Book Collection Project

DB Issues

PSN 8/16/06

Status of Project

- As of last month, draft first-cut project pages on SCC website. (with help from Mike Powell)
  - Focused on organization
  - Only small changes since (not uploaded)
  - [Brief review as partial basis discussion](#)

- Then, useful email discussion with Randall, Paul, Bernard about classification issues

- Which suggested (at least to me) that the subject of a Book Collection DB couldn't be neglected further ➔

Book Collection DB?

- **Why?** Unlike other SCC collections of software-related materials, can't maintain or present data in linear list

- **Why not same as museum DB?**
  Requirements for on-line collection different
  - Need different "metadata"
  - need different browse, search, presentation and maintenance mechanisms.

- But the two must be related

Outline

- Classification
- Strawman RDB schema

Why classify?

- To shelve the physical books.
  - For this need stable system with single subject assignments (a book can be in only one place).
  - Best if no knowledge of subject matter required for decisions. So use an existing classification.
  - Randall's suggestion of the L of C subject codes looks good (better than Dewey).
- To provide a basis for organizing the on-line collection
  - Allow it to be presented in an understandable way
  - Give succinct grasp of what each book in a list is about.
  - Provide a basis for selecting groupings for annotation, e.g. subject/period
  - Provide basis for study of the history: what kinds of things were published when.

Classification Issues for Online Collection

- <u>For many books, no single satisfactory primary subject</u>
  - What is the subject of a "history of X"? X? or "History"?
  - Paul MacJones: allow use of several subjects for books, with each subject being a node in a shallow hierarchy.

    (A book can be classified as both "a/b/c" and "d/e/f")

- <u>Class structures & classifications unstable over time</u>
  - basic works in a subfield at first not easily classifiable
  - work in one field later adapted for work in another,

    Initial tendency to group adaptations with source

- <u>Handling systematic multi-class memberships.</u>
  - E.g, if book in class X/history then always also in class History, or History/x.
  - Can handle by lattice structure, and/or rules.

Classification Issues, ctd.

Use existing system & classifications?

- Library of Congress subjects?
  - single hierarchy, stable, some assignments bad
  - (but good lists of multiple subjects within book entries)
- ACM Subjects?
  - multiple subject assignments, shallow hierarchies,
  - But:
  - One subject primary, often too many secondaries
  - Class hierarchies revised infrequently
  - Some early classifications bad

Use existing classification info? yes, as bases for decisions

Strawman DB Schema

Classification

Categories: <category id, long name>
                    <ALG, Algorithms>
                    <TH.ALG, Algorithms from theoretical perspective.>
                    <TH, Computer Science Theory>

CategoryLattice (1:n) < category id, parent id>
            Dotted category names imply one path in lattice.
            If a category is x.y, the implied path is x/x.y

            So: <TH.ALG, TH> is implicit
                    <TH.ALG, ALG>   is not

BookSubjects (1:n) <ISBN, category id, default?)
            <ISBNxx, TH.ALG, T>

Strawman DB Schema, ctd

Links to physical collection:

    Accessions (n:1) <u>Museum accession number</u>, ISBN>

    Shelving: <ISBN, LofC Subject>

Basic Entry

Books: < ISBN, originators, auth/ed?,

          title, volume, publisher, edition, year >

Annotation Links:

BookAnnote: <ISBN, filename>

Group Annotation

       GroupAnnote: <GAID, filename, title>

       GACat (multi) <GAID, category>

       GAPeriod   <GAID, start year, end year>

Notes

- Implementation issues
- Presentation issues

Out of time.

End

Goals as of July/2006

- **Organizing publication list by subject/date**
  - **Why?**
  - **For understandable presentation**
  - **To provide foci for annotation, not just of individual but of groups of related books.**
- **Selecting significant pubs or pub groups**
- **Obtaining context entry for each**
- Helping to obtain their editions for the CHM
- Providing online versions
  - when copyright arrangements can be made