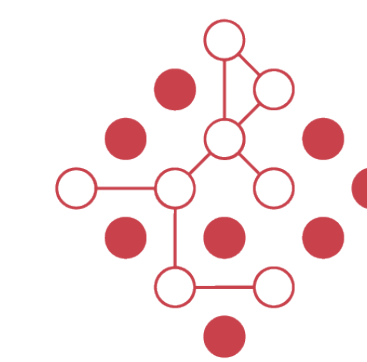


EVALUATING TAG RECOMMENDATIONS FOR E-BOOK ANNOTATION USING A SEMANTIC SIMILARITY METRIC

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

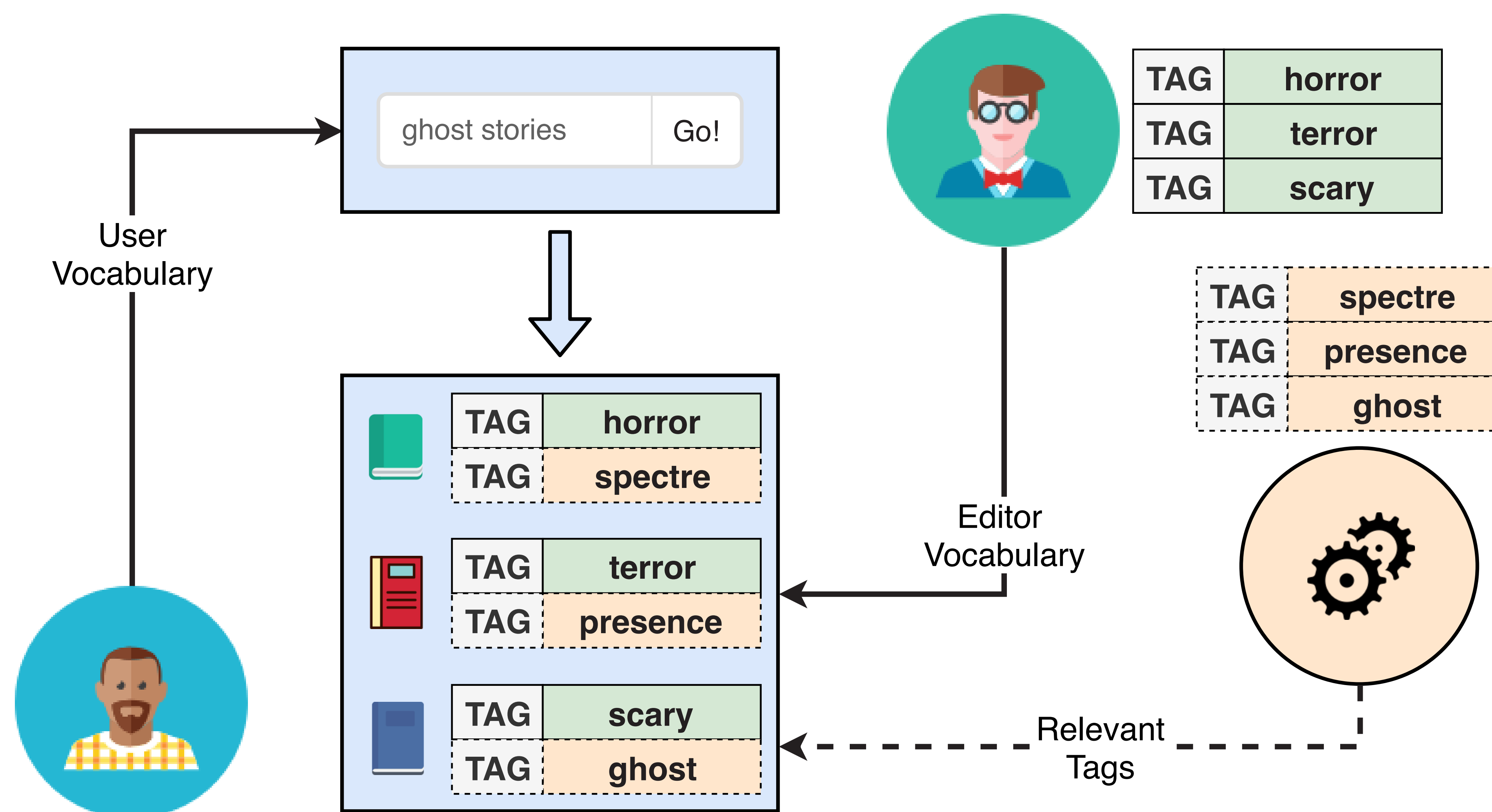
- 13 German publishers, namely:
 - Kunstmann, Delius-Klasnig, VUR, HJR, Diogenes, Campus, Kiwi, Beltz, Chbeck, Rowohlt, Droemer, Fischer and Neopubli
- Amazon search query logs for 12 months, i.e., November 2017 to October 2018

Train	#
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

E-BOOK ANNOTATION

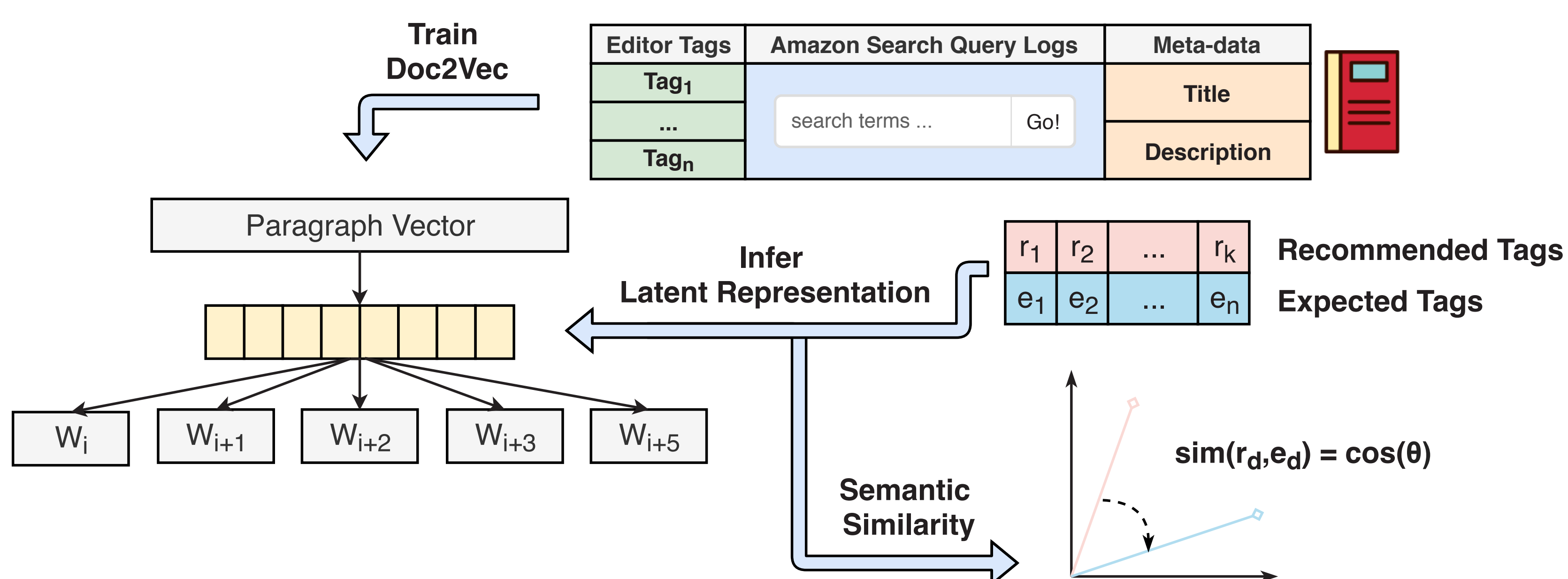
Idea:

- **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**.

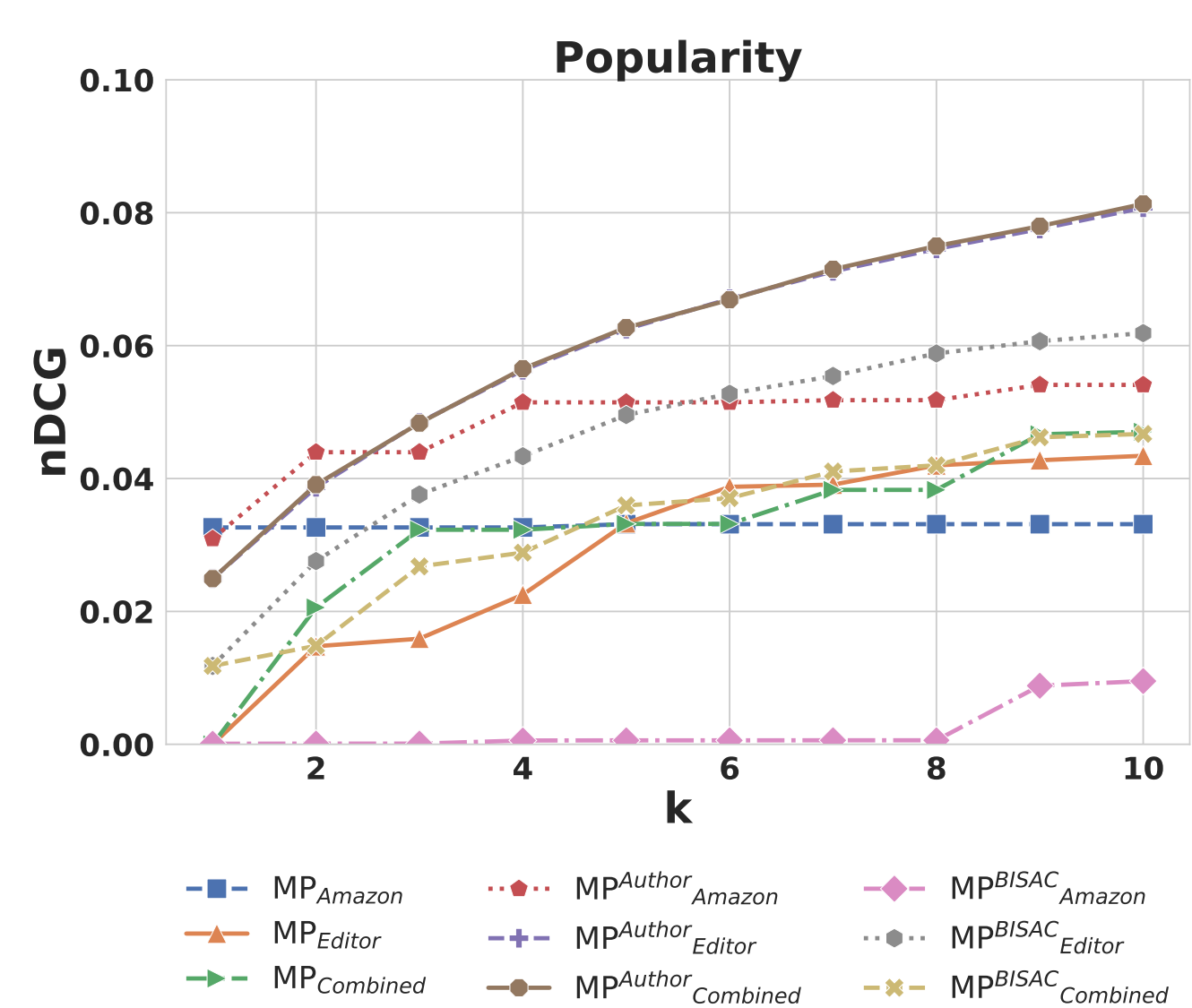


SEMANTIC SIMILARITY METRIC

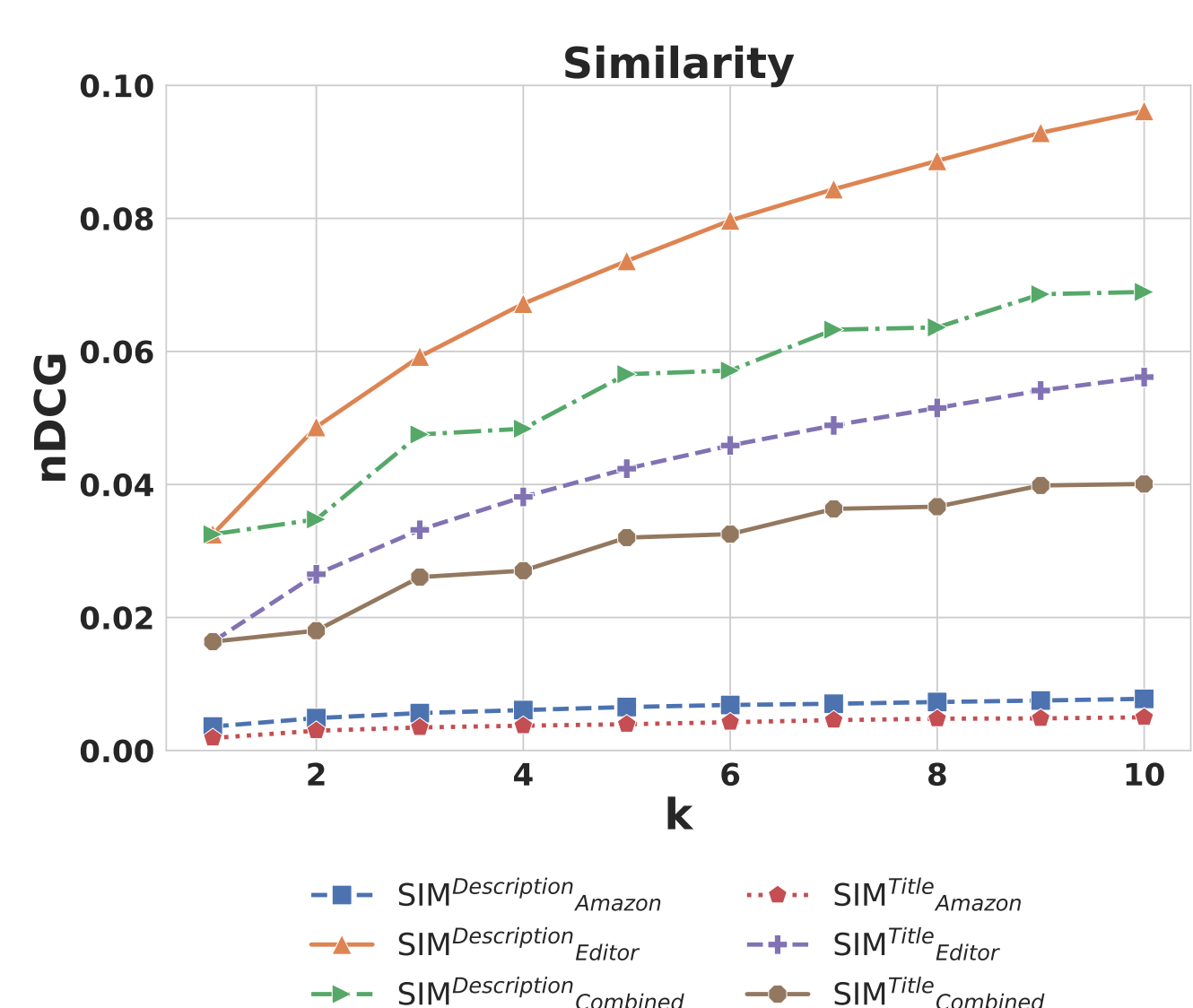
Learn the **semantic relationships** from the editor’s and user’s **vocabulary**. Use it to **compare** how **semantically similar** the recommended tags are to the expected tags.



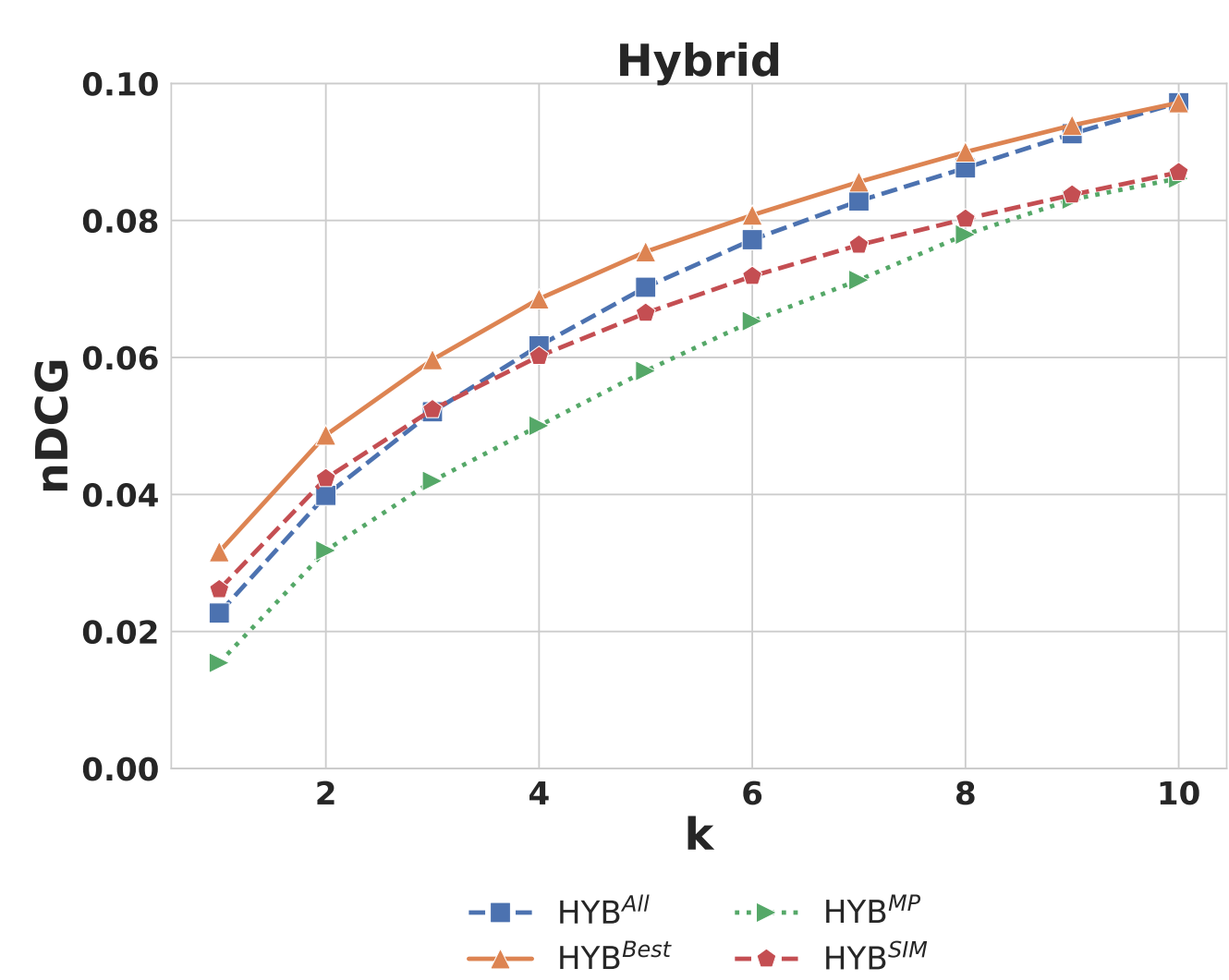
ACCURACY EXPERIMENT



(a) Popularity-based approaches



(b) Similarity-based approaches



(c) Hybrid approaches

BEYOND-ACCURACY EXPERIMENT

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

Approaches that **do not provide a high accuracy**, 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.

