

# Task Force on Pathfinder Indexing - Final Report

May 31, 2001

## **Introduction**

The Task Force on Pathfinder Indexing was created to look at various requests to improve search results in the Gladis Database as it is presented in Pathfinder. The members were asked to examine these requests and sift through the various implications of making such changes based on their experience as programmers, reference librarians, and catalogers. It should be acknowledged that a number of features of Sitesearch, upon which Pathfinder is based, limit the ability to make certain changes, thus the recommendations below only address programming enhancements that can be achieved locally. There were other issues, such as the confusion that results from the use of the term "phrase search," which the Task Force felt went beyond the charge of examining indexing issues. We did, however, make one recommendation to alter the use of this term and that alteration appears in Recommendation 3.

Numerous suggestions and requests were examined and the committee recommends their implementation as described below. All of the recommendations listed below, if accepted, can be programmed at one time; thus the list does not prioritize the recommendations. The adoption of these recommendations will require an extensive revision of the Pathfinder help and information screens, a review of which has not been made in some time.

An experimental database will be created so that all Library staff can view these recommendation both by performance and by the look-and-feel of the Pathfinder screens should such changes be enacted.

## **Problem/Recommendation**

1) The problem of "too many hits" for voluminous personal authors.

Recommendation: It is difficult to use Pathfinder to find individual works of voluminous personal authors since the personal author search as it is now constructed includes a number of MARC fields other than the 100 field for the first (or main) personal author. One way of resolving this issue would be to create new search options that limit the author search only to the 100 field and to label this search appropriately. We recommend the name First Personal Author Keyword and First Personal Author Phrase search to differentiate these searches from the current options. Since this search has specific applications, we recommend its appearance only from the Full Feature search screen with appropriate "help" screens directing the patron to this option when the current author search results in "too many hits".

Pros: A search using this new index would allow patrons to shrink the result to those records where the author is the first (or main) personal author. The benefits can be

demonstrated by the example of a patron who would like to locate a copy of Shakespeare's Hamlet avoiding translations, adaptations, or other variants that are retrieved in the current author search.

Cons: A limited search such as the one proposed would impede searches for works having several authors when the patron does not know which author is listed first, and where cataloging rules prescribe a concept of authorship which patrons may not understand. Some confusion can be avoided by labeling this search as First Personal Author, alerting the patron to the fact that it has a very specific application. To reduce false results, the Task Force recommends offering this search option only in the Full Feature drop down menu. Addition of these search options would also increase the list of options available on the Full Feature Search drop down menu which some feel is already unwieldy.

2) The problem of "too many hits" for Title searches.

Recommendation: The Title index includes a number of MARC fields apart from the main title and its variants. Among these are the Uniform Title, the Series fields, and the added entry title fields. Each of these has its advantages and uses. Inclusion of Uniform Titles assists patrons in the field of Music, where it is difficult to find works of major composers. Inclusion of the Series fields benefits those disciplines where searching on words in a series and a volume number is the quickest way to the desired record. [It should be noted that many patrons do not understand the concept of "series" and frequently search series as a title.] The Task Force recommends leaving the content of the Title index as it now stands and creating two new indexes, which contain only the main title and its variants (245, 246, and 247). We recommend adding these two new indexes to the Full Feature index list and calling them Main Title Keyword and Main Title Phrase.

Pros: The maintenance of the Title search in its present form would assure continuity in Pathfinder searching that many patrons expect. The option of adding the two Main Title indexes will provide a way to decrease the number of titles that are retrieved due to the Uniform Title and Series indexes being excluded from this specialized index.

Cons: Even with the changes proposed, certain searches may still retrieve too many hits. Since this search is more limiting than the current title search, it would only be an appropriate option in specific situations. Addition of these search options would increase the list of options available on the Full Feature Search drop down menu which some feel is already unwieldy

3) Linking all subfields of phrase indexes

Recommendation: The Task Force recommends altering all phrase indexes in the following manner. Instead of indexing each subfield individually, all subfields would be concatenated up to 70 characters. The subject phrase index is currently done this way,

and there have been requests to have the Uniform Title phrase index work in this manner. We suggest improving all phrase indexes by concatenation to 70 characters.

Pros: Changing the phrase indexes as described above will allow searches across subfields for more precise retrievals. Browsing is also more natural looking and Gladis-like. Changing the Title Phrase index in this manner would create a true "exact title" search and thus we recommend that the "Title Phrase" label on the Quick Search screen be replaced with either "Exact Title" or "Beginning of Title". This revision would eliminate the confusion created by the term "title phrase" which patrons do not intuitively understand.

Cons: The proposed change would eliminate the ability to search by subtitle (it will no longer be indexed separately) which may be the way a patron remembers the title of the work. The same applies to other subfields in the other phrase indexes. *However, patrons would still be able to search the subfields by using the corresponding keyword index.* The phrase indexes may also respond more often with "too many truncated words" since more individual index entries are created when all subfields are linked. Since the phrase indexes have always been sensitive to truncation, this may go unnoticed.

#### 4) Expansion of Uniform Title search

Recommendation: Currently the Uniform Title index covers only uniform titles in the 130, 240, 243, 730, and 799, but uniform titles are also found in some other fields, especially 7xx \$. The group recommends including these additional fields in the UniformTitle index.

Pros: More thorough searches of Uniform Titles such as those that appear in records for sound recordings.

Cons: There can often be a larger search result which might frustrate some users.

#### 5) Non-retrieval of special collections materials due to current leader code definitions.

Recommendation: The Bancroft library has requested that MARC leader byte 06 ("BLT" in Gladis, "TYPE" in OCLC) be used in Pathfinder to define record formats as "manuscripts" or "manuscripts & archival collections" when the first values are p, t or b. These codes are defined in MARC as "mixed materials (p)" and "manuscript language material (t)". [Prior to format integration, b was defined as "archival and manuscript control" in MARC-AMC records. Although this code is invalid in new cataloging, it needs to be included to retrieve older records.] Pathfinder currently retrieves code t as "Books" and code p as "Mixed Materials", therefore materials cataloged since format integration are retrieved under these formats, while identical types of material cataloged prior to Format Integration are retrieved as "Manuscripts".

Pros: Searching with the new definition of format limit "Manuscripts" would retrieve all manuscript language material and archival collections, not just those cataloged prior to

1996 (Format Integration). This will be true for manuscripts and collections at all campus units, including Bancroft, Environmental Design, Ethnic Studies, Water Resources, and any unit with archival collections and manuscripts.

Cons: Some "mixed materials" might not strictly be manuscripts or archival collections, but false hits resulting from this are expected to be minimal, and perhaps even non-existent. [A search of non-Bancroft "mixed materials" records seems to retrieve archival collections and oral histories.] Also, it should be noted that an item cannot be retrieved as more than one format. So items with leader code "tm", "tc", and "td" will no longer be retrieved with a "Books" search as they were before, and items with leader codes "pm", "pc", and "pd" will no longer be retrieved with a "Mixed Materials" search.

#### 6) Treatment of hyphens or dashes in Pathfinder indexes

Recommendation: Change programming of Pathfinder keyword indexes to treat hyphens and dashes the same as they are handled in Gladis and Melvyl, that is treat the hyphen as a space. Phrase indexes remain unchanged, that is retaining embedded hyphens.

Pros: Users of all systems can be given the same instructions as to the handling of hyphens and dashes. Current usage of Gladis and Melvyl indicates that patrons are comfortable with the way these systems handle the punctuation. We investigated the use of hyphens in Hebrew and Arabic and our informants said that users searching these two languages would not be impeded by this change.

Cons: Searching formulas in titles which have a hyphen or a dash may present problems, and the URL keyword index will be an exception to the rule because its special indexing routine leaves the dashes in. It seemed likely that the number of users performing these kinds of searches in Pathfinder was small, and the problems they would encounter could be alleviated by proper help screens on the use of the hyphen and dash. We also consulted the East Asian Librarians about the use of this punctuation in Chinese and Japanese. Users of these languages would tend to search hyphenated words as one word, which would produce incorrect results. For Chinese, the impending switch to the Pinyin system might alleviate this problem. For Japanese, users would have to be instructed to perform these searches as separate words.

#### 7) Indexing of Relators (\$4 subfield in headings).

Recommendation: Relators describe the roles of personal, corporate, and conference authors. For example, an author such as Benjamin Britten could be a composer, conductor, or a performer. The Combined Author/Relator Index would include personal, corporate, and conference authors, and added entries, linking all subfields up to 70 characters, and including the \$4 if it is within the first 70. Because the features of this index are best utilized as a "browse", we recommend that it be made available only from the Browse screen.

Pros: A browse of this index would show a break down of the author by role, and may even be useful as a headings file substitute, a feature of GLADIS which is not available in Pathfinder.

Cons: If the \$4 relator fell beyond the 70 character limit, it would not be retrievable.

#### 8) New indexes for searching Maps

Recommendation: Add new search options for field 034 \$b (Coded Cartographic Mathematical Data) and 052 \$a and \$b (Geographic Classification Code)

Pros: Enhances the searching of cartographic materials

Cons: None that are evident

#### 9) Enhance Call Number Keyword searching by adding indexing for the Call Number Suffix (i.e. Holding Suf note)

Recommendation: By adding indexing for the Call Number Suffix, patrons would be able to find materials which have storage or special location designations contained in this Gladis field. (e.g. Bioscience TX747.U5 Locked Case 1)

Pros: This enhancement is easy to program and can assist locating materials that have special shelving or storage conditions.

Cons: None that are evident.

#### 10) Changes to searching of Music Publishers

Recommendation: Pathfinder currently has three indexes which treat the music publisher number in different ways. The indexes get little use, and we recommend that two of them, mn: and mn=, be removed and the present mo: index be renamed to mn:.

Pros: Two fewer indexes in the Full Featured list.

Cons: None that are evident

#### 11) Limiting to format "Dissertation"

Recommendation: Currently Pathfinder cannot limit to the format "dissertation" as does Melvyl. By indexing the 502 field in the MARC record Pathfinder would be able to limit to this format.

Pros: Enhanced format limits would help patrons specifically looking for dissertations.

Cons: None that are evident.

Submitted by:

- James Eason, Bancroft Library
- Janet Garey, Systems Office
- Aija Kanbergs, Teaching Library
- Corliss Lee, Teaching Library
- Lee Leighton, ex officio
- Ralph Moon, ex officio
- Cathy Moreno, Technical Services
- John Roberts, Music Library
- Cathy Scheiman, Technical Services
- Allan Urbanic (Chair), Area Studies