

A Conceptual Data Mining Model (DMM) used in Selective Dissemination of Information (SDI): a case study of Strathmore University library

Mr. Ambayo Jackson Alunga

Assistant Lecturer, The Technical University of Kenya

Email: jalungar@gmail.com

Prof. Ismail Ateya Lukandu

Academic and Research Director, Strathmore University

Email: iateya@strathmore.edu

Abstract

Rationale - The process of locating and acquiring relevant information from libraries is getting more complicated due to the vast amount of information resources one has to plough through. To serve users purposefully, an academic library should be able to avail to users the tools and services that lessen the task of searching for information.

Design - The research proposed a two-phase data mining through analysing the access behaviour of users. In the first phase, the Ant Colony Clustering Algorithm was used as the data mining method and separated users into several clusters depending on access records used. The clusters were in the form of course groupings. Users who have similar interests and behaviour were collected in the same cluster. In the second phase, the user records in the same cluster were analysed further. The second phase relied on association which was used to discover the relationship between users and information resources, users' interests and their information access behaviour.

Findings - It was ascertained that although users were able to locate and retrieve the information they needed, it was not up to the degree of satisfaction they expected. Furthermore, it took them some time to acquire the information. Using data mining together with selective dissemination of information would enable users to access relevant information without promptly thus saving time and other resources.

Practical implications - The mining of user data within library databases would facilitate a better understanding of user needs and requirements leading to the development and delivery of specialised and more fulfilling services.

Originality - The proposed DMM model is original as it is one of a kind that suggests integrating SDI with data mining in libraries.

Keywords

Data mining, bibliomining, Selective Dissemination of Information (SDI), information needs, knowledge discovery in databases (KDD), academic libraries

Citation: Alunga, J.A. and Lukandu, I.A. (2016). A Conceptual Data Mining Model (DMM) used in Selective Dissemination of Information (SDI): a case study of Strathmore University library. *Regional Journal of Information and Knowledge Management*, 1 (2),12-21.

**Published by the Regional Institute of Information and Knowledge Management
P.O. Box 24358 – 00100 – Nairobi, Kenya**

1 Introduction

Information services such as Selective Dissemination of Information (SDI) and current awareness have been available to make the process of information searching and location in libraries easier. But these services stress on the documents and not the information content. This document-led view of manual systems has several limitations which have led the stakeholders to seek other options. The search for options has turned attention to the potential of information and communication technologies (ICTs) in academic libraries resulting in their increased application. The increasing use of the Internet and other digital media has led to the burgeoning of library information content and resources.

This expanded view of libraries leads to a number of insights: First, libraries are typically embedded within larger institutions. Corporate libraries serve their corporations; academic libraries serve their universities; and public libraries which are run by representatives serve taxpaying communities. Secondly, libraries play a pivotal role within their institutions as repositories and providers of information resources. In the provider role, libraries represent in microcosm the intellectual and learning activities of the people who comprise the institution. These facts provide the basis for the strategic importance of a library. To be more effective, an information service system is needed that can identify user needs and how well those needs are served. Such a service exists in bibliomining which reveals insights that have meaning in the context of the library's host institution. This in turn leads to the acquisition of relevant information resources that are targeted at the users' needs.

To be able to understand the concept of data mining the term "data" is first of all looked at as defined by Laudon and Laudon (2007:14) who describe it as "streams of raw facts representing events occurring in organisations or the physical environment before they have been organised and arranged into a form that people can effectively understand and use." In a library setting the user's name, identification number, and information resource's class number are examples of data. The term mining is perceived as the excavation of data. As pointed out by Connolly and Begg (2005:1245), data mining is described as "the process of extracting valid, previously unknown, comprehensible, and actionable information from large databases and using it to make crucial business decisions". Roiger and Geatz (2003) have defined data mining as the process of employing one or more computer learning techniques to automatically analyse and extract knowledge from data contained within a database. Data mining in libraries has been discussed by researchers in the context of Web mining. Researchers such as Banerjee (1998) have explored the means by which data mining can help the library. In discussing possible applications, he explains that full-text dynamically changing databases tend to be better suited to data mining technologies. As the Web is a full-text, dynamically changing database, it is indeed appropriate to use these technologies to analyse it. In addition researchers such as Chau (1999) have explored Web mining (data mining on the World Wide Web) as a tool to help the user find information online. Not only can Web mining be used to create better search tools, but it can also be used to track the searching behaviour of users. By tracking this information, librarians can create better Web sites and reference tools (Chau, 1999).

2 Rationale of study

The abundance of information makes it difficult for a user to be able to locate relevant information in time or even know of its existence. Searching for information is in itself a difficult activity. This fact has been observed by Chen and Chen (2006) in their research that concluded that since library collections have been increasing day by day, it is difficult for readers to find the books which interest them as well as representative booklists. Yet, this should not be so. Inadequate resources have reduced library staffs' ability to contact patrons about the latest and relevant or delinquent information materials.

The importance of a well-equipped academic library and its role in research cannot be underestimated. According to Chen, Lin and Wu (2004), even a well-stocked academic library can face serious challenges hampering the effective use of the information resources contained within it.

3 Methodology

Through this study, the researchers sought to determine the information needs of the academic library users; design a user profile that can be used for selective dissemination of information; propose a data mining technique which supports the selective dissemination of information process; and design a conceptual data mining model to be used in selective dissemination of information in a university library.

A case study approach was adopted for the collection and analysis of data. Random sampling technique facilitated the choice of 100 library users and library staff from whom data was collected using a combination of self

administered and researcher administered questionnaires. Additional data was collected through a review of relevant literature. Data was analysed and presented using descriptive statistics, cross tabulations and graphs, by the use of Statistical Package for Social Sciences (SPSS) version 12.

4 Findings and discussions

The findings of the study are presented and discussed hereunder:

Information needs of academic library users

It was ascertained that there is a demand for relevant information amongst the library clientele with academic, financial and employment information being the highest sought. Figure 1 below summarises the information needs identified by the respondents.

It is evident from the foregoing that the majority (99%) of the respondents needed information for academic purposes. The other important information needs identified by the respondents include research (97%), financial information (94%), and job/employment (63%). The prioritisation of academic and research information is understandable because learning and research are the main reasons academic library users visit their libraries. The need for financial information is due to the fact that most of the students in Strathmore University are pursuing business related courses. Employment information is also important because of the high rate of unemployment in Kenya.

Information seeking methods

The preferred information access pathways were: OPAC, lecturer referrals, classmate referral, reading list and use of shelf lists. The

majority (91%) of the respondents relied on the library's online catalogues to identify and locate information materials. A cross tabulation of information seeking methods and success rate revealed that OPAC and classmate referrals were the most highly used information seeking mechanisms. The use of librarians and shelf lists did not register strongly with library users when it came to their use in the search for

information materials within the library. The users preferred to have a system that recommends to them the availability of relevant information materials. They also desired to have the ability to do their own searching through the automated library catalogue which scored the highest when it came to the relevancy of resources sought through it.

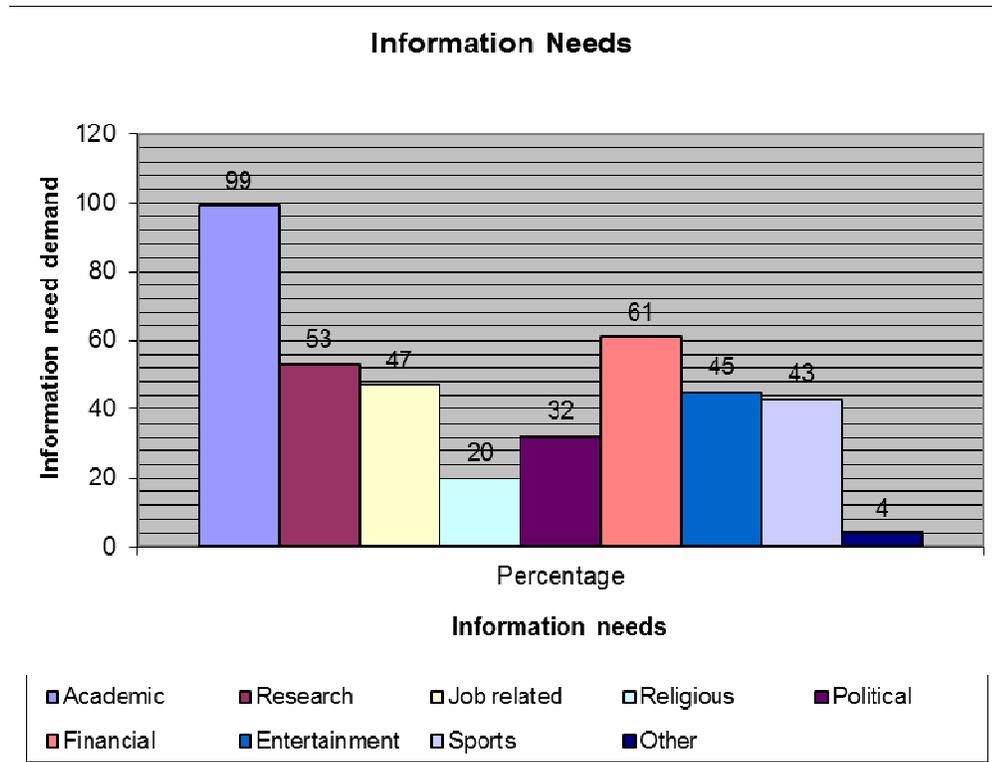


Figure 1 – Information needs

Information services

The knowledge and use of information services at the library included reference, photocopying, printing, lending and short loan. The use of current awareness services and SDI which would notify users on the existence of relevant and latest information resource availability registered marginal use and knowledge. Strathmore University library uses Koha - an integrated library system (ILS) and the first open source ILS. The library uses the latest Koha

version 3.0.0 that runs on Linux. The client side OPAC interface is written in XHTML whereas the librarians' interface has the same but requires a JavaScript enabled browser to run properly.

User profiles

Strathmore University library does have their users' profile contained within a database. However, the actual profiling is done through the grouping of individuals according to the course they are taking. The profiling has a collection of all the resources that have been

used by a particular user. This takes care of the user's borrowing record. It was established that the library uses this user profile to match user needs to the available library resources. However, the method was not efficient in identifying these needs accurately. Therefore, SDI was not effectively used to dispatch relevant information to the patrons thereby hindering its impact. Nonetheless, the existing mechanism enabled the setting up of a system that could be updated regularly by the user in terms of requesting for resources and inputting of data on the interests that could be helpful in identifying their needs. It is in this line that the researchers propose a data mining tool that can match the user profile to the information about resources contained within the database, and, selectively disseminating the relevant information to the appropriate individual users.

The research findings established that there were information needs that are not adequately catered for by the information services available within the Strathmore University library. As for those needs that are catered for, there is too much information that is to be sifted through. This sifting is a time consuming exercise for the majority of patrons who sometimes end up not having acquired the information they needed. It is in this response that a system that could be applied in the matching of information needs and the available resources within the library was envisioned.

5 Recommendations

A data mining model that can be used in the selective dissemination of information is hereby proposed. The purpose is to link the users' information needs to the available and relevant information materials. This matching requires

technologies much like search engines that are specialised at rummaging through library databases and mining bibliographical entries and user details to come up with what can be the closest to determining and anticipating information use patterns and demands within libraries.

In view of the research findings that information needs of the library patrons do exist and having established what they are, the preferred means of seeking information and taking into consideration the library system in use, the suggestions hereunder can go a long way in establishing the proposed model.

User profile

The user profile will consist of the keywords that characterise the library user. The profile indicates the name, registration number, reading history, contact details, topics of interest and areas of interest that are not academic. The library already keeps details of users in an electronic form. Therefore, there is no need to design a new user profile since all the elements needed for the construction of a user profile are contained within the personal details records.

Selective Dissemination of Information

The library can embrace SDI as a current awareness service. This entails informing users of the presence of information materials that are relevant to them through e-mail. Already the library has user profiles in terms of groups and individuals who can make suggestions on the kind of information materials they would like to see in the library. This enables the library to not only tell what a group needs but also the interests of a particular individual. For the SDI service to be effective, it requires the support of

a user profiling service that is system generated and modified by the user.

Data Mining Technique to be used in SDI

The research proposed a two-phase data mining approach through analysing the access behaviour of users. In the first phase, the Ant Colony Clustering Algorithm is applied as the data mining method. It is used to separate users into several clusters depending on access records used. The clusters were in the form of course groupings in which users who have

similar interests and behaviour were collected in the same cluster. In the second phase, the user records in the same cluster were analysed further. The second phase relied on association which was used to discover the relationship between users and information resources, users' interests and their access behaviour. To effectively maximise on the process of providing relevant information needs to users, the two processes (SDI and data mining) are combined to form a single service. The model of this service is as indicated in Figure 2 below.

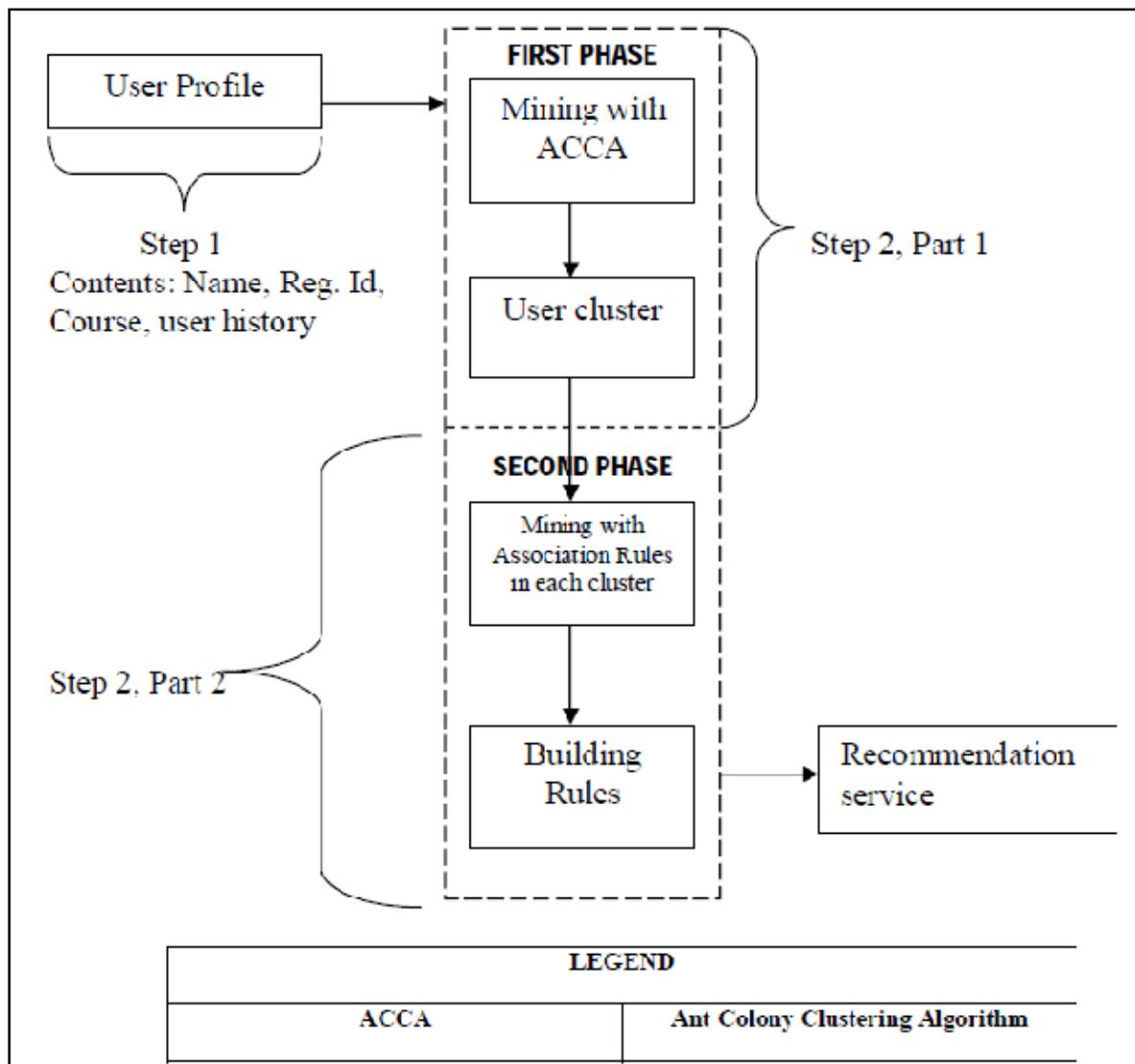


Figure 2 - SDI Data Mining Model

User Profiling

The user interest profile is a list of index terms that indicates the area of user's interests, needs, intentions and the different specificities that characterise the user. This way, the user profile constitutes a determining element to improve the relevance of the answers during a search session within a database. The modelling of the user profile and the way to adapt it to different users who do not have a precise idea of the information they seek, enables a personalised access to the contents of scientific documents. The user profile is composed of attributes containing the user preferences associated to values that are obtained from the user's web page. This information contains the user's name, course being undertaken, registration number and the reading history. It is in this instance that individuals are also asked to give their recommendations on books and other information materials they would like to have in the library.

The process of mining with Ant Colony Clustering Algorithm is then employed to separate users into several clusters depending on access records held by the library in the user profile. Users who have similar interests and behaviour are collected in the same cluster (the cluster could be course taken, for instance,

BBIT, BCOM, MCOM, ACCA, MSc, CBIS or area of research being undertaken). The interest could be in the form of course taken, information needs category, interests and hobbies, which, in the case of a single user are all integrated to form a single user profile. A single user profile has an individual's name, course taken, registration number, contact details, hobbies and reading history.

The first step is to initialise the parameters. A set of artificial ants is positioned on the first job according to an initialisation rule. Each ant constructs its own cluster. Once the ants have completed their clusters, each cluster's variance (CV_{intra}) is calculated. The percentages of the furthest nodes are chosen to be regrouped into the cluster with the shortest distance. If the new variance (CV_0_{intra}) is smaller than CV_{intra} that means the nodes in the updated cluster are more similar than the nodes in the previous cluster. While applying new clusters, an ant simultaneously updates the amount of pheromone on its visited paths (by applying the local updating rule). After all of the ants have built solutions, the pheromone trails on the paths of the global best cluster are modified again (by applying the global updating rule) up to the current iterations. The process is terminated after predefined iterations, as shown in Figure 3.

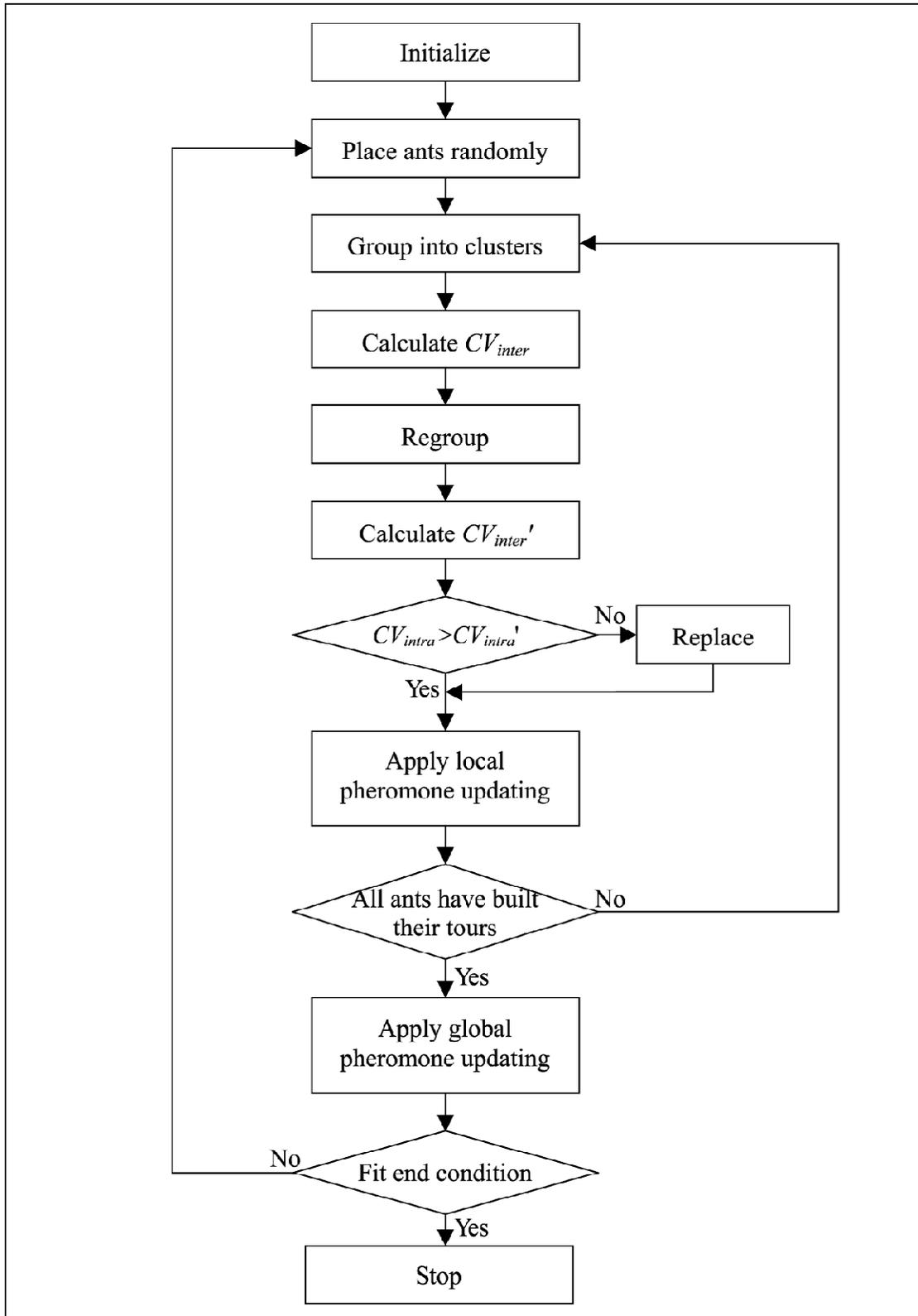


Figure 3 - Ant Colony Clustering Algorithm flow chart

Association

The second phase for the data mining method is to find out the patterns of relationships in each cluster by association rules. Association rules as the data mining method are applied to discover the associations among users' interests and information materials held within the library. Then, this is compiled to determine the user information needs.

Recommendation service

It is from these needs that any occurrence of a relevant material to the user is sent to them via the SDI service to alert them of its existence. The SDI service provides each subscriber with a periodic, tailored listing of new documents that have been entered into the file and that are likely to apply to the user's interest as a professional. The means of informing users about the existence of relevant information materials is through their electronic mail that is contained within their user profile. The

recommendation system architecture is shown in Figure 4. On logging into the respective electronic mailboxes, the user finds a notification detailing the existence of an information resource that is of relevance to them. The user then has a choice of reserving the document through the library website or not.

6 Conclusion

The ability of libraries to effectively serve their clients relies on their ability to know what the patrons' needs are and how to meet them. The proposed data mining and SDI model is a tool that can aid in the achievement of such a task.

7 Future research

As the title of the research paper implies this is a study towards the design of a model. Its practicality can be determined by the actualisation of such a system. It is in this that the researchers propose and encourage a development of the system in future.

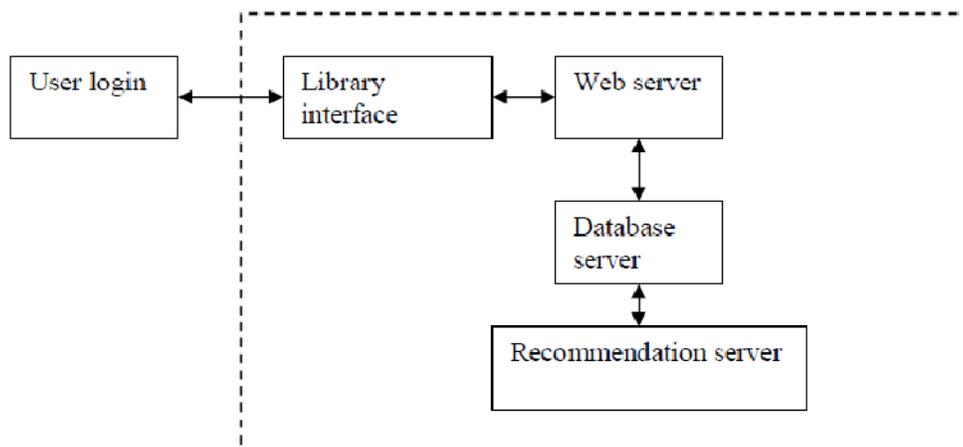


Figure 4 - Recommendation Service: email approach

8 Implications of findings

Though data mining has been used successfully for several years in the scientific and business communities for tracking the behaviour of individuals and groups, processing medical information, and a number of other applications, its use in libraries is limited. Data mining offers two major advantages to libraries: 1) it can provide faster and more thorough access to materials than manual cataloguing because user profiles' help determine user needs that are matched to the available resources; and 2) it can be used by employees or users with basic computer and analytical skills; so people can more easily find what they need without requiring the assistance of highly skilled staff.

9 References

Banerjee, K. (1998). Is data mining right for your library? *Computers in Libraries* 18(10):28-31.
Chau, M. (1999). Web mining technology and academic librarianship: human-machine

connections for the twenty-first century. *First Monday* 4(6) Retrieved 27 March 2017 from: http://www.firstmonday.dk/issues/issue4_6/cha.

Chen, C. & Chen, A. (2006). Using data mining technology to provide a recommendation service in the digital library. Retrieved 27 March 2016 from <http://www.emeraldinsight.com/0264-0473.htm>.

Connolly, T. & Begg, C. (2005). *Database systems; a practical approach to design, implementation, and management*. Harlow, England: Pearson Education Ltd.

Laudon, K. & Laudon, J. (2007). *Management information systems: managing the digital firm*. Upper Saddle River, NJ.: Pearson Education.

Roiger, R. & Geatz, M. (2003). *Data mining: a tutorial-based primer*. Boston: Pearson education.