# Finding the optimal topic sequence for online courses using SERPs as a Proxy

Sylvio Rüdian
Humboldt-Universität zu Berlin,
Weizenbaum Institute
Berlin, Germany
ruediasy@informatik.hu-berlin.de

Niels Pinkwart
Humboldt-Universität zu Berlin,
Weizenbaum Institute
Berlin, Germany
niels.pinkwart@hu-berlin.de

## ABSTRACT

Finding the optimal topic sequence of online courses requires experts with lots of knowledge about taught topics. Having a good order is necessary for a good learning experience. By using educational recommender systems across different platforms we have the problem that the connection to an ontology sometimes does not exist. Thus, the state of the art recommenders can suggest courses with an optimal order within a platform. But on a more global view, a recommendation across different platforms with optimal order is not existing as long as no ontology was defined or courses are not connected to an existing ontology. Nowadays experimental approaches manipulate the learning paths to find the optimum. As this can impact the learning experience of participants, this approach is ethically unacceptable. To overcome this problem, we propose a data-driven approach using the search engine result pages (SERPs) of Google. In our experiment, we used pair-wise search queries to get access to web pages, those 38.000 texts were used to test some NLP metrics. 10 different metrics were examined to create an optimal order that was compared to the optimal sequence defined by experts. We observed that the Gunning Fog Index is a good estimator to determine the optimal order within a cluster of topics.

## Keywords

Course Sequencing, educational recommender system, web search, adaptive courseware, personalization.

## 1. INTRODUCTION

Providing the optimal sequence of topics in online courses is of high interest because it influences the learning outcome as well as motivation. Lots of MOOCs are existing, but in which order they should be done is defined by experts and this is a time-consuming procedure. Large-scale educational recommender systems [1] suggest online courses across different platforms. Creating an optimal sequence based on an ontology is an easy solution as an ontology includes the optimal order, defined by human experts. This can be done within single platforms, but an ontology across different courses across several platforms is not existing. McCrae et al. [2] state that it “is difficult to link to ontologies”. The willingness to create a connection of own online courses to an existing public ontology is low as this is expensive due to manual work on the one hand but can also result in course recommendations of other suppliers on the other hand, which does not meet the interests of the suppliers.

The optimal sequence is missing in recommender systems as long as no manually created large-scale ontology or optimal sequence exists. Recommender systems only provide a ranking based on how well the suggested courses fit into the user's learning situation. There are existing approaches, e.g. linked data to create a structured semantic web [3]. Their idea is to create a network that contains the meaning of the data. But there is the problem that the semantic web is limited to specific domains. If the networks have not been created for the topics that we need, we cannot use them. Besides, the structure in the semantic web is designed to understand relationships between objects, not whether there is a dependency from the educational perspective. Further on, there is the problem that topics for online courses often consist of multiple words to describe the topic or concept. Finding the correct corresponding concept within the semantic web can be challenging.

Having an optimal order of online courses is of high interest in online education as many topics require the knowledge of subtopics. Knowledge dependencies can be modeled by experts manually on the one hand, but this is a cost-intense procedure that requires lots of knowledge about the taught topics and provided courses as well. On the other hand, the world wide web is full of contents of different quality. Every topic that can be taught can be found there, but the contents of web pages are still not used for topic sequencing in education. Crawlers get access to all the texts and companies like Google define an order of pages related to a search query. Within a search engine, we get access to all pages that they define to satisfy the user intent [4]. Using this large number of pages for each topic could be beneficial in creating optimal topic sequences for online courses.

An optimal order is very important for a good learning experience in online courses. We define an optimal order as the sequence of course topics where each topic should be taught when all pre-requirements are fulfilled based on the previous courses. As long as topics are taught where the requirements are missing, the dropout rate will be high. Using courseware (single parts of a course) [5] to generate a new online course it is important to have an optimal order. Otherwise, the participant cannot understand the topic because of missing knowledge. The same problem exists in AI-generated learning paths of online courses, which must be consistent according to the fundamental didactical method of starting teaching basics, not with specialized knowledge.

Sylvio Rüdian and Niels Pinkwart “Finding the optimal topic sequence for online courses using SERPs as a Proxy”. 2021. In: Proceedings of The 14th International Conference on Educational Data Mining (EDM21). International Educational Data Mining Society, 317-322. https://educationaldatamining.org/edm2021/
*EDM ’21* June 29 - July 02 2021, Paris, France

Observing the world wide web, we can get a variety of texts on any topic. We want to use this already existing large set of pages to find the optimal topic order of online courses with an experimental approach. To access all the pages with corresponding texts that we need, we use the search engine Google, especially the Search Engine Result Pages (SERPs) [6]. It is known that Google uses the semantic web in the background, depending on the search query, which helps to overcome the challenge to find the optimal corresponding concept of the semantic web using the search engine as a proxy [7]. This is beneficial for the case that for specific domains no linked data is existing – as the search engine tries to provide related resources, even if they are misspelled and the search engine can give results for queries that they have never seen before.

Using topics as keywords results in a list of pages that satisfy the user intent according to Google [4]. This can be used as a base for having access to different features for ordering course topics. SERPs help to understand the popularity of how many pages are indexed by the search engine, which could be an indicator for good sequencing as less specialized topics are existing compared to general basic topics. Besides, having a lookup for two topics in one search, we get pages that contain both keywords, those frequency or deviation could be an indicator for finding the optimal sequence. If we observe online courses, then we usually have an increasing difficulty level. Using the complexity of texts could help to estimate the optimal order based on the difficulty.

In this paper, we concentrate on the three research questions:

1) Is the SERP popularity of all topics a good indicator to find an optimal topic sequence for online courses?

2) Is the topic frequency of page texts that are listed within the SERPs an estimator to determine the optimal topic sequence in online courses?

3) Does ordering topics' texts by text difficulty metrics result in a sequence that is appropriate to be used as a sequence in online courses?

## 2. RELATED WORK

Brusilovsky et al. [5] define this problem as "sequencing of lessons" where each lesson is connected to a topic. This contains numerous chunks of educational material, ranging from videos and texts to different interactive tasks. The authors use a domain concept structure, that is stored independently from teaching materials. Each concept needs to be linked to the teaching material. It has the advantage of being able to use the courseware to generate a personalized online course according to the interests and knowledge gaps of a learner. This approach is comparable to using an ontology that needs to be defined by experts, based on rules and graph representation. It is the fundamental model to define an optimal sequence of online courses but requires the creation of the ontology by experts.

S. Fischer [8] uses an ontology knowledge base, namely a "knowledge library" to create an optimal course sequencing. Therefore, they use modularized media content as courseware together with metadata that describes the link to the ontology model. With that, they have access to a taxonomy that can be used to create a good ordering of topics as well as generating questions with right and wrong answers (depending on the granularity of the ontology). The modular resources can be used to generate courses, according to the knowledge gaps of learners.

Xu et al. [9] propose to learn from users providing specific course sequences for testing and use their performance to create an optimal sequence for new users. While this approach works it has the disadvantage that it requires real test users which may perform badly within the scenario. Doing this in a field study is acceptable but using real students is not sustainable from an ethical point of view. We want to emphasize that we do not want to use this experimental user behavior data as this is ethically not acceptable.

Cucuringu et al. [10] used already captured student participation in courses to create pair-wise comparisons using ranking aggregation to create a global ranking. This ranking proposes an order of how courses should be taken by students. One major problem is incomplete data as some pairings are not existing for a comparison.

S. Morsy [11] states that a global ranking of online courses cannot be used for personalized recommendations. But having a global ranking can be helpful to determine which courses should be done in which order. Combining this knowledge with personalized courses or topic recommendations is helpful as the course dependencies (e.g. what knowledge is necessary to understand a topic) are the same for personalized recommendations, which are filtered by topics/concepts that the learner is already aware of. Thus having a global ranking can be beneficial for personalization as well.

Using the information of chosen courses by students and their performance is a good way to determine an optimal course sequence. A major limitation with that approach is the limitation of data and to have access to chosen courses and the resulting performance. This approach does not comply with the GDPR as the information on whether students passed or failed an exam is classified to be sensitive personal data, that cannot be accessed for course sequencing in general [12]. Thus, their application does not work in a real-world scenario in the EU. Based on the limitations of being dependent on user performance or manually created ontologies, we propose a new methodology to create an optimal order of online courses, based on their topic.

## 3. METHODOLOGY

As we learned from Rüdian et al. [13]: Even if experts are scoring the same results of educational tasks, their scores vary among each other. If we observe the order of topics, then we know that there is not always a perfect solution regarding the whole sequence because of ambiguous expert opinions. In the pre-study, four experts (AI instructors) had the task to create the optimal order of 20 AI-related topics to be taught within online courses. We used the following topics: *neural networks, voice recognition, chatbots, Linux, data visualization, Python, statistic basics, part-of-speech tagging, LSTM, data preparation, deep learning, TensorFlow, object recognition, Naïve Bayes, natural language processing, ethical principles, clustering, reinforcement learning, cross-validation, and regression*. The resulting sequences are then used to make a pair-wise comparison to understand the overlap across instructors and to see where we have a high overlap. The pair-wise sequence score S is defined as followed: For every topic A and B of the expert sequence with $A \neq B$ and every topic C and D of the sequence derived by the algorithms or another expert with $C \neq D$ we count all hits where $(A < B \text{ and } C < D)$ or $(A > B \text{ and } C > D)$. Thus, the topics have the same order within both sequences. This number is divided by the number of possible combinations, defined as S.

Some topics have dependencies; e.g. neural networks should be introduced before teaching LSTM or natural language processing should be taught before starting with part-of-speech tagging; others do not have strong relations and can be taught somewhere, e.g. ethical principles or Linux.

The idea of the main study is to compare the sequences created by instructors with algorithmic ones. Therefore, it is a good fundament to measure tutors' decisions among each other first to define accuracy as a gold standard that we want to achieve with our methodology. Thus, we need a realistic generalizable accuracy that we should achieve instead of over-optimizing our approach with high accuracy that is the optimum for a sequence of one expert only. Besides, the pre-study identifies the optimal order of the topic subsets that are the same for all experts. The order of these topic subsets will be compared to the order that we get from our algorithms to see how well the algorithms perform in a real-world scenario.

Our approach is to use the Search Engine Result Pages of Google (SERPs) and we derive different metrics based on the results. A search engine can be used to find web pages that are related to given keywords. One of the main purposes of the search engine Google is to satisfy the user intend by providing a list of web pages that are related to the search query [4]. Thus, using it allows us to get pages that have a high authority according to the Google ranking algorithm, which is, according to them, a metric of high quality. We use this list of pages with our topics as search queries to understand the popularity, the number of user searches, the complexity of topics, and which topics have a semantic connection. These metrics are then used to create a sequence, based on a linear order of the observed data. These sequences are compared to the experts' ones to understand whether there is a connection between our metrics and the optimal order, defined by experts.

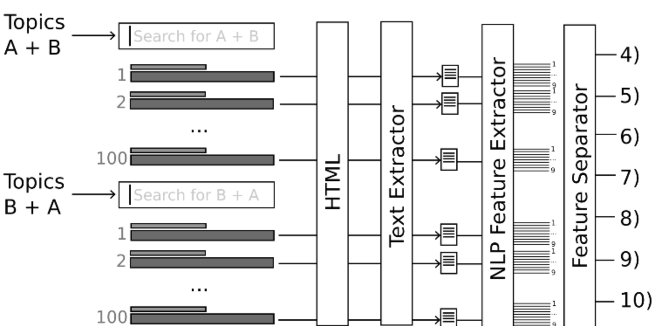


**Figure 1. Pipeline for pair-wise search, extraction, and separation of features.**

The approach of using a search engine as a basis has the advantage that we do not need to do experiments with students where they may be badly influenced due to bad testing sequences. Thus we are independent and can use our approach on a larger scale. That makes our approach more practically usable. We use different data as a basis, use them to rank our 20 topics, and compare the order with the instructors' ones. Our approaches are the following:

1) We use the number of topic results that are estimated by the search engine by searching for every keyword separately.

2) We use the number of topic results pair-wise keyword combinations and observe the number of estimated results.

3) We use the keyword search estimator and rank our topics according to the estimated search amount.

4) We use the first 100 results of all pair-wise keyword combinations and count how often both keywords within the 100 listed pages exist.

5) We use the first 100 result pages as in 4), search for both keywords on the pages, and summarize how often each keyword occurs at first in the text.

We use the 100 result page texts and apply three algorithms to estimate the text complexity, namely 6) Flesch-Reading-Ease (FRE) [14], 7) RIX [15], and 8) Gunning Fog Index (GFI) [16]. Then we use the 100 result page texts with basic NLP metrics: 9) The type-token ratio (TTR) and 10) the number of words per sentence (NoW). We assume that observing how many pages are existing in combination helps to identify topics that have a semantic connection. Using the information on how many pages are existing gives hints about the popularity (1,3), where for complex topics mostly less content exist than for basics. Observing the complexity of the contents (4-7) could help to identify the difficulty level of topics to find the optimal sequence. Figure 1 visualizes the method for 4)-10) to get features based on a pair-wise topic search.

The Flesch-Reading-Ease Index is based on the "Standard Text Lessons in Reading" [17] and is calculated from the average sentence and word length [14]. The main idea of the Gunning Fog Index is to reduce the complexity in newspapers as a kind of warning system for authors that texts are not "unnecessary complex". Therefore, the author uses the sentence lengths, the number of syllables, easy words, and hyphenated words to estimate the complexity of a text [18]. The "Regensburger Index" (RIX) uses difficulty parameters like passive, sentence complexity, and predications to derive the complexity [15]. All approaches differ in the selection of features that are used to create the indexes.

Finally, we use a random forest regressor [19] to predict the pair-wise sequence, using the data of 4)-10) to estimate the feature importance to support our findings. To get all the data, including the SERPs, all pages, and the estimated search amount, we use a commercial web crawler for SERPs[1]. This is necessary as the pair-wise lookup of 20 keywords results in $20 * 20 - 20 = 380$ searches, where we need to download 100 web pages each, resulting in 38.000 files. A simple crawler that we used in our lab before, was banned after 20 crawls, thus using a commercial one is the most efficient option.

Each data source 1) – 10) is then used to create a ranking of topics, based on their linear order. These sequences are compared with the expert ones to find the optimal feature that can be used in a real-world setting. To compare the sequences of the experts with the algorithmic ones, we use a pair-wise topic comparison to test whether the order is the same in both sequences and summarize the hits. Thus we can compute the overlap that represents the accuracy in our experiments.

## 4. RESULTS

The overlaps across the expert sequences range from 0.6 to 0.8 (Table 1). Thus we have an orientation of the resulting overlap that can be achieved with our approach at maximum. While the overall sequences defined by experts are partly different, we identified some partial sequences that are identical across all expert-based rankings and use them as ground truth. We detected some matching sequences of topics:

A = ["data preparation" → "data visualization" → "clustering"], B = ["neural networks" → "deep learning" → "LSTM"], and C = ["natural language processing" → "part-of-speech tagging" → "voice recognition" → "chatbots"],

[1] https://seorld.com/crawler

where [A → B] means that topic A needs to be explained before topic B. This makes sense as each topic mostly requires knowledge of the previous one(s), e.g. "neural networks" have to be introduced first and after that, "LSTM" can be explained. We use the three clusters (A, B, and C) to visualize whether our rankings make sense in a real-world scenario as the overlap of sequences defined by a number only is too abstract. All in all, in our pre-study we can conclude that we identified three clusters using sequences of four experts.

**Table 1. Pair-wise sequence overlaps of 4 AI experts.**

| | Expert 1 | Expert 2 | Expert 3 | Expert 4 |
|---|---|---|---|---|
| Expert 1 | - | .60 | .65 | .80 |
| Expert 2 | - | - | .65 | .65 |
| Expert 3 | - | - | - | .75 |
| Expert 4 | - | - | - | - |

Then we used all the different data points that we got from the crawler separately and created a sequence based on their linear order. Table 2 shows all the results of our experiments. We calculated the pair-wise overlap to compare estimated sequences with the expert ones. Also, we tested whether our partial sequences of the topic sets in A, B, and C have the same order as defined by our experts.

Observing 1) - 3) we can answer the first research question as these metrics represent the popularity of topics within the SERPs. The ordered list of topic pages is not a good indicator to find an optimal topic sequence for online courses. Thus, popularity is not a good indicator of course sequencing.

**Table 2. Overlap of sequences with four experts (E1…E4) and the information on whether the orders of our clusters A, B, and C are the same as defined by experts.**

| Approach | E1 | E2 | E3 | E4 | A | B | C |
|---|---|---|---|---|---|---|---|
| 1) | .55 | .40 | .45 | .50 | No | No | No |
| 2) | .60 | .50 | .40 | .50 | Yes | No | No |
| 3) | .45 | .45 | .40 | .53 | No | No | No |
| 4) | .53 | .63 | .53 | .53 | No | No | No |
| 5) | .58 | .53 | .53 | .40 | No | Yes | No |
| 6) FRE | .35 | .50 | .55 | .50 | No | No | No |
| 7) RIX | .50 | .50 | .50 | .45 | No | Yes | No |
| **8) GFI** | **.55** | **.65** | **.60** | **.60** | **Yes** | **Yes** | **Yes** |
| 9) TTR | .45 | .40 | .40 | .40 | No | No | No |
| 10) NoW | .50 | .50 | .50 | .50 | No | No | No |

Observing the pair-wise searches in 4) and 5) we can conclude, that topic frequency within the related texts is also not a good indicator to get an optimal sequence of topics, which answers the second research question. We limited the search to exact matches. Further, using n-grams or other methods to detect variants could be beneficial.

We identified the Gunning Fog Index as an estimator to create an optimal order. This answers our third research question. Using this metric for text complexity is the most robust feature to create a good sequence of topics in our experiment. Also, the order we got from our clusters is the same as in the sequence that we got by using the GFI. This is very important for a practical educational environment as the orders of topics that have a taxonomy with knowledge dependencies need to be done correctly. The overlap with the expert sequences ranges from 0.55 to 0.65, which is acceptable as the overlap of sequences across experts was in the range of 0.6 and 0.8. The remaining text complexity metrics (6, 7, 9, 10) are not as robust as the GFI.

To get more insights into the importance of the identified predictor, we use the random forest regressor [19] as we investigate linear

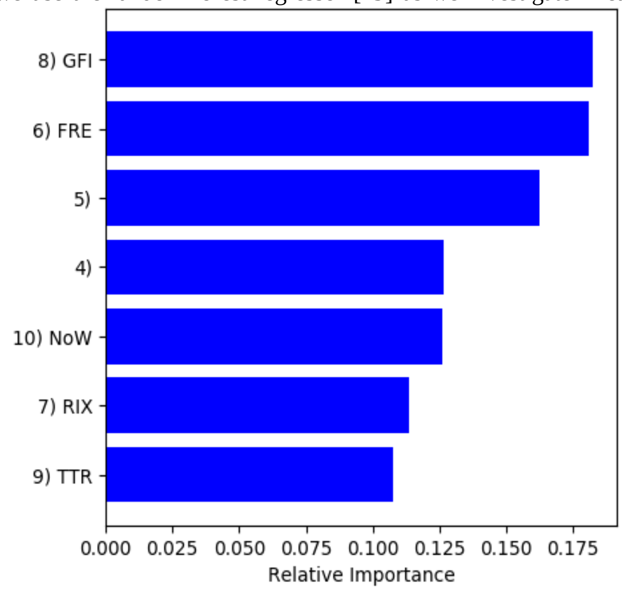


**Figure 2. Relative importance of features according to the random forest regressor.**

features only and – for future work – we want to identify features of high importance that also work for non-linear dependencies to predict the optimal sequence. Therefore, we use all pair-wise approaches (4-10) to train a decision tree using the random forest regressor. As prediction target, we used all pairings orderings (e.g. topic "neural networks" needs to be taught before topic "LSTM"). This is a classical approach to predict the ordering of items, based on different features. Figure 2 displays the relative importance of features that we got. The most relevant feature is the Gunning Fog Index (GFI), which performs best in our experiments as well.

## 5. DISCUSSION

Automated analysis of the pair-wise SERPs and the text complexity using the GFI can help to assist instructors during planning course sequences. From an ethical point, doing experiments with students is not justifiable as it could corrupt the learning outcome as a bad implication. As our approach is independent of experimenting with users, this method can be applied on a large scale. Combining this approach with educational recommender systems, we can provide a sequence of topics, based on the topic set that we get from the recommender system, even if no ontology is defined in the background. Using the text complexity helps to start with topics that can be explained more easily than the following ones. Having automatic composed online courses based on courseware, it can be beneficial to use the third party data of SERPs to find an optimal order. This is an important step to create personalized online courses that are adaptive to the knowledge level, where no pre-defined ontology exists.

There are various fields of application where we can use our approach. This method can be used for planning lecture sequences at school or university, based on the complexity of taught topics. It is the same in preparing new lectures, based on existing learning material, that can be composed in an optimal order. Besides, the curricula at universities could be optimized, where students participate in courses of different universities. Having a recommendation for a good order on which courses should be visited at which point of time is beneficial.

From a practical point of view, it is important to note that the number of searches, while using a commercial crawler, is a cost factor. If $n = \text{“}number\ of\ topics\text{”}$, then the number of searches $C = n^2 - n$, having pair-wise searches $A + B$ and $B + A$ (with A and B being topics of the list). This is necessary as the SERP list of $A + B$ is not the same as $B + A$. Also, the search query $A + B$ returns 100 pages that need to be crawled to get the texts. In our experiment, this results in 16Gb of data, having 38.000 texts of 20 topics with pair-wise searches, where the metric has to be derived for each text. The required storage grows exponentially with the number of topics. The length of the resulting topic list of educational recommender systems can be limited in general, thus it is not a problem, but it is important to limit the list first, before finding the optimal order to avoid the need for large storage and high computational capacities. Besides, using the first 10 results of the SERPs instead of 100 reduces the crawling budget as well as computing time, but makes the approach less robust.

In our experiment, we conclude that commercial popularity and the estimated search amount are no indicators for a good topic sequence. Independently from the intention of the paper, using popularity is a helpful metric to get insights into trends about what people are searching for. Online course suppliers can use this information to create online courses for a large audience, those sizes can be estimated with the search popularity. As data-driven approaches, e.g. AI-related decisions require lots of participants, offering online courses that are of high interest can help to get the required number of participants to have enough training data for AI methods. From the researchers' perspective having popular courses is of high interest to obtain AI decisions with a high statistical significance. Sources like the Semantic Web do not provide this additional information.

As this is ongoing research, the next step is to create a comparison of the identified cluster sequences with sequences that can be derived using the semantic web as proposed by Toman & Weddell [20]. This real-world experiment can show the applicableness in the field of education. If this method results in similar sequences, we recommend using an already existing semantic network and in case of missing concepts, we can use our method as a fallback.

Observing the overlap of expert sequences, we can see that they are quite diverse. Finding an “optimum in education” is mostly a trade-off between different opinions of experts. We used the sequences to detect partial sequences that are similar across all experts. In the future, all topics should have a description of the taught contents to reduce the variety of sequences. Examining the detected partial sequences, we can see that these topics have a semantic connection and some topics have knowledge dependencies. In a future scenario, we recommend finding clusters of topics first and then use a text complexity metric like the GFI to get the optimal order. Otherwise, there might be a switch of topics, those order is good while looking at the complexity only, but could be confusing on a more global view. From the didactical perspective, switching between different topics in the learning path that have little semantic coherence is not recommended.

In this paper, we focused on AI-related topics to present our research at an early stage. It is of high interest to compare our approach to data from another domain. We assume that the GFI as a complexity metric can be used as an indicator for a useful order as well. But it is important to note, it remains possible that GFI randomly happened to give a good result. Thus extending the experiment to different domains is necessary to give a final and scalable recommendation.

Besides, we assumed that using text difficulty metrics will result in nearly the same order as their task is identical. Observing the results, we can see that there are major differences in the resulting order. The GFI is used to estimate how many years of formal education the reader needs to understand the text on the first reading [16]. In our case, it was the best and most practical metric. Looking at Figure 2, the Flesch-Reading-Ease is also of high importance but failed to create the optimal order of our three clusters (Table 2). Comparing the GFI with FRE, both metrics are based on syntactical features. The GFI is enriched with contextual features like “easy words”. This enrichment could be a reason why this index works best in our experiment. Besides, other textual metrics need to be taken into account for testing. Semantical features could be used as well as text entailment. In the future, combining these metrics can be beneficial, e.g. at training a neural network with all metrics to use non-linear dependencies, that were not examined in this paper yet. Textual metrics must be used carefully as they are “just” formulas for judging the complexity of texts [21]. The methods cannot be used to judge the appropriateness of contents or whether the content is correct. Thus, selecting learning material of high quality is important and the metrics are not useful in the selection process.

The proposed approach depends on the SERPs of Google. Having a high fluctuation of rankings within the SERPs could change the feature's importance. As Google regularly updates their algorithms within a core update twice a year, rankings may change [22]. As we use the first 100 results we assume that the approach is robust because there are only minor changes if we consider the set of the first 100 pages. It is debatable that high-ranking Google results contain web pages of high authority, it can be discussed whether the first 100 resulting pages are a good resource for educational purposes and whether they are trustworthy. Instead, they are likely to be optimized for search engines, e.g. by search engine optimizers that create contents with ingoing links of high authority web pages aiming to have high rankings. There is the problem, that often texts of competitors are re-written for new pages to rank for similar terms. Thus, many texts with similar contents can be found. Besides, the SERP came from multiple contributors, they may include low-quality texts from commercial sources and web pages that block search engines are systematically excluded.

We use Google as a proxy to get access to the web pages that contain the texts that we are working with. The same can be done with other search engines. Alternatively, being limited to resources those contents are created by editors of publishing houses for education may be biased as the complexity of texts also depends on the writing style of authors. Using a resource like the first 100 texts results in a more robust view to avoid this bias due to averaged data. It can be discussed whether Google is a good source for characterizing academic terms because SERPs might be too inclusive and therefore noisy. Based on our experiment we could see that a text difficulty average of the gathered data can be a good indicator. Whether this is the case, in general, has to be examined in further experiments. Besides, we could use educational materials from publishing houses. In general, they are not publicly accessible, which increases the costs if we want to use them. Further research can examine whether resources like Wikipedia or Web of Science can be used with similar metrics to determine the optimal order.

Limiting the approach to the header of courses could generally lead to wrong conclusions if the courses do not cover the topic that was given in the headline. In our experiments, we used topics as keywords only. Using the course description or the course(ware) content itself to obtain more rich information for having richer keywords could be beneficial, that will be addressed in further experiments. Besides, we did not consider synonyms, which should be observed in future studies because using different words (even synonyms) results in different SERPs.

## 6. CONCLUSION

In this paper, we propose different strategies to use texts that we got using a search engine to find the optimal order of online course topics. The pre-study has shown that the optimal topic sequences differ among experts. But we can also observe that there are partial topics that have the same order in all expert sequences. We identified them to define a gold standard and to check for the practical usefulness. The sequences derived by our approaches were compared to the expert ones and the order of the partial topics. The commercial popularity, that can be derived by searches in search engines is not an indicator of a good topic sequence. Searching for pair-wise topics and comparing the text complexity of the SERPs' web pages' texts can be used as an indicator for creating a plausible order of taught topics within online courses.

We identified the Gunning Fox Index as the most robust metric for topic sequencing. We can conclude that this feature helps to find the optimal sequence for automatic composed online courses to personalize them ethically without using students giving them randomized learning paths that could impair their learning experience as well as their learning outcome.

## 7. ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DII127 (Weizenbaum-Institute). The responsibility for the content of this publication remains with the authors.

## 8. REFERENCES

[1] N. Manouselis, H. Drachsler, K. Verbert and E. Duval, "Recommender systems for learning," Springer Science & Business Media, 2012.

[2] J. McCrae, D. Spohr and P. Cimiano, "Linking Lexical Resources and Ontologies on theSemantic Web with lemon," in *Extended Semantic Web Conference*, Springer, 2011.

[3] P. Jain, P. Hitzler, P. Z. Yeh, K. Verma and A. P. Sheth, "Linked Data is Merely More Data," in *D. Brickley; V. K. Chaudhri;H. Halpin; and D. McGuinness, editors, Linked Data Meets Artificial Intelligence*, Menlo Park,CA, AAAI Press, 2010, p. 82–86.

[4] W. Gross, T. McGovern and R. Sturtevant, "Search engine using user intent". USA Patent US20060064411A1, 2005.

[5] P. Brusilovsky and J. Vassileva, "Course sequencing techniques for large-scale web-based education," in *Int. J. Cont. Engineering Education and Lifelong Learning, Vol. 13, Nos.1/2*, 2003, pp. 73-94.

[6] D. Kelly and L. Azzopardi, "How many results per page?: A Study of SERP Size, Search Behavior and User Experience," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015, p. 183–192.

[7] T. Steiner, R. Troncy and M. Hausenblas, "How Google is using Linked Data Today andVision For Tomorrow," in *Proceedings of Linked Data in the Future Internet at the Future Internet Assembly*, Ghent, 2010.

[8] S. Fischer, "Course and Exercise Sequencing Using Metadata in Adaptive Hypermedia Learning Systems," in *ACM Journal of Educational Resources in Computing, Vol. 1, No. 1*, ACM, 2001, pp. Article #3, 21 pages.

[9] J. Xu, T. Xing and M. v. d. Schaar, "Personalized Course Sequence Recommendations," IEEE, 2016.

[10] M. Cucuringu, C. Z. Marshak, D. Montag and P. Rombach, "Rank Aggregation for Course Sequence Discovery," 2017.

[11] S. Morsy, "Learning Course Sequencing for Course Recommendation," 2018.

[12] EU, "Processing of special categories of personal data," in *Art. 9 GDPR*, 2018.

[13] S. Rüdian, J. Quandt, K. Hahn and N. Pinkwart, "Automatic Feedback for Open Writing Tasks: Is this text appropriate for this lecture?," in *DELFI 2020 - Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.*, 2020, pp. 265-276.

[14] T. v. d. Brück and J. Leveling, "Parameter Learning for a Readability Checking Tool," DOI 10.5281/zenodo.2552414, 2019.

[15] J. Wild and M. Pissarek, "RATTE - Regensburger Analysetool für Texte," Universität Regensburg, 2016, pp. 1-12.

[16] R. Gunning, "The fog index after twenty years," in *Journal of Business Communication 6.2*, 1969, pp. 3-13.

[17] F. R, "A new readability yardstick," in *J. Appl Psycho 1* , 1948.

[18] R. Gunning, The Technique of Clear Writing, McGraw-Hill, 1952.

[19] L. Breimann, "Random Forests," in *Machine Learning 45*, 2001, p. 5–32.

[20] D. Toman and G. Weddell, "On Order Dependencies for the Semantic Web," in *Conceptual Modeling*, 2007, pp. 293-306.

[21] J. W.Pichert and P. Elam, "Readability formulas may mislead you," in *Patient Education and Counseling 7(2)*, 1985, pp. 181-191.

[22] Seorld, "Google Update," 05 10 2020. [Online]. Available: https://seorld.com/blog/seo/google-update. [Accessed 21 10 2020].