Catherine McGowan

17:610:553:93

Term Project Final

PROJECT TITLE: Targeted Information – A Social Digital Library Design Proposal

ABSTRACT

How can the gap between a social digital world and a digital library be bridged? In this paper, I will propose a digital library that offers Targeted Information to users to bridge the gap between the digital social worlds of users and the information stored within the databases of a digital library. The proposed methods will employ data mining of a user's browsing history to identify subjects and key terms that are current and relevant to the active and present interests of the user. Once the key terms are collected from the user's browsing history, the website of a digital library will auto-generate recommended titles for the user to explore. This proposal is an attempt to connect a user's Internet social activity with the content available within a digital library.

1. INTRODUCTION

The Pew Research Center performed a study in 2016 and determined that 79% of all Internet users are Facebook users. (Greenwood et al, 2016) This statistic is vital because it illustrates an essential characteristic of users and their Internet usage and behaviors. Facebook is a website and application designed for social interaction between users, blogs, media, etc. The digital world that Facebook creates for individuals fosters an environment for social activity and exchange, and information consumption to take place. Social activity or interaction covers a multitude of subjects that can range from photos of a user's latest meal to thoughts on the latest political event featured in the news. This is a

centralized location for identity formation, community building, and cultural exchange to occur for individuals. Accordingly, the statistic that 79% of all Internet users are Facebook users infers that 79% of Internet users are participating in the formation of identity, development of community, and exchange of culture.

Digital libraries contain content that ranges from entertaining works of fiction to informative, scholarly articles. These creative and scholarly works are reflections of human identity formation, community construction, and cultural exchange and theory. However, the content of digital libraries is quarantined to the framework of the digital library. The individual must consciously choose to seek information (or entertainment) from a library and leave the digital world she is socially, emotionally, and intellectually actively participating in. So, how can the gap between a social digital world and a digital library be bridged? In this paper, I will propose a digital library that offers Targeted Information to users to bridge the gap between the digital social worlds of users and the information stored within the databases of a digital library. The proposed methods will employ data mining of a user's browsing history to identify subjects and key terms that are current and relevant to the active and present interests of the user. Once the key terms are collected from the user's browsing history, the website of a digital library will auto-generate recommended titles for the user to explore. This proposal is an attempt to connect a user's Internet social activity with the content available within a digital library.

2. OBJECTIVES

The purpose of this digital library design proposal is to intersect the daily Internet based activities of users with information stored within a digital library. Targeted

Information will auto-populate on the digital library website to create suggested content available in the digital library, customized specifically for the user based on her Internet activity. Users will have the ability to explore the recommended titles through two layers of filtering. The basic filtering layer will allow the user to accept or reject a recommendation. The advanced filtering layer will allow the user to review the key terms that were retrieved by means of data mining the user's browsing history and select or deselect the key terms the user wishes to shape the title recommendations. This design proposal will also include data mining of the digital library itself, relative to the borrowing history of its patrons. Patrons who are identified as members and have a login and / or library card will have content suggested to them based on their borrowing history fused with other patrons' similar borrowing activity. Finally, all searchable content available within the digital library will contain abstracts that users can share across social media platforms to contribute to conversations and create meaningful connections between social activity and digital libraries.

3. LITERATURE REVIEW

There are several approaches that have been researched to create title recommendations and topic filtering for users. I will discuss two major approaches in this section. The first approach is commonly referred to as content-based or collaborative filtering. In this approach the user must have a profile and actively participate in the shaping of her interests profile. This method is a long-form approach and requires continual maintenance of interest preferences to remain relevant as the user's interests

evolve. If the user does not maintain the interest preferences, the functionality

performance for title recommendation is low.

> "The main challenges of content-based filtering approaches are the identification
> of item features from content or description and the development of user interest
> profiles that distinguish preferred items or disliked ones." (San-Yih et al, p. 262)

Furthermore, the user's interests can have characteristics that ebb and flow. Some

interests are permanent, and some are short-term respective to her current studies or

personal life. The permanence of content-based or collaborative filtering does not allow

for a malleability or adaptability of evolving user interests.

The second approach is referred to as the association rule. In this approach, the

user's borrowing history is mined for data that determines relationships or topic clusters

to generate title recommendations.

> "Association rule Mining is a popular data mining method and well researched
> method for discovering interesting relationship between variables in large
> databases." (Jomsri, p. 131)

While this method works to find correlations between subjects and titles that may not be

an obvious connection, or a conscious subject search performed by the user, it still fails to

recommend titles that are relevant to the user's current activity. It requires accessing a log

of user borrowing history to recommend titles, which requires the user to maintain an

active profile within the digital library.

The approaches outlined above both fail to recommend titles that are relevant to a

user's social activity that occurs outside of the digital library. Active Internet users are

socializing and seeking information regularly outside of the framework of a traditional

digital library. The proposed design and integration methods outlined below are

suggestions for taking the first step to bridge social activity with formal information

seeking to create Targeted Information that is relevant to the present tense Internet-based conversations, thoughts, and activities of a user.

4. DESIGN INTEGRATION AND METHODS: Targeted Information

In order to define what Targeted Information is, I will briefly explain *targeted advertising*. Users encounter targeted advertising daily on the Internet. These ads are auto-generated, for example, in a user's Facebook news feed or featured on the right side bar of a Google search, based on recent user activity. The user activity is identified and processed by accessing user cookies stored in a user's Internet browser. Keywords are pre-assigned and stored in the coding of targeted ads, so when the JavaScript and PHP codes are scanning the cookies of the user, the program is searching for keyword matches between the keywords assigned to the ad and the keywords found in the user cookies. When the matches are found, the relative ad is auto-generated for the user to see.

The goal is to apply the same coding and functionality of targeted advertizing to Targeted Information. In this coding and programming design, the digital library would use the database of keyword search terms that are attributed and stored as metadata to all searchable content within the digital library, and it would search for keyword matches found within the user's cookies. When keyword matches are found, a list of recommended titles would populate for the user to explore—this is Targeted Information. This process would be the first step in creating a connection to the available content of the digital library that is relative to the recent activity of the user. For example, if the user recently explored a web page for a news article on healthcare reform, the programming

function of a digital library can generate Targeted Information on that topic for the user

to explore.

Adding Targeted Information capability does not require a complete digital

library website redesign or overhaul. The following is a list of the basic requirements for

adding Targeted Information functionality:

1. Add a viewable results box for Targeted Information auto-population. [See

   Fig.1]

2. Apply JavaScript and PHP code to the page or pages where the Targeted

   Information results box is located.

3. Connect Targeted Information results hyperlinks to search the database for the

   user to access the recommended titles.

Figure 1 Design Proposal of Targeted Information Recommended Titles View
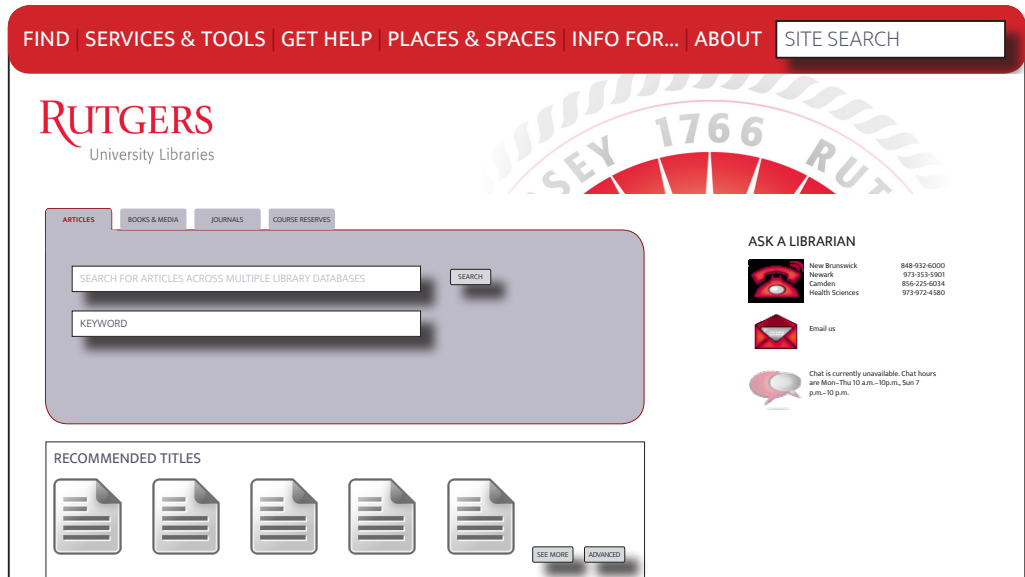
4.1 DESIGN INTEGRATION AND METHODS: Title Recommendation

Internet users are accustomed to product recommendations on retail websites such as Amazon. The product recommendations are generated unique to each user based on the following criteria: purchase history, recent searches, and the purchase history of other customers with similar products. The purpose of auto-populating product recommendations is to encourage the user to explore more items beyond the scope of the intended visit. This is known as building the basket for the customer—maximize the potential for increasing the total transaction sale. In traditional brick and mortar retailers, this would be referred to as up-selling.

There are two functional layers that digital libraries can employ for incorporating algorithmic Title Recommendation lists into their websites. The primary functional layer requires building a database of all titles borrowed by patrons and creating relationships between the two. What are the titles a patron has borrowed in her user history? What other patrons have borrowed one or more of the same titles? What are the subject keywords that are assigned to these shared titles? What other titles are assigned these same subject keywords? The goal here is to find the relationships shared between the borrowing history of multiple patrons and auto-populate Title Recommendation lists. These lists would populate on the existing pages of titles. For example, if a user searches for the title *Fahrenheit 451*, the page would auto-populate a list of recommended titles based on the history of the users and subject keywords assigned to other titles available in the digital library.

The second functional layer is an addition to the primary functionality described above and requires the application of a user login for each patron. When the user is

logged in, the Title Recommendation lists will auto-populate relative to the user's

borrowing history. Borrowing history is data that already exists in most digital libraries,

so the purpose is to access it in ways that are engaging to the patrons and make the results

feel personalized and tailored to their interests. In this scenario, Title Recommendation

lists can also include the results of Targeted Information, which would be a combination

of the user cookie information pulled from the user's web browser, the borrowing history,

and relative titles that match the keywords associated with both.

The Title Recommendation lists would be viewable in two formats for the user.

The first format would be the simple view of the Title Recommended list. [See Fig. 2]

The user can scroll through the list of recommended titles, read the abstracts or overviews

of the titles, and choose to reject or accept these suggestions. Acceptance and rejection of

the recommended titles would be in the form of two buttons: Add to List and Not

Interested. When the user selects Add to List, the title is added to a reading list that can

be compiled during the user visit. (This function is similar to a shopping cart for e-

commerce websites.) When the user selects Not Interested, the title is removed from the

list, allowing the user to manually refine the Title Recommendation results.

Figure 2 Design Proposal of Title Recommendation List Simple View
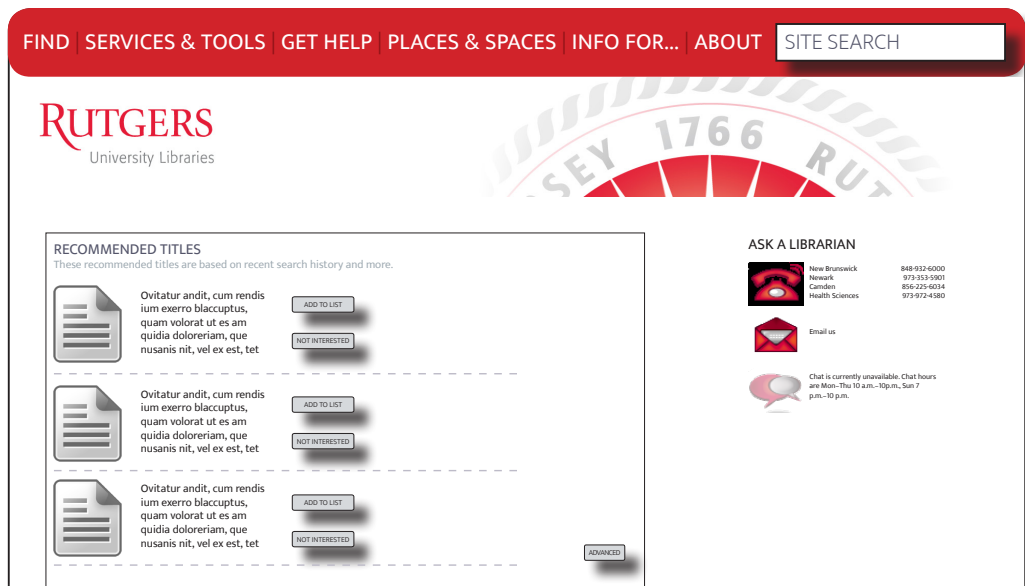
The second format would be the advanced view of the Title Recommended list. [See Fig. 3] In the advanced view, the user can review all key terms that were retrieved from the PHP function that searched and retrieved recent browser history terms. These retrieved key terms would be presented in a list format with a checkbox or radio button to select or deselect. In this step the user can select the key terms that will be searched for in the digital library to refine the search results.

Figure 3 Design Proposal of Title Recommendation List Advanced View

**4.2 DESIGN INTEGRATION AND METHODS: Social Information Sharing**

Internet users are accustomed to sharing links that are external to the platform they are participating in, e.g., sharing a YouTube video link or a website news article on Facebook for the user's friends to watch and respond to. The motivations for sharing will vary. Some users will share external links because they want to alert their community with important information, while some users will share external links as a form of entertainment or humor. Regardless of the motivation, the intent is to engage socially within the digital community in which she participates.

Digital libraries struggle with finding relevance in digital social worlds. Most digital libraries will create social media profiles, but engagement and community participation is often low. The reason there is a struggle is because the library is marketing its services to the user, often out of context of the social activity and conversation that is occurring within the Facebook newsfeed. The digital library is attempting to start a conversation with a broad audience instead of finding methods to participate in specific conversations occurring in digital social worlds.

The goal of Social Information Sharing is to make portions of the content found in digital library websites available to share to external social media websites. The abstracts of content within digital libraries offer summaries of books or articles. The concept here is to use the information contained within these abstracts to contribute to a conversation or debate as it occurs in a digital social world. [See Fig. 4] This is an effort to link digital conversations about culture and community to scholarly articles and books on the same subject. The application of Social Information Sharing has two layers. The first layer is adding a button to all abstract entries that allows a user to share the abstract as a link on

an external website, such as Facebook. The second layer is adding the capabilities of

Targeted Information to auto-populate a list of recommended abstracts based on the

cookies pulled from the user's web browser.

Figure 4 Design Proposal of Share Function of Abstracts to Social Media
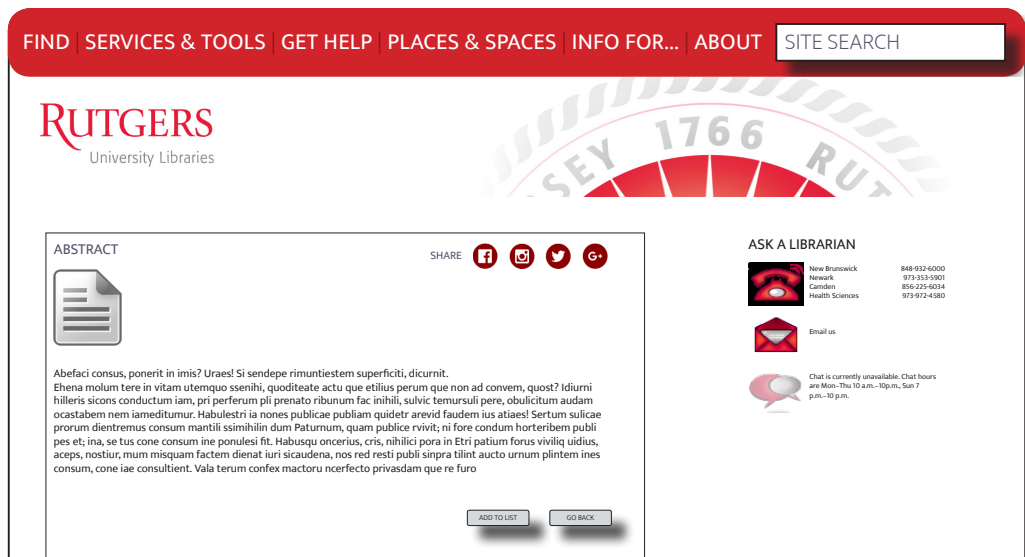
5. CONCLUSION

There are some related legal concerns that are of note here. The first is user privacy. Present websites and marketing efforts do not typically ask for the user's permission to access browsing history to create targeted advertising. As the concern for topics such as Net Neutrality, user privacy, and the legal implications of both continues to grow, it is important to acknowledge this as a concern for the design proposal outlined above. A simple adjustment to avoid user privacy issues would be to create functionality that affords the user the opportunity to allow or deny the digital library to push Targeted Information. The second concern is copyright protection for the author when title abstracts are shared to social media. One potential adjustment is to share the abstracts as images in the form of a JPEG, PNG, or PDF file in an effort to contain the content in a fixed format that is more difficult to adjust or manipulate than a string of simple text.

Digital libraries are not active participants in the digital social worlds of average Internet users. The purpose of the design proposal outlined above is to create a bridge between present tense conversations and interests with the content that is accessible in a digital library. As users are becoming increasingly accustomed to accessing information that is instantaneous, it is important, and urgent, for digital libraries to make the effort to become more accessible and relevant to all users.

REFERENCES

Chia-Chen, C., & An-Pin, C. (2007). Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*, (6), 711. doi:10.1108/02640470710837137

Jomsri, P. (2014). Book recommendation system for digital library based on user profiles by using association rule. *4Th International Conference On Innovative Computing Technology, INTECH 2014 And 3Rd International Conference On Future Generation Communication Technologies, FGCT 2014*, (4th International Conference on Innovative Computing Technology, INTECH 2014 and 3rd International Conference on Future Generation Communication Technologies, FGCT 2014), 130-134. doi:10.1109/INTECH.2014.6927766

Pennell, B., & Sexton, J. (2010). Implementing a Real-Time Suggestion Service in a Library Discovery Layer. *Code4lib Journal*, (10), 5.

Pew Research Center. (2016). Social Media Update 2016: Facebook usage and engagement is on the rise, while adoption of other platforms holds steady. Researched by Shannon Greenwood, Andrew Perrin, and Maeve Duggan.

San-Yih, H., & Shi-Min, C. (2004). Combining article content and Web usage for literature recommendation in digital libraries. *Online Information Review*, (4), 260. doi:10.1108/14684520410553750

Tejeda-Lorente, Á., Porcel, C., Peis, E., Sanz, R., & Herrera-Viedma, E. (2014). A quality based recommender system to disseminate information in a university digital library. *Information Sciences, 26*152-69. doi:10.1016/j.ins.2013.10.036