

# Considering Navigational Value of Keywords in the Process of Navigation Leads Selection

Róbert Móra

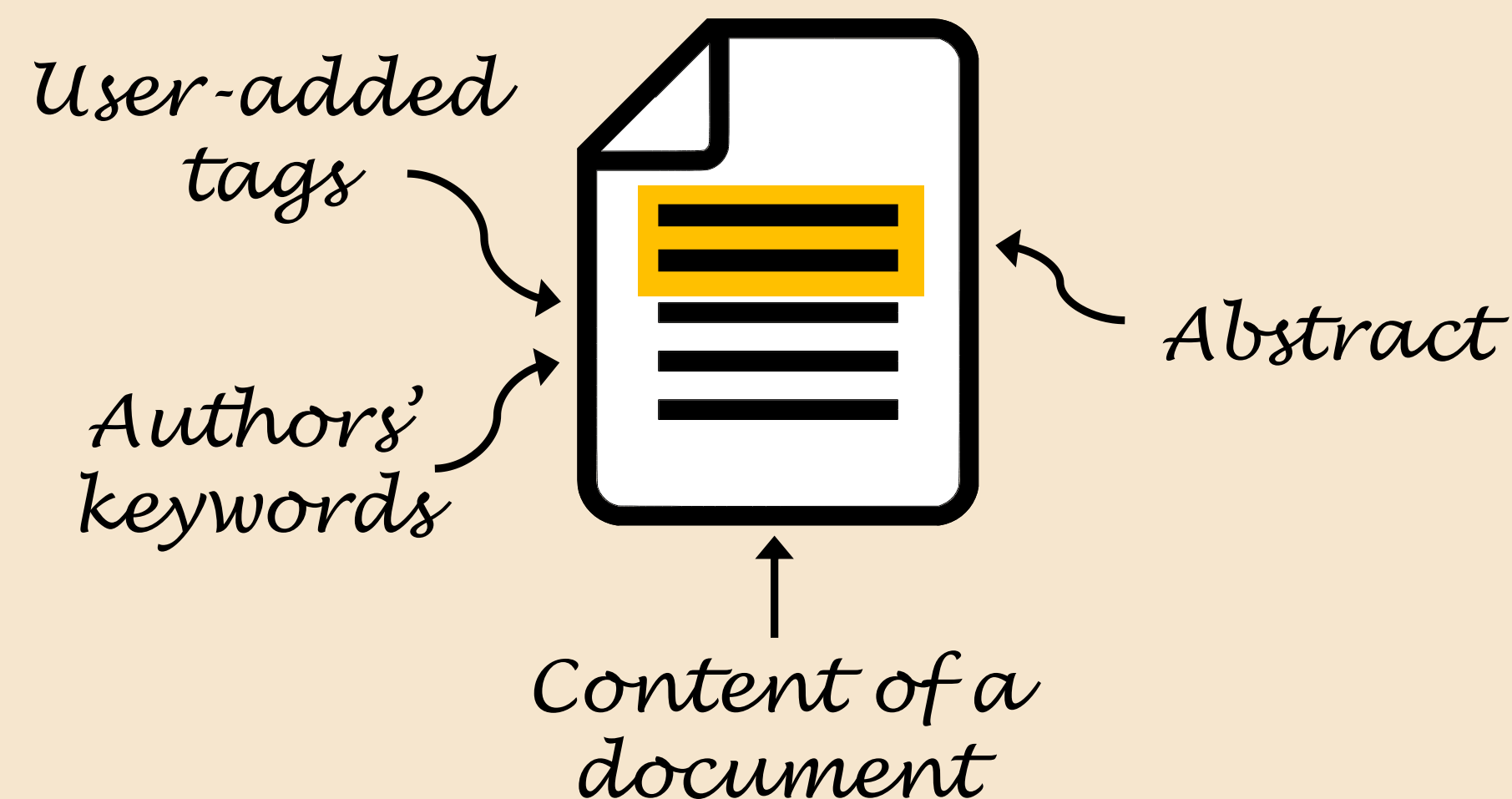
Supervisor: Mária Bielíková

## Exploratory Search in Digital Libraries

- Novice researcher scenario
- Navigation leads
  - important terms automatically extracted from the documents
  - visualized under summary (abstract) of a document
  - filtering of information space

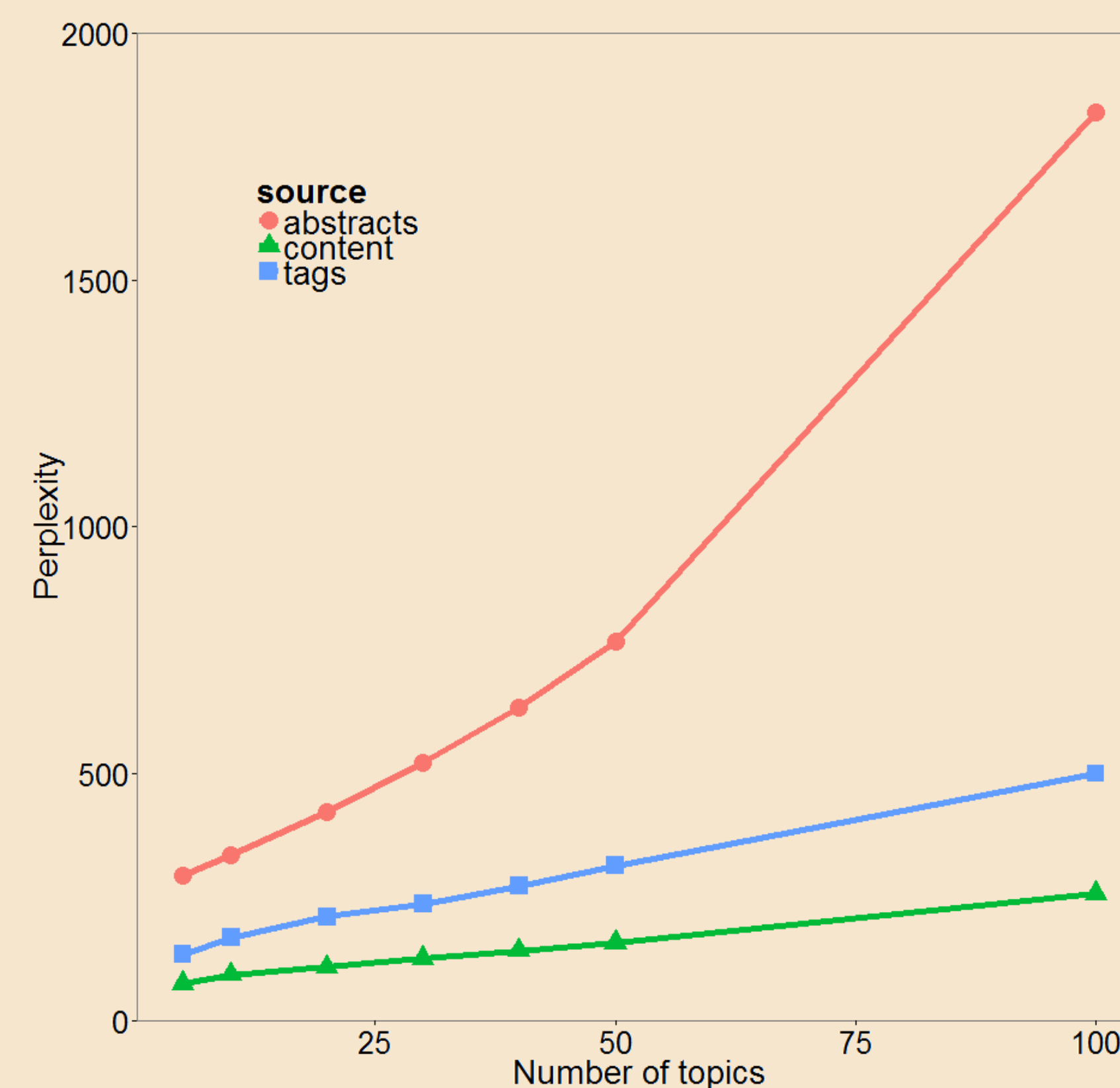
## Selection of Navigation Leads

- Document relevancy  $R_D$ 
  - how the term is relevant for the document
- Navigational value NV
  - how the term represents the information subspace
  - using LDA (Latent Dirichlet Allocation)
- Clustering of the documents
  - documents belong to multiple clusters
  - depth-first exploration
  - exploration of related topics
- Various sources of keywords

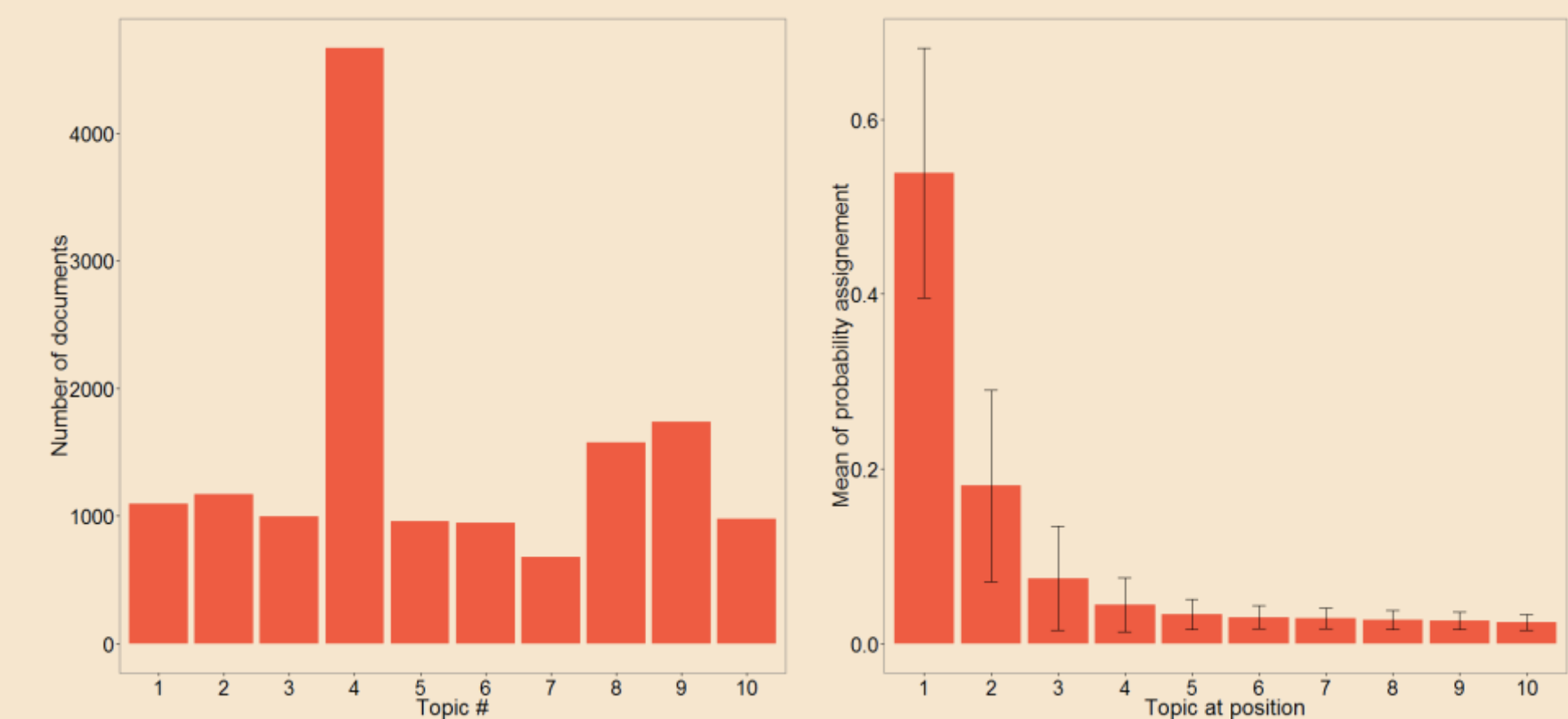


## Evaluation

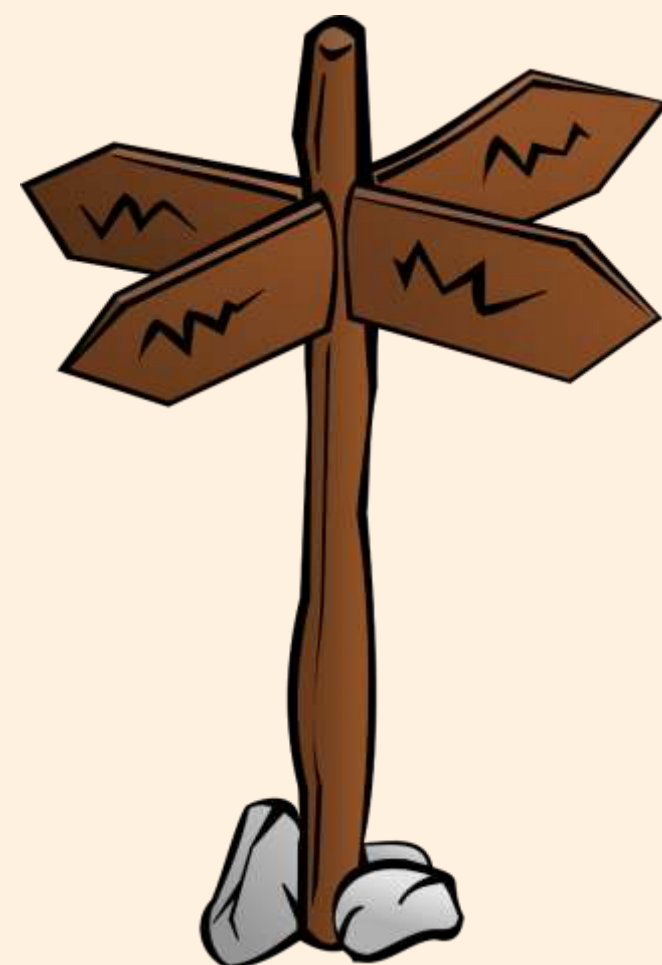
Q1: Are there any differences when using different sources of keywords for LDA? Which ones provide the best results?



Q2: Is the topic modelling employing LDA a suitable approach for identifying the keywords relevant for navigation?



Abstracts		Content + Tags	
Topic #1	Topic #2	Topic #1	Topic #2
recommendation user system social item	model system software use development	recommender system collaborative filtering adult person recommend	social network semantic web ontology social media linked data



## Conclusions

- The resulting clusters cover different topics in the documents
- The terms representing the topic clusters are descriptive and relevant for the topics
- There are significant differences in the quality of the resulting clusters when using different keyword sources
  - content and user-added tags are the best ones
- Next step: a user study in Annota employing the eye-tracking