Final Presentation

Jingheng Pan
jingheng.pan@studium.uni-hamburg.de

Adrian Lindloff
adrian.lindloff@studium.uni-hamburg.de

Lijunnan Bai
lijunnan.bai@studium.uni-hamburg.de

Structure

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

1. What is the Adaptive Storyfinder?
2. How did the project go?
3. Core functionality
4. Demo
5. Outlook

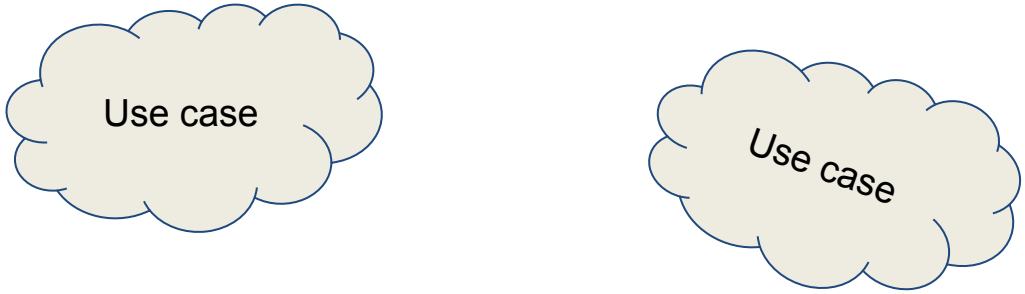
Motivation

Use cases:

- People want to see stuff about things that they like
  → discover interesting new information

- A person would like to read an article and focus on desirable sections
  → made aware of interesting parts

Motivation



=> Goal of the Adaptive Storyfinder:

- Suggest reading material which matches the preferences
  → User preference-oriented ranking of search results

- Find interesting parts of a website according to user preferences

“Adaptive Storyfinder”

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

What is the Adaptive Storyfinder?

- A System which leverages user preferences to fit the users needs
- ”Adaptive” Storyfinder assists in finding interesting content

Why “adaptive”?

- Different users get different answers tailored to their interests

How did the project go?

Getting Started

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

- A lot of struggles on finding approaches on what exactly to do

- What do we want to implement? Where do we want to take the idea?

  → ~~Behavioral data~~

  → ~~Eyetracking~~

=> User-related content to determine user preferences

Technology

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Backend:

- FastAPI
- Docker
- PyTorch for ML + HuggingFace

Frontend:

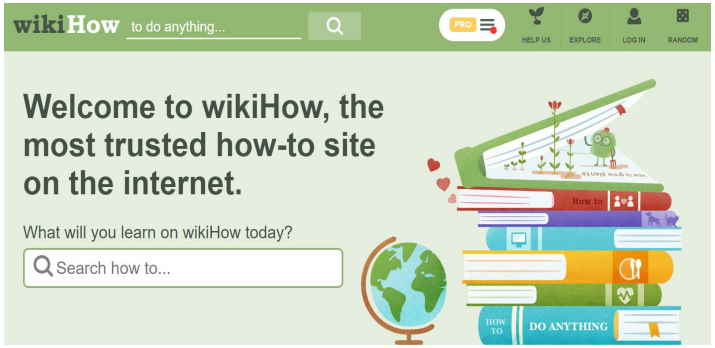
- Vue.js with mainly Vuetify & Bootstrap

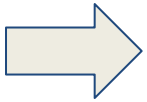
Article Database

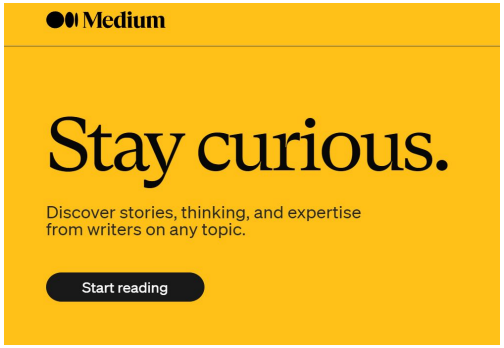
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Article database:

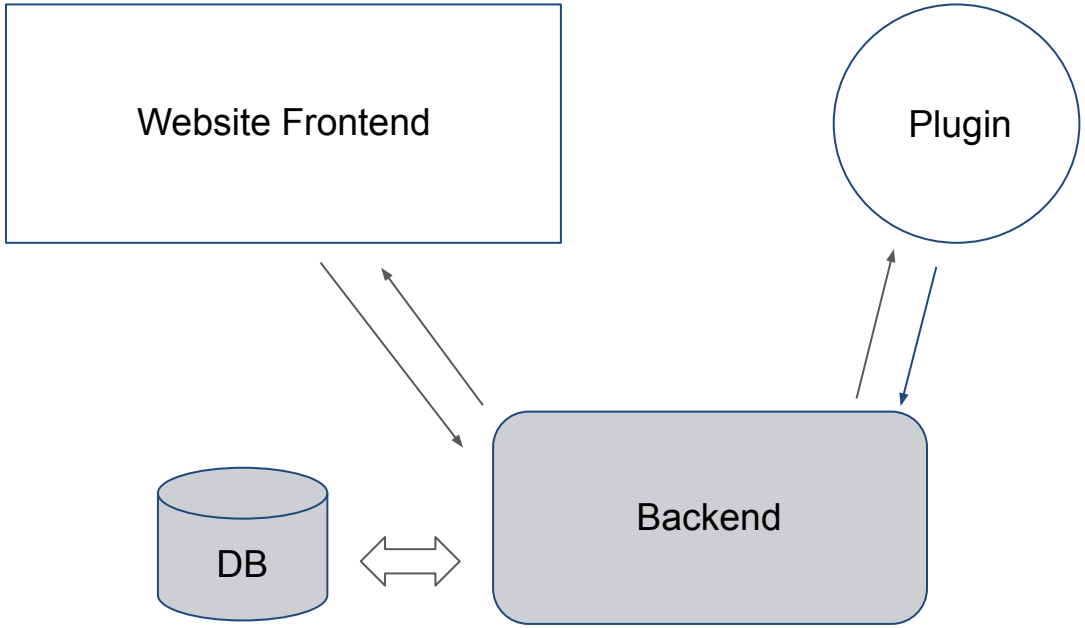
- From WikiHow API & dataset to a medium dataset







Architecture



Backend deals with:
- User Management
- User Histories

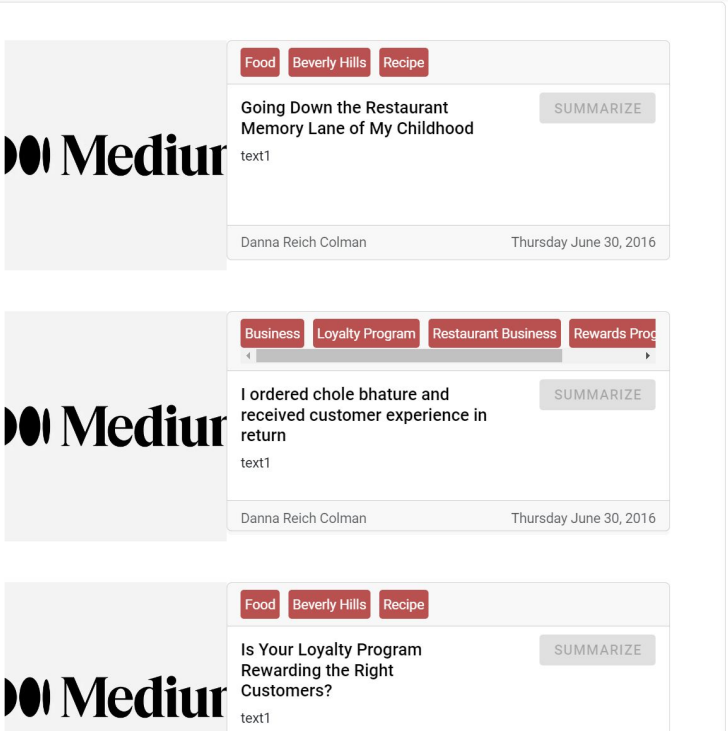
For the Website Frontend:
- Re-ranking
- Topic recommendation

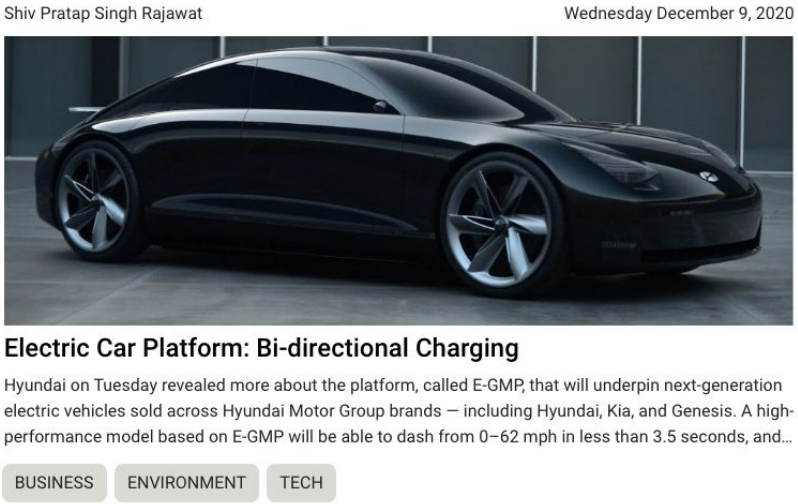
For the plugin:
- Highlighting of paragraphs

Core functionality

- Search feature
- History feature
- Topic feature
- Plugin

Search feature



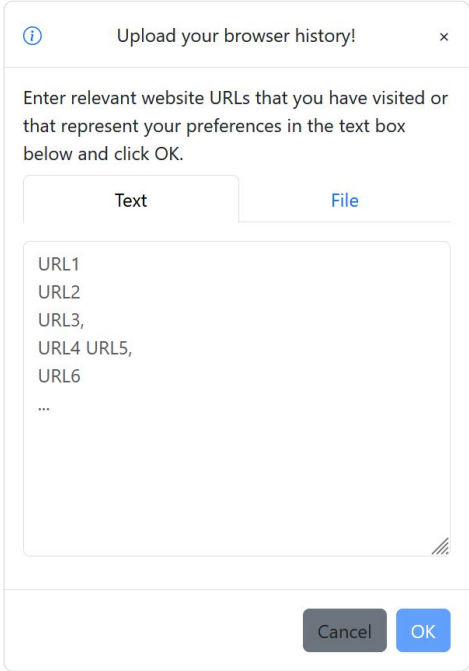


History feature

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

- A browser history partly contains a users interests

  → can be leveraged to enhance the pursuit of interesting new articles



- Users can specify their favorite websites
- Users can upload their favorite websites
- Users can automatically send their chrome history

History feature

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

- Management of user histories

Topic feature

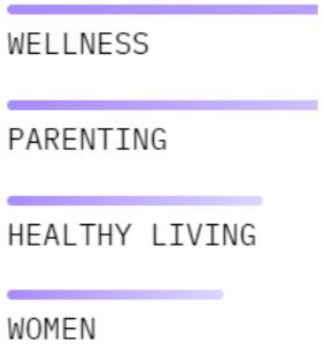
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Inspiration: Predefined tags in the original dataset do not scale efficiently

Topic model:

- Using Bert-based pre-trained model

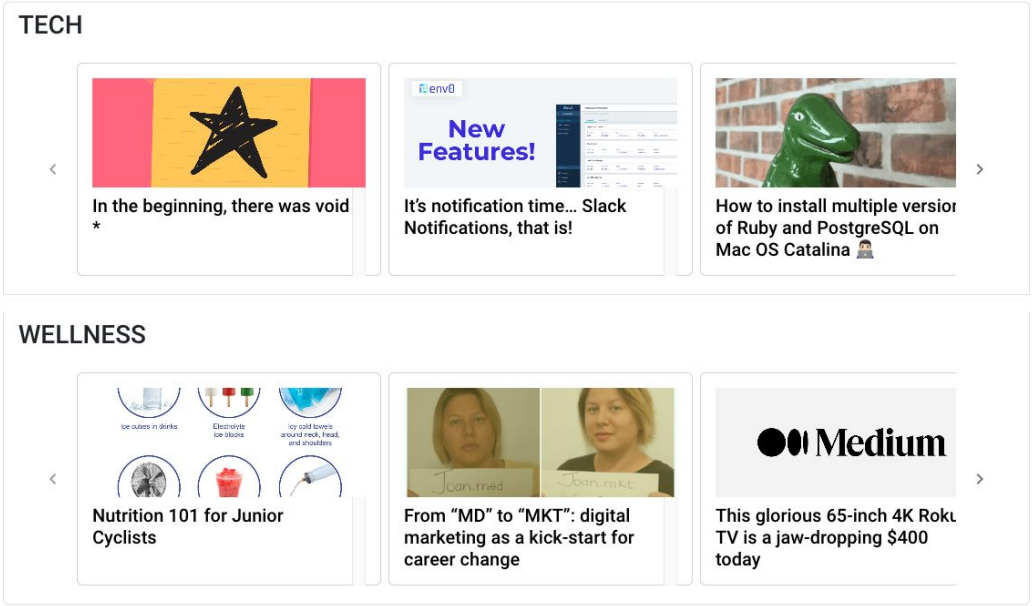
  → News-category-classification (42 categories)



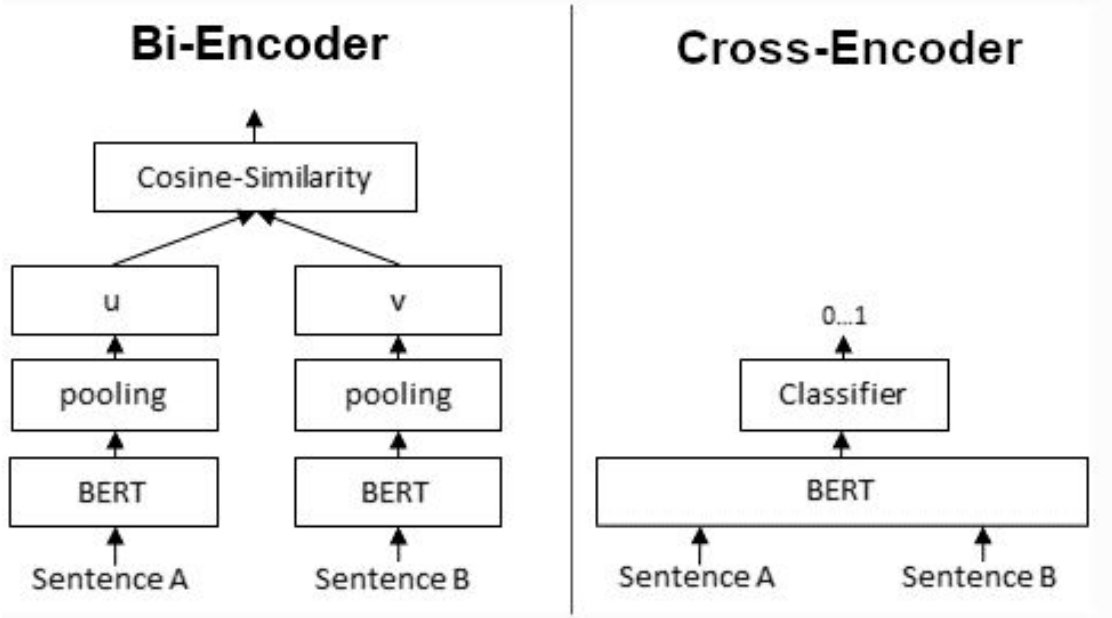
- Applied to our dataset → Categorize websites consistently
- Reliable way to capture interest groups

  → Used to recommend new articles & to aid in the re-ranking process

Topic feature



Sentence-transformers



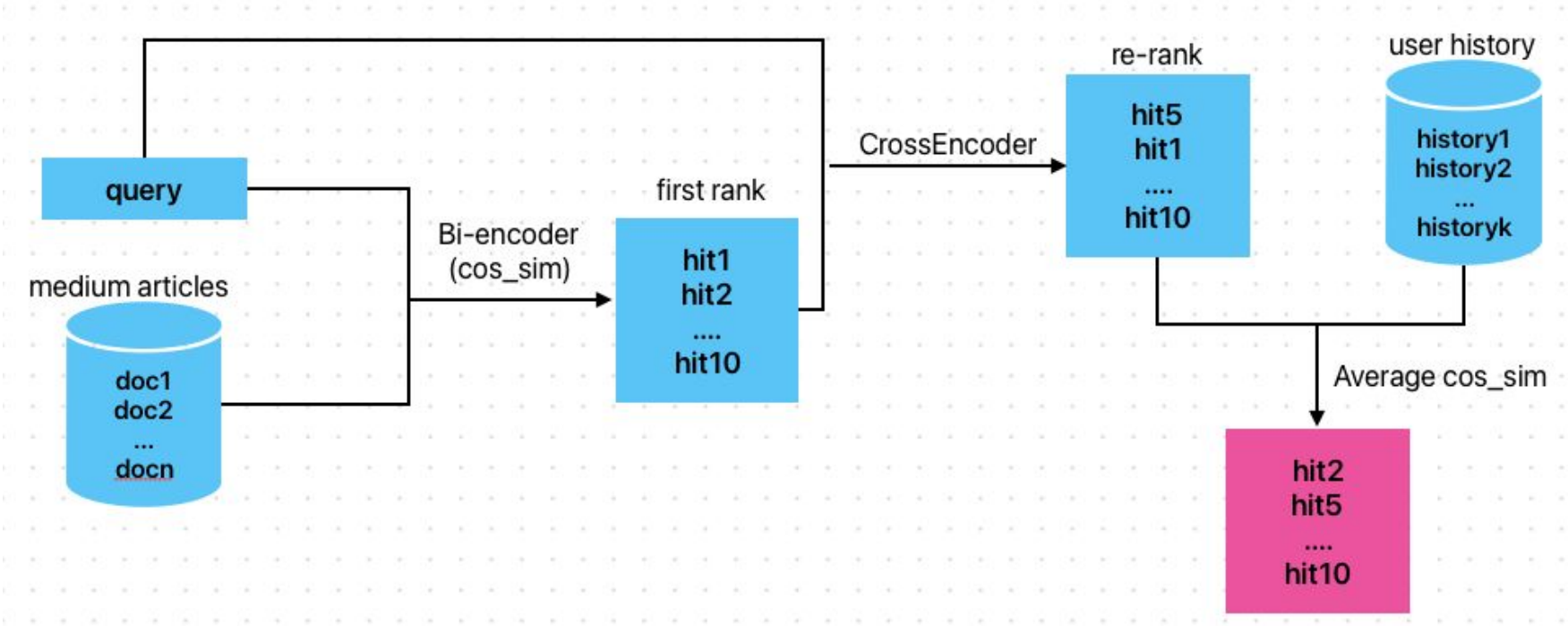
Bi-encoder:

- Pass sentence independently
- Output: embeddings

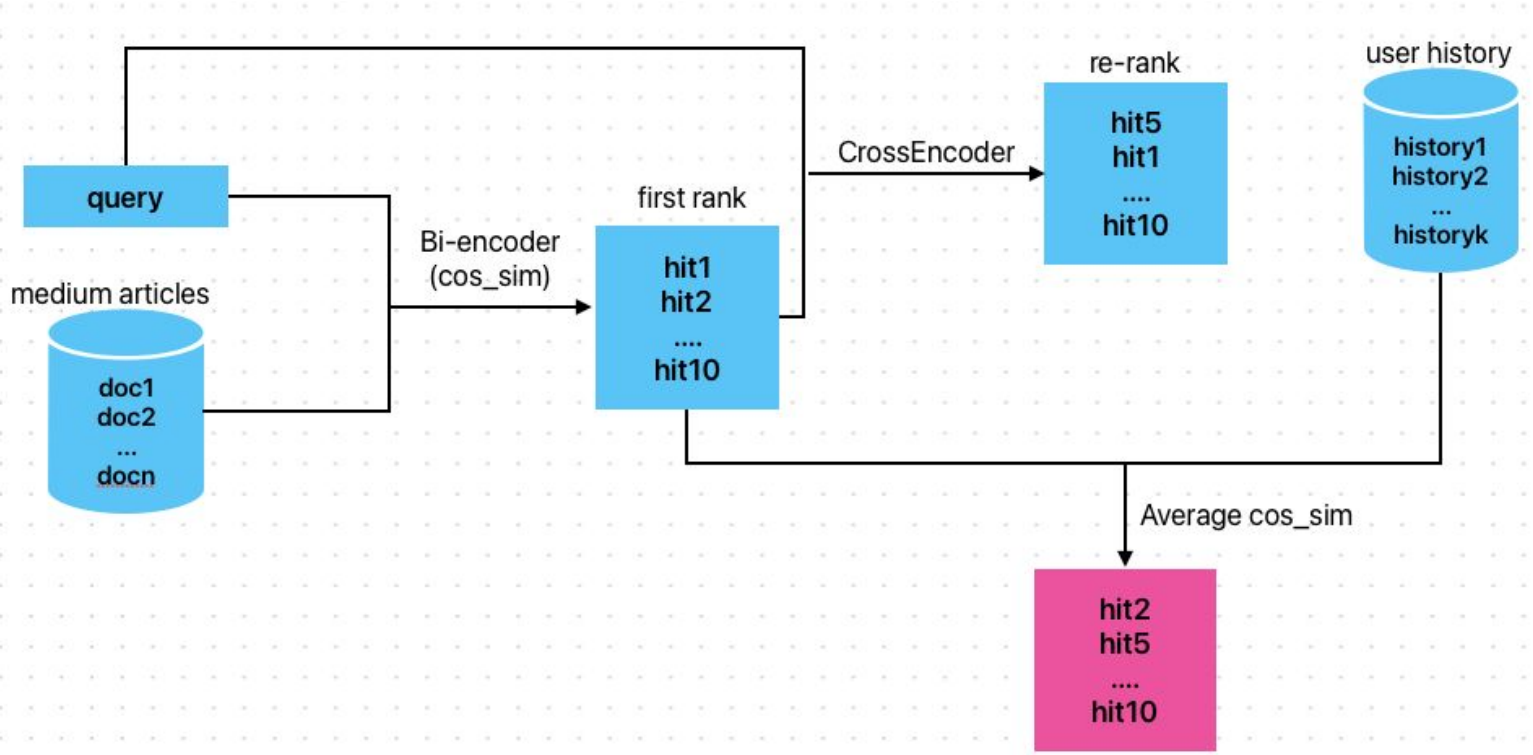
Cross-encoder:

- Pass sentence simultaneously
- Output: possibility of the similarity of input sentences

Ideally previous schema



Real previous schema



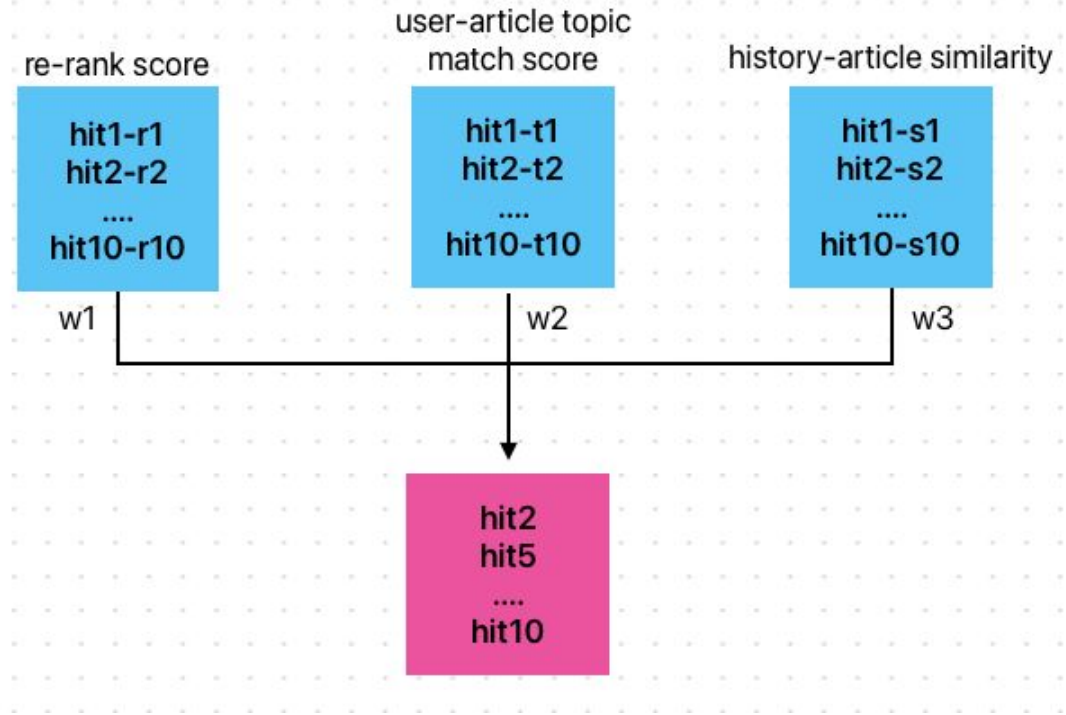
Search Pipeline

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

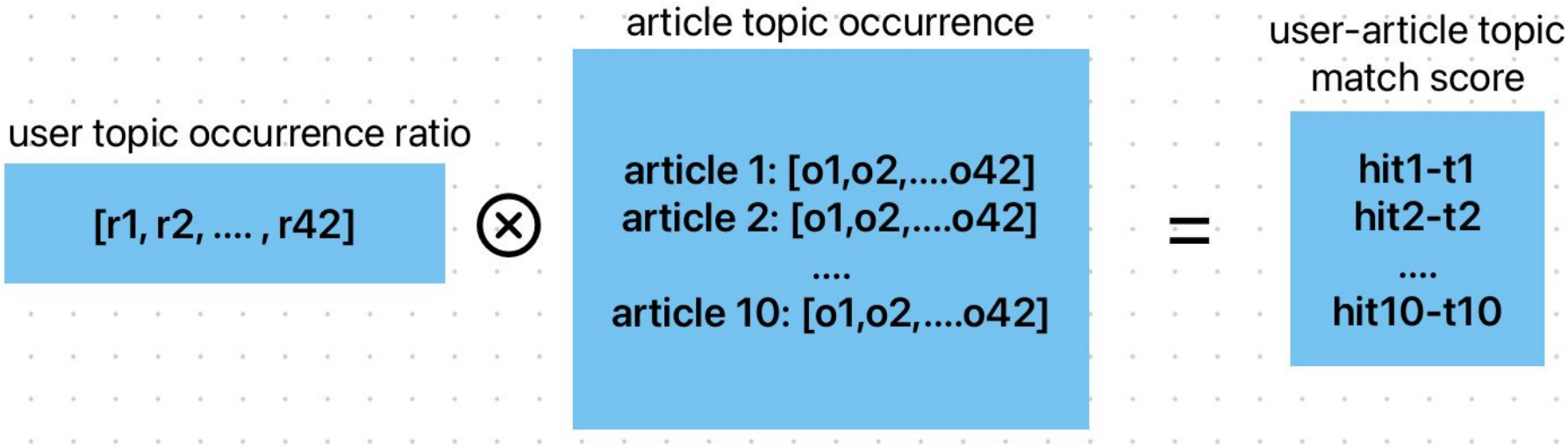
Old pipeline:

- Query search: Re-ranker for query retrieval
- Preference search: Cosine-similarity between history and candidates
- Preference search has nothing to do with the results from query search

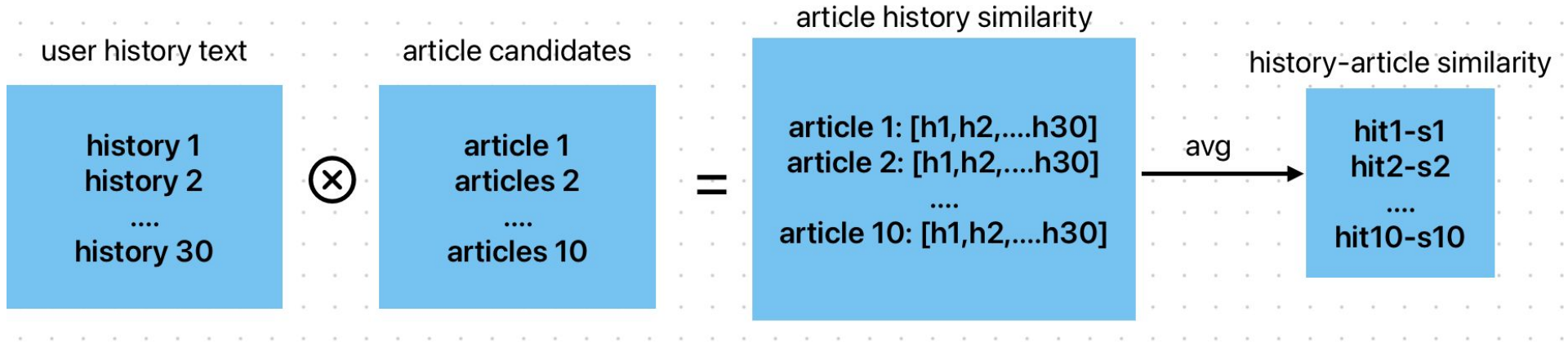
Optimized schema



User-article topic match score



history-article similarity



Search Pipeline

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Old pipeline:

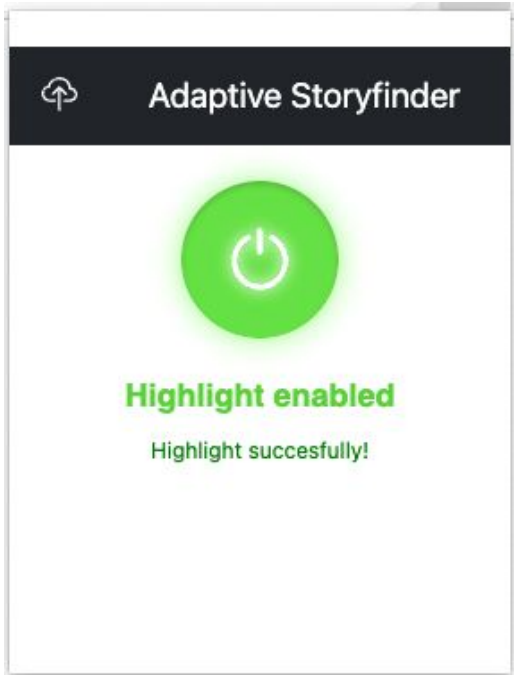
- Query search: Re-ranker for query retrieval
- Preference search: Cosine-similarity between history and candidates
- Preference search has nothing to do with the results from query search

New pipeline:

- Query search: Re-ranker for query retrieval
- Preference search: consider **re-rank score, history similarity and topic similarity score**

Plugin

- Highlight the paragraphs
  - Paragraphs that might be interesting to user
  - According to history-paragraphs cos-similarity
- Automatically upload browser history
- Shares User Authentication with the Website



Demo

Limitation

- Pipeline
  - Scale of the dataset
  - Re-Rank weights trade-offs
  - User-article topic match score is less convincing due to topic sparsity of articles
- Topic Model
  - Not very accurate to assign a topic
- Plugin
  - Paragraphs-article similarity is insufficient to suggest interesting paragraphs to user

Outlook

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

- Include other sources of information beyond the medium dataset

  → expand dataset with more sources

- Generate AI Images for articles

- Implement user/ data privacy or at least inform the users about the scope of data usage

Any Questions?