LibraryVis: Towards Visually Understanding Library Resource Usage

Rahul Kamal Bhaskar, Mahshid Marbouti, S. M. Waliur Rahman, Julia Paredes, Ostap Orobets, Amir Karbalaei, Xuezhang Hu, Craig Anslow, Frank Maurer

Abstract—Understanding library resource data is challenging. Knowing which resources are the most and least used can help libraries save costs. We developed LibraryVis a web-based visual analytics application to explore the usage of library digital resources (e.g. conference papers, journals, e-books, and online databases) at our University. LibraryVis will allow librarians to make better decision about what resources to purchase and help inform the design of future business visualization applications.

Index Terms—Business Visualization, Information Visualization, Library Analytics, Visual Analytics.

1 Introduction

University libraries contain large collections of books and other resources, which can be loaned out to students and staff. Libraries work on a limited budget and often need to make decisions to buy or cancel a subscription to get the right to distribute resources. To make better decisions, librarians need to understand which resources are relevant to users and which ones are rarely used. Some researchers have explored visualization techniques to gain insight from the vast amounts of information collected by libraries to better support students, staff, and librarians.

The Bohemian bookshelf [4] and Springer [1] are visualization tools designed for the public to explore what is available in a library. Springer [1] uses visualizations to identify the most active countries, downloads at certain times, and trending topics in real time.

Xu et al. developed a tool to explore large digital collections [5]. The tool is targeted to archivists and makes use of visualizations to support decision making about allocation of resources. Data can be filtered on different aspects such as type of files, to identify trends in the collection.

JiscLAMP is an ongoing project focused on analytics for UK libraries [3]. The main focus is a data visualization dashboard to analyze data collected by libraries on a daily basis to improve the services provided.

This paper discusses the design of LibraryVis, a web based visual analytics application that supports librarians of our University in the exploration and understanding of resources usage in terms of popularity, location, and time. With the implementation of LibraryVis we expect librarians to make better purchasing decisions of resources when allocating their budget.

- *Rahul Bhaskar, Mahshid Marbouti, Waliur Rahman, Julia Paredes, Ostap Orobets, Craig Anslow, Frank Maurer are from the Department of Computer Science at University of Calgary, Canada. E-mail: rbhaskar@ucalgary.ca*
- *Amir Karbalaei and Xuezhang Hu are from the Department of Electrical and Computer Engineering at the University of Calgary, Canada.*

2 LibraryVis

LibraryVis is a web based visual analytics application to analyze the usage of digital resources (e.g. conference papers, journals, e-books, and online databases) of a library. LibraryVis is developed using JavaScript visualization APIs, as APIs are light weight and integrates to the existing web application with minimum modification in code. LibraryVis was designed to address the following business needs:

- **RQ1** How to review the overall status of the authentication system?
- **RQ2** Which resources are most and least utilized by the users of the library?
- **RQ3** Are there any unauthorized access to the digital resources through the Authentication System?
- **RQ4** What are the trends in digital resources utilization at the University library?

Design: To answer these needs, we deployed interactive visualizations to analyze digital resources usage data provided by the authentication system (i.e. EZproxy)of our University. EZproxy is a system which provides users (i.e. students and faculties of our University) with access to the digital resources subscribed by the library.

In a business environment, visualizations can have explanatory or exploratory goals [2]. The explanatory type is meant to direct analysis to a defined path in nature while the exploratory type helps an analyst to view a dataset from many dimension or compare multiple datasets with each other. Both explanatory and exploratory types of visualizations were considered in the design of LibraryVis.

We adopted an Agile software development methodology to develop LibraryVis. Weekly meetings were held in order to discuss the requirements with the librarians. Development process was initiated by creating paper prototypes based on the business needs. Prototypes were shared with the librarians and were improved based on their feedback. A list of associated tasks was created in order to implement the prototypes. Tasks were implemented based on the priorities from the librarians. The work was divided into

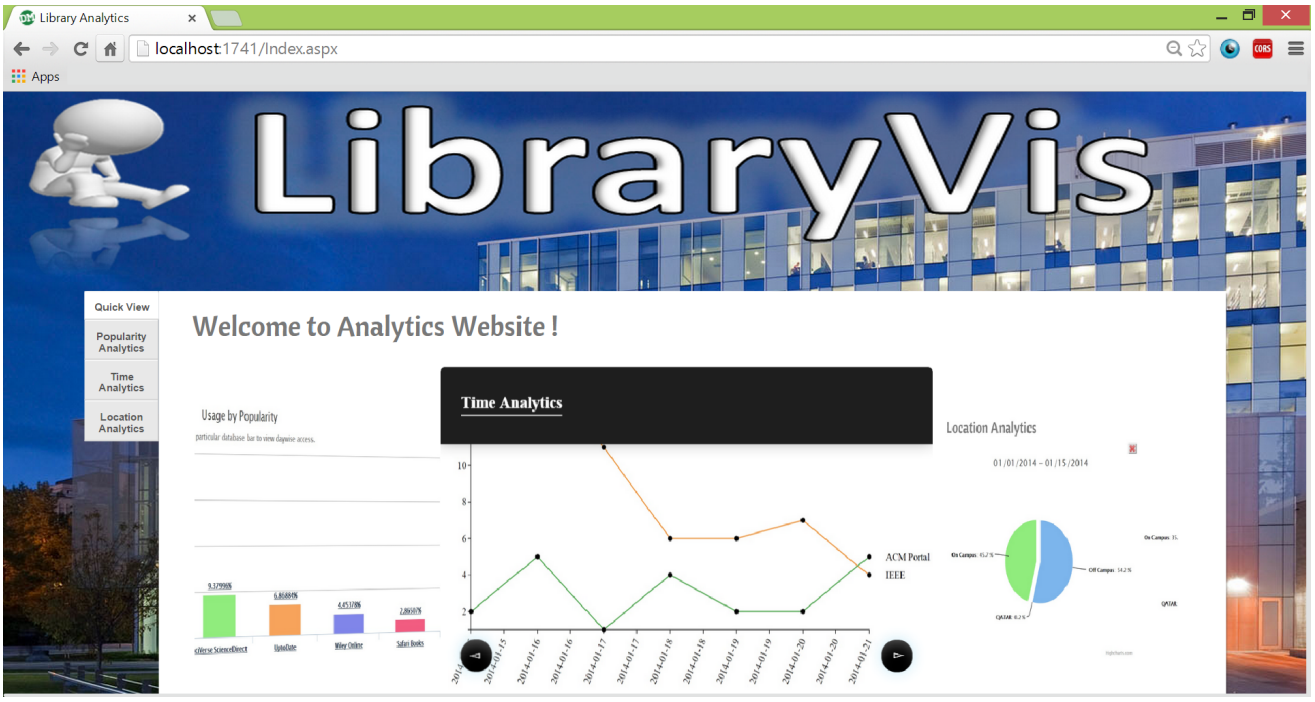


Fig. 1: LibraryVis: Display Quick View visualization that shows summary of popularity, time and location analytics.

multiple iterations. At the end of each iteration a demo session was arranged to show the librarians progress on the project and to discuss enhancement to the visualizations. The final prototype was evaluated by conducting a usability study involving the librarians.

Interactive Visualizations: LibraryVis contains different types of interactive visualizations. The goal of using these visualizations is to help the librarians with addressing their business needs (i.e. RQ1–4).

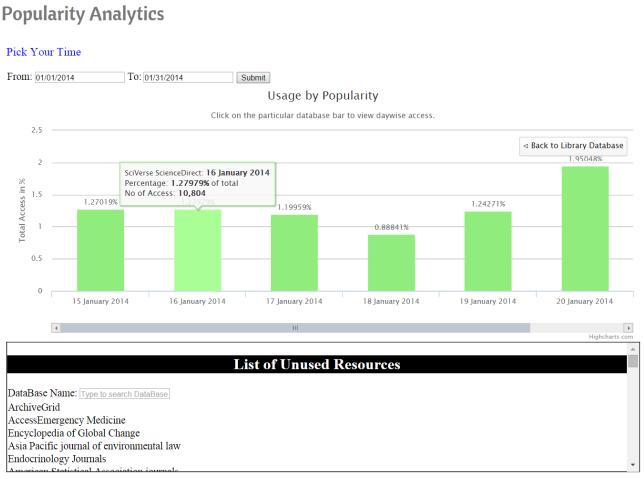
RQ1 – Quick View: A summarized version of the analysis is needed in order to see the overall status of the EZproxy for a period of time. As a result a dashboard (i.e. Quick View) was designed which showed the overall status of the EZproxy at one glance. Figure 1 provides a snapshots of the quick view containing visualizations from popularity, time, and location analytics.

RQ2 – Popularity Analytics: The librarians need to know whether the amount paid for a subscription to a digital resource is worth purchasing or not. This can be achieved by checking the number of times a resource is being accessed by the users of EZProxy or utilization percentage of the resources. Figure 2(a) shows the popularity analytics tab which displays a bar chart showing aggregate usage of different resources over a selected time period. In order to provide more detailed information some properties were encapsulated in sub charts. Detailed information is displayed by updating the existing graph and displaying the date wise usage of the selected database in the graph. There is a button in the visualization to toggle from the detail to the summary version of analysis.

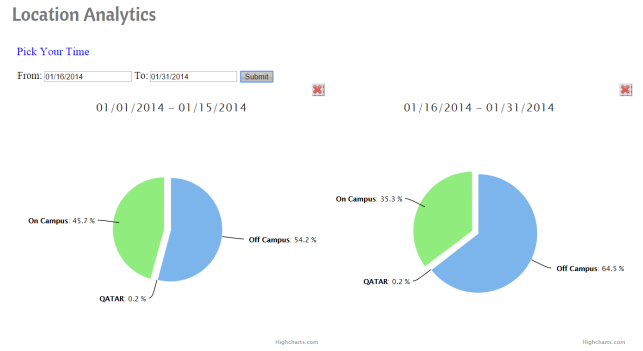
Suppose the librarians are seeking insight about the most and least utilized resources. For this analysis the librarians have to navigate to the popularity analytics tab in LibraryVis and enter a time period for the analysis. A visualization (i.e. bar chart) is displayed afterwards showing resource usage in decreasing order. Furthermore, the librarians are able to check on which day the particular digital resource is used by clicking on bars from the graphs and get the updated visualization which contains date wise usage of the digital resource. Figure 2(a) shows the detail view of the selected resource ScienceDirect. If the resource is not being utilized then it will be listed at the bottom of the page among the unused resources.

RQ3 – Location Analytics: Librarians want to classify from what location a resource is being accessed from (i.e. on-campus, satellite-campus or off-campus access). Access is determined by the users IP address, on-campus if the resource is accessed by the university IP address, satellite-campus if by the satellite campus IP address otherwise it is an off-campus access. The analysis can be performed by comparing and contrasting the percentage or sum of accesses by locations for two time periods. Figure 2(b) shows the location analytics visualization which displays multiple pie charts showing analysis for the EZproxy usage for on-campus, satellite-campus, and off-campus.

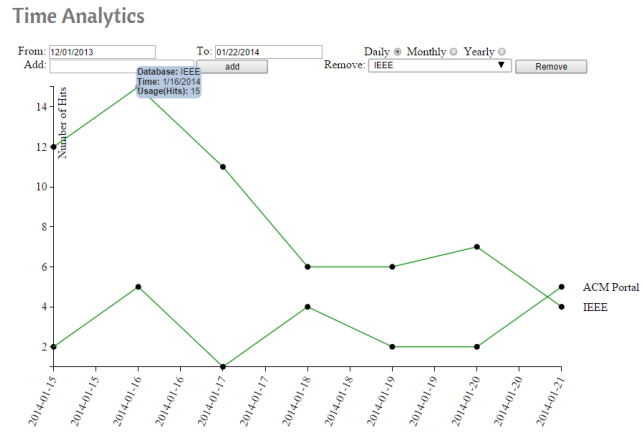
Suppose a librarian wants to seek insight about resource accessed in order to check fraud for over utilization from different locations. This can be accomplished by navigating to the location analytics tab and submitting multiple time queries for which they want to compare and contrast. The result of the queries displays multiple visualizations which show the percentage of access by locations. Figure 2(b) shows the location wise usage trend for 1-15 January 2014 (i.e. on-campus: 45.7%, satellite-campus: 0.2% and off-



(a) Popularity Analytics.



(b) Location Analytics.



(c) Time Analytics.

Fig. 2: LibraryVis: example visualizations.

campus: 54.2%) and 16-31 January 2014 (i.e. on-campus: 35.3%, satellite-campus: 0.2% and off-campus: 64.5%).

RQ 4 – Time Analytics: Figure 2(c) shows the time based analytics which consist of a line chart that shows comparative analysis on the usage pattern of the different digital resources over time. The librarians can compare and contrast resource usage by selecting multiple resources. This visualization can be updated to show aggregated values (i.e. grouping of data on daily, monthly and yearly basis).

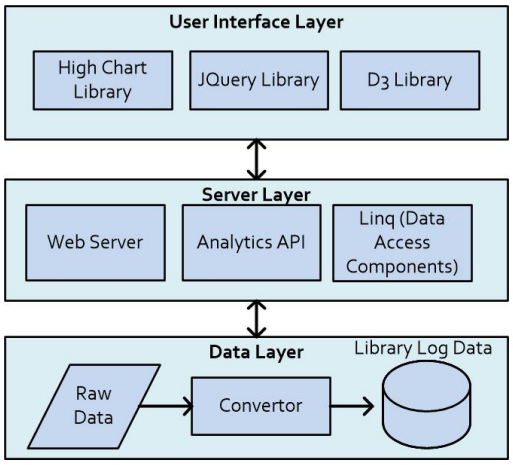


Fig. 3: LibraryVis Architecture

Suppose a librarian wants to compare the usage of two resources over time. This can be accomplished by navigating to the time analytics visualization and submitting a query for the different resources. Figure 2(c) displays two line which compares two resources IEEE and ACM portals usage trend which help the librarians to understand how many times one resource is used in comparison to others.

Architecture and Implementation: Figure 3 shows the architecture of LibraryVis, which has three layers. The data layer uses a relational database for storing EZProxy data. Information was available in text format. Prior to uploading the data it was parsed. Database updates were performed manually. Raw data was collected from the log files generated by the Ezproxy. Converter was used to process, filter and uploaded data into database (i.e. library log data). The server layer provides data processing. The data required for analysis is queried using LINQ queries. Further queried data is used by the Analytics API which was used to process queried data for extracting insights. Finally the processed data is sent to the user interface layer via a web server. The user interface layer provides visualizations of the queried data in the web interface, using different APIs (i.e. Highcharts and D3). Web interface was implemented using HTML, JavaScript, and JQueryUI. The user interface layer and server layer communicate via Ajax.

3 Evaluation

To evaluate the design of LibraryVis, we conducted a usability study at the last stage of the prototype. The participants included two librarians, one visualization expert and a system analyst. During the study, participants were given the software with a set of instructions consisting of 9 tasks designed to cover all the major areas of LibraryVis. Each task was a request to request a visualizations on the website and to use it while providing feedback on what went good and what went wrong. The tasks included :

1) Using time analytics tab, add resource to be visualized on the chart.
2) Remove the added resource in the time analytics tab.

3) Take a few minutes to test the features provided in the Time Analytics tab and report any problems you find in the tab.
4) Using popularity analytics tab, Find the name of the most used publication accessed for journals, papers or any other resource name in the past 3 months.
5) Using popularity analytics tab, find the most used resource.
6) Using popularity analytics tab, find least used resource.
7) Using location analytics tab, find what percentage of resources have been accessed "off campus" and "on campus"?
8) Using location analytics tab, select a different timeline and try to compare results of the two timeline.
9) Using location analytics tab, try to remove one chart and add another one.

The study lasted for 20 minutes where the participants completed all instructions and provided the development team with valuable feedback on their experience using LibraryVis. Participants' comments helped us discover some features in the UI that they had difficulty in working with. For task 1, participants had to select a database and then add the according visualization, however, they expected the selected database will be added automatically.

> *"Seems strange to select and then also have to press add."* Participant 1.
> *"I needed to be told to click add."* Participant 2.

The comments also helped us improve some features according to participants preferences. For example while selecting different time periods, participants preferred some dates will be selected automatically.

> *"Whenever 'from date' is selected 'to date' should be updated automatically."* Participant 3.

Another example was that participants liked to see some additional information on the screen.

> *"I would like to see unused library resources along with used resources."* Participant 2.

Another interesting point was, after working with one tab, working with another tabs of the application was easier and faster to them and finally they could perform all the tasks and answer the questions correctly. Overall, the feedback was a great opportunity for us to observe potential future users of the system. This allowed us to have a better understanding of their way of interacting with the application and how to make the application more effective for their business needs.

4 Discussion

Developing LibraryVis required iterative feedback from the librarians, who played significant role in explaining domain as well as in creating effective visualizations for viewing usage of digital resources. An Agile methodology was used to gather requirements, develop, and prioritize features based on themes. It is observed that agile approach along with participation of the customer is very effective in developing product, as you get regular feedback which help in improving design while development is still in progress.

The visual analysis provided by LibraryVis makes the work of the librarians more efficient in terms of time spent analyzing their resource usage. Before LibraryVis the librarians had to go over logs with millions of records which made it difficult to filter the important information for relevant analysis. The value of LibraryVis is that it allows librarians to use their expertise to analyze data that was sometimes hidden or not easily accessible in these logs. LibraryVis summarizes and displays the most relevant information for the librarians, but is flexible enough to get some input from their users.

The University had a simple website to show stats about resources usage. This website wasn't interactive which made it difficult for the librarians to analyze the data when presented with unnecessary information for the tasks they wanted to complete. The successful development of LibraryVis showed the University that interactive visualizations not only made it easier to complete the tasks, but as shown in our usability studies and in out meetings it was easier to analyze the data and come up with new questions to be further analyzed.

The usability studies showed that separating the visualizations by themes had some limitations when exploring data. For instance, librarians couldn't simultaneously filter and see a resource usage by time, location, and popularity. Although the visualizations were not novel, the integration of visualizations into one application allowed discovery of trends for effectively understanding the use of resource.

For future work, we are integrating LibraryVis into the library system at the University to facilitate the business needs of librarians. The first thing we need to do is to automatize the parser so that every time we have new logs, we are able to upload this data to our databases to perform fast queries. Since we had our initial prototype working with a local database, we need to migrate our data to the University's servers. This will allow the data to be more secure, and faster to access. Once LibraryVis is deployed, we would like to conduct some field studies to analyze the usability of the application in the context of librarians day to day jobs. We are also planning to expand the LibraryVis by adding reporting features and fraud detection visualizations to detect when accounts are being misused.

Librarians have to make decisions to allocate resources effectively to serve users better. We presented LibraryVis which allows librarians to visualize information about resource usage so they can make better informed decisions when determining what digital resources to purchase.

Acknowledgments

We would like to thank John Brosz, Andrew Pasterfield, Rene Reaume, and Helen Clarke from the Taylor Family Digital Library.

References

[1] Springer realtime. http://realtime.springer.com/.
[2] FusionCharts. Principles of data visualization. http://www.fusioncharts.com/whitepapers/downloads/Principles-of-Data-Visualization.pdf.
[3] B. Showers. Jisc lamp: Shedding light on library data and metrics. http://infteam.jiscinvolve.org/wp/2013/02/01/jisc-lamp-shedding-light-on-library-data-and-metrics/, 2013.
[4] A. Thudt, U. Hinrichs, and S. Carpendale. The bohemian bookshelf: Supporting serendipitous book discoveries through information visualization. In *Proc. of CHI*, pages 1461–1470. ACM, 2012.
[5] W. Xu, M. Esteva, S. D. Jain, and V. Jain. Analysis of large digital collections with interactive visualization. In *Proc. of VAST*, pages 241–250, 2011.