Library Waves: A Biannual Peer Reviewed Journal of Library and Information Science
Volume 12, Issue 1 (January-June, 2026); ISSN: 2455-2291 (Online); Website: www.librarywaves.com

What Do Programmers Ask About DSpace? An Analysis of Technical Questions Using Machine Learning

***Sourav Mazumder[#]** ****Kishore Dey** *****Dr. Tapan Barui**

***** Librarian, Ghoksadanga Birendra Mahavidyalaya, Cooch Behar (West Bengal) - 736171, INDIA; E-mail: smazumderlis91@gmail.com
****** Librarian, Karunamoyee Library, Barasat, Kolkata (West Bengal) - 743248, INDIA; E-mail: kishore.ju29@gmail.com
****** Assistant Professor, Department of Library and Information Science, University of North Bengal, Darjeeling (West Bengal) - 734013, INDIA; E-mail: tapanbarui@nbu.ac.in
# Corresponding author.

Received: 30 May 2026 ***Accepted: 30 June 2026***

Abstract

The growing use of digital repositories has increased the importance of understanding technical challenges associated with repository platforms. This study examines user-generated questions (2010-2025) related to DSpace collected from Stack Overflow. A total of 512 questions were analysed using K-means clustering, a machine learning algorithm for identifying themes or topics. The results reveal that developers mostly asked questions about "indexing configuration," "API usage," and "interface customisation." Additionally, a co-occurrence network map of question tags explored relationships among technologies such as Java, XMLUI, and Solr. The study presents statistics on question scores, answer counts, and comment counts, which reveal patterns in community responses. The findings of the study have practical implications for developers, DSpace users, Library and Information Science professionals, and repository managers in improving documentation and user support to address issues encountered while using DSpace.

Keywords: User questions analysis, Digital Library Software, Text Mining, Unsupervised Machine Learning, Information Seeking Behaviour.

1. Introduction

Scholarly communication has been characterised by dynamic changes, leading to the introduction of Institutional Repositories (IRs) as important mechanisms for managing scholarly content in digital form (Lynch, 2003). IRs have been important in creating, archiving, preserving, and disseminating digital scholarly works since the beginning of the new millennium in connection with the global open-access revolution. The Budapest Open Access Initiative has played a key role in this regard (Costa & Leite, 2016). IRs have grown extensively over the past two decades in the context of mandates by funders, universities, and governments, with steady growth worldwide as indicated by OpenDOAR and ROAR (Singh, 2025). IRs are useful in enhancing visibility and access to research as well as ensuring compliance with open access policies and preservation.

Among repository systems, DSpace is a widely used open-source software for deploying IRs (https://roar.eprints.org/view/software/ accessed on 18/04/2026). It was developed jointly by the MIT Libraries and Hewlett-Packard in 2002 (Wikipedia contributors, 2026). It provides

functionalities such as metadata management using Dublin Core, interoperability via OAI-PMH, and customised workflow management. The flexibility and scalability of DSpace have led to its extensive deployment in academia (MacKenzie Smith et al., 2003).

IRs rely heavily on adequate technological infrastructure; however, many institutions face major challenges related to bandwidth, outdated hardware, and power supply (Okoroma, 2018). Furthermore, deploying DSpace presents technical barriers, including software dependency conflicts, migrating full-text content, and integration issues with authentication systems (Dobariya & Doctor, 2024).

Online Q&A platforms play a crucial role in information seeking by peer support and community responses (Jin et al., 2022). These are important tools for users to obtain personalised, experience-based answers through collective intelligence (Ayoub & Wani, 2025). Among various Q&A platforms, Stack Overflow has emerged as a widely used resource for programmers and developers seeking solutions to programming-related problems (Rosen & Shihab, 2016; Wan et al., 2019; Wijerathna et al., 2021; Zhu et al., 2022).

We can explore more studies on technical questions from Stack Overflow, but they are limited to a specific domain, such as DSpace. The present study analyses DSpace related questions from Stack Overflow to identify major themes. The study applies text-mining techniques, including clustering and tag network analysis, to explore user-generated questions and their underlying relationships. The findings aim to provide insights into common technical issues and contribute to improving user support and documentation.

2. Review of Literature

This section reviews existing and some closely related studies. Treude et al. (2011) examined the effectiveness of questions posted on Stack Overflow for code reviews and conceptual questions. Vasilescu et al. (2013) explored the correlation between Stack Overflow and GitHub activity. Barua et al. (2014) analysed Stack Overflow discussions using LDA (Latent Dirichlet Allocation) to discover topic trends. Rosen and Shihab (2016) investigated questions posted by mobile developers on Stack Overflow regarding development techniques and issues. Simeonov and Stanchev (2019) reviewed the DSpace software platform, focused on data storage and indexing systems for creating open-access repositories in India, and on the platform's role in maintaining the BulDML and BGOpenAIRE repositories. Wan et al. (2019) extracted machine learning-related questions from Stack Overflow and Q&A websites to determine topic distribution and trends, focusing on Scikit-learn, TensorFlow, and Keras.

Uddin et al. (2020) showed that software developers preferred the mined usage scenarios in Opiner over official API documentation. Lee and Su (2021) proposed a method for recommending API classes from natural language queries using Doc2Vec and Latent Dirichlet Allocation to analyse queries collected from Stack Overflow and identify the top 10 Java API classes. Wijerathna et al. (2021) analysed Stack Overflow data to identify design context knowledge and introduced an automated mining approach called DPC Miner. Zhu et al. (2022) studied the relationship between OSS and CQA in the context of open software development on Stack Overflow. Khan et al. (2024) used SESE (Software Engineering Stack Exchange), PMSE (Project Management Stack Exchange), LDA (Latent Dirichlet Allocation) and Stack Overflow to identify trends and challenges in different software development approaches.

IRs can be defined as information sources for storing, protecting, and disseminating digital materials, and they present challenges related to copyright and content providers' permissions. According to Fatima and Tahira (2009), one of the main barriers to the creation of digital libraries is the lack of technological competencies of LIS specialists, which must be improved to overcome such difficulties. Metadata management was also identified by Chapman et al. (2009) as one of the key challenges facing institutions running institutional repositories. Russel and Day (2010) noted that low submission levels in institutional repositories remain a major challenge for many organisations, suggesting that further action is needed to encourage their use. Da Rosa and Lamas (2012) argued that developing countries have several obstacles when setting up their institutional repositories. These included a lack of technological capability and poor internet connectivity. Sterman (2014) has listed several challenges associated with establishing IRs, including a lack of technological know-how, funding issues, institutional barriers, copyright issues, staff shortages, and content acquisition. Formanek (2021) discussed the need to upgrade legacy DSpace systems due to support expiration and the technical and operational complexities of migration. In another study, Formanek (2021) highlighted the significance and complexities of the Search Engine Optimisation process within the DSpace repository system. Asary (2024) found that while DSpace is effective for digital repositories, it requires significant customisation to meet the complex demands of industrial content management. Dobariya and Doctor (2024) noted technical issues in DSpace deployment and system compatibility.

Several studies have explored coding-related questions to understand developer issues and trends. However, queries to DSpace have received little attention. This study addresses this gap by examining DSpace related questions posted on Stack Overflow to identify the major issues and themes discussed by programmers.

3. Research Objectives

This study mainly focuses on the following four objectives:

- To analyse DSpace related questions on Stack Overflow.
- To examine user engagement patterns statistically.
- To identify major themes of the questions using a clustering technique.
- To explore tag relationships through network analysis.

4. Methods

This study follows a quantitative exploratory approach using text mining techniques based on machine learning. This section is divided into three parts, as follows:

4.1 Data Collection

Data was collected from Stack Overflow Data. To retrieve relevant DSpace related questions, a SQL query (Box 1) was executed on the Posts containing the tag dspace. The query retrieved information of 528 records (accessed on 10/03/2026), including post ID, title, text, tags, creation date, score, answer count, and comment count. Some records were not DSpace related questions, though. So, these were excluded. The final dataset, containing 512 questions, was used for analysis, and the question titles were considered for inclusion in the text corpus. Body text may provide greater context but also includes explanatory elements, such as lines of code, which could affect the analysis.

```
SELECT
   Id,
   Title,
   Tags,
CreationDate,
   Score
AnswerCount
CommentCount
FROM Posts
WHERE Tags LIKE '%dspace%'
ORDER BY CreationDate DESC;
```

Box 1: SQL query for data retrieval

4.2 Text Pre-processing

The retrieved question titles were cleaned and prepared for analysis by converting them to lowercase and removing the term “dspace” to avoid bias. The processed text was transformed into numerical representations using TF-IDF with the scikit-learn (Pedregosa et al., 2011) TfidfVectorizer, with English stopwords removed and a maximum feature limit of 5000. This step enabled the textual data to be used for further clustering analysis.

4.3 Cluster Analysis

This study uses K-means (Jin & Han, 2010) to identify major thematic patterns in the dataset. The number of clusters was set to 10 (k = 10) to capture diverse thematic patterns in the dataset. Each cluster was represented by the top ten keywords. The clusters were further labelled with a topic by human judgment. Dimensionality reduction and cluster visualisation were performed using Principal Component Analysis (PCA) (Figure 1).

In addition, this study presents statistical analyses of scores, comments, and answers. The count of answers can indicate the difficulty level of questions.

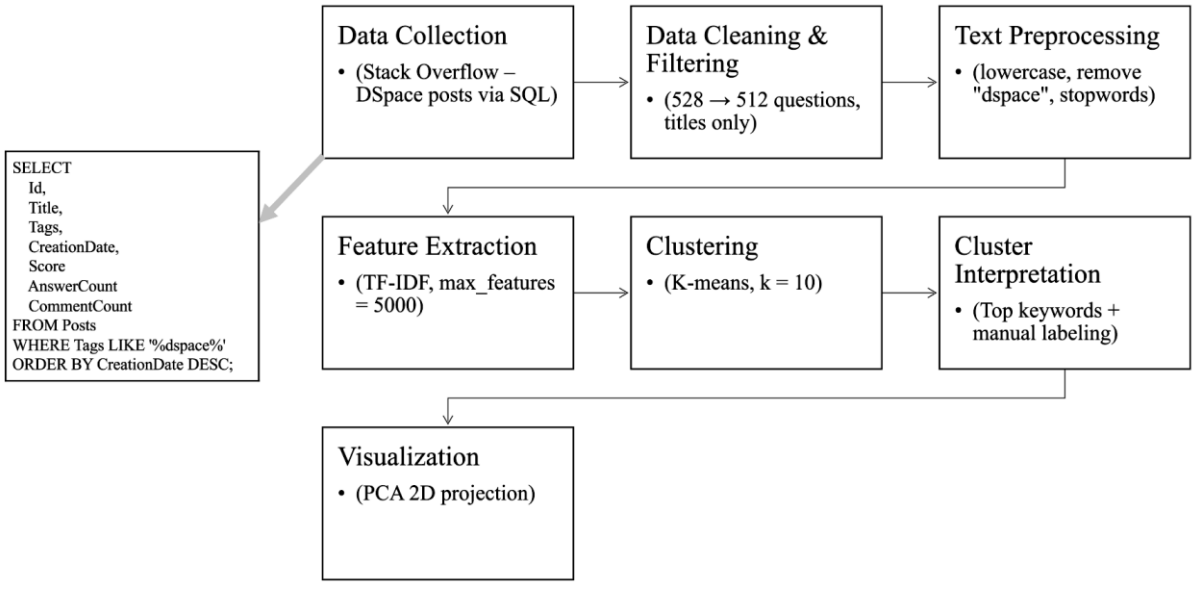

Figure 1: Workflow of the study

5. Results and Discussion

5.1 Trends of DSpace related questions

The trends of DSpace technical questions is shown in Figure 2. There is a trend of growth, with a peak and then a decline. It started with 2 questions in 2010, slowly increased to 7 in 2012 and 8 in 2013, and then spiked at 38 questions in 2014 and peaked at an all-time high 110 questions in 2015. This activity continued in 2016 (n = 93) and 2017 (n = 81). The spike in 2015, and continuing high activity in 2016 and 2017, possibly implies increasing adoption and higher customization of DSpace repositories. Despite increased query volume in 2017, the number of searches has sharply dropped after 2018, with only 41 and 39 questions in 2018 and 2019, respectively. They fell further to 31 and 15 in 2020 and 2021, before smaller fluctuations in 2022 and 2023, then fell drastically to 4 and 2 in 2024 and 2025. A total of 512 questions were parsed for these years. Most questions were from 2014 to 2017, followed by a apparent decline. This suggests these years were the most active for programmers' communication, searching, and discussion of issues. The decline may be attributable to system stability, better documentation, or the use and sharing of other repository sites and technologies.

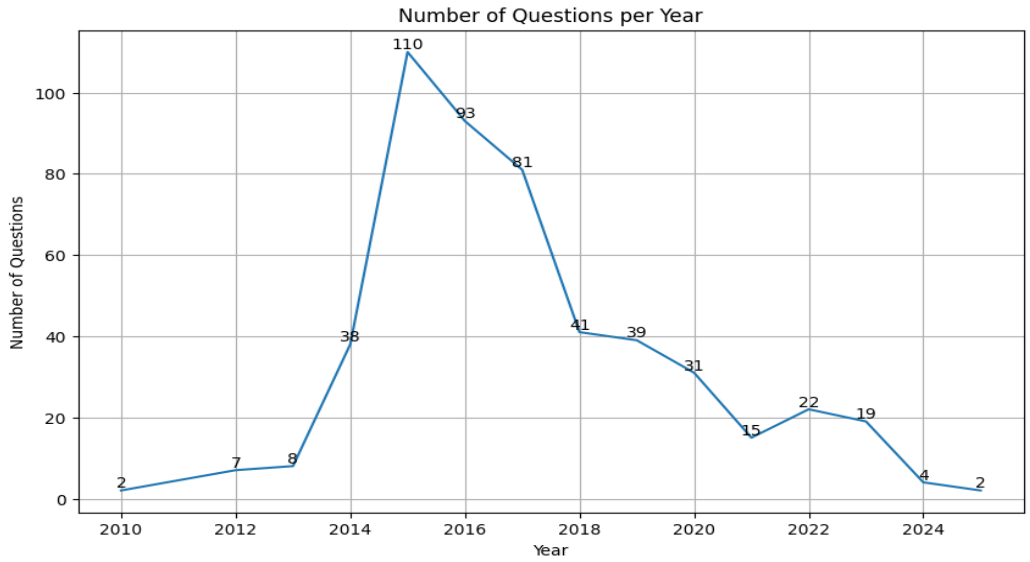

Figure 2: Trends of DSpace related Questions on SO

There are potential reasons why this trend may have occurred. First, over the years, DSpace has been continually developed into a more reliable and robust system; hence, the number of serious problems encountered during work with it has decreased. Second, due to the emergence of better developer documentation and other resources such as GitHub, developers have given the latter system dominance and popularity among educational institutions worldwide. Moreover, the proliferation of AI-based services such as ChatGPT has changed how developers find information online. Now, in addition to using forums, developers can get quick help from conversational AI assistants who provide solutions based on the context.

5.2 Statistical Analysis

The quantitative analysis of technical questions and the engagement (shown in Figure 3) in DSpace provides further insights. The histogram of question scores (Figure 3 (A)) shows that

a large proportion of question scores are low, with a large number concentrated around 0 and 1, meaning many questions were seen by the community but not rated high enough. This could be due to niche technical issues or community disinterest to vote. However, some questions tend to have high scores consistently, implying they are somehow more relevant.

The frequency distribution of answer counts (Figure 3 (B)) shows that the majority of questions have a single answer. Very few questions received two or more answers. This indicates that the level of giving answers is low for most questions. The potential reason could be either that the first answer is a satisfactory answer, or the DSpace developer community is small and specialised , with few people able to suggest alternatives. In addition, the distribution of comment counts (Figure 3 (C)) is evidence that most questions have zero comments. In a few cases where the counts are high, the comments indicate iterative engagement.

Lastly, the boxplot (Figure 3 (D)) shows the variation of score, number of answers, number of comments. For example, the variation in number of comments indicates that while most questions had relatively few comments, some questions had many comments. The median values of these metrics were also low, confirming that there is limited community engagement in the technical questions.

Overall, the statistics suggest that questions on the DSpace tend to be quickly answered and require little discussion, suggesting the presence of a specialised programmers and expert-oriented support community.

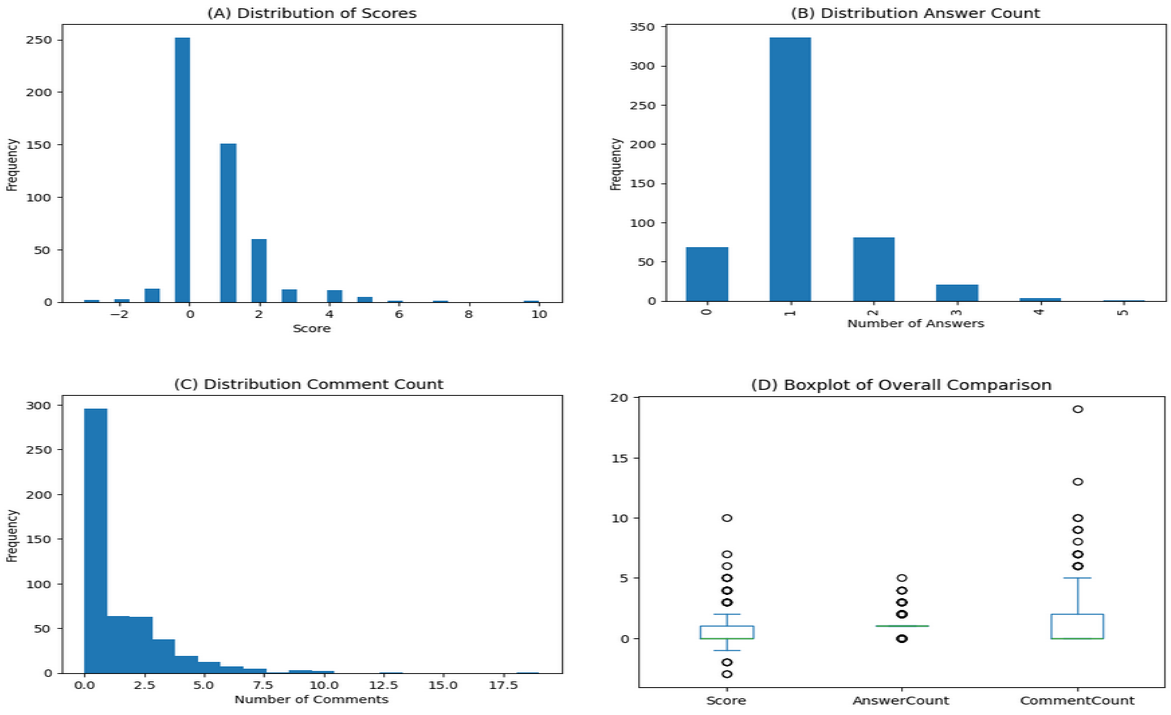

Figure 3: Subplots of statistical analysis

5.3 Cluster analysis

Clusters of themes in the questions are shown in Table 1 and Figure 4. The most dominant cluster is “Indexing Configuration” (n = 236), which exceeds all other clusters. Hence, programmers faced challenges related to indexing, Solr configuration, and collection

management. Clusters like “API Usage” (n = 50), “Interface Configuration” (n = 37), and “UI Theme Customisation” (n = 36) appeared in the questions. These reflect efforts to improve user interfaces, customisation, and integration. The results support Asary’s (2024) observations on the need for extensive customisation in DSpace. Similarly, the present study identifies a considerable proportion of questions related to customisation. These were followed by “Build Errors and Deployment Problems” (n = 31) and “Batch Import and File Upload Issues” (n = 31), which indicate operational and system-level difficulties, as well as large-scale content-handling issues. Other clusters, such as “Metadata Management” (n = 26), “Content Display” (n = 25), and “JSPUI Customisation and Metadata Settings” (n = 23), show consistent concerns related to content organisation, retrieval, and interface-level adjustments. The findings align with Chapman et al. (2009), who also reported challenges associated with metadata management. Finally, “Installation Issues” (n = 17) appears to be the least frequent cluster. These results reveal that DSpace related questions are mainly concentrated on indexing and configuration, and customisation and integration.

Table - 1: Major themes of questions and their representative keywords

Clusters	Top Keywords	Cluster Labels (Themes)
0	‘solr', 'items', 'collections', 'collection', 'oai', 'add', 'java', 'community', 'tomcat', 'assetstore'	Indexing Configuration
1	data', 'metadata', 'edit', 'using', 'user', 'pmh', 'form', 'repository', 'oai', 'input	Metadata Management
2	installation', 'issue', 'windows', 'working', 'displaying', 'installing', 'importing', 'jspui', 'page', 'shibboleth'	Installation Issues
3	item', 'view', 'retrieve', 'display', 'invert', 'xmlui', 'collection', 'add', 'content', 'access'	Content Display
4	error', 'build', 'ant', 'internal', 'failed', 'fresh_install', 'tomcat', 'cris', 'fresh', 'install'	Build Errors and Deployment Problems
5	search', 'theme', 'custom', 'mirage', 'xmlui', 'facet', 'sidebar', 'results', 'box', 'customize'	UI Theme Customization
6	page', 'xmlui', 'discovery', 'feedback', 'home', 'community', 'blank', 'logo', 'metadata', 'add'	Interface Configuration
7	jspui', 'different', 'set', 'customization', 'metadata', 'item', 'use', 'multiple', 'themes', 'file'	JSPUI Customization and Metadata Settings
8	import', 'file', 'fails', 'batch', 'oai', 'error', 'administrator', 'collection', 'xmlui', 'uploading'	Batch Import and File Upload Issues
9	using', 'api', 'new', 'rest', 'bitstream', 'adding', 'class', 'specific', 'submission', 'metadata'	API Usage

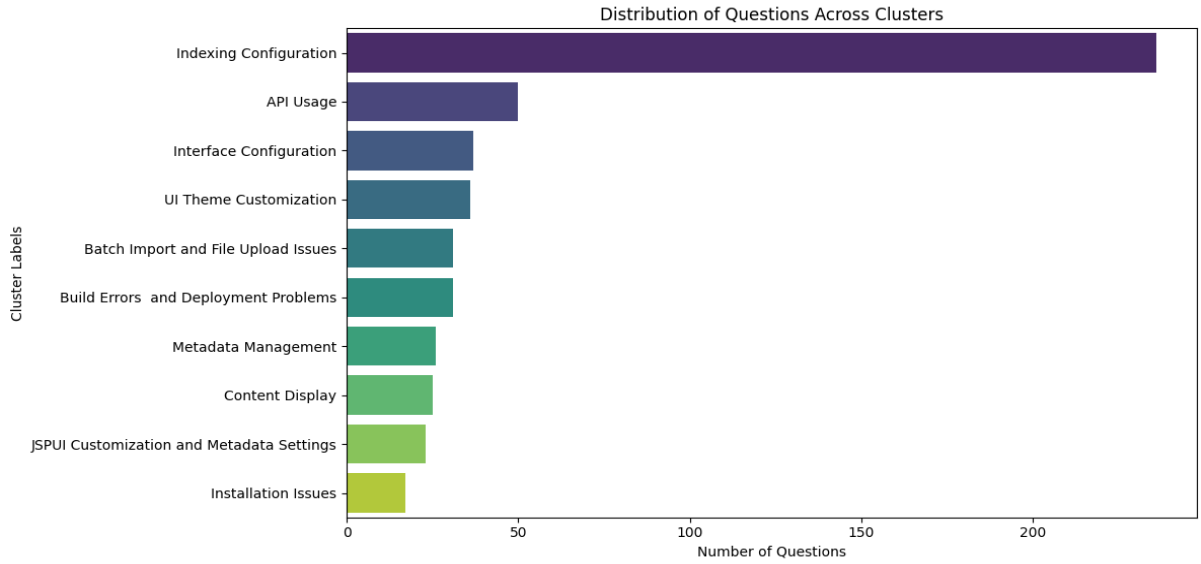

Figure 4: **Frequency count of clusters**

The visualisation in Figure 5 shows a scatter plot of questions after *K*-means clustering and PCA dimensionality reduction. Most data points are densely situated in the centre. Many questions have characteristics and overlap across clusters. Some clusters (on the right side of PC1) are separated from the dense zone, suggesting clearer thematic differences within those groups. On the other hand, clusters near the centre are overlapping, indicating closely related topics. A few clusters appear more scattered and dispersed, showing variability within those topics, while others are more compact, indicating more consistent themes.

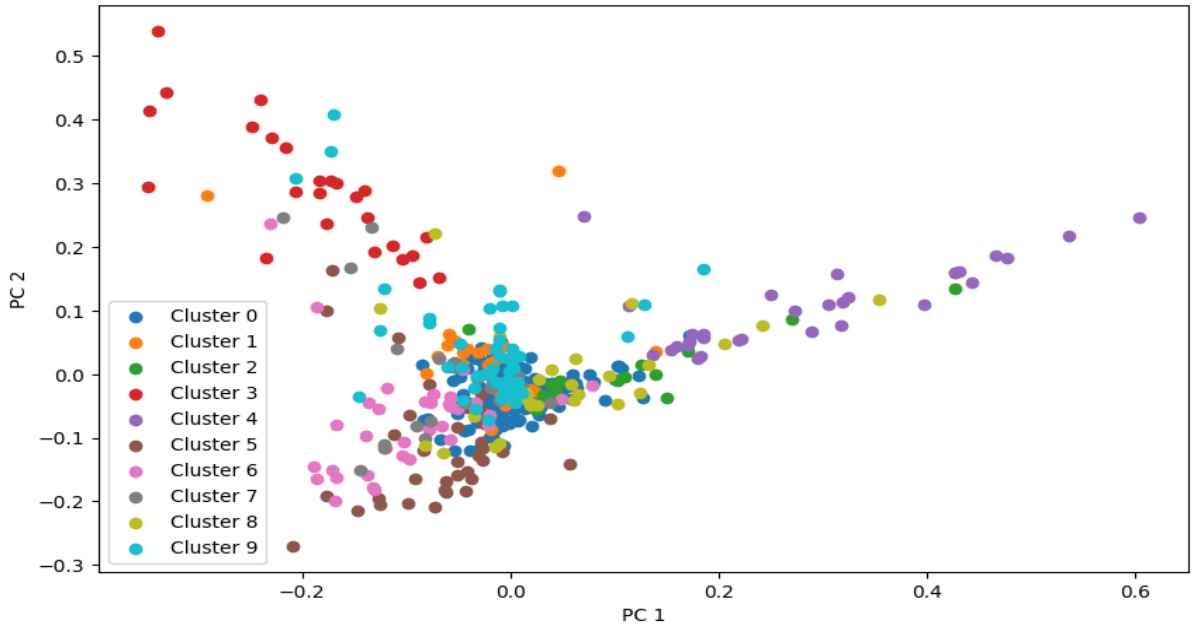

Figure 5: PCA visualisation

The network graph (Figure 6) represents the co-occurrence relationships among the top 50 tags (key technologies) used in DSpace related questions. At the centre, “dspace” is connected to the core, linking with other technologies such as “java,” “xmlui,” “solr,” “tomcat,” and “apache-cocoon.” These are the fundamental components of DSpace development and configuration. A closer look at some backend technologies (e.g., java, sql,

postgresql, maven, ant) indicates system development and database management aspects. On the other hand, UI customisation technologies (e.g., xmlui, jsp, css, mirage, user-experience) reflect interface design and customisation techniques. Further, the graph shows the inter-connections between metadata and interoperability-related components (e.g., oai, oai-pmh, metadata). Some other connections, including file handling and specialised usage of DSpace (e.g., pdf, bitstream, angular, saas, simulink).

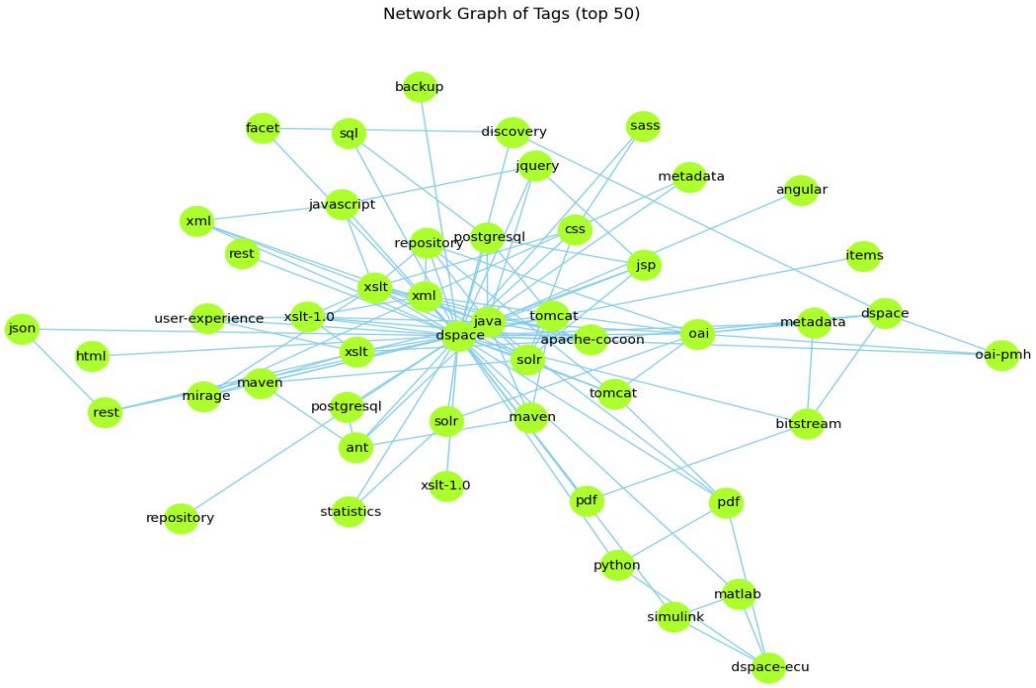

Figure 6: Network graph of the top-50 tags in the questions

6. Conclusion

This study analysed DSpace-related questions on Stack Overflow to identify major technical issues. The findings reveal that developers or DSpace users mostly asked questions about indexing configuration, API usage, and user interface customisation. Moreover, the visualisations (see Figure 5 and Figure 6) show the clustering of questions, their relationships, and nature. The results provide useful insights for developers and DSpace users to understand common problems, which can help with troubleshooting. The study is based solely on a single platform, which may not represent all DSpace users. Furthermore, clustering of title text data may reveal overlapping themes. Future studies may analyse the body of the questions for deeper analysis. But researchers need to be careful when analysing body text, as it may contain code snippets.

Acknowledgement

Grammarly and ChatGPT have been used only for language enhancement and proofreading purposes.

References

1. Arshad Ahmad et al. (2018). A survey on mining Stack Overflow: Question and answering (Q&A) community. *Data Technologies and Applications, 52*(2), 190-247. https://doi.org/10.1108/DTA-07-2017-0054
2. Asary, G. (2024). Evaluating DSpace for an industrial application. http://hdl.handle.net/11427/40751
3. Ayoub, A., & Wani, Z.A. (2025). Users perception of social Q&A sites: An empirical study of information seeking and sharing behaviour in SCIS environments. *Digital Library Perspectives, 41*(3), 465-496. https://doi.org/10.1108/DLP-08-2024-0128
4. Bajaj, K. (2014). Mining Stack Overflow for questions asked by web developers: An empirical study (Master's thesis, University of British Columbia). https://doi.org/10.14288/1.0167068
5. Barua, A., Thomas, S.W., & Hassan, A.E. (2014). What are developers talking about? An analysis of topics and trends in Stack Overflow. *Empirical Software Engineering, 19*(3), 619-654. https://doi.org/10.1007/s10664-012-9231-y
6. Beyer, S., Macho, C., Di Penta, M., & Pinzger, M. (2019). Mining Stack Overflow for API recommendation: An empirical study. *Journal of Systems and Software, 147*, 91-104. https://doi.org/10.1016/j.jss.2018.10.010
7. Bi, T., Liang, P., Tang, A., & Xia, X. (2021). Mining architecture tactics and quality attributes knowledge in Stack Overflow. *Journal of Systems and Software, 180*, 111005. https://doi.org/10.1016/j.jss.2021.111005
8. Broulím, J. (2025). Digital Libraries as a Powerful Concept for Storing Anonymized Data. In Digital Libraries Across Continents. Routledge.
9. Chapman, J.W., Reynolds, D., & Shreeves, S.A. (2009). Repository metadata: approaches and challenges. *Cataloging & Classification Quarterly ,47*(3-4), 309-325.
10. Chen, H., & Zhang, Y. (2014). Functionality Analysis of an Open Source Repository System: Current Practices and Implications. *The Journal of Academic Librarianship, 40*(6), 558-564. https://doi.org/10.1016/j.acalib.2014.09.012
11. Costa, M.P. da, & Leite, F.C.L. (2016). Open access in the world and Latin America: A review since the Budapest Open Access Initiative. *Transinformação, 28*, 33-46. https://doi.org/10.1590/2318-08892016002800003
12. da Rosa, I.B., & Lamas, D.R. (2012). Designing mobile access to DSpace-based digital libraries. In Proceedings of the International Working Conference on Advanced Visual Interfaces (pp. 745-748). ACM. Retrieved from https://dl.acm.org/doi/abs/10.1145/2254556.2254699
13. Dobariya, N., & Doctor, G. (2024). Restructuring and Upgradation of the DSpace-based Institutional Repository of CEPT University Library: A Case Study. *The Journal of Electronic Theses and Dissertations, 3*(1). https://doi.org/10.52407/HXMD3094
14. Faseeh, M., & Jamil, H. (2024). Revolutionizing duplicate question detection: A deep learning approach for Stack Overflow. *IgMin Research, 2*(1), 1-5. https://doi.org/10.61927/igmin135
15. Fatima, N., & Tahira, M. (2009). Institutional repositories: Challenges and opportunities.
16. Fontão, A., Ábia, B., Wiese, I., Estácio, B., Quinta, M., & Dias-Neto, A.C. (2018). Supporting governance of mobile application developers from mining and analyzing technical questions in Stack Overflow. *Journal of Software Engineering Research and Development, 6*(8). https://doi.org/10.1186/s40411-018-0052-6

17. Formanek, M. (2021). Solving SEO Issues in DSpace-based Digital Repositories: A Case Study and Assessment of Worldwide Repositories. *Information Technology and Libraries, 40*(1). https://doi.org/10.6017/ital.v40i1.12529
18. Formanek, M. (2023). DSpace 7 Benefits: Is It Worth Upgrading? *Information Technology and Libraries, 42*(3). https://doi.org/10.5860/ital.v42i3.16209
19. Jin, J., Zhang, T., & Yan, X. (2022). Why do users continually seek knowledge in online Q&A communities? An empirical investigation. *Information Discovery and Delivery, 51*(1), 1-12. https://doi.org/10.1108/IDD-04-2021-0042
20. Jin, X., & Han, J. (2010). K-Means Clustering. In C. Sammut & G. I. Webb (Eds.), Encyclopedia of Machine Learning (pp. 563-564). Springer US. https://doi.org/10.1007/978-0-387-30164-8_425
21. Khan, A.A., Khan, J.A., Akbar, M.A., Zhou, P., & Fahmideh, M. (2024). Insights into software development approaches: Mining Q&A repositories. *Empirical Software Engineering, 29*, 8. https://doi.org/10.1007/s10664-023-10417-5
22. Lee, W.K., & Su, M.T. (2021). Mining Stack Overflow for API class recommendation using Doc2Vec and LDA. *IET Software, 15*(5), 308-322. https://doi.org/10.1049/sfw2.12023
23. Library Philosophy and Practice (e-journal).
24. Lynch, C.A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. Portal: *Libraries and the Academy, 3*(2), 327-336. https://doi.org/10.1353/pla.2003.0039
25. MacKenzie Smith, M., Mary Barton, Mick Bass, Margret Branschofsky, Gregory McClellan, & David Stuve (2003). DSpace: An open source dynamic digital repository. *D-Lib Magazine, 9*(1). https://doi.org/10.1045/january2003-smith
26. Munir, H., Linåker, J., Wnuk, K. et al. (2018). Open innovation using open source tools: a case study at Sony Mobile. *Empir Software Eng, 23*, 186-223. https://doi.org/10.1007/s10664-017-9511-7
27. Okoroma, F.N. (2018). Technological and Infrastructural Issues of Institutional Repositories in University Libraries in Nigeria: And the Way Forward. *The International Information & Library Review, 50*(3), 251-262. https://doi.org/10.1080/10572317.2018.1491710
28. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*(85), 2825-2830.
29. Ponzanelli, L., Bacchelli, A., & Lanza, M. (2014). Leveraging crowd knowledge for software comprehension and development. Proceedings of the IEEE International Conference on Software Maintenance and Evolution, 57-66. https://doi.org/10.1109/ICSME.2014.21
30. Rosen, C., & Shihab, E. (2016). What are mobile developers asking about? A large-scale study using Stack Overflow. *Empirical Software Engineering, 21*(3), 1192-1223. https://doi.org/10.1007/s10664-015-9379-3
31. Russell, R., & Day, M. (2010). Institutional repository interaction with research users: a review of current practice. *New review of academic librarianship, 16*(S1), 116-131.
32. Sahu, T.P., Nagwani, N.K., & Verma, S. (2016). An empirical analysis on reducing open source software development tasks using Stack Overflow. *Indian Journal of Science and Technology, 9*(21). https://doi.org/10.17485/ijst/2016/v9i21/89198

33. Simeonov, G., & Stanchev, P. (2019). Using DSpace platform for creation of open access local repositories. *Digital Presentation and Preservation of Cultural and Scientific Heritage, 9*. https://doi.org/10.55630/dipp.2019.9.23
34. Singh, D. (2025). Open access institutional repositories in India: Insights from the OPENDOAR. *Cybrarians Journal, 77*, 14-23. https://doi.org/10.70000/cj.2025.77.690
35. Sterman, L. (2014). Institutional Repositories: An Analysis of Trends and a Proposed Collaborative Future. *College & Undergraduate Libraries, 21*(3-4), 360-376.
36. Treude, C., Barzilay, O., & Storey, M. A. (2011). How do programmers ask and answer questions on the web? (NIER track). Proceedings of the 33rd International Conference on Software Engineering, 804-807. https://doi.org/10.1145/1985793.1985907
37. Trisal, A., & Mandloi, D. (2021). Machine learning: An overview. *International Journal of Research - Granthaalayah, 9*(7), 343-348. https://doi.org/10.29121/granthaalayah.v9.i7.2021.4120
38. Uddin, Khomh, & Roy. (2020). Mining API usage scenarios from Stack Overflow. *Information and Software Technology, 122*, 106277. https://doi.org/10.1016/j.infsof.2020.106277
39. Vasilescu, B., Filkov, V., & Serebrenik, A. (2013). Stack Overflow and GitHub: Associations between software development and crowdsourced knowledge. Proceedings of the International Conference on Social Computing, 188-195. https://doi.org/10.1109/SocialCom.2013.35
40. Wan, Z., Tao, J., Liang, J., Cai, Z., Chang, C., Qiao, L., & Zhou, Q. (2019). Large-scale empirical study on machine learning related questions on Stack Overflow. *Journal of Zhejiang University (Engineering Science), 53*(5), 819-828. https://doi.org/10.3785/j.issn.1008-973X.2019.05.001
41. Wang, S., Lo, D., & Jiang, L. (2020). PostFinder: Mining Stack Overflow posts to support software developers. *Information and Software Technology, 127*, 106367. https://doi.org/10.1016/j.infsof.2020.106367
42. Wijerathna, L. et al. (2021). Mining and relating design contexts and design patterns from Stack Overflow. *Empirical Software Engineering, 27*(1). https://doi.org/10.1007/s10664-021-10034-0
43. Wikipedia contributors. (2026). DSpace. In Wikipedia. https://en.wikipedia.org/w/index.php?title=DSpace&oldid=1334257719
44. Zhu, W. et al. (2022). An empirical study of question discussions on Stack Overflow. *Empirical Software Engineering, 27*(6). https://doi.org/10.1007/s10664-022-10180-z

■■■