Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Automated Matching of Exercises with Knowledge components

Zejie TIAN[a*], Brendan Flanagan[b] , Yiling DAI[b] & Hiroaki Ogata[b]
[a]Graduate School of Informatics, Kyoto University, Japan
[b]Academic Center for Computing and Media studies, Kyoto University, Japan
*tianzejie123@gmail.com

Abstract: As intelligent tutoring systems (ITS) fast develop, they have been implemented in many classrooms combing with various tutorial tools. Especially, matching exercises with knowledge components is the fundamental task of many applications, such as automatic recommendation, student knowledge tracing etc. However, manually labelling educational data is labor intensive and time consuming. Therefore, a range of machine learning methods has been proposed to address this problem, while few of them focusing on Japanese educational dataset in the real high school. In this paper, we leverage natural language processing techniques with Keyphrase extraction methods based on Japanese math exercises. We evaluate the model performance with several state-of-the-art methods and how it works in a real educational task. The results show that our methods outperform several state-of-art methods and can effectively save time for managing math exercise.

Keywords: Keyphrase extraction, nature language processing, knowledge components

1. Introduction

Towards the last decade, with the fast development of AI and educational techniques, Intelligent Tutoring System (ITS) and Learning Management Systems (LMS) are widely used for helping students to promote their learning performance. They focused on finding an "optimal path" towards a set of skills, also known as Knowledge components (KCs). Therefore, properly matching exercises with knowledge components is important task for exercise-based applications such as personalized recommendation (Hou, Y., Zhou, P., Xu, J., et al, 2018) (Hou, Y., Zhou, P., Wang, T., et al, 2016), knowledge tracing (Piech, C., Bassen, J., Huang, J., et al, 2015) (Vie, J. J., & Kashima, H. ,2019) etc. Currently, in many schools, exercises sets are still manually labeled by experts. However, manual labelling costs quantity of time and requires expert labors. As the number of exercises fast grows, it requires automatically understand the meaning of exercises and labelling it with suitable knowledge components. Therefore, many research have been proposed to solve this problem. Usually the solution of this problem can be framed as the multinomial classification (Shen, J. T., Yamashita, M., Prihar, E., et al, 2021) which assigns the most relevant KCs label to the exercises. Prior research solutions included supervised learning methods (Pardos, Z. A., & Dadu, A., 2017) (Hage, H., & Aimeru, E., 2006), unsupervised learning methods (Desmarais, M. C., 2012) and deep learning methods (Shen, J. T., Yamashita, M., Prihar, E., et al, 2021) (Huang, T., & Li, X., 2021). However, existing methods most focus on English exercises and seldom of them based on Japanese exercises. Therefore, there are few of linguistic based analyzing rules can be referenced. Besides, most of them require a large number of datasets. Usually in the real school teaching environment, it is hard to collect as much data as Online learning platform and few of them focusing on reducing expert checking time by developing helpful tools. But it needs to blend the bridge to help expertise saves time for managing ITS.

Toward these challenges, in this paper, we propose a new model combing the word embedding based method with the Keyphrase extraction method. Both models may have some strength, word embedding based methods has great ability for representing and understanding the semantic meaning of text. Keyphrase extraction-based methods can utilize the domain knowledge to category the data which has great interpretability. Especially, for the Japanese, the fine-grained information can be extracted according to the structure of grammar. Combing both of them can be complementary to each method. Word embedding based method can help the model to understand the semantic meaning of exercises from words and with the help of Keyphrase extraction method, it can focus on the "key concepts" of each exercise and improve the model interpretability. With the experience of word embedding based method, it can also improve the model performance of Keyphrase extraction method. We conduct the experiment to test how well it can assist experts for classifying the exercises and comparing the model performance with several state-of-art models. The experiments show that our methods can effectively save the expert oversights time and outperform several state-of-models. By showing the effectiveness of the system, my research paved the way towards a complete automatic knowledge component annotation system.

2. Related work

2.1 Matching exercises

There are wide range of researches of been posted in the last decades. Some researches organized exercises by finding similarity between them which also called as finding similar exercises. Generally, the similar exercises serve the same educational purpose and indicate the same KCs (Del Solato, T., & Du Boulay, B., 1995). Some prior works leveraged texts and concepts in the exercises to compare the similarity. VSM (Tsinakos, A., & Kazanidis, I., 2012) is one of the models by representing the exercises as TF-IDF and calculated the similarity by text distance methods. But they were lack of understanding of semantic meaning of exercises and accuracy was low. Recently, some deep learning-based methods have been proposed to solve this problem. (Liu, Q., Huang, Z., Huang, Z., et al, 2018) extended the resources by utilizing texts, images and KCs for systematically exploit the semantic information of data. (Huang, T., & Li, X., 2021) utilized BERT methods to release the risk of lacking of labeled exercises.

Some researches proposed to match exercises with the KCs of textbook. (Matayoshi, J., & Lechuga, C., 2020) used natural language processing and machine learning methods to match the ITS exercises with certain textbook content and the textbook content belongs to specific topics. By introducing the text book data and explanation of each exercise can effectively help model understand and classify the ITS exercises. In this research, we will utilize the high-performance ability of deep learning model and extension the resource according to the Keyphrase.

2.2 Keyphrase extraction

Keyphrases are the set of words which can effectively highlight the meaning of documents. Extraction of Keyphrases can help us quickly analyzing and organizing the huge amount of data, which can provide concepts or the theme of documents. (Chau, H., Labutov, I., Thaker, K., et al, 2021) proposed to use the machine learning methods by using the POS tagger to annotate the Keyphrases in the educational textbooks. (Alzaidy, R., Caragea, C., & Giles, C. L., 2019) used the deep learning approach to show the outperformance of the Bi-LSTM-CRF model. There are also some related researches (Siddiqi, S., & Sharan, A., 2015) used linguistic based approaches or specific domain knowledge to extract the related Keyphrases in the documents. By knowing or extracting the relationship of Keyphrases can also help us to build up the knowledge graph (Wong, W., Liu, W., & Bennamoun, M., 2012). Construction of educational knowledge graph can also effectively help some downstream task such as automatic recommendation and knowledge tracing etc. (Nakagawa, H., Iwasawa, Y., & Matsuo, Y., 2019) (Tong, S., Liu, Q., Huang, W., et al, 2020) To the best of our knowledge, those methods are still not fully exploit in the educational area and also there are few rule-based methods to reference for the Japanese math data.

3. The proposed model

In this paper, we propose an ensemble method combing both word embedding based method and Keyphrase extraction. We utilize Word2Vec (Church, K. W., 2017) to transform the exercises into the vector space for understanding the semantic meaning of exercises. Moreover, the Keyphrase extraction can help to indicate the key knowledge components of each exercises. The overall structure of our model showed in figure 1.

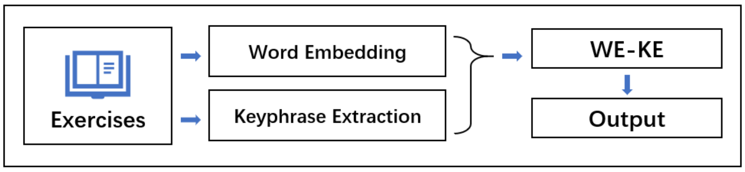

Figure 1. The Overview of Our Proposed Model.

The whole process can be seen as input the exercise text and output the most one of relevant KCs. We use the R_{WE} to represent the result of the word embedding based method and R_{KE} to represent the result of the Keyphrase extraction method. For each single model, the output is the vector of relevant probability for each exercise. We are combining both model results as the input feature for the final model. We will illustrate details in the following sections.

3.1 Word embedding based method

In our system (Flanagan, B., & Ogata, H., 2018), the exercises are in the digital format stored in the separate PDF file. Comparing with HTML file (Matayoshi, J., & Lechuga, C., 2020), it is hard to completely parse and represent the function information within math exercises. Therefore, we only utilize the text information extracting by the Pdf2text (pdftotext, 2021) Python library. Next, we preprocess the text data by utilizing the Nagisa (nagisa, 2020) Python library to segment text into words. During the process, we also filtered the noisy and stop words such as number, function etc. Then, we transformed the words into vector space which length is 300 by utilizing the FastText (Joulin, A., Grave, E., Bojanowski, P., et al, 2016) method. For representing the whole exercises sentences, we calculate the average of word vectors. It can be a simple and effectively way to represent the exercises, because the math questions usually contain few synonyms and have clearly topic. Moreover, we treated the task as the multi-classification problems and trained by machine learning methods: Input the vector representation of exercises and output the probability for each KCs' labels.

3.2 Keyphrase extraction

According to our observation, math concepts can be obviously found in the most of Japanese junior high school math exercises text. For example, Japanese words "因数分解(factorization)" in the exercises showed in figure 2 reveals that it is a question about factorization. Therefore, we collected those key math concepts from text books and expertise suggestions. Then we detecting each exercise about whether it has specific math concepts and save it in the Keyphrase list. As for the example question, the Keyphrase list should be ['2元2次式(binary quadratic)', '因数分解'(factorization)].

37. 2元2次式の因数分解 Factorization of binary quadratic form

次の式を因数分解せよ。Factorization of the following formulas

(1) $x^2-xy-2y^2-x-7y-6$ (2) $3x^2+7xy+2y^2-5x-5y+2$

Figure 2. The Math Exercises about Factorization in the Junior High School Math.

Inspired by the work (Wu, W., Li, H., Wang, H., & Zhu, K. Q,2012), for better extracted the relationship between Keyphrase and flourish the description of exercises, we extracted the relationship of KCs by Japanese linguistic rules. Japanese words 'の' often plays the role of possessor and modifier, usually it can deliver the detail information of KCs such as "整数の加法(Addition of integer)" which shows that the exercise is about addition of integer. 'を' usually uses to mark the object of word, which often shows the purpose of exercises such as “式を計算” (function calculation). ‘と’ is like ‘and’ in the English used to connect similar word together. Conclusion of linguistic pattern and example shows in Table 1.

Table 1. The Example and Pattern of Linguistic Rules

Pattern	Example
_の_	多項式の次数 (Degree of polynomial)
_と_の_	乗法と除法の混じった計算 (Calculation with a mixture of product and division)
_の_と_	多項式の加法と減法 (Adding and subtracting polynomials)
_を_	同類項をまとめ (Summarize similar terms)

In addition, for better construction of relationship, we filtered the results by setting up the stop words and assigned nearest concepts to the pronoun. For distinguishing the key concepts and modifier, we assign the key concepts according to the key math concepts list extracting from the textbooks. As for the example of figure 2, we store the results as a tuple list: [(2元2次式(binary quadratic), 因数分解(factorization)), (式(formula), 因数分解(factorization))].It should be noticed that the order in the tuple didn’t influence the result.

We assume that *if exercises belong to the same KCs then it must have similar key concepts and key concepts relationship.* According to this assumption, we calculate the similarity score between exercise as follows:

$$sim(E^1, E^2) = \frac{E^1_{keywords} \cap E^2_{keywords}}{E^1_{keywords} \cup E^2_{keywords}} + \frac{E^1_{KR} \cap E^2_{KR}}{E^1_{KR} \cup E^2_{KR}}$$

Keywords is the key concepts set and KR is the key concepts relationship set. We calculate the average of similarity scores between exercises in the test set and a batch of exercises belong to certain knowledge components in the training se and store the final results into a matrix.

3.3 Ensemble model

The ensemble model was showed in the figure 1. Results from word embedding based method can help model better understand the semantic meaning of exercises from text and results from the Keyphrase extraction can help better focus on the key topic of exercises. The similar exercises should contain similar semantics meaning and key concepts. For better unite the two models results, we simply average weighted the both model results (Zhou, Z. H., 2012):

$$O = \mathrm{ML}((R_{WE} + R_{KE})/2, \lambda)$$

O is final output of model predictions which can be represented as $O(o_1, o_2, \dots, o_n)$ where n equals to the number of KCs. For the model input, R_{WE} is the result of word embedding based method, R_{KE} is the result of Keyphrase extraction method and λ is the hyperparameter for normalization to avoid overfitting. ML is the machine learning model in our study is based on XGBoost model. We show the detail analysis in the experiment session.

4. Experiment

4.1 Datasets and experiment settings

To better evaluate our model, we conduct our experiment on the Japanese math exercises collected in the real Junior high school. The datasets contained 830 exercises included four main topics: geometry, function, statistic and probability. Currently, our task is to classify the exercises into 13 detailed KCs. Follow the prior works, we measure the model performance with the macro F1-score and Accuracy which calculated as (TP+TN)/ (TP + TN + FP + FN). We utilize the XGBoost model for the word embedding based methods and final state. To measure the generalizability of our proposed model, we apply 5-fold cross-validation at the KCs level, which means we split exercises of each KCs into 5 distinct groups and both training and testing datasets are selected from those.

We compared our model with the following prior works. They all use the same input data.

- Vector space model (VSM) (Tsinakos, A., & Kazanidis, I., 2012): The method transformed exercises into vector space and compared the cosine similarity between them. We chose it as the baseline model for comparison.
- Support vector machine (SVM) (Karlovčec, M., Córdova-Sánchez, M., & Pardos, Z. A., 2012): SVM model has a great ability to measure high dimensional input features and suitable for the classification of exercises with dense concepts.
- XGBoost (Chunamari, A., Yashas, M., Basu, A., et al, 2022): XGboost is the scalable machine learning models which is built up on the tree boosting. It shows great performance in many kinds of classification problems.
- Neural network (NN) (Patikorn, T., Deisadze, D., Grande, L., et al, 2019): Neural network is one the classic machine learning model and shows promise performance in the classification of exercises job.
- Long short-term memory network (LSTM) (Tong, W., Tong, S., Huang, W., et al, 2020): LSTM is one of the deep learning methods and can help to memorize the relationship of each input features.

The experiment results showed in the table 2. According to the results, our model F1-score of 0.7897 and Accuracy of 0.7957 for the matching of ITS exercises with knowledge components. For the others, the accuracy of VSM is 0.5579, SVM is 0.7073, XGBoost is 0.7287, NN is 0.7287, LSTM is 0. 7409. The F1-score of VSM is 0.5559, SVM is 0.6864, XGBoost is 0.7137, NN is 0.7127, LSTM is 0.7382. This performance significantly outperformed then prior state-of-art works.

Table 2. Experiment of Our Proposed Model

Evaluation	Model					
	VSM	XGBoost	SVM	NN	LSTM	WE-KE
Accuracy	0.5579	0.7287	0.7073	0.7287	0.7409	0.7957
Macro F1-score	0.5559	0.7137	0.6964	0.7127	0.7382	0.7897

In summary, our model outperforms then other model across junior high school's Japanese math exercises datasets. The results show that effectively combing both semantic meaning extracting from word embedding based methods and Key concepts comparison can help model to improve the performance. In the next session, we will do some ablation study to see the performance of each single methods.

4.2 *Ablation study*

In order to better analysis the contribution of each part of our model. We conduct the ablation study to evaluate, the results show in table 3.

According to the results, both methods have promising performance. It can show that word embedding based methods can extract the meaningful semantic information and Keyphrase extraction can effectively organize the exercises by keywords. It should be noticed that the experiment also shows that Keyphrase can effectively show the main topic of exercises, so that the rule-based model performance can be comparable with machine learning method. Moreover, the ensemble model can better analyze both model results and outperform then each single model.

Table 3. *Ablation Study*

Evaluation	Model		
	WE	KE	WE-KE
Macro F1-score	0.7137	0.7350	**0.7897**
Accuracy	0.7287	0.7195	**0.7957**

Besides, we also conduct the experiments of different word vectorization methods by using TF-IDF (Aizawa, A., 2003). This method related on the term frequency and inverted document frequency. The results of experiment showed in table 4. According to the results, the accuracy of VSM is 0.7165, SVM is 0. 8140, XGBoost is 0. 8140, NN is 0. 8171, LSTM is 0.8065. The F1-score of VSM is 0.7129, SVM is 0.8095, XGBoost is 0.8100, NN is 0.8156, LSTM is 0.8025. Our model the F1-score is 0.8364 and accuracy is 0.8384. Our model performance with the TF-IDF vectorization also outperforms others and higher than using Word2Vec. Due to the feature's length of TF-IDF is 866 which almost same as the size of experiment data. Therefore, the model is easily overfitting. We keep using Word2Vec in our main experiment. In conclusion, the ensemble of two methods can provide stable results and outstanding performance in the matching exercises with KCs task.

Table 4. *Experiment of Model with TF-IDF*

Evaluation	Model					
	VSM	XGBoost	SVM	NN	LSTM	WE-KE
Macro F1-score	0.7129	0.8100	0.8095	0.8156	0.8025	**0.8364**
Accuracy	0.7165	0.8140	0.8140	0.8171	0.8065	**0.8384**

4.3 Human test

In this session, we will evaluate our model with expert to check whether our model can effectively help them to save time for checking. Because this work is the fundamental of other ITS implementation such as automatically recommendation and knowledge tracing etc. Therefore, it is important to guarantee the correctness of matching exercises with KCs. It is inevitable to invite the expert into this process. For better filter the available knowledge components, we showed the top 3 results in each selection box according to the model prediction as showing in figure 4. Table 5 showed the comparison of Accuracy@K.

Table 5. *The Accuracy of top k*

Evaluation	Model					
	VSM	XGBoost	SVM	NN	LSTM	WE-KE
Accuracy@2	0.6829	0.8445	0.8445	0.8994	0.8598	**0.8994**
Accuracy@3	0.7378	0.9146	0.9132	0.9421	0.9177	**0.9543**

According to the results, our model Accuracy@2 is 0.8994 and Accuracy@3 is 0.9543 both of them are out perform then others. The result also indicates that the true results are most probably included in the top-ranking results. It can effetely save user time for searching.

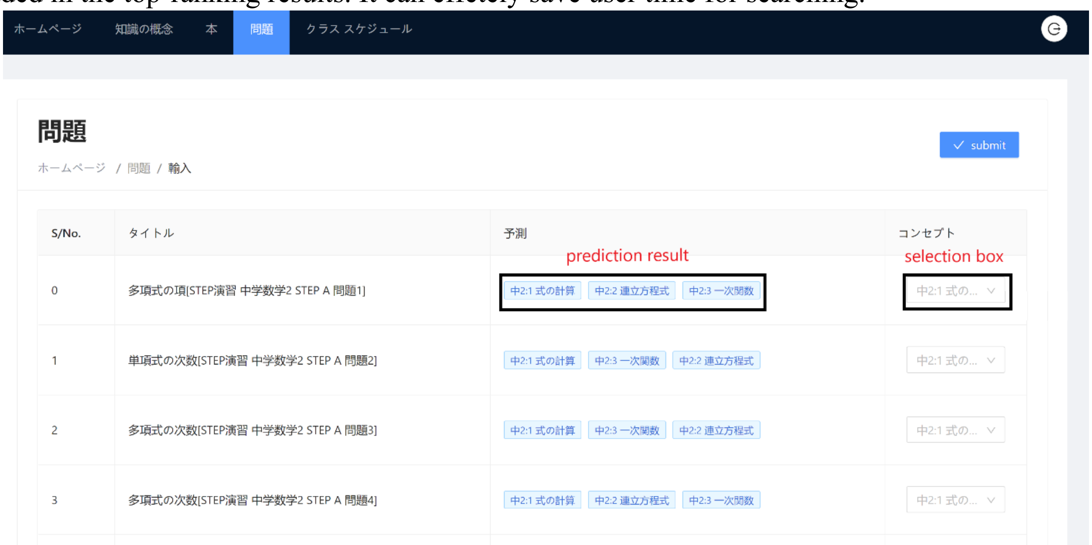

Figure 4. Shortcut of Knowledge Management System

In the figure 4, we develop the simple knowledge portal interface to allow user upload the data and receive the model prediction results. In the "prediction" column, it lists the top three related knowledge components according to the model result. It can help users quickly to find the available choice. For each selection box, the knowledge components are ranked according to the prediction probability and the highest one is default selected by the system. If without the model predictions, all the knowledge components rank by default orders. We conduct preliminary experiment based on it. We invited 6 participants who are the Ph.D. or master's degree students whose researching area is about the education and with good math ability. Then, randomly assign them to different experiment groups and they will not know which group they are in. One group provides the model prediction results and the other require them to assign the KCs to exercises by hands. We split the Junior high school datasets as 6:4 for training datasets and testing datasets than randomly extracted 50 exercises from the testing dataset for the human test.

The results showed in table 6. According to the result, with the help of our prediction results, expert took average 6 minutes less than no prediction function assist. We also calculate the recall as whether the true label in the filtered option list and it has around 0.98. It suggested that almost all the true labels were included in the filter option list and no need for searching. Furthermore, the precision@1 means how many true labels were ranked in the first option. The result was 0.84 and most of true labels were selected as default results. In conclusion, our matching exercises with KCs portals can effectively help experts saving their time for checking the classification results.

Table 6. *The Evaluation Result of Human Test*

Evaluation	Model		
	Time(m)	Recall	Precision@1
Prediction	**6**	0.98	0.84
No-prediction	8	-	-

5. Discussion

Although currently work shows promising performance when matching the exercises with the knowledge components, there still has many challenging and problems needed to be solved. First, parse the pdf file may lose more information then HTML format file. For example, the function information can be completed extracted from HTML file, but hard for pdf file. In many math exercises, the function information can also helpful for extracting the KCs information (Patikorn, T., Deisadze, D., Grande, L., et al, 2019). Second, the geometry is one of the important part for math course and the graph of those exercises are also delivered the important information (Liu, Q., Huang, Z., Huang, Z., et al, 2018). Therefore, heterogenous input features should also be concerned. Moreover, The datasets in our experiments are not plenty enough for comprehensively showing our models performance and the performance of real application. We will try to include more data into our experiments. Besides, our KCs labeled are not enough in the real educational tasks . We should also need to consider the overfitting problem described in the (Patikorn, T., Deisadze, D., Grande, L., et al, 2019). Since, there are a large number of near-identical problems. The model may only remember some specific words, but not the true features of exercises. In collusion, in the future work, we will try to expand our datasets with more KCs and datatype. We will also consider how to utilize the heterogenous data information. For the human test, it also needs to invite more expertise and teachers for the experiment test. We will provide more experiment result and detailed data in the future research report.

6. Conclusion

In this paper, we proposed to use an ensemble method for automatic labeling exercises with specific knowledge component. Comparing with the past researches, our method can maintain the high-performance on the small data size. We try to combine the advantages of word-embedding based model which has great ability for understanding the semantic meaning of texts and Keyphrase extraction model which built up on the domain knowledge and can be easily interpretable. According to the experiment result, the ensemble model outperforms the baselines such as: XGBoost, SVM, LSTM etc. Moreover, it also better than each single model performance. For better prove the practical effectiveness, we develop the knowledge portal system which embedded with the proposed method and provided simple interface to help experts check and update the model results and relevant knowledge components information. Then, conduct the preliminary study of the human test. The results show that with the help of model, it can save the experts checking time.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) 20H01722, JSPS Grant-in-Aid for Scientific Research (Exploratory) 21K19824, NEDO JPNP20006 and JPNP18013.

References

Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. *Information Processing & Management, 39*(1), 45-65.

Alzaidy, R., Caragea, C., & Giles, C. L. (2019, May). Bi-LSTM-CRF sequence labeling for keyphrase extraction from scholarly documents. In *The world wide web conference* (pp. 2551-2557).

Chau, H., Labutov, I., Thaker, K., He, D., & Brusilovsky, P. (2021). Automatic concept extraction for domain and student modeling in adaptive textbooks. *International Journal of Artificial Intelligence in Education, 31*(4), 820-846.

Chunamari, A., Yashas, M., Basu, A., Anirudh, D. K., & Soumya, C. S. (2022). Quora Question Pairs Using XG Boost. In *Emerging Research in Computing, Information, Communication and Applications* (pp. 715-721). Springer, Singapore.

Church, K. W. (2017). Word2Vec. *Natural Language Engineering, 23*(1), 155-162.

Conde, A., Larrañaga, M., Arruarte, A., Elorriaga, J. A., & Roth, D. (2016). litewi: A combined term extraction and entity linking method for eliciting educational ontologies from textbooks. *Journal of the Association for Information Science and Technology, 67*(2), 380-399.

Del Solato, T., & Du Boulay, B. (1995). Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education, 6*, 337-378.

Desmarais, M. C. (2012). Mapping question items to skills with non-negative matrix factorization. *ACM SIGKDD Explorations Newsletter, 13*(2), 30-36.

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal, 10*(4), 469-484.

Hage, H., & Aimeru, E. (2006, March). ICE: A system for identification of conflicts in exams. In *IEEE International Conference on Computer Systems and Applications, 2006.* (pp. 980-987). IEEE.

Hou, Y., Zhou, P., Wang, T., Yu, L., Hu, Y., & Wu, D. (2016). Context-aware online learning for course recommendation of MOOC big data. *arXiv preprint arXiv:1610.03147.*

Hou, Y., Zhou, P., Xu, J., & Wu, D. O. (2018, April). Course recommendation of MOOC with big data support: A contextual online learning approach. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 106-111). IEEE.

Huang, T., & Li, X. (2021). An Empirical Study of Finding Similar Exercises. *arXiv preprint arXiv:2111.08322*.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Karlovčec, M., Córdova-Sánchez, M., & Pardos, Z. A. (2012, June). Knowledge component suggestion for untagged content in an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems* (pp. 195-200). Springer, Berlin, Heidelberg.

Liu, Q., Huang, Z., Huang, Z., Liu, C., Chen, E., Su, Y., & Hu, G. (2018, July). Finding similar exercises in online education systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1821-1830).

Matayoshi, J., & Lechuga, C. (2020). Automated Matching of ITS Problems with Textbook Content. In *iTextbooks@ AIED* (pp. 17-28).

Nagisa (0.2.7). (2020, July 7). Retrieved April 25, 2022, from: https://github.com/taishi-i/nagisa

Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019, October). Graph-based knowledge tracing: modeling student proficiency using graph neural network. In *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 156-163). IEEE.

Pardos, Z. A., & Dadu, A. (2017, July). Imputing KCs with representations of problem content and context. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 148-155).

Patikorn, T., Deisadze, D., Grande, L., Yu, Z., & Heffernan, N. (2019, June). Generalizability of methods for imputing mathematical skills needed to solve problems from texts. In *International Conference on Artificial Intelligence in Education* (pp. 396-405). Springer, Cham.

Pdftotext (2.2.2).(2021, November 24). Retrieved April 25, 2022, from: https://github.com/jalan/pdftotext

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems, 28*.

Shen, J. T., Yamashita, M., Prihar, E., Heffernan, N., Wu, X., McGrew, S., & Lee, D. (2021, June). Classifying Math Knowledge Components via Task-Adaptive Pre-Trained BERT. In *International Conference on Artificial Intelligence in Education* (pp. 408-419). Springer, Cham.

Siddiqi, S., & Sharan, A. (2015). Keyword and keyphrase extraction techniques: a literature review. *International Journal of Computer Applications, 109*(2).

Tiana, Z., Zhengc, G., Flanaganb, B., Mic, J., & Ogatab, H. BEKT: Deep Knowledge Tracing with Bidirectional Encoder Representations from Transformers.

Tong, S., Liu, Q., Huang, W., Huang, Z., Chen, E., Liu, C., ... & Wang, S. (2020, November). Structure-based Knowledge Tracing: An Influence Propagation View. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 541-550). IEEE.

Tong, W., Tong, S., Huang, W., He, L., Ma, J., Liu, Q., & Chen, E. (2020, November). Exploiting Knowledge Hierarchy for Finding Similar Exercises in Online Education Systems. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 1298-1303). IEEE.

Tsinakos, A., & Kazanidis, I. (2012). Identification of conflicting questions in the PARES system. *International Review Of Research In Open And Distributed Learning, 13*(3), 297-313.

Vie, J. J., & Kashima, H. (2019, July). Knowledge tracing machines: Factorization machines for knowledge tracing. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 750-757).

Wong, W., Liu, W., & Bennamoun, M. (2012). Ontology learning from text: A look back and into the future. *ACM Computing Surveys (CSUR), 44*(4), 1-36.

Wu, W., Li, H., Wang, H., & Zhu, K. Q. (2012, May). Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 481-492).

Zhou, Z. H. (2012). Ensemble methods: foundations and algorithms. CRC press.