E-book data mining: real information behavior of university academic community

Azra Rafique
National Library of Engineering Sciences, University of Engineering and Technology, Lahore, Pakistan and
The Department of Information Management, University of the Punjab Quaid-i-Azam Campus, Lahore, Pakistan, and
Kanwal Ameen and Alia Arshad
Department of Information Management, University of the Punjab, Lahore, Pakistan

Received 21 April 2020
Revised 25 January 2021
Accepted 1 March 2021

Abstract

Purpose – The purpose of this study is to examine the evidence-based use patterns of Higher Education Commission (HEC) subscribed e-books databases by the academic community at institutions of higher education in Pakistan. The study also investigates the differences in usage based on points of access, scholarly disciplines and gender of users.

Design/methodology/approach – A transaction log analysis (TLA) method was used to explore the use patterns of HEC e-books databases at the University of the Punjab, utilizing the SAWMILL analytical software and MS Excel.

Findings – The results of the study showed that the use of e-books was at a growing stage at the university. Male boarding students of the university were more active users of e-books as compared to their female counterparts. The Central Library of the University of the Punjab, the Department of Zoology and the Institute of Communication Studies were the most frequently used access points for the e-book users, and the Faculties of Sciences, Life Sciences, Economics and Management Sciences had the most active e-book readers. Furthermore, it was found that the HEC e-books databases were only being used during the four months of summer vacations at the university.

Research limitations/implications – The study provides only descriptive use frequencies rather than a deep log analysis of e-books usage.

Practical implications – This research provides important practical implications for examining the evidence-based use patterns of e-books databases' users at the higher educational level. The research suggests that HEC should maintain subscriptions of the most required e-books databases and that the information professionals should conduct orientations and information literacy programs to enhance the utilization of these subscribed databases among female boarder students and those faculties where they were being used less frequently.

Originality/value – The study is the second part of the first phase of a Ph.D. project. This is the first large scale study conducted in a developing country which reports the close to actual, approximate use patterns of e-books based on raw transaction logs of local cache servers at the higher academic level.

Keywords Electronic books, Log analysis, Data mining, Use patterns, HEC e-books databases

Paper type Research paper

Introduction

The rapid changes in information and communication technology (ICT) have revolutionized every field of life. These changes have forced the libraries to transform their information material from paper to electronic format by either digitizing collections or subscribing to

The authors would like to thank Network Administrators, Muhammad Hamid and Rana Imran Nazir (PU, Lahore) in providing the log data. The authors would also appreciate the IT expert Fahad Islam (Karachi) for his continuous help in the log analyzing process and handling the compatibility and data analysis issues in the process to complete the first phase of the authors' Ph.D. project.

Library Hi Tech
Vol. 41 No. 2, 2023
pp. 413-431

0737-8831
DOI 10.1108/LHT-07-2020-0176

online resources. These steps have also enabled libraries to solve their space issues and to improve their services according the users' expectations regarding 24/7 access and access to information at their door-steps. Generally, a major portion of a library collection consists of physical books. E-books not only help in overcoming the physical space limitations of a library, they can also be used by multiple users simultaneously and enable remote access to information. Tenopir (2003) reviewed over 200 international studies conducted during 1995–2003 on electronic library resources, including e-journals, e-books, digital/online libraries, websites and the Internet. She reported that there had been a rapid increase in the use of electronic resources all over the world, and that e-books were rapidly becoming one of the most important electronic information resources.

The operational definition of an e-book for this study is "a book in electronic format available online through the internet". Nicholas (2008) opined that if library professionals did not monitor the usage behavior of their virtual users, they would not be able to provide them access to the required e-resources and services. Becker (2015) presented that while e-book research had become a popular topic for research, there was much confusion and conflict among researchers due to the widespread variations found among research studies on e-books. The author advised that it was, therefore, better to make "major collection decisions" based on one's own institution's recent data. Furthermore, Becker (2015) considered the year 2010 as "year zero" for e-books as that was the year Amazon's Kindle e-reader became a bestselling product and e-books gained mainstream success worldwide.

Literature review revealed that numerous studies have been conducted to investigate the phenomenon of e-books by utilizing online surveys, interviews and focus groups etc. These studies have, however, reported contradictory results. While some note an increase in the use of e-books, others have reported a minimal use of e-books among library users (Anuradha and Usha, 2006; Jamali *et al.*, 2009; Lonsdale and Armstrong, 2010; Nicholas *et al.*, 2008; Noorhidawati and Gibb, 2008; Rowlands *et al.*, 2007; Yalman, 2015). Cassidy *et al.* (2012) conducted a survey and found that the graduate students and faculty members at the Sam Houston State University (SHSU) in East Texas had a lower preference for e-books when compared to print books. Several studies conducted through Log Analysis method to explore the use of e-books (Ahmad, 2015; Ahmad and Brogan, 2012; Ahmad *et al.*, 2014; Connaway, 2001; Freeman and Saunders, 2016; Goodwin, 2014; Link, 2012; Mckay and Buchanan, 2016; Nicholas *et al.*, 2008) have found encouraging developments in the use of e-books. To identify the digital use behaviors of users, Nicholas *et al.* (2014) found logs to be useful in identifying the digital use behavior of academic researchers as well as being good indicators of the information seeking process of users and their level of use of search engines like Google and Google Scholar as discovery platforms.

On the other hand, in the Pakistani scenario, little attention has been paid to investigating the use of e-books. Jadoon *et al.* (2011) highlighted that there was moderate confidence in the process of online searching and downloading of medical books and full text articles among self-reporting medical students. Khan *et al.* (2016) investigated the use of e-books among agricultural, engineering and social science university students in Khyber PakhtunKhwa, Pakistan, whereas, Rafiq (2014) and Rafiq and Warraich (2016) investigated the use of e-books among medical undergraduate students in Lahore. These studies reported little or growing use of open access e-books among medical students. The literature review also exposed the dearth of research done on the use of e-book databases, particularly on the evidence-based actual use of e-books databases in the sub-continent and developing countries like Pakistan. The usage levels of different individual e-books databases have been investigated through log analysis method in various developed countries. However, this kind of technical analysis has rarely been undertaken in Pakistan and other developing countries, where most studies on e-books have relied on self-reporting by respondents through surveys, focus groups or interviews. Only a few studies were found to be evidence based using log

analysis method, and most were project-based. Thus, the local and developing countries' scenario was lacking evidence-based studies on the use of e-book databases which could be used for collection management and improving services in libraries.

The purpose of this study was to identify the evidence-based use patterns of Higher Education Commission (HEC) subscribed e-books databases. The patterns investigated through transaction log analysis (TLA) included variations based on differences in access points, scholarly communities and gender of users. The usage patterns of various e-books databases based on real statistics are crucial for their subscription management. It was hoped that the log analysis of HEC subscribed e-books databases would reveal a reasonable approximation of their actual use patterns prevailing in the University of the Punjab, Pakistan; a university which has been ranked second in the overall top ten universities ranking in Pakistan in the year 2015 (HEC, 2016). The results might be helpful for HEC in making decisions regarding budget allocation and access management of its e-books databases. The results may be helpful for library professionals in their future planning and coordination with HEC for their own collection development and user service facilitation with respect to e-books.

This study aims to answer the following research questions:

RQ1. Which are the frequently used HEC e-books databases by the community at the University of the Punjab?

RQ2. What are the accessing points and use frequencies of HEC e-books databases?

RQ3. What are differences in the e-book use patterns among various scholarly communities?

RQ4. What is the gender-based difference in the e-book use patterns among boarding students?

Literature review

In depth literature review was conducted for the purpose of the study. HEC ebrary (219 journal articles), Emerald Insight with other scholarly journal databases and Google Scholar were explored using various terms like "e-books use", "log analysis + e-books", "ebooks". Key studies on e-books' use in international and local scenario were investigated. HEC Summon brought 80 scholarly journal articles on e-books usage during the last five years.

Increasing use of e-books was reported by Connaway (2001) while describing the netLibrary e-book model. This model was described keeping in view the eight important evaluation elements for academic e-book usage identified by the "e-book task force for the University of California digital library". In a survey through e-mail, Rowlands *et al.* (2007) assessed the academic users' awareness levels and perceptions regarding e-books, as well as their existing e-books usage. The researchers found a highly patterned book discovery behavior, using various searching strategies at different times to meet various information needs. Moreover, forty eight percent of the users preferred reading on screen and the availability and ease of use of e-books was identified as a favorite feature of e-books.

For a Joint Information Systems Committee (JISC) funded project, Nicholas *et al.* (2008) used Survey Monkey to investigate the use of e-books in general, JISC collection text and general use of library for policy makers, publishers and information managers useful for future planning. The researchers found that even though 62% of the respondents were e-books users they were dissatisfied with the number of textbooks available as e-books. Moreover, the teachers and students reported reading only parts of an e-book instead of completely reading it from start to finish. They simply skimmed the e-books to search for their required information. In continuation of this project, Jamali *et al.* (2009) reported that nearly twelve thousand respondents were using e-books because for their online accessibility, searchability, affordability, portability

and because they could be used by multiple users simultaneously. On the other hand, the researcher found that about five thousand respondents were struggling to read the books on screen and felt the need of information literacy skills instructions from the university library staff.

Under the UK JISC National e-book Observatory funded project, Lonsdale and Armstrong (2010) conducted focus groups and interviews to explore the promotion of e-textbooks in the UK. Sixty-one individuals from the library staff, students and academics were selected from fifteen universities of the UK. The researchers found that while the library professionals were using several promotional tools, they were not always achieving their targets because of a lack of a formal promotion strategy. While the publishers were actively promoting their e-resources among the library and academic staff, the students were left feeling unaware of such promotions.

Searching and viewing patterns of e-books use in the academic library of J. N. Desmarais Library at Laurentain University, Canada, were explored by Lamothe (2010). Lamothe noted that the library's e-books collection size increased from a single e-book in 2002 to over 60,000 in 2008. This was the result of varied purchasing patterns employed of the library managers; from bulk purchasing to selective purchasing and then back to bulk purchasing. It was further noted that the searching and viewing of e-books increased more rapidly as compared to the library's e-books' collection size. During selective purchasing, "highest viewings per e-book and searches per e-book ratios" were recorded. Moreover, e-books usage was directly proportional to the library's collection size. Strong co-relation (0.99) appeared between searches and viewing.

Ahmad and Brogan (2012) explored the digital and reading behavior of e-books users at the Australian Academic and Research Libraries through a deep log analysis. They found that in the year 2010, only 32% of the university population was accessing the e-books and merely 20% of the users were reading one or more e-books. Similarly, another study was conducted by Ahmad *et al.* (2014) to see the difference in the information seeking behavior of computer literate users and computer illiterate users of e-books. The researchers suggested an e-book user's information seeking behavioral profile and validated it through log data analysis. They reported a significantly different information seeking behavior of e-book early adapters at the Edith Cowan University Library as compared to other users. These individuals were computer users with advanced technological skills and experience, generally known as power users. An interesting comparison of identical titles of books both in print and electronic format was conducted by Goodwin (2014) to determine the format preference of users. Use statistics from e-DSC platforms and circulation statistics of the same titles in print were compared. The study found that the users from the humanities and social sciences preferred print format of the titles greater than or equal to their electronic versions.

Potnis *et al.* (2018) investigated the effects of five personal, organizational, environmental and technical factors on use of e-books by the undergraduate students at a land-grant university in the Southern US through survey and open-ended questions. The study suggested eight activities for librarians for increasing the use of e-books. These activities included improving the image of libraries, active involvement of students in selection and management of e-books, maintaining diverse collection of e-books, e-books with students' preferred features, promotion of existing collection, user friendly website for e-books' searching, provision of student-centered experience of e-books usage and presence of expert librarian to help students in using e-books.

Eight weeks' experience of e-books use by the iSchool graduate and Ph.D. students of library and information science and information management was recorded in online diaries by Tracy (2018). The study revealed that more than seventy nine percent respondents used e-books in an Internet browser and over sixty three percent respondents felt easy or very easy to use the e-books and they just skimmed for specific information.

In another study, using the same e-storybook by preschool learners at Taiwan kindergarten schools, it was found by Weng *et al.* (2019) that experimental group had significantly higher overall engagement and learning achievement than the control group. Kim *et al.* (2019) explored three-dimensional usage of the National Library of Korea, Sejong based on four years' activity logs of actual users. The researchers found increased usage of the library especially in the months of January, February, July and August. Moreover, higher usage during the weekends as compared to weekdays and afternoon was also appeared in the results.

In an empirical study, Liu *et al.* (2020) tested and compared the "effect of informational cascades on online reading behaviour of free and paid e-books" of a Chinee's well-known online reading platform. The study found that information cascades were more prominent for subscribed e-books than for free e-books. Furthermore, online readers' choice was "significantly affected by book ranking after controlling for cumulative clicks and word-of-mouth (WOM) volume" for both subscribed and free e-books.

In the local context, studies by Rafiq (2014) and Rafiq and Warraich (2016) found greater use of open access e-books as compared to the HEC subscribed e-books among students for academic purposes. In another study, Khan *et al.* (2016) investigated the use of e-books through survey research and reported that the use of e-books among agricultural, engineering and social science students of KPK universities was quite low. Arshad and Ameen (2015) investigated the usage patterns of Punjab University Library website by using TLA. The study highlighted some useful aspects of the actual usage statistics of the website. However, these studies were not particularly focused on the usage patterns of HEC e-books.

While several studies have been conducted in developing countries, including Pakistan, on the use e-books, they have been limited to surveys or interviews. The authors could not find any study on e-books' use patterns via raw log data analysis to determine the approximately actual use patterns in developing countries like Pakistan. This study will fill the gap in local and international literature and provide approximately actual use statistics of e-books use.

HEC e-books databases

HEC has been subscribing to more than 30 e-journal databases and 42,000+ e-books costing over Rs. 900 million annually in order to improve and increase the research output of universities in Pakistan (HEC, 2013). HEC e-books can be accessed through ebrary-eContent Platform where thousands of e-books and their chapters are available to Pakistani students and researcher studying and working in the country's institutions of higher education. These e-books are related to all major disciplines of study such as engineering, medical sciences, agriculture, biomedicine, life sciences, business and economics, humanities, social science, law and astronomy. E-books by renowned publishers like McGraw-Hill, SpringerLink and Project MUSE are available on the ProQuest platform. Although, ProQuest provides access to theses, dissertation and scholarly articles as well, only usage of e-books available in the ProQuest database is included in this study.

Methodology

Investigation of the use pattern of e-books databases is conducted through data mining using TLA. Based on the quantitative results of this phase, the searching behavior of scholars while accessing the required e-resources and any issues arising during the process will be explored through interviews. The study population included all HEC e-books databases users (students, academics and staff) at the University of the Punjab. The University of the Punjab has 44,754 students (out of which 7,509 are hostel residents) enrolled on campus in about 350 academic programs. There are 1,172 teaching faculty members and 5,392 staff members in the university (PU Fact Book, 2017). The detailed raw log data acquisition and processing

have already been explained by Rafique *et al.* (2019). The definition of various terms or patterns like hits, visitors, sessions with duration and page-views are defined as under:

(1) Hits: Hit is a single action taken by a user at a website like a click on the link that corresponds to one line of data in the web server log file. (www.deep-software.com, 2021)

(2) Page view: Each webpage accessed/downloaded by the user considered as a page view.

(3) Visitors: Due to confidentiality, the IP addresses were treated as visitors or access points. A visitor might be an individual person, or a system in a public place being used by several persons (e.g. in a library or computer labs).

(4) Session: A session is the duration of the starting and ending time of a specific task on web.

(5) Duration: Duration is the total time spent during sessions by users while using the e-books.

One-year (1st February 2017 to 31st January 2018) transaction log data (size = 481 GB) from the cache servers installed at the university were taken on a bimonthly basis. While using databases' domain names as keywords through UBUNTO operating system, data related to all e-journal databases and e-books' use were segregated from other raw log data of Internet use like e-mail, social media and so on. From that extracted log data, HEC subscribed e-books databases' use data, comprising of a 6.47 MB file, were separated after being cleaned three times for parsing. Links available on the university's webpage (http://pu.edu.pk/page/show/Ebooks.html) for HEC ebrary/digital library, like ProQuest (2,855+ titles), McGraw Hill Collections (670+ titles), Project Muse (5,277 titles) and Springer e-books (7,705 titles) were used to segregate the relevant log data. The keywords "e-book" and "ebook" were used to segregate the usage statistics of e-books from other raw log data. Facebook and freely accessed e-books' usage data were also included in the segregated log data which was diverting from the objective of the study, therefore such data were cleaned.

All entries with zero second session duration were excluded from data analysis. They were excluded because even though they were most likely only a hit on a webpage or a pop-up, they could affect the true results of the study. There were 350 top level domains identified in the log analysis which were checked one by one for data cleaning. Two hundred and fifty-one domains were considered to be irrelevant (like common-place-book.com, ee-books.com and so on) and were deleted from the log data before analysis. Data cleaning process was repeated three times for objective results and for ensuring the reliability and validity of the process. Following is the typical example of the raw log data extracted out of the access log servers installed at the University of the Punjab.

The above example in Figure 1 includes the date (02/Jul/2017) and exact time (10:32:28 + 0500) of the session, the proxy of the client system/ IP/ access point (1**.**.*00.250) and the action/code (TCP_MISS/200). The method (GET) for accessing the needed information and URL (http://ebookcentral.proquest.com/auth/lib/hec-ebooks/maintenance.action?) were also part of the log data string.

To explore the frequencies of e-book databases' use, TLA was conducted with the help of an IT expert using SAWMILL log analyzing software (version 8.7.9.4). The use patterns of

02/Jul/2017:10:32:28 +0500 1**.**.*00.250 TCP_MISS/200 GET
http://ebookcentral.proquest.com/auth/lib/hec-ebooks/maintenance.action?

Figure 1.
Example of raw log data of the log servers

varied users from different scholarly faculties were identified with respect to page-views, number of sessions, session duration and number of visitors. Database accessing activity was observed in only 99 days, in which the users appeared to be using e-books databases like ProQuest e-books, HEC ebrary, Project Muse, McGraw-Hill and Springer e-books. A total of 21,011 page-views were logged spanning 2,543 sessions in 174 h. During the duration of the study, 13,000 visitors got 50,420 hits while accessing their required e-books. Entries with very low session duration like zero seconds as well as those from other domains like Amazon.com etc. were removed to avoid misleading results. The IP numbers were replaced with their respective department names or locations. This was done according to the IP addressing documentation maintained by the university's I.T. center for identifying the various access points and ensuring confidentiality of individual user/ system.

Significance of the study

There has been no previous study conducted in Pakistan that has reported actual use frequencies of HEC subscribed e-books databases so far. The results of this study might be useful for HEC decision makers in planning access management of their databases as well maintenance of e-books databases subscriptions. It might also assist information professionals in improving their services and offering information literacy or training programs for their users.

Data analysis

Overview of e-books databases use patterns

An overview might be helpful for a quick and easy understanding of HEC e-books databases usage before reporting of the results of the study.

TLA technique was used for analyzing the raw log data to explore the use patterns of HEC e-books databases. The log analysis revealed that HEC e-books databases remained active for 99 days in the year 2017–18 (from June 11, 2017 to September 24, 2017). Table 1 indicates that a total of 50,420 hits and 21,011 session events (page-views) were conducted by 1,301 visitors (systems) during 2,543 sessions over 174 h. By dividing these values with 99 active days, an average per day was determined for daily use per visitor as shown in Table 1. Out of the 1,301 users, 789 were repeat users; those who frequently repeated their visits more than once. Less than half ($n = 512$) of the users accessed the e-books databases only once and never returned. As the access log recording software had already been customized because of storage issues, therefore, the used data size and user names were excluded from recording of log data in cache servers. That is why the log files were showing the data size in 0 Bytes.

Use of e-books databases

Table 2 shows the frequencies of e-book databases usage. The results show that ProQuest was the frequently used database for using e-books. It had the largest number of ($n = 1940$)

Metrics	Total freq.	Daily avg. (99 days)	Daily use/visitor
Hits	50,420	509.3	38.88
Page views	21,011	212	16.2
Visitors	1,301	13.1	1
Size	0 B	0	0
Sessions	2,543	25.7	2
Session duration	174.21 h	01.46 h	0.08 min

Note(s): 7 days 6 h 21 min = 174.21 h = 10,453 min/99 = 106 min per day

Table 1. Overview of e-books log data analysis (99 days)

sessions lasting almost 148 h, whereas, HEC ebrary (279 sessions) and HEC digital library (158 sessions) e-books platforms/ databases were the second and third most frequently used databases. Users spent over eight hours using HEC ebrary and almost six hours on the HEC digital library database. HEC ebrary access was routed through the website of the University of the Punjab, whereas, HEC digital library database could be accessed directly through the Internet.

Accessing points and use-patterns of HEC e-books databases

The access points of e-books databases were identified through the Internet Protocol (IP) based addresses in the TLA. There were 72 access points identified, including various academic and general administrative departments within the university as well as the male and female students' hostels.

Different accessing points of e-books databases. The results of the study depicted that the Central Library of the university remained the most prominent access point in terms of hits (34%) and page-views ($n = 7{,}005$). Twenty-four visitors consumed over 69 h in 478 sessions at the Central Library. The Department of Zoology and the Institute of Administrative Sciences (IAS) came on second and third place respectively, according to the number of page-views ($n = 817$ and $n = 719$ respectively). IAS department users spent more time accessing the e-books databases (over nine hours) as compared to those from the Zoology Department (over five hours). On the other hand, the Applied Psychology Department was the least utilized access point with only four page-views by a single visitor over two sessions. The Islamic Studies Department and the Institute of Communication Studies (ICS) were also with limited activity in terms of e-books' databases usage. For an easier comparison, all the departments were divided into two groups. The first group comprised of departments under the faculties of Science, Life Sciences, Engineering and Technology, Economics and Management Sciences and Social Science. All other faculties like Arts and Humanities, Islamic Studies, Law, Commerce and Education and administrative departments like Central Library, Administration Block, Examination, Accounts, Jamia Masjid and IT Center/Server Room were clustered separately in the second group.

Figures 2 and 3 highlight the differences in the use of e-books among the Departments of Humanities, Islamic Studies, Commerce, Education and general places like Central Library, Hostels and the Admin Block. The results show that the Main Library, Department of Philosophy and the Institute of Education and Research (IER) were more active according to page-views and visitors as compared to others. Similarly, the users utilizing these access points conducted more sessions and spent more hours as compared to other access points such as the departments of art and humanities subjects. The University of the Punjab's Central Library remained the most prominent access point among all the academic and administrative departments including the executive club, student hostels and the mosque. Figures 2 and 3 also portray the use of HEC e-books databases at the different access points

Table 2. Usage of e-books databases

E-books databases	Sessions	Session duration (hours)
ProQuest	1940	147.57
HEC ebrary	279	8.29
HEC digital library	158	5.36
Project MUSE e-books	96	7.16
e-Books on PU library website	26	1.19
McGraw-Hill medicine e-books	22	0.34
McGraw-Hill engineering books	16	1.33
Cambridge University Press	6	1.37

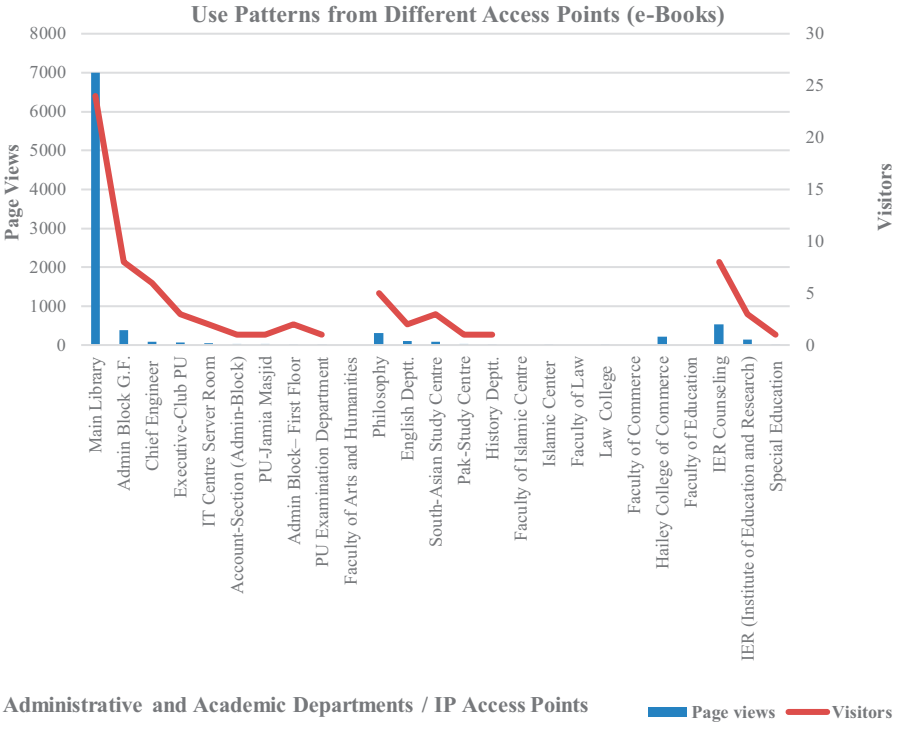

Figure 2.
Various administrative and academic' scholarly communities (use of e-books)

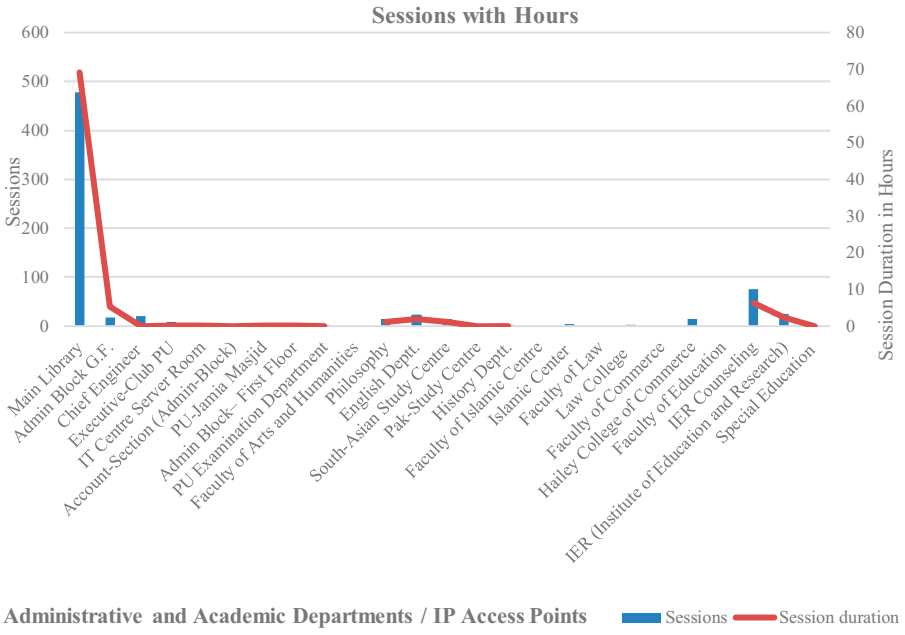

Figure 3.
Various administrative and academic scholarly communities (use of e-books)

like the Admin block, Punjab University Central Library, Accounts, Jamia Masjid, Examination department, including faculties of Arts and Humanities, Islamic Studies, Law, Commerce and Education etc. in terms of total visitors, page-views and number of sessions with their duration.

The differences in the use patterns of various faculties like Science, Social Sciences, Engineering and Technology, Life Sciences and Biological Sciences were observed through the log analysis as shown in Figures 4 and 5. Both figures show that the Faculty of Life Sciences remained the most active access point of HEC e-books databases among science subjects, followed by the Faculty of Economics and Management Sciences and the Faculty of Sciences.

Moreover, it was found through the log analysis that the users were accessing e-books databases more frequently from the science faculties as compared to the administrative departments and other faculties like the Faculty of Arts and Humanities, Islamic Studies, Law, Commerce and Education.

Access points other than the academic and administrative departments were also identified in the TLA, e.g. girls ($n = 11$) and boys ($n = 13$) hostels were also being used for accessing e-books databases. An IP address was combined for both male and female hostels which had eleven page-views by two visitors in three sessions spending twenty-two minutes. Log analysis of boarding male and female students at the hostels may be helpful in investigating the gender-based differences in e-books databases use.

The log analysis presented in Table 3 shows that in the girls' hostels, female boarders had 3,477 hits (7%) with 1,548 page-views during the study year. Seventy-five females consumed over fourteen hours in conducting 271 sessions. Whereas, male boarding students had 6,218 hits (12%) with 4,214 page-views. Sixty-seven male hostel boarders conducted 198 sessions in nearly fourteen hours.

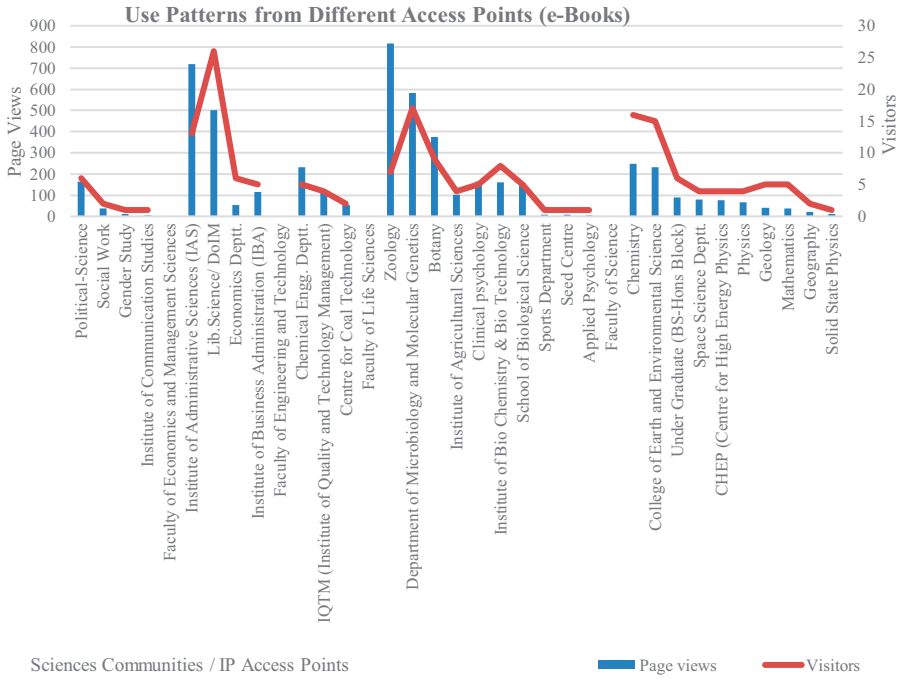

Figure 4.
Usage patterns of various science subjects' scholarly communities

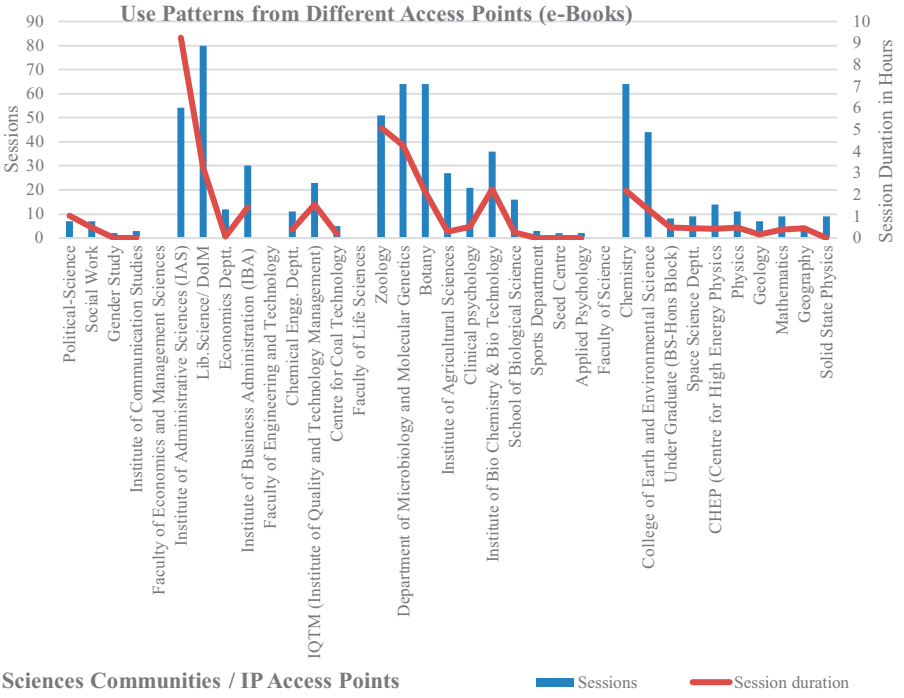

Figure 5.
Usage patterns of various subjects' scholarly communities

Source IP	Page views	Visitors	Sessions	Session duration
Total girls hostels	1,548	75	271	14.27
Total boys hostels	4,214	67	198	13.54

Table 3.
The use of e-books at girls and boys hostels

Gender-based difference in e-books databases use. Gender-based difference could only be identified between the boys' and girls' hostels' use of e-books. More females (n = 75) were active in the eleven girls' hostels as compared to the sixty-seven male students who were active in the thirteen boys' hostels. However, the boarding male students had more than double the number of page-views (n = 4,214) than the female students (n = 1,548), while conducting a lower number of sessions (n = 198) as compared to the female students (n = 271).

Various use frequencies of e-books databases. Log analyzing software "SAWMILL" was used for exploring the various use frequencies like daily or monthly use and for conducting session path analysis of e-books usage.

Daily use of e-books. For observing the daily use frequencies, the TLA was used to explore the data for important patterns. Only 99 days, from June 11, 2017 to September 24, 2017, showed any e-books usage. All other days did not show a single click for HEC e-books, even though users did access free e-books during those days. The free e-books use was disregarded as it was beyond the objectives of this study. The analysis highlighted that users had the most page-views (n = 1,869) while accessing the e-books databases on the August 21st, followed by August 7th (n = 1,711) and July 10th (n = 1,249) in the year 2017. On the other

hand, August 17th, August 23rd and July 20th had the highest number of visitors ($n = 69, 68, 67$ respectively). In terms of number of sessions conducted, July 20th, August 7th and July 18th remained the busiest days. The users spent more time accessing the databases on July 20th, August 15th and July 13th. Figures 6 and 7 show graphic frequencies of e-books databases' usage on a daily basis.

Monthly use of e-books. Analysis of the use of e-books databases on a monthly basis was important as it yielded information that could be used to improve services by targeted policy making, planning training courses, making decisions regarding database subscriptions. The log analysis revealed that the August, 2017, was the most active month of e-books databases usage with higher values in terms of page-views, number of visitors and sessions as compared to other months. On the other hand, the month of July in the same year had more activity as compared to August in terms of total number of hits and time spent in using e-books (Table 4, from February 1, 2017 to January 31, 2018).

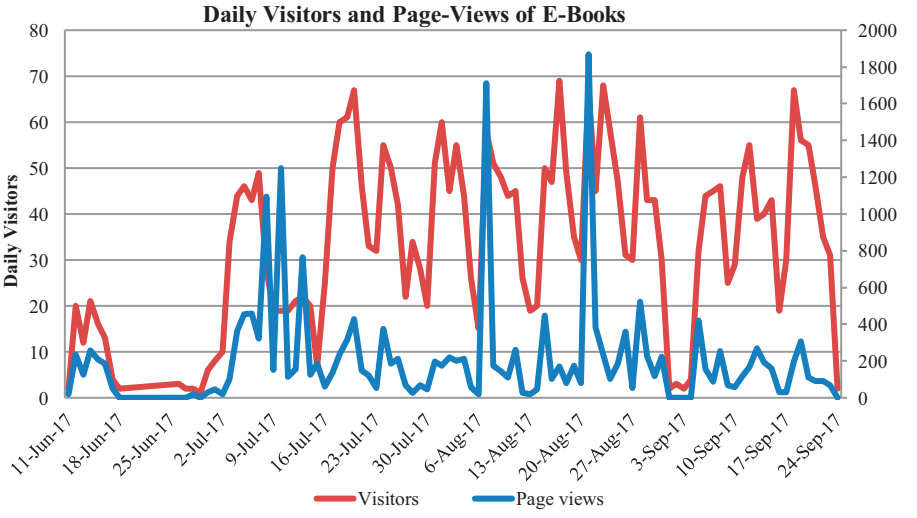

Figure 6.
Daily usage graph of HEC e-books databases

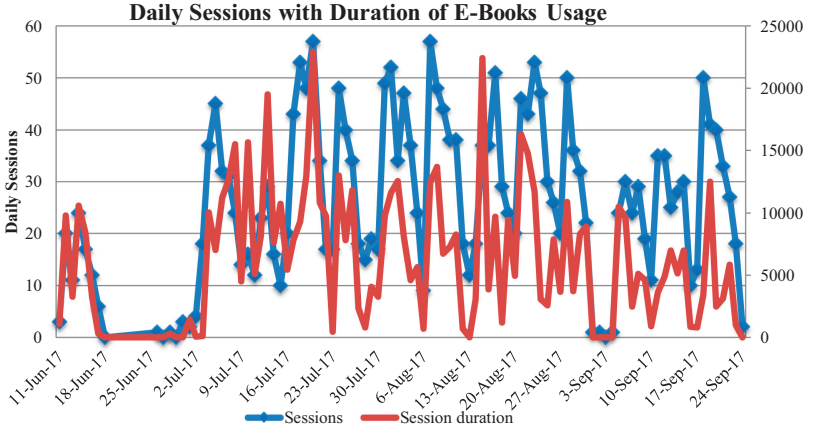

Figure 7.
Daily usage graph of HEC e-books databases

Figure 8 illustrates that the month of July 2017 remained the most active month in terms of e-books databases use with the highest number of hits (n = 24,562). Whereas, June 2017 was the least busy month with just 5,252 hits. Other eight months of the study i.e. from February, 2017 to mid-June, 2017 and October, 2017 to end of January, 2018, did not even have a single hit, page-view or visitor of HEC e-books databases. This might be because the HEC e-books can only be read online and cannot be downloaded in PDF form to be read at home or in leisure time. Therefore, the users may have more time to read them online during the summer vacation months.

Session path analysis (e-books). Session path analysis was another use pattern that showed the preferences of the e-books databases users. Session path was the website address or a starting point from where the users started their sessions for reading e-books from a database. Table 5 shows that ProQuest database (search.proquest.com) was the most popular database among e-books users where 1,463 sessions were started from this path. The second most frequently used path was HEC ebrary, with 176 sessions starting from there. Lastly, the third most frequently used pathway was the official website of the university itself. The results revealed that overall, ProQuest was most popular among e-books databases users with more than 65% (n = 1,658) sessions starting from this path and its various links.

Patterns of various scholarly communities. In order to find the differences among the use patterns of various scholarly communities the departmental names were replaced in the table with their IP addresses according to the IP list maintained by the IT department of the university as mentioned earlier. The departments were grouped under their relevant faculties. Ten scholarly faculties with several departments and various administrative branches under "general departments" were thus identified to show their contribution toward the e-books databases usage.

Month	Hits	Page views	Visitors	Sessions	Session duration in H.
Jun/2017	5,252	1,130	74	98	10.15
Jul/2017	24,562	8,289	587	840	72.36
Aug/2017	14,422	8,712	687	1,078	66.16
Sep/2017	6,184	2,880	498	527	25.14
Total	50,420	21,011	–	–	174.21

Table 4.
Monthly use of HEC e-books databases

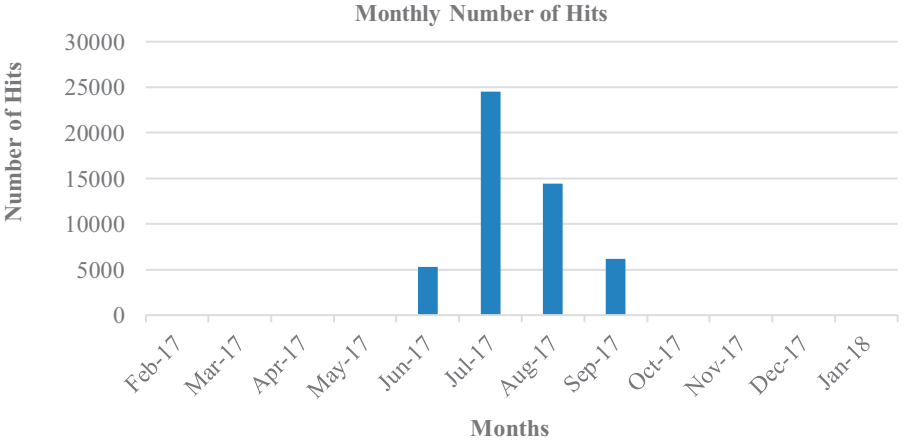

Figure 8.
Bar graph monthly use of HEC e-books databases

Table 5.
E-books databases' top 10 Sessions paths started at. . .

Out of 2,543 sessions, . . . Reset/Collapse all		
1,463	Started at	http://search.proquest.com
176	Started at	http://site.ebrary.com
95	Started at	http://pu.edu.pk
72	Started at	http://pqdtopen.proquest.com
67	Started at	http://www.proquest.com
48	Started at	http://muse.jhu.edu
30	Started at	search.proquest.com:443
26	Started at	http://www.digitallibrary.edu.pk
26	Started at	http://gateway.proquest.com
23	Started at	http://www.e-booksdirectory.com
265 more rows	(+10 + 20+50 + 100+200)	

Table 6 shows that the Faculties of Life Sciences, Economics and Management Sciences and Science had more frequent users of e-books databases as compared to the Faculties of Education, Arts and Humanities, Engineering and Technology, Commerce, Behavioral and Social Sciences, Law and Islamic Studies in terms of page-views with higher number of visitors and sessions.

Table 7 portrays the differences between various faculties' scholarly use of HEC e-books databases. It shows that the Faculties of Life Sciences and Economics and Management Sciences were more frequently using e-books as compared to other communities in terms of number of page-views and visitors. Not a single e-book database user was observed from the Faculty of Oriental Learning in the analysis or graphic presentation (Table 7). The statistics of general departments and all hostels were excluded from analysis in this section because the users could not be categorized according to their faculty or subject of study.

In order to obtain an average statistic per user of a faculty, the cumulative frequencies of all the hits, page-views, sessions and session duration of a faculty were divided by the number of visitors from that faculty. It was observed that a small number of users from the Faculty of Education and Faculty of Arts and Humanities were responsible for a high number of page-views, consequently the ranking of scholarly faculties was seen to change in Table 8. However, the Faculty of Life Sciences was still ranked third according to the average per user page-views and session duration, and even in this average use per visitor statistic, the Faculties of Islamic Studies and Law remained at the bottom of the list. Thus, the results

Table 6.
Usage patterns of various scholarly communities

Access points (IPs)/Faculties	Page views	Visitors	Sessions	Session duration
Faculty of Life Sciences	2,378	58	289	18.03
Faculty of Economics and Management Sciences	1,390	50	176	16.33
Faculty of Science	992	65	199	9.46
Faculty of Education	693	12	101	9.46
Faculty of Arts and Humanities	539	12	58	4.26
Faculty of Engineering and Technology	401	11	39	2.55
Faculty of Commerce	224	8	14	2.15
Faculty of Behavioural and Social Sciences	207	9	17	1.48
Faculty of Law	14	2	2	0.00
Faculty of Islamic Studies	8	1	1	0.00
Total	6,846	228	896	65.32

Note(s): The values are presented in descending order according to page-views

depicted that according to the average use per visitor, the Faculty of Education and the Faculty of Arts and Humanities remained on the top of the list. Other scientific, technological, behavioral and managerial faculties remained in the middle, while those related to Islamic studies and law remained at the bottom of the list according to average page-view per visitor.

Limitations and delimitations of the study

The size of the actual data used and the identity of individual users could not be found due to the previous customization of log recording software of the University of the Punjab. Furthermore, due to financial and time constraints, the log analyzing software (SAWMILL) could not be customized or up-graded for a deep log analysis i.e. logs of most read titles and searching preferences of users, e.g. simple, advanced or keywords searching etc. The study was, therefore, limited to the TLA of only HEC subscribed e-books' databases available on-campus at the university.

Discussions

The objectives of the study were to explore the use patterns of HEC subscribed e-books databases at institutes of higher education, users' frequent access points for accessing e-books' databases and to identify differences in use patterns based on gender and scholarly affiliation of users.

The findings revealed that while there were more than fifty-one thousand enrolled students, staff and teaching faculty at the University of the Punjab during the year 2017–18, only 2.55% (n = 1,301) of the users had accessed HEC e-books databases during the whole

Various faculties	Page views	Visitors
Faculty of Life Sciences	2,378	58
Faculty of Economics and Management Sciences	1,390	50
Faculty of Science	992	65
Faculty of Education	693	12
Faculty of Arts and Humanities	539	12
Faculty of Engineering and Technology	401	11
Faculty of Commerce	224	8
Faculty of Behavioural and Social Sciences	207	9
Faculty of Law	14	2
Faculty of Islamic Studies	8	1

Table 7. Differences between various faculties' scholarly usage of the HEC e-books databases

Scholarly faculties (IP based)	Page views	Sessions	Session duration
Faculty of Education	57.8	8.4	0.49
Faculty of Arts and Humanities	45	4.8	0.22
Faculty of Life Sciences	41	5	0.19
Faculty of Engineering and Technology	36.5	3.54	0.16
Faculty of Commerce	28	1.75	0.17
Faculty of Economics and Management Sciences	27.8	3.52	0.20
Faculty of Behavioral and Social Sciences	23	1.89	0.15
Faculty of Science	15.3	3	0.09
Faculty of Islamic Studies	8	1	0
Faculty of Law	7	1	0

Table 8. Average use per visitor of each faculty

year. The findings, however, did show that the use of e-books had been growing steadily in the university. More than fifty thousand hits and over twenty-one thousand page-views were recorded in just four months (June to September) of the year under study. During the remaining eight months, there was minimal activity seen. This was most likely because the users were intensively engaged in classes or exams during these months and could not spare continuous time to read e-books online. Another reason for lack of activity during these months could be the users' inability to download e-books from the HEC databases in portable document format (PDF) for continuous at home reading. Each user had an average of sixteen page-views daily; over eight minutes during two sessions. Seven hundred and eighty-nine visitors revisited (sessions = 2,031) the databases to read e-books. This was four times more than one-time sessions (n = 512). These results support the earlier findings by Khan *et al.* (2016), Rafiq (2014) and Rafiq and Warraich (2016); that the use of e-books was still in the growing stage. Similarly, e-books' use has also been reported to be at a very early stage by Tenopire (2003) and Noorhidawati and Gibb (2008).

In more than one-third (n = 1,940 out of 2,543) of the sessions, e-books databases users preferred ProQuest. HEC ebrary and digital library were their second and third preference respectively, for accessing e-books. HEC ebrary routes through the university's web domain whereas the HEC digital library can be accessed directly through the Internet. These results exhibit the familiarity of the users with these platforms for accessing or reading e-books. The Central Library of the university was the most frequently used access point of e-books databases with thirty four percent hits or over seven thousand (1/3 of total) page-views. The e-books users might feel that the main library as the most comfortable and accessible point with ready availability of information professional staff's guidance and support for accessing and reading online e-books. It was also an access point that remained open during summer vacations as well. The Department of Zoology (n = 817) and the Institute of Administrative Sciences (IAS, n = 719) were the second and third more frequently used access points respectively. IAS department had users spending two hours more with lesser page-views as compared to the Zoology department. The reasons behind this difference might be explored for designing orientation or literacy programs by information professionals and access management decisions by the HEC authority.

The analysis also revealed interesting results regarding the behavior of boarding students in university's hostels. More male boarding students accessed the e-books databases as compared to the female students, and even though they conducted less sessions and spent less time accessing the databases, they had nearly double the number of hits and page-views as compared to their female counterparts. When compared to the female boarding students, the male boarding students were more active and more frequent users of e-books databases. These results depicted that female boarders might need more help and guidance for using e-books databases. This should be further explored and addressed to enhance the e-books use among female boarding students. This finding is similar to that of like Rowlands *et al.* (2007) who reported that there were more male e-book users than females in an online survey-based study.

As compared to the general and other academic departments like the Faculties of Arts and Humanities, Law and Education; the departments of various sciences like the Faculties of Life Sciences, Economics and Management Sciences and Faculty of Sciences remained the more active communities in the university for using online e-books. But the average use per visitor of each faculty presented a different picture in which Faculty of Education, Arts and Humanities and Life Sciences exhibited better e-books usage as compared to the Faculty of Law, Islamic Studies and Faculty of Science. Such type of diversity by subject was also highlighted in a survey-based study on e-books by Rowlands *et al.* (2007).

The e-books' readers preferred to utilize their free time or summer vacations between June to September for reading e-books. These statistics indicate that the users at the university

found the months of July and August to be most favorable for accessing and reading online e-books. These timings might be taken into consideration when planning system maintenance and system upgrading activities in order to avoid disruptions to service during these times. These findings were different from those of Ahmad and Brogan (2012) who had reported higher e-books use levels during the semesters as compared to summer vacations. Moreover, they had also reported that e-books were being accessed throughout the year instead of only during certain months. The finding of the current study showed that specific days like August 21st, August 7th and July 10th, 2017 were eventful respectively according to page-views, whereas, August 17th, August 23rd and July 20th had more visitors. Similarly, the users spent more time on these databases on July 20th, August 15th and July 13th. These results disclose that users availed the summer vacation period to read e-books when they were free from exams or any other assignments and had plenty of time to concentrate on reading e-books online. Similarly, these results differed from the findings of Ahmad and Brogan (2012) who reported that users accessed e-books on a daily basis.

Conclusion

It is concluded that the use of e-books by the academic community at the institutes of the higher education was still in the growth phase in Pakistan. The results of the study highlighted that the users were accessing e-books from platforms such as ProQuest, HEC ebrary and HEC digital library more than from e-books database established by e-book publishers, e.g. Project MUSE, McGraw-Hill and Cambridge University Press. There were more re-visitors than one-time visitors. The users at the University of the Punjab were accessing e-books mostly from the Central Library of the university, the Department of Zoology and the Institute of Administrative Sciences. The communities of Life Sciences, Economics and Management Sciences and Sciences were more active than those of Arts and Humanities, Law and Education. However, according to the average use per visitor, the Faculty of Education and the Faculty of Arts and Humanities were the topmost e-books users. Males boarding students were better e-book users than females boarding students and PDF was the most popular format among users. The users also read more HEC subscribed e-books during their summer vacations as compared to the whole year.

Implications

The study results imply that the use of HEC subscribed e-books databases is limited in certain communities and is in the growth phase overall. It can be improved by information professionals and HEC decision makers through creation of awareness regarding this facility among those communities who are found to be using it less. Trainings or information literacy instructional programs may also improve the use of e-books databases among female students and those university communities that were using them to a lesser degree. HEC can also facilitate users by providing off-campus access to expand the use of e-books throughout the whole year. Moreover, HEC can play its role in access management by adding the required e-books/e-textbooks into the HEC ebrary/digital library. HEC or funding agencies may provide log analyzing software useful for deep log analysis like most read titles, searching preferences like simple or advanced/keywords searching and so on for future studies. The study of e-books usage is crucial for designing information tools for library management systems that can facilitate users. Usage statistic might also be useful for international publishers in developing and offering competitive and attractive subscription packages to institutions. Furthermore, by enhancing the e-books usage among all communities, a striking market for the international e-book publishers can be created.

References

Ahmad, P. (2015), "E-book adoption in academic and research libraries", Ph.D. Unpublished PhD thesis, Edith Cowan University, Perth.

Ahmad, P. and Brogan, M. (2012), "Scholarly use of e-books in a virtual academic environment: a case study", *Australian Academic and Research Libraries*, Vol. 43 No. 3, pp. 189-213.

Ahmad, P., Brogan, M. and Johnstone, M.N. (2014), "The e-book power user in academic and research libraries: deep log analysis and user customisation", *Australian Academic and Research Libraries*, Vol. 45 No. 1, pp. 35-47.

Anuradha, K. and Usha, H. (2006), "Use of e-books in an academic and research environment: a case study from the Indian Institute of Science", *Program*, Vol. 40 No. 1, pp. 48-62.

Arshad, A. and Ameen, K. (2015), "Usage patterns of Punjab University Library website: a transactional log analysis study", *The Electronic Library*, Vol. 33 No. 1, pp. 65-74.

Becker, B.W. (2015), "Ebooks in the library: the current state of research", *Behavioral and Social Sciences Librarian*, Vol. 34 No. 4, pp. 230-233.

Cassidy, E.D., Martinez, M. and Shen, L. (2012), "Not in love, or not in the know? Graduate student and faculty use (and non-use) of e-books", *The Journal of Academic Librarianship*, Vol. 38 No. 6, pp. 326-332.

Connaway, L.S. (2001), "A web-based electronic book (e-book) library: the netLibrary model", *Library Hi Tech*, Vol. 19 No. 4, pp. 340-349.

Freeman, R.S. and Saunders, E.S. (2016), *E-book Reading Practices in Different Subject Areas: An Exploratory Log Analysis*, Academic E-Books, Purdue, pp. 223-248.

Goodwin, C. (2014), "The e-Duke Scholarly Collection: e-book v. print use", *Collection Building*, Vol. 33 No. 4, pp. 101-105.

Higher Education Commission (2013), *Annual Report: 2012-13*, Islamabad.

Higher Education Commissions (2016), *5th Ranking of Pakistani Higher Education Institutionns (HEIs) 2015*.

Jadoon, N.A., Zahid, M.F., Mansoorulhaq, H., Ullah, S., Jadoon, B.A., Raza, A., Hussain, M., Yaqoob, R. and Shahzad, M.A. (2011), "Evaluation of internet access and utilization by medical students in Lahore, Pakistan", *BMC Medical Informatics and Decision Making*, Vol. 11 No. 1, p. 37.

Jamali, H.R., Nicholas, D. and Rowlands, I. (2009), "Scholarly e-books: the views of 16,000 academics: results from the JISC national e-book observatory", *Aslib Proceedings*, Emerald Group Publishing, London, Vol. 61, No. 1, pp. 33-47.

Khan, A., Bhatti, R. and Khan, A. (2016), "E-books usage by agricultural, engineering and social science students in selected universities of Pakistan: an empirical assessment", *The Electronic Library*, Vol. 34 No. 6, pp. 958-973.

Kim, T.Y., Gang, J.Y. and Oh, H.J. (2019), *Spatial Usage Analysis Based on User Activity Big Data Logs in Library*, Library Hi Tech, Sejong.

Lamothe, A.R. (2010), "Electronic book usage patterns as observed at an academic library: searches and viewings", *Partnership: the Canadian Journal of Library and Information Practice and Research*, Vol. 5 No. 1, pp. 1-22.

Link, F.E. (2012), "Are we there yet? An analysis of e-book equivalent coverage in highly-circulated titles at The College of New Jersey Library", *Collection Building*, Vol. 31 No. 4, pp. 132-135.

Liu, Q., Zhang, X. and Li, Y. (2020), "The influence of information cascades on online reading behaviors of free and paid e-books", *Library and Information Science Research*, Vol. 42 No. 1, p. 101001.

Lonsdale, R. and Armstrong, C. (2010), "Promoting your e-books: lessons from the UK JISC national e-book observatory", *Program*, Vol. 44 No. 3, pp. 185-206.

Mckay, D. and Buchanan, G. (2016), "You can check it out but it will never leave: characterising ebook borrowing patterns", *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 2016, ACM, Melbourne, pp. 203-212.

Nicholas, D. (2008), "If we do not understand our users, we will certainly fail", *The E-Resources Management Handbook*, Vol. 1, pp. 122-129.

Nicholas, D., Rowlands, I., Clark, D., Huntington, P., Jamali, H.R. and Ollé, C. (2008), "UK scholarly e-book usage: a landmark survey", *Aslib Proceedings*, Emerald Group Publishing, London, pp. 311-334.

Nicholas, D., Clark, D., Jamali, H.R. and Watkinson, A. (2014), "Log usage analysis: what it discloses about use, information seeking and trustworthiness", *International Journal of Knowledge Content Development and Technology*, Vol. 4 No. 1, pp. 23-37.

Noorhidawati, A. and Gibb, F. (2008), "How students use e-books-reading or referring?", *Malaysian Journal of Library and Information Science*, Vol. 13 No. 2, pp. 1-14.

Potnis, D., Deosthali, K., Zhu, X. and McCusker, R. (2018), "Factors influencing undergraduate use of e-books: a mixed methods study", *Library and Information Science Research*, Vol. 40 No. 2, pp. 106-117.

Punjab University (2017), "PU Fact Book", The University of the Punjab, Lahore, pp. 1-55.

Rafiq, S. (2014), *Medical Students' Perceptions towards the Use of Electronic Books*, M.Phil, The University of the Punjab, Lahore.

Rafiq, S. and Warraich, N.F. (2016), "Utilization of e-books among undergraduate medical students at Lahore", *Pakistan Journal of Information Management and Libraries*, Vol. 17, pp. 191-200.

Rafique, A., Ameen, K. and Arshad, A. (2019), "Use patterns of e-journals among the science community: a transaction log analysis", *The Electronic Library*, Vol. 37 No. 4, pp. 740-759.

Rowlands, I., Nicholas, D., Jamali, H.R. and Huntington, P. (2007), "What do faculty and students really think about e-books?", *Aslib Proceedings*, Emerald Group Publishing, London, pp. 489-511.

Tenopir, C. (2003), *Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies*, Report of Council on Library and Information Resources, Washington, DC.

Tracy, D.G. (2018), "Format shift: information behavior and user experience in the academic e-book environment", *Reference and User Services Quarterly*, Vol. 58 No. 1, pp. 40-51.

Weng, C., Otanga, S., Weng, A. and Tran, K.N.P. (2019), "Effects of tangrams on learning engagement and achievement: case of preschool learners", *Journal of Computer Assisted Learning*, Vol. 36 No. 4, pp. 458-467, doi: 10.1111/jcal.12411.

Yalman, M. (2015), "Preservice teachers' views about e-book and their levels of use of e-books", *Procedia-Social and Behavioral Sciences*, Vol. 176, pp. 255-262.

About the authors

Azra Rafique is a PhD student at the Department of Information Management, the University of the Punjab, Lahore, Pakistan and currently working as Senior Librarian at National Library of Engineering Sciences, University of Engineering and Technology, Lahore, Pakistan since 2004. The author has 21 years working experience in well reputed automated academic libraries like University of the Central Punjab and University of Engineering and Technology, Lahore. Azra Rafique is the corresponding author and can be contacted at: azra_haroon@hotmail.com

Prof. Dr. Kanwal Ameen (PhD) is currently serving as the Vice Chancellor of the University of Home Economics, Lahore. She has served as Chairperson and Professor at the Department of Information Management, the University of the Punjab, Lahore, Pakistan and has a vast experience in the field of Library and Information Management teaching and scholarly research.

Dr. Alia Arshad is currently Assistant Professor at the Department of Information Management, the University of the Punjab, Lahore, Pakistan.