Evaluating Tag Recommendations for E-Book Annotation Using a Semantic Similarity Metric

Emanuel Lacic*, Dominik Kowald*, Dieter Theiler, Matthias Traub,
Lucky Kuffer, Stefanie Lindstaedt, Elisabeth Lex

{elacic,dkowald,dtheiler,mtraub}@know-center.at,
lucky.kuffer@hgv-online.de, slind@know-center.at, elisabeth.lex@tugraz.at

* Both authors contributed equally to this work.

Summary

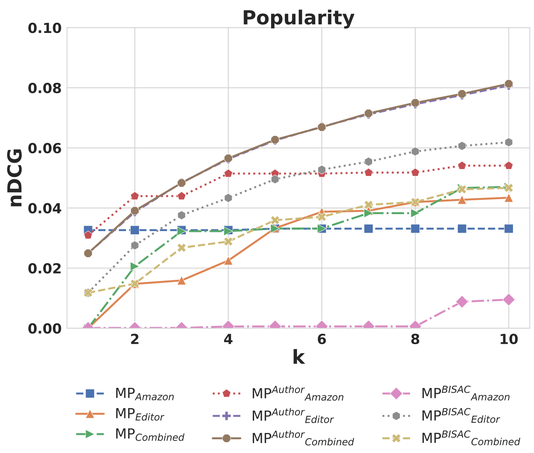
- Provide tag recommendations that **incorporate** both the **vocabulary** of the **editors** and e-book **readers**.
- Tag recommenders can provide **poor accuracy** performance but still **deliver semantically relevant** results.

Data

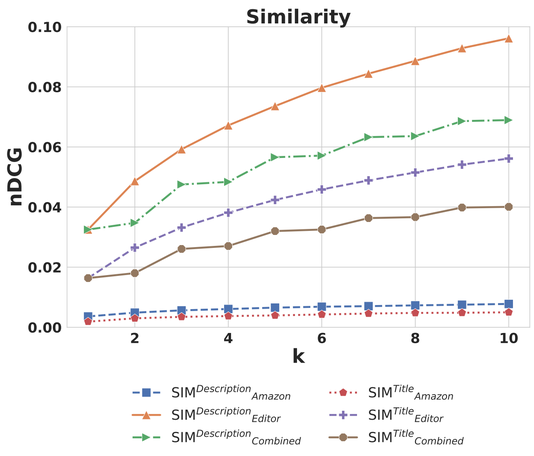
- 13 German publishers, namely:
  - Kunstmann, Delius-Klasnig, VUR, HJR, Diogenes, Campus, Kiwi, Beltz, Chbeck, Rowohlt, Droemer, Fischer and Neopubli
- Amazon search query logs for 12 months, i.e., November 2017 to October 2018

Train Train	#
Editor e-books	48,705
Amazon e-books	21,243
Editor Vocabulary size	114,707
Amazon Vocabulary size	8,240
Test Set	**#**
E-books	2,896
Vocabulary size	33,663

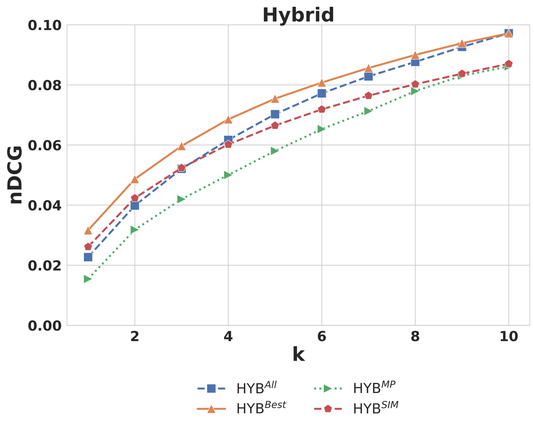
Accuracy Experiment



(a) Popularity-based approaches



(b) Similarity-based approaches

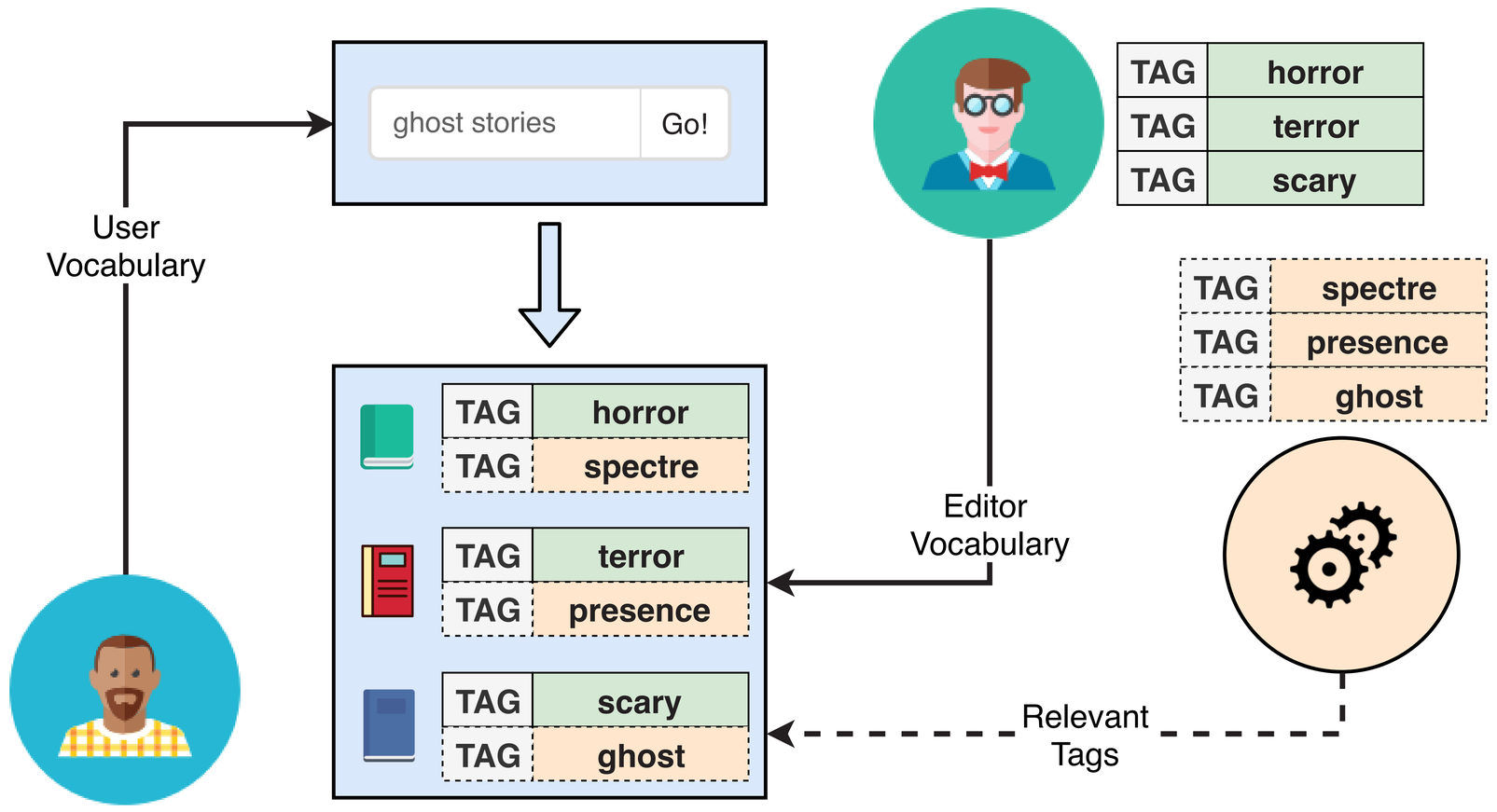


(c) Hybrid approaches

E-Book Annotation

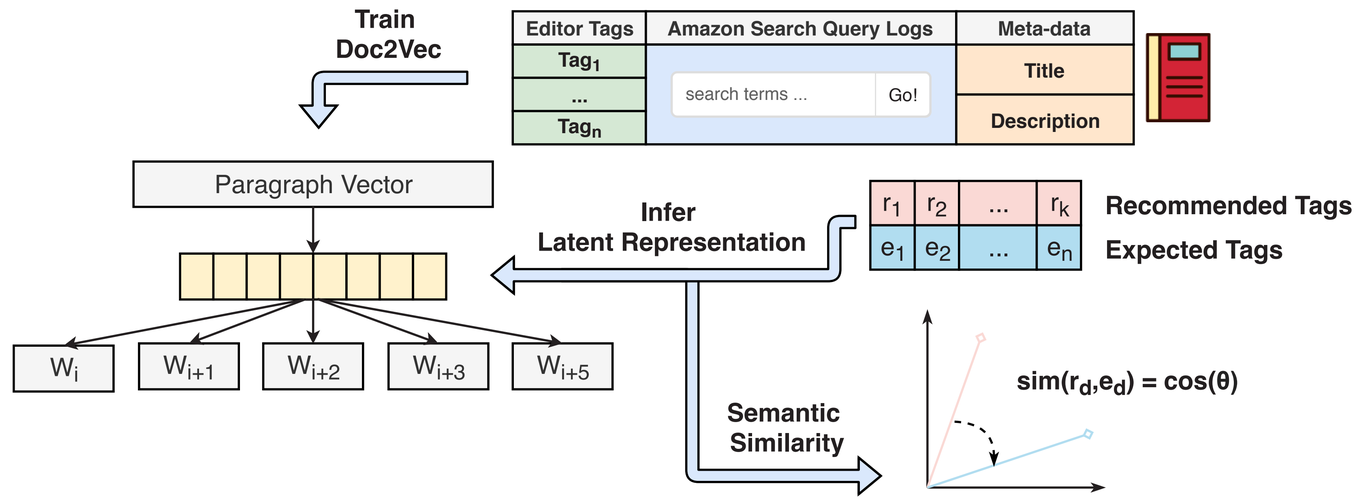
Idea:

- **Support editors** in the e-book annotation process with tag recommendations.
- **Mimic** the **vocabulary of users** in Amazon, who search for e-books.
- **Measure** not only **exact "hits"** of our recommendations, but also **semantic matches**.



Semantic Similarity Metric

Learn the **semantic relationships** from the editor's and user's **vocabulary**. Use it to **compare how semantically similar** the recommended tags are to the expected tags.



Beyond-Accuracy Experiment

The semantic similarity measure helps us **interpret** the recommendation **quality**.

Approaches that **do not provide a high accuracy** could still result in tag recommendations that are **semantically related** at a **high degree**.

Combining both data sources (i.e., **vocabularies**) **enhances the quality** of tag recommendations for annotating e-books.