RFC-009: Library Directory System

Abstract

This RFC specifies the Library Directory System for the Library of Cyberspace: a searchable catalog of research papers with natural language queries, fuzzy matching, and beautiful output. Finding knowledge should be delightful.

E Pluribus Unum

```
╔══════════════════════════════════════════════════════╗
║  □ Cyberspace Library Directory                        ║
║  Finding knowledge should be joyful                    ║
╚══════════════════════════════════════════════════════╝
```

Motivation

The Library of Cyberspace contains 421+ research papers spanning 50 years of computing history. Without a catalog:

- **Discoverability**: How to find Lamport's clock paper?
- **Connections**: What relates to SPKI?
- **Context**: When was this written? By whom?

The Directory System provides:

1. **Natural language search** - "papers by Lampson about protection"
2. **Multi-index access** - By author, topic, year, collection
3. **Fuzzy matching** - "Lamp" finds Lampson and Lamport
4. **Beautiful output** - Unicode, emoji, formatted results
5. **S-expression export** - Machine-readable catalogs

Data Model

Document Record

```
(define-record-type <document>
  (make-document title authors year topics file-path
```

```scheme
                    collection source-url description)
  document?
  (title document-title)
  (authors document-authors)      ; List of strings
  (year document-year)            ; Integer or #f
  (topics document-topics)        ; List of strings
  (file-path document-file-path)
  (collection document-collection)
  (source-url document-source-url)
  (description document-description))
```

Collection Record

```scheme
(define-record-type <collection>
  (make-collection name path documents description)
  collection?
  (name collection-name)
  (path collection-path)
  (documents collection-documents)  ; List of documents
  (description collection-description))
```

Library Record

```scheme
(define-record-type <library>
  (make-library collections documents
                author-index topic-index year-index collection-index)
  library?
  (collections library-collections)
  (documents library-documents)
  (author-index library-author-index)     ; Hash table
  (topic-index library-topic-index)        ; Hash table
  (year-index library-year-index)          ; Hash table
  (collection-index library-collection-index))
```

Index System

Four hash table indexes for O(1) lookup:

```
author-index:     "Lamport" → [doc1, doc2, doc5, ...]
topic-index:      "cryptography" → [doc3, doc7, doc12, ...]
year-index:       1978 → [doc1, doc4, ...]
collection-index: "lampson" → <collection>
```

Building Indexes

```scheme
(define (build-author-index documents)
  (let ((index (make-hash-table equal?)))
    (for-each
      (lambda (doc)
        (for-each
          (lambda (author)
            (hash-table-update!/default
              index
              (normalize-author author)
              (lambda (docs) (cons doc docs))
              '()))
          (document-authors doc)))
      documents)
    index))
```

Query Interface

Natural Language Parsing

```scheme
(parse-query "papers by Lampson about protection")
;; => (query (author "Lampson") (topic "protection"))

(parse-query "from 1978")
;; => (query (year 1978))

(parse-query "SPKI certificates")
;; => (query (topic "SPKI") (topic "certificates"))
```

Recognized Patterns: - by <name> → author search - about <topic> → topic search - from <year> → year search - in <collection> → collection filter - Bare words → topic search

Fuzzy Matching

```scheme
(fuzzy-match "Lamp" author-list)
;; => ("Lampson" "Lamport")

(fuzzy-match "crypto" topic-list)
;; => ("cryptography" "cryptographic" "encryption")
```

Matching strategies: 1. Exact match (highest priority) 2. Case-insensitive match 3. Prefix match 4. Substring match

Query Functions

```scheme
(find-by-author library "Lampson")
(find-by-topic library "SPKI")
(find-by-year library 1978)
(find-by-collection library "lampson")
(find-related library document)
```

REPL Interface

Interactive Session

```
$ csi -s directory-repl.scm
```

```
╔══════════════════════════════════════════════════╗
║  ⬡ Cyberspace Library Directory                   ║
║  "E Pluribus Unum" - Finding knowledge should be joyful  ║
╚══════════════════════════════════════════════════╝
```

```
Loaded 312 documents from 18 collections
Type 'help' for search tips, 'quit' to exit

⬡ > papers by Rivest

⬡ Found 3 documents:

 [1] SDSI: A Simple Distributed Security Infrastructure
     ⬡ R. Rivest, B. Lampson
     ⬡ 1996  ⬡ lampson

 [2] SPKI Certificate Theory
     ⬡ C. Ellison, B. Frantz, B. Lampson, R. Rivest, ...
     ⬡ 1999  ⬡ lampson

⬡ > about SPKI

⬡ Found 8 documents:
...

⬡ > stats

Library Statistics:
  ⬡ Documents: 312
  ⬡ Collections: 18
  ⬡ Authors: 245
```

```
   ▢ Topics: 89
   ▢ Years: 1971-2024

▢ > quit
Happy exploring! ▢▢
```

Commands

Command	Description
help	Show search tips
stats	Library statistics
collections	List all collections
authors	Show author index
topics	Show topic index
years	Show year distribution
quit	Exit REPL

INDEX.md Parser

Each collection contains an INDEX.md with paper listings:

```
# Lampson Collection

Butler Lampson's papers on protection, capability systems, and distributed computing.

## Papers

1. **Protection** (1971)
   - Author: Butler Lampson
   - Topics: capability, protection, access control
   - Source: ACM Operating Systems Review

2. **Hints for Computer System Design** (1983)
   - Author: Butler Lampson
   - Topics: systems design, engineering
   - Source: ACM Operating Systems Review
```

Parser Output

```
(parse-index-file "lampson/INDEX.md")
;; =>
(collection
  (name "lampson")
  (description "Butler Lampson's papers...")
```

```
(documents
  ((title "Protection")
   (authors "Butler Lampson")
   (year 1971)
   (topics "capability" "protection" "access control"))
  ...))
```

Export Formats

S-expression Export

```
(export-to-sexp library "catalog.sexp")
```

Output:

```
(library
  (metadata
    (document-count 312)
    (collection-count 18)
    (generated "2026-01-06"))
  (collections
    (collection
      (name "lampson")
      (documents
        (document
          (title "Protection")
          (authors ("Butler Lampson"))
          (year 1971)
          (topics ("capability" "protection")))
        ...))
    ...))
```

HTML Export

```
(export-to-html library "catalog.html")
```

Generates browsable web catalog with: - Alphabetical author index - Topic tag cloud - Year timeline - Search functionality

Markdown Export

```
(export-to-markdown library "catalog.md")
```

Documentation-friendly format for README files.

Design Philosophy

Super Nice and Helpful

- Natural language understanding
- Fuzzy matching catches typos
- Suggestions for empty results
- Encouraging prompts

Beautiful Output

- Unicode box drawing (   )
- Emoji visual markers (   )
- Consistent formatting
- Progressive disclosure

Hone In On What You Need

- Multi-strategy search
- Progressive filtering
- Cross-references
- Related paper discovery

Security Considerations

No Execution

The directory system is read-only: - Parses INDEX.md files (trusted) - No code execution - No network access - No file modification

Path Traversal

File paths validated: - Must be within library root - No `..` components - Normalized before use

---

Implementation Notes

Dependencies

- `srfi-1` - List utilities
- `srfi-13` - String utilities
- `srfi-69` - Hash tables
- `chicken file` - Directory scanning

Performance

- Index building: O(N) documents
- Query: O(1) hash lookup + O(M) result filtering
- Fuzzy match: O(K) candidates

Extensibility

Future enhancements: - Full-text PDF search (OCR) - Citation graph visualization - BibTeX export - Web API server - Semantic search with embeddings

---

References

1. Dublin Core Metadata Initiative
2. Library of Congress Subject Headings
3. BibTeX format specification
4. Unicode Standard for box drawing

Changelog

- **2026-01-06** - Initial specification

Implementation Status: Complete **Test Status:** Passing (test-directory.scm) **Documents Indexed:** 421+