

Search and Browse Interface in Digital Library

I. Objectives

The interface of a Digital Library plays a vital role in discovering its content. Browse and search interface connects a user to digital libraries. As such success of a digital library lies with the design, search and browse functionality of the digital library. The objective of this module is to discuss the following aspects of interface designing in digital library with examples:

- Interface designing of Digital Library, some basic principles;
- Searching functionalities provided in varieties of digital libraries and common search options;
- Browsing options provided in digital library with different architecture;
- Search results representations.

II. Learning Outcomes

After going through this lesson, the learner would gain knowledge about the basic principles and characteristics of interface designing, common search and browse options provided by varieties of digital libraries. In addition to basic, advanced and expert search options, learner would also be introduced to the concepts of faceted and meta search options. The Learner would also be imparted knowledge about the representation of search results that are retrieved using search and browse options.

III. Structure

1. Introduction
2. Layout & Designing
 - 2.1. Content overview
 - 2.2. Search Interface
 - 2.2.1. Simple Search/Basic Search
 - 2.2.2. Advanced Search
 - 2.2.3. Expert Search
 - 2.2.4. Faceted Search
 - 2.2.5. MetaSearch
 - 2.3. Browse Interface
 - 2.4. Display of Retrieved information
 - 2.5. Specialized Digital Libraries (Video, Audio, Images etc.)
3. Summary

1. Introduction

In case of traditional libraries, the role of a librarian is to enrich and organize the library collections so as to enable users to locate individual items easily. Likewise, the primary aim of a digital library is to guide its users to quickly identify the most reliable and suitable digital items whether they are stored in its own collection physically or to provide access to content from other digital libraries located at a remote location. In traditional library, the users search for their desired information sources by searching the physical catalogue or the online catalogue (OPAC) in case of automated libraries and locate them on shelves. But in case of digital libraries, the digital items may be available on a number of servers distributed on different physical locations. The interface of a digital library is the sole medium which directly interacts with the users for discovering the collection and content stored in a digital library and catering the information needs of the users.

As Galitz (2002) defines it, "the user interface is the part of a computer and its software that people can see, hear, touch, talk to, or otherwise understand or direct." Interface designs are guided by an assessment of user needs and aimed to maximize interaction with primary resources and support both browsing and analytical search strategies.

2. Layout & Designing

Elaborating a conceptual model for interface design Arms(0)ⁱ emphasizes that interface design encompasses what appears on the screen and how the user manipulates it; among its aspects are fonts, colors, logos, keyboard controls, menus, and buttons. Functionally design specifies the functions that are offered to the user. Typical functions include selecting parts of a digital object, searching a list or sorting results, obtaining help, and manipulating objects that have been rendered on the screen. These functions are made possible by the data and metadata that are provided by the digital library, and by the underlying computer systems and networks. Conceptual model of interface design, therefore consists of the following components:

Conceptual model	Interface design
	Functional design
	Data and metadata
	Computer systems and networks

Dillon (2002)ⁱⁱ lists five questions that designers of digital library interfaces should be addressing:

- How do we attract users to our resources, and make them stay?
- What will bring a user back to our resources again?
- How do I build an interface that supports a richer comprehension or appreciation of the contents?
- What makes the material more learnable by users?
- Can novices learn from viewing an expert's construction of an information space?

Since user interface for a digital library must display large volumes of data most effectively and efficiently, a user of a digital library should be presented with a:

- One or more overlapping windows that can be resized and rearranged.
- Intuitive interface to query and retrieve large amount of data spread through a number of resources
- Ability to change the user's perspective from high-level summarized information down to a specific paragraph of a document.

Tedd (2005)ⁱⁱⁱ opines that the interfaces were designed according to the principles that users should maximize their interactions with information resources and minimize their attention to the system itself, and that both browsing and search strategies should be supported for effective and efficient use of digital content available in a digital library.

Galitz (2002)^{iv} formulated the following sets of design principles that he argued should be applied to the design of a user's interface:

- Interface design should be aesthetically pleasing and attractive to the eye, as interactions primarily are in the visual realm.
- Visually, conceptually and linguistically clear and unambiguous.
- Compatible with the users and the tasks to be accomplished. Moreover, it should be compatible with earlier versions of the system, or any other similar kinds of systems (in theory, this would mean that all digital library interfaces would follow a standard design).
- Comprehensible, that is, easily learned and understood.
- Configurable, that is, easy to personalize configure and re-configure.
- Consistent in the sense of look, feel and execution; the same action should always give the same result.
- Controllable by the user, so that actions result from explicit user requests, are performed quickly, and are interruptible; the user should feel that he/she is in charge of the interface.
- Direct in the ways in which tasks are accomplished; the effect of actions on objects should be visible.
- Efficient, by minimizing eye and hand movements.
- Familiar, by using concepts and language that users should know, using real-world metaphors, and building upon users' existing knowledge.
- Flexible to the differing needs of users (in terms of their knowledge and skills, experience, personal preferences, and habits).
- Forgiving of common and unavoidable human errors; preventing errors whenever possible; and providing constructive messages in case of errors.
- Predictable on the part of users who should be able to anticipate the natural progression of each task.
- Recoverable by allowing reversible actions.
- Responsive to user requests, with visual, textual or auditory acknowledgement.
- Simple.

- Transparent, so that the workings inside the computer or database remain invisible to users.

Galitz emphasizes that although these principles taken together represent the design ideal, in practice trade-offs will be required between some of the individual principles. The desire to maintain compatibility with earlier versions of the interface, for example, may clash with the desire better to meet many of the other principles; efficiency may clash with flexibility, and so on.

In a nutshell, Galitz (2002) emphasizes, "the best interface is one that is not noticed, one that permits the user to focus on the information and task at hand, not the mechanisms used to present the information and perform the task." The best digital library interfaces, then, are not the ones that on first encounter impress users with the most vivid colours, the most attention-grabbing icons, or the most intricate screen layout; rather, they are those that unobtrusively allow users, no matter what are their personal characteristics or their task in hand, to find what they are seeking quickly, accurately and with the least effort.

Sastry & Reddy (2009)^v proposed the following principles for user interface design for digital libraries for an effective user interaction and implementation:

- Simple : The digital library user interface should be simple and straightforward so that the basic functions are easily noticeable to the users.
- Support: The digital library user interface should provide users to control over the DL; it has to enable the users to accomplish tasks using any sequence of steps that they would naturally use. It should be more on event driven rather than menu driven.
- Familiar: The user interface of digital library should be familiar to its users, i.e, the users should not require special training to perform any task.
- Informative Feedback: The user interface of digital library must provide informative feedbacks to its user during various tasks performed by the user.
- Design Dialogues to Yield Closure: Informative feedback at the completion of a series of actions.
- Prevent Errors: The user interface should not allow users to make any serious errors. Alternately the system can be designed as insensitive to errors. It should detect the user errors and offer simple, constructive and specific instruction.
- Multimedia Support: User interface of digital library should support multimedia information.
- Profile Based Support: The digital libraries may support profile creation and provide customized services based on user preferences.
- Lithe and Simple: The user interface of the digital library should be lithe and simple without having heavy and unnecessary graphics which may slow the loading of content and create disinterest among users.
- Pan and Zoom Support: It should support the basic Pan and Zoom features.
- Accuracy: It should provide as much accurate information as possible as a poor display of information, spelling errors and grammatical errors may affect the credibility of the digital library

- Efficient Searching with NLP support: Digital Library should provide efficient search mechanisms with the excellent search interface. It should provide natural language analysis and processing techniques for effective and user-friendly searching.
- Support of semantic approach and Resource Description Frame Work (RDF) Technologies
- Sharing and Reusing of Information: It should support sharing of the content by various mechanisms.
- Multilingual Support: It should provide multilingual support for searching and displaying regional content.
- Platform Independent: The user interface should be platform independent and work effectively in all environments.
- Future Plug-ins Support: It should support plug-ins for future developments and interacting with other systems.

Digital library generally enables users to provide some common actions for discovering the content and material stored in the archive. The following is a brief of the search and browse functionality available in Digital libraries.

2.1. Content Overview

It is a common practice to provide a brief overview of the content and collection with coverage of the materials available in the digital library. It can be simple text base or can be represented with visual animations and pictures.

2.2. Search Interface

The objective of any search is to encapsulate a user's information need in one or several words - the query - and display the resulting matched items.^{vi} Most digital libraries provide a simple search box as available in any web-based search engine in which users can input their search query or keywords.

2.2.1. Simple Search/Basic Search

Simple Search, often called "Basic Search" is the most common feature available in any digital library which is the most preferred entry point for discovering its content. An example is shown in Fig 1, from the JSTOR Archive, an archive of various types of content from journal articles to books, images, etc. It gives a simple search box for making any free text search query which searches through all metadata fields including full-text content.

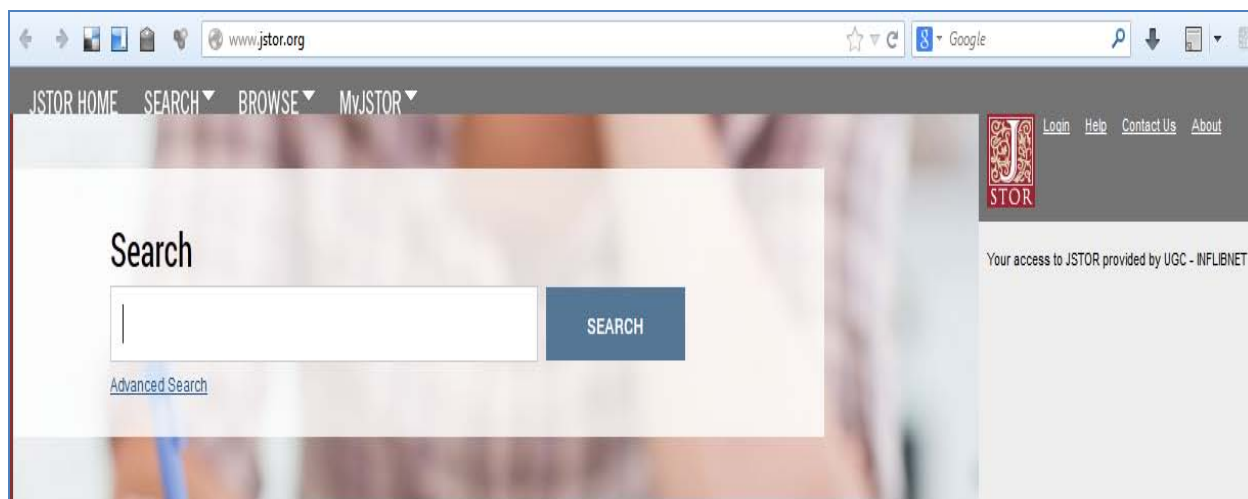


Figure 1 Simple Search box of JSTOR Archive

Some digital library provides simple search functionality along with options to restrict the search query to specific key metadata fields. An example shown in Figure 2 for ACS publications which provides search for anywhere and in key fields, e.g. Title, Author and Abstract.

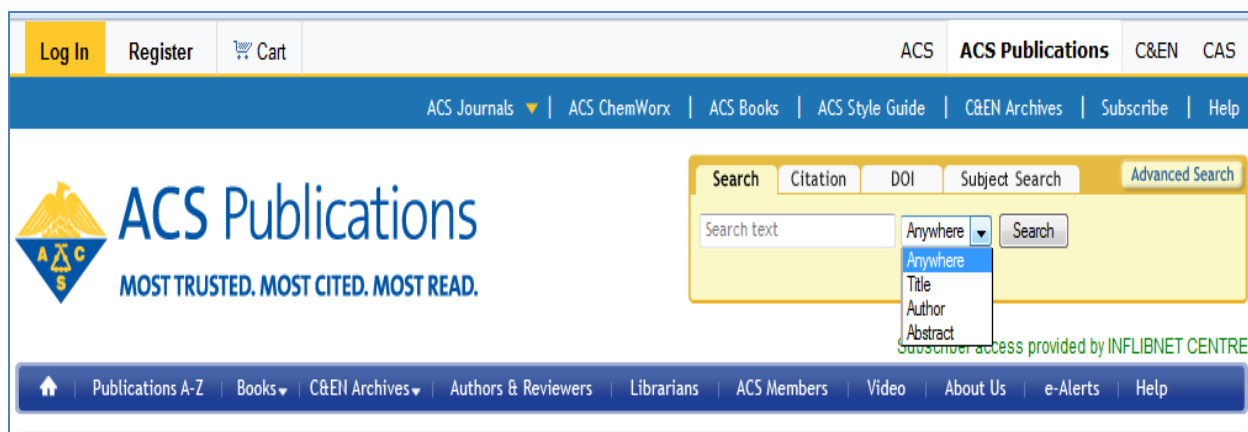


Figure 2 Simple Search box for ACS Publications

The e-thesis library of Library and Archives Canada, as shown in Figure 2, provide search for Name, Title, Keyword, Note, ISBN etc.

Figure 3 E-thesis library of Library and Archives Canada

Most digital library allow the basic search techniques using simple search box as given below:

Boolean Searching: The Boolean search operators, AND, OR and NOT, are used to broaden or narrow the search results.

The Wildcard (?) and Truncation (*) Symbols: The wildcard (?) and truncation (*) symbols are used to create searches where there are unknown characters, multiple spellings or various endings.

Proximity Search: Proximity Search gives results that contain two or more terms that appear within a specified number of words (or fewer) apart in the database(s). The proximity operator is placed between the terms that are to be searched, e.g. "HTML" <NEAR> publishing will search for documents that contain the word "HTML" and "publishing" within close proximity of each other (either before or after), i.e. it might fetch terms like “HTML and Electronic Publishing”, Electronic Publishing Using HTML”, or Publishing Electronic Text with HTML”.

Grouping Terms together Using Parentheses: Parentheses can be used to control a search query. Without parentheses, a search is executed from left to right. However, words enclosed in parentheses are searched first. Parentheses allow you to define the way the search will be executed. The left phrase in

Parentheses is searched first; then, based upon those results, the second phrase in parentheses is searched. Detailed description of various search techniques is discussed in a separate module.

2.2.2. Advanced Search

Digital libraries also provide "Advanced Search" functionality for expert users that provides either:

- i) Multiple search boxes to specify their search query/keyword in relation to specific fields like subject, creator, abstract, title, collection type, time period, geographic location, full-text etc;
- ii) Plain text box that allows a user to construct his/her search strategy using Boolean operators, proximity operators, parentheses, etc.

An example for advanced search screen of ScienceDirect as shown in Fig 4 shows the multiple search input box with the option to select the fields related to search string and selection option for Boolean operators between the fields. It also provides functionalities for limiting the search query to category of material (books or journals), selection of subject category and date range (publication years) of the content.

Figure 4 Advanced Search Screen for ScienceDirect

An example of the HathiTrust Digital library as shown in fig 5, which provide the option to add additional fields in a search query and limiting facilities for publication years, language and country of origin.

The screenshot shows the HathiTrust Digital Library's Advanced Full-text Search page. At the top left is the HathiTrust logo. A yellow 'LOG IN' button is at the top right. The main search area is titled 'Advanced Full-text Search:' and includes a link to 'Search Tips'. Below this are two search input fields with dropdown menus for 'all of these words' and 'in'. The first dropdown is set to 'Full-Text + All Fields' and the second to 'Title'. Below the search fields is a link to '+ Add a pair of search fields', which is circled in red. Underneath is a 'Limit to:' section. It includes a 'Full view only' checkbox, a 'Year of publication' dropdown set to 'During or after', and a text input field. Below this are two dropdown menus: 'Language' and 'Limit to Original Format'. The 'Language' dropdown is set to 'All' and the 'Limit to Original Format' dropdown is set to 'All'. Both dropdowns are highlighted with a red rectangle. At the bottom are 'Search' and 'Clear/reset' buttons.

Figure 5 Advanced Search Screen for HathiTrust Digital Library

2.2.3. Expert Search

Some digital libraries like ScienceDirect, JSTOR, etc. provides Expert Search facilities which does not provide a separate search box, instead provide an option to make complex search in a simple search box or text box. In Expert search, users have to formulate their query by using field tags, boolean operators, wild card characters and parentheses.

An example query for JSTOR interface is:

Query: ti:"american economic review" AND ty:fla NOT au:robert

The above query will search for articles of Type (ty) full length article(fl) where the article title (ti) will contain the phrase "american economic review" which is are not authored (au) by Robert.

Likewise, users can make complex queries with flexibility of which fields to query for and in which sequence. The abbreviations for different fields varies from interface to interface like for querying the article field one has to write the field tag as "ti" in JSTOR, where as for ScienceDirect one has to write the field tag as "ttl". As such for making such queries, one has to first refer the help manuals provided by the digital library for creating a complex query.

2.2.4. Faceted Search

With the invention of search engines, like Lucene which generates indices from metadata and full-text and retrieve results faster by searching the indices, newer discovery features are added for searching digital libraries. Faceted search provides search and discovery functionality and may include next generation search features such as relevance ranking, spell checking, tagging, enhanced content, search facets. The facets can be deployed as tools to refine a large number of search results and to narrow down search strategy to specific interest of the users. Faceted search also ensures that there is no null result for the user.^{vii}

In a nutshell search interface for digital libraries should include the following functionalities:

- Simple search : with search option for searching on all bibliographic fields, grouping results by archive/collections, sorting the search results on various fields
- Advanced search: with focus on searching specific fields with more complex search queries and filter options
- Full-text Search: include indexing of full-text content of the items along with the metadata fields for deeper search
- Intelligent search: provide intelligent search functionality like auto suggestions, spell checkers, similar item suggestions

2.2.5. MetaSearch

Digital libraries are not limited to a single repository of digital objects. The Contents of a digital library comes from different digital repositories accessible through library portals and resource discovery gateways. To cater to these requirements, digital libraries should incorporate standard and popular federated search protocols for exploring the contents stored in digital libraries. Some important search and retrieval protocols for incorporating search via federated search or meta-search solutions in digital libraries are as follows:

- Z39.50: Z39.50 is an ANSI / NISO standard for information storage and retrieval. It is a protocol which specifies data structures and interchange rules that allow a client machine to search databases on a server machine and retrieve records that are identified as a result of such a search. Z39.50 protocol is used for searching and retrieving bibliographic records across more than one library system. This protocol is not used by the Internet search engines. It is more complex and more comprehensive and powerful than searching through http. Z39.50 has been extended to allow system feedback and inter-system dialogue. Like most applications working under client-server environment, Z39.50 needs a Z39.50 client program on one end, and a Z39.50 server program on the other end.
- The name Z39 came from the ANSI committee on libraries, publishing and information services which was named Z39. NISO standards are numbered sequentially and Z39 is the 50th standard developed by the NISO. The current version of Z39.50 was adopted in 1995 superseding earlier versions adopted in 1992 and 1988.
- SRU/SRW: Search and Retrieval via URL (SRU) and Search and Retrieval Web Service (SRW) are Web Services-based protocols for querying Internet indexes or databases and

returning search results. The web services are two types, i.e, REST (Representational State Transfer) and SOAP (Simple Object Access Protocol). The SRW uses SOAP protocol and the SRW uses REST protocol for information retrieval. viii

- **NISO Metasearch XML Gateway (MXG):** MXG is proposed as an alternate to Z39.50 protocol and is based on the SRU protocol. The NISO MXG is a low-barrier-to-entry method for content providers to expose their content to metasearch application.ix
- **OpenSearch:** It is a way for websites and search engines to publish search results in a standard and accessible format suitable for syndication and aggregation. The OpenSearch is built on XML and supports a mechanism for telling a deep web search engine how to query it and the search results data are retrieved in a highly structured format. As such, search results are easy to process and display by a federated search service.x
- In many cases, centralized index are also maintained by harvesting the metadata of relevant sources of information. Some major standards for harvesting of metadata which should be provided for discovery of digital library content are as follows:
 - a. **OAI-PMH:** Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides a mechanism for repository interoperability. Using OAI-PMH, the service providers can make request to the repositories to harvest metadata. This is mostly used by the repositories and some e-journals providers also expose their content using it. ^{xi}
 - b. **METS:** Similar to OAI-PMH in purpose and function, METS supports XML-encoded metadata harvesting, but unlike OAI-PMH, METS can harvest both metadata and object. ^{xii}

Details of various standards including standards related to search and discovery of digital library content are discussed in a separate module.

2.3. Browse Interface

Another important aspect of information retrieval in digital libraries is the browsing process. The browsing process is to retrieve the desired item from the digital library where the detailed information of the object is not known or there is a lack of clarity in information need of user. Browsing is a vital part of the information-seeking process, allowing information seekers to meet ill-defined information needs and find new information. A common definition of browsing is an exploratory information-seeking strategy relying heavily on serendipity and being used to meet an ill-defined information need. Browsing is easily shown to be a vital part of the information-seeking process and very effective when combined with searching. ^{xiii}

Browsing can be provided by different formats with different entry points. It can be provided through metadata based alphabetical browsing like author, title, keywords/ subject

categorization (for example the Library of Congress classification scheme), publisher, publication year etc.

Image browsers (as available from Smithsonian Libraries collection (<http://library.si.edu/digital-library/art-and-design>)) are effective means for retrieving information in digital libraries containing images and video galleries where search queries may not be very effective in the retrieval of the items. The image browsers can be created in this method where users can view thumbnails for the items in a group of 10, 20 and so on.

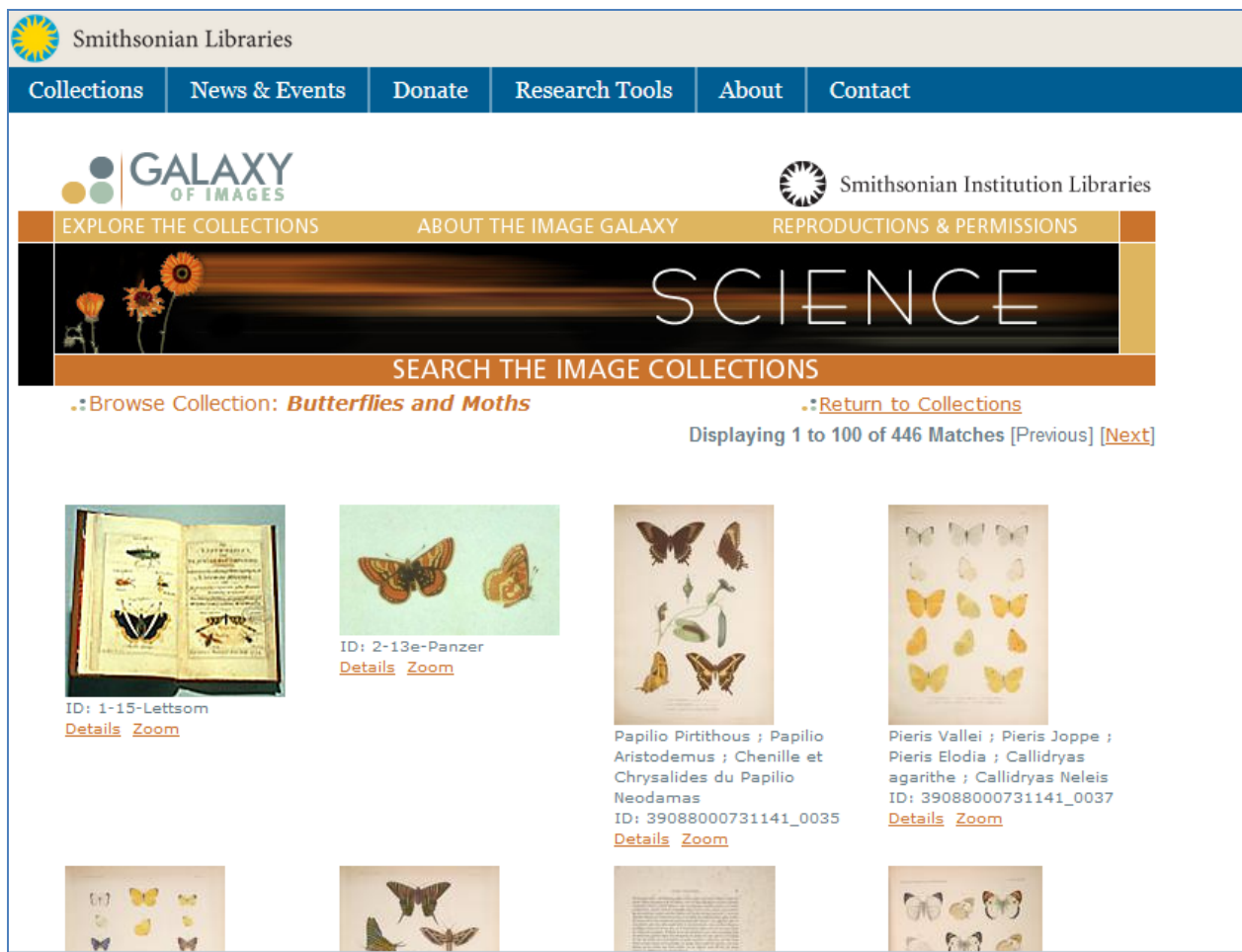


Figure 6 Image Browser in Smithsonian Libraries collection

Page browser are also a kind of browser which have an important usability in collections where digitized pages are individual images for manuscripts or journals. A typical interface of page browser can be viewed in Digital library of Villanova University (<http://digital.library.villanova.edu/>) which provide a thumbnail of individual pages on the left panel, enabling the user to browse between the index page and the page images to go from one image to the next and zoom in for detailed view.

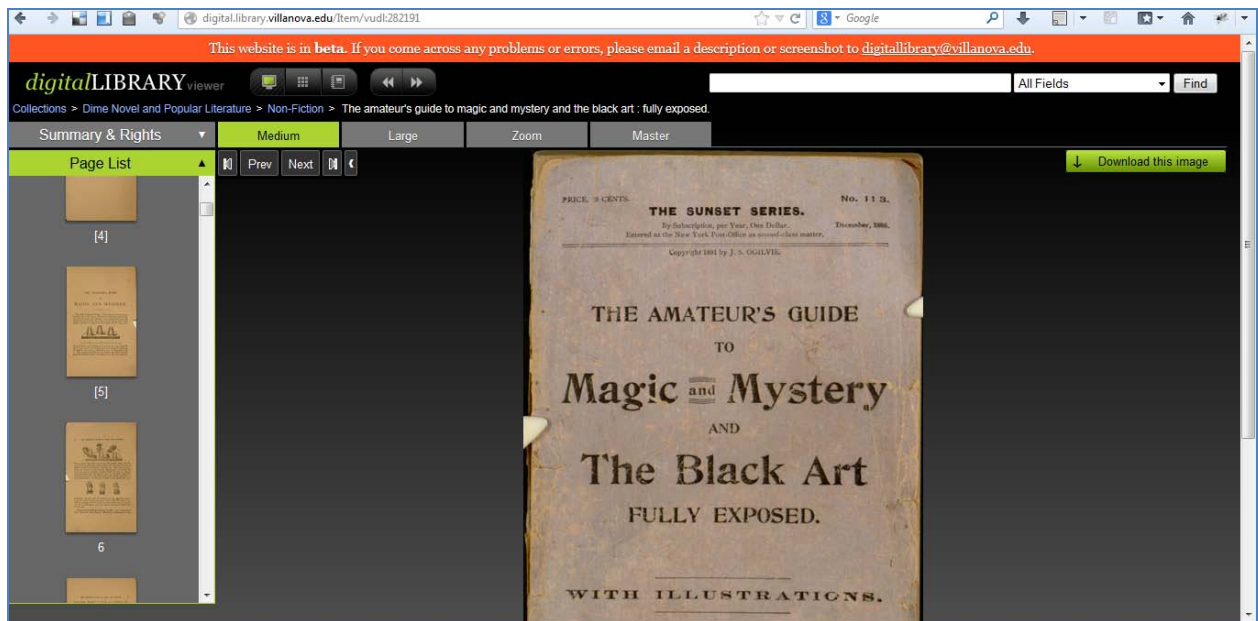


Figure 7 Page Browser view in Digital library of Villanova University

A modern semantic browser approach can also be provided using semantic browsing tools which will provide self organized maps and hierarchical structure of phrases based for browsing content based on semantic data. An elaborated account of semantic web and digital libraries is presented in another module.

2.4. Display of Retrieved information

Searching and browsing involves an iterative process for discovering content of digital libraries. It may happen that the required content can be retrieved in a single search query or the user may need to query several time and refine the search results for finding what exactly the user is seeking for. In order to refine the search query and make decisions of these kinds, the needs to quickly evaluate the search results. This is where the display of the retrieved information plays a key role.

The retrieved search results must be displayed with precision and brevity so that maximum search results can be displayed in a single window enabling users to select the relevancy of the content as early as possible.

An example of the search result page of Ebrary is shown in fig -8 that shows the brief details of an e-book with metadata of title, author, publisher, publication date and keyword etc. along with a cover image of the book for easy identification. It also provides an option to view the table of content (TOC) of the book on the same screen for further enhancement of the search result.

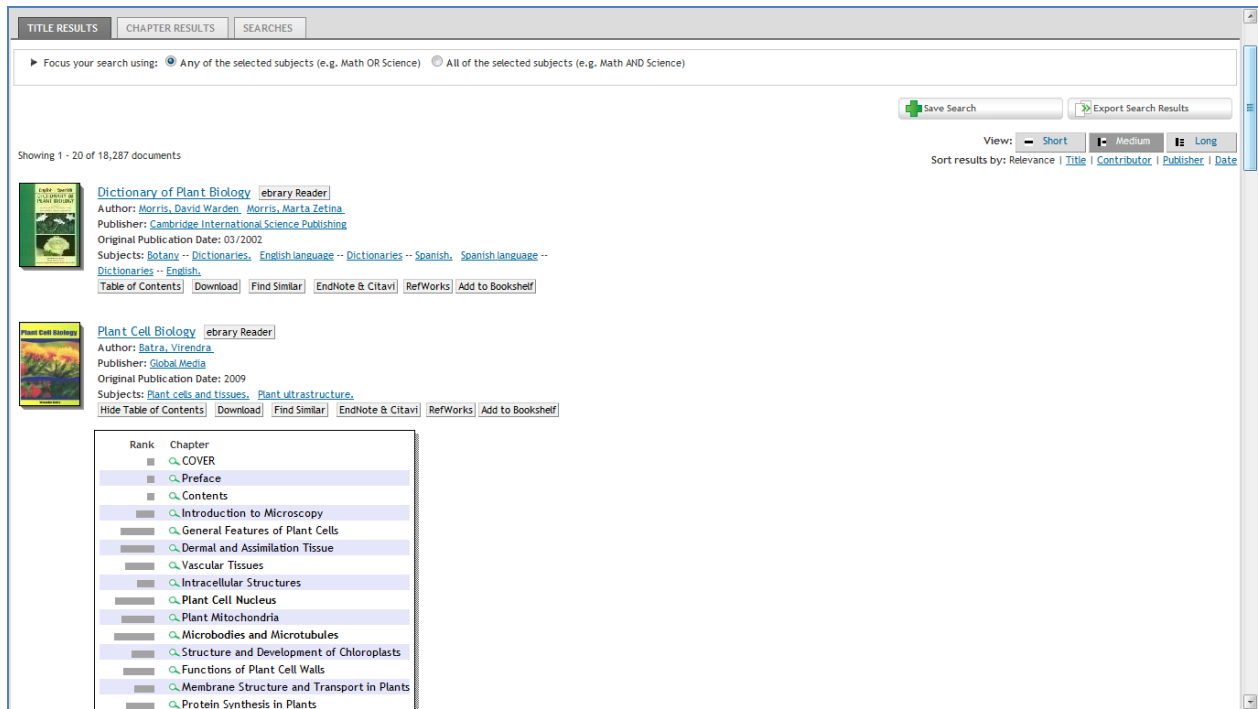


Figure 8 Search Result page for E-brary

Another example taken from HathiTrust Digital library (Fig 9) shows the search results with facets on the left side based on various fields like subject, author, language, place of publication, publication year etc. for refinement of the search results for narrowing the search query incase of large search result sets.

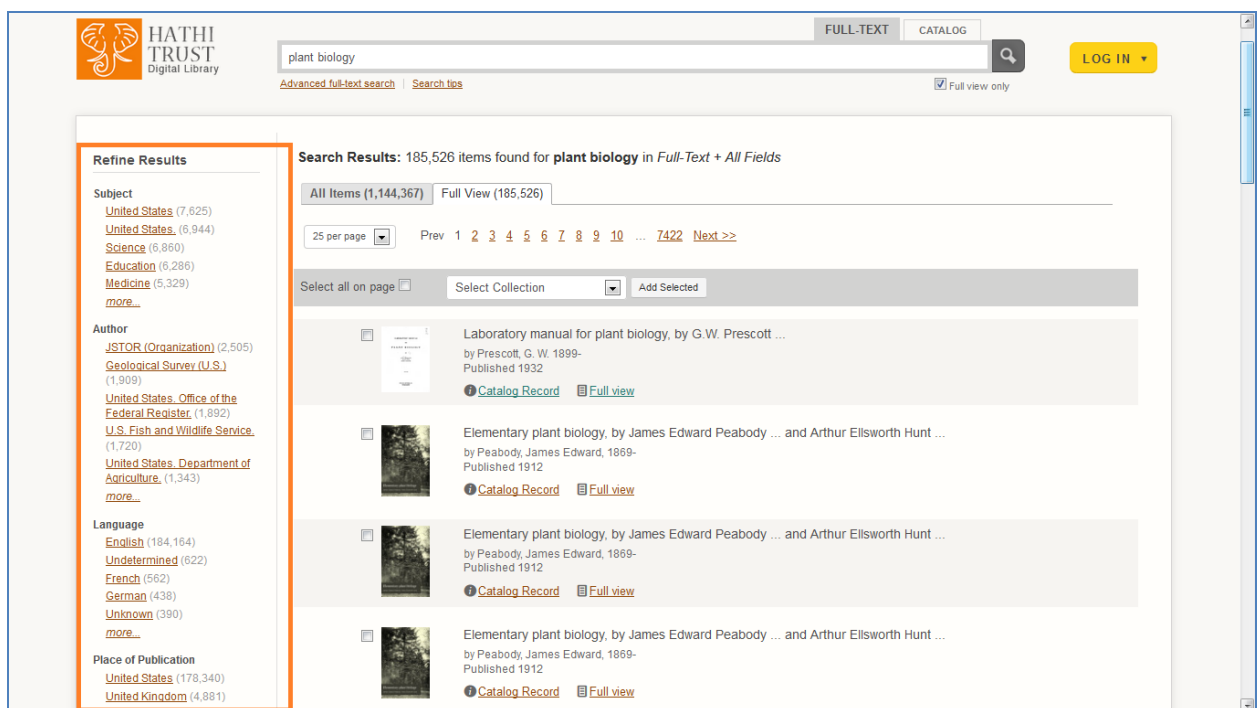


Figure 9 Search Result Screen for HathiTrust Digital Library

The JSTOR Archive, as displayed in Fig 10, also provide a very informative result set display where the basic information on search results are listed with a option to view a preview of the

item with abstract level display along with a view of the full-text content highlighting the searched key words for understanding the context of the keywords available in the item.

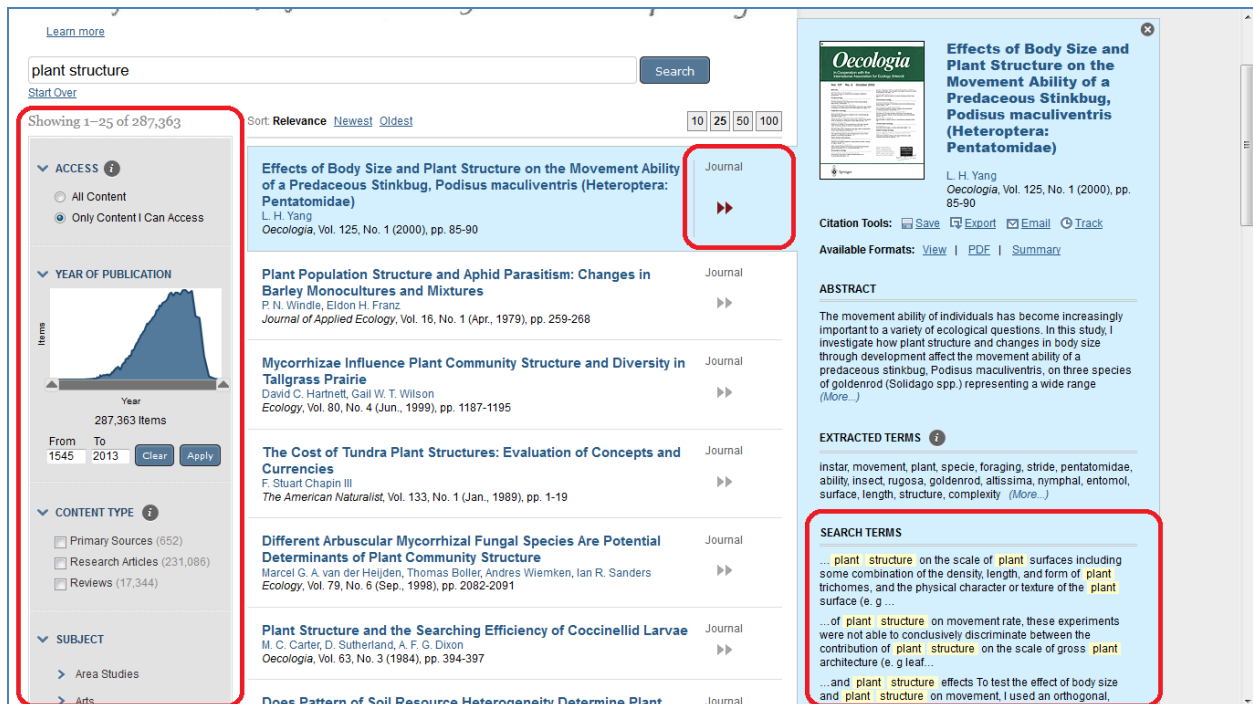


Figure 10 Search Result for JSTOR Archive

2.5. Specialized Digital Libraries (Video, Audio, Images etc)

Some specialized digital libraries contain various non-textual content (multimedia contents) like still images, videos and audio, etc. These specialized collections pose different challenges for the designers. Often in visual (image) content digital libraries browsing can be an effective mechanism where thumbnails of reduces size can be displayed on any page. Sound clips can be browsed using sampling at specific time intervals. But browsing may not be an effective way of retrieving these content where the collection is very large.

3. Summary

Interface design for searching and navigation is an important aspect in digital libraries. Browse and search interface connects a user to digital libraries. To expose the valuable content stored in a digital library, it should be designed in view of the targeted audiences. The support for simple search to sophisticated advanced query search should be taken care of based on the requirements of various types of users. The search and browse interface should have interoperability functionality with supported standards and protocols like Z39.50, SRU/SRW, Opensearch, OAI-PMH, METS etc. for exposing the content in aggregated services. Digital libraries with specialized non textual content pose challenges before the designer to find new options for easy searching and identification of the content stored.

References

1. Arms, William Y. (2000) Digital Libraries. MIT Press, . p 158,
<http://site.ebrary.com/id/2001012?ppg=158>
2. Dillon, A. (2002) Technologies of information: HC and the digital library. In J.M. Carroll, ed. Human-Computer interaction in the New Millennium. Boston: ACM Press, 457-474
3. Tedd, Lucy A... Digital Libraries: Principles and Practice in a Global Environment., Berlin, DEU: K. G. Saur, 2005. p 129.,
<http://site.ebrary.com/lib/inflibnet/Doc?id=10256544&ppg=148>
4. Galitz, W. O. (2002) The Essential Guide to User Interface Design. 2nd ed. New York: Wiley.
5. Hanumat G. Sastry, Lokanatha C. Reddy, User Interface Design Principles for Digital Libraries, International Journal of Web Applications, Volume 1 Number 2 , June 2009
6. Tedd, Lucy A.. Digital Libraries : Principles and Practice in a Global Environment., Berlin, DEU: K. G. Saur, 2005. p 129.,
<http://site.ebrary.com/lib/inflibnet/Doc?id=10256544&ppg=148>
7. Walker, J, (2006). New resource discovery mechanisms. The e-resources management handbook, 78-89
8. Morgan, Eric Lease. (2004). An Introduction to the Search/Retrieve URL Service (SRU). Ariadne 40. Available: <http://www.ariadne.ac.uk/issue40/morgan/>. (Accessed 20th Nov 2013)
9. NISO RP-2006-02, NISO Metasearch XML Gateway Implementers Guide. Available: <http://www.niso.org/publications/rp/RP-2006-02.pdf>. (Accessed 20th Nov 2013)
10. Content access basics - Part III – OpenSearch. Available: <http://federatedsearchblog.com/2008/01/04/content-access-basics-part-iii-opensearch/>. (Accessed 20th Nov 2013)
11. Open Archives Initiative Protocol for Metadata Harvesting. Available: <http://www.openarchives.org/pmh/>. (Accessed 20th Nov 2013)
12. Gibson, Ian, Goddard, Lisa and Gordon, Shannon. (2009). One box to search them all Implementing federated search at an academic library. Library Hi Tech. 27 (1), pp. 118-133.
13. D. McKay et al.: Enhanced browsing in digital libraries: three new approaches to browsing in Greenstone, Int J Digit Libr (2004) 4: 283–297 / Digital Object Identifier (DOI) 10.1007/s00799-004-0088-6