J. SLIS, 17(4): 77-95, Winter 2025 slis.scu.ac.ir

Shahid Chamran University of Ahvaz

Journal of Studies in Library and Information Science
(*JSLIS*)

https://doi.org/10.22055/slis.2024.40034.1874

Research Paper

Clustering Digital Resources for Collection Development and Improving Archive and Digital Library Services

Shahnaz Khademizadeh[1*], Faezeh Sadat Tabatabai Amiri[2], Mohammadreza Nasiri[3]

1. Associate Professor, Department of Information Science, Faculty of Education and Psychology, Shahid Chamran University, Ahvaz, Iran.
2. PhD student, Department of Information Science, Shahid Chamran University, Ahvaz, Iran.
3. Graduate of Master, Department of Information Science, Qom University, Iran.

Article Info.

Received: 16 August 2024 **Accepted**: 9 September 2025

Abstract

Background and Objectives: Digital libraries generate large volumes of transaction and search-log data that can reveal users' actual information needs and support evidence-based collection development, digitization priorities, resource organization, and personalized services. The Ayatollah Boroujerdi Digital Library in Qom holds distinctive collections, including manuscripts, hot-metal and lithographic publications, textbooks, theses, and other rare resources. Despite the value of these collections, conventional usage statistics provide limited insight into relationships among the subjects searched by users. This study therefore aimed to identify clusters of subject use and association patterns among users' searches in the library's digital databases. More specifically, it examined which subjects were frequently used together and how the resulting patterns could contribute to acquisition, collection development, further digitization, improved service quality, and the future design of recommender systems.

Methodology: This cross-sectional analytical study examined ten years of digital-library usage data covering the period 2011–2021. The dataset was extracted from the NOSA digital-resource archive system. The original databases were stored in SQL format and the extracted records were converted to CSV files for analysis. Data preparation and cleaning included separating date and time values recorded in a single field, converting ordered subject labels into codes, and removing records associated with IP addresses generating more than 50 searches. This procedure was intended to distinguish searches by individual users from activity originating from shared systems or internal library computers. Subjects assigned to each resource were retrieved through resource codes and consolidated into principal subject categories. A two-step clustering procedure was applied because it is suitable for large datasets and can accommodate both categorical and quantitative data. Association-rule mining using the Apriori algorithm was then employed to identify dependencies and co-occurrence patterns among searched subjects. IBM Modeler version 18 was used for clustering and association-rule analysis. For network visualization, the data were transformed into a matrix and analyzed with the bibliometrix package in R. The clustering analysis concerned searched subjects rather than the physical type or digital format of resources.

Findings: PDF was the dominant format, accounting for 98.35% of the resources used. Manuscripts represented 99.91% of the records for which resource type was identifiable, demonstrating the central role of this collection in digital-library use. According to the recorded counts, 709,337 of the 1,048,575 visits, approximately 68%, were made by guest users, whereas 339,238 were associated with non-guest users. Overall use increased markedly over the study period, reaching 245,187 recorded uses in 2018 and 244,333 in 2020. December, April, and January had the highest aggregated use, and the busiest daily period was between 10:00 a.m. and 12:00 noon. Two subject clusters were identified. Cluster 1 covered approximately 56% of the data and was characterized by jurisprudence, principles of jurisprudence, prayer, literature, medicine, mysticism, history, ethics, and sciences. Cluster 2 accounted for approximately 44% and included theology, Hadith, syntax, philosophy, logic, poetry, astronomy, prayers, interpretation, and miscellaneous subjects. Network analysis showed that jurisprudence, syntax, theology, principles of jurisprudence, and Hadith were among the most prominent and highly connected subjects. Jurisprudence had particularly strong co-search relationships with logic, ethics, medicine, Hadith, history, and syntax. The association rules showed confidence values of approximately 55% for some relationships involving jurisprudence, most notably its relationships with logic, ethics, and medicine. By contrast, subjects such as divination, elegies concerning martyrdom, commentary, and arithmetic attracted comparatively little attention.

Discussion: The findings demonstrate that digital-library logs can provide managers with evidence of actual user behavior that is not readily available through surveys or aggregate visit statistics. The predominance of manuscripts and the concentration of use in the first subject cluster suggest that digitization, description, and collection-development activities should give particular attention to jurisprudence and its related fields. At the same time, the relatively limited use of the library's rare hot-metal and lithographic holdings indicates a need for better online promotion, standardized organization, and improved discoverability. Because a large proportion of users access the library as guests, encouraging optional membership could make it possible to analyze longitudinal interests and provide more personalized services without restricting public access. The co-search patterns can also inform subject navigation, "related resources" features, and recommender systems. Nevertheless, the results should be interpreted as associations rather than causal relationships. They reflect one library, depend on the quality of subject coding and log data, and do not incorporate resource format into the clustering model. Future research should compare Apriori with techniques such as k-means and test combined clustering and association-rule approaches in other digital-library environments.

Keywords: Clustering, Digital resources, Collection development, Acquisition, Digital library, Data mining, Ayatollah Boroujerdi library.

***Corresponding author:** Email: s.khademi@scu.ac.ir

How to Cite: Khademizadeh, S., Tabatabai Amiri, F. S., & Nasiri, M. (2025). Clustering digital resources for collection development and improving archive and digital library services *Journal of Studies in Library and Information Science*, 17(4), 77-95. https://doi.org/10.22055/slis.2024.40034.1874

Authors retain the copyright and full publishing rights.
Published by Shahid Chamran University of Ahvaz. This article is an open access article licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0)

فصلنامه
مطالعات کتابداری و علم اطلاعات. سال ۱۷، شماره ۴، زمستان ۱۴۰۴

مقاله پژوهشی

خوشه‌بندی منابع دیجیتالی در راستای مجموعه‌گستری و بهبود خدمات آرشیو و کتابخانه دیجیتالی

شهناز خادمی زاده[1*]، فائزه السادات طباطبایی امیری[2]، محمدرضا نصیری[3]

۱. استاد گروه علم اطلاعات و دانش شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید چمران اهواز، اهواز، ایران
۲. دکتری، گروه علم اطلاعات و دانش شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید چمران اهواز، اهواز، ایران
۳. دکتری علم اطلاعات و دانش شناسی، دانشگاه قم، ایران

اطلاعات مقاله

تاریخ دریافت: ۱۴۰۳/۰۵/۲۶ تاریخ پذیرش: ۱۴۰۴/۰۶/۱۸

چکیده

هدف: پژوهش حاضربدنبال این هدف است که باتوجه به موضوعات مورد جستجوی کاربران،انواع مختلف منابع دیجیتالی اعم کتابخانه درموضوعات متنوع راخوشه‌بندی نمایدتاگامی درجهت بهبودمجموعه‌سازی ومجموعه‌گستری وبهبود کیفیت خدمات کتابخانه دیجیتالی برداشته شود.

روش‌شناسی: پژوهش یک مطالعه تحلیلی ازنوع مقطعی بودکه درطول دوره ۱۰ساله‌ی راه‌اندازی کتابخانه دیجیتالی آیت‌الله العظمی بروجردی درقم از سال ۱۳۹۰ تاسال ۱۴۰۰ صورت پذیرفت.جامعه آماری،داده‌های ۱۰ساله‌ی مربوط به کتابخانه دیجیتالی استخراج شده در مردادماه ۱۴۰۰ از نرم‌افزار نوسا است که پایگاه‌های کتابخانه SQL بوده و داده‌های استخراج شده به فرمت CSV تبدیل شدند. تکنیک‌داده‌کاوی قوانین انجمنی (الگوریتم Apriori) به منظور تعیین وابستگی منابع اطلاعاتی موردجستجوی کاربران در گروه‌های مختلف موضوعی،مورداستفاده قرارگرفت.پس ازاستخراج داده‌ها،آماده‌سازی وپاکسازی داده‌ها انجام شد.برای خوشه‌بندی و قوانین انجمنی ازنرم‌افزار مدلر نسخه۱۸ و برای رسم نمودارهای شبکه‌ای،داده‌ها تبدیل به ماتریس شده و ازکتابخانه bibliometrix درنرم‌افزارR استفاده شد.

یافته‌ها: براساس استفاده کاربران ازموضوعات مختلف منابع در پایگاه‌های کتابخانه دیجیتالی،دوخوشه استخراج شده که خوشه۱حدود ۵۶ درصد و خوشه۲ حدود ۴۴ درصدداده‌ها را پوشش می‌دهند. درخوشه۱، موضوعات فقه،اصول،دعا،ادبیات،طب،عرفان،تاریخ،اخلاق وعلوم بیشتر استفاده شده است. درخوشه۲نیز، موضوعات کلام،حدیث،نحو،فلسفه،منطق،شعر،نجوم،ادعیه،تفسیرومتفرقه بیشترمورداستفاده واقع شده‌اند.یافته دیگر اینکه بیشترین میزان وقوع همزمان جستجو توسط کاربران مربوط به موضوع فقه باموضوعات منطق،اخلاق،وطب است.بعلاوه، زمانی که یک فرد درابتدا موضوع فقه راجستجوکرده،پس از آن،به احتمال ۵۵درصد موضوع طب را نیزمورد مداقه قرارداده است.

نتیجه‌گیری: این مطالعه فرصت مناسبی را برای مدیران فراهم می‌آورد تا با شناسایی وشناخت جامعه کاربری،هماهنگی میان برنامه‌ریزی‌های مؤثر وتصمیم‌گیری‌های استراتژیک رابا نیازهای واقعی کاربران شکل داده وزمینه بهره‌مندی حداکثری ازمنابع منحصر بفرد اطلاعاتی این کتابخانه رادرسطح ملی و بین‌المللی فراهم آورند.ازاین‌رو،یافته‌های پژوهش علاوه براینکه لزوم توجه بیشتر مدیران کتابخانه درمجموعه‌سازی و تلاش در جهت دیجیتال‌سازی بیشتر منابع مربوط به حوزه‌های موضوعی مذکور را نمایان می‌سازد، می‌تواند در ارائه سیستم‌های توصیه‌گر در کتابخانه، مفید واقع شود.

کلیدواژه‌ها: خوشه‌بندی، مجموعه‌گستری، کتابخانه دیجیتالی، داده‌کاوی، کتابخانه آیت‌الله بروجردی.

*نویسنده مسئول
ایمیل: s.khademi@scu.ac.ir

مقدمه و بیان مسئله

کتابخانه‌ها از جمله سازمان‌هایی هستند که با پدیده کلان‌داده مواجه‌اند. بر اساس قانون پنجم علم کتابداری رانگاناتان، «کتابخانه سازمانی در حال رشد است»؛ از این رو، حجم داده‌های کتابخانه‌ها با سرعت بسیار زیادی در حال افزایش است. برای مدیریت کارآمد و مؤثر کتابخانه و نیز گسترش مجموعه و خدمات آن، خودکارسازی کتابخانه و ایجاد کتابخانه دیجیتال اجتناب‌ناپذیر است. ظهور کتابخانه‌های دیجیتال که داده‌های دیجیتالی را ذخیره می‌کنند، جست‌وجو را آسان‌تر ساخته است. برای مدیران کتابخانه مهم است که بیندیشند چگونه می‌توانند خوانندگان را در یافتن آنچه نیاز دارند، راهنمایی کنند (Chen and Chen, 2007). بنابراین می‌توان گفت که با توسعه فناوری‌های نوین اطلاعاتی، کتابخانه‌ها از جمله کتابخانه‌های دیجیتال و مراکز اطلاعاتی، در تلاش هستند تا کمیت و کیفیت خدمات خود را به کاربران بهبود بخشند. از سوی دیگر، در دسترس بودن حجم عظیمی از اطلاعات، ابزارها و روش‌های سنتی را برای پاسخ‌دهی دقیق و سریع به کاربران بدون اتلاف وقت و هزینه‌ی آنان، ناکارآمد ساخته است. از این رو، به‌کارگیری فنون و روش‌هایی همچون داده‌کاوی برای دستیابی کارآمد به داده‌ها و استخراج دانش معنادار از آنها از اهمیت بالایی برخوردار است (Bussaban and Kularbphettong, 2014). این فناوری نوین و قابل‌اعتماد که پتانسیل تسهیل وظایف کتابخانه‌ها و مراکز اطلاعاتی را دارد، ابزاری مناسب برای پیش‌بینی رفتار آینده کاربران به منظور اتخاذ تصمیم‌های دانش‌بنیان به شمار می‌رود (Mishra and Mishra, 2013). داده‌کاوی می‌تواند با تحلیل تراکنش‌های کاربران، کتاب‌های مرتبط را به آنان ارائه دهد و در نتیجه کیفیت خدمات کتابخانه را بهبود بخشد و فرایندهای مدیریت کتابخانه را تسهیل کند (Yi et al., 2014; Zhang, 2014). اصطلاح کتاب‌کاوی یا داده‌کاوی برای کتابخانه‌ها، نخستین بار توسط Nicholson and Stanton (۲۰۰۳) برای توصیف ترکیبی از انباره‌داده، داده‌کاوی و کتاب‌سنجی به کار گرفته شد. این اصطلاح برای ردیابی الگوهای رفتاری در میان کاربران و کارکنان کتابخانه، تغییرات رفتاری و الگوی تراکنش‌های سامانه‌های کتابخانه و نیز استفاده از منابع اطلاعاتی در سراسر کتابخانه به کار می‌رود (Nicholson and Stanton, 2006). کتاب‌کاوی مفهوم جدیدی نیست، بلکه برای تسهیل جست‌وجوی عبارت‌های «کتابخانه» و «داده‌کاوی» در بافت کتابخانه‌ها ابداع شده است (Siguenza-Guzman et al., 2015). همانگونه که اشاره شد، کتاب‌کاوی ابزاری مهم برای کشف اطلاعات مفید کتابخانه‌ای از رکوردهای داده‌ای کتابخانه، به منظور پشتیبانی از تصمیم‌گیری است (Kao, Chang and Lin, 2003). Shiheh (2010) و Prakash, Chand and Gohel (2004) به کاربرد داده‌کاوی در کتابخانه برای مرور کلی فعالیت‌های کتابخانه، از جمله ارائه اطلاعات مفید و ضروری برای تأمین نیازهای مدیریت کتابخانه در تصمیم‌گیری، نظارت بر عملکرد کارکنان، شناسایی نقاط ضعف و پیش‌بینی نیازهای آینده کاربران اشاره کرده‌اند. به طور کلی می‌توان گفت که برای بهبود کیفیت تعامل بین کتابخانه و کاربران آن، به‌کارگیری ابزارهای داده‌کاوی در کتابخانه‌ها ارزشمند محسوب می‌شود (Chang and Chen, 2006). یکی از شواهد تعامل کاربران با کتابخانه‌ها، از جمله کتابخانه‌های دیجیتال، در نتایج جست‌وجو منعکس می‌شود. اطلاعات ارزشمندی که از مجموعه‌داده‌های کتابخانه دیجیتال استخراج می‌شود، با یکپارچه‌سازی راهبرد کتابخانه، جست‌وجو را بهبود می‌بخشد (Chang and Chen, 2006). طراحی مؤثر سامانه‌ها و به ویژه کمک به کاربران در یافتن آسان‌تر اطلاعات، درک نحوه جست‌وجوی کاربران حائز اهمیت است. بررسی رفتار کاربران به درک بهتر آنان و نیازهای اطلاعاتی‌شان منجر می‌شود. در نتیجه، کتابخانه قادر خواهد بود کاربردهای کاربرمحور مناسب‌تری ارائه دهد. برای دستیابی به این هدف، می‌توان با استفاده از تجربه مشابه سایر کاربران، نیازها و مسائل خاص هر کاربر را پیش‌بینی کرد (Kovacevic, Devedzic and Pocajt, 2010).

کتابخانه آیت‌الله بروجردی در قم یکی از کتابخانه‌هایی است که به دلیل دارا بودن منابع ارزشمند اطلاعاتی شامل نسخه‌های خطی، کتاب‌های چاپ سربی، کتاب‌های سنگی و سایر منابع اطلاعاتی ویژه، جایگاهی خاص در میان مخاطبان و

پژوهشگران نسخ‌های خطی و حوزه‌های مرتبط دارد. با آگاهی از این موضوع، مدیران مربوطه برنامه‌ریزی دقیقی برای دیجیتالی‌سازی این منابع اطلاعاتی ارزشمند و منحصربه‌فرد آغاز کرده‌اند تا دسترسی ملی و بین‌المللی پژوهشگران را به این منابع فراهم آورند. پژوهش‌های گوناگونی همچون Krishnamurthy and Balasubramani، (2014) Bussaban and Kularbphettong و (2011) Yu نشان داده‌اند که استفاده از فناوری داده‌کاوی در کتابخانه‌ها به حفظ کاربران و تأمین منابع درخواستی آنان، خرید منابع کارآمدتر، افزایش منابع استنادی علمی، غنی‌سازی مجموعه‌سازی و افزایش استفاده از منابع اطلاعاتی می‌انجامد. بنابراین، پرسشی که در این پژوهش مطرح می‌شود این است که چگونه می‌توان از طریق فنون داده‌کاوی، انواع مختلف منابع دیجیتال شامل نسخه‌های خطی، چاپ سربی، سنگی و سایر منابع منحصربه‌فرد کتابخانه را در موضوعات گوناگون خوشه‌بندی کرد؟ به عبارت دیگر، چگونه می‌توان از فنون داده‌کاوی برای بهبود گسترش مجموعه و مجموعه‌سازی به منظور ارتقای کیفیت خدمات کتابخانه دیجیتال آیت‌الله بروجردی بهره برد؟ یافته‌های این پژوهش فرصت مناسبی را برای مدیران فراهم می‌آورد تا برنامه‌ریزی مؤثر و تصمیم‌گیری‌های راهبردی را با نیازهای واقعی کاربران هماهنگ سازند و با شناسایی و شناخت جامعه کاربران، حداکثر استفاده را از منابع اطلاعاتی منحصربه‌فرد کتابخانه در سطح ملی و بین‌المللی به عمل آورند. از این رو، یافته‌های پژوهش بر ضرورت توجه بیشتر مدیران کتابخانه به مجموعه‌سازی و تلاش برای دیجیتالی‌سازی هرچه بیشتر منابع مرتبط با این حوزه‌های موضوعی تأکید دارد و همچنین می‌تواند در ارائه سامانه‌های توصیه‌گر در کتابخانه مفید واقع شود.

مرور مبانی نظری و پژوهش‌های مرتبط

بر اساس مرور پیشینه پژوهش‌های مربوط به کاربرد فنون داده‌کاوی در ارزیابی کتابخانه‌ها، می‌توان مطالعات را به‌طور کلی در چهار گروه اصلی دسته‌بندی کرد، همان‌گونه که Siguenza et al (2015) نیز اشاره کرده‌اند. گروه نخست شامل آن دسته از پژوهش‌هایی است که به بررسی دیدگاه درونی کتابخانه‌ها پرداخته و خدمات کتابخانه را با استفاده از فنون داده‌کاوی تحلیل می‌کنند. این مطالعات به ارزیابی آن جنبه‌هایی اشاره دارند که به‌عنوان بخشی از خدمات، قابلیت‌ها و تسهیلات کتابخانه‌ها نظیر طرح سازمان‌دهی، تجهیزات الکترونیکی، کارکنان کتابخانه، امکانات و غیره در نظر گرفته می‌شوند. دیدگاه درونی سامانه کتابخانه، کاربرد داده‌کاوی را در تحلیل موضوعات مرتبط با فرایندها و خدماتی که در کتابخانه انجام می‌شوند، هنگام تحلیل خدمات مد نظر قرار می‌دهد. این دسته شامل مطالعات زیر است: (Decker and Höppner, 2006)؛ (Weiner, 2009)؛ (Yi, 2011)؛ (Siriprasoetsin et al., 2011)؛ (Emmons and Wilkinson, 2011)؛ (Zweibel and Lane, 2012)؛ (Yi, 2012)؛ (Sewell, 2013)؛ (Tempelman-Kluit and Pearce, 2014)؛ (Zhang and Wang, 2016)؛ (Ruixiang, 2018).

در گروه دوم، از فنون داده‌کاوی برای ارزیابی سامانه کتابخانه از دیدگاه بیرونی استفاده شده و کیفیت مورد تحلیل قرار گرفته است. به عبارت دیگر، در این مطالعات کیفیت مجموعه و خدمات توسط خود کاربران سنجیده می‌شود. این پژوهش‌ها با بررسی ادراکات کاربران، به ارزیابی مناسب‌بودن و قابلیت‌استفاده‌پذیری منابع فیزیکی و دیجیتال می‌پردازند. روش‌های ارزیابی کیفیت خدمات و مجموعه شامل جمع‌آوری آمار، صندوق‌های پیشنهادات، آزمون قابلیت‌استفاده‌پذیری وب، قابلیت‌استفاده‌پذیری رابط کاربری و نظرسنجی‌های رضایت‌سنجی است. مطالعات (Decker and Hermelbracht, 2006)؛ (Siriprasoetsin et al., 2011)؛ (Papavlasopoulos and Poulos, 2012)؛ (Zhang and Wang, 2016)؛ (Li, 2017) را می‌توان در این گروه جای داد.

مطالعات گروه سوم، کتابخانه را از دیدگاه درونی مجموعه کتابخانه‌ای در نظر گرفته و مجموعه‌ها را تحلیل می‌کنند. در این دیدگاه، ارزیابی سودمندی مجموعه کتابخانه مدنظر قرار می‌گیرد. مطالعات بسیاری را می‌توان در این دسته جای داد که از جمله آنهاست: (Kao, Chang and Lin, 2003)؛ (Nicholson, 2003)؛ (Wu, 2003)؛ (Will, 2006)؛ (Chen and Chen, 2006)؛ (Nicholas et al., 2006)؛ (Chen and Chen, 2007)؛ (King et al., 2007)؛ (Chen et al., 2008)؛ (Li and Chen, 2008)؛ (Kovacevic, Devedzic and Pocajt, 2010)؛ (Tosaka and Weng, 2011)؛ (Emmons and Wilkinson, 2011)؛ (Zweibel and Lane, 2012)؛ (Papavlasopoulos and Poulos, 2012)؛ (Long and Wu, 2012)؛ (Yang, 2012)؛ (Koulouris and Kapidakis, 2012)؛ (Zhang and Wang, 2013)؛ (Li, 2014)؛ (Krishnamurthy and Balasubramani, 2014)؛ (Bussaban and Kularbphettong, 2014)؛ (Wang, 2014)؛ (Renaud et al., 2015)؛ (Li, 2017)؛ (Jin and Sheng, 2017)؛ (Liu, 2018)؛ (Roy et al., 2018)؛ (Yi, Chen and Cong, 2018)؛ (Ruixiang, 2018)؛ ژو (Zhou, 2020).

و سرانجام در گروه چهارم، آن دسته از مطالعاتی جای می‌گیرند که از فنون داده‌کاوی برای تحلیل مجموعه‌های کتابخانه از دیدگاه بیرونی بهره گرفته و رفتار کاربران را هنگام دستکاری سامانه کتابخانه از طریق تحلیل کاربرد، ارزیابی می‌کنند. برای نمونه، از تعاملات کاربران با سامانه و تحلیل لاگ‌های تراکنش برای مطالعه ترجیحات کاربران با هدف شخصی‌سازی خدمات کتابخانه استفاده می‌شود. مطالعات گوناگونی به این موضوع پرداخته‌اند که از جمله می‌توان به موارد زیر اشاره کرد: (Pu and Yang, 2003)؛ (Decker and Hermelbracht, 2006)؛ (Decker and Höppner, 2006)؛ (Chen and Chen, 2006)؛ (Nicholas et al., 2006)؛ (Chen and Chen, 2007)؛ (King et al., 2007)؛ (Li and Chen, 2008)؛ (Chen et al., 2008)؛ (Weiner, 2009)؛ (Finnell and Fontane, 2010)؛ (Kovacevic, Devedzic and Pocajt, 2010)؛ (Yu, 2011)؛ (Emmons and Wilkinson, 2011)؛ (Todorinova et al., 2011)؛ (Tosaka and Weng, 2011)؛ (Long and Wu, 2012)؛ (Shieh, 2012)؛ (Zhang and Wang, 2013)؛ (Ahmad, Brogan and Johnstone, 2014)؛ (Hájek and Stejskal, 2014)؛ (Tempelman-Kluit and Pearce, 2014)؛ (Jomsri, 2014)؛ (Bussaban and Kularbphettong, 2014)؛ (Samson, 2014)؛ (Krishnamurthy and Balasubramani, 2014)؛ (Renaud et al., 2015)؛ (Jin and Sheng, 2017)؛ (Liu, 2018)؛ (Roy et al., 2018)؛ (Yi, Chen and Cong, 2018)؛ (Ochilbek, 2019)؛ (Zhou, 2020)؛ (Prehanto et al., 2020)؛ (Silwattananusarn and Kulkanjanapiban, 2020)

بررسی دقیق مطالعات فوق نشان می‌دهد که تعداد پژوهش‌هایی که از فنون داده‌کاوی برای تحلیل قابلیت‌استفاده‌پذیری و مجموعه کتابخانه استفاده کرده‌اند، بیشتر است و پس از آن، تحلیل خدمات و ارزیابی کیفیت قرار دارند. با این حال، باید توجه داشت که برخی از مطالعات به‌طور همزمان وارد حوزه تمام این دسته‌بندی‌ها شده‌اند؛ برای مثال، هم مجموعه و هم استفاده از کتابخانه را تحلیل کرده‌اند. در پژوهش حاضر، با توجه به مجموعه منحصربه‌فرد نسخه‌های خطی، چاپ‌های سربی و سنگی و سایر منابع در موضوعات گوناگون که در کتابخانه دیجیتال آیت‌الله بروجردی در اختیار کاربران قرار می‌گیرد، کوشش شده است تا مجموعه کتابخانه از طریق فنون داده‌کاوی مورد کاوش قرار گیرد. علاوه بر این، تلاش شده است تا با خوشه‌بندی منابع دیجیتال موجود در پایگاه‌های داده کتابخانه، اقداماتی برای ارائه خدمات بهتر به کاربران از سراسر کشور و سراسر جهان انجام شود و نیز در جهت بهبود مجموعه‌سازی و توسعه مجموعه کتابخانه دیجیتال گام برداشته شود.

روش‌شناسی

پژوهش حاضر یک مطالعه تحلیلی مقطعی است که در بازه زمانی ده‌ساله ساخت کتابخانه دیجیتال آیت‌الله بروجردی در قم، از سال ۱۳۹۰ تا ۱۴۰۰ (۲۰۱۱ تا ۲۰۲۱) انجام شده است. جامعه آماری شامل داده‌های ده‌ساله مربوط به کتابخانه دیجیتال

است که در تیرماه ۱۳۹۹ (ژوئیه ۲۰۲۰) از نرم‌افزار مخزن منابع دیجیتال نوسا استخراج شده‌اند. پایگاه‌های داده به فرمت SQL بودند و داده‌های استخراج‌شده به فرمت CSV تبدیل شدند. در پژوهش حاضر، به‌پیروی از بسیاری از مطالعات دیگر در این حوزه، از تکنیک قواعد انجمنی و الگوریتم a-priori برای تحلیل مجموعه کتابخانه استفاده شده است. بنابراین، از میان فنون گوناگون داده‌کاوی، از قواعد انجمنی (الگوریتم a-priori) برای تعیین وابستگی متقابل منابع اطلاعاتی جست‌وجوشده توسط کاربران در موضوعات مختلف بهره گرفته شد. یکی از پرکاربردترین تکنیک‌ها برای کاوش در مجموعه، داده‌کاوی با قواعد انجمنی است.

در فرایند داده‌کاوی، پس از استخراج داده‌ها، عملیات آماده‌سازی و پالایش انجام شد؛ به‌طوری که داده‌ها به فرمت مناسب تبدیل گردیدند؛ برای مثال، تاریخ و زمان استفاده از پایگاه‌های داده که در خروجی نرم‌افزار در یک فیلد قرار داشتند، از یکدیگر تفکیک شدند و موضوعات منابع نیز به‌دلیل ترتیبی بودن داده‌ها، به کد تبدیل شدند. افزون بر این، داده‌های مربوط به آی‌پی‌های سامانه‌هایی که بیش از ۵۰ جست‌وجو از طریق آنها انجام شده بود، حذف گردید تا داده‌های مربوط به کاربر واقعی از داده‌های مربوط به کاربرانی که از سامانه‌های اشتراکی و سامانه‌های داخلی کتابخانه برای جست‌وجو استفاده می‌کردند، متمایز شود.

موضوعاتی که برای هر منبع تعیین شده بود، بر اساس کد منبع استخراج و به یک موضوع اصلی تبدیل شدند تا مشخص گردد کاربران چه موضوعاتی را جست‌وجو می‌کنند.

برای پالایش داده‌ها از نرم‌افزار اکسل و برای خوشه‌بندی و قواعد انجمنی به‌ترتیب از نرم‌افزار IBM Modeler 18 استفاده شد. به منظور ترسیم شبکه‌های گرافی، داده‌ها به ماتریس تبدیل و از کتابخانه کتاب‌سنجی (bibliometrix) در نرم‌افزار R استفاده گردید.

بر اساس داده‌های پیشین، صرفاً موضوعات خوشه‌بندی شدند و نوع و قالب منابع در خوشه‌بندی لحاظ نشد. به این معنا که خوشه‌بندی موضوعات بر اساس استفاده کاربران از موضوعات مختلف منابع در پایگاه‌های داده کتابخانه دیجیتال صورت گرفته است. برای خوشه‌بندی، از روش دو مرحله‌ای استفاده شد. این روش برای داده‌های حجیم و استفاده همزمان از داده‌های ترتیبی و کمی قابل کاربرد است.

یافته‌ها

جدول ۱ قالب‌ها و انواع منابع مورد استفاده توسط کاربران را نشان می‌دهد.

جدول ۱. قالب‌ها و انواع منابع مورد استفاده

Table 1. Formats and Types of Resources Used

برچسب سطرها	مجموعه	نسخه خطی	جمع کل	
txt	۳۷۹	۰	۳۷۹	٪۰/۰۴
pdf	۴۵۴	۹۰۸۵۷۰	۹۰۹۰۲۴	٪۹۸/۳۵
(pdf) فایل مخفی	۰	۱۴۹۱۸	۱۴۹۱۸	٪۱/۶۱
جمع کل	۸۳۳	۹۲۳۴۸۸	۹۲۴۳۲۱	
	٪۰/۰۹	٪۹۹/۹۱		

همان‌گونه که در جدول ۱ مشاهده می‌شود، بیش از ۹۸٪ از منابع امانتی به فرمت PDF هستند و بیش از ۹۹٪ از کاربران از نسخه‌های خطی استفاده کرده‌اند.

جدول ۲. وضعیت کاربران عضو و مهمان

Table 2. Membership Status of Registered and Guest Users

وضعیت مهمان	تعداد	درصد
نادرست	۳۳۹۲۳۸	٪۳۲
درست	۷۰۹۳۳۷	٪۶۸
مجموع	۱۰۴۸۵۷۵	

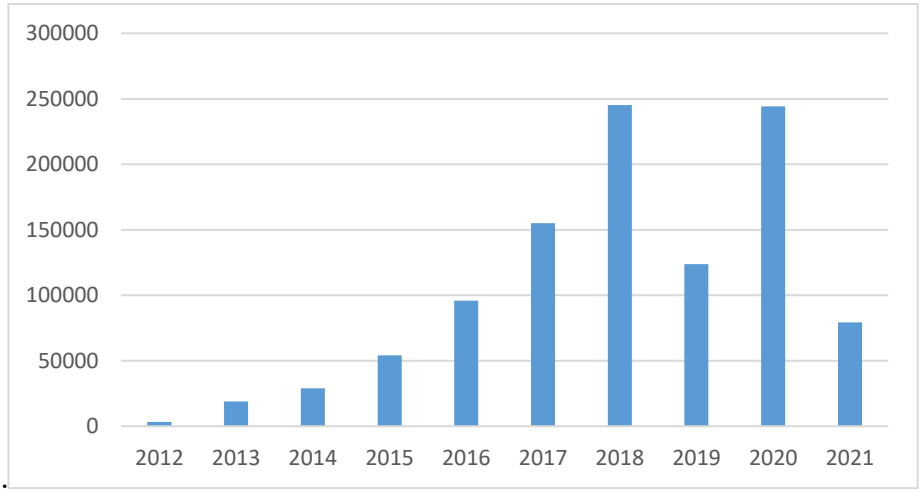
جدول ۲ نشان می‌دهد که حدود ٪۸۶ از کاربران به‌عنوان مهمان به کتابخانه دیجیتال مراجعه کرده‌اند و عضو نیستند.

جدول ۳. میزان استفاده از کتابخانه در بازه ۱۰ ساله

Table 3. Library Use over a Ten-Year Period

برچسب سطرها	تعدادalg_ID
۲۰۱۲	۳۱۸۸
۲۰۱۳	۱۸۸۱۰
۲۰۱۴	۲۸۸۸۶
۲۰۱۵	۵۴۰۶۴
۲۰۱۶	۹۵۹۱۳
۲۰۱۷	۱۵۵۱۲۷
۲۰۱۸	۲۴۵۱۸۷
۲۰۱۹	۱۲۳۸۰۸
۲۰۲۰	۲۴۴۳۳۳
۲۰۲۱	۷۹۲۵۹
جمع کل	۱۰۴۸۵۷۵

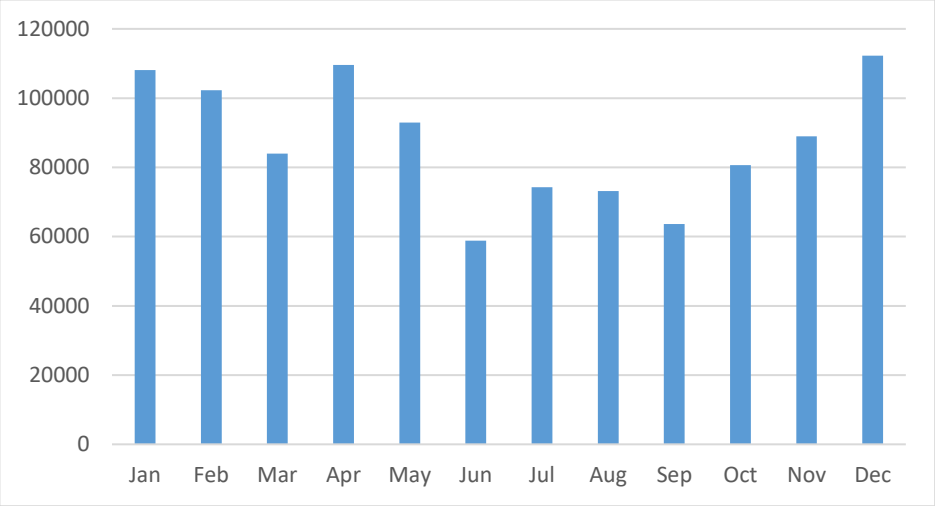
یافته‌های جدول ۳ و نمودار ۱ نشان می‌دهد که در بازه ۱۰ ساله ایجاد کتابخانه دیجیتال، میزان استفاده از کتابخانه در سال‌های اخیر افزایش یافته است.



نمودار ۱. میزان استفاده از کتابخانه در بازه ۱۰ ساله

Chart 1. Library Use over a Ten-Year Period

جدول ۴ و نمودارهای ۲ و ۳ میزان استفاده از کتابخانه را در ماه‌های مختلف سال نشان می‌دهند، به‌طوری که به‌ترتیب در ماه‌های دسامبر، آوریل و ژانویه، بیشترین تعداد کاربران به پایگاه‌های داده کتابخانه دیجیتال مراجعه کرده‌اند.



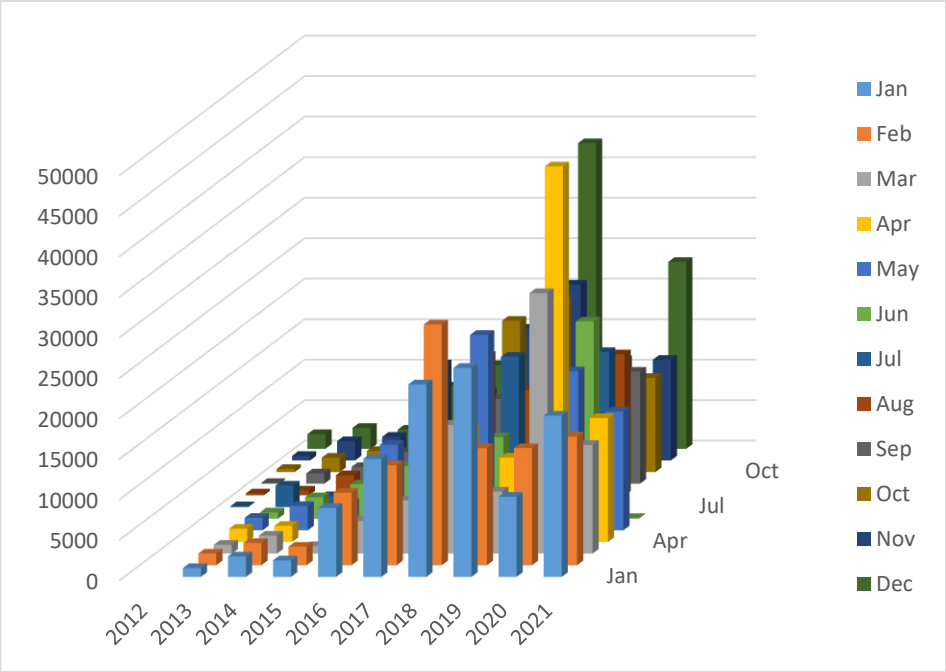
نمودار ۲. میزان استفاده از کتابخانه در ماه‌های مختلف سال

Chart 2. Library Use across Different Months of the Year

جدول ۴. میزان استفاده از کتابخانه در ماه‌های مختلف سال در بازه ۱۰ ساله

Table 4. Library Use across Different Months over a Ten-Year Period

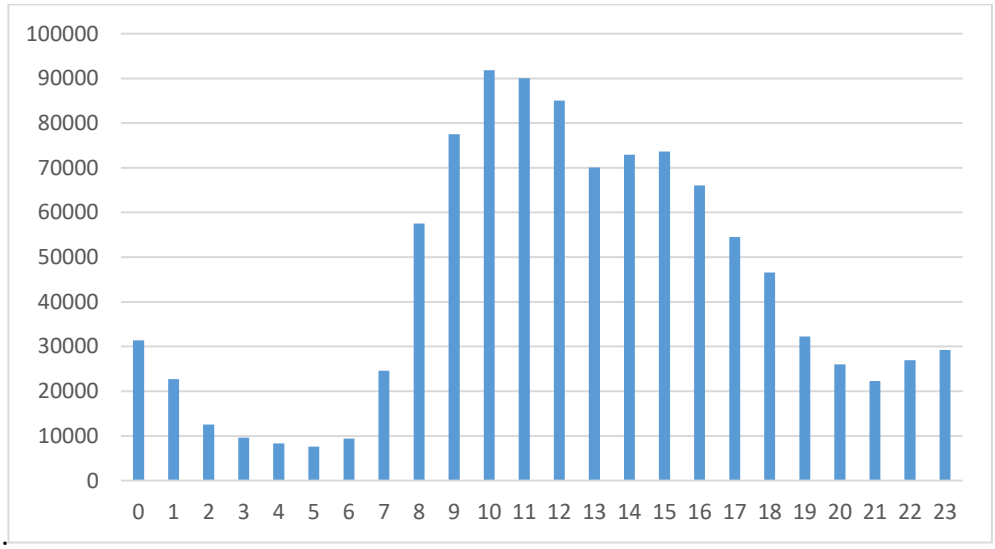
برج سب سطره ا	ژانویه	فوریه	مارس	آوریل	می	ژوئن	ژوئیه	اوت	سپتامبر	اکتبر	نوامبر	دسامبر	جمع کل
۲۰۱۲							۹۹	۲۳۷	۹۶	۳۸۷	۵۵۷	۱۸۱۲	۳۱۸۸
۲۰۱۳	۱۰۸۸	۱۴۴۲	۱۰۸۸	۱۶۳۲	۱۵۳۵	۷۶۸	۲۶۳۶	۶۰۰	۱۲۷۸	۱۷۷۴	۲۳۸۷	۲۵۸۲	۱۸۸۱۰
۲۰۱۴	۲۵۲۶	۲۷۴۹	۲۲۱۳	۱۹۸۵	۲۹۹۹	۲۶۴۰	۱۳۳۷	۲۴۹۳	۲۰۵۸	۲۶۰۳	۲۹۲۱	۲۳۶۲	۲۸۸۸۶
۲۰۱۵	۲۰۳۷	۲۲۸۴	۹۳۰	۲۱۱۵	۳۴۰۱	۴۲۴۵	۳۳۸۵	۲۶۲۷	۳۹۲۲	۹۵۸۰	۱۱۸۵۴	۷۶۸۴	۵۴۰۶۴
۲۰۱۶	۸۵۱۰	۸۹۶۳	۴۰۲۳	۶۹۵۹	۱۰۵۵۷	۶۴۶۹	۹۱۲۹	۶۶۱۷	۶۹۳۹	۸۹۸۹	۸۴۱۷	۱۰۳۴۱	۹۵۹۱۳
۲۰۱۷	۱۴۵۷۳	۱۲۳۴۸	۶۵۷۷	۱۰۸۶۷	۱۱۰۷۳	۷۱۶۸	۱۱۹۷۸	۱۷۲۱۱	۱۰۴۴۸	۱۸۶۸۸	۱۶۲۷۴	۱۷۹۲۲	۱۵۵۱۲۷
۲۰۱۸	۲۳۷۶۵	۲۹۷۱۶	۱۵۹۰۷	۱۳۹۸۶	۲۴۱۲۰	۱۰۰۴۷	۱۸۵۳۷	۱۲۹۸۹	۱۵۳۸۷	۲۱۲۹۴	۲۱۷۰۶	۳۷۷۳۳	۲۴۵۱۸۷
۲۰۱۹	۲۵۸۰۵	۱۴۴۳۹	۷۶۴۷	۱۰۴۱۲	۴۹۹۵	۳۰۷۳	۸۰۰۰	۱۲۹۷۴	۹۶۴۸	۵۶۸۹	۱۲۳۹۶	۸۷۳۰	۱۲۳۸۰۸
۲۰۲۰	۹۸۸۹	۱۴۴۶۶	۳۲۱۶۱	۴۶۳۵۰	۱۹۶۲۶	۲۴۳۳۵	۱۹۱۳۳	۱۷۴۱۱	۱۳۸۳۹	۱۱۶۲۳	۱۲۴۲۹	۲۳۰۷۱	۲۴۴۳۳۳
۲۰۲۱	۱۹۹۰۱	۱۵۸۹۰	۱۳۴۲۵	۱۵۲۹۱	۱۴۶۶۰	۹۲							۷۹۲۵۹
جمع کل	۱۰۸۰۹۴	۱۰۲۲۹۷	۸۳۹۷۱	۱۰۹۵۹۷	۹۲۹۶۶	۵۸۸۳۷	۷۴۲۳۴	۷۳۱۵۹	۶۳۶۱۵	۸۰۶۲۷	۸۸۹۴۱	۱۱۲۲۳۷	۱۰۴۸۵۷۵



نمودار ۳. میزان استفاده از کتابخانه در ماه‌های مختلف سال در بازه ۱۰ ساله

Chart 3. Library Use across Different Months over a Ten-Year Period

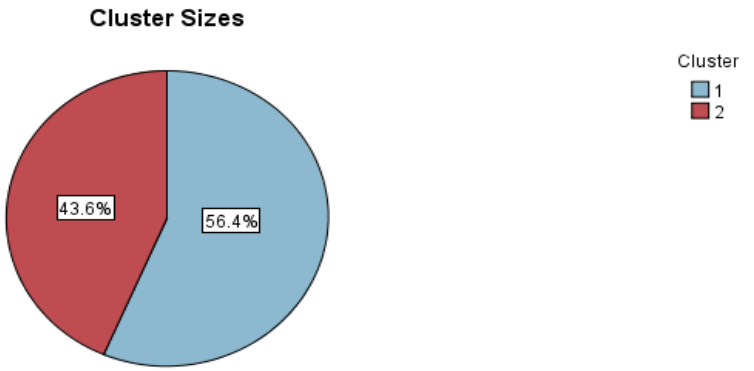
نمودار ۴ میزان استفاده از کتابخانه را در ساعات مختلف روز نشان می‌دهد. بر اساس آن، در ساعت ۱۰ تا ۱۲ ظهر، بیشترین تعداد کاربران به کتابخانه دیجیتال مراجعه کرده‌اند.



نمودار ۴. میزان استفاده از کتابخانه در ساعات مختلف روز

Chart 4. Library Use at Different Hours of the Day

با تکیه بر داده‌های پیشین، در پژوهش حاضر صرفاً موضوعات جست‌وجوشده توسط کاربران خوشه‌بندی شدند و نوع و قالب منابع تأثیر قابل‌توجهی بر خوشه‌بندی نداشت. همان‌گونه که در شکل ۱ مشاهده می‌شود، دو خوشه از جست‌وجوی موضوعی کاربران در پایگاه‌های داده کتابخانه دیجیتال استخراج شد: خوشه ۱ حدود ۵۶٪ و خوشه ۲ حدود ۴۴٪ از داده‌ها را پوشش می‌دهد.

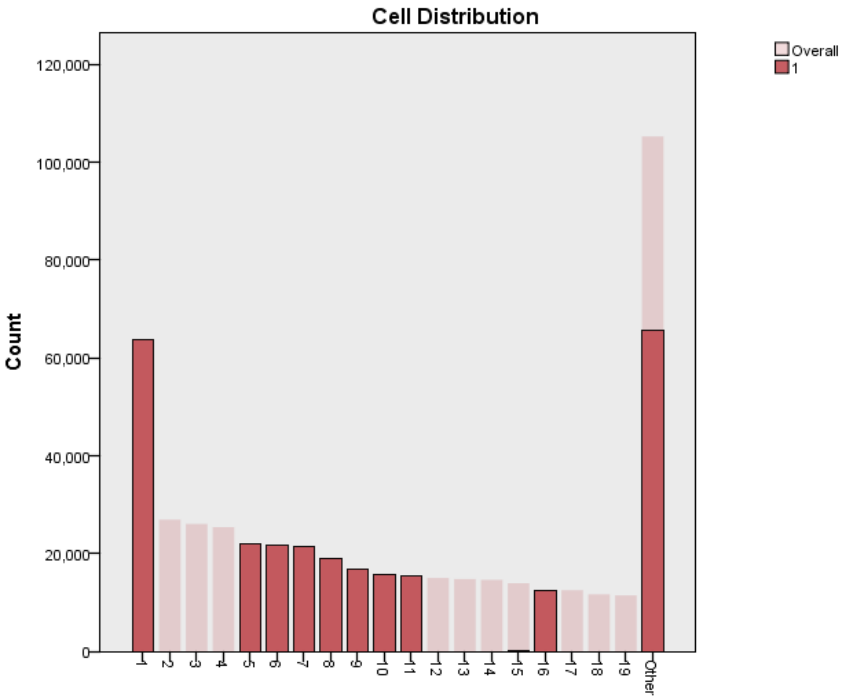


Size of Smallest Cluster	211775 (43.6%)
Size of Largest Cluster	274045 (56.4%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.29

شکل ۱. خوشه‌بندی موضوعی منابع بر اساس جست‌وجوی موضوعی کاربران

Figure 1. Subject Clustering of Resources Based on Users' Subject Searches

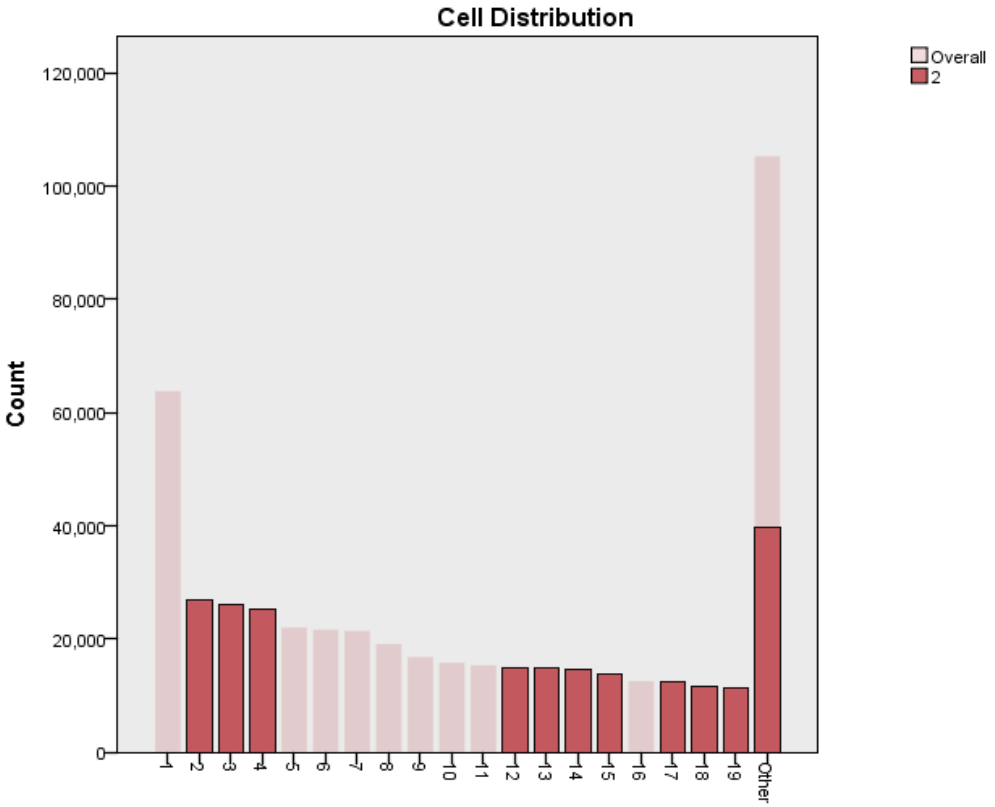
نمودار ۵ موضوعات منابع پرکاربرد توسط کاربران را در خوشه اول نشان می‌دهد. بر اساس این نمودار، پرکاربردترین موضوعات در خوشه اول عبارتند از: فقه، اصول فقه، ادعیه، ادبیات، طب، عرفان، تاریخ، اخلاق و علوم.



نمودار ۵. موضوعات منابع پرمصرف توسط کاربران در خوشه

Chart 5. Most Frequently Used Resource Subjects in Cluster 1

در نمودار ۶، موضوعات منابع پرمصرف در خوشه دوم ارائه شده است. همان‌گونه که مشاهده می‌شود، خوشه دوم عمدتاً شامل موضوعاتی نظیر کلام، حدیث، نحو، فلسفه، منطق، شعر، نجوم، ادعیه، تفسیر و موضوعات متفرقه است.

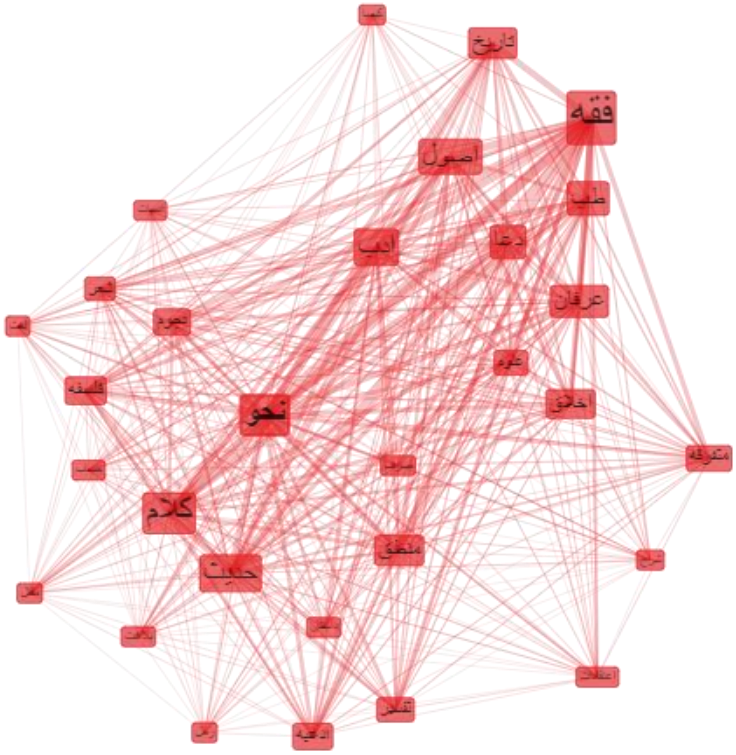


نمودار ۶. موضوعات منابع پرمصرف توسط کاربران در خوشه ۲

Chart 6. Most Frequently Used Resource Subjects in Cluster 2

همان‌گونه که پیشتر ذکر شد، موضوعات بر اساس استفاده کاربران از موضوعات مختلف در کتابخانه دیجیتال، در خوشه‌های جداگانه دسته‌بندی شدند؛ به این معنا که کدام موضوعات در جست‌وجوی کاربران بیشتر به‌طور هم‌زمان رخ داده‌اند. این موضوع نشان می‌دهد که هر کاربر هنگام جست‌وجوی هم‌زمان موضوعات، به چه موضوعات دیگری بیشتر تمایل دارد.

شکل ۲ رابطه بین موضوعات جست‌وجوشده توسط کاربران را نشان می‌دهد که در جدول ۵ نیز قابل مشاهده است. همان‌گونه که روابط بین موضوعات نشان می‌دهد، بالاترین میزان جست‌وجوهای هم‌زمان کاربران مربوط به موضوع فقه با موضوعات منطق، اخلاق و طب است. از سوی دیگر، موضوعات برجسته‌تری همچون فقه، نحو و کلام ارتباط بیشتری با سایر موضوعات دارند و اندازه قلم موضوعات نشان می‌دهد که کاربران بیشتر به دنبال آن موضوعات هستند. به این ترتیب، موضوعاتی نظیر فقه، اصول فقه، نحو و حدیث بیشتر توسط کاربران جست‌وجو می‌شوند. در همین حال، برخی موضوعات مانند پیش‌گویی (رمل)، مقتل، شرح و حساب، توجه کمتری دریافت کرده و به ندرت جست‌وجو شده‌اند.



شکل ۲. رابطه بین موضوعات جست‌وجوشده توسط کاربران

Figure 2. Relationships among Subjects Searched by Users

این موضوع را می‌توان به‌دقت در قواعد انجمنی بررسی کرد. در مرحله بعد، از قواعد انجمنی برای بررسی خوشه‌های موضوعی استفاده شده است.

جدول ۵ موضوعاتی با میزان پشتیبانی ۱۰ تا ۱۶ درصد را ارائه می‌دهد. در شاخص اطمینان، موضوعاتی با میزان بالاتر از ۵۰٪ استخراج شده‌اند. برای مثال، مشاهده می‌شود که در ۵۵٪ موارد، منطق همراه با فقه جست‌وجو شده است و بالعکس؛ یعنی این موضوعات به‌طور هم‌زمان جست‌وجو شده‌اند.

جدول ۵. موضوعات جست‌وجوشده به‌طور هم‌زمان توسط کاربران

Table 5. Subjects Searched Simultaneously by Users

نتیجه	مقدمه	%پشتیبانی	%اطمینان
فقه	منطق	۹/۷	۵۵/۲
فقه	اخلاق	۹/۱	۵۵/۱
فقه	طب	۱۱/۵	۵۴/۹
فقه	حدیث	۱۳/۶	۵۴/۰
فقه	تاریخ	۱۰/۳	۵۳/۶
فقه	نحو	۱۶/۰	۵۱/۸
فقه	ادعیه	۱۲.۰	۵۱/۲
فقه	عرفان	۱۰.۹	۵۱/۰
فقه	اصول فقه	۱۳.۴	۵۱/۱

یافته‌های جدول ۶ نشان می‌دهد که وقتی کاربر ابتدا فقه را جست‌وجو می‌کند، با احتمال ۵۵٪ به احتمال زیاد طب را نیز جست‌وجو می‌کند.

جدول ۶. موضوعات جست‌وجوشده توسط کاربران بر اساس ترتیب جست‌وجو

Table 6. Subjects Searched by Users Based on Search Sequence

نتیجه	مقدمه	پشتیبانی %	اطمینان %
فقه	طب	11.5	54.9
فقه	حدیث	13.6	54.0
فقه	نحو	16.0	51.8
فقه	ادعیه	12.0	51.2
فقه	اصول	13.4	50.1
فقه	ادبیات	14.3	48.4
فقه	کلام	18.0	46.2
کلام	فقه	28.6	29.1
نحو	فقه	28.6	29.0
حدیث	فقه	28.6	25.7
ادبیات	فقه	28.6	24.1
اصول فقه	فقه	28.6	23.5
طب	فقه	28.6	22.0
ادعیه	فقه	28.6	21.6

بحث

بر اساس دسته‌بندی ارائه‌شده در این مقاله از کاربرد داده‌کاوی در ارزیابی کتابخانه‌ها، یعنی تحلیل خدمات، تحلیل کیفیت، تحلیل مجموعه، و تحلیل کاربرد و استفاده، پژوهش حاضر در آخرین دسته قرار می‌گیرد. بنابراین، این مطالعه از نظر کاربرد داده‌کاوی در کتابخانه‌ها با مطالعات مذکور همسو است. به‌ویژه، از نظر تحلیل مجموعه و نیز تحلیل کاربرد و استفاده، و نیز از نظر فن داده‌کاوی به‌کاررفته، یعنی قواعد انجمنی، با مطالعات (Roy et al., 2018)؛ (Yi, Chen and Cong, 2018) ؛ (Renaud et al., 2015) ؛ (Krishnamurthy and Balasubramani, 2014) ؛ (Bussaban and Kularbphettong, 2014) ؛ (Long and Wu, 2012) ؛ (Prehanto et al., 2020) ؛ (Silwattananusarn and Kulkanjanapiban, 2020) ؛ (Yu, 2011) همخوانی دارد. افزون بر این، این مطالعات نیز از رفتار امانت و مرور کاربران برای اهداف گوناگون استفاده کرده‌اند؛ برای نمونه در مطالعه (Roy et al., 2018) از آن برای خوشه‌بندی منابع مختلف داده استفاده شده است.

از سوی دیگر، پژوهش حاضر از نظر فن به‌کاررفته، یعنی قواعد انجمنی، با مطالعات (Decker and Höppner, 2006)؛ (Zhang and Wang, 2013)؛ (Li, 2014)؛ (Zhou, 2020)؛ (Liu, 2018)؛ (Chen et al., 2008)؛ (Li and Chen, 2008) ؛ (Chen and Chen, 2007) ؛ (Pu and Yang, 2003) ؛ (Jomsri, 2014)همسو است که از این فن برای اهداف گوناگونی همچون ارائه الگوهای دسترسی برای کاربران و تحلیل استفاده، شخصی‌سازی خدمات کتابخانه و ارائه سامانه‌های توصیه‌گر کتاب به کاربران بهره برده‌اند.

یافته‌های پژوهش حاضر که با هدف استفاده از فنون داده‌کاوی برای خوشه‌بندی انواع مختلف منابع دیجیتال کتابخانه دیجیتال آیت‌الله بروجردی در قم، به منظور بهبود مجموعه‌سازی و توسعه مجموعه و کیفیت خدمات کتابخانه دیجیتال انجام شد، نشان داد که فرمت PDF مناسب‌ترین قالب برای کاربران است، زیرا حجم فایل‌های دیجیتالی‌شده نسخه‌های خطی بسیار

زیاد است و حفظ ساختار آنها نیز حائز اهمیت است. همچنین از میان منابع گوناگون در پایگاه‌های مختلف کتابخانه، یعنی نسخه‌های خطی، چاپ سربی، سنگی، کتاب‌های درسی و پایان‌نامه‌ها، نسخه‌های خطی بیش از سایر منابع مورد استفاده قرار گرفته‌اند. با توجه به ظرفیت بالای منابع بسیار منحصربه‌فرد این کتابخانه در زمینه چاپ سربی و سنگی که به گفته متخصصان حتی از کتابخانه ملی جمهوری اسلامی ایران نیز برتر است، به نظر می‌رسد که نیاز به اطلاع‌رسانی گسترده از سوی مدیران و مسئولان کتابخانه از طرق گوناگون در وب وجود دارد و باید به استانداردسازی فرایند سازماندهی این منابع توجه ویژه‌ای معطوف شود؛ چرا که به‌نظر می‌رسد با وجود منحصربه‌فرد بودن و غنای منابع و خدمات ارائه‌شده به کاربران، این کتابخانه در میان جامعه مخاطبان شناخته‌شده نیست.

بر اساس یافته‌ها، حدود ٪۸۶ از کاربران به‌عنوان مهمان به کتابخانه دیجیتال مراجعه کرده‌اند و عضو نیستند. نکته مهم این است که مدیران و مسئولان سیاستی را در پیش گرفته‌اند که بر اساس آن، کاربران برای استفاده از خدمات کتابخانه دیجیتال ملزم به عضویت نیستند و با ورود به عنوان مهمان می‌توانند از تمامی خدمات بهره‌مند شوند. با این حال، به نظر می‌رسد اتخاذ رویکردی مبتنی بر عضویت در کتابخانه، مزایایی را برای کتابخانه در تحلیل لاگ‌های کاربران به منظور ارائه خدمات شخصی‌سازی‌شده به کاربران بر اساس نیازها و علایق آنان به همراه دارد که در زمینه بهبود خدمات به کاربران نمی‌توان از آن چشم‌پوشی کرد.

یافته دیگر این است که بر اساس تحلیل داده‌ها در بازه ۱۰ ساله، تعداد استفاده از کتابخانه دیجیتال در طول سال‌ها افزایش یافته است که این امر در نتیجه افزایش تمایل کاربران به دریافت خدمات دیجیتال، آگاهی کاربران از وجود خدمات دیجیتال به دلیل اطلاع‌رسانی توسط مسئولان کتابخانه و نیز مشکلات ناشی از بحران کرونا و محدودیت دسترسی به کتابخانه در دوران قرنطینه دور از انتظار نیست.

نمودار مربوط به میزان استفاده از کتابخانه در ساعات مختلف روز نشان می‌دهد که استفاده از کتابخانه دیجیتال در ساعاتی که کتابخانه برای حضور فیزیکی باز است، بیشتر است. این امر می‌تواند به این دلیل باشد که کاربران بیشتر از پایگاه‌های داده کتابخانه‌های دیجیتال در محیط کتابخانه استفاده می‌کنند؛ موضوعی که می‌تواند توسط مدیران برای اتخاذ تصمیم‌های مدیریتی مورد توجه قرار گیرد.

از میان دو خوشه استخراج‌شده بر اساس موضوعات منابع، خوشه ۱ که شامل فقه، اصول فقه، ادعیه، ادبیات، طب، عرفان، تاریخ، اخلاق و علوم است، ٪۵۶ از منابع را در بر می‌گیرد؛ که نشان‌دهنده میزان استفاده و گرایش به این موضوعات است. نسخه‌های خطی و متون کهن با درون‌مایه‌های یادشده، منابع اصلی و دست‌اول برای پژوهشگران به شمار می‌روند که استفاده از آنها را اجتناب‌ناپذیر می‌سازد؛ بنابراین، بهتر است مدیران در مجموعه‌سازی و توسعه مجموعه به این منابع توجه ویژه‌ای داشته باشند. به این ترتیب، مدیران کتابخانه می‌توانند از این موضوعات در مجموعه‌سازی و توسعه مجموعه بهره گیرند تا مجموعه کتابخانه را برای ارائه بهتر خدمات به کاربران غنی‌تر سازند.

یافته مهم دیگر پژوهش نشان می‌دهد که کاربرانی که موضوع کلی فقه را جست‌وجو کرده‌اند، به‌طور هم‌زمان برخی موضوعات جزئی‌تر مانند منطق، اخلاق و طب را نیز جست‌وجو کرده‌اند. از سوی دیگر، بر اساس ترتیب جست‌وجو، اگر کاربری ابتدا نسخه خطی را در موضوع فقه جست‌وجو کرده باشد، با احتمال ٪۵۵ موضوع جزئی‌تر طب را نیز جست‌وجو کرده است. با توجه به کلیت موضوع فقه، بدیهی است که پس از جست‌وجوی مذکور، کاربران جست‌وجوی خود را تخصصی‌تر می‌کنند، برای مثال با موضوعات جزئی‌تر دیگری همچون طب. افزون بر این، این یافته با توجه به افزایش توجه به مسائل طب اسلامی در سال‌های اخیر، مسائل نوپدید مطرح‌شده در حوزه طب و نیاز به ورود فقه در حل و پاسخگویی به آنها، و این واقعیت که استخراج احکام فقهی بر اساس منابع دست‌اولی است که به صورت نسخه‌های خطی در دسترس هستند،

شایان تأمل است. از این رو، یافته‌های پژوهش بر ضرورت توجه بیشتر مدیران کتابخانه به مجموعه‌سازی و تلاش برای دیجیتالی‌سازی هرچه بیشتر منابع مرتبط با این حوزه‌های موضوعی تأکید دارد و همچنین می‌تواند در ارائه سامانه‌های توصیه‌گر در کتابخانه مفید واقع شود.

توصیه می‌شود در پژوهش‌های آینده از سایر الگوریتم‌های داده‌کاوی مورد استفاده در تحلیل مجموعه کتابخانه، نظیر k-meansاستفاده شود. همچنین می‌توان از ترکیبی از روش‌های داده‌کاوی مانند a-priori و k-means نیز برای تحلیل بهره برد. علاوه بر این، توصیه می‌شود روش‌های گوناگون داده‌کاوی برای تحلیل مجموعه‌داده‌های کتابخانه با یکدیگر مقایسه شوند تا مدیران بتوانند بهترین روش متناسب با محیط خود را انتخاب نمایند.

References

Ahmad, P., Brogan, M., & 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*, 45 (1): 35–47. http://dx.doi.org/10.1080/00048623.2014.885374

Bussaban, K. and Kularbphettong, K. (2014). Analysis of users' behavior on book loan log based on association rule mining. World Academy of Science, Engineering and Technology, *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 8 (1): 18-20.

Chen, A. P. and Chen, C. C. (2006). A new efficient approach for data clustering in electronic library using ant colony clustering algorithm, *The Electronic Library*, 24 (4): 548-559.

Chen, C. C. and Chen, A. P. (2007). Using data mining technology to provide a recommendation service in the digital library, *The Electronic Library*, 25 (6): 711-724.

Chen, R. S., Tsai, Y. S., Yeh, K., Yu, D. and Bak-Sau, Y. (2008). Using data mining to provide recommendation service, *Wseas Transactions on Information Science and Applications*, 5 (4): 459-474.

Decker, R., & Hermelbracht, A. (2006). Planning and evaluation of new academic library services by means of Web-based conjoint analysis. *The Journal of Academic Librarianship*, 32 (6): 558–572. http://dx.doi.org/10.1016/j.acalib.2006.06.016.

Decker, R., & Höppner, M. (2006). Strategic planning and customer intelligence in academic libraries. *Library Hi Tech*, 24 (4): 504–514. http://dx.doi.org/10.1108/07378830610715374.

Emmons, M., & Wilkinson, F.C. (2011). The academic library impact on student persistence. *College and Research Libraries*, 72 (2): 128–149.

Finnell, J., & Fontane, W. (2010). Reference question data mining: A systematic approach to library outreach. *Reference and User Services Quarterly*, 49 (3): 278–286.

Hájek, P., & Stejskal, J. (2014). Library user behavior analysis- Use in economics and management. *WSEAS Transactions on Business and Economics*, 11: 107–116.

Jin, W. and Sheng, H. (2017). Research on management of libraries in universities and colleges based on k-means clustering algorithm under big data environment, *Revista de la Facultad de Ingeniería*, 32 (8): 177-181.

Jomsri, P. (2014). Book recommendation system for digital library based on user profiles by using association rule, Innovative computing technology (INTECH), *Fourth International Conference on the Innovative Computing Technology (INTECH), IEEE*, Luton: 130-134.

Kao, S. C.; Chang, H. C.; & Lin, C. H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining, *Information Processing and Management*, 39 (1): 133–147. Doi:10.1016/S0306-4573(02)00019-5.

King, J. D., Li, Y., Tao, X., & Nayak, R. (2007). Mining world knowledge for analysis of search engine content. *Web Intelligence and Agent Systems*, 5 (3): 233–253.

Koulouris, A., & Kapidakis, S. (2012). Policy route map for academic libraries' digital content. *Journal of Librarianship and Information Science*, 44 (3): 163–173. http://dx.doi.org/10.1177/0961000612444299.

Kovacevic, A.; Devedzic, V.; & Pocajt, V. (2010). Using data mining to improve digital library services, *The Electronic Library*, 28 (6): 829 – 843, Doi: 10.1108/02640471011093525

Krishnamurthy, V.; and Balasubramani, R. (2014). An association rule mining approach for libraries to analyses user interest, Intelligent computing applications (ICICA), *International Conference on Intelligent Computing Applications, IEEE*, Coimbatore: 122-125.

Li, J. and Chen, P. (2008). The application of association rule in library system, Knowledge Acquisition and Modeling Workshop. *International Symposium on Knowledge Acquisition and Modeling Workshop, IEEE,* Wuhan: 248-251.

Li, X. (2014). An algorithm for mining frequent itemsets from library big data. *Journal of Software*, 9 (9): 2361–2365. http://dx.doi.org/10.4304/jsw.9.9.2361-2365.

Li, y. (2017), Research on application of BP neural network in library electronic resource quality evaluation, *Revista de la Facultad de Ingeniería*, 32 (5): 605-613.

Liu, y. (2018), Data mining of university library management based on improved collaborative filtering association rules algorithm, *Wireless Personal Communications*, 102 (4): 3781-3790.

Long, X. and Wu, Y. (2012), Borrowing data mining based on association rules, *International Conference on Computer Science and Electronics Engineering (ICCSEE), IEEE*, Hangzhou, 2: 239-242.

Mishra, R.N. and Mishra, A. (2013). Relevance of data mining in digital library. *International Journal of Future Computer and Communication*, 2 (1): 10-14.

Nicholas, D., Huntington, P., Monopoli, M., & Watkinson, A. (2006). Engaging with scholarly digital libraries (publisher platforms): The extent to which “added-value” functions are used. *Information Processing and Management*, 42 (3): 826–842. http://dx.doi.org/10.1016/j.ipm.2005.03.019.

Nicholson, S. (2003). The bibliomining process: Data warehousing and data mining for library decision-making. *Information Technology and Libraries*, 22 (4), 146–151.

Ochilbek, R. (2019), Using data mining techniques to predict and detect important features for book borrowing rate in academic libraries, *15th International Conference on Electronics, Computer and Computation (ICECCO), IEEE*: 1-5.

Papavlasopoulos, S., & Poulos, M. (2012). Neural network design and evaluation for classifying library indicators using personal opinion of expert. *Library Management*, 33 (4/5): 261–271. http://dx.doi.org/10.1108/01435121211242308.

Prehanto, D.R., Indriyanti, A.D., Permadi, G.S., Vitadiar, T.Z. and Jayanti, F.D. (2020), Library book modeling data using the association rule method with Apriori algorithm in determining book placement and analysis of book loans, *International Journal of Advanced Science and Technology*, 29 (5): 1244-1250.

Pu, H., & Yang, C. (2003). Enriching user-oriented class associations for library classification schemes. *The Electronic Library*, 21 (2): 130–141. http://dx.doi.org/10.1108/02640470310470507.

Renaud, J., Britton, S., Wang, D. and Ogihara, M. (2015), Mining library and university data to understand library use patterns, *The Electronic Library*, 33 (3): 355-372.

Roy, D., Ding, C., Jin, I. and Thomas, D. (2018). Association rule based clustering of electronic resources in university digital library, *International conference on theory and practice of digital libraries*, 349-353.

Ruixiang, O. (2018), The purchasing management and practice of sci-tech books in university library based on back-propagation artificial neural network, *International workshop on advances in social sciences (IWASS), Hong Kong*: 714-719.

Samson, S. (2014). Usage of e-resources: Virtual value of demographics. *The Journal of Academic Librarianship*, 40 (6): 620–625. http://dx.doi.org/10.1016/j.acalib.2014.10.005.

Sewell, Robin R. (2013). Who is following us? Data mining a library's Twitter followers. *Library Hi Tech*, 31 (1): 160–170. http://dx.doi.org/10.1108/07378831311303994.

Shieh, J. C. (2012). From website log to findability. *The Electronic Library*, 30 (5): 707–720. http://dx.doi.org/10.1108/02640471211275747.

Siguenza-Guzman, L.; Van Den Abbeele, A.; Vandewalle, J.; Verhaaren, H.; & Cattrysse, D. (2015). A holistic approach to supporting academic libraries in resource allocation processes. *The Library Quarterly: Information, Community, Policy*, 85 (3): 295–318.

Silwattananusarn, T.; kulkanjanapiban, P. (2020). Mining and Analyzing Patron's Book-Loan Data and University Data to Understand Library Use Patterns. *International Journal of Information Science and Management (IJISM)*, 18 (2): 151-172.

Siriprasoetsin, P., Tuamsuk, K., & Vongprasert, C. (2011). Factors affecting customer relationship management practices in Thai academic libraries. *The International Information and Library Review*, 43 (4): 221–229. http://dx.doi.org/10.1016/j.iilr.2011.10.008.

Tempelman-Kluit, N., & Pearce, A. (2014). Invoking the user from data to design. *College and Research Libraries*, 75 (5): 616–640. http://dx.doi.org/10.5860/crl.75.5.616.

Todorinova, L., Huse, A., Lewis, B., & Torrence, M. (2011). Making decisions: Using electronic data collection to re-envision reference services at the USF Tampa Libraries. *Public Services Quarterly*, 7 (1–2): 34–48. http://dx.doi.org/10.1080/15228959.2011.572780.

Tosaka, Y., & Weng, C. (2011). Reexamining content-enriched access: Its effect on usage and discovery. *College and Research Libraries*, 72 (5): 412–427. http://dx.doi.org/10.5860/crl-137.

Yang, S. T. (2012). An active recommendation approach to improve book-acquisition process. *International Journal of Electronic Business Management*, 10 (2): 163–173.

Yi, C.; Xia, Y.; and Zhang, Z.Y. (2014). Study the personal push service of university library based on big data mining, *Advanced Materials Research*, 998/999: 1261-1265.

Yi, Z. (2011). Planning change in the information age: Approaches of academic library directors in the United States. *International Journal of Knowledge, Culture and Change Management*, 10 (12): 155–176.

Yi, Z. (2012). Conducting meetings in the change process: Approaches of academic library directors in the United States. *Library Management*, 33 (1/2): 22–35. http://dx.doi.org/10.1108/01435121211203293.

Yi, K., Chen, T. and Cong, G. (2018). Library personalized recommendation service method based on improved association rules. *Library Hi Tech*, 36 (3): 443-457.

Yu, P. (2011). Data mining in library reader management, *International conference on network computing and information security, IEEE,* Guilins, 2: 54-57.

Wang, P. (2014), Research on the prediction method for library reader flow based on evolutionary neural network. *International Conference on Vehicle and Mechanical Engineering and Information Technology (VMEIT), Applied mechanics and materials, Beijin*g, 543/547: 2128-2132.

Weiner, S. (2009). The contribution of the library to the reputation of a university. *The Journal of Academic Librarianship*, 35 (1): 3–13. http://dx.doi.org/10.1016/j.acalib.2008.10.003.

Will, N. (2006). Data-mining: Improvement of university library services. *Technological Forecasting and Social Change*, 73 (8): 1045–1050. http://dx.doi.org/10.1016/j.techfore.2006.05.006.

Wu, C. H. (2003). Data mining applied to material acquisition budget allocation for libraries: Design and development. *Expert Systems with Applications*, 25 (3): 401–411. http://dx.doi.org/10.1016/S0957-4174(03)00065-4.

Zhang, C. (2014). Research on construction of universities sports management information system based on data mining, *Applied Mechanics and Materials*, 687/691:2702-2705.

Zhang, Q. S., & Wang, X.Y. (2013). Research of personalized information service based on association rules. *Advanced Materials Research*, 760-762: 1800–1803. http://dx.doi.org/10.4028/www.scientific.net/AMR.760-62.1800.

Zhang, Q. and Wang, P. (2016). A study on readers' satisfaction of university library based on bp neural network. *Journal of Computational and Theoretical Nanoscience*, 3 (9): 6005-6010.

Zhou, Y. (2020), Design and implementation of book recommendation management system based on improved Apriori algorithm. *Intelligent Information Management*, 12 (3): 75-87.

Zweibel, S., & Lane, Z. B. (2012). Probing the effects of policy changes by evaluating circulation activity data at Columbia University Libraries. *The Serials Librarian*, 63 (1): 17–27. http://dx.doi.org/10.1080/0361526X.2012.687850.