Prerequisite or Not Prerequisite? That's the Problem! An NLP-based Approach for Concept Prerequisites Learning

Chiara Alzetta[•◇], Alessio Miaschi[⋆◇], Giovanni Adorni[•], Felice Dell'Orletta[◇], Frosina Koceva[•], Samuele Passalacqua [•], Ilaria Torre[•]
[•]DIBRIS, Università degli Studi di Genova, [⋆]Dipartimento di Informatica, Università di Pisa,
[◇]Istituto di Linguistica Computazionale "Antonio Zampolli", Pisa - ItaliaNLP Lab
{chiara.alzetta,frosina.koceva}@edu.unige.it, alessio.miaschi@phd.unipi.it,
samuele.passalacqua@dibris.unige.it, {ilaria.torre,adorni}@unige.it,
felice.dellorletta@ilc.cnr.it

Abstract

English. This paper presents a method for prerequisite learning classification between educational concepts. The proposed system was developed by adapting a classification algorithm designed for sequencing Learning Objects to the task of ordering concepts from a computer science textbook. In order to apply the system to the new task, for each concept we automatically created a learning unit from the textbook using two criteria based on concept occurrences and burst intervals. Results are promising and suggest that further improvements could highly benefit the results.[1]

Italiano. *Il presente articolo descrive una stategia per l'identificazione di prerequisiti fra concetti didattici. Il sistema proposto è stato realizzato adattando un algoritmo per ordinamento di Learning Objects al compito di ordinamento di concetti estratti da un libro di testo di informatica. Per adeguare il sistema al nuovo scenario, per ogni concetto stata automaticamente creata una unità di apprendimento a partire dal libro di testo selezionando i contenuti sulla base di due differenti criteri: basandosi sull'occorrenza del concetto e sugli intervalli di burst. I risultati sono promettenti e lasciano intuire la possibilità di ulteriori miglioramenti.*

1 Introduction

Personalised learning paths creation is an active research topic in the field of education (Chen, 2009; Kurilovas et al., 2015; Almasri et al., 2019). The most fundamental issue behind this task is the need to understand how educational concepts are pedagogically related to each other: what information one has to study/know first in order to understand a given topic. In this paper we focus on such relations, i.e. *prerequisite relations*, between educational concepts of a textbook in English and we present a method for their automatic identification. Here, we define *concepts* all the relevant topics extracted from the textbook and we represent them as single or multi word terms.

Automatic prerequisite extraction is a task deeply rooted in the field of education, whose results can be easily integrated in many different contexts, such as curriculum planning (Agrawal et al., 2016), course sequencing (Vuong et al., 2011), reading list generation (Gordon et al., 2017), automatic assessment (Wang and Liu, 2016), domain ontology construction (Zouaq et al., 2007; Larranaga et al., 2014) and automatic educational content creation (Lu et al., 2019). Several methods have been devised to extract prerequisite relations (Liang et al., 2015; Pan et al., 2017a; Liang et al., 2018b), however they were mainly focused on educational materials already enriched with some sort of explicit relations, such as Wikipedia pages, course materials or learning objects (LOs). More challenging is identifying prerequisites when no such relations are given and textual content is the only available resource.

In 2019, we proposed two methods to identify prerequisite relations between concepts without using external knowledge or even pre–defined relations. The former method (Adorni et al., 2019) is based on burst analysis and temporal reasoning on concepts occurrence, while the latter (Miaschi et al., 2019) uses deep learning for learning object ordering. Both these methods extract prerequisite relations form textual educational materials without using any form of structured information.

In this work, we adapt the system for learning object ordering described in Miaschi et al. (2019) to the task of sequencing concepts in a textbook according to their prerequisite relations. For training and testing our system we relied on a new version of PRET (Alzetta et al., 2018), a gold dataset manually annotated with prerequisite relations between educational concepts. Moreover, since the classifier was designed to acquire learning objects as input, we automatically created a learning unit[2] for each concept according to two different criteria: (i) considering all sentences showing an occurrence of the concept, (ii) considering burst intervals (Kleinberg, 2003) of each concept extracted according to the strategy of Adorni et al. (2019).

The remainder of the paper is organised as follows. First, we present related work (Section 2) and the dataset used for the experiments (Section 3). Section 4.1 presents the classifier, while Burst analysis is described in Section 4.2 and the experimental settings in Section 4.3. Results and discussion are reported in Section 4.4, while error analysis is illustrated in Section 5. Section 6 concludes the paper.

Our Contribution. In this paper: (i) we use a deep learning-based approach for prerequisite relation extraction between educational concepts of a textbook; (ii) we test the impact of creating learning units for each concept according to different criteria and without relying on any explicit structured information, such as Wikipedia hyperlinks; (iii) we show the effectiveness of our approach on real educational materials.

2 Related Work

Datasets annotated with prerequisite relations are built mainly considering two types of data: course materials, acquired from MOOCs (Chaplot et al., 2016; Pan et al., 2017a; Pan et al., 2017b; Gasparetti et al., 2018; Roy et al., 2018) or university websites (Liang et al., 2017; Li et al., 2019), and educational materials in a broader sense, such as scientific databases (Gordon et al., 2017), learning objects (Talukdar and Cohen, 2012; Gasparetti et al., 2018) and textbooks (Wang et al., 2016). The most common approach for prerequisite annotation is to ask experts to evaluate all possible pairs generated from the combination of selected concepts (Chaplot et al., 2016; Wang et al., 2016; Li et al., 2019) or a random sample of that set (Pan et al., 2017b; Gordon et al., 2017; Gasparetti et al., 2018). The dataset presented by Wang et al. (2016) is the one we consider most closely related to ours, since it shows prerequisite relations between relevant concepts extracted from a textbook. However, in their dataset a matching with a Wikipedia page was a strict requirement for concept selection. Contrary to previous works, we asked experts to build the concept pairs if a prerequisite relation was observed while reading a textbook, regardless the existence of a corresponding Wikipedia page for the concepts. Hence we allowed for more subjectivity, without restricting experts' evaluation to a predefined list of items.

For what concerns prerequisite learning approaches, initial work in this field relied on graph analysis (Vassileva, 1997; Brusilovsky and Vassileva, 2002) or, more recently, on link-based metrics inferred from the Wikipedia graph of hyperlinks between pages (Liang et al., 2015). Talukdar and Cohen (2012) made the first attempt to apply machine learning techniques to prerequisite prediction: hyperlinks, hierarchical category structure and edits of Wikipedia pages are the features of a MaxEnt classifier. Similarly, Gasparetti et al. (2018) use Wikipedia hierarchical category structure and hyperlinks. Similarly to our approach, (Liang et al., 2018a; Liang et al., 2018b) integrated text–based features for prerequisite learning, but reported graph–based features as more informative.

Contrary to the above methods, we assign a higher informative value to the textual content referring to a concept and we use this only to extract the features for the classifier. Moreover, we combine the classifier with the burst algorithm (Kleinberg, 2003), which selects the most relevant textual content related to a concept from the textual material. This choice makes our method suitable for prerequisite learning on educational contents also when structured graph information is not available.

3 Dataset

For our experiments we relied on a novel version of PRET dataset (Alzetta et al., 2018), PRET 2.0, a dataset manually annotated with prerequisite relation between educational concepts extracted from

[2]Learning unit is meant here as learning content, with no reference to units of learning in curricula and tables of content.

a chapter of a computer science textbook written in English (Brookshear and Brylow, 2015).

In this novel version, five experts were asked to re–annotate the same text indicating any prerequisite concept of each relevant term appearing in the text. The set of relevant terms was extracted with the same automatic strategy described in Alzetta et al. (2018), but this time the list was manually validated by three experts in order to identify a commonly agreed set of concepts, which resulted in a terminology of 132 concepts. Besides these terms, each expert could independently add new concepts to the terminology when annotating the text if he/she regards them as relevant. Consequently, experts produced different sets of concept pairs annotated with prerequisite relations since 221 new concepts were manually added during the annotation process.

The final gold dataset results from the combination of all annotations, thus considering as positive pairs (i.e. showing a prerequisite relation) all pairs of concepts annotated by at least one expert. The manual annotation resulted in 25 pairs annotated by all five experts, 46 annotated by four experts, 83 by three, 214 by two and 698 by only one annotator, for a total of 1,066 pairs.

2,349 transitive pairs were also automatically generated and added to the dataset: if a prerequisite relation exists between concepts A and B and between concepts B and C, we add a positive relation between A and C to increase the coherence of annotation. In order to obtain a balanced dataset for training our deep learning system, negative pairs were automatically created by randomly pairing concepts and adding them as negative examples if they were missing in the dataset. Overall, the final dataset consists of 353 concepts and 6,768 relations.

4 Method and Experiments

In this Section we present our approach for learning prerequisites between educational concepts.We trained and tested the same deep learning model on three datasets generated from PRET 2.0 that vary with respect to the criterion used for retrieving textual content of each concept in the dataset. As a result, we were able to study performance variations of the classifier given different input data.

Task. We tackle the problem of concept prerequisite learning as a task of automatic binary classification of concept pairs: given a pair of concepts (A,B), we predict whether or not concept B is a prerequisite of concept A.

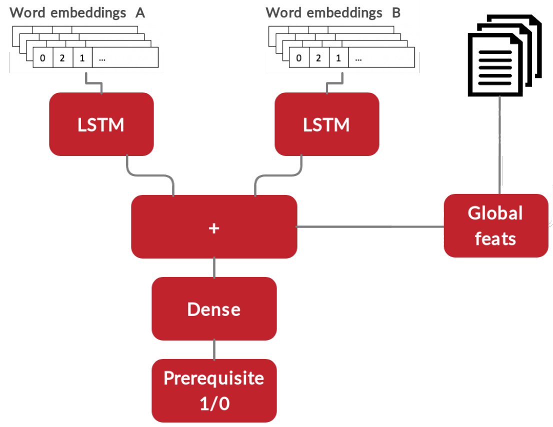

Figure 1: Method workflow.

4.1 Classifier

The system used to predict whether or not two concepts show a prerequisite relation is the deep learning architecture described in Miaschi et al. (2019). Specifically, we relied on the model which uses pre-trained word embeddings (WE) and global features automatically extracted from the dataset.

The system architecture (see Figure 1) is composed of two LSTM-based sub-networks with 64 units, whose outputs are concatenated and joined with a set of global features. The input of the two LSTM-based sub-networks corresponds to the pre-trained WE of concept *A* and *B* respectively. The output layer consists of a single Dense unit with sigmoid activation function. The pre-trained WE were computed using an English lexicon of 128 dimensions built using the ukWac corpus (Baroni et al., 2009). Global features were devised to extract linguistic information from learning units of both concepts in a pair, such as mentions to the other concept of the pair or the Jaccard similarity between textual contents of the two learning units.

For the complete list of global features, refer to Miaschi et al. (2019).

4.2 Burst Analysis

Burst analysis is based on the assumption that a phenomenon might become particularly relevant in a certain period along a time series, most likely because its occurrence rises above a certain threshold. Such periods of increased activity of the phenomenon are called "burst intervals" and can be modelled by means of a two state automaton in which the phenomenon is in the first state if it has a low occurrence, but then it moves to the second state if its occurrence rises above a certain threshold, and eventually it goes back to the first state if its occurrence goes below the threshold (Kleinberg, 2003).

Given its nature, this kind of analysis is highly employed for detecting events from data streams (Fung et al., 2005; Takahashi et al., 2012; Kleinberg, 2016). When applied to textual data – e.g., for text clustering (He et al., 2007), summarization (Subasic and Berendt, 2010) or relation extraction (Yoon et al., 2014; Lee et al., 2015) – the linear progression of the text acts as the time series, hence burst intervals correspond to sequences of sentences where a given term is particularly relevant. In Adorni et al. (2019) burst analysis was used to detect the bursting intervals of concepts along a textbook chapter: for each term, the burst algorithm identified a unique or multiple burst intervals of various length (i.e. a different number of sentences involved in each interval). Temporal reasoning (Allen, 1983) was then employed to find prerequisite relations between concepts.

In this work we use the burst intervals retrieved as described in Adorni et al. (2019) to select relevant content of the textbook for each concept. Our intuition is that burst intervals should capture the most informative portions of text for each concept from the entire textbook content. Note that for this experiment we only used the bursts detected with the first phase of the algorithm described in (Adorni et al., 2019), i.e. the temporal reasoning is not employed here.

4.3 Experimental Settings

Since our deep learning model was designed to find prerequisite relations between learning objects, we had to adapt our classification algorithm to the task we deal with in this work, namely ordering concepts from a textbook. To this aim, we created learning units for each concept of PRET 2.0 dataset and we used them as input for the classifier.

In order to verify the impact of different input data, we tested different strategies for the creation of learning units. Hence, content related to each concept was retrieved according to two different criteria: (1) considering all sentences where a certain concept occurs (Occurrence Model); (2) considering burst intervals for each concept. The latter is further divided into two cases depending on the appearing order of burst intervals: *(i)* burst intervals reflect their linear order along the text (Burst Intervals Model); *(ii)* burst intervals are reordered, having the most relevant burst interval as first (Most Relevant Burst Interval Model). The most relevant burst interval is defined as the first burst interval that exceeds the average length of all the bursts of that concept (Adorni et al., 2019; Passalacqua et al., 2019).

The resulting datasets show different learning unit dimensions: Burst Intervals models produce learning units with an average length of 534 tokens, while those considered for the Occurrence Model are smaller, with 250 tokens on average. While global features consider the entire content of the learning unit, for all models WE are computed only for the first n sentences. We tried different length of n: 5, 10, 15 and 30.

Results in terms of F-Score and accuracy were compared against a Zero Rule algorithm baseline.

Model	Emb. Dim.	F-Score	Accuracy
Occurrence	5	73.75	69.65
	10	**74.79**	**70.36**
	15	73.7	69.19
	30	73.11	67.97
	avg	73.84	69.30
Burst Intervals	5	71.75	65.54
	10	**73.91**	**69.49**
	15	72.97	67.77
	30	71.37	65.06
	avg	72.5	66.96
Most Relevant Burst Interval	5	**73.06**	**67.8**
	10	72.04	66.52
	15	71.58	64.43
	30	71.49	64.48
	avg	72.04	65.80
Baseline		66.66	50

Table 1: Classification F-Score and Accuracy values for the three models with varying number of sentences considered for lexical features. Average and baseline values are also reported.

4.4 Experiments Results and Discussion

Results reported in Table 1 show satisfying performances of our system that outperforms the baseline in all configurations. Best results are obtained by the Occurrence Model using 10 sentences to compute lexical features. In general, computing the WE on 10 sentences or less allows to obtain better performances in all settings. This could be due to the fact that the definition of a concept and its contextualisation with respect to other concepts are generally discussed by the author of the book when the concept is first mentioned in the text. Thus, sentences containing the first occurrences of the term seem to be the most informative for this task. To assess this hypothesis, we manually inspected sentences containing the first mention of each concept. The analysis revealed that 36.3% of the observed sentences contained a concept definition, thus supporting our intuition that the first mention is relevant for concept contextualisation.

The results obtained using the Burst Interval Model are slightly worse, although comparable, probably because, since burst intervals do not necessarily capture all the occurrences of a concept, in some cases the first mentions could be missing from the learning unit. The lowest scores are predictably those obtained using the Most Relevant Burst Interval Model: changing the order of the sentences penalises the system since the temporal order often plays an important role when a prerequisite relation is established between two concepts. Several algorithms exploit a time-based strategy for prerequisite extraction relying on the temporal nature of this relation (Sosnovsky et al., 2004; Adorni et al., 2018) and the analysis of human annotations suggests that the direction of this relation (i.e. A is prerequisite of B or vice-versa) tends to be highly correlated with the temporal order of the two concepts (Passalacqua et al., 2019). Besides, the most relevant burst is not necessarily the first burst interval for that concept and, for this reason, it could contain less relevant information about the concept and its prerequisites. Interestingly, the best results for this model are obtained considering only 5 sentences for computing WE, probably because the system has less chance of observing a lexicon related to other concepts.

If we look at the variation of accuracy values with respect to the classifier confidence (see Figure 2), we observe that our system shows an expected behaviour. In fact, at high confidences correspond high accuracy scores, while at confidence around .5 (12.66% of dataset pairs) we notice that the classifier is more unsure of its decision, obtaining results below the baseline. It should be noted also that the majority of concept pairs (25%) have been classified with a confidence value around .6, while the pairs obtaining the highest confidence value (i.e. equal to 1) are only 1.21%.

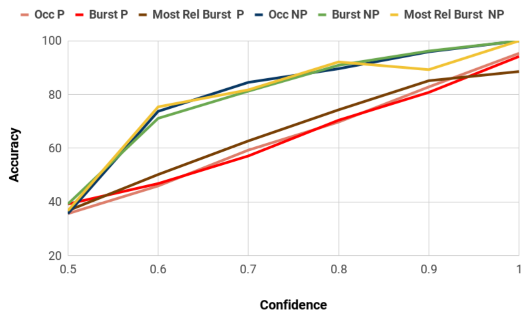

Figure 2: Variation of accuracy values wrt the classifier confidence for pairs labelled as prerequisite (*P*) and non prerequisite (*NP*) in all models considering 10 sentences to compute lexical features.

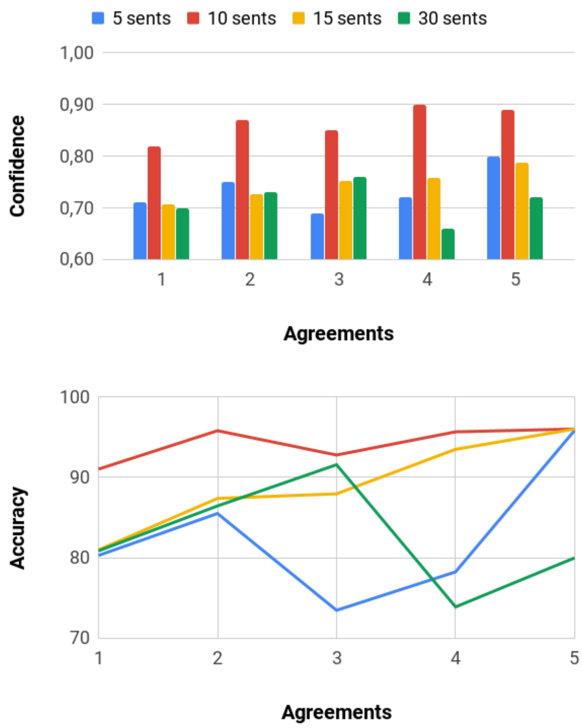

Figure 3: Variation of confidence (on top) and accuracy (on bottom) wrt the agreement value for the Occurrence Model (all possible embeddings length are considered).

The graphs in Figure 3 show the variation of confidence and accuracy values with respect to the annotators agreement. We report results only for the Occurrence Model since it is the one that obtained the best scores during classification. As we can see, the concept pairs for which all the annotators agree on tend to obtain higher confidence and, consequently, the classifier shows the best performances. The only exception is the model that computes WE using the first 30 sentences, which obtains instead the best scores on the pairs annotated by only 3 experts. The reason for this behaviour will be explored in future work.

5 Error Analysis

This Section compares the results obtained by the three models (i.e. Occurrence, Burst Interval and Most Relevant Burst Interval) when considering 10 sentences for computing WE.

The overall number of pairs assigned with a wrong label by the classifier is quite similar across each setting: 1,835 pairs for the Occurrences model, 1,923 for the Burst Interval model and 2,089 for the Most Relevant Burst model. Moreover, we observe that among these pairs more than 80% were classified as "prerequisite", suggesting that the system overestimates the prerequisite relation, assigning the label also to non–prerequisite pairs.

Focusing the analysis on relations that are annotated as prerequisites in the dataset, we observe how their prediction varies across models. 126 pairs were assigned with a wrong "non-prerequisite" label by all models showing similar average confidence values: 0.66, 0.66 and 0.62 for Occurrences, Burst and Most Relevant Burst model respectively. This result suggests that these pairs are particularly complex to classify. Conducting a deeper analysis on this subset, we notice that 85.71% (108) of the pairs are transitive pairs automatically generated (see Section 3). Such type of relations seems thus harder to classify than manually annotated ones and might require a different set of features to be recognised considering also that they represent more distant relations. Furthermore, consider that the remaining 18 pairs (14.28%) are manually annotated relations with low agreement values: 15, 2 and 1 were annotated by one, two and three annotators respectively.

6 Conclusion

In this paper we tested a deep learning model for prerequisite relation extraction in a real educational environment, using a dataset (PRET 2.0) built starting from a computer science textbook. The results demonstrated the effectiveness of our system, suggesting that it is possible to infer prerequisite relation out of textual educational material without using any form of structured information. Nevertheless, further work needs to be done, particularly for improving the performances of our system in a out-of-domain scenario, namely using concept pairs of a different domain during testing. Moreover, it could be useful to investigate the use of transitive relations and to study more accurately their impact on the system's performance. In addition, in order to identify prerequisite relationships while taking into account different types of relations (e.g. transitive ones) it could be interesting to frame our task as a ranking or multi-classification task rather than a binary classification one. Further analysis is also required to investigate the effect of using different numbers of sentences for creating WE. We plan also to explore the impact of using temporal reasoning on concept pairs (Adorni et al., 2019), which has not been considered in this work.

References

Giovanni Adorni, Felice Dell'Orletta, Frosina Koceva, Ilaria Torre, and Giulia Venturi. 2018. Extracting dependency relations from digital learning content. In *Italian Research Conference on Digital Libraries*, pages 114–119. Springer.

Giovanni Adorni, Chiara Alzetta, Frosina Koceva, Samuele Passalacqua, and Ilaria Torre. 2019. Towards the identification of propaedeutic relations in textbooks. In *International Conference on Artificial Intelligence in Education*, pages 1–13. Springer.

Rakesh Agrawal, Behzad Golshan, and Evangelos Papalexakis. 2016. Toward data-driven design of educational courses: A feasibility study. *Journal of Educational Data Mining*, 8(1):1–21.

James F Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11).

Abdelbaset Almasri, Adel Ahmed, Naser Almasri, Yousef S Abu Sultan, and Ahmed Y Mahmoud. 2019. Intelligent tutoring systems survey for the period 2000-2018. *IJARW, International Journal of Academic Engineering Research (IJAER), 3 (5):21-37.*

Chiara Alzetta, Forsina Koceva, Samuele Passalacqua, Ilaria Torre, and Giovanni Adorni. 2018. Pret: Prerequisite-enriched terminology. a case study on educational texts. In *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it 2018.*

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226.

Glenn Brookshear and Dennis Brylow, 2015. *Computer Science: An Overview, Global Edition*, chapter 4 Networking and the Internet. Pearson Education Limited.

Peter Brusilovsky and Julita Vassileva. 2002. Course sequencing techniques for large-scale web-based education. *Int. Journal of Continuing Engineering Education and Life-long Learning*.

Devendra Singh Chaplot, Yiming Yang, Jaime G Carbonell, and Kenneth R Koedinger. 2016. Data-driven automated induction of prerequisite structure graphs. In *EDM*, pages 318–323.

Chih-Ming Chen. 2009. Ontology-based concept map for planning a personalised learning path. *British Journal of Educational Technology*, 40(6):1028–1058.

Gabriel Pui Cheong Fung, Jeffrey Xu Yu, Philip S Yu, and Hongjun Lu. 2005. Parameter free bursty events detection in text streams. In *Proceedings of the 31st international conference on Very large data bases*, pages 181–192.

Fabio Gasparetti, Carlo De Medio, Carla Limongelli, Filippo Sciarrone, and Marco Temperini. 2018. Prerequisites between learning objects: Automatic extraction based on a machine learning approach. *Telematics and Informatics*, 35(3):595–610.

Jonathan Gordon, Stephen Aguilar, Emily Sheng, and Gully Burns. 2017. Structured generation of technical reading lists. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 261–270.

Qi He, Kuiyu Chang, Ee-Peng Lim, and Jun Zhang. 2007. Bursty feature representation for clustering text streams. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 491–496.

Jon Kleinberg. 2003. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397.

Jon Kleinberg. 2016. Temporal dynamics of on-line information streams. In *Data Stream Management*, pages 221–238. Springer.

Eugenijus Kurilovas, Inga Zilinskiene, and Valentina Dagiene. 2015. Recommending suitable learning paths according to learners preferences: Experimental research results. *Computers in Human Behavior*, 51:945–951.

Mikel Larranaga, Angel Conde, Inaki Calvo, Jon A Elorriaga, and Ana Arruarte. 2014. Automatic generation of the domain module from electronic textbooks: method and validation. *IEEE transactions on knowledge and data engineering*, 26(1):69–82.

Seulki Lee, Youkyoung Park, and Wan C Yoon. 2015. Burst analysis for automatic concept map creation with a single document. *Expert Systems with Applications*, 42(22):8817–8829.

Irene Li, Alexander R Fabbri, Robert R Tung, and Dragomir R Radev. 2019. What should i learn first: Introducing lecturebank for nlp education and prerequisite chain learning. *Proceedings of AAAI 2019*.

Chen Liang, Zhaohui Wu, Wenyi Huang, and C Lee Giles. 2015. Measuring prerequisite relations among concepts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1668–1674.

Chen Liang, Jianbo Ye, Zhaohui Wu, Bart Pursel, and C Lee Giles. 2017. Recovering concept prerequisite relations from university course dependencies. In *AAAI*, pages 4786–4791.

Chen Liang, Jianbo Ye, Shuting Wang, Bart Pursel, and C Lee Giles. 2018a. Investigating active learning for concept prerequisite learning. *Proc. EAAI*.

Chen Liang, Jianbo Ye, Han Zhao, Bart Pursel, and C Lee Giles. 2018b. Active learning of strict partial orders: A case study on concept prerequisite relations. *arXiv preprint arXiv:1801.06481*.

Weiming Lu, Pengkun Ma, Jiale Yu, Yangfan Zhou, and Baogang Wei. 2019. Metro maps for efficient knowledge learning by summarizing massive electronic textbooks. *International Journal on Document Analysis and Recognition (IJDAR)*, pages 1–13.

Alessio Miaschi, Chiara Alzetta, Franco Alberto Cardillo, and Felice DellOrletta. 2019. Linguistically-driven strategy for concept prerequisites learning on italian. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 285–295.

Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang. 2017a. Prerequisite relation learning for concepts in moocs. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1447–1456.

Liangming Pan, Xiaochen Wang, Chengjiang Li, Juanzi Li, and Jie Tang. 2017b. Course concept extraction in moocs via embedding-based graph propagation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 875–884.

Samuele Passalacqua, Frosina Koceva, Chiara Alzetta, Ilaria Torre, and Giovanni Adorni. 2019. Visualisation analysis for exploring prerequisite relations in textbooks. *First Workshop on Intelligent Textbooks*.

Sudeshna Roy, Meghana Madhyastha, Sheril Lawrence, and Vaibhav Rajan. 2018. Inferring concept prerequisite relations from online educational resources. *31st AAAI Conference on*

Innovative Applications of Artificial Intelligence (IAAI-19).

Sergey Sosnovsky, Peter Brusilovsky, and Michael Yudelson. 2004. Supporting adaptive hypermedia authors with automated content indexing. In *Proceedings of Second International Workshop on Authoring of Adaptive and Adaptable Educational Hypermedia at the Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004), Eindhoven, the Netherlands*.

Ilija Subasic and Bettina Berendt. 2010. From bursty patterns to bursty facts: The effectiveness of temporal text mining for news. In *Proceedings of ECAI 2010: 19th European Conference on Artificial Intelligence*, pages 517–522.

Yusuke Takahashi, Takehito Utsuro, Masaharu Yoshioka, Noriko Kando, Tomohiro Fukuhara, Hiroshi Nakagawa, and Yoji Kiyota. 2012. Applying a burst model to detect bursty topics in a topic model. In *International Conference on NLP*, pages 239–249. Springer.

Partha Pratim Talukdar and William W Cohen. 2012. Crowdsourced comprehension: predicting prerequisite structure in wikipedia. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 307–315. Association for Computational Linguistics.

Julita Vassileva. 1997. Dynamic course generation. *Journal of computing and information technology*, 5(2):87–102.

Annalies Vuong, Tristan Nixon, and Brendon Towle. 2011. A method for finding prerequisites within a curriculum. In *EDM*, pages 211–216.

Shuting Wang and Lei Liu. 2016. Prerequisite concept maps extraction for automatic assessment. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 519–521. International World Wide Web Conferences Steering Committee.

Shuting Wang, Alexander Ororbia, Zhaohui Wu, Kyle Williams, Chen Liang, Bart Pursel, and C Lee Giles. 2016. Using prerequisites to extract concept maps from textbooks. In *Proceedings of the 25th acm international on conference on information and knowledge management*, pages 317–326. ACM.

Wan C Yoon, Sunhee Lee, and Seulki Lee. 2014. Burst analysis of text document for automatic concept map creation. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 407–416. Springer.

Amal Zouaq, Roger Nkambou, and Claude Frasson. 2007. An integrated approach for automatic aggregation of learning knowledge objects. *Interdisciplinary Journal of E-Learning and Learning Objects*, 3(1):135–162.