X. Artifact Appendix

A. Abstract

Our artifacts include the source code used to build and annotate RemembERR and the annotated RemembERR database. We also provide code of all experiments described in this paper, and provide a Docker image to make reproducing the results easier. Further, we added an example script to encourage readers to write their own queries. Note that generating RemembERR from scratch is a lot of work: parsing and annotating involved tens of hours of high-focus work for two humans. Reproducing the experiments is quick (<1 h).

The `Readme.md` file in our repository contains detailed instructions.

B. Artifact check-list (meta-information)

- **Data set**: The RemembERR database
- **Run-time environment**: Python3
- **Hardware**: Any Linux machine
- **Output**: Figures and numbers
- **Experiments**: All data shown in the paper
- **How much disk space is required?** ≈3 GB (including the software dependencies)
- **How much time is needed to prepare workflow**: <1h
- **How much time is needed to complete experiments**: <1h
- **Publicly available?**: Yes
- **Code licenses** (if publicly available)?: GPLv3
- **Data licenses** (if publicly available)?: GPLv3
- **Workflow framework used**?: Luigi (Python-based, pip package)
- **Archived (provide DOI)**?: Yes.

C. Description

1) How to access: https://github.com/comsec-group/rememberr

2) Hardware dependencies: None.

3) Software dependencies: We provide instructions for Ubuntu. The apt dependencies are the following:

build-essential
libpoppler-cpp-dev
build-essential
libpoppler-cpp-dev
software-properties-common
python3.8-dev
libgl1
libglib2.0-0
software-properties-common
git
cm-super
dvipng
texlive-latex-extra
texlive-fonts-recommended
python3.8
python3-pip
python3-distutils
python3-apt.

The pip dependencies are the following:
camelot-py==0.10.1
colorama==0.4.4
luigi==3.0.3
numpy==1.22.3
openpyxl==3.0.9
pandas==1.4.2
pdftotext==2.2.2
pikepdf==5.1.1
readchar==3.0.5
matplotlib==3.5.1
opencv-python==4.5.5.64

See `Readme.md` or `requirements.txt` in our repository for details.

4) Data sets: RemembERR (provided as part of our artifacts).

5) Models: None.

D. Installation

Clone the repository and install the apt and Python dependencies. You may use a Python virtual environment:

```
python3 -m venv /.venv/rememberr
source /.venv/rememberr/bin/activate
```

Alternatively, you may find all tools preinstalled in the Docker image that we provide.

E. Experiment workflow

Follow the instructions in `Readme.md`.

F. Evaluation and expected results

Numbers will be provided in stdout and figures in the directory specified in `Readme.md`.

G. Experiment customization

We provide an example custom script that bootstrap the learning process of how to use the database for custom analyses. Please refer to `Readme.md`.

H. Notes

None.

I. Methodology

Submission, reviewing and badging methodology:

- https://www.acm.org/publications/policies/artifact-review-badging
- http://cTuning.org/ae/submission-20201122.html
- http://cTuning.org/ae/reviewing-20201122.html