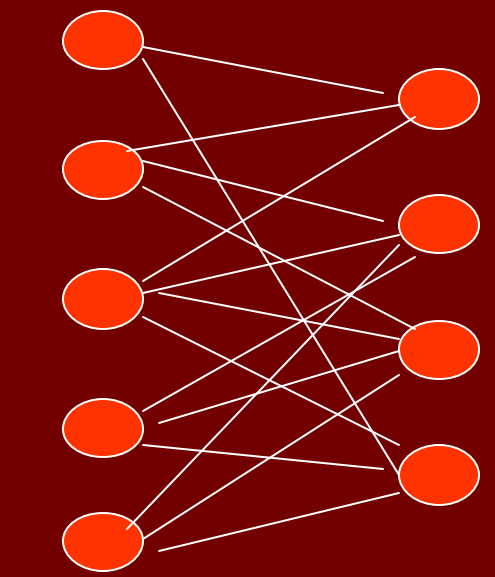
- Can we exploit the graph structure of the FDL to learn about its contents?
- Internet search engines infer information about web pages based on hyperlink structure
  - if A points to B then A thinks B is important

"Take advantage of current web search methodologies and apply them to our library, while capturing and exploiting characteristics of the Formal Methods search space."

Hubs and Authorities

Hubs are pages that point to a lot of good Authorities.
Authorities are pages pointed to by a lot of good Hubs.

- Graph theoretic approach
  - nodes = pages; edges = links
- Sparse adjacency matrix A
  - hubs = 1st eigenvector of AA^T
  - authorities = 1st eigenvector of A^TA
- SubCommunities
  - nonprinciple eigenvectors
  - densely connected bipartite graphs



Kleinberg '98 HITS

Application to the FDL

- Intuition
  - authorities are key definitions, basic functions
  - hubs are theorems that span a particular theory
  - subcommunities
- Design Choices
  - Nodes are all rules, definitions, theorems
  - Edges are transitive closure of the logical dependencies
- Levels of Authority

Preliminary Results

- Nuprl standard library authorities are key definitions
  - all, member, implies, prop, and, iff
  - subcommunities were contained within man-made directories
- Integer library authorities are definitions and basic theorems
  - divides_wf, gcd_p_wf, gcd_p_sym, gcd_wf, comb_for_gcd_p_wf, gcd_p_zero, gcd_p_shift"
- Integer library subcommunities
  - "eqmod_weakening" "eqmod_transitivity" "coprime_iff_ndivides" "eqmod_fun"
  - "gcd_p_neg_arg_a" "gcd_p_neg_arg_2" "gcd_elim" "gcd_is_divisor_2"
  - "decidable_or" "decidable_int_equal" "comb_for_not_wf" "spread5"
  - "comb_for_segment_wf" "select_listify_id" "comb_for_map_wf" "map_select"

Related Challenges

- Can graph theoretic search services help automate
  - classification of objects?
  - creation of content by suggesting related theorems, perhaps also using other kinds of dependency links?
- Desire to combine with pattern matching or name-based searches
- How should the interface be designed and which services are most desired?

Accounting

- Information is not equal to Knowledge
- Account for truth in a way that is machine checkable
- Users may not agree on what is acceptable

Certificate System

- Have 2 components
  - Pointer to object it certifies
  - Pointer to the certificate kind, which contains a program that when executed checks it's claim
- Desire: The client may infer from a certificate of a formal proof that it conforms to specific methods of inference, and could test this claim
- Multiple certificates
  - may point to the same certificate kind
  - may point to the same object
  - the kind can point to other objects

Accounting Challenges

- Designed and documented, not yet implemented
- Overhead of storing certificates for each object
- Engaging the community to provide content
- Deepening understanding of certificate uses in practice

Proof Sharing

Formal sharing, 1 logic	Informal sharing	Formal sharing, 2 logics
PVS — PVS	Coq — document, heuristic	HOL — Nuprl* Restricted

What are other potential scenarios of proof sharing, are their needs met?

Outreach Challenge more content and services, greater accessibility