https://www.researchgate.net/publication/365202930

Abstract
Systematic review of available literature on explainable recommendation systems in education and their reproducibility.
First part employs an NLP-powered toolkit that automates a large portion of review process. Quantitaive review of all literature followed by qualitative of few selected articles. Analyzed and compared based on purpose of recommendations, tools and techniques used, and wether the research is easy or hard to reproduce.

2 Methodology
To identify the work that has already been completed on our topic of interest, we will use an NLP toolkit designed for automating systematic, scoping, and rapid reviews [24]. Requires structured input of data consisting of keywords, properties, property groups, required relevance, included sources, and beginning and ending years.
Keywords, as input parameteres, serve as search terms.
Properties are words or phrases that must appear in title, abstract and keywords. Organized in property groups to address synonyms and variants. Each property group must be represented in the title
or abstract of the paper for it to be included in the results.
Relevance
is configured using a dedicated input parameter, which is an integer that specifies the minimum number of occurrences of each property group for an article to be considered relevant. In our case, the relevance setting was set to 2. This means that words or phrases from at least 2 property groups must be presents in the article's title or abstract for it to be considered relevant. Parameter value of 2 has shown to be the most accurate. Lowering to 1 results in many irrelevant papers. Rising to value 3 and more is just too big restriction, articles in results too low.
The scope of the returned articles is also constrained by the required input parameters: beginning and ending years.
The scope of the returned
articles is also constrained by the required input parameters: beginning and ending years. There is a comprehensive description of all input parameters in the original paper [24].
PRISMA, the "Preferred reporting items for systematic reviews and metaanalyses - PRISMA statement" [17], [18] is the methodology used for selecting and processing the articles. The PRISMA methodology aims to standardize these types of surveys.
Figure 1 PRISMA Workflow
Quantitative analysis of articles, number of articles over years and which contain particular keywords. Graphs to illustrate findings. After that qualitative investigation. Ascertain wether or not research presented in the articles was pertinent to educational recommendation systems and wether or not the methodologies described are both explicable and reproducible. Articles were included in the qualitative
analysis after manual inspection of their relevance to the topic.

The result of the NLP toolkit used in this paper emerges with the production of multiple files. These files include the visualization of the results as
charts in vector PDF files, CSV files that contain all of the articles that have been filtered, files containing only articles segmented by different properties and property groups, as well as a BibTeX file that contains all of the article data
and helps to simplify the process of citing sources. The charts that are obtained

provide a summary of the findings according to a variety of metrics, including the number of articles by country, the number of articles per keyword, the number of articles per year, and the number of articles per source (library). More charts can be easily generated from the data included in the CSV result files, since these files contain the complete data for each provided article, such as publisher, publication year and place, authors, their affiliations, origin countries and emails, the relevance factor for each included articles, number of citations, DOI numbers and other soft data.

Reproducibility of each paper is graded on 5-point Likert scale. Grading is done based on the ease of replicating the results outlined in the paper, the data accessibility and the clarity of the implemented algorithms. A 5 point reproducibility score means that the research is performed on publicly available dataset, its results can easily be replicated in our own environment and the results verified. A 4 point score means that the research is thorough and clear, but the dataset was not included and is not publicly accessible. A 3 point grade means that the dataset is not included and the explanation is vague but relies on popular concepts and can easily be assembled using common knowledge of ML techniques. Repdocibility score 2 means that data is missing and used techniques are not well explained, but the research can be reproduced although the process would be very hard and much guessing would be needed. Finally, a reproducibility score of 1 means that the research process is impossible to be replicated with the given data and information.

3. Results

When we compared the number of citations that each of the chosen articles has with the given reproducibility score, we made a discovery that was both interesting and remarkable. It would appear that the likelihood of an article being cited by others increases in proportion to the degree to which it provides a detailed description and is simple to reproduce.

4. Conclusion

An interested finding that emerged from this research is the correlation between the number of citations certain research paper has and the ease of reproducing the same paper. It turns out that researchers prefer citing papers which can be easily understood and which are suitable for reproducing in the own research process. Having access to a publicly accessible dataset when reproducing certain research can prove crucial for replicating the results.