Friedrich-Schiller-Universität Jena · Bauhaus-Universität Weimar · UNI A Universität Augsburg University · Technology Arts Sciences TH Köln · UNI KASSEL VERSITÄT

🏆 Information Retrieval Shared Task 🏆 Guidelines

A Dataset and a Search Engine for the IR and ACL Anthologies

Version 1.0

In this shared task, we will create an IR experimentation dataset for querying two repositories of scientific publications in the areas of information retrieval, natural language processing, and computational linguistics: the IR anthology and the ACL anthology. We will also create information retrieval systems for retrieving relevant documents for queries, and evaluate them on a blind test set. We will have prices for a range of categories: best systems for particular categories, best software engineering, best documentation, and best analysis. But do not forget: ***you all win*** *a great deal of experience with information retrieval research!*

Part 1: Topic Construction

In an Information Retrieval (IR) dataset, a *topic* typically refers to a subject or theme around which information is organized or retrieved. It represent an information need that a user might have when searching the dataset. In our case, a topic is expressed using a single *query*. In our shared task, each student/team should create 1-3 topics and submit them in the form of a file called `topics.xml` with the following structure.

ℹ Ask your lecturer / tutors about your university's concrete assignment and mode of submission.

```
<topics>
<topic number = "1">
    <title>stemming for arabic languages</title>
    <description>Which papers focus on improving stemming in arabic languages?
    </description>
    <narrative>Relevant papers include research on stemming methods for arabic
    languages or how to improve those methods. Papers that focus on stemming in
    other languages are not relevant, as well as papers that do not focus on
    stemming. </narrative>
</topic>
...
</topics>
```

When constructing your topic(s), imagine that you are searching literature for a term paper on a scientific topic related to information retrieval or natural language processing. After you have picked a subject for your imaginative term paper (don't worry, you won't have to write it, at least not in this class :)), think about how you would query the IR and ACL anthologies for this topic. Go to the respective website and try out your queries. Read a bit about these topics if you are not already familiar with them, and refine your information need. The more specific your information needs, the more interesting our dataset will be!

- The `title` is the string that is going be used as the query, i.e., it should contain a brief keyword-style query.
- The `description` field contains a natural language question expressing the same information need.
- The `description` field contains a detailed instruction which types of papers should be considered relevant/nonrelevant to this topic. Try to be as specific as possible when detailing the information need.

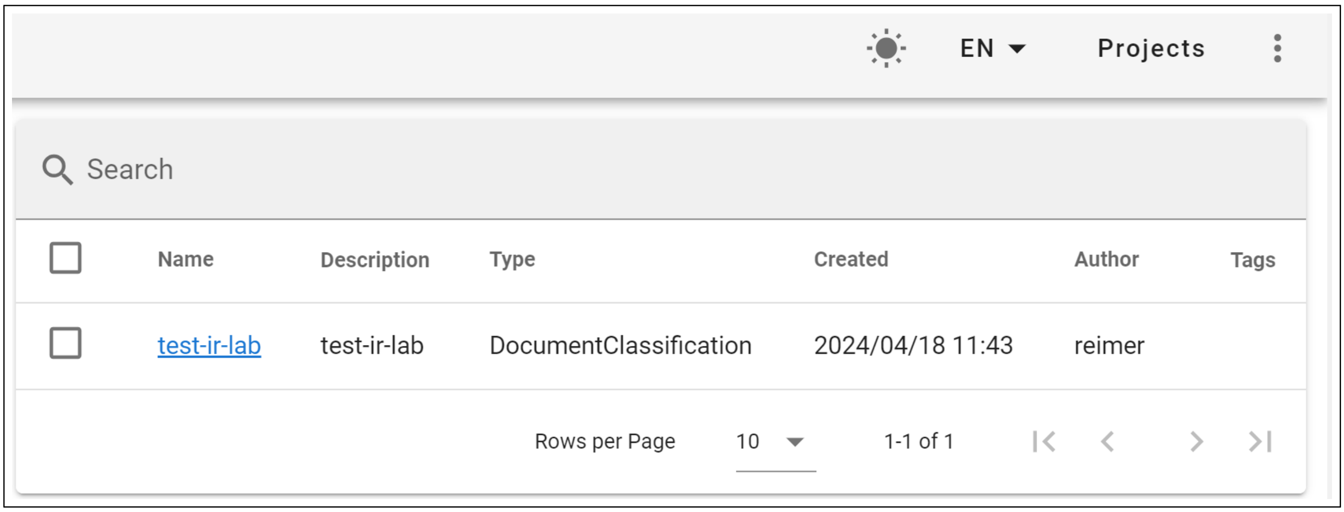
Part 2: Relevance Labeling

For each topic/query, the results of a baseline retrieval system consisting of the combination of a few well-established ranking methods will be provided by the University of Jena. For labeling documents with regard to whether they are relevant to a query, we use the web-based annotation system doccano [1]. The set of documents retrieved by the baseline system for each topic/query has been uploaded to this system for you.

Log in to the doccano installation at: https://doccano.web.webis.de/auth

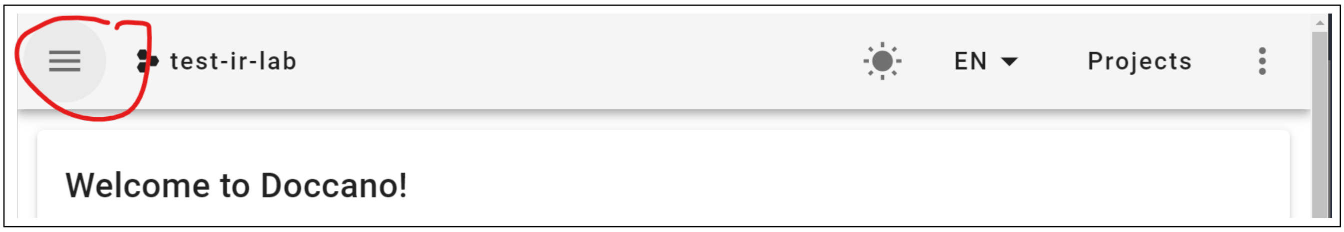
ℹ Your login details will be provided by your lecturer / tutor.

After logging in, you can see an overview of the projects that are assigned to you:



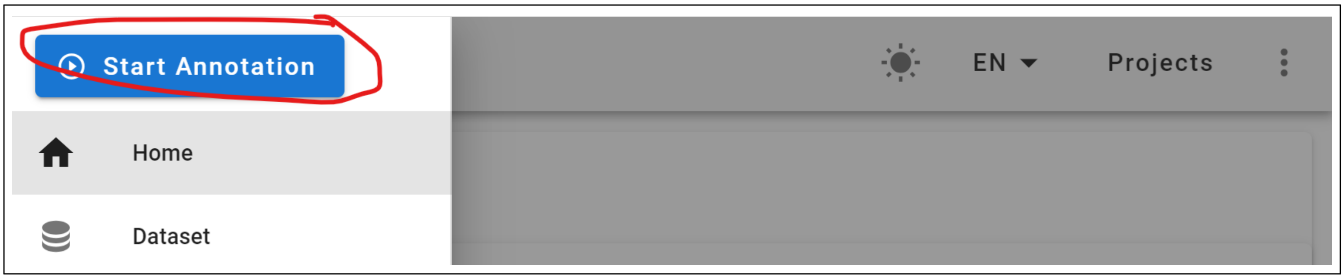
□	Name	Description	Type	Created	Author	Tags
□	test-ir-lab	test-ir-lab	DocumentClassification	2024/04/18 11:43	reimer	

Select a project. Next, if the side menu is not already expanded, click on the menu button.

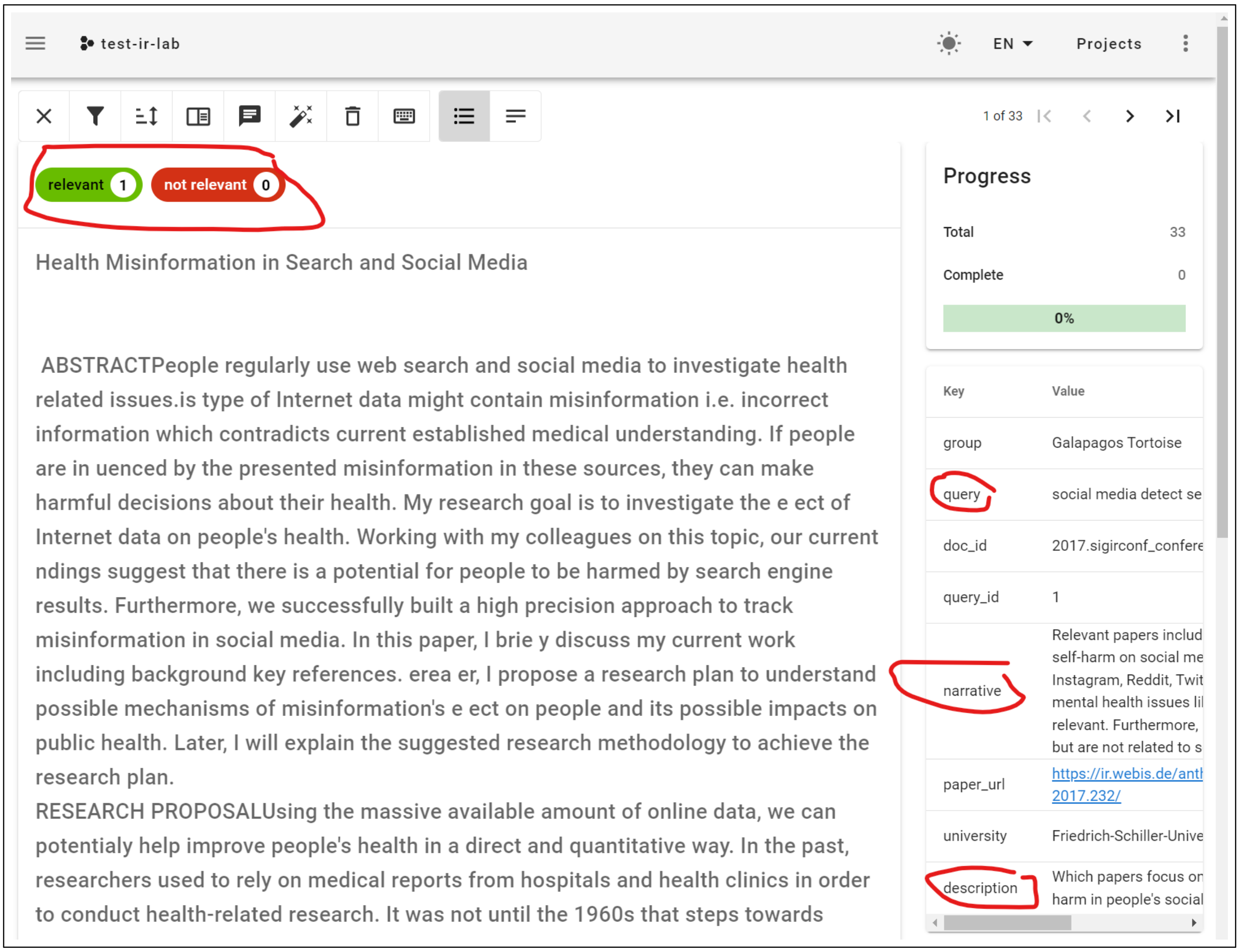


Welcome to Doccano!

Click on the blue “Start Annotation” button.



The first document will be displayed. On the right-hand side, you can see the query, narrative, and description. Your task is to judge whether the displayed document is relevant to this information need or not. On the left-hand side, you can see the abstract of the paper. Read it carefully. If in doubt, go to the full paper (the link is provided via `paper_url`). After determining whether or not a paper is relevant to the topic, click on one of the buttons at the top. You can then navigate to the next topic using the arrow buttons at the top right-hand side of doccano.



Health Misinformation in Search and Social Media

ABSTRACTPeople regularly use web search and social media to investigate health related issues.This type of Internet data might contain misinformation i.e. incorrect information which contradicts current established medical understanding. If people are influenced by the presented misinformation in these sources, they can make harmful decisions about their health. My research goal is to investigate the effect of Internet data on people's health. Working with my colleagues on this topic, our current findings suggest that there is a potential for people to be harmed by search engine results. Furthermore, we successfully built a high precision approach to track misinformation in social media. In this paper, I briefly discuss my current work including background key references. Thereafter, I propose a research plan to understand possible mechanisms of misinformation's effect on people and its possible impacts on public health. Later, I will explain the suggested research methodology to achieve the research plan.
RESEARCH PROPOSALUsing the massive available amount of online data, we can potentialy help improve people's health in a direct and quantitative way. In the past, researchers used to rely on medical reports from hospitals and health clinics in order to conduct health-related research. It was not until the 1960s that steps towards

Progress

Total 33

Complete 0

0%

Key	Value
group	Galapagos Tortoise
query	social media detect se
doc_id	2017.sigirconf_confere
query_id	1
narrative	Relevant papers includ self-harm on social me Instagram, Reddit, Twit mental health issues lil relevant. Furthermore, but are not related to s
paper_url	https://ir.webis.de/antl 2017.232/
university	Friedrich-Schiller-Unive
description	Which papers focus on harm in people's social

References

[1] Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. doccano: Text annotation tool for human, 2018. URL https://github.com/doccano/doccano. Software available from https://github.com/doccano/doccano.