

Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Research on English Reading Comprehension Material Recommendation System under Text Similarity Algorithm

Ruili Chu^{1,†}

1. College of Weinan Normal University, Weinan, Shaanxi, 714099, China.

Submission Info

Communicated by Z. Sabir

Received February 1, 2025

Accepted May 7, 2025

Available online June 5, 2025

Abstract

With the development of information technology and the change of people's education concept, personalized learning is getting more and more attention. The traditional classroom is difficult to realize the need of personalized English reading comprehension material recommendation, this paper designs a fusion English reading comprehension material recommendation system based on collaborative filtering algorithm and improved text similarity algorithm. Aiming at the problem that the traditional text similarity algorithm ignores the user's differentiated attention to the information of English reading comprehension materials in text matching, the Inter-TF-IDF algorithm, which integrates the calculation of concentration and dispersion, is proposed. Combining this algorithm and the user-based collaborative filtering algorithm, the fusion recommendation system of this paper is proposed. The system utilizes the collaborative filtering algorithm to get the preliminary recommendation results of English reading comprehension materials, and then utilizes the improved Inter-TF-IDF algorithm to calculate the similarity between the text of the English reading comprehension materials and the text browsed by the user, and selects the English reading comprehension materials with a high degree of similarity as the final recommendation results. The overall recommendation accuracy of the system in this paper is maintained at a high level of 0.82-0.95. In the actual application effect, it significantly improves the English scores of students using the system, and obtains a high degree of satisfaction from the students. The recommendation system in this paper has a good promotion and application prospect in the field of English learning.

Keywords: Collaborative Filtering Algorithm; Text Matching; Text Similarity Algorithm; Inter-TF-IDF Algorithm.
AMS 2010 codes: 97B20

[†]Corresponding author.

Email address: 619362073@qq.com

ISSN 2444-8656



https://doi.org/10.2478_amns-2025-1029



© 2025 Ruili Chu, published by Sciendo.



This work is licensed under the Creative Commons Attribution alone 4.0 License.

1 Introduction

Nowadays, every stage of education and other fields pay great attention to cultivate the comprehensive ability of students, with the continuous reform and development of the new curriculum standard of English, English reading comprehension is also gradually emphasized [1-2]. English reading not only examines the memorization process of students' knowledge acquisition, but also pays more attention to the examination of students' comprehensive assessment ability, thinking ability and reasoning ability [3-4]. Cultivating students' higher-order thinking ability in English reading teaching plays an important role, which is an ability that needs to be focused on in the current teaching process. Based on the complexity, flexibility, and diversity of the comprehensive application of the English language subject, a large amount of in-depth reading is indispensable for further improving students' English learning level and guiding them to explore the cultural connotations behind different language forms on this basis [5-7]. In order to further improve students' reading quality and practicality of reading, it is necessary to select appropriate reading comprehension materials from a wide range of English reading materials. Suitable reading materials can meet the needs of students' characteristics, increase students' learning interest, expand vocabulary, improve the sense of language and grammatical comprehension, and strengthen students' linguistic thinking, so as to improve the effect of reading and create favorable conditions for the cultivation of students' core literacy [8-10]. For this reason, it is necessary to develop a system that intelligently recommends English reading comprehension materials.

In the era of networking, algorithms change people's work and life, in which text similarity algorithms are becoming more and more important and their application is becoming more and more common in the era when the network is increasingly penetrating into every aspect of people's lives. As long as there is knowledge or information in the environment, it is possible to use the algorithm. The most typical current applications are in the fields of intelligent translation, answering system, knowledge retrieval, document classification, etc., and the applications in each of these fields are the most basic ones [11-13]. All other kinds of applications are based on the text similarity algorithm. The algorithm is able to realize matching between different information and find the target information, which solves the contradiction between massive knowledge and precise demand, fast retrieval demand and computational efficiency, and time-consuming and laborious manual operation and efficient and fast automatic machine computation [14].

English has been emphasized for a long time in the countries of second language scholars, and the recommendation of materials has become a hot topic as a result. There are relatively few studies on special recommendation of English reading materials, but there is no lack of research on the recommendation of various materials in English. Takii et al [15] used information retrieval technology to formulate an interpretable English material recommendation method, realizing the recommendation of tracking the dynamic change of students' learning, which makes the difficulty of recommendation in line with the students' current level of knowledge. Bai and Xian [16] introduced natural language processing (NLP) technology to build an automatic classification and recommendation system for English teaching materials, which is carried out by collaborative filtering and content-based filtering in the recommendation part, with the former based on the user's interactive behaviors in the system to identify the recommendation and the latter through the features extracted by NLP, and the combination of the two constitutes a hybrid recommendation module, which enhances the effectiveness of the recommendation. Zhang [17] proposes a method for the automatic recommendation of English educational resources, which is automatically update the education-related materials in major English resource libraries through a multi-objective particle swarm optimization algorithm, which cleans, recognizes, and classifies these materials via NLP technology. Escobar-Acevedo et al [18] proposed an automatic classifier under the Common European Framework of Reference for Languages (CEFR) hierarchy, which can be used to classify

the English texts in web resources, and sort them according to the degree of difficulty to achieve specific recommendation.

In reading material recommendation, Liu et al [19] used students' historical reading comprehension assessment as a reference, collaborative filtering was used to evaluate the historical reading comprehension of other students at the same level, and a reading comprehension recommendation system was constructed through machine learning, which could perform feature extraction on the material text for assessing students' reading comprehension level. Wang et al [20], based on the theory of inter-district development, and with ChatGPT's support, developed a personalized English reading comprehension support system, which provided reading comprehension materials that were generated with reference to previous materials, and the difficulty level of the materials was benchmarked against the students' current level. Wu and Huan [21] designed a personalized English reading recommendation system based on adaptive learning, which included basic vocabulary searching, marking notation, highlighting, and other functions, and most importantly, the system recommends reading materials for students' individual needs after analyzing students' ability level, article difficulty and relevance. Cheng and Wang [22] used machine learning to design an adaptive recommendation model for collated fragmented reading resources, analyzed students' needs under a generative decision tree algorithm, and carried out adaptive recommendation of reading resources.

In this paper, a fusion English reading comprehension material recommendation system is designed by combining a user-based collaborative filtering algorithm and an improved text similarity calculation method, which improves the original text similarity algorithm in terms of concentration and dispersion, and increases the algorithm's emphasis on the features of lexical item distributions within independent categories of text and the features of lexical item distributions between categories. The cosine similarity algorithm is used to estimate the cosine value of the vector space angle between samples, and the value is used to indicate the similarity between samples. The functional boards of the system were designed from two functional levels: front-end interaction subsystem and back-end management subsystem. The specific value of this paper's work is verified by the performance validation of the algorithm and the investigation of the application effect of the English reading comprehension material recommendation system.

2 English reading comprehension material recommendation system based on fusion recommendation

2.1 Collaborative Filtering Algorithm

Collaborative filtering algorithm (CF) is a widely used algorithm in the field of recommendation algorithms, mainly based on the target user's past behavioral information, through the nearest neighbor algorithm, to recommend to the target user the current content that may be of interest to the target user, the common ones are the user-based collaborative filtering (User CF) [23] and the item-based collaborative filtering (Item CF) [24].

User based collaborative filtering algorithm is one of the earliest recommendation algorithms, if users have similar preference profile then they like similar items. The process of User CF algorithm is depicted in Fig. 1.

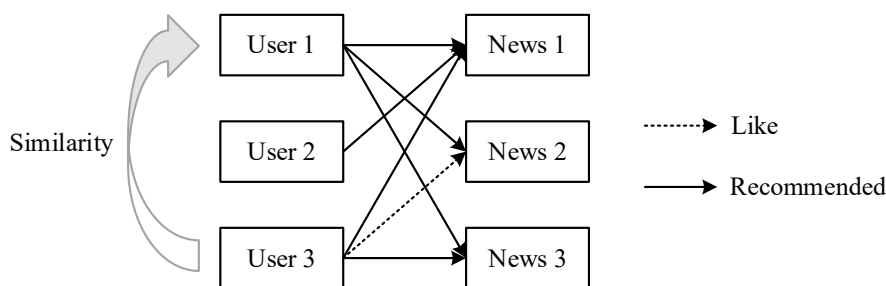


Figure 1. Schematic diagram of User CF algorithm

From the figure, we can see that user 3 is the target user who needs to make recommendation, and he is interested in English reading comprehension material 1 and English reading comprehension material 3, user 1 is interested in English reading comprehension material 1, English reading comprehension material 2 and English reading comprehension material 3, user 2 is interested only in English reading comprehension material 1, and user 1 and the target user 3 are interested in more English reading comprehension materials together. Therefore, user 1 and target user 3 have higher similarity and belong to similar users, and the English reading comprehension material 2 that user 1 is interested in is recommended to the target user. User CF algorithm is mainly divided into two steps:

- 1) Calculate the similarity between the target user and other users through the target user's historical behavior information, and construct a nearest neighbor set based on the similarity.
- 2) According to the target user's nearest neighbor set, find out the English reading comprehension materials that the target user has not clicked on, calculate the target user's interest level in these English reading comprehension materials, and recommend the first N English reading comprehension materials with high interest level to the user.

User CF algorithm is relatively simple to implement, does not take into account the content information of the English reading comprehension materials and the relevant attributes of the user, with the increase in the number of users, the calculation of user similarity becomes difficult, and the efficiency of the algorithm decreases.

Item CF predicts the items that target users are interested in by analyzing the relationship between different items and constructing the similarity set of items.

User 1 is interested in English reading comprehension material 1, English reading comprehension material 2 and English reading comprehension material 3, user 2 is interested in English reading comprehension material 1 and English reading comprehension material 2, and user 3 is the target user and is interested in English reading comprehension material 1. It can be found that users who are interested in English reading comprehension material 1 are interested in English reading comprehension material 3, and it is considered that English reading comprehension material 1 and English reading comprehension material 3 are similar, so English reading comprehension material 3 is recommended to the target user. Item CF algorithm is also mainly divided into two steps:

- 1) Calculate the similarity between English reading comprehension materials through the user's behavioral information on English reading comprehension materials, and construct the nearest neighbor set of similar English reading comprehension materials.

- 2) Through the set of nearest neighbors of English reading comprehension materials and the user's historical behavior, find out the English reading comprehension materials that the user has not clicked, calculate the user's interest in these English reading comprehension materials, and generate a recommendation list for the user.

2.2 TF-IDF algorithm

The TF-IDF [25] algorithm is a common statistical method often used to assess the importance of lexical items in a corpus, which can be used as text feature extraction. As in equation (1):

$$W_{i,j} = \frac{n_{i,j}}{\sum_i n_{i,j}} \log \frac{N}{N_i} \quad (1)$$

Where $n_{i,j}$ represents the number of occurrences of word j in document i , N represents the total number of documents in the corpus, and N_i represents the number of documents containing word j . The algorithm considers that the importance of the lexical items is not only related to the frequency of the lexical items in the text, but also related to the distribution of the lexical items in the documents. The TF-IDF algorithm is often applied in the tasks of text categorization, sentiment analysis, and real-time prediction of multiclassification.

2.3 Inter-TF-IDF algorithm

Although the above TF-IDF-based algorithm achieves better classification results on several tasks, it ignores the distribution of lexical items within a certain category of text and the distribution of lexical items between categories. This shortcoming is not prominent on the tasks of text categorization or sentiment analysis, because all the analyzed text information comes from the same side of the text, and there is no process of comparing the two sides of the text. However, in the search task, since the text to be analyzed is derived from two different texts, one from the textual information of the English reading material and the other from the learner's information, more attention needs to be paid to the distribution of lexical items among categories. For example, a learner may only be interested in a certain category of English reading materials, so the relevant lexical items of that category of English reading materials should have a higher weight in characterizing that learner. Based on this, this paper proposes an improvement method for the traditional TF-IDF algorithm in terms of concentration and dispersion.

2.3.1 Concentration

This paper argues that not only are lexical items more representative when they occur more frequently, but they are also more representative when they occur in only a few categories of text. For example, for learners focusing on English reading materials in a certain category domain, the descriptive information should be more frequent in terms of lexical items related to that category. If a lexical item is concentrated only in texts related to a certain category of English reading materials, then the word is more representative, so the concentration can be expressed by the formula:

$$D_c = \frac{|C|}{|C_i| + \alpha} \quad (2)$$

where $|C|$ denotes the number of English reading comprehension materials, and $|C_i|$ denotes the number of English reading comprehension material categories in which lexical items i appear in the descriptive information of the English reading comprehension materials. The size of $|C|$ is constant in the case of a given data set. The smaller $|C_i|$ and the larger D_c indicate that the lexical item is more representative. α is a model hyperparameter, which is to prevent the denominator from being zero.

2.3.2 Dispersion

This paper argues that if the lexical item occurs only in a certain category of textual information, the lexical item is strongly representative. However, if the lexical item appears in only a few texts in this category, the lexical item is less representative. For example, in the descriptive information of English reading comprehension materials in a particular domain, descriptive information of other classes will appear less frequently. In this paper, we use dispersion to describe this feature with the following formula:

$$D_s = \frac{|S_i|}{|S|} \quad (3)$$

Where $|S_i|$ indicates the number of occurrences of lexical item i in the descriptive text of a certain type of English reading comprehension material, and $|S|$ indicates the number of descriptive texts of that type of English reading comprehension material. The smaller the value of $|S|$, the larger the value of $|S_i|$, and the larger the value of D_s , the greater the dispersion of the lexical item and the more representative it is.

2.3.3 Inter-TF-IDF algorithm

Based on the above definition and analysis of concentration and dispersion, and combining the advantages of traditional TF-IDF, this paper proposes an improved TF-IDF algorithm, Inter-TF-IDF, which is calculated as shown in Equation (4):

$$\begin{aligned} Inter - TF - IDF &= D_c * D_s * TF - IDF \\ &= \frac{|C|}{|C_i| + \alpha} * \frac{|S_i|}{|S|} * \frac{n_{i,j}}{\sum_i n_{i,j}} \log \frac{N}{N_i} \end{aligned} \quad (4)$$

The weight of a word item is the product of concentration, dispersion and TF-IDF. Due to the consideration of concentration and dispersion, when a lexical item is more representative in the text, its Inter-TF-IDF calculated value will be larger compared to the TF-IDF calculated value.

2.3.4 Similarity calculation

After the processing of text information extraction algorithm, the similarity of two samples needs to be calculated to compare different samples. In this paper, the cosine similarity algorithm [26] is used to calculate the cosine value of the angle between the two samples in the vector space to represent the similarity of the two samples, the closer the cosine value is to 1, the more similar the two samples are. Its calculation formula is shown in (5):

$$\cos\theta = \frac{a \cdot b}{|a| * |b|} \quad (5)$$

Where a and b denote the feature vectors of the learner and the English reading comprehension material respectively, $a * b$ denotes the dot product of the learner and the English reading comprehension material vectors, and $|a| * |b|$ denotes the product of the vector modes. In the English reading comprehension material recommendation system, by calculating the cosine similarity between the English reading comprehension material feature vector a and the learner vector b , the system recommends the English reading comprehension material with the top ten similarity to the learner's descriptive text to the learner.

2.4 Fusion Recommender System Based on Text Similarity

The fusion recommendation algorithm proposed in this paper is a fusion of user-based collaborative filtering algorithm and improved text similarity algorithm. Firstly, the user-based collaborative filtering algorithm is used to get a preliminary recommendation result, and then the user browsing records are obtained from the front-end of the webpage, and then the improved text similarity algorithm is used to calculate the similarity between the text of the English reading comprehension materials in the preliminary recommendation list and the text of the users' browsing of the English reading comprehension materials, and the English reading comprehension materials with a high degree of similarity in the preliminary recommendation list are used as the final recommendation result. The structure of the fusion recommendation system based on the similarity of this text is shown in Figure 2.

The algorithm flow is as follows:

Input: user-English reading comprehension material rating matrix.

Output: a collection of recommended English reading comprehension materials.

In the first step, the user-English reading comprehension material rating matrix is utilized to obtain a preliminary list of recommended English reading comprehension materials according to the formula of collaborative filtering algorithm in this paper.

The second step is to extract the feature information of an English reading comprehension material in the preliminary recommended list and integrate it into an information text.

In the third step, the feature information of users browsing English reading comprehension materials is integrated into one information text.

In the fourth step, the similarity of two texts is calculated using the Inter-TF-IDF algorithm for improving text similarity algorithm mentioned in this paper.

In the fifth step, the preliminary recommended English reading comprehension material IDs and their corresponding text similarities are saved in the form of key-value pairs in the dictionary, corresponding to the keys and values in the dictionary, respectively.

In the sixth step, repeat the second to fifth steps until all the English reading comprehension materials in the preliminary recommendation list have been computed, and a dictionary collection L about English reading comprehension material-text similarity is obtained.

Step 7, sort the dictionaries in the set L in descending order of their values.

In the eighth step, the top N ranked ones are selected as the final recommendations.

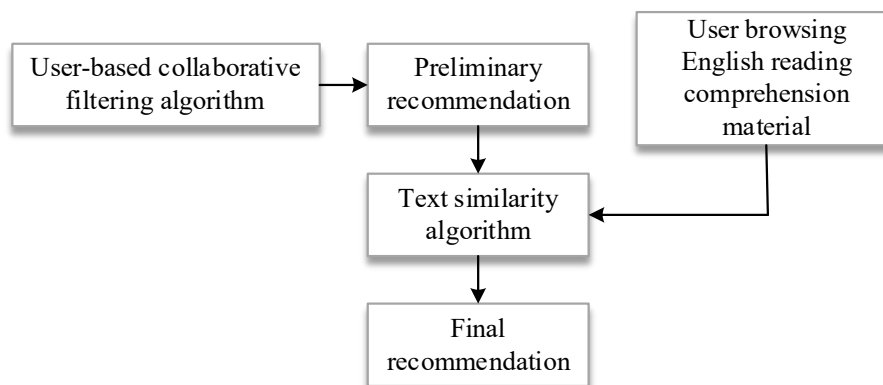


Figure 2. Structure diagram of fusion recommendation system based on text similarity

2.5 System Functional Design

The personalized learning recommendation tool for English reading comprehension materials includes two parts: the front-end interaction subsystem and the back-end management subsystem. The front-end interaction subsystem is mainly used for students to interact with the tool for personalized learning of English reading comprehension materials, while the back-end is mainly used for administrators to manage their own information, students' information, and information about the resources of English reading comprehension materials. The overall functional diagram of the personalized learning recommendation tool for English reading comprehension materials is shown in Figure 3.

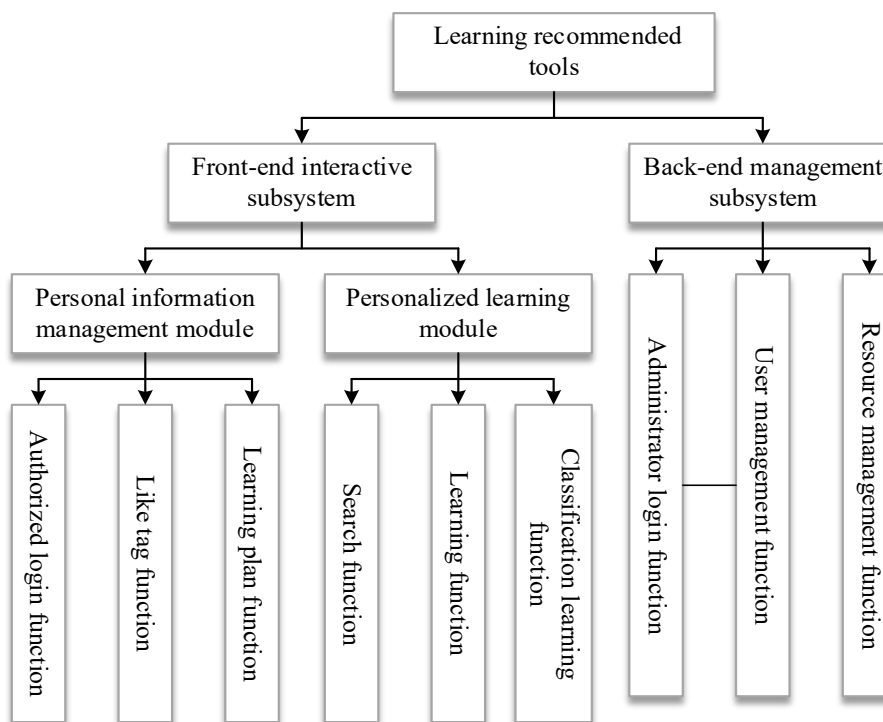


Figure 3. Overall function diagram of learning recommendation tool

2.5.1 Functional design of front-end interaction subsystems

The personal information management module includes the preference label function, learning plan function and authorization login function. In the preference label function and authorization login function, users can modify the preference label and authorize login respectively. In the study plan function, users can view and delete their own study plans.

The Personalized Learning Module for English Reading Comprehension Materials includes three functions: English Reading Comprehension Materials Search, English Reading Comprehension Materials Learning and English Reading Comprehension Materials Classification Learning. In these three functions, users can study the details of English reading comprehension materials, and the tool will recommend corresponding English reading comprehension materials for users based on the recommendation algorithm mentioned above, and users can rate or join the study plan for the English reading comprehension materials they are interested in. In the English reading comprehension material search function, users can enter the keywords of the English reading comprehension material they want to find and select the corresponding type of information to search.

1) Authorized Login Function

Users will use the authorization login function in the initial login to the applet and in the personal information management module. When logging into the app for the first time, the client of the app will prompt the user to choose the account information, i.e., whether to use the WeChat account information for authorization login or to use the visitor account for login, and after the choice is completed, the user will enter the home page of the app. At the same time, the visitor user can also be authorized in the personal information management module to switch to the authorized login identity to use all the functions of the app.

2) Preference Label Function

Users can use this function in the personal information management module to add or delete tags for the types of English reading comprehension material topics they are interested in. When the user uses this function in the personal information management module, the applet will determine whether the user is an authorized user, and if so, it will enter the preference label modification page. If not, the user will be prompted for authorization. When the user enters the preference tag modification page, the applet client accesses the database, reads the user's preference tag information and returns it. When the user makes changes to the preference label, the data will be entered into the database.

3) Learning Program Function

The study plan function is divided into two parts, the first part is the part to view and modify the study plan in the personal information management module, where the user is able to view and delete the English reading comprehension materials that he/she has added to the study plan. The second part is to add a new study plan, which will be called when the user clicks the Add Study Plan button on the English reading comprehension material details page. When the user uses this function in the Personal Information Management module, the applet client will check the user's identity. Only when the user is an authorized user can he/she use the study plan function. That is to say, when a user uses the study plan function, the applet client will access the server and return the study plan information of the user, and the user's modifications to the study plan will be deposited on the server.

4) English reading comprehension material search function

Using the English reading comprehension material search function, users are able to search for the exact English reading comprehension material they need. Specifically, users can enter specific keywords into the search box and then select the type of keyword, i.e. whether the keyword belongs to the “author”, “subject” or “original text”. Then the applet will access the server’s database according to this information, and the database will return the list of English reading comprehension materials queried. This ensures that the English reading comprehension materials retrieved are more comprehensive and meet the needs of learners, reduces redundant English reading comprehension materials that users don’t need, and improves the accuracy of English reading comprehension materials search.

5) Learning Functions of English Reading Comprehension Materials

The English reading comprehension materials learning function is mainly for learners with low motivation to learn English reading comprehension materials. The purpose of designing this function is to stimulate the interest of these learners in learning English reading comprehension materials by providing them with the original English reading comprehension materials that they may be interested in, thus enhancing their motivation for learning English reading comprehension materials and enabling them to stay and continue learning English reading comprehension materials. At the same time, for learners who have a certain interest and foundation in learning English reading comprehension materials, directly providing the original English reading comprehension materials that satisfy their interest in learning can also facilitate the learning of this part of the learners.

6) English Reading Comprehension Material Classification Function

This function is designed for learners who want to concentrate on learning English reading comprehension materials of a certain topic type. In the English reading comprehension materials classification learning function, learners can click into the English reading comprehension materials classification learning interface, and then choose the type of English reading comprehension materials they are interested in to study.

2.5.2 Functional design of the back-end management subsystem

The administrator privileges of the applet are given by the applet developer. After the administrator logs into the background of the applet, the administrator is able to manage the user’s account and English reading comprehension material resources. Therefore, the background functions of this tool include three functions: administrator login, user management and English reading comprehension material resource management.

1) Administrator Login Function

The administrator authority of the applet is given by the applet developer. After the applet developer authorizes the administrator, the administrator will get an account and password, using which, through the administrator login function, the applet administrator can log in the backend of the applet.

2) English reading comprehension material resource management function

In the English reading comprehension material resource management function, the administrator is able to add, delete, check and change the author, original text, translation and other information of the English reading comprehension material within the scope of authority, and can also operate the label of the English reading comprehension material.

3) User Management Function

In the user management function, the administrator is able to perform operations such as adding, deleting, querying and modifying user information according to the permissions given by the developer of the applet.

3 Algorithm recommendation experiment process and analysis

3.1 Recommended accuracy for different numbers of materials

In order to evaluate the accuracy of the fusion algorithm based on text similarity proposed in this paper, UserCF algorithm, CNN algorithm and this paper's algorithm were compared and experimented, and three more representative datasets of English reading comprehension materials were selected, namely, containing reading comprehension questions extracted from Chinese secondary school students' English exams-the RACE dataset, the one generated based on the Wikipedia article, containing multi-quality, diverse English text dataset-SQuAD, and the English version of the CMRC dataset. The results of the comparison experiments are shown in Figures 4, 5, and 6, respectively. The experimental data show that the English reading materials recommended by the fusion algorithm based on text similarity proposed in this paper have a high accuracy rate, and also maintain a high recommendation accuracy rate when the number of recommended English reading materials is 10, and the overall recommendation accuracy rate is between 0.82 and 0.95, while when the number of recommended English reading materials is 10, the UserCF algorithm and RNN algorithm's English reading comprehension material recommendation accuracy rate is 0.4 and 0.3 respectively, the recommendation results are less satisfactory.

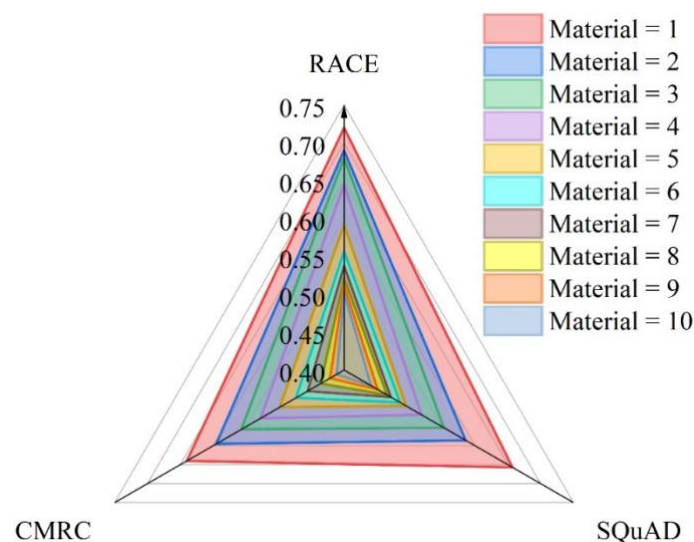


Figure 4. Experimental results of UserCF algorithm

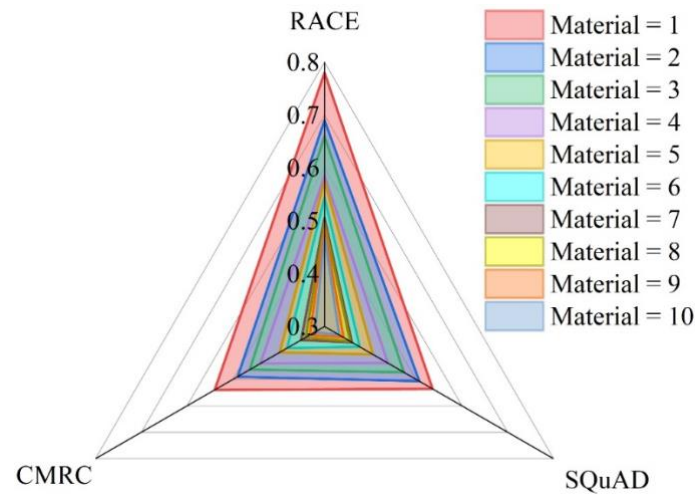


Figure 5. Experimental results of CNN algorithm

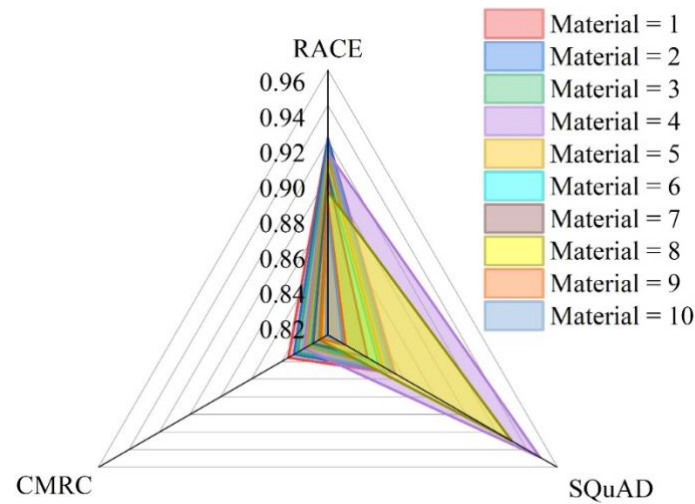


Figure 6. Experimental results of this article algorithm

3.2 Comparison of accuracy of different algorithms

With the increase of the number of iterations, the RMSE value of this paper's algorithm shows a decreasing trend, which indicates that the prediction error of the model is gradually decreasing, and at the number of iterations of 20, the RMSE value of this paper's method decreases to 0.2, and the accuracy of the prediction results is gradually improving. Figure 7 shows the trend of the RMSE value of the algorithm, from which it can be seen that the RMSE value of the UserCF model also shows a decreasing trend with the increase in the number of iterations. However, the RMSE value of this paper's algorithm decreases significantly faster than that of UserCF, which means that this paper's algorithm is able to converge to a lower error level more quickly during the iteration process. Therefore, with the same number of iterations, the prediction accuracy of this paper's algorithm is usually higher than that of UserCF and CNN.

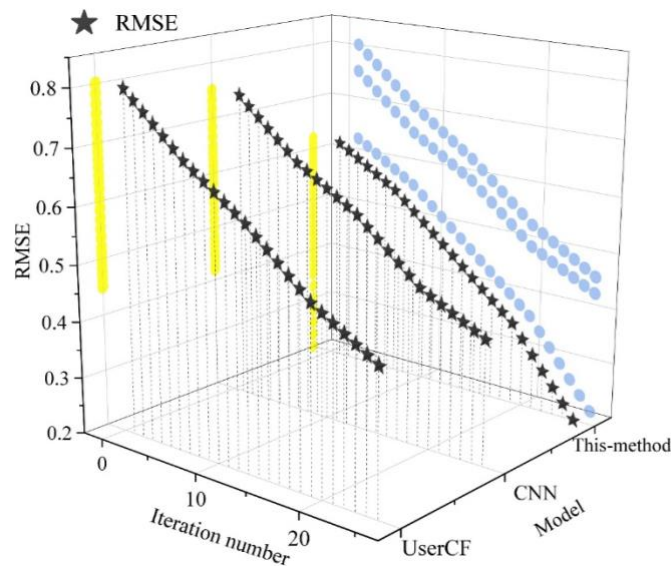


Figure 7. Algorithm accuracy comparison

4 Application of a recommendation system for English reading comprehension materials

4.1 Performance testing

The performance test of the system is mainly used to test whether the system can provide highly reliable and fast response service. In terms of response time, this paper uses the method of calling interfaces and adopts test cases with concurrency of 100, 500 and 1000 to test the response performance of the system. After testing, when the concurrency reaches 1000, the average response time of each functional module of the system can be controlled within 850ms, which is in line with the current stage of development requirements. The concurrent response test results are shown in Table 1.

Table 1. Concurrent response test results

Functional module	Concurrent quantity	Mean response time/ms
Home management	100	305
	500	302
	1000	322
Reading comprehension material learning module	100	215
	500	300
	1000	410
Recommended function	100	510
	500	755
	1000	845
Interactive function	100	285
	500	319
	1000	359

In terms of reliability, this paper conducts security tests from three aspects: user information security, user authority security and system ecological security, and the test results are shown in Table 2. The test results of the test items are all passed, proving the reliability of the system in this paper.

Table 2. Safety test results

Test module	Test item	Test results
User information security	The user password is not available	Pass
	The user can't enter the account by the back key after the user exits the login	Pass
	User interaction history is not visible	Pass
User authority security	Ordinary user is isolated from administrator privileges	Pass
	Ordinary user conflicts with ordinary user privileges	Pass
	The average user conflicts with the administrator user privileges	Pass
Ecosystem security	The administrator deleted the operation to return	Pass
	Automatic backup of system database content	Pass

4.2 Analysis of Student Learning Effectiveness

In this study, “English reading comprehension” is selected as the teaching content, and 100 students in the class of 2020 in a high school are the subjects of the study, including 50 students in the control class and 50 students in the experimental class. As to whether the two groups of students can be compared, a pre-test was administered to the students of the two majors, which included the students' gender, age, and basic English reading comprehension. If the bases of the two classes are comparable it is possible to do a comparative experiment, if the gap between the bases of the two classes is relatively large, it is not possible to do a comparative experiment. By analyzing the pre-test data, the results show that the two classes have comparable foundations and are able to do the comparison experiment.

The text similarity-based English reading comprehension material recommendation system is used for teaching practice by comparing the six posttest scores of the control group and the experimental group obtained from the comparison experiment. The six 6 times comparison experiments instead of 1 time comparison experiments were conducted, on the one hand, because of the cumulative comparison experiments of the effect of the text similarity-based English reading comprehension material recommendation system, and on the other hand, in order to exclude the chance and to make the data obtained more convincing. The article on the 6 times of post-test scores of paired samples T-test to get Table 3, Table 3 for the experimental group and the control group post-test scores T-test, through the table we can find that the post-test 1, the experimental group and the control group's scores are comparable, the post-test 2 to the post-test 6 experimental group more and more obviously higher than the control group's scores. There may be two reasons why the experimental group did not outperform the control group for the first time: first, the students in the experimental group may be unfamiliar with the functions of the operating system, resulting in an insignificant learning effect and the system playing a lesser role, whereas the subsequent students will be proficient in the use of the system's functions. The second is that the experimental group started to learn the recommended English reading comprehension materials only after the posttest 1, and the effect of learning will be reflected in the subsequent scores, and the students in the experimental group at the first posttest had not yet learned the English reading comprehension materials recommended by the system, and were not able to play the role of the text-similarity-based English reading comprehension materials recommendation system. We can observe that the overall

performance of the experimental group is higher than that of the control group, so the data show that the text similarity-based English reading comprehension material recommendation system in this paper is able to recommend personalized English reading comprehension materials, achieve personalized learning for students, and improve students' English learning performance, so the text similarity-based English reading comprehension material recommendation system is able to achieve accurate English reading comprehension material recommendation. Recommendation.

Overall, as the number of times the text similarity-based English reading comprehension material recommendation system is used increases, the experimental group's performance is higher than that of the control group and the degree of higher than the control group is more and more significant, and the Sig. value of the fourth, the fifth and the sixth time is 0.000, which proves that the text similarity-based English reading comprehension material recommendation system is able to give the students personalized learning of English reading comprehension materials, improve the The effectiveness of the text similarity-based English reading comprehension material recommendation system has been verified, which provides a basis for the promotion of the text similarity-based English reading comprehension material recommendation system, and also shows that the text similarity-based English reading comprehension material recommendation system can, to a certain extent, solve the problem of the traditional classroom's difficulty in taking care of every student and satisfying the students' needs. It also shows that the text similarity-based English reading comprehension material recommendation system can, to a certain extent, solve the problem that traditional classrooms are difficult to take care of each student and satisfy students' individual learning needs.

Table 3. Test results of the experimental group and the control group

Postmeasure	Group	N	Mean	Standard deviation	Standard error	Sig.
After test 1	Experimental group	50	75.14	3.865	0.832	0.365
	Control group	50	75.38	3.758	0.875	
After test 2	Experimental group	50	76.51	4.158	0.795	0.026
	Control group	50	76.23	4.025	0.824	
After test 3	Experimental group	50	77.65	3.895	0.821	0.022
	Control group	50	77.54	3.738	0.811	
After test 4	Experimental group	50	76.89	2.899	0.793	0.000
	Control group	50	77.12	3.634	0.785	
After test 5	Experimental group	50	78.94	2.468	0.768	0.000
	Control group	50	79.14	2.577	0.722	
After test 6	Experimental group	50	80.92	2.165	0.744	0.000
	Control group	50	77.85	2.466	0.716	

4.3 Analysis of student satisfaction

In order to collect feedback on learners' satisfaction with the learning process of the English reading comprehension material recommendation system based on text similarity, a system satisfaction questionnaire was designed in this paper. The questionnaire contained ten questions, each of which was rated on a five-point scale, and investigated three aspects, namely, the functional usability of the English reading comprehension material recommendation system, satisfaction with the recommendation results, and trust. The questionnaire was distributed online and answered online, and some of the specialized questions were explained before filling out the questionnaire, in order

to ensure the authenticity and validity of the data as much as possible. A total of 100 questionnaires were distributed online and 100 were recovered, with an effective recovery rate of 100% and an answer rate of 100%. After summarizing the statistics, the average scores of each question in the questionnaire are shown in Table 4.

Table 4. English reading material recommendation system satisfaction score

Index	Topic	Average score (1-5)
Availability	Using this system, I know what the next English reading material to study is.	3.44
	The system is suitable for my recommended English reading material.	4.08
	It is appropriate for the system to help me plan the learning order of English reading materials.	3.54
	The recommended reasons for recommending English reading materials are reasonable.	3.66
Satisfaction	By learning, I master (improve) the knowledge or skills I need.	3.82
	The follow-up English reading materials and resources recommended by the system are appropriate and satisfactory.	3.77
	Using this system, I can easily understand the knowledge architecture of English reading materials.	2.89
	The recommended function of English reading materials optimizes my learning process and improves the learning efficiency.	3.35
Trust	I am willing to study in the order of English reading materials recommended by the system.	4.07
	I am willing to recommend this English reading material system to other learners.	4.16

Questions 1, 2, 3, and 4 in the questionnaire investigated the usability aspects of the recommendation function of the English reading comprehension material recommendation system, and the results of the survey are shown in Figure 8. From the mean scores and the statistical distribution graphs of the data, more than 65% of the learners thought that the sequence of English reading comprehension materials and the difficulty of the resources recommended by the system were appropriate, and 14% of the learners thought that the learning sequence of the English reading comprehension materials planned by the system was not appropriate. Since the selection of English reading comprehension materials from static uniformity to dynamic personalization may bring some learning confusion to learners, Question 1 specifically investigated the impact of this. The results of the survey show that about 57% of the learners think that using the system can make it clearer what they will learn next in terms of English reading comprehension, and about 25% think that the effect is average, but 18% think that they are not clear about what they will learn next, which suggests that the dynamic learning path of English reading comprehension materials may bring about different degrees of learning confusion, or the system lacks the necessary hints and explanatory information, and therefore the system lacks the necessary information. This suggests that the dynamic learning path of English reading comprehension materials may bring about different degrees of learning confusion, or the system lacks the necessary hints and explanatory information, so the system should continue to be optimized in terms of the guiding nature of the learning function and the explanatory nature of the recommendations. Depending on the source of the recommended candidates, different summaries of the reasons for the recommendations are provided. The purpose of displaying the reason for recommendation is to increase learners' trust in the recommendation results. According to the feedback results of Q4, more than 50% of learners recognized the current reason for recommendation.

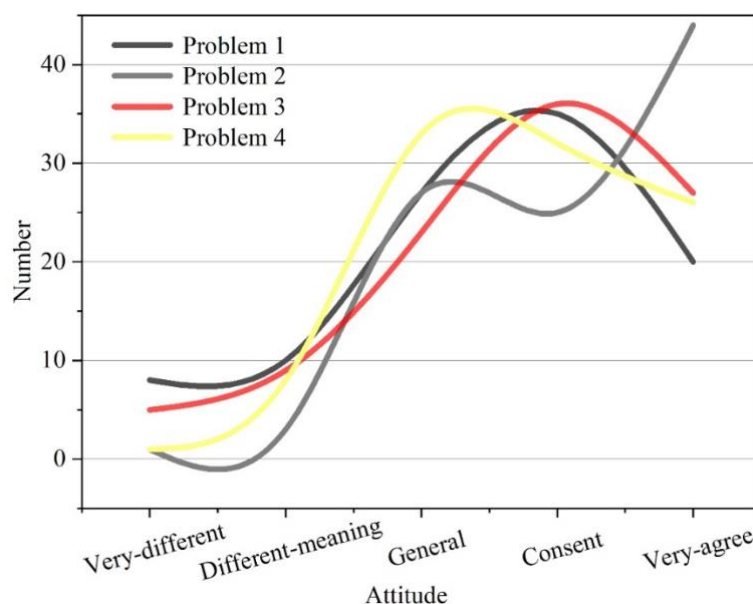


Figure 8. Statistical distribution of usability problems

Questions 5, 6, 7 and 8 in the questionnaire were investigated in terms of satisfaction with the effectiveness of the recommendations and the learning experience, and the results of the survey are shown in Figure 9. After analyzing the mean scores and the statistical distribution graphs of the data, it can be learned that 68% of the learners think that the follow-up English reading comprehension material resources recommended by the system are satisfactory, and 69% of the learners think that they have mastered or improved the required knowledge or skills after personalized learning of English reading comprehension materials based on the system. It is worth noting that question 7 showed that about 35% of learners felt that they were not able to easily understand the structure of the knowledge system in terms of English reading comprehension materials, which should be taken seriously. Preliminary analysis suggests that this phenomenon may be due to: (1) a more fine-grained division of English reading comprehension materials, with more types of relationships between English reading comprehension materials, which makes the structure of the themes of English reading comprehension materials more complex compared to the traditional structure. (2) The learning paths of the English reading comprehension materials are dynamically recommended and therefore changed with the learning situation, causing confusion in understanding the overall structure of the knowledge of the English reading comprehension materials. Question 8 shows that 30% of the learners think that the effect of improving the learning efficiency of English reading comprehension materials is average, that is, compared with other recommendation systems, the recommendation function of this system does not optimize the learning process of English reading comprehension materials, and 50% of the learners think that this system improves their learning efficiency, which shows that although the recommendation algorithm and system function of English reading comprehension materials based on text similarity algorithm proposed in this paper have a certain degree of effectiveness, there is still some room for optimization and improvement. In particular, such learners who give negative feedback should be further tracked, reproduced and analysed in their learning progress logs, and conducted individual interviews if necessary.

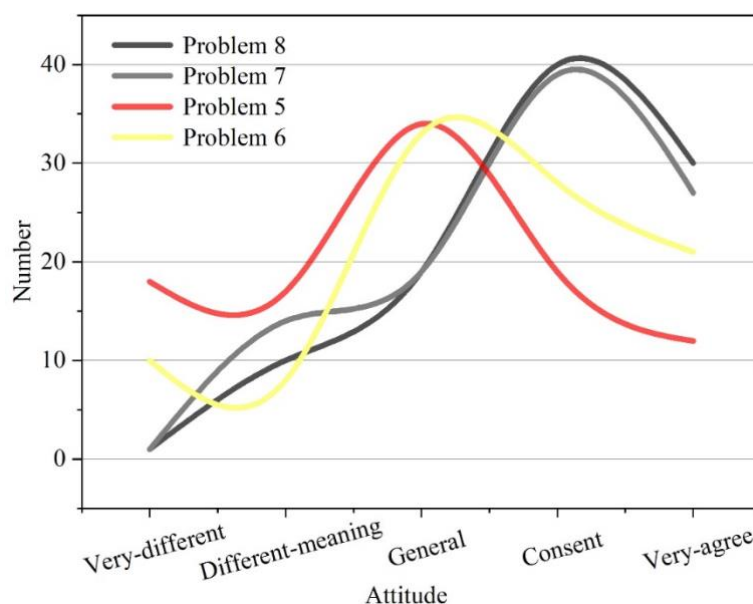


Figure 9. Statistical distribution of satisfaction problems

The last two questions in the questionnaire investigated the aspect of recommendation trust, and the results are shown in Figure 10. More than 78% of learners are willing to follow the order of knowledge points of English reading comprehension materials recommended by the system by default, and more than 80% of learners are willing to recommend this system to other learners. This basically indicates that the recommendation function and other learning functions of this system are trusted by most learners, the user experience is acceptable, and the fusion recommendation system for English reading materials based on text similarity algorithm is basically effective and feasible.

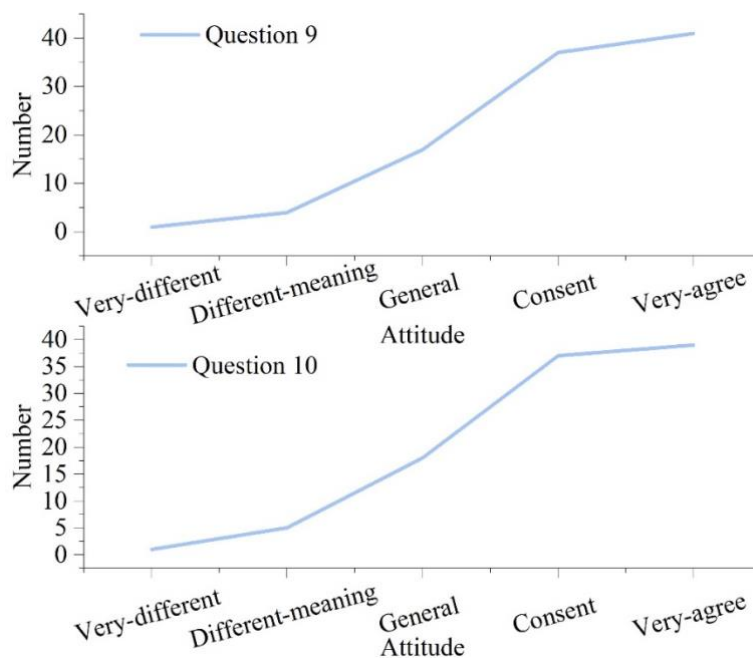


Figure 10. Statistical distribution of trust problems

5 Conclusion

The methods used in this paper mainly include collaborative filtering algorithm and text similarity calculation method, which are improved on the traditional text similarity calculation method, and a fusion English reading comprehension material recommendation system based on improved text similarity is proposed.

In this paper, the fusion collaborative filtering recommendation algorithm and Inter-TF-IDF algorithm are proposed, which effectively overcomes the cold-start problem existing in the traditional algorithm and improves the accuracy of English reading comprehension material recommendation. In the experimental part, the effectiveness of the improved fusion algorithm is verified on RACE, SQuAD and CMRC datasets, and the overall recommendation accuracy is kept at an excellent level of 0.82-0.95.

On the basis of the previous work, this paper takes English reading comprehension material recommendation as an example, and designs the practical application experiments of the English reading comprehension material recommendation system in this paper, and the experiments prove that, through this system, students only need to provide their own learning behaviors and preferences, and the system can help students to quickly find English reading comprehension materials in line with their own learning goals, and significantly improve the students' English scores, and in the fourth post-test experiment The Sig.2 of the control group and the experimental group is 0.000.

The paper concludes with an evaluation of the effectiveness of the proposed English reading comprehension material fusion recommendation system by questionnaire method. The experimental results prove that the system can improve learner satisfaction and has an acceptable good performance on the task of recommending English reading comprehension materials.

Funding

Talent Project of Weinan Normal University "A Cross-Linguistic Comparative Study of Verbs of 'Eating' from a Typological Perspective" (Project No.: 2022RC05).

References

- [1] Wang, Q., & Luo, S. (2019). Shifting from teaching the subject to developing core competencies through the subject: The revised senior middle school English curriculum standards (2017 edition) in China. *Second handbook of English language teaching*, 109-134.
- [2] Long, D. (2019, September). The Path for Cultivation of College Students' Comprehensive Quality from the Perspective of General Education. In *The 4th International Conference on Economy, Judicature, Administration and Humanitarian Projects (JAHP 2019)* (pp. 738-742). Atlantis Press.
- [3] Singh, R. K. V., & Shaari, A. H. (2019). The analysis of Higher-Order Thinking skills in English reading comprehension tests in Malaysia. *Geografia*, 15(1).
- [4] Ismail, N. M., Muliawati, I., & Setiyana, R. (2022). THE IMPACTS OF MODELED-READING STRATEGY ON STUDENTS'REASONING SKILL ABILITY. *JURNAL ILMIAH DIDAKTIKA: Media Ilmiah Pendidikan dan Pengajaran*, 23(1), 41-53.
- [5] Gramley, S., Gramley, V., & Pätzold, K. M. (2020). *A survey of modern English*. Routledge.
- [6] Meng, J. (2017). An Empirical Study on the Application of Cooperative Learning to Comprehensive English Classes in a Chinese Independent College. *English Language Teaching*, 10(2), 94-99.

- [7] Ramadhani, Z., & Refnaldi, R. (2024). Cultural Sense Analysis of Reading Text for the Seventh Grade in the Junior High School English Textbook: "English for Nusantara". *Journal of Languages and Language Teaching*, 12(3), 1431-1442.
- [8] Cahyo, S. D., Umam Muslim, M. R., Rahman, A. N., & Pratolo, B. W. (2019). Needs Analysis of Islamic-Based English Reading Material for the Muhammadiyah Junior High School. *International Journal of Evaluation and Research in Education*, 8(2), 286-292.
- [9] Sudipa, N. (2020). The Effectiveness of Environment-based Reading Material in Increasing Students' Vocabulary in EFL Learning in the Eighth Grade of Sumbawa District Junior High School. *Journal of Language Teaching and Research*, 11(5), 793-800.
- [10] Kung, F. W. (2019). Teaching second language reading comprehension: the effects of classroom materials and reading strategy use. *Innovation in Language Learning and Teaching*, 13(1), 93-104.
- [11] Gu, W., Yuan, Y., & Ye, S. (2024, September). AI Translation Analysis Based on Text Features and Similarity Calculation. In *Proceedings of the 2024 9th International Conference on Cyber Security and Information Engineering* (pp. 624-628).
- [12] Abimanyu, C. G., ER, N. A. S., & Badung, B. I. (2020). Automatic essay answer rating using the text similarity method. *Jurnal Elektronik Ilmu Komputer Udayana p-ISSN*, 2301, 5373.
- [13] Reddy, G. S., & Rajinikanth, T. V. (2017). A text similarity measure for document classification. *IADIS International Journal on Computer Science and Information Systems*, 12(1), 14-25.
- [14] Prasetya, D. D., Wibawa, A. P., & Hirashima, T. (2018). The performance of text similarity algorithms. *International Journal of Advances in Intelligent Informatics*, 4(1), 63-69.
- [15] Takii, K., Flanagan, B., Huiyong, L. I., Yuanyuan, Y. A. N. G., & Ogata, H. (2022, November). Explainable English Material Recommendation Using an Information Retrieval Technique for EFL Learning. In *International Conference on Computers in Education*.
- [16] Bai, Y., & Xian, H. (2024). Automatic Classification and Recommendation for English Teaching Materials Based on Natural Language Processing. *Journal of Circuits, Systems and Computers*.
- [17] Zhang, D. (2024). An Automatic Recommendation Method for English Language Education Resources Based on Particle Swarm Algorithm. *International Journal of High Speed Electronics and Systems*, 2540292.
- [18] Escobar-Acevedo, A., Guerrero-García, J., & Guzmán-Cabrera, R. (2022). A Model Text Recommendation System for Engaging English Language Learners: Facilitating Selections on CEFR. *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 14(3).
- [19] Liu, M. C., Lin, W. Y., & Tsai, C. L. (2022, July). Computer-Aided Response-to-Intervention for Reading Comprehension Based on Recommender System. In *International Conference on Artificial Intelligence in Education* (pp. 512-518). Cham: Springer International Publishing.
- [20] Wang, X., Zhong, Y., Huang, C., & Huang, X. (2024). ChatPRCS: A personalized support system for English reading comprehension based on ChatGPT. *IEEE Transactions on Learning Technologies*.
- [21] Wu, T. T., & Huang, S. H. (2023). An Adaptive and Personalized English Reading Recommendation System. In *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy* (pp. 3015-3033). Cham: Springer International Publishing.
- [22] Cheng, J., & Wang, H. (2021). Adaptive algorithm recommendation and application of learning resources in English fragmented reading. *Complexity*, 2021(1), 5592534.
- [23] Yun Wu, Yiqiao Li & Ren Qian. (2017). NE-UserCF: Collaborative Filtering Recommender System Model based on NMF and E2LSH. *International Journal of Performability Engineering*(5), 610-619.
- [24] Chen Shuangyu, Zhang Bin, Chen Junhao, Zhou Zihan, Shen Yibo & Xiang Weitao. (2024). Reproduction Study of Item-CF Algorithm Based on Penalty Factor Correction. *Academic Journal of Computing & Information Science*(5),

- [25] Yang Zhanhai, Wang XuAn, Qiu Mingyue, Hou Senlin & Wu Yuqiang. (2025). Account of Spatio-Temporal Characteristics in Customs Anti-Smuggling Intelligence Acquisition: A Combined LSTM+CRF Model Using TF-IDF and Levenshtein. *International Journal of Data Warehousing and Mining (IJDWM)*(1), 1-20.
- [26] Aristoteles, Muhammad Umaruddin Syam, Tristiyanto & Bambang Hermanto. (2024). Implementation of Cosine Similarity Algorithm on Omnibus Law Drafting. *International Journal of Advanced Computer Science and Applications (IJACSA)*(4).

About the Author

Ruili Chu (1980.4-), female, native place: Chengcheng, Shaanxi, unit: Weinan Normal University, professional title: associate professor, education: doctoral degree, master's supervisor, main research direction: theoretical linguistics.