The importance of data mining, user information behaviour and interaction audit for information literacy

Patrick Ajibade and Ndakasharwa Muchaonyerwa

Introduction

One of the important roles of the library and any librarian is to facilitate information discovery and often perform information literacy training. Because most academic libraries' data reside in the cloud and databases, data mining skills must be embedded in library training. Basic data mining, text mining, data analytics techniques as well as programming language skills are becoming prominent in the curricula of most library and information science (LIS) schools in the developed world and is a must for future librarians.

One of the critical roles of any librarian is to provide useful information to the library manager for an effective communication and decision-making purposes. The interaction audit of the users and their information behaviour is important to gather this information. These insights cannot be manually generated in a library that services over 20,000 users. It will require the ability to understand information behaviour and to provide data analysis and data analytics and text mining knowledge. The aid of a programming language will be essential because some of the data that we gather will be in the form of text and words; therefore, text mining and sentimental analysis are very vital.

Data mining literacy

Data literacy is the ability to use computational knowledge and soft skills to locate, aggregate and disseminate information. Using programs and software or cloud-based services is vital to distinguished librarians as knowledge workers. All information specialists need to be able to retrieve any data information or data set. The ability to use emerging technologies will only increase efficiency, turnaround time and the ability to increase the precision ratio of the information and the volume. Data mining in this knowledge society era is an additional ability to further information literacy within the library context (Berendt, 2012). Because it provides the ability to mine data and text through knowledge discovery (Olmer, 2008). The library has consistently adopted data mining (Katsurai and Joo, 2021). It was further discussed that this technology and skill is vital in developing knowledge products (Haravu and Neelameghan, 2003) and for knowledge organisations. Besides the use of data mining for knowledge discovery is mostly applied to bibliometric analysis within the LIS field (Jayasekara and Abu, 2018). It has been discussed that student online interaction can be understood by mining such information from various sources (He, 2013) to provide library management intelligence for decision-making through integrated mining exercises (Li and Wu, 2006). In most developing countries libraries with limited resources and financial capabilities to expend on large collection to devise means to minimise wasteful expenditure. Even the libraries in the developed world could limit the potential information overload by not curating items and resources that are not often consulted or used.

Information audits

The ability of librarians to understand the nature of problems either from the system audit perspective or information perspective enables them to be agile in daily operation. Information system audits is often conducted within the banking sector because of the sensitive nature of the information being processed (Wahyudi and Deswandi, 2016). Nevertheless, the same concept could be useful in the library and information service sector.

Just as blockchain technology provides immutability to records (Ahmad *et al.*, 2021), information and system audits provide solid foundation for an agile information governance. And in the context of library, provincial library information systems have been documented through audits (Pradana *et al.*, 2019). System audit and analysis would also help to understand system functions and how an individual interacts with the system. All these analyses could be done separately without any recourse to actual user data. The sentiment analysis could be done independently by using the data and text mining to perform the sentiment analysis.

Sentiment analysis

Recently, data mining is one of the techniques used to discover some of the information that exists within databases. One of the advantages of data mining is the potential to validate knowledge discovery (Gul *et al.*, 2021). It provides a logical pattern and actionable intelligence for the library services to understand large set of textual data. One of the attributes of the online or virtual learning space have in common with the library management systems is the online environment in which these infrastructures are built to facilitate knowledge and information discovery. Text mining has recently been used to

support massive online learning platforms (Dina *et al*., 2021). Because of large volume, libraries must be competent in big data collection, technologically enhanced searching and information retrieval.

Brahimi *et al*. (2016) note that data mining can help understand chat feeds (and Twitter feeds) polarity, the library can use the same techniques to understand users' sentiments. Text mining has been found to be useful in various contexts and fields of study. For example, the Python programming language was used to analyse social networks (Vidya Kumari and Kavitha, 2018), for opinion (Dhanalakshmi *et al*., 2016) or information mining (Gopal *et al*., 2011). The techniques have been used to study a social media platform (Li *et al*., 2020), just as the library could use it to study user behaviour. Depending on the use and context of research, mining student or faculty member opinion regarding the library information resources (Mathimagal and Jayalakshmi, 2021) could be the next innovative approach in achieving optimal resource usage and return on investment. The analysis of users, user interaction and content engagement because of user-generated content (UGC) could help the collection and acquisition department to strategically modify their strategies.

The library must also shift its thinking paradigm and begin to reason as a business entity that is looking to improve return on investment *vis-à-vis* content acquisition, engagement and utilisation. This strategy could also help the library understand any shift in training performance and user's perception (Noferesti and Shamsfard, 2016). It will enable the library to generate performance indicators (Saura *et al*., 2019) for the library services through the user UGC data (Pavaloaia *et al*., 2019). Irrespective of how valuable the application of sentiment analysis in the library context is, the ability to mine and contextualise a large set of data will be critical for an efficient information services. Information sensing (Ajibade, 2016) and big data analytics skill would enable the information librarian (Ajibade, 2016) to enhance their knowledge-sharing ability (Ajibade *et al*., 2019). Because most libraries now have social media accounts, performing sentiment analysis would be beneficial, just as Gupta *et al*. (2021) strategies to analyse TikTok user content. There are many techniques and programming languages that can mine quantitative and qualitative data. Python might be one of the beginner's friendly choice. Surya Gunawan *et al*. (2020) used the program to perform public information analysis.

Mining public sentiments has been done before (Yan *et al*., 2020) and this can be used in the library context to find out what information users are mostly concerned with and the traffic and duration of time spent could help the library. This information could help them plan information training services or use such for collection development purposes. An information librarian could open a library training Twitter feed, collect real-time reaction and collate questions during training. The ability to quickly gather data and concerns could help the librarian resolve some of the participants' concerns. Albayrak and Gray-Roncal (2019) have explored this technique in another context to predict an event or monitor real-time data. This process will reduce the likelihood of unsatisfactory training sessions. It will minimise the need for a patron to seek further training or wait till the next training session to get some of the needed information.

Web interface audit and user interaction

Information system audits could provide useful information regarding problem that might be hard to understand if this audit is not conducted. From the root source (starting point), many branches (menu and sub menus) were analysed. The information architecture of the web interface was provided from the diagnosis. Initially, this type of analysis is usually conducted by the system librarian or library management system developer to understand the performance of the LMS or any online resources. One of the ideas behind this analysis was that diagnosing system and performing information audits should be routine task. Therefore, users' interaction analysis and UGC feedback would make libraries much more efficient and tailor their information literacy training based on business intelligence from their patron's data.

Having a fast, responsive, interactive LMS is crucial to encourage continual user access because of the low bandwidth and unavailability of mobile data networks in some remote areas. This graphical information is just one of the snippets of information and insights we have generated and the ability to create more would be necessary for libraries with limited resources.

Concluding remarks

The alignment of information technology with the library management depends on key factors. Firstly, the skills requisite to handle technical and technological requirements are vital. One of the suggested programming skills is Python as its syntax could be easily understood compared with other syntax for data or text mining. The training and retraining of the librarian to handle the system and information audits must be continuously carry out. This training enables librarians to perform robust information audits and business intelligence for decision-making that enhances optimal use of resources. The insight that in-depth data provides can help the library in designing responsive information literacy training. Furthermore, the analysis could help the library identify whether lack of engagement has anything to do with how the training is conducted, the format/types of resources or the user interface that is difficult to navigate.

REFERENCES

Ahmad, D., Lutfiani, N., Ahmad, A.D.A.R., Rahardja, U. and Aini, Q. (2021), "Blockchain technology immutability framework design in E-government", *Jurnal Administrasi Publik: Public Administration Journal*, Vol. 11 No. 1, pp. 32-41.

Ajibade, P. (2016), "Validating information sensing in a South African university as an impetus to improved information management practice and performances", *Journal of Social Sciences*, Vol. 48 No. 3, pp. 225-238.

Ajibade, P., Ondari-Okemwa, E. and Matlhako, M. (2019), "Information technology integration for accelerated knowledge sharing practices: challenges and prospects for small and medium enterprises", *Problems and Perspectives in Management*, Vol. 17 No. 4, pp. 131-140.

Albayrak, M.D. and Gray-Roncal, W. (2019), "Data mining and sentiment analysis of real-time Twitter messages for monitoring and predicting events".

Berendt, B. (2012), "Data mining for information literacy", *Intelligent Systems Reference Library*, Vol. 25, pp. 265-297.

Brahimi, B., Touahria, M. and Tari, A. (2016), "Data and text mining techniques for classifying Arabic tweet polarity", *Journal of Digital Information Management*, Vol. 14 No. 1, pp. 15-25, available at: www.scopus.com/inward/record.uri?eid=2-s2.0-84969780164andpartnerID=40andmd5=57719abacdba72ad8002405b928bfc8b

Dhanalakshmi, V., Bino, D. and Saravanan, A.M. (2016), "Opinion mining from student feedback data using supervised learning algorithms".

Dina, N.Z., Yunardi, R.T. and Firdaus, A.A. (2021), "Utilizing text mining and feature-sentiment-pairs to support data-driven design automation massive open online course", *International Journal of Emerging Technologies in Learning (iJET)*, Vol. 16 No. 1, pp. 134-151, doi: 10.3991/IJET.V16I01.17095.

Gopal, R., Marsden, J.R. and Vanthienen, J. (2011), "Information mining – reflections on recent advancements and the road ahead in data, text, and media mining", *Decision Support Systems*, Vol. 51 No. 4, pp. 727-731, doi: 10.1016/j.dss.2011.01.008.

Gul, S., Bano, S. and Shah, T. (2021), "Exploring data mining: facets and emerging trends", *Digital Library Perspectives*, Vol. 37 No. 4, pp. 429-448, doi: 10.1108/DLP-08-2020-0078.

Gupta, R., Sameer, S., Muppavarapu, H., Enduri, M.K. and Anamalamudi, S. (2021), "Sentiment analysis on Zomato reviews", Paper presented at the 2021 13th International Conference on Computational Intelligence and Communication Networks (CICN).

Haravu, L.J. and Neelameghan, A. (2003), "Text mining and data mining in knowledge organization and discovery: the making of knowledge-based products", *Cataloging and Classification Quarterly*, Vol. 37 Nos 1/2, pp. 97-113, doi: 10.1300/J104v37n01_08.

He, W. (2013), "Examining students' online interaction in a live video streaming environment using data mining and text mining", *Computers in Human Behavior*, Vol. 29 No. 1, pp. 90-102, doi: 10.1016/j.chb.2012.07.020.

Jayasekara, P.K. and Abu, K.S. (2018), "Text mining of highly cited publications in data mining".

Katsurai, M. and Joo, S. (2021), "Adoption of data mining methods in the discipline of library and information science", *Journal of Library and Information Studies*, Vol. 19 No. 1, pp. 1-17, doi: 10.6182/jlis.202106_19(1).001.

Li, J., Xu, Q., Cuomo, R., Purushothaman, V. and Mackey, T. (2020), "Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study", *JMIR Public Health and Surveillance*, Vol. 6 No. 2, p. e18700.

Li, Q. and Wu, Y.F.B. (2006), "Information mining: integrating data mining and text mining for business intelligence".

Mathimagal, N. and Jayalakshmi, S. (2021), "Intellectual behaviour of student based on education data determined by opinion mining".

Noferesti, S. and Shamsfard, M. (2016), "Using data mining techniques for sentiment shifter identification".

Olmer, P. (2008), "Knowledge discovery: data and text mining".

Pavaloaia, V.D., Teodor, E.M., Fotache, D. and Danileţ, M. (2019), "Opinion mining on social media data: sentiment analysis of user preferences", *Sustainability (Switzerland)*, Vol. 11 No. 16, p. 4459, doi: 10.3390/su11164459.

Pradana, H.A., Gede Putra Pertama, P.P., Sujono, Mayasari, M.S., Andrika, Y., Juniawan, F.P. and Sylfania, D.Y. (2019), "Audit of provincial library information system based on COBIT 4.1".

Saura, J.R., Palos-Sanchez, P. and Grilo, A. (2019), "Detecting indicators for startup business success: sentiment analysis using text data mining", *Sustainability (Switzerland)*, Vol. 11 No. 3, p. 917, doi: 10.3390/su11030917.

Surya Gunawan, T., Aleah Jehan Abdullah, N., Kartiwi, M. and Ihsanto, E. (2020), "Social network analysis using Python data mining".

Vidya Kumari, K.R. and Kavitha, C.R. (2018), "Data mining for the social awareness of the social networks".

Wahyudi, M. and Deswandi, A. (2016), "Audit information systems core banking system using ITIL V.3 case study on BTPN Sharia bank", *Journal of Theoretical and Applied Information Technology*, Vol. 87 No. 1, pp. 38-46, available at: www.scopus.com/inward/record.uri?eid=2-s2.0-84966930091andpartnerID=40andmd5=2ad00f34106b339a2d4d90bf4b8094c8

Yan, Y., Chen, J. and Wang, Z. (2020), "Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations", *Applied Geography*, Vol. 123, p. 102306, doi: 10.1016/j.apgeog.2020.102306.

FURTHER READING

Ajibade, P. and Mutula, S.M. (2020), "Big data research outputs in the library and information science: South African's contribution using bibliometric study of knowledge production", *African Journal of Library, Archives and Information Science*, Vol. 30 No. 1, pp. 49-60.

Balagué, N., Düren, P., Juntunen, A. and Saarti, J. (2014), "Quality audits as a tool for quality improvement in selected European higher education libraries", *The Journal of Academic Librarianship*, Vol. 40 No. 5, pp. 529-533, doi: 10.1016/j.acalib.2014.01.002.

Bashorun, M.T., Olarongbe, S.A., Bashorun, R.B. and Akinbowale, A.T. (2022), "Information literacy competence and use of electronic information resources among undergraduates in university of Ilorin, Nigeria", *Mousaion: South African Journal of Information Studies*, Vol. 40 No. 1, p. 5.

Granado, K.C., Romero, C.O.R. and Rodríguez, M.E.C. (2015), "Information audit as a management tool at the library in the University of Sancti Spiritus", *Revista Cubana de Informacion en Ciencias de la Salud*, Vol. 26 No. 2, pp. 107-124, available at: www.scopus.com/inward/record.uri?eid=2-s2.0-84929396032andpartnerID=40andmd5=860830506e6b6250119b4442136abeb8

Kang'ethe, S. and Ajibade, P. (2016), "Validating the fact that effective information packaging and dissemination is a strong tool to mitigate the effects of HIV/AIDS in selected African countries", *Journal of Human Ecology*, Vol. 55 No. 3, pp. 221-226.

Suwarjo, S., Haryanto, H., Wuryandani, W., Mahfuzah, A., Hidayah, R. and Erviana, V.Y. (2022), "Digital literacy analysis of elementary school teachers on distance learning instructional process in Yogyakarta", *AL-ISHLAH: Jurnal Pendidikan*, Vol. 14 No. 2, pp. 1145-1156.

Patrick Ajibade (*ajibadep.sa@gmail.com*) is based at the Department of Information Studies, University of Fort Hare, Alice, South Africa.

Ndakasharwa Muchaonyerwa (*NMuchaonyerwa@ufh.ac.za*) is based at the Department of Library and Information Science, Faculty of Social Sciences and Humanities, University of Fort Hare Alice, South Africa.