Practicals Specification - IR experiment

Expectations

We expect you to:
- define roles (responsibilities) within your group
- work with advanced concepts of IR (e.g., neural retrieval)
- perform an IR experiment and evaluation
- submit the code for reproducibility (e.g., in a public GitHub repo)
- report the results as a presentation

And of course learn along the way.

Detailed Description

30pts
Group size: 4 people
~ 50h/person

Perform an IR experiment.
Submit the code for reproducibility (e.g., in a public GitHub repo).
Report the results as a presentation.

An example of transformer-based IR will be provided in class.

Programming Language: Python

Typical experimental steps

- Define the goal of the experiment (e.g., in form of a research question)
- Dataset selection (use existing one)
- Dataset preprocessing and analysis (insights should inform experiment)
- Establish feasible baseline(s) (required for validation)
- Implement IR model
- Optionally implement metrics
- Evaluate models (with suitable metrics) and analyze the results

Improve and repeat.

Submission

- Prepare a Repository (code, plots for analysis, small datasets or link to dataset).
- Prepare slides.
- Put slides (PDF) for presentation in Repo.
- Submit both the link to the public Repo and the slides as PDF in TC.

Additional notes:
- Recommended platform: GitHub.
- To submit the code privately, contact in advance (at least 1 week) and provide a reason.

Submission deadline: 2023-01-10 at 00:00 for all groups.
Timeslot selection will be opened in advance.

Focus for grading

- We want a clean setup, regarding steps performed (but not necessarily regarding code style, e.g., we don't expect specific code conventions like PEP 8).
- While we want advanced methods, elegance is greater than complexity. Hence, focusing on simple but effective aspects might be preferred.
- Clear result presentation and discussion. Results do not necessarily have to be positive.
- Learning: we want you to learn a lot on this project.

Points to consider

- The setup must be valid. For instance, data leakage (e.g., test dataset used for training) is a major issue.
- Plain numbers (e.g., in form of a table) are not results, but require an analysis/interpretation.
- Negative results are also valid. However, they require a careful analysis of the results and validation of the experiment. For instance, an advanced model performing similarly to random suggests a methodological issue.

Tips

- Jupyter is generally recommended. However, for long-running tasks scripts might still be better.
- Numpy or PyTorch is recommended instead of plain python.
- If a dataset is too large, create a smaller subset and report their statistics in the presentation.
- If anything is unclear ask for help early (e.g., in TeachCenter).

Basic presentation format

8min (sharp) + 2min discussion

Title page must include:
- Name of the project (which you can define on your own)
- Group number
- Group members + team roles
- Link to repository

The content of the presentation should be structured as follows:
- Intro (RQ/motivation)
- Data + Methods (+ optional theoretical background)
- Results (+ analysis/interpretation)
- Conclusion (incl. limitations/biases)

No separate overview slide or transition slides.

Repository

Should be public (e.g., GitHub, Zenodo). **Be proud of your work.**
If this is for some reason not possible, you must let us know in advance. The code will then need to be shared privately.

Must include:
- Code for reproducibility
- Plots (other separate or within Jupyter notebooks)
- Brief description of the dataset (e.g., location)
- Optionally: the dataset itself

We will list the projects on the https://socialcomplab.github.io/ website (links to repo, but no code).
Additionally, we plan to highlight top projects on the website and social media.

IR topics

The typical example is Web Search. An extensive list of topics is provided at the IR conference - SIGIR: https://sigir.org/sigir2022/call-for-papers/

The topics span (among others):
- algorithms
- content analysis
- machine learning
- behavior modeling
- evaluation
- fairness and biases
- domain-specific applications

A focus thereby is the analysis and evaluation of IR systems. An image generator in itself is not a task of IR and its evaluation is likely not feasible for this course.

Also, we provided a separate overview of resources (e.g., project ideas and datasets).