

Information Retrieval Lab

Welcome to the Information Retrieval Class 2023

Agenda

1. Organization
2. Project Info
3. Hands-on Tutorial

Organization

Communication

- ❑ Slides, Announcements & Materials will be available at
`https://temir.org/teaching/information-retrieval-ss23/information-retrieval-ss23.html`
- ❑ Main Communication channel is email
 - Official announcements via Mail (check your student mails regularly!)
 - Optional Discord for group communication ¹

¹server “*webis-lectures*”, mail us for access code

Organization

Grading

Two assessments:

- ❑ Project: **Deadlines: 02.05.23, 06.06.23, 27.06.23**
- ❑ Exam: **Date: TBA**

Project and Exam are mandatory and worth 50% each.

- ❑ to receive a final grade, must receive a passing grade in both.

Organization

Project

You will create a domain-specific search engine for a collection of information retrieval publications, collected over the past decades, called the IR Anthology.

Organization

Project

- ❑ Project contents:
 - Building and evaluating a domain-specific information retrieval system
 - Data handling, indexing, implementation of suitable retrieval models, evaluation of search quality
- ❑ Three times throughout the term (for each milestone) you will hand in a jupyter notebook containing
 - Code
 - For each code snippet a description of what the code does in text form
 - A reflection on your process

The project consists of 3 Milestones:

- ❑ Data (Deadline: 02.05.23)
- ❑ Methods (Deadline: 06.06.23)
- ❑ Analysis (Deadline: 27.06.23)

Organization

Preliminary Term Overview + Milestones

Milestones: 1) Data, 2) Methods, 3) Analysis

- ❑ Tutorials: Docker, Tira 1 [week 1: 04.04.23]
Homework: repeat tutorials + (python + jupyter notebook tutorials if need be)
- ❑ Tutorial: Tira 2, Milestone 1) Data [week 2: 11.04.23]
Homework: work on Milestone 1) Data
- ❑ Q&A [~ week 3: 18.04.23]
Homework: work on Milestone 1) Data
- ❑ Tutorial: Milestone 2) Methods [~ week 5: 02.05.23]
DEADLINE Milestone 1) Data, Homework: work on Milestone 2) Methods
- ❑ Q&A [~ week 6: 09.05.23]
Homework: work on Milestone 2) Methods
- ❑ Tutorial: Milestone 3) Analysis [~ week 10: 06.06.23]
DEADLINE Milestone 2) Methods, Homework: work on Milestone 3) Analysis
- ❑ Q&A [~ week 11: 13.06.23]
Homework: work on Milestone 3) Analysis
- ❑ Q&A [~ week 13: 27.06.23]
DEADLINE Milestone 3) Analysis

Organization

Outline of Todays Tutorial

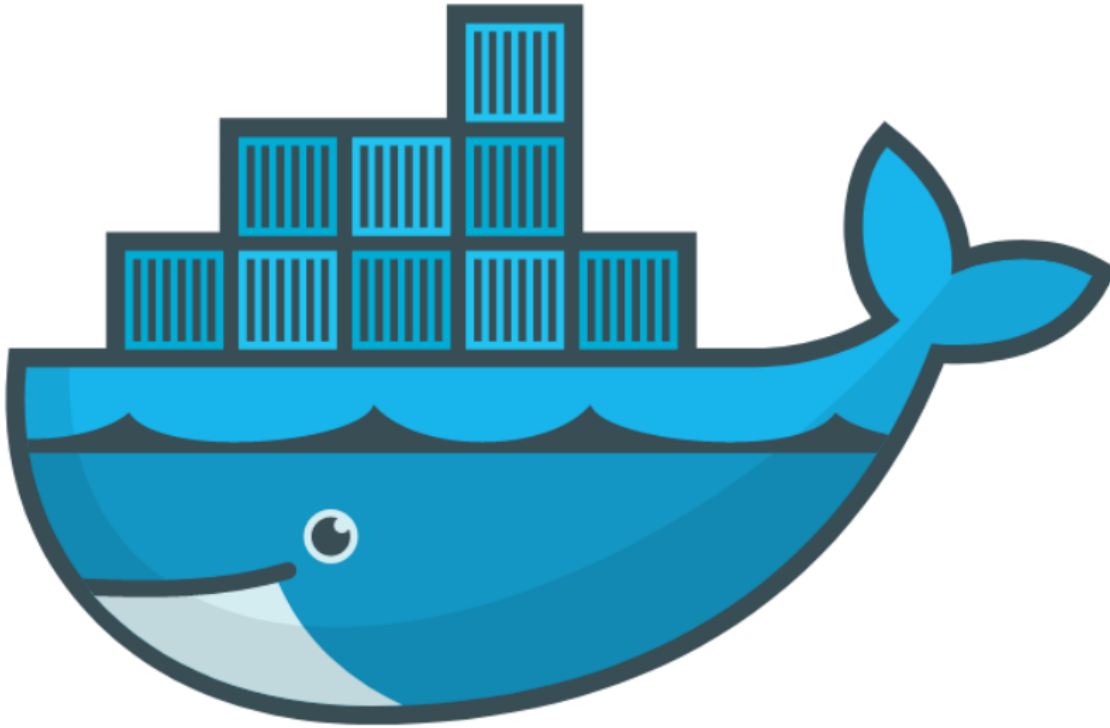
- ❑ We will use modern libraries and tools
 - Docker for deployment
 - `ir_datasets` for data wrangling
 - PyTerrier for declarative retrieval pipelines

Agenda:

- ❑ Docker Tutorial
- ❑ Project Tutorial

Docker Tutorial

Docker Basics



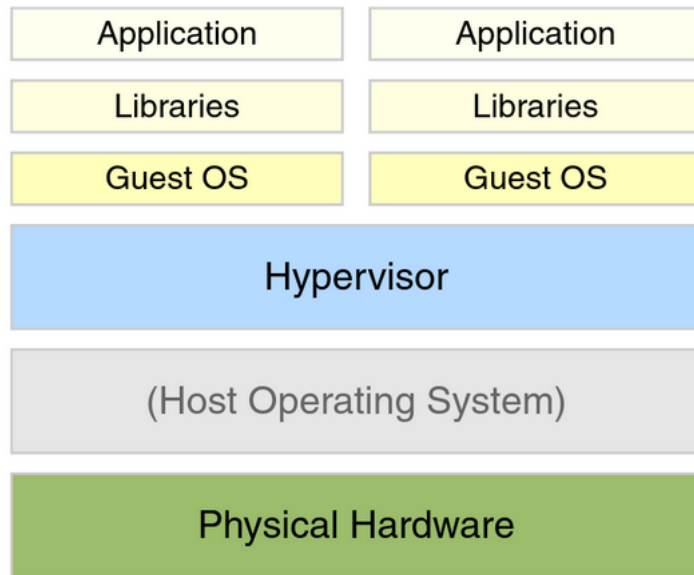
- ❑ Goal: If you can start/stop your jupyter notebook everything is fine
- ❑ <https://docs.docker.com/get-docker/>
- ❑ We will provide all required commands

Docker Tutorial

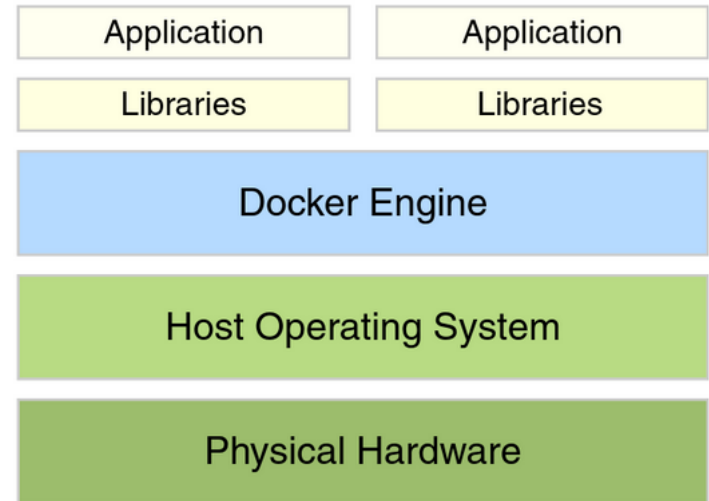
Use Cases for Docker

- ❑ Run guest systems as containers
- ❑ Shipping and running micro services as portable images
- ❑ Exploring and experimenting with new technologies
- ❑ Encapsulation mechanism to deploy applications in parallel without conflicts

Virtual Machines vs Docker



Virtual Machines



Docker

Docker Tutorial

Example Docker Commands

- ❑ Visit hub.docker.com
- ❑ We use the `bash` and `webis/tira-ir-starter-pyterrier` images

Docker Tutorial

Example Docker Commands

- ❑ Visit hub.docker.com
- ❑ We use the `bash` and `webis/tira-ir-starter-pyterrier` images

Bash Image

```
docker run --rm -ti bash
```

- ❑ `--rm`: Remove container after completion
- ❑ `-ti`: Attach stdin and stdout
- ❑ **ToDo**: Run above comand without `-ti`. What happens?
- ❑ **ToDo**: Write text to some file, restart the container. What happens?

Docker Tutorial

Example Docker Commands

- ❑ Visit hub.docker.com
- ❑ We use the `bash` and `webis/tira-ir-starter-pyterrier` images

Bash Image

```
docker run --rm -ti bash
```

- ❑ `--rm`: Remove container after completion
- ❑ `-ti`: Attach stdin and stdout
- ❑ **ToDo**: Run above comand without `-ti`. What happens?
- ❑ **ToDo**: Write text to some file, restart the container. What happens?

Bash Image With Volume Mounts

```
docker run --rm -ti -v $PWD:/bla bash
```

- ❑ `-v <HOST_PATH>:<CONTAINER_PATH>`: Mount the directory `<HOST_PATH>` on the system to the directory `<CONTAINER_PATH>` within the container
- ❑ **ToDo**: Write text to some file so that it is persistent.

Docker Tutorial

Jupyter Notebook and PyTerrier Pipelines with Docker

- ❑ We have prepared a Docker image with all reasonable libraries/frameworks preinstalled

```
docker run --rm -ti -p 8888:8888 \
    -v $PWD:/workspace/ \
    webis/tira-ir-starter-pyterrier:0.0.1-base \
    jupyter notebook --allow-root --ip 0.0.0.0
```

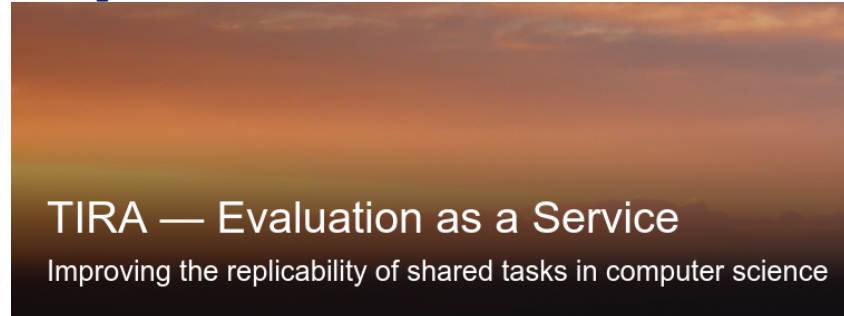
- ❑ `-p <HOST_PORT>:<CONTAINER_PORT>`: Map port `<HOST_PORT>` on the system to the port `<CONTAINER_PORT>` within the container
- ❑ `jupyter notebook --allow-root --ip 0.0.0.0`: The command executed in the container. This command starts a Jupyter notebook.
- ❑ **ToDo:** Play around with Python in the notebook for a few minutes

Project Tutorial

Evaluation and Prototyping with TIRA

You will use TIRA.io for prototyping and the evaluation of the search engines.

`http://tira.io`

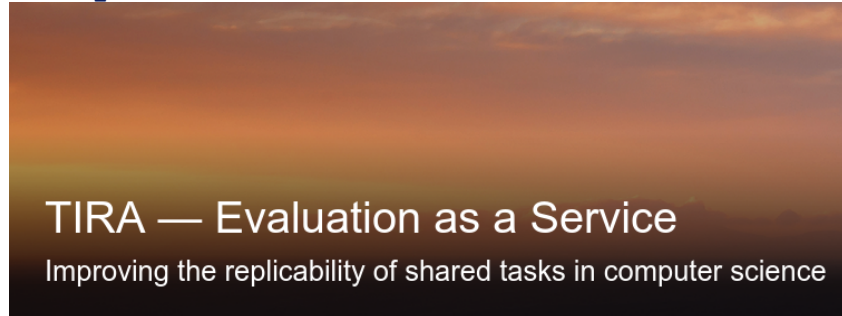


Project Tutorial

Evaluation and Prototyping with TIRA

You will use TIRA.io for prototyping and the evaluation of the search engines.

`http://tira.io`



Procedure:

1. Implement Docker images to handle data access and retrieval
2. Upload image to dedicated image registry in TIRA
3. Everything is executed in a Kubernetes cluster

Project Tutorial

Tutorial

<https://github.com/tira-io/ir-experiment-platform/tree/main/ir-datasets/tutorial>

- ❑ Covers the full pipeline (data, retrieval, analysis)
- ❑ Creates all artifacts

Homework

- ❑ You must work in groups of between 3 and 5 people.
- ❑ **Homework until next week:**
 - Find Group!
 - Mail us (1) your individual group name, (2) all group members, and (3) put all of your members in CC.
 - Theresa: theresa.elstner@uni-leipzig.de
 - Harry: harry.scells@uni-leipzig.de
 - You can get a Discord channel per group (optional, request this per email)