## Topic Modelling Process

### Tweaking k

Working with topic modelling requires the researchers to determine how many clusters of documents (k) the model should cluster documents into. This is an iterative process and requires the researcher to understand the outputs of the process.

### Output Interpretation

Topic models provide valuable insights into documents by clustering them. Reseachers are still required to interpret the results and label the outputs. Therefore they need understanding on the topics in the corpus. Novice researchers do not have this by defenition and require document exploration to attain it.

### Accuracy Assessment

Topic models can make mistakes in clustering the documents they work with. Precision and recall are indfications of model performance on clustering documents. It requires insight into the input dataset to assess performance.

## Related Document Acquisition

### Keyword Iteration

Finding the right keywords to acquire documents can be difficult. Students during interviews pointed they have an iterative process. They adjust their search terms according to results of the search and the documents they read.

### Document Availability

Availability of papers was said to be an issue for document acquisition. Students report that companies hold valuable information and papers are behind paywalls not supported by their institution credentials. Availability also varied depending on research topic

## Relevant Information Extraction

### Information Density

Students reported searching for information relavant to their projects . The density of this new information varies among documents. Too high density and reading becomes difficult. Too low and documents are not worth reading.

### Document Styles

Document formatting varies among academic documents. Students preferred specific type of formatting and writing styles. Formatting was more important for larger documents. A clear abstract and well indexed section were the most important aspects of papers read.

## PaperScout: Document Exploration

### Document Navigation

Paperscout, the tool i designed, has the function to navigate quickly to specific parts of the document. Doing so is done according to some popular scouting methods including the abstract and conclusion. Document naviagtion becomes more efficient with better topic indexation.

### Topic Indexing

Indexing topics is provided to see at a glance which topics are discussed in a paper. Added is an indication of the length of the section discussing the topic. Together with document navigation this makes focussed scouting of documents more convenient and efficient.

### Normalizing Document Formatting

During testing a new potential feature was discussed. Some documents that students come across do not feature sufficient formatting in their eyes. For these documents normalizing the formatting by overwriting the original format can improve readability for students.
Older and longer documents are more likely to fall outside the preferred or even readable category for students.


Jesper M. van den Berg
Improving Document Exploration for Students as Novice Reseachers During their Literature Review
15-03-2023
Strategic Product Design

**Committee** Senthil Chandrasegaran
Dirk Snelders

**Faculty of Industrial Design Engineering**

**Delft University of Technology**