

DALNET Search Engine Task Force

Search Interface Recommendations

First Draft August 23, 2000

- I. Search Interface: Functions & Features
- II. Search Interface: Design Principles & Examples
- III. Search Engine Integration: Library Vendor Products

Addendum A: DALNET iPAC Survey – Selected Functions & Features

Addendum B: Resources I – Compiled by DALNET Search Engine Task Force

Addendum C: Resources II – Compiled by Pat Ensor, University of Houston

Addendum D: Search Engine Task Force - Charge

Addendum E: Market Requirements Document - Children's Public Access Catalog

Task Force Members

Robert Chapman, Asst Manager, Bibliographic Division, Acquisitions Specialist
Detroit Public Library

Adriene Lim, Systems Librarian, DALNET Liaison, Technical Services & Systems
Wayne State University

Vanessa Middleton, Electronic Services Librarian & Research Support Center Coordinator
Wayne State University

Leo Papa, Task Force Chair, Associate Dean for Technical Services & Systems
University of Detroit Mercy

Sherman Zhang, Assistant Professor, Library & Information Science Program
Wayne State University

Resource Persons

Deborah Adams, Botsford General Hospital

Duryea Callaway, Wayne State University

Mary Howarth, Oakland Community College

Rhonda J. McGuinnis, Wayne State University

Scott Muir, DALNET Project Leader, Wayne State University

Donna Roe, University of Detroit Mercy

Karen M. Turbolino, VA Medical Center

I. Functions & Features

- I. A. Ease of Use**
- I. B. Workflow**
- I. C. Spell Check**
- I. D. Help**
- I. E. Pre-defined searches**
- I. F. Concept Searching and Synonym Recognition**

II. Design Principles & Examples

II. A. Ease of Use

- II. A. 1. Naturalness**
- II. A. 2. Consistency**
- II. A. 3. Relevance**
- II. A. 4. Supportiveness**
- II. A. 5. Flexibility**

II. B. Concept Searching and Synonym Recognition

- II. B. 1. Thesauri**
- II. B. 2. Statistical and Numerical Approaches**
- II. B. 3. Juxtaposition**
- II. B. 4. Operators**

II. C. Result Ranking and Sorting

- II. C. 1. KnowledgeCite Library**
- II. C. 2. Publication time and date ranking**
- II. C. 3. Title and author ranking**
- II. C. 4. Northern Light**
- II. C. 5. AltaVista**
- II. C. 6. Google**
- II. C. 7. Excite**
- II. C. 8. Infoseek**

III. Search Engine Integration

- III. A. 1. Ex Libris USA**
- III. B. 2. Endeavor Information Systems**

Addendum A: DALNET iPAC Survey – Selected Functions & Features

Addendum B: Resources I – Compiled by DALNET Search Engine Task Force

Addendum C: Resources II – Compiled by Pat Ensor, University of Houston

Addendum D: Search Engine Task Force Charge

Addendum E: Market Requirements Document: Children's Public Access Catalog

I. Recommended Features

I. A. Ease of Use

- The interface should automatically remove duplicate items, that have been retrieved in multiple searches, from the search history,
- The interface results display should highlight keywords in titles and summaries
- The interface should provide the option to save and consult retrieved documents at any point during the search process
- The interface should provide the option to change the results display by selecting different visual configurations, prompts, toolbars, and by resizing
- The interface should provide the option to customize the display of search results by using different sort functions
- The interface should provide a zoom mode to enlarge the view of results
- The interface should highlight different types of query terms in different colors, depending on their source, e.g., thesaurus terms might be green, user-entered terms red, and terms extracted from documents blue
- The interface should provide the main options for a search session in an easy to use navigation bar, which de-emphasizes or disables functions that are inapplicable at a particular point in a search session

I. B. Workflow

- The interface should not allow a search to result in dead-ends, zero hits, or null (empty) results sets
- The interface should provide help messages and screens that support progressive refinement of each search, until a result set of items most closely related to the query terms is retrieved
- The interface should allow a combination of keyword and browse functions that increase the probability some information will be retrieved
- The interface should give the user clues about why a search resulted in so few (or no) items, suggesting alternative query methods rather than reporting no results
- The interface should indicate ambivalent queries to the user, for example, indicate a distinction between "whales" and "Wales"
- The interface should offer the user follow-up strategies to achieve successful information retrieval
- The interface should automatically correct or point out a query's syntactical requirements and errors
- The interface should maintain a transaction log which will help the user analyze unsuccessful queries
- The interface should support the following functions at all points in a session:
 - More: Expand a query
 - Extra Words: Add extra words to the query
 - Titles: View the titles list from the full document display
 - New: Enter a new search query

- Print, or Mail: Output full details of selected documents
- End: Exit search interface
- Help: On-line help option

I. C. Spell Check

- The interface should spell check against a minimum of 25,000 words per second
- The interface should provide multiple, simultaneous spell checks, i.e., it should accept input from many clients at the same time
- The interface should support stemming, i. e. searching a root word minus any prefixes and suffixes for alternate words or spellings, and display results as a new search set
- The interface should perform approximate searches, identifying and listing words with similar spellings
- The interface should provide Intelligent suggestions for misspelled words, using word lists to locate alternatives for typographical and phonetic errors, and display these options as a new search set
- The interface should utilize a word list, created by the library, which provides alternatives for misspelled words
- The interface should utilize a minimum 100,000 word, multilingual list that can be checked against American and British English alternative spellings
- The interface should perform multiple word-list checks, allowing the user to choose between a default list, a library created list, a legal list, a medical list and foreign language lists, etc.
- The interface should automatically sort suggested alternative terms in order of the most probable correct word, i.e., the word which most closely matches the original query term
- The interface should provide an option to detect and auto-correct: misspelled words using an on/off switch
- The interface should provide an option to detect and auto-correct misspelled words, with or without case-sensitivity, using an on/off switch
- The interface should provide an option to detect and auto-correct repeated or double-word errors using an on/off switch

I. D. Help

- The interface should provide a search wizard or advisor to help construct search terms and phrases
- The interface should provide definitions and examples of search terms, which describe the use of truncation, wildcard, Boolean, proximity, phrase searching, abbreviations, acronyms, broader terms and narrower terms, etc.
- The interface should provide definitions and examples of category or subject searching
- The interface should provide an option to search for grammatical articles, allowing retrieval of terms and phrases such as "Vitamin A" and "The Babe" by using an on/off switch

- The interface should use checkboxes to limit searches by date range, language, age group, geographic area, item format, location, and collection, etc.
- The interface should retain, and allow the user to select and display, results from different search history sets
- The interface should retain and permit the display of unlimited search history sets until user signs off
- The interface should allow the selection and display of a combined search history result set
- The interface should allow the selection and combination of previous search term sets for new query
- The interface should allow search term set modification, that is, the adding, deleting and moving terms between sets to, in order to create new search sets
- The interface should allow the user to save search history and results sets as files in folders and to send them as e-mail
- The interface should permit the display of successful searches, based on terms that are similar or related to an intended query, conducted by previous users
- The interface should provide the option to select canned searches which have been pre-defined by the library
- The interface should provide directional aids that provide links to special collections and services
- The interface should provide a list of error messages that include explanations and suggested corrections
- The interface should facilitate natural language (cgi-bin) query input from remote users, allowing them to request the assistance of a librarian during the search process
- The interface should provide a glossary of descriptions for all search terms

I. E. Pre-defined searches

- The interface should provide pre-defined searches, performed with hidden input fields that define search criteria (such as database field and value), so that specific results will be displayed when the search button is clicked. This could be used for natural language terms and common phrases not easily associated with Library Subject Headings. [See the Modular Gateway Interface (MGI) at: <http://www.pageplanetsoftware.com/mgiuserguide/index.html>]
- The interface should provide related, hierarchical, display tables for the user. For instance, if someone types in "monkeys" a hierarchical table would display a list of more specific, related subjects, i.e., gorillas, chimpanzees, baboons, etc.
- The interface should incorporate user-defined search mechanisms that allow them to refine their search until a desired answer is achieved
- The interface should incorporate limiters that reduce the number of retrieved items
- The interface should produce a list of words related to those entered by the user as soon as the query is initiated
- The interface should provide the user with an option to access a list of synonyms
- The interface should provide the user with an option to access word combinations generated by previous user requests

- The interface should utilize spider technology to provide three categories of search terms based on the initial query: primary keywords, related words, and forbidden words, which could be used to refine an advanced search
- The interface should incorporate focused crawler technology in order to seek pages that are relevant to pre-defined sets of topics
- The interface should incorporate a topical taxonomy, for both the Library of Congress and Dewey Subject Headings, to reflect common phrase requests and natural language requests
- The interface should display natural language terms and Library of Congress Subject Headings as taxonomies. [See for example Yahoo!, The Open Directory Project, The Virtual Library, or the Mining Company]
- The interface should utilize key phrase indexes to improve browsing capabilities. [Keyphind search engine technology automatically extracts key phrases that form the basic units of indexing and presentation, allowing users to interact with a collection at the level of topics and subjects rather than words and documents. Key phrases are returned instead of documents lists. See the article entitled, "Improving Browsing in Digital Libraries with Keyphrase Indexes," available at <http://www.cs.usask.ca/faculty/gutwin/1999/keyphind-journal/keyphind-12-final-TR.html>]
- The interface should make categories of multi-phrase terms, not simply word level indexing, available to users
- The interface should allow the user to save, access and retrace the path history of previously navigated topic tree hierarchies
- The interface should provide a set of related terms that users can add to their initial query terms for better results
- The interface should organize the classification schemes for end-users differently than those for the classifier, and provide more than one scheme for users to browse and navigate, both before and after retrieval
- The interface should organize information in knowledge classes, representing different branches or subject areas, so that terms selected from a knowledge class will provide the user with cross-references to semantically related terms
- The interface should index successful searches and store them in a text file so that users researching similar topics can access them
- The interface should institute topic trees that guide users deeper into a specific topic
- The interface should provide a relational table of natural language terms which allow a user to select related library subject headings
- The interface should create a searchable database of commonly phrased search terms
- The interface should provide a graphical display that maps relational terms, such as broader and narrower terms, which the user can select for subsequent queries

I. F. Concept Searching and Synonym Recognition

- The interface should provide a concept-search tool that helps users target their queries by defining their concepts [See Gavagai Technology, Inc, <http://www.gavagai.net>]

- The interface should provide a concept-search tool that clearly highlights and displays the parts of the document that embody the desired concept
- The interface should provide a flexible concept searcher that is easy to navigate, permits the user to define and save concepts, and can execute searches using those concepts along with relevant key words
- The interface should allow complex concept searches to be easily saved and rerun at a later time
- The interface should facilitate similarity searching, based on one or several feature representations and approximated evaluation techniques
- The interface should utilize semi-structured data models and respective query mechanisms
- The interface should facilitate information retrieval based on vague query formulation and uncertain document representations, by considering document structure and hyperlinks
- The interface should support relevance feedback and other interaction mechanisms that support implicit query reformulation, for both for text and multimedia document collections
- The interface should cluster documents based on content

I. G. Relevance

- The interface should support query expansion: terms should be extracted from records the user judges relevant to a particular query, added to the user's original query terms, and weights should be calculated according to a probabilistic model
- The interface should report the degree of match for retrieved citations, based a maximum possible weight for a record containing all the query terms.
- The interface should support query expansion, or the construction of a new search for similar documents, based on documents judged relevant by the user. [Words and phrases are extracted from the relevant documents, weighted, and added to the existing search terms to generate an expanded query]
- The system should support interactive query expansion, extracting words and phrases from relevant documents and presenting them to the user who decides which of them to include or exclude in a new search
- The interface should support automatic query expansion, extracting words and phrases from relevant documents and presenting the user with a list of document titles that highlights the words and phrases used to find them [In automatic query expansion the terms extracted from relevant documents are used to produce a new document set, but the user never sees the terms and has no control over how the expansion operates]
- The interface should support query enhancement by allowing the user to type in more words or phrases during the session
- The interface should support query enhancement by offering suggested terms, from thesauri or documents, which may be relevant to the query and potentially able to improve the success of the search
- The interface should support weighting of items according to the number of words they have in common with the query and the overall frequency of those words in the thesaurus as a whole. [The thesaurus match routine must also

handle lead-in terms, i.e., those that are not themselves used as document descriptors, but are may be synonyms of preferred terms that are descriptors]

- The interface should support a ranking threshold algorithm which provides document scores, i.e., that can partition the result list into documents that match the query ``well'', ``fairly well'', and ``not very well'', and to drop those which fall below a certain threshold
- The interface should allow the user to record a Yes or No relevance judgment on any document
- The interface should trigger a procedure to reweigh the original query terms, and/or the new extracted terms, for each positive relevance judgment made by the user

II. Design Principles & Examples

II. A. Ease of Use

Library Web sites are becoming information portals to a wide variety of resources. Wayne State University's Library Web site, for example, provides access to a number of electronic databases as well as a WebPac for locating library materials. The availability of different types of resources within DALNET institutions raises the issues of scope, consistency and compatibility for any search tools that may be introduced by *epixtech*. It is advisable that a unified interface be developed for all of electronic resources, so that DALNET users will have minimum cognitive load. There should be no need for DALNET users to learn how to use multiple search interfaces. The ease of use of an interface is a general system requirement that is normally based on user studies. A DALNET search interface should include the following features:

- The interface should achieve the correct balance between automatic and user control during a search session.
- The interface should promote users' ability to make subjective evaluations without hindering what the system does well, i.e. to process quantitative data in predefined ways.
- The interface should consider the characteristics of its potential users and their information-seeking tasks.
- The interface should use one main window with a number of subsidiary areas, whenever possible, rather than use overlapping windows.

Naturalness

Search methods should seem appropriate to the task being performed by particular users. Research in information science, for example, has found that users with social sciences and humanities background tend to employ an exploratory approach to seeking information while users with science and engineering background tend to perform direct searches. Allowing this latter group of users to enter search terms would seem appropriate, whereas, the availability of categories and directories that allow users to browse databases may seem more natural to the former.

Consistency

A system component should reinforce user expectations learned from previous interactions with it or with similar systems components; the user should not be expected to learn one method for one module of the system and another method for other modules. Ideally, all of the different search tools accessible via a library's Web site should have a unified interface.

Relevance

Similar to the concept of elegance that is applied to scientific theories, the best search and retrieval interface is one that requires the minimum of user input and system output necessary for the completion of the user's task.

Supportiveness

The user interface should provide adequate information for the user to operate the system and to perform the desired task. Specific instructions, guides or forms should be available to help users formulate input requests. There should also be adequate mechanisms to provide users with feedback about whether their queries are well formed or appropriate, as well as the status of the request as it is being processed.

Flexibility

The user interface should accommodate different user requirements, preferences, and levels of performance. The interface for the DALNET search engine should allow adults and young children to select from (or default to) custom interfaces since an integrated display of search results from the library WebPAC, multiple databases, and Web sites, may be appropriate for one group but not the other. The consortium model presents additional complexity to the question of how flexibility relates to scope: users with affiliations to multiple DALNET institutions may wish to conduct searches on all or only some of those institutions at various times. Users with affiliation to one DALNET institution may wish to retrieve results from all institutions or only its affiliate institution at various times. The ability to provide all of the above-mentioned options, through multiple user interfaces and/or selection options, is desirable.

II. B. Concept Searching and Synonym Recognition

Concept searching and synonym recognition differ from keyword searching, in that such search software attempts to retrieve relevant hits regardless of whether a user's exact search terms are present in descriptive metadata, assigned subject headings, or within the text of the material itself. The search mechanism, in other words, tries to search for what users *mean*, not necessarily what they have explicitly included in their queries.

Thesauri

Thesauri methods of concept searching offer the highest level of recall and precision, but require either the manual or automatic creation and maintenance of thesauri. The manual creation of thesauri necessitate a large amount of human effort and time, while the automatic creation of thesauri demands advanced technology, such as artificial intelligence, associative neural networks, statistical algorithms, and the like.

Concept search tools can use the same types of predefined thesauri as catalogers and indexers. According to the Center for Technology in Government (CTG), search software can read a text, transform and compare every word in it to its own internalized knowledge base, instantly cross-reference each recognized concept to all other related terms, and generate a result. The "internalized knowledge" can be established thesauri such as Library of Congress Subject Headings and the National Library of Medicine Medical Subject Headings, or thesauri developed by experts to be synonymous with various terms contained in the resources. The drawback to these approaches, however, is that large amounts of time and effort are needed to develop, maintain and update such thesauri.

Automated concept search tools can produce thesauri using complicated forms of artificial intelligence and statistical algorithms, with human beings assuring quality control. Such methods offer the greatest promise for future integrated searches across disparate types of resources and large collections. See, for example, "Automatic Subject Indexing Using an Associative Neural Network," by Yi-Ming Chung, William M. Pottenger, and Bruce R. Schatz (*Digital Libraries*, '98. Pittsburgh, PA: Association for Computing Machinery, c1998).

Statistical and Numerical Approaches

Some statistical methods do not require the indexing of established thesauri to do their work. According to CTG, such search software can generate a list of terms that are statistically related to those in the user's query. These words would have a "significant degree of co-occurrence" with the query terms. After generating the aforementioned list, the concept search operation performs a conventional search using the original query and the related terms. As you can imagine, one drawback to this type of statistical method is the routine delivery of false hits.

According to Pam Dyer, the Excite Web search engine uses a proprietary, numerical approach that tags search concepts in order to build searches. Excite determines meaning by calculating the frequency with which certain important words appear in source documents. When several words or phrases, tagged to signal a particular concept, appear close to each other in a text the search engine concludes by statistical analysis that the piece is "about" a certain subject. For example, the word "heart," when used in a medical/health context, would be likely to appear with such words as coronary, artery, lung, stroke, cholesterol, pump, blood, attack, and arteriosclerosis. If the word heart appears in a document with other words such as flowers, candy, love, passion, and valentine, a very different context is established, and the search engine returns hits on the subject of romance. This type of combined predefined-concept and statistical method of concept searching seems to work well, according to many sources.

Juxtaposition

An interesting approach to finding implicit related materials with searches is accomplished by the Arrowsmith software (<http://kiwi.uchicago.edu>), developed by the University of Chicago. Right now, the developers have focused on using it with Medline searches, but the process may be promising for other applications as well. First, the user creates two input files that result from two different searches in the database and submits these files (A and C) to Arrowsmith. Arrowsmith then creates a list of phrases and terms common to the two files, providing a source for intermediate linkages, B, between A and C. The user edits the B-List, and then when the B-List is resubmitted, Arrowsmith creates two subsets of Medline records, AB and BC, with both juxtaposed to help the user notice any possible A-C relationships.

Operators

In Primary Source Media's search engine, concept operators, such as (!) or (LIKE) can be applied to all or part of a query (<http://www.ddrs.psmmedia.com>). Whenever operators appear the operations are applied to the smallest discrete unit of query expression that immediately precedes it. For example, *Power!, corruption and lies!* will retrieve pages that include the following terms: *authority*, because it is a concept related to power, *corruption*, because it is a designated search term, and *falsity*, because it is a concept related to lies. Primary Source Media does not describe the mechanisms behind this proprietary search feature.

Primary Source Media's search engine also facilitates the application of fuzzy operators (~) to handle variant word forms or misspelled words to all or part of a user query (<http://www.ddrs.psmmedia.com>). Wherever the operator appears," according to the instructions, "the program will search for words that are spelled like that word. For example: *~color* will retrieve works with *colour* and *colours*." The same type of operator could be used to search for related or synonymous words, but this would again require either the creation of thesauri or some other automatic means with which the system would be able to index the synonyms. Some systems accomplish this by, again, generating a list of terms that are statistically related to the words in the query, i.e. words that have a significant degree of co-occurrence with the query terms.

II. C. Result Ranking and Sorting

Ranking and sorting functions, which organize the results of a search, have been popular practices in commercial information retrieval systems and Web search engines. Ranking and sorting of records in the search result set is based upon a record's relevance to a search query. Relevance ranking is necessary for keyword or subject searches and should be set as their default ranking and sorting method. There are several relevance ranking methods and when choosing one for the DALNET search engine, the features of DALNET databases, i.e., bibliographic databases, should be considered. Ranking results by publication time, author, and title is relatively simple, whereas, relevance ranking involves complicated information retrieval techniques and algorithms. Based on the current Web search engine practice, the following ranking and sorting functions are suggested for the DALNET search engine.

KnowledgeCite Library from Silver Platter, <http://www.knowledgocit.com>, is an example of a product that provides relevance ranking on a variety of criteria, including: 1) the location of a search term within a document or record, 2) the proximity of search terms, 3) the age of the document, 4) whether the document includes all, rather than some, of the search terms, 5) the frequency of appearance of the search terms within the document, and 6) a term weight assigned to a search term based on its uniqueness with respect to the collection or database as a whole.

Publication time and date ranking

Rank or sort the records in the search result set in reverse chronological order by publication time: by year for books, by year and month of issue for serials, and by date for Web documents. These ranking methods should be made available as an option in all search modes, i.e., allow users to rank results this way whether they have conducted a title, author, keyword, or subject search. Northern Light and Infoseek are examples of Web search engines that sort by date.

Title and author ranking

Titles sorts organize the records in the search result set in ascending alphabetical order by title. This ranking method should be the default for title searches and should also be made available as an option for all other search modes. Author sorts organize the records in the search result set in ascending alphabetical order by authors' names, with last name first. Author ranking should be the default ranking method for author searches and should be available for all other search modes. A search engine that allow users sort by title or author can be found at:
<http://wgate.bibsys.no/search/pub?lang=E>

Northern Light [Reviewed by Greg R. Notess]

In general, hits are sorted in order of relevance. Of far more importance is Northern Light's Custom Search Folders that organize the full set of search results into subject, source, document type, or language folders. The subject folders use subject terms from a hierarchy developed by librarians, however; they appear to be automatically assigned. Folders are subdivided until less than 26 records are contained in an individual folder. When using the Power Search function, results can be sorted by date, in reverse chronological order, as well as by relevance.

AltaVista [Reviewed by Greg R. Notess]

If the query of a Simple Search matches a question in the Ask Jeeves database, the answers are either displayed at the top or bottom of the results list depending on the query. If the query matches one or more records in the RealNames Internet Keyword database, these results are displayed next. Web database results are displayed according to AltaVista's relevance ranking formula, based on the location of terms in a document (such as in the title), the proximity of multiple search terms to each other, and the frequency of the terms. Results are clustered by site, based on the exact host name, so that only one record per site appears on the main results page. In other

words, "name.com" and "www.name.com" would be two separate clusters even though they belong to the same domain. Users can see additional hits from a specific site by selecting the option to display "More pages from this site". Open Directory categories may be listed at the bottom under the option to "Find What You're Looking for in".

The results of an Advanced Search are not sorted unless one or more terms are included in the ranking keywords box. There is no apparent order to the list when the ranking keyword box is empty, although the order may remain consistent on subsequent searches. Results from the Advanced Search are clustered by site, based on the exact host name, if this option is selected in a checkbox.

Google

Results are sorted by relevance using PageRank, which assesses the link structure of documents containing the required search terms. Google also uses text-matching techniques to analyze how similar linked pages are to those containing all search terms. A special caching feature saves a copy of each page that meets these criteria, so, if a link you need no longer exists on the Web, "Show Matches" will display the last saved version of the page. Pages are grouped by site, designated by indenting the subsequent hits from the same site, which is a form of clustering.

Excite

By default, results retrieved by this robot-based index are grouped by relevance score or site. Excite has an option to sort the top 40 hits by site but there is no option to sort results alphabetically or by date. Concept searching and keyword searching are performed with a best match method. Boolean operators, nesting, and phrase searching features are available and the Power Search feature offers simplified Boolean searching with a "form". The Search Wizard presents related terms.

Infoseek

Infoseek provides a best match system with relevance ranking, Boolean operators, and phrase searching. Results can be narrowed by one command and in subsequent steps. Some automatic query expansion features are provided. The following fields may be queried: page URL, host name, page title, and linking/citing pages. Metadata in a certain format indexed and displayed. The default sort for Infoseek is by relevance score and by site. Relevance is determined by the location of the search terms, i.e., those highest in the document, the frequency of the words within the document, and the unusualness of the search words, i.e., uncommon words are rated higher. Link analysis is also used for relevance ranking. All pages from a specific site are grouped together, but clicking on the "Ungroup Results" option will re-sort them by relevance. The grouped results can also be sorted by date.

III. Search Engine Integration: Library Vendor Products

III. A. Ex Libris USA: <http://www.exlibris-usa.com>

Aleph 500

Aleph 500 is a Web interface that can be modified for unique groups of End Users who can enter the system as a Guest or login to a customized profile. End Users may customize the screen and record formats for display of query results. The product help screens, menus and messages support 17 languages. Searching on a wide range of configurable search browse indexes, End User results can be limited by date, range, language and other filters. Results sets can be combined using a variety of operators, and a scope feature allows the results of one set to be used as the basis for future searches.

MetaLib <http://www.aleph.co.il/metalib/index.html>

MetaLib translates an End User's profile and query into a parallel broadcast search of distributed catalogs and databases and locates pertinent documents including links to full-text documents. A Resource Store contains a list of databases and resources, including collection descriptions and scope, which serve as a prospectus for End Users. MetaLib Supports MARC and non-MARC formats as well as Z39.50, http, and various collection type protocols (for catalogs, databases and archives). User registration, licensing and copyright control guarantee the integrity of institutional policies, while document delivery services provide links to local and remote databases in which desired resources are stored. Query results are merged and de-duplicated to reduce redundancy of information presented to users, and context-sensitive links to services, databases and local holdings are provided. The user interface, language, display and search environment as well as retention of search results can be customized to create a personal user profile. This profile can be linked to an E-Basket in which the End User can store a personal catalog selected documents. MetaLib accepts queries in multiple languages.

SFX <http://www.aleph.co.il/sfx/index.html>

SFX, or Context Sensitive Reference Linking, allows selection of links to services using a wide variety of formats and protocols. Based on the OpenURL specification, <http://216.229.137.107/OpenURL/openurl.html>, and installed on a local or Ex Libris hosted server, SFX contains links to resources and services and provides uniform, vendor-independent measures of use. End Users navigate through institutionally defined electronic collections implemented through customized hooks to specific databases. Note: Ex Libris is listed as a client of Index Data, <http://www.indexdata.dk>, a consulting and software-development enterprise that specializes in solutions for distributed information management using open, vendor-independent information retrieval standards which help groups of people or organizations to share their information as seamlessly as possible. Their Links page, <http://www.indexdata.dk/links>, includes references to Z39.50 Information, Metadata (including Dublin Core, the Basic Semantic Registry and the Global Information Locator), and Record Syntaxes (including USMARC, UKMARC and UNIMARC authorities), and the W3C's Extensible Markup Language (XML).

III. B. Endeavor Information Systems: <http://www.endinfosys.com>

Citation Server

Citation Server facilitates the parallel searching of multiple local and remote (Internet) databases containing citations, abstracts and full-text, with links to local acquisitions, serials and circulation information. The product supports access through the Z39.50 protocol, using MARC 856 Tags for linking to URL's and helper applications for multimedia display. Databases can be grouped according to their subject, or the proximity and cooperative arrangements of institutions. End Users are informed if the library holds the desired resource and whether the user can be authenticated for access it in local databases, or through a remote library or commercial database. Libraries have a great degree of control over the display of item records, including which fields and sub fields appear, the display order, field labels and spacing.

enCompass

enCompass generates full-text indexing, as well as easy entry of administrative metadata and the automated generation of descriptive metadata, for digital resources. It features simultaneous retrieval of MARC and XML metadata resources (EAD, Dublin Core, TEI, etc.) that can be collected and managed in multiple repositories. Resources are organized in a hierarchical series of Collections, Repositories, Containers, and Objects, along with the corresponding metadata for each. Full-text and metadata for resources are accessible to staff through a keyword-, phrase-, and qualified keyword-indexes, and object identifiers. End Users access resources by means of a single user-interface utilizing two modes of keyword and context-based searching: Direct Searching returns container and object listings for relevant matches available in user-selected collections, while Resource Discovery allows users to browse the metadata of all collections, returning a hierarchical display of collections and containers. End Users may execute a Direct Search at any point during the Resource Discovery process.

Addendum A: DALNET iPAC Survey - Selected Functions & Features

General Features

- 10 User must be able to select from a list of available databases or groups of databases.
- 10 Help must be context-sensitive and available at any point in the search process.
- 9 The system must provide users access to the bibliographic records of all consortium member libraries as well the user's host library.
- 8 The system must allow movement through the entire results of a search history, allowing the exploration of alternative paths at any step of the process without retracing the search.
- 7 The user will be able to create a personalized profile for authenticated tasks.
- 5 The user interface must support multiple languages with language sort capability

Indexes

- 10 The system will support adjacency and proximity searching.
- 10 Browse functionality is supported.
- 9 The system will support complex nested searches.
- 9 The system must allow the use of library-specified stop words taken from the library system database when desired.
- 9 The system must allow flexible sorting capabilities.

- 8 The system shall allow patrons to search across different indexes, e.g., author and subject.
- 8 Examples, wizards or help for standard search options will be available.
- 6 The system must allow for a library-specified synonym list.

General Searching

- 10 The user can cancel a search at any time during the search process.
- 10 The system will allow use of Boolean operations, "and", "not" and "or".
- 9 The system will support redirection of searches through hyperlinks.
- 9 The system will provide easy to follow prompts throughout the query process.
- 9 The system shall facilitate narrowing the scope of a search narrowing by language, date, and fixed fields, etc.
- 9 The system must permit browsing of the catalog, including the authority file
- 9 The system must enable patrons to optionally search the resources of a specific default library, or group of libraries, as well as those of the entire consortium
- 9 The system must automatically include "see" and "see also" references in the alphabetical listing of the authority.
- 9 The system must allow the user to apply an existing search string to additional targets, i.e. databases, repositories or Web sites, without re-keying the string.
- 9 Every search will produce a result: a default of "no records found" would be returned for a null search.
- 8 The system will provide feedback about the status and percent completed of various processes such as searching, sorting and limiting.
- 8 The system will display the index entries that most closely resemble the terms entered, i.e., based on the best available form of a query if confronted, for example, with a misspelled word or a term that does not exist in an index.
- 8 The system will allow implicit Boolean searches, i.e., the "and" operator will be implicit for word combinations.
- 7 The system will allow use of multiple nested operators.
- 7 Relevancy shall be determined by a free text assessment of search terms.
- 6 The system will retain search terms in a query until the user enters another set of terms.

Search Results

- 10 Single hit sets defaults to the full view.
- 10 Results lists may be sorted by location, i.e., my location first.
- 9 The system will display the library, location, current location, collection, item call number, item status, and item barcode for each item in the results set.
- 9 The full view provides a complete display of information available for individual records.
- 9 Searches may be broadened by truncation.
- 9 Full view will show copy status for items.
- 8 The results list will allow display of 856 tags.
- 8 Searches may be broadened by adding the plural form of terms.
- 8 Records may be displayed in a labeled format.
- 7 The system will display the circulation status of items in the results set.
- 7 The brief view will support the display of multiple records.
- 7 Display of an authority browse may be collapsed to show broader categories.
- 7 A child window will be created if a hyperlink resource, such as an 856 tag, is selected from a results list.
- 6 The user will be able to sort, limit, and add search terms from the brief view.
- 6 The user will be able to change filters from the brief view.

Navigation

- 9 The user may easily navigate to the first and last records of a results set.
- 9 The system will support navigation by means of mouse.
- 9 The system will support navigation by means of menus.
- 9 The next title command will be available in the full view of results sets.
- 8 The user may easily select multiple records with a mouse or the keyboard.
- 8 The system will support navigation by means of icons representing frequently used functions.
- 6 The system will support navigation by means of icons representing general categories and subjects.
- 6 The system will support multiple simultaneous results sets.

Session Control

- 10 Users may end the search session, and erase search history, by shutting down the browser.
- 9 Search history may be cleared by selecting a clear button, or by the workstation timeout
- 7 The search interface will be integrated with the user authentication scheme used by the library and the consortium.

Search History

- 9 Two or more search strings can be combined and edited for a new search.
- 8 The system will save all search strings for the duration of the search session.
- 7 Searches can be saved to a file by default or by user selection.
- 5 An entire search history will persist across databases until cleared

Patron Empowerment

- 9 Users can create booklists, save them to a file, print and e-mail them.

Printing

- 10 Results sets for brief records may be printed, saved to disk or clipboard, and e-mailed.
- 10 Results sets for full records may be printed, saved to disk or clipboard, and e-mailed.

Electronic Dictionary

- 9 A server-based thesaurus will be available to users.
- 8 A server-based dictionary will be available to users.
- 8 Misspelled search terms will be underlined

Broadcast Searching

- 10 The system will facilitate searching across multiple types of databases.
- 10 Detailed descriptions of all databases will be available.
- 9 The system will be able to display the number of hits found in each database or system.
- 9 The system will be able to display an integrated, de-duplicated list of hits from all databases or systems.
- 8 The system will search any Z39.50 compliant server.
- 8 The system will provide hooks to holdings records.
- 8 The priority order of databases to be searched will be controlled locally, i.e., at the level of the user interface.
- 7 The system will support Z39.50 scan or browse function.
- 7 The system will support Z39.50 access to non-MARC record formats.

System Administration

- 10 Each member library's user interface must be customizable using simple configuration tools or standard web page scripting languages.
- 8 Individual libraries will be able to define search menus and prompts.
- 8 The library will be able to predefine sorts and limits.
- 7 Search types may be defined separately for each database.

Survey Respondents:

Deborah Adams, Botsford General Hospital
Duryea Callaway, Wayne State University
Robert D. Chapman, Detroit Public Library
Mary Howarth, Oakland Community College
Adriene Lim, Wayne State University
Rhonda J. McGuinnis, Wayne State University
Vanessa Middleton, Wayne State University
Leo Papa, University of Detroit Mercy
Donna Roe, University of Detroit Mercy
Karen M. Turbolino, VA Medical Center

Addendum B: Resources I – Compiled by DALNET Search Engine Task Force

Browsing and Searching the Internet
http://www.lub.lu.se/netlab/documents/nav_menu.html

Comparison Search
<http://coverage.cnet.com/Content/Features/Dlife/Search/ss01.html>

DMOZ Open Directory Project
http://dmoz.org/Computers/Internet/WWW/Searching_the_Web/Search_Engines/

Gregg Notess
<http://www.notess.com>

Kansas City Public Library: Introduction to Search Engines
<http://www.kcpl.lib.mo.us/search/srchengines.htm>

Search Engine Watch
<http://www.searchenginewatch.com/>
<http://www.searchenginewatch.com/resources/reviews.html> Search Engine Reviews

ADAM: Art, Design, Architecture & Media Information Gateway
<http://adam.ac.uk/>

A searchable catalogue of 2546 Internet resources, carefully selected and catalogued by professional librarians for the benefit of the UK Higher Education community.

Huwe, Terence K. "New Search Tools for Multidisciplinary Digital Libraries." ONLINE, vol. 23, no. 2, March 1999. <http://www.onlineinc.com/onlinemag/OL1999/huwe3.html>

[T]he products described in this article offer new strategies for both the super users of the library and research communities and for casual users. Each product has a slightly different focus, both in target audiences and suites of features, but all tackle the challenge of merging distinct knowledge bases using new search tools. The first example, the Knowledge Cite Library, is also the most ambitious. Database Advisor, a simple (but elegant) finding aid, is ideal as a low-cost solution. Pharos, <http://uias.calstate.edu>, is the California State University's custom-designed, system-wide information utility that weaves multidisciplinary finding aids throughout its interface. Northern Light, a popular new entry in both the Internet and commercial markets, offers a single solution to the serious researcher, combining a series of retrieval processes.

Kelly, Brian. WebWatch: UK University Search Engines. Ariadne, iss. 21, 20-Sep-1999.
<http://www.ariadne.ac.uk/issue21/>

An analysis of search engines used by University and Colleges web sites as given in the HESA list [3] was carried out during the period 16 July - 24 August 1999. Information was obtained for a total of 160 web sites.

Koch, Traugott. Browsing and Searching Internet Resources.
http://www.lub.lu.se/netlab/documents/nav_menu.html

Which search service to start with? General recommendations. WWW, including multi-protocol services. Combined and collected search engines. Detailed description of important search services. Typology. Robot-based WWW Catalogs. Indexing full document content. Regional services. Subject based, robot generated search services. Indexing various elements from the documents. List- and template-based WWW Catalogs. Fee-based search services. Indexes for individual types of Web resources and/or other protocols. Gopher (only). Library catalogs (OPACs). Search for OPACs and connect. Search records from OPACS through HTML Forms
BROWSING: Internet resources by type, location, subject, and subject service, chronologically, and at random.

Koch, Traugott. Internet search services. Translation of a paper given at: "Further on the way to the virtual library! Libraries are using the Internet," Erste INETBIB-Tagung in der UB Dortmund, 11.3, 1996. <http://www.lub.lu.se/tk/demos/DO9603-meng.html>

In order to evaluate the different search services it is necessary to organize them into a typology or taxonomy. The possibilities and weaknesses will be briefly compared and described for the most important services. Still more important are the weaknesses and problems common for all attempts to index the Internet. The problems of the quality of the input, the technical performance and the general problem of indexing virtual hypertext are shown to be at least as difficult as the different aspects of harvesting, indexing and information retrieval. To conclude, some of the developments made in the area of retrieval services will be mentioned, in relation to descriptions of the contents of documents and standardization efforts.

Search Engine Section. ONLINE, vol. 23, no 3, May 1999.
<http://www.onlineinc.com/onlinemag/OLtocs/OLtocmay4.html>

- Search Engines in the Internet Age. Greg R. Notess
- Web Search Engines: Features and Commands. Randolph Hock .
- Results Ranking in Web Search Engines. Martin P. Courtois and Michael W. Berry.
- Choosing an Intranet Search Engine. Darlene Fichter
- The Future of Web Search. Chris Sherman
- NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval. Susan Feldman

- Meta Search Engines. Nancy Garman
- A Web Search Trifecta: Keeping Tabs on Search Engine Features and Technology. Bill Mickey
- On the Net - Rising Relevance in Search Engines. Greg R. Notess

Search Engine Showdown

<http://searchengineshowdown.com/>

Features of the Web Search Engines. Statistics. Search Strategies. Reviews. News. Usenet. Phone Numbers. Subject Directories. Email Lists. Multiple Search Engines. Links and Readings.

Search Engine Watch

<http://searchenginewatch.com/>

Search Engine Submission Tips. Web Searching Tips. Search Engine Listings. Reviews, Ratings & Tests. Search Engine Resources. Search Engine News.

Semio

<http://www.semio.com/>

Mentioned in SLA "Portals" videoconference, Semio's suite of information categorization solutions: automatically extracts key concepts from large volumes of text-based content; structures information into logical categories, drawing connections between concepts, and provides user-friendly access to the structured information.

Copernic 2000 <http://www.copernic.com/>

Addendum C: Resources II – Compiled by Pat Ensor, University of Houston

An Expert Web Searcher

Compiled by Pat Ensor Ensor@dt.uh.edu

Director of Library Services, University of Houston-Downtown

Subject Guides

About.com <http://www.about.com/> Human guides select, organize, and annotate sources, and it shows. A great place to start with a broad topic.

Open Directory <http://www.dmoz.org/> Another human-edited subject resource with well-chosen sources. The other place I cover for a broad topic. Also searched in Google <http://www.google.com>

Alta Vista Directory <http://www.altavista.com> Choose a category in order to search only the directory. Combines LookSmart and Open Directory.

Yahoo <http://www.yahoo.com/> I use for news, looking up organizations, and financial profiles.

Search Engines

Northern Light <http://www.northernlight.com/> Big, sophisticated, good organization of results, rarely covered in metasearch engines.

Google <http://www.google.com/> Big, good relevance ranking, special features.

Fast SEARCH <http://www.alltheweb.com> Big, quick.

Alta Vista <http://www.altavista.com> For searches where a proximity operator is needed.

Metasearch Engines and Search Engine Collections

Ixquick <http://www.ixquick.com> The best metasearch engine I've found – translates your search syntax, doesn't run them on engines that can't accept them, eliminates duplicates, assigns stars for number of appearances in top ten. Fast SEARCH, Inktomi, Alta Vista included.

MetaCrawler <http://www.metacrawler.com> Pretty good relevance ranking. Google and Alta Vista included.

Search.com <http://www.search.com> Strength is in the number of sites you can search at once and the metasearch subsets. No Fast SEARCH, but Inktomi and Alta Vista.

Fossick.com WebSearch Alliance <http://fossick.com> Chooses engines to metasearch dynamically, sometimes including Fast Search, Northern Light, Alta Vista, and Inktomi. Great collection of specialty sites. Can do country-specific searches.

Multimedia Searching

Alta Vista <http://www.altavista.com>

Lycos Richmedia <http://richmedia.lycos.com>

Fossick Online Multimedia and Digital Image Search <http://fossick.com/Multimedia.htm> Lots more multimedia sites.

The Invisible Web

For a good definition of the Invisible Web, go to the Help function of InvisibleWeb.com.

AlphaSearch – Gateway to the Academic Web <http://www.calvin.edu/library/searreso/internet/as/>

Direct Search – Search Tools and Directories <http://qwis2.circ.qwu.edu/~gprice/direct.htm>

INFOMINE: Scholarly Internet Resource Collections <http://infomine.ucr.edu/search.phtml>

InvisibleWeb.com <http://www.invisibleweb.com/>

Lycos Searchable Databases http://dir.lycos.com/Reference/Searchable_Databases/

WebData.com <http://www.webdata.com/webdata.htm>

The Big Hub <http://www.thebighub.com>

Search Engine News

About.com's Web Search <http://websearch.about.com/internet/websearch/mbody.htm>

ResearchBuzz! Great Internet Research News and Information <http://www.researchbuzz.com/>

Search Engine Guide: Search Engine News
<http://www.searchengineguide.com/searchengineneeds.html>

Search Engine Showdown <http://www.searchengineshowdown.com/>

Search Engine Watch <http://www.searchenginewatch.com/>

Addendum D: Search Engine Task Force Charge
April 18, 2000

Search Engine Task Force

The Search Engine Task Force will be responsible for working with *epixtech* to deliver the search engine enhancements in the DALNET contract. The task force will work in conjunction with the Children's OPAC Task Force to improve the web-based searching of Horizon. The Task Force will look at web search engines, library catalogs, and other relevant products to identify desirable features and develop functional specification for inclusion in the DALNET search engine. These specifications should be complete by August 30, 2000. The Task Force will work with *epixtech* to establish a time-line and to review developments. The Task Force will report to the DALNET Steering Committee through the DALNET Project Manager who will also coordinate the work of the two Task Forces.

Primary focus should go to the DALNET specified enhancements in the contract that requires. These functions are to be incorporated into the DALNET web-based search engine.

- Concept-searching
- Spell-checker
- Result set ranking
- Pre-defined searches

The task force may also consider additional functionality suggested by *epixtech* - after the completion on the contractual specifications. This second phase can more fully address the needs of the DALNET Information Hub incorporating the searching of Horizon catalogs, DALNET bibliographic databases, DALNET metadata elements, and the World Wide Web. This list includes, but is not, limited to:

- Synonym recognition
- No Dead End workflow
- High ease of use
- Online help

Addendum E: Market Requirements Document: Children's Public Access Catalog
Riccardo Ferrante, Product Manager, June 16, 1999

Overview

The purpose of this document is to outline in general the key functionality and user interface (UI) features that would be considered essential for a marketable PAC product designed specifically for use by children ages 7 and higher.

Market Requirements

For its initial consideration, the children's PAC (KPAC) will be targeted for distribution in North America, specifically the United States and Canada. The potential for disparity between metaphors easily understood in the USA and other English speaking countries not in direct proximity, and subsequently non-English speaking countries, suggests a secondary analysis of the initial product metaphors prior to any expansion of distribution beyond the USA and Canada.

Functionality

Concept-searching. The PAC will use subject headings to identify concepts being searched. If a 'no records found' result occurs, the pac will suggest a different topic based on the subjects recently searched.

Spell-checker. Based on the language settings of the Windows terminal in use, the pac underlines misspelled search terms, using a convention identical or at least similar to MS Word. Misspelled search terms must be acceptable to the pac.

Synonym recognition. Pac should recognize synonyms for the search term and present the option to review a list of synonyms to the end-user after the first search is created.

No Dead End workflow. Every search will produce a successful result, including the 'no records found' screen. Users should have the option to modify an unsuccessful search or send the search terms to their web search engine of choice.

Results set ranking. Relevancy shall be determined by a free text assessment of the search terms.

Key UI Features

High ease of use. A very high value is placed on the ability of an end-user to approach the PAC for the first time with success after very little instruction, if any.

Pre-defined searches. The UI should present a series of pre-defined searches clearly marked. Library staff should be able to define, modify, and/or delete the searches.

Large font presentation, smaller font option. On launch or reset, this pac should use a 14-16 pitch font. The font setting should be capable of being decreased or enlarged easily by the average end-user.

Icon driven presentation. For the youngest segment of end-users, the icon metaphors should communicate clearly the major points of each page.

Online help. Online help screens should be easy to access and available for all major areas of the PAC.

Constraints & Dependencies

Analysis is required to provide the necessary information in this section.