ISSN (Print) : 0972-2467
ISSN (Online) : 0976-2477
Journal of Information and Knowledge, Vol 60(2), April 2023, p.105-112
DOI: 10.17821/srels/2023/v60i2/170969

Enhancing OPAC Through Geospatial Data: Stackmaps and Text Visualization in Koha

Soumik Kerketta[1] and Mondrita Mukhopadhyay[2]*

[1]Librarian, Surendralal Das Teachers' Training College, Howrah – 711227, West Bengal, India; soumik.kerketta.bdy@gmail.com
[2]Research Scholar, Department of Library and Information Science, University of Kalyani, Kalyani – 741235, West Bengal, India; mondrita24c@gmail.com

Abstract

Physical stacks in academic libraries, despite the advent of digital repositories, remain important to users, especially in countries like India where physical resources hold considerable value. This study seeks to develop a system that enables users to locate books physically by integrating stackmaps functionalities with Koha OPAC. In addition, the study showcases how an open-source text analytics server can be incorporated inside an OPAC in Koha to generate various word-level visualizations by analyzing a text corpus, including the identification of geospatial features such as place names. This research aims to contribute to the advancement of information retrieval and visualization techniques in OPACs in academic libraries and to improve the user experience in locating physical resources. (The video abstract of this paper may be found at: https://youtu.be/q940TUkcTTE).

Keywords: Dreamscape, Geospatial Data, Koha, Stackmaps, Text Analysis, Voyant

1. Introduction

The library catalogue is a fundamental component in fulfilling the mission of libraries and aligns with the Fourth Law of Library Science, which emphasizes the importance of saving the time of users and advancing knowledge within the user community. In today's rapidly evolving network environment, it is necessary to reconsider how the user experience can be improved in discovering, locating, and obtaining materials in an Online Public Access Catalogue (OPAC). However, the legacy of the OPAC search has resulted in complicated search interfaces and generally does not provide any aid to users to navigate library collections, particularly in a large academic library where resources are distributed in many buildings and, within a building, on many floors. Traditionally, libraries have placed signs on different floors in buildings and created static floor maps to help users find documents and specific collections distributed throughout physical spaces. Many libraries have since attempted to add directional floor maps to their OPACs, generated dynamically based on call numbers (mainly class numbers), and with the advancement in text processing and text visualization technologies, libraries are also exploring the use of these tools in generating text analytics-based services in OPAC based on the text corpus of a given document, e.g., the full-text of an open access book. This research is an attempt to develop a prototype for the Central Library at the University of Kalyani that can provide dynamic location information for stacks along with different text analytics services if the full-text corpus of a document is available and linked in the catalogue database (MARC tag 856). In such an enhanced OPAC, users can directly access a stackmap from the Web-OPAC based on the call number of a retrieved record. It is quite feasible to create different dynamic maps (animated GIF files) for various call numbers (for example, DDC-based class numbers) by considering the distribution of the stacks in the library floors and then linking the resources dynamically through a JavaScript based on the

**Author for correspondence*

call number of a retrieved document. This study shows that in the same OPAC interface, it is also feasible to link a back-end text analytics server (like Voyant, an open-source text analytics server) for generating different text analytics-based services based on supplied text corpora, including displaying map-based movement between places mentioned in the given text.

2. Objectives

This research is a quest to find the answers to the following questions- a. Is it possible to locate a book in the stacks of a given library based on the call number of the book and reflect it visually in OPAC through suitable floor maps; b. Is it possible to read a book's text corpus, visually represent the results of text analysis, including mentioned geographical locations on a map, and then reflect this in the library OPAC? The two groups of objectives based on the questions asked are as follows -

- To develop a JavaScript to integrate stackmaps in Koha (latest stable release 22.11) OPAC through - a. Call number-based integration, and b. Display of a stackmap in a Lightbox.
- To integrate a text analytics server in Koha OPAC for - a. Text corpus-based analysis services; and b. Integration of the results of text analysis in Koha OPAC.

3. Review of Literature

Enhancing Koha OPAC through geospatial data is the primary focus of this research, and therefore, it needs to discuss initiatives related to various aspects of map-based library OPACs and the inclusion of text analytics services. In the early 2000s, a research study reported the development of a database of library stackmaps and the integration of maps generated through GIS software in a library OPAC to indicate particular locations of every individual book in a library (Xia, 2004). Since then, many libraries have experimented with integrating 2D as well as 3D maps into OPAC, which can visualize the physical structure of a library stack as well as the information structure of the subject coverage (Kaizer and Hodge, 2005). Integrating stackmaps into a library catalogue is not a new idea, and such initiatives have been reported during the period from 2005 to 2022 like the reports for – linking online library maps to OPAC by New York University's law library (Paulus-Jagric, 2006) and the initiative of Wichita State University (WSU) library to enhance users' experience in locating library materials in their OPAC. (Li and Deng, 2008). Later, the WSU library developed its library mapping application in 2010. The main mechanism for running this map-based document locating was based on a set of dynamic stackmaps modelled on WSU libraries (Gallagher, 2010). Researchers opined that integration of stackmaps in a library OPAC helps users to find their required document quickly (Chase, 2014). The development of animated stackmaps and integration of those maps in the EBSCO discovery service interface has been reported by Illinois State University (McMillin *et al.,* 2016). There are several other relevant studies on the incorporation of a variety of geospatial content types, from raster imagery to scientific vector data to georeferenced scanned maps, development of a data model for repositories and its integration with geospatial datasets using the Hydra framework (Durante and Hardy, 2015). Linda Hill (2006), in her book, provides an in-depth and comprehensive introduction to the concept of georeferencing, which is the process of assigning geographic coordinates to a particular object, feature, or data point. The book covers various aspects of georeferencing, including its history, methods, and applications in various fields such as cartography, geography, and remote sensing and provides an excellent overview of the fundamental concepts of georeferencing, including datum, projection, and coordinate systems. The book is well-structured and provides a step-by-step approach to georeferencing, making it accessible to readers with different levels of expertise in the field. Abresch *et al*., (2008) in their book entitled *Integrating Geographic Information Systems into Library Services: A Guide for Academic Libraries* provide a comprehensive guide to integrating GIS technology into academic library services. The book is written by a team of experienced academic librarians and GIS professionals who offer practical advice and insights into the benefits of using GIS in libraries. The book begins with an introduction to GIS technology and its applications in academic libraries. It then moves on to discuss collection development strategies, user education, and spatial analysis. The authors provide examples of how GIS can be used to support teaching and research activities in various academic disciplines, including geography, environmental studies, and social sciences. A strength of the book is its practical approach, providing step-by-step guidance on how to select GIS software, build a GIS collection, and develop user education programs. The book also includes case studies from academic libraries that have successfully integrated GIS into their services. The authors also provide a detailed overview of the key metadata standards and best practices for describing spatial data, including the FGDC Content Standard for Digital Geospatial Metadata and the ISO 19115 standard, providing a practical guide to implementing metadata standards and best practices in a library setting. The authors provide examples of how libraries can use metadata to enhance the discovery and use of geospatial data by their users. The chapter also covers issues related to data quality, data integration, and data interoperability, all of which are critical issues in the field of geospatial data management (Hanson and Heron, 2008). A follow-up research has been carried out to develop a MARC-formatted authority dataset for Indian Geo-administrative units given the inadequate coverage of Indian place names in global authority datasets. It starts with an authenticated place name file in CSV format and applies data wrangling tools and techniques to fetch geospatial data and other related datasets from open-access data sources to develop a geographic name authority file for Indian place names with geocoordinate data values. Later, this research also demonstrates how that authority dataset can be implemented in an open-source ILS and how retrieval features of a library discovery system can be enhanced through a geodetic search interface by utilizing that authority dataset. The entire methodology is based on open data, open-source software, and open standards (Mukhopadhyay and Mukhopadhyay, 2022).

Information retrieval in libraries of any type or size to date centres around only textual search despite tremendous advancements in digital information representation and retrieval. A study attempted to enhance the information retrieval features of a typical library search system by fusing geodetic search capabilities into it. The prototype framework applies Solr-based open-source library discovery software, namely VuFind as a retrieval system, Leaflet, an open-source JavaScript library for interactive maps, OpenStreetMap as a cartographic data provider available under the Open Data Commons Open Database License (OdbL), and a set of MARC-formatted bibliographic records on Antarctica processed in open source Koha ILS. The MARC tags 034 (in subfields $d, $e, $f, and $g) have been used. The tag 034 is used to represent the longitude and latitude of the places that the documents dealt with. The tags 651 and 653 are used to represent place names. The tag 651 takes place names from authority lists, whereas tag 653 takes values in the uncontrolled form (Mukhopadhyay and Mukhopadhyay, 2018). One research introduced the concept of Geographical Information Retrieval (GIR) providing an overview of the field of GIR, which is concerned with the development of techniques for retrieving geospatial information from large, heterogeneous databases. The authors describe the challenges involved in GIR, including the need to deal with complex queries, the variety of data formats and scales, and the difficulty of integrating geospatial and non-geospatial information. The authors provide an overview of the key techniques used in GIR, including text-based and content-based retrieval, spatial indexing, and spatial query languages. They also discuss the various applications of GIR, including environmental management, disaster response, and urban planning. One of the strengths of this paper is that it provides a clear and concise overview of the key issues and techniques involved in GIR, making it accessible to readers with various levels of expertise in the field. The authors provide a comprehensive review of the literature in the field and offer insights into the current state of the art and future research directions (Jones and Purves, 2008). Antelman *et al*., (2006) in their paper discuss the new functionality that has been enabled, the implementation process and system architecture, the assessment of the new catalogue's performance, and future directions.

4. Design and Methods

The stackmap plan of this study is based on a real-world scenario considering the layout of the library floors and stack positions of the Central Library of the University of Kalyani (CL-KU). The library has a three-storied building and houses approximately 1,69,128 books, 1,650 reference tools, 165 current journals (Print), 7,312 bound volumes of journals, 2,884 theses awarded by the university, 52 letters of Rabindranath Tagore to Parul Devi in original and so on. The open-access stack rooms occupy 278 double-sided stacks distributed on three floors. The books are arranged using the Dewey Decimal Classification (DDC) 23rd edition-based call numbers. The stacks are organized in a linear format by main classes, from 000 to 999. The CL-KU presently has SOUL-based OPAC but

Table 1. Research design

Group A: Integration of stackmaps	Group B: Integration of text analytics services
Use the Javascript-based approach as developed in the University of Hartford Library; Develop a set of dynamic stackmaps modelled on Wichita State University Libraries; Tune the Javascript plugin to handle DDC-based Call Numbers; Develop stackmaps for Central Library, Kalyani University (a three-storied building); This Javascript-based plugin for Koha OPAC consists of two main components: a collection of images (dynamic), and a script that retrieves these images; A CSS file which styles the map thumbnail and lightbox; and Maps are integrated into OPAC by adding the script tag to the OPACUserJS preference.	Installation and configuration of back-end language analysis tool (here Voyant – an open source text analytics server); Creation of corpus in local Voyant server; Development of linking mechanism: At Voyant end, the corpus (full-text document in Koha linked in the tag 856) must produce a unique URL; The linking script at Koha end needs to create a "Language analysis" tab conditionally in OPAC if tag 856 provides the analysis URL; Fetching data visualization from related corpus available at Voyant end in the given OPAC tab with the facility to navigate; Koha side tuning of OPACUserCSS, OPACUserJS and OPAC-detail.tt files

planning to switch to Koha with an RFID system in the year 2023. Under such circumstances, this research study may help CL-KU to enhance Koha OPAC by integrating stackmaps and text analytics-based services.

The design of this research study may be viewed from two major approaches as given in Table 1. It gives a bird's eye view of the entire methodologies related to two groups of works (Groups A and B).

4.1 Group A Methodology

Multiple maps are drawn based on the different physical locations of the documents in the Central Library, Kalyani University (call number-wise arrangement of documents in the library). The prototype presently supports up to the third summary of Dewey Decimal Classification (DDC, 23rd edition is in use). All these animated maps (in 'gif' format) are finally placed in a hierarchical order by

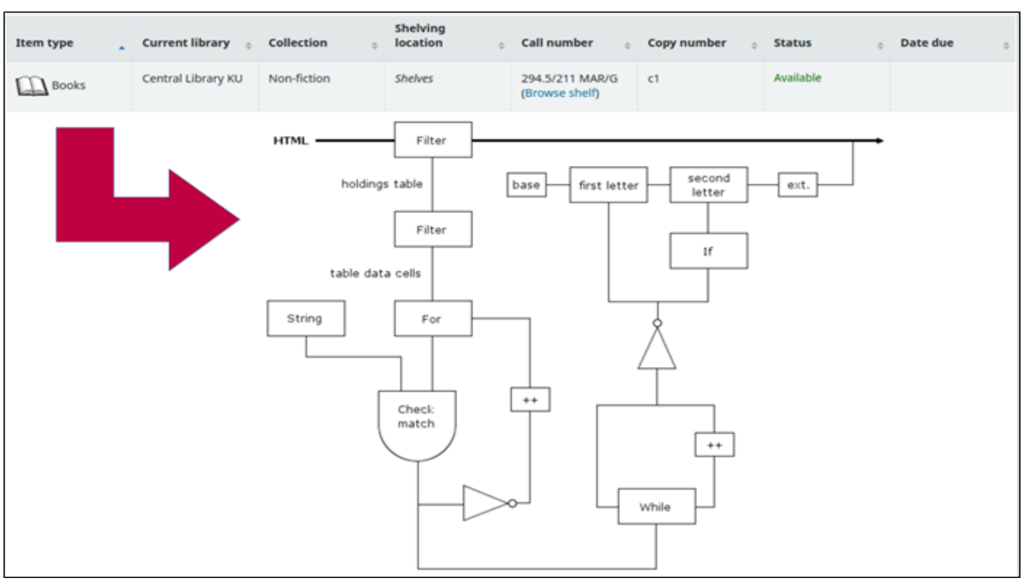

Figure 1. Concept - from call number to stackmap linking in Koha OPAC[1,2].
(Sources: University of Hartford Libraries[1] and Koha Shelf Map RFC[2]).

following the DDC third summary (like *shelves > 3 > 36* – to include all documents starting with the main class 36 like 361, 362 and so on) inside the document root of the Apache server (*/var/www/html/*) to ensure web-enabled access of the maps. The process involves many steps and starts with the drawing of basic floor maps of the selected library Stackmaps for this study have been drawn using GNU Paint based on different physical locations of the documents in the library (800×600 pixel size). In the next step, the location of the book is represented dynamically by using a human-like figure from the entry point to the book location. Finally, these multiple maps (against a given location) are merged in proper order into a single animated gif map by using Ezgif online utility (https://ezgif.com/) to generate a dynamic map for each call number-based location. After the map-building process for all the 3rd summary of DDC (000 to 999) is over, the next step is to develop a mechanism to read the call number of a retrieved book in the Koha OPAC and link it to the corresponding stackmap available under the Apache document root (*/var/www/html/kohamaps/*). The logic part of the script is illustrated in Figure 1.

This research study starts with the approach suggested by the University of Hartford Libraries in their Koha development project (https://libweb.hartford.edu/koha/maps/concepts.asp) and fine-tuned the proposed structure to manage DDC-based class number (in place of LC-based classification). It also takes into consideration the Shelf map RFC (https://wikiup-community.org/wiki/Shelf_Map_RFC) of the Koha community. The JavaScript as modified works in the following manner:

- A user searches for an item in the Koha OPAC and finds it.
- On the OPAC, the user sees a link to a map that shows the location of the item in the library.
- When the user clicks on the link, a script begins to work.
- The script finds the table that holds the location information for the item (Holdings tab in Koha OPAC – Figure 1) and scans through the table cells using a For loop and string matching until it finds the one with the call number.
- It then parses the call number to find the subject class and subclass.
- Using the subject digits, it builds a URL with a specific format.
- The maps must be organized in a specific way, with all maps of the same class in a folder named after the first digit of the class (*shelves > 3 > 36* – to include all documents starting with the main class 36 like 361, 362 and so on).
- The program is fast, taking only about 0.0005 seconds to find and display the correct map in a manner like - *http://localhost/kohamaps/maps/library6/shelves/9/92/920.gif*.

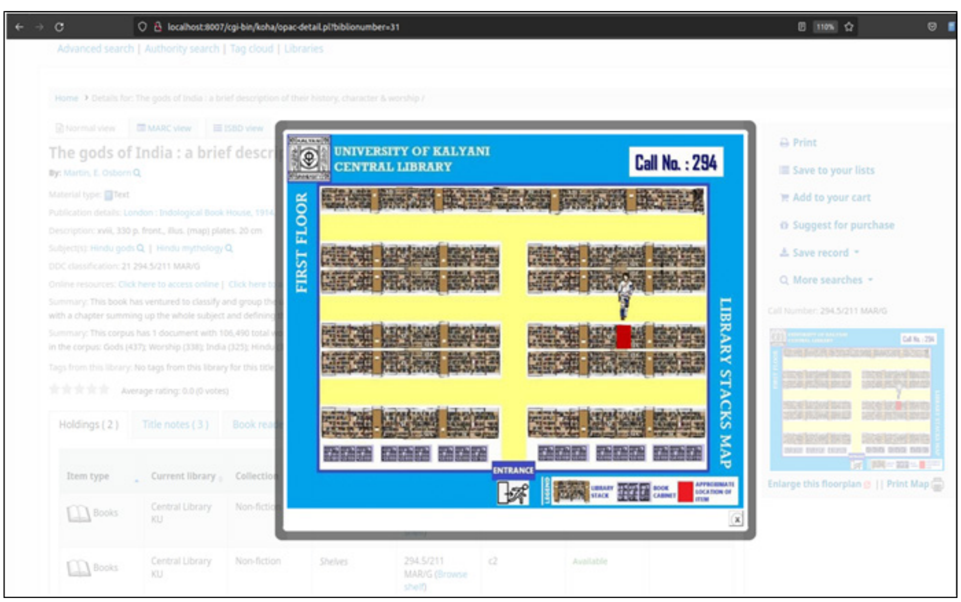

Figure 2. Implementation- call number-based stackmap display in a lightbox in Koha OPAC.

This research study has taken the clues from the advice given in Koha RFC and the approach recommended by the University of Hartford Libraries in their Koha development project and fine-tuned these recommended approaches to manage DDC-based class numbers in place of LC-based classification. The modified JavaScript works by searching for an item in the Koha OPAC and finding its location through a specific URL format that is built based on the subject digits of the call number. The maps must be organized in a particular way, and this program is incredibly fast to find and display the correct map (Figure 2). This approach could potentially improve the efficiency and accuracy of library OPACs that use DDC-based class numbers for cataloguing resources.

4.2 Group B Methodology

This part of the methodology is essentially based on the model proposed in the research study of Mukhopadhyay and Dutta (2020) that describes the processes of integrating an open-source text analytics tool, namely Voyant, with Koha. This study dealt with Koha 19.11 (the then stable release) and demonstrated how different text analysis services can be generated from a text corpus and can be integrated in real-time in a Koha OPAC. Our study adopted the approach as recommended in this paper but fine-tuned the entire process for the present stable release of Koha (version 22.11) and included a new geospatial service in a separate tab in Koha OPAC, namely Dreamscape. It is a tool in Voynat for the representation of texts in a geospatial context. Dreamscape aims to detect mentions of locations, particularly city names, within texts and uncover recurrent connections between these locations. This aids in identifying the movement of people, ideas, goods, or other phenomena. It is important to note that the concept of travel is subject to interpretation and scrutiny. While a sequence of locations may or may not have significance, Dreamscape strives to facilitate their study. A significant limitation of Dreamscape is that current methods for automatically identifying locations in texts tend to produce a substantial number of errors. These errors may involve falsely identifying a non-location as a location (false positive) or neglecting to tag an actual location (false negative). Due to the complexities of language and the limitations of computers in understanding its meaning, such inaccuracies are to be expected. However, the methodology may be presented as follows-

4.2.1 Selection of ILS - Koha

The identification and selection of a front-end Integrated Library System (ILS) requires consideration of several factors. These include the need for open source, web-enabled, Unicode-compliant, and REST/API support, as well as the ability to handle selective occurrences of a repeatable MARC tag/subfield and conditional display control. Koha, NewGenLib, and Evergreen are possible candidates, but Koha is the recommended choice due to its flexibility in API handling and user base.

4.2.2 Selection of Text Analytics Tool - Voyant

The identification and selection of a back-end language analysis tool/service is also a critical consideration. The chosen software must be open source, web-enabled, Unicode-compliant, and offer data visualization support, as well as the ability to selectively export result display URLs for linking in the bibliographic framework. Voyant is the preferred software option as it supports a wide array of file types, different analysis services, improved visualization facilities, and local installation options.

4.2.3 Linking Koha and Voyant

The linking mechanism between Koha and Voyant must have several key attributes, such as URL-based linking and the ability to record corpus data URL along with other links using a repeatable MARC tag/subfield. The linking script at the Koha end must create "Text analysis" and "geospatial" tabs conditionally in OPAC, if the MARC tag/subfield provides an analysis URL, and then fetch data visualization from the corresponding corpus available at the Voyant end in the given OPAC tab with the facility to navigate in other analysis URLs.

4.2.4 Activities at Koha and Voyant Ends

On the Koha side, Koha's OPACUserCSS can be utilized to create a class that can be called in OPACUserJS, which can indicate the availability of the corpus and control the OPAC display. OPACUserJS can be used to create links to the book reader if the first occurrence of 856 $u is present, to the corresponding corpus in Voyant if the second occurrence of 856 $u is present, and to call opac-detail.tt to create tabs depending on the presence of (a) and/or (b). On the Voyant side, a corpus can be created

Figure 3. Text analysis and geospatial features in Koha OPAC.

in the local Voyant server, and a unique corpus ID can be generated for easy identification.

In summary, the features of the linking mechanisms include the selection of the OPACUserCSS and OPACUserJS APIs in Koha to populate the system preferences with the script that will link Koha and Voyant, and the use of the MARC tag 856 sub-field "u" to include locally available URLs for language analysis service. The linking script has two logical sections: a linking algorithm to connect the book reader, if the first occurrence of 856 $u is populated in the MARC record (loads the book reader), a linking algorithm to connect text analysis, if the second occurrence of 856 $u is available in the MARC record (different text analysis services), and a mechanism to connect geospatial tab, if the third occurrence of 856 $u is available in the MARC record (for Dreamscape service). The following data entry rules need to be followed:

- Enter the full-text location in the first occurrence of the tag 856 subfield u;
- Enter text analysis URL (with term Cirrus as default) in the second occurrence of the tag 856 subfield 'u'; and
- Enter the geospatial analysis URL (DreamScape) in the second occurrence of the tag 856 subfield u.

The results of the mechanisms as discussed concerning the group B methodology are illustrated here in Figure 3.

5. Conclusion

In conclusion, this research study addresses physical stacks in academic libraries, especially in countries like India, where they hold considerable value for users. By integrating stackmaps functionalities with Koha OPAC and incorporating an open-source text analytics server, this study developed a prototype for the Central Library at the University of Kalyani to improve the user experience in locating physical resources. The proposed system provides dynamic location information for stacks, and different text analytics services if the full-text corpus of a document is available and linked in the catalogue database. This research contributes to the advancement of information retrieval and visualization techniques in OPACs of academic libraries and showcases the potential of using text analytics-based services in improving the user experience. Overall, this study demonstrates the potential for innovation and improvement in library services through the integration of technology.

6. Note and Acknowledgment

As it is too technical to report in the text the entire methodologies related to the integration of stackmaps and text analytics services in Koha OPAC, the authors have developed a video abstract (duration 2 minutes 48 seconds without audio) for providing readers with an overview of the enhanced OPAC services in Koha (latest stable release 22.11) based on the methodologies and techniques reported in this research study. The video abstract can be accessed here – https://youtu.be/q940TUkcTTE. The authors gratefully acknowledge the contributions made by Prof. Parthasarathi Mukhopadhyay, Department of Library and Information Science, Kalyani University in designing the architecture of integration and for guiding us in developing the scripts for fusing stackmaps and text analytics services in Koha OPAC.

7. References

Abresch, J., Hanson, A., Heron, S. and Reehling, P. (2008). Integrating geographic information systems into library services: A guide for academic libraries. IGI Global. https://doi.org/10.4018/978-1-59904-726-3

Antelman, K., Lynema, E. and Pace, A.K. (2006). Toward a 21st century library catalog. Information Technology And Libraries, *25*(3), 128-139. https://doi.org/10.6017/ital.v25i3.3342

Durante K. and Hardy D. (2015). Discovery, management, and preservation of geospatial data using Hydra. Journal of Map and Geography Libraries, *11*(2), 123-154. https://doi.org/10.1080/15420353.2015.1041630

Gallagher, P. (2010). Map it@ WSU: Development of a library mapping system for large academic libraries. The Code4Lib Journal, *10*. Available from: https://journal.code4lib.org/articles/3072.

Hanson, A. and Heron, S. (2008). Describing geospatial information. In: Integrating geographic information systems into library services: A guide for academic libraries. IGI Global, 82-113. https://doi.org/10.4018/978-1-59904-726-3.ch004 PMid:18948138

Hill L. L. (2006). Georeferencing: The Geographic Associations of Information. MIT Press. https://doi.org/10.7551/mitpress/3260.001.0001

Jones C. B. and Purves R.S. (2008). Geographical information retrieval. International Journal of Geographical Information Science, *22*(3), 219-228. https://doi.org/10.1080/13658810701626343

Kaizer, J. and Hodge, A. (2005). AquaBrowser library: Search, discover, refine. Library Hi Tech News, *22*(10), 9-12. https://doi.org/10.1108/07419050510644329

Krenelka Chase, A. (2014). Mapping the library for increased user engagement. Library Hi Tech News, *31*(4), 14-16. https://doi.org/10.1108/LHTN-04-2014-0026

Li, H. and Deng, S. (2008). Location and shelf mapping from OPAC search results: With reference to Wichita State University. New Library World, *109*(3/4), 107-116. https://doi.org/10.1108/03074800810857577

McMillin, B., Gibson, S. and MacDonald, J. (2016). Mapping the stacks: Sustainability and user experience of animated maps in library discovery interfaces. Journal of Electronic Resources Librarianship, *28*(4): 219-231. https://doi.org/10.1080/1941126X.2016.1236537

Mukhopadhyay, P. and Dutta, A. (2020). Language Analysis in Library OPAC designing an Open Source Software based Framework for Bibliographic Records in Mainstream and Tribal Languages. DESIDOC Journal of Library and Information Technology, *40*(5), 277-285. https://doi.org/10.14429/djlit.40.05.16034

Mukhopadhyay, P. and Mukhopadhyay, M. (2018). Search enhancements in library discovery: Towards a new dimension of GIS based information retrieval. In IASLIC 28th National Seminar, Visva-Bharati, Santiniketan; *57*, pp. 280-291.

Mukhopadhyay, P. and Mukhopadhyay, M. (2022). Developing geodetic search interface through auto-generation of geographic name authority datasets. Technological Advancements in Library Service Innovation. IGI Global, 59-81. https://doi.org/10.4018/978-1-7998-8942-7.ch004 PMCid:PMC9858612

Paulus-Jagric, D. (2006). Online law library maps. Law Library Journal, *98*(4), 691-702.

Xia, J. (2004). GIS in the management of library pick-up books. Library Hi Tech, *22*(2), 209-216. https://doi.org/10.1108/07378830410543520