RESEARCH ARTICLE

International Journal of Computer Communication and Informatics

DOI: 10.34256/ijcci2516

Semantic Tag Clustering to Alleviate the Cold Start Problem in Learning Resource Recommendation: A Case Study on Delicious Dataset

Anisha Poly [a,*], P.K. Nizar Banu [a]

[a] Department of Computer Science CHRIST(Deemed to be University)Bangalore, India
* Corresponding Author: anisha.poly@res.christuniversity.in

Received: 23-01-2025, Revised: 02-04-2025, Accepted: 18-04-2025, Published: 30-04-2025

Abstract: This study explores a methodology for recommending learning resources, demonstrated through a case study on the Delicious dataset. Tags, representing keywords assigned to describe content, are semantically clustered using K-Means. Sentence Transformers are employed to generate dense vector representation of these tags, enabling more effective clustering. The system identifies meaningful tag groups to deliver relevant recommendations, even in the absence of user interaction history, effectively addressing the cold-start problem through predefined tag profiles. The proposed methodology personalizes resource recommendations for Deaf and Hard of Hearing (DHH) learners by leveraging their profile and resource Meta data. It enhances resource search during the cold start phase by identifying the most relevant tag cluster that matches with the learner's search query and retrieving preferred content based on the learner profile. Future extensions could incorporate dynamic preferences that evolve over time, enabling more adaptable and personalized recommendations. This work provides a robust foundation for clustering the resources based on their semantic meaning, thereby improving content-based search and retrieval of relevant learning resources.

Keywords: DHH, Tags, Semantic Search, Sentence Embedding, Personalized Learning, Tag Clustering.

1. Introduction

The advent of web technology with the rapid development of Internet, has significantly increased the information availability but has also led to an inflation of digital information, making the information retrieval a tedious task. The recommendation engines help personalize the web experience by filtering and presenting the information that a user might be interested in. Personalization plays an important role in eLearning platforms as well, ensuring the delivery of the right resources to the right user. So, the key question that need to be addressed here is how to identify the right resources specific to the user that drive the personalization. Personalized retrieval of the highly relevant resources from a pool can be achieved through collaborative methods, based on user interaction data, or by identifying similar-minded users. It also can be content based, analysing user profiles or resource metadata, or hybrid, combining both the techniques.

ELearning platforms are gaining immense popularity nowadays due to, its flexibility, adaptability, and accessibility anytime and anywhere, beyond the traditional classroom settings. They make education more democratic, catering to the diverse needs of learners and tailoring to their individual needs [1]. This makes the Deaf and hard of hearing (DHH) learners who experience hearing loss in speech frequencies, important stakeholders in eLearning. Deaf Individuals experience the hearing loss in speech frequencies in both ears 70DB while hard of hearing individuals typically experiencing the hearing loss ranges from 60 to 70DB [2]. Combined with the influence of educational settings, and the deafness state of family resulting in various preferences for communication and learning for DHH learners [3]. This diversity emphasize the limitations of one-size-fits-all learning approach for DHH population and underscores the need of more flexible e-learning solutions for them [4].

To ensure flexibility in eLearning solutions that accommodate the diverse needs of DHH learners, eLearning platforms should be enriched with digital learning content, typically referred to as learning objects (LO) [5]. The implementation of personalization lies in delivering the most relevant learning objects to users by considering the individual preferences of DHH learners in accessibility, communication and learning styles. Hence assigning the most suitable learning content based on the learner's choices and adjusting its difficulty according to their progress is one of the challenges in creating a personalized and adaptive eLearning system. Storing and managing of LOs becomes a significant task in the eLearning domain, as it helps learners quickly access relevant learning materials based on their preferences, which in turn boosts their engagement in online learning.

The LOs characterised by rich metadata offer a promising way to organize content, improve discoverability, and make the retrieval more robust [6]. The metadata often serves as keywords that describe the content or descriptors providing information about the type of content - such as exercise, quizzes, questionnaire, simulation, or readings – accessibility features, such as sign language, subtitles, or transcripts, difficulty levels, or modality, such as videos, text-based content, interactive elements or auditory formats. These metadata often described as tags, can be static or dynamic and allows the user to assign their own opinion about the LO [7] for their future retrieval. So, tags are the effective way to tackle the cold start problem for new users, where the insufficient interaction data for recommendations. They enhance keyword-based search by effectively retrieving resources and personalizing recommendations by capturing the user interest that evolve over time.

Many studies in the literature leverage static and dynamic tags for hearing users [9-11], but their potential remains underutilized for Deaf and Hard of Hearing (DHH) learners. The excessive amount of information on educational platforms can lead to disengagement among DHH learners. Assigning tags to learning resources helps filter DHH-accessible content effectively and personalize recommendations based on individual learning preferences.

This work bridges the gap by proposing a methodology for recommending eLearning resources through keyword searches, leveraging tag clustering and semantic similarity. The organization of this study is as follows: The Introduction provides an overview of how resource metadata aids in recommendations within an eLearning environment. The Related Works section discusses existing recommendation studies that utilize tags and tag-based clustering. The Methodology section outlines the approach applied to the Delicious dataset to deliver relevant recommendations without requiring interaction data. The Results and Discussion section presents the findings and insights derived from the study. Finally, the Conclusion and Future Scope highlight the key takeaways and potential directions for extending this research.

2. Related Works

Tags are user-contributed metadata that can help address the cold-start problem often faced by recommender systems. In an eLearning environment, tags can be categorized into two types: local and global [8]. Global tags are provided by the course creator at the time of uploading the learning object. These tags, such as title, description, or format, remain consistent across all users and are relatively static. For example, a learning object might be tagged with 'science,' 'grade12,' or 'HTML.' Local tags, on the other hand, are assigned by users to describe a learning object based on their interactions with the learning resources. These tags evolve over time, reflecting individual user input and preferences. Examples of local tags include 'easy,' 'interesting,' or 'interactive'.

T. B. Lalitha and P. S. Sreeja [1] enhance the recommendation of eLearning resources by automatically classifying online learning resources into different groups using clustering algorithms such as K-means, DBSCAN, and agglomerative clustering. They found that K-Means outperforms all the other methods they employ a combination of Term Frequency-Inverse Document Frequency (TF-IDF) and TextRank for keyword extraction. The text of the online learning content is first segmented using word segmentation techniques, and stop words are removed to clean the text. Keywords are then extracted from the resources and ranked to filter the most important words. Using the extracted tags, different clustering models are created with the aforementioned algorithms to classify the resources. These classifications are used to recommend resources based on keyword searches.

Tag clustering enables the grouping of resources based on descriptors provided for the learning objects. It helps identify and group resources with semantically related tags, making resource organization and retrieval more efficient. The clusters can be refined over time as users interact more with the resources. This refinement can consider tag weights associated with users, such as visiting frequency, to determine the most frequently used tags for different users. This approach helps personalize recommendations by adapting to evolving user preferences and behaviors [9].

According to the research [10], the tags are used to identify the topics by clustering the tags along with their semantic relationship and usage patterns. This helps to identify the latent preferences of users and their interaction with the resources on the topics. So, by clustering tags into semantically related topics helps to optimize the recommendation process.

Tags combined with the time information is explored to improve the recommendation results [11]. This study uses the tags as a way to identify the user's perceptions and categorization of content that indirectly captures the semantic relationships between tags and incorporates time-based weightings to track changes in user preferences over time. In the study conducted by the [12], Tags are used to refine recommendations by reducing the dimensionality of the tag space through grouping semantically related tags using the Latent Dirichlet Allocation (LDA) technique. This approach helps in organizing tags, identifying semantic similarities, and recommending items that align with users' past interactions.

The study [13] proposes an automatic method to generate multidimensional labels for educational resources by leveraging feature words extracted using Term Frequency-Inverse Document Frequency (TF-IDF) and TextRank algorithms. These feature words are utilized as tags to cluster the educational resources. Within each cluster, the relationships between the tags are analysed to ensure the selection of the most relevant tags. The extracted tag words are then used to automatically label the educational resources.

The studies discussed above aim to enhance eLearning resource recommendations by exploring the effective use of tag information. All of these studies seek to reduce the tag space by applying semantic relationships or usage patterns to form meaningful groups, thereby making resource retrieval more efficient. When a learner searches for a document, either keyword-based filtering among resources or collaborative filtering based on user interaction patterns—or a combination of both—can be employed to improve retrieval efficiency.

This survey underscores the significance of tags and their semantic clustering in resource retrieval, demonstrating their effective application in eLearning applications for our focused group of DHH learners. To cluster the tags based on their semantic similarity, and since we have numerical embeddings of the tag data, K-means is a more suitable technique. This is because K-means minimizes intra-cluster variance, which ensures that semantically similar tags are grouped together based on their vector representations. In contrast, methods like DBSCAN are more sensitive to the selection of parameters such as epsilon (ε) which defines the neighborhood size, and MinPts, the minimum number of points required to form a cluster. If these parameters which can result in most tags being classified as noise. Additionally, the tag embeddings being high dimensional, where all points are nearly equidistance from each other, making it harder to identify meaningful clusters. The tag-based semantic grouping, followed by filtering techniques, ensures the retrieval of learning resources that are highly aligned with their personal preferences and needs.

3. Methodology

The proposed work aims to streamline the selection of the most appropriate resources based on keyword search, often considered as tags, from the most similar cluster of resources. The resources are tagged with keywords and clustered based on these tag values, ensuring that the most semantically related learning resources fall into the same cluster. This approach can be easily adapted for DHH learners, as they tend to be more comfortable with short keywords. Additionally, this method remains beneficial even if tag details are not available for the user. In such cases, static keywords—metadata assigned to the resources—can be used to provide initial recommendations based on preferences encoded in the user profile.

3.1 Dataset

For the experimental analysis, a well-known bookmarking dataset, Delicious is taken as it includes the web bookmarks by users. The dataset is released in the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011) [14, 15]. The dataset statistics are given in the Table 1 [16].

Table 1. delicious data statistics

Total no: of Users	1867
Total no: of URLS	69226
Total no: of Tags	53388
Total no: of Bookmarks	104799

Table 2. Filtered Elearning Data Statistics

Total no: of URLs	44501
Total Unique Keys	11850
Total Unique Tags	515
Average Tags per Bookmark	3.79
Maximum Tags per Bookmark	13
Minimum Tags per Bookmark	1

Here, each bookmark is based on the URL, a web page link with the following metadata: title and user assigned keywords. For the proposed work, the tag file that contains the set of tags available in the dataset and bookmarks file that contains the details of bookmarks and bookmark_tags file contains the tags assigned to the bookmarked URLs, and the number of times the tags assigned to the particular resource.

3.2 Data Preprocessing

Data preprocessing is the first and important step to prepare the data before applying any learning algorithm. As the Delicious dataset contains more generic tag words like "potential"," self-awareness","confirmation" etc to describe the bookmarks. To align with the objective of the work, eLearning related tag words from the 'tag file' are filtered first.

To identify the tags related to the eLearning content by checking the semantic similarity between a set of predefined seed keywords and tags present in the tag dataset. A sentence transformer model was employed to get the word embeddings for both the seed keyword and actual tag data [17]. This model is well-suited for capturing the semantic meaning of text, enabling efficient comparison between the tags and the seed keywords. Next, cosine similarity was computed between the embeddings of tag words and seed key words. This is chosen over the other metrics due its efficacy to capture the semantic closeness. Unlike Euclidean distance, which is sensitive to magnitude differences, cosine similarity measures the closeness of two vectors in the embedding space, providing a reliable metric for assessing semantic similarity. Tags with a similarity score above the threshold value of 0.5 were considered relevant to eLearning content and were included in the filtered tag list. The threshold value 0.5 was selected to ensure the balance between retrieving the relevant tags while minimizing noise such as unrelated matches or exclusion of potential tags. Then dropping of duplicates tags and normalization of tags were performed.

The next step involved merging the filtered tags with the bookmark_tag file to associate the corresponding bookmarks with the filtered tags. The tags in the merged data were then transformed into keys, with each key representing a combination of tags assigned to a particular bookmark. Since multiple tags can be assigned to the same bookmark, the tags were grouped based on their respective bookmarkID. For example, for bookmarkID 7, the tags assigned included "literacy," "education," "learning," "technology," and "resources." To ensure consistency, the tags in each key were sorted alphabetically, as the order of tags holds no significance. For instance, the key "education software teaching technology" is equivalent to "software teaching technology education." Sorting eliminates discrepancies due to tag order.

Finally, duplicate checks were performed on the keys to retain only unique combinations resulting in 11,850 bookmarks with unique combination of keys. This step ensures the integrity and quality of the data, which is essential for meaningful clustering and subsequent analyses. The most frequent keywords includes education, web, technology, learning, resources, and programming. A detailed statistics of filtered dataset and tag frequency distribution is shown in table 2 and figure 1, respectively.

3.3 Tag Clustering Using Semantic Similarity

3.3.1 Sentence Embedding

To compute the semantic relations between the tags in our dataset, we utilized the Sentence Transformer model: all-MiniLM-L6-v2, which effectively captures contextual and semantic similarities in textual data [18]. The effectiveness of this model was verified through experiments on a synthesized dataset of 300 tags, where it demonstrated strong performance in capturing the semantic meaning of the tags, as well as facilitating accurate clustering.

In this process, the unique set of keys (combinations of tags) from the extracted dataset were input into the model. This model generates 768-dimensional dense embeddings, where each embedding vector encodes the semantic relationships and contextual meaning of the corresponding tag within the key. These embeddings provide a numerical representation of the tags, which makes it easier to measure their similarity.

To ensure the consistency and accuracy of the similarity computation, the embeddings were normalized. Normalization ensures that each embedding vector is scaled uniformly, which is crucial for improving the quality of clustering and similarity search. The scaling phase significantly contributes to the robustness of the clustering algorithm, ensuring that similar tags are grouped together, and enhancing the overall quality of the semantic similarity measures.

3.3.2 K-means Clustering

The K-means clustering was performed on the normalized embeddings of the unique keys. The elbow method was employed to determine the optimal number of clusters (k). The graph illustrates the relationship between the number of clusters (on the x-axis) and inertia (the sum of squared distances, on the y-axis). Inertia is used to measure how well the data points are separated within the clusters. Initially, inertia is high, and as more clusters are added, inertia tends to decrease because the distances between points are reduced. The 'elbow' is the point where the curve begins to flatten, indicating that adding more clusters no longer significantly improves the model. The optimal number of clusters is chosen at the elbow point, where there is a balance between the number of clusters and the model's ability to avoid overfitting. Here, we chosen the elbow at '15' where the curve starts to flatten out and the plot shown in the figure 2.

K-means clustering, a centroid-based partitioning method, was applied to group the bookmarks based on the tags assigned to them. K-means is a simple and powerful clustering model that performs well on tag embeddings and groups the data based on Euclidean distance [19]. Here, K was selected as 15 based on the elbow plot, which suggested the optimal number of clusters and identified meaningful clusters based on the semantic similarity between the tags. So, the related resources are grouped in the same cluster. For example, the bookmarks tagged with javascript, jquery, Ajax and html are grouped in one cluster while the bookmarks related with social media, face book, and internet will be grouped into another cluster. Figure 3 shows the snapshot of clustering result.

3.3.3 Topic Modelling

To describe the clusters and assign meaningful names, we employed topic modeling, as no ground truth is available for the bookmarks. Latent Dirichlet Allocation (LDA) is a commonly used method to uncover the latent relationships among the tags [20]. After clustering the resources, the centroids of the tag clusters represent the average vector of the tags within each cluster, the theme of the cluster. By analyzing the data points closely related to the cluster centroids, we can uncover the nature of each cluster. Analyzing these closest points helps identify the topics or concepts associated with each cluster, allowing us to interpret the cluster labels accordingly. Table 3 shows the cluster analysis and labelling using topic modelling.

3.3.4 Similarity Search

The similarity search component implements the keyword-based search in the tag clusters and will retrieve the bookmarks that has similar keys to the searched query. The system incorporates the pertained sentence transformer model for the semantic embedding's of the searched keyword and employs the cosine similarity to match the search tags in the resource clusters. The steps involved are:

3.3.4.1 Tag Embedding Generation

Semantic embeddings are generated for a seaerch query using the sentence transformer model employed in the key embeddings in the previous step. This embedding, a high-dimensional vector, captures the essence of the search tag. Despite their visible differences, this method ensures that similar tags are positioned close together in a vector space, reflecting their semantic relationships. Before further processing, the embeddings are normalized to unit length to ensure they are compatible.

3.3.4.2 Identification of Relevant Cluster

The power of clustering is utilized here to reduce the search space and the system leverages this predefined resource clusters, each identified by a cluster label. By employing the same embedding model, the vector space of cluster labels are generated and cosine similarity is applied to find a semantic match between the cluster and search keyword. The clusters that are most relevant to the search query are identified by the highest similarity score above the threshold value. Clusters do not pass the similarity score will eliminate from the search space.

3.3.4.3 Resource Filtering by Cluster

After identifying the top N relevant clusters, the system filters the bookmarks or resources to include only the resources linked to these clusters. This process guarantees that the recommendation space is concentrated solely on the most relevant resources.

3.3.4.4 Cosine Similarity Calculation with the Filtered Resources

The embeddings of the keys of resources are retrieved from the precomputed set of resource embeddings following the filtration of resources from the relevant clusters. The cosine similarity is then calculated between the key embeddings and search keyword embeddings. This process assesses the semantic closeness between the search term and the resources within the relevant clusters.

3.3.4.5 Ranking and Recommendation Generation

The top-N resources are generated based on their similarity score against the search keywords and the resources with highest similarity scores will be selected and recommended to the learners.

4. Results & Discussion

This section discusses the results of clustering and similarity search.

4.1 Cluster Analysis

The clustering employed was K-Means and was able to form 15 clusters as shown in table 3, each representing a range of related topics. Certain clusters, such as Cluster 0 and Cluster 9, focus on web development technologies, whereas others, like Cluster 1, clearly prioritize e-learning and education. But there exists some overlap and redundancy between the clusters in terms of top tags. Cluster 0 and 9 contain both the tags related to javascript, web, and programming. But the clusters don't have any tags in common, but the Euclidean distance between their centroids is less and has high cosine similarity. This suggests that both the clusters share the common topic, web development but with different focuses such JavaScript and HTML software.

Additionally, cluster 1 and 3, which focus on education and teaching, also share similarity in terms of their tags, though they focus on the slightly different topics such as lesson plan and general teaching. Cluster 4 (social media), characterized by tags related to Facebook and social media, appears quite specialized; however, it could intersect with other clusters that include topics related to the internet or technology. Cluster 5(Education and Teaching) includes various topics related to education, schools, and teaching. While this cluster is useful as it stands, it could potentially be subdivided into more specific subclusters, such as teaching resources versus learning resources or K–12 education in comparison to higher education.

Overall, the K-means clustering was able to identify the broad themes among the tagged bookmarks but still there is a room for the further refinement of clusters. This can be done by employing splitting, or merging clusters, or experimenting with the other portioning or hierarchical algorithms.

4.2 Similarity Seacrh Results

We carried out a similarity search using the query 'machine learning and data science' to find the most pertinent resources by analyzing tag similarities within the clusters. The results we got:

Recommendation1: Data mining, forecasting and bioinformatics competitions on Kaggle

Similarity Score: 0.703409

Cluster: Computer Science, AI, and Data Science

Description: This resource provides insights into bioinformatics competitions on Kaggle, which align with machine learning applications in data science.

Recommendation 2: Preserving science: what to do with raw research material?

Similarity Score: 0.695902

Cluster: Computer Science, AI, and Data Science

Description: A discussion on how raw research materials in data science should be preserved, which complements machine learning projects that handle large datasets.

Recommendation 3: Home | OpenHazards

Similarity Score: 0.690175

Cluster: Computer Science, AI, and Data Science

Description: An introduction to hazard management systems, which may be of interest to data scientists working with risk data in machine learning models.

Here the recommendations are aligned with the theme of the search keywords and suggested the resource ranging from practical applications to theoretical aspect. Even though the semantic-based keyword clustering and similarity search significantly improves the search result for a query, challenges such as scalability issues in large dataset, semantic drift overtime, lack of explainability, and resource-intensive training remain. These can be mitigated by dimensionality reduction techniques, fine tuning the models using new search queries and user interactions, providing example-based solutions to the recommendations, and using the precomputed embeddings instead of computing them for every query [21].

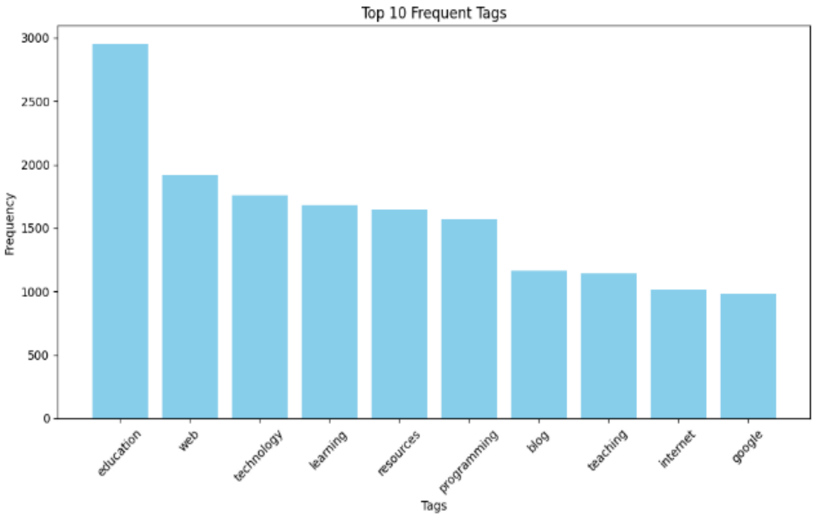

Figure 1. Top frequent tags

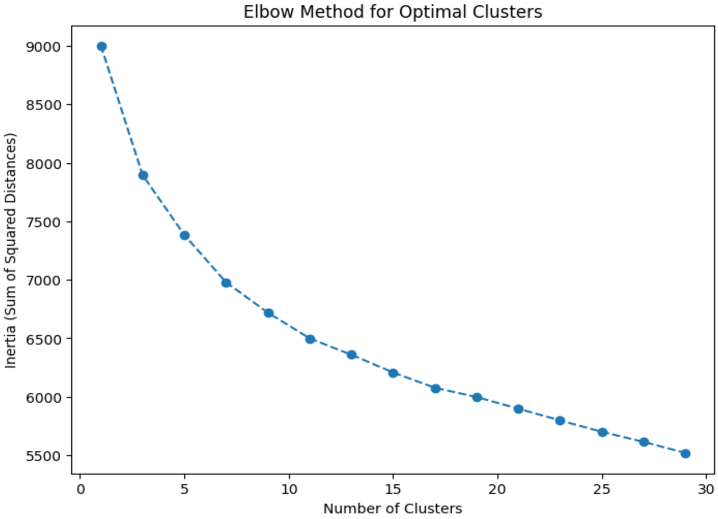

Figure 2. Elbow plot for optimal number of clusters

bookmarkID	key	cluster
7	curriculum education learning lessonplans lessons literacy resources technology	3
9	books kids literacy reading	8
11	education software teaching technology	6
15	books	8
16	education lesson lesson plans lessonplans teaching	3
17	curriculum education lessonplans math resources science teaching	3
18	curriculum education lessonplans lessons resources teaching technology	6

Figure 3. Snapshot of clustering results

4.3 Discussion

The proposed methodology is effective to handle the cold start problems through pre-defined tag profiles. Even though the current system only captures the keywords describing the bookmarks, other tags such as content type, modality, difficulty level etc. could be easily added to the resource metadata and the resources can be grouped based on the combination of keywords added to the resources. So, if a user visiting the eLearning platform for the first time, these precomputed tag profiles will be able to deliver the preference-based contents based on the similarity check between the user search keywords and tag profiles. This will enhance the user experience right from the initial interaction with the learning platform.

Due to its ability to handle new users with minimal interaction data, the system provides flexibility in offering valuable recommendations tailored to their specific needs and preferences. This approach is beneficial not only for new users but also for inactive users who lack sufficient interaction data with the platform.

Table 3. Cluster Analysis And Labelling

Cluster	Closest Tags	Cluster Label
0	javascript, jquery, ajax, html	Web Development and Programming
1	elearning,education,learning, online	E-Learning and Education Resources
2	Blog,technology web,internet	Technology and Blogging
3	lesson plan, math, resources, science	Educational Resources and Lesson Plans
4	facebook,internet socialmedia,business	Social media and Marketing
5	education, culture,resources school	School and Classroom Teaching

6	education software teaching technology	Educational Technology and Resources
7	Software, coding, data, software	Programming and Computer Science
8	books kids literacy reading,	Literacy and Educational Books
9	online web webapp , html	Web Programming and Web Development
10	Internet,google,yahoo,chrome	Internet Search and Online Technologies
11	Blog, education, technology, training videos	Educational Videos and YouTube
12	Academics, android, google. mobile technology	Mobile and Web Technologies
13	brain science, ai books, machine learning, datavisualization	Computer Science, AI, and Data Science
14	marketing writing, business,career jobs, economic science	Business, Economics, and Marketing

By reducing the search space through clustering tags, the system can dynamically adapt to evolving user preferences as learners progress through the platform. Incorporating user tagging weights and identifying similar users with comparable interaction patterns within clusters can further enhance the recommendations.

As the number of tags in the learning environment increases, the clustering algorithms like K-Means might struggle with scalability. It requires the pairwise similarity to group the resource and this may be computationally expensive and time consuming. So, as the system grows optimization techniques might helpful to reduce the computation time and still holds the clustering quality.

5. Conclusion & Future Scope

The methodology discussed here effectively addresses the challenges posed by the cold-start problem, making it valuable for new or inactive users by providing meaningful recommendations even in the absence of interaction data. The use of predefined tag profiles and clustering techniques ensures the delivery of relevant resources, even when a user's interaction history is unavailable. Furthermore, clustering-based reduction in the tag space enhances the system's search efficiency and scalability. This methodology is particularly well-suited for Deaf and Hard of Hearing (DHH) learners, enabling recommendations tailored to their preferences in accessibility, communication modes, and learning styles.

As learners progress through the system, their preferences and learning goals may evolve. The current system, relying on predefined rules, should be enhanced to capture this dynamic nature. For instance, a learner initially interested in video-based materials may later prefer interactive content. Similarly, if a learner struggles with a specific topic, the system should recommend simpler resources or alternative formats. Conversely, advanced learners should be directed toward more challenging content. Refining user profiles to account for this dynamic behavior is crucial to ensuring recommendations remain adaptable and relevant.

Another potential area for improvement is the integration of collaborative filtering methods with the current tag clustering system. Incorporating user tagging weights into the resource matrix would allow the system to dynamically adjust cluster profiles and identify neighboring users with similar interaction histories. This capability could enable the system to adapt recommendations in real time, aligning with evolving user preferences. Such a hybrid approach would not only enhance the system's ability to meet individual learning needs but also foster greater engagement by providing accurate and relevant suggestions tailored to users at a granular level.

Overall, the proposed system lays a robust foundation for effective personalized and adaptive recommendations. With significant potential for refinement, particularly through the incorporation of dynamic user preferences and collaborative filtering, the system promises to become even more effective in addressing diverse learner needs while maintaining relevance and adaptability over time.

References

[1] Lalitha, T.B., Sreeja, P.S. (2024) .Keyword Extraction and Pattern Model Identification on Online Learning Contents for Classification to Enhance Microlearning Concepts and Obtain Personalized eLearning Contents, *International Journal of Engineering Trends and Technology*, 72(3), 230–248. https://doi.org/10.14445/22315381/IJETT-V72I3P121

[2] Government of India, Department of Empowerment of Persons with Disabilities, (2016). The Rights of Persons with Disabilities Act.

[3] Singh, S. (2019). Revisiting Inclusion of Children with Hearing Impairment: Issues and Possibilities, in Disability, Inclusion and Inclusive Education, *Springer*, 157–174. https://doi.org/10.1007/978-981-15-0524-9_8

[4] Poly, A., Banu, P.K.N. (2024). Discovering the Micro-Clusters from a group of DHH learners. *In Advances in educational technologies and instructional design book series*, 159–175. https://doi.org/10.4018/979-8-3693-0868-4.ch010

[5] Joy, J., Renumol, V. (2021). Mapping of learning style with learning object metadata for addressing cold-start problem in e-learning recommender systems. *International Journal of Learning Technology*, 16(4), 267. https://doi.org/10.1504/ijlt.2021.121364

[6] Mendez, N.D.D., Morales, V.T., Vicari, R.M. (2016). Learning object metadata mapping with learning styles as a strategy for improving usability of educational resource

repositories. IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje, *IEEE*, 11(2), 101–106. https://doi.org/10.1109/rita.2016.2554038

[7] Klašnja-Milićević, A., Ivanović, M., Vesin, B., Budimac, Z. (2017). Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques. *Applied Intelligence*, 48(6), 1519–1535. https://doi.org/10.1007/s10489-017-1051-8

[8] Vairavasundaram, S., Varadharajan, V., Vairavasundaram, I., & Ravi, L. (2015). Data mining-based tag recommendation system: an overview. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, 5(3), 87–112. https://doi.org/10.1002/widm.1149

[9] Selvakumar, S., Inbarani, H., Shakkeel, P.M. (2016). A Hybrid Personalized Tag Recommendations for Social E-Learning System, *International Journal of Control theory and applications*, 9(2) , 1187–1199.

[10] Pan, X., Zeng, X., Ding, L. (2022). Topic optimization–incorporated collaborative recommendation for social tagging. *Data Technologies and Applications*, 58(3), 407–426. https://doi.org/10.1108/dta-11-2021-0332

[11] Jo, H., Hong, J., Choeh, J.Y. (2023). Improving Collaborative Filtering Recommendations with Tag and Time Integration in Virtual Online Communities. *Applied Sciences*, 13(18), 10528. https://doi.org/10.3390/app131810528

[12] Bogers, T. (2018). Tag-Based recommendation. In Lecture notes in computer science, 441–479. https://doi.org/10.1007/978-3-319-90092-6_12

[13] X. Pan and Y. Shi, Automatic Generation of Multidimensional Labels of Educational Resources Based on Grey Clustering In *International Conference on E-Learning, E-Education, and Online Training*, Springer Nature Switzerland, 160-174. https://doi.org/10.1007/978-3-031-51471-5_11

[14] ACM Recommender Systems. (2020). *RecSys* . https://recsys.acm.org/recsys20

[15] GroupLens. (2013). *HetRec2011 dataset*. https://grouplens.org/datasets/hetrec-2011

[16] Waskow.Canada.Ca, Slot Online 24 Jam Deposit Pulsa & Judi Terpercaya, Delicious.com, 2021. http://www.delicious.com

[17] Devika, R., Vairavasundaram, S., Mahenthar, C.S.J., Varadarajan, V., Kotecha, K. (2021). A deep learning model based on BERT and sentence transformer for semantic keyphrase extraction on big social data. *IEEE Access*, *9*, 165252–165261. https://doi.org/10.1109/access.2021.3133651

[18] Khosa, S., Mehmood, A., Rizwan, M. (2023). Unifying sentence transformer embedding and Softmax voting ensemble for accurate news category prediction. Computers, 12(7), 137. https://doi.org/10.3390/computers12070137

[19] Merlini, D., Rossini, M. (2021). Text categorization with WEKA: A survey. *Machine Learning With Applications*, *4*, 100033. https://doi.org/10.1016/j.mlwa.2021.100033

[20] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L. (2018). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

[21] Patwardhan, N., Marrone, S., Sansone, C. (2023). Transformers in the real world: A survey on nlp applications. *Information*, 14(4), 242. https://doi.org/10.3390/info14040242

Funding

No funding was received for conducting this study.

Conflict of interest

The Author have no conflicts of interest to declare that they are relevant to the content of this article.

About The License