

Semantic User Profiling of Scholars (PeerJ CompSci 2016): Supplementary Material

- [Semantic Publishing](#)

Supplementary material for our PeerJ CompSci submission on Semantic User Profiling. Note: files are provided here for review purposes and will be moved to Github over the next weeks.

Data

Knowledge Base

The KB files are provided as supplements, in [N-Quads format](#). The files are generated with Jena's `tdbdump` command, but should load fine into other, non-Jena triplestores as well. To load them into a new KB, create an empty directory (e.g., `/tmp/tdb`) and issue:

```
1. tdbloader --loc=/tmp/tdb triples.nq
```

For more information on `tdbloader`, please refer to Apache Jena's [TDB Command-line Utilities](#) page.

Software

All software described in the paper is available under open source licenses.

Required Third-Party Software

- You need a [JDK](#) (Java 7 or better), as well as an installation of [Apache Ant](#).
- You also need [Apache Jena](#), version 2.13 or better.
- If you want to run the text mining pipeline, you must have GATE, version 8.1 or better, installed. See the [GATE homepage](#) for instructions on how to install and run GATE.
- Also, for the text mining pipeline, you need access to a [DBpedia Spotlight](#) installation (RESTful interface). Since the output highly depends on the model used for Spotlight, if you want to reproduce our results, you will need to use the exact same model as in our paper: [en_2+2](#).
- For the automatic evaluation tool, you need additionally [Ivy](#)

Text Mining Pipeline

The text mining pipeline described in the paper is provided as a [ZIP file](#). Unzip the downloaded package on your workstation and follow the instructions below to reproduce our experiments: (*Note: the pipeline has been tested on Linux and MacOS X*):

1. You must have created a TDB-based triplestore and loaded our mapping rules as described above.

2. Start GATE (v8.1 or better). Choose `File ? Restore Application` from `File`. Then browse to where you unzipped the pipeline and choose the provided `.xgapp` file.
3. Once the pipeline is loaded, you can open any document.
4. Create a new corpus and add the documents.
5. Double-click on the `Semantic_Profiling` pipeline under `Applications`. Choose the new corpus you created above from the dropdown and click `Run this Application` button.
6. Once the pipeline is finished you can open the documents and examine their annotations.
7. In order to examine the generated triples, **first close the GATE application**. Then you can either check the triples using Jena's `tdbdump` command or publish it through a [Fuseki server](#).

Automatic Evaluation Tool

The evaluation in the paper is based on the provided responses from the user study participants, exported from LimeSurvey. Our evaluation tool then takes these results and computes the various metrics provided in the paper. You can download the source code as a [zip archive](#).

The tool can be started from the command line in the 'Analysis'-folder with the ant task '`ant run`'. The tool processes all files in the `data` folder (original exported Limesurvey files in `xlsx` format) and creates an output folder `result` including the analyzed files in the format `original_file_name'_results_metrics.xlsx`.

The currently computed metrics are: *MAP*, *Precision@rank* and *nDCG* as described in the paper.

Attachment	Size
ScholarLens.zip	5.6 MB
Analysis.zip	458.99 KB



Except where otherwise noted, all original content on this site is copyright by its author and licensed under a [Creative Commons Attribution-Share Alike 2.5 Canada License](#).

Source URL (retrieved on 2026-01-09 14:17):

<https://www.semanticsoftware.info/semantic-user-profiling-peerj-2016-supplements>