Information Retrieval Lab

Welcome to week 2 of the Information Retrieval class 2023

© Gienapp/Elstner/Fröbe/Scells 2023

Agenda

1. Organization

2. Project Info: Milestone 1

3. Hands-on Tutorial: How to do last weeks tutorial on TIRA

© Gienapp/Elstner/Fröbe/Scells 2023

Organization

Did you find groups?
If not -> EMAIL US

- Theresa: theresa.elstner@uni-leipzig.de
- Harry: harry.scells@uni-leipzig.de

© Gienapp/Elstner/Fröbe/Scells 2023

Project Info: Milestone 1 – Data

General

You will create a domain-specific search engine for scientific papers on Information Retrieval research.

Overall Goal:

- Process a raw dataset that you will use as the basis for your domain-specific information retrieval system
- Dataset: “ IR-Anthology” contains abstracts of scientific papers on information retrieval over the past decades.
  `https://ir.webis.de/anthology/`
- Download dataset from here
  `https://files.webis.de/teaching/ir-ss23/`

© Gienapp/Elstner/Fröbe/Scells 2023

Tangent

IR Evaluation

© Gienapp/Elstner/Fröbe/Scells 2023

Project Info: Milestone 1 – Data

What to do

- Process the documents in the raw “IR-Anthology” dataset into format compatible with Milestone 2
  - i.e., `ir_datasets`
- Create *topics* that represent several information needs
  - for inspiration, see content from lecture slides

© Gienapp/Elstner/Fröbe/Scells 2023

Project Info: Milestone 1 – Data

What to do

The processed dataset will consist of:

1. the document collection in `.jsonl`-format, a form consistent with `ir_datasets`, e.g., like so

```
{"doc_id": "0001", "text": "How quickly daft jumping zebras vex."}
{"doc_id": "0002", "text": "Quick fox jumps nightly above wizard."}
{"doc_id": "0003", "text": "The jay, pig, fox, zebra and my wolves quack!"}
```

2. your custom topics for your dataset in TREC XML-format, e.g., like so:

```
<topics>
  <topic number="1">
    <title>fox jumps above animal</title>
    <description>What pangrams have a fox jumping above some animal?</description>
    <narrative>Relevant pangrams have a fox jumping over an animal (e.g., an dog). Pangrams
        containing a fox that is not jumping or jumps over something that is not an animal are
         not relevant.</narrative>
  </topic>
  <topic number="2">
    <title>multiple animals including a zebra</title>
    <description>Which pangrams have multiple animals where one of the animals is a zebra?</
        description>
    <narrative>Relevant pangrams have at least two animals, one of the animals must be a zebra
        . Pangrams containing only a zebra are not relevant.</narrative>
  </topic>
</topics>
```


©Gienapp/Elstner/Fröbe/Scells 2023

Project Info: Milestone 1 – Data

What to do

- Register your processed dataset into `ir_datasets` (like in last week's tutorial)
- You must register it using the name **`ironthology-<team>`**
  - Where **`<team>`** is your team name **in TIRA**

Reminder: check out further resources on the course webpage (python, jupyter, commandline tutorial)

© Gienapp/Elstner/Fröbe/Scells 2023

Project Info: Milestone 1 – Data

Submission in TIRA

- You will receive two invite links
  - One link to invite you to your group in TIRA
  - One link to invite you to the TIRA submission page

© Gienapp/Elstner/Fröbe/Scells 2023

Project Info: Milestone 1 – Data

What to hand in

Hand in one jupyter notebook per group that performs these steps on TIRA

- We have created a template notebook with all cell types that must be contained here `https://temir.org/teaching/information-retrieval-ss23/template-notebook.ipynb`
- Your jupyter notebook will produce the input for Milestone 2
- We show you how to do this in the first two tutorials

DEADLINE FOR THIS NOTEBOOK 02.05.23 (approx. 3 weeks from now)

© Gienapp/Elstner/Fröbe/Scells 2023

Hands-on Tutorial: How to do last weeks tutorial on TIRA

TIRA tutorial

- https://www.tira.io/t/how-to-organize-a-ir-task-with-ir-datasets/1444
- Follow these steps now, in class, or in your own time.

Did you find groups?
If not -> EMAIL US

- Theresa: theresa.elstner@uni-leipzig.de
- Harry: harry.scells@uni-leipzig.de

© Gienapp/Elstner/Fröbe/Scells 2023

SHARKI Survey

- https://umfrage.uni-leipzig.de/index.php/715264?lang=en
- https://umfrage.uni-leipzig.de/index.php/715264?lang=de

© Gienapp/Elstner/Fröbe/Scells 2023