Taguette: open-source qualitative data analysis

Rémi Rampin[1] **and Vicky Rampin**[1]

1 Independent Researcher

Summary

Taguette is a free and open-source computer-assisted qualitative data analysis software (CAQDAS) (Knowledge Bank, 2018) package. CAQDAS helps researchers using qualitative methods to organize, annotate, collaborate on, analyze, and visualize their work. Qualitative methods are used in a wide range of fields, such as anthropology, education, nursing, psychology, sociology, and marketing. Qualitative data has a similarly wide range: interviews, focus groups, ethnographies, and more.

Statement of Need

Taguette fills a specific research need for qualitative researchers who cannot afford access to the software to do their work. For commercial CAQDAS packages, the lowest subscription price is 20 USD/month, and the lowest desktop application price is 520 USD (Knowledge Bank, 2018). There have been fewer than twenty open-source CAQDAS packages available **ever**, and fewer than five are being currently maintained, including Taguette.

Taguette directly supports qualitative inquiry of text materials (see Figure 1). It is unique in that it provides a free and open-source tool for qualitative researchers who want real-time collaboration (see Figure 3). Taguette has already been used in multiple research publications, which we have compiled in a Zotero library (*Taguette Zotero Library*, n.d.), and also is being self-hosted by research institutions on behalf of their communities (example: Digitalization Research Cluster, Leiden University).

Taguette

Taguette is a web application written in Python (Python Software Foundation, 2021) with the Tornado Web Framework (Facebook Inc and contributors, n.d.). It is designed to run both on a desktop machine, in single-user mode, or on a server, where it allows real-time collaboration. In addition, we have been running a server at app.taguette.org for anyone to use since March 2019, where we have about 2,000 monthly active users. Taguette is multiplatform, with installers provided for MacOS and Windows, a Docker image, and on the Python package Index (PyPI). It is available in 7 languages and has been downloaded over 12,000 times.

Importing Documents

Work in Taguette begins with importing a document. We support a variety of text formats, including HTML, RTF, EPUB, PDF, DOCX, Markdown, and more. Documents are converted to HTML using the `ebook-convert` command, part of the Calibre ebook manager (Goyal &

contributors, n.d.) or wvWare (McNamara & contributors, n.d.) for old Microsoft Word 97 .doc documents. A copy of Calibre is included in our installers so that users don't have to set up any additional software. After conversion, the document is sanitized to remove unwanted formatting and embedded media, and avoid security issues such as cross-site scripting.

Analysis

After a user has imported a document into Taguette, they can then qualitatively highlight sections of text (see Figure 1). Those highlights are organized in hierarchical tags that can be created, merged together, and recalled at will (see Figure 2). Data for all projects including documents, tags, and highlights is stored in a SQL database, which allows for easy exploration and scripting should the user need to go beyond the capabilities offered by our interface. In single-user mode, Taguette automatically creates a SQLite database in the user's home directory, and performs schema migrations automatically when a new version of Taguette is installed. Taguette can also use the other SQL backends supported by SQLAlchemy (Bayer, 2012).
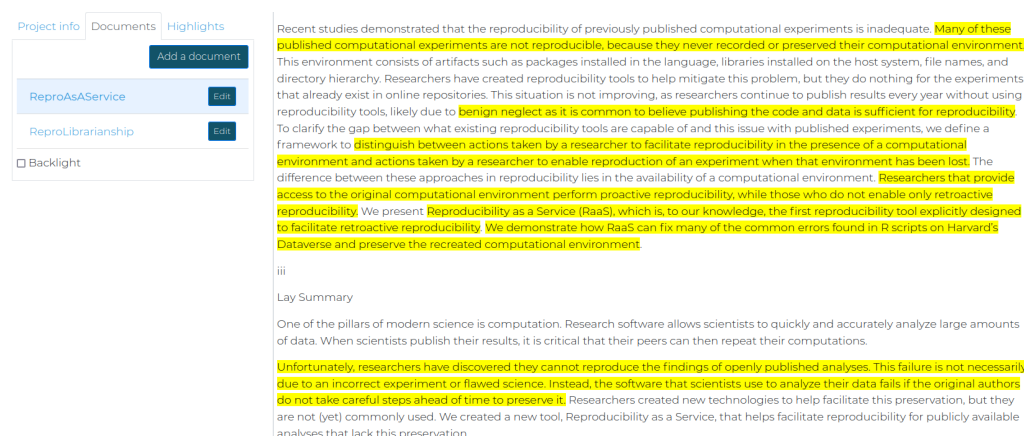

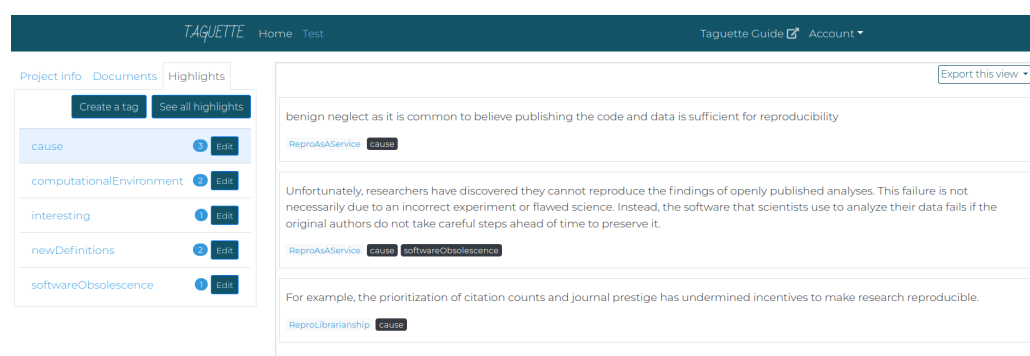**Figure 1:** Document view, where highlights are created and associated with tags.

Figure 2: List of highlights for a given tag.

Live collaboration

The multi-user version of Taguette allows for live collaboration of multiple users in a single project. It is possible to add other accounts as collaborators to your project, with a choice

of permissions: some users can only tag, some can change documents, and others have full control including adding or removing collaborators.
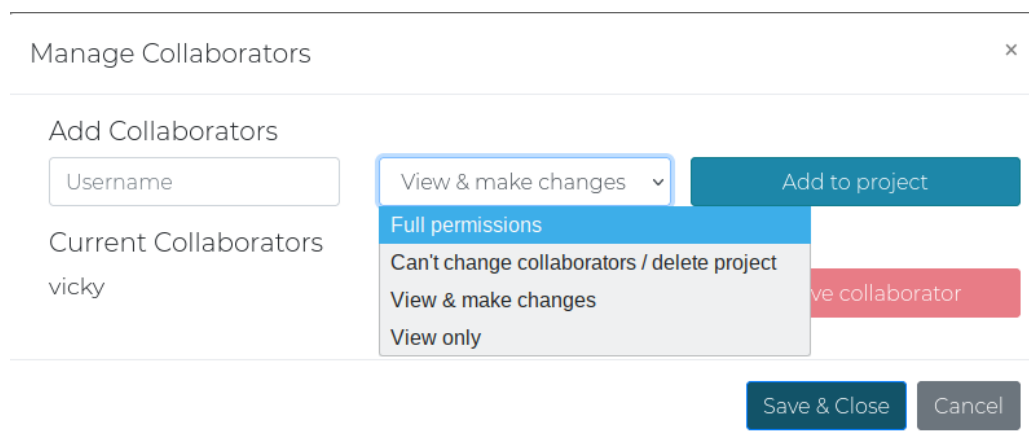

Figure 3: Adding collaborators through the interface.

From then on, any change made by a different user is reflected immediately to the other users. This allows for faster annotation of large projects, without having to exchange partially processed documents via email for example. Taguette is currently the only free and open-source CAQDAS package that supports this.

Exporting

Taguette offers a variety of exporting options. A user can export a codebook as a document or spreadsheet, which is the list of all the tags, with their description and the number of associated highlights, throughout the project. Another option is to export a highlighted document, where the sections highlighted by the user are marked and each annotated with the associated tags. Finally, it is possible to export a list of all the highlights across documents, either for all tags or for a specific tag or hierarchy of tags (see Figure 4).
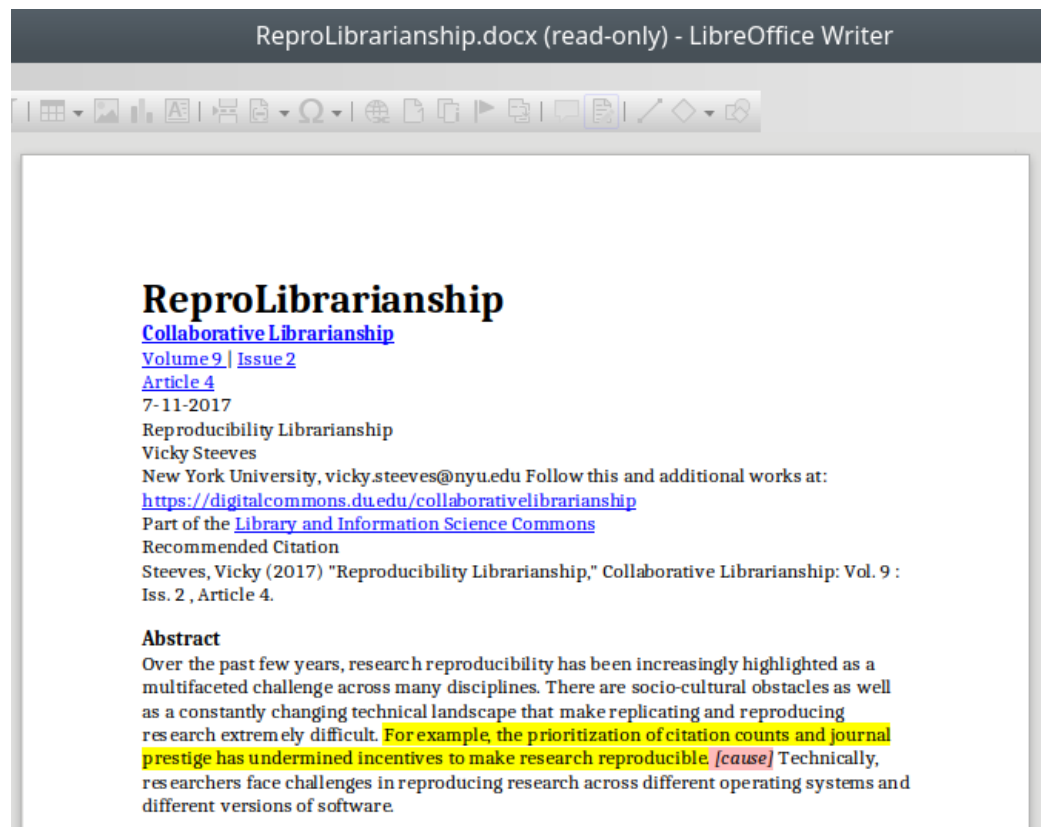
Figure 4: A highlighted document exported from Taguette and opened in LibreOffice.

It is also possible to export a project as a SQLite3 database (Hipp, 2000), in Taguette's native schema, that contains all the information necessary to continue work on another instance of Taguette. It is even possible to import them on our hosted version, app.taguette.org, or to export from there to a local copy. Older versions of the schema are automatically recognized and converted to the latest version if needed.

Related Work

Other currently maintained open-source CAQDAS packages include: QualCoder (Curtain, n.d.), qcoder (Elin Waring et al., n.d.), and qdap (Rinker et al., n.d.). qcoder and qdap are both R packages that support qualitative analysis of text, and require knowledge of R and RStudio to use. Both provide an interface to use the results of qualitative analysis with the rest of the R ecosystem. QualCoder is a desktop application (made with Python and PyQt5) that allows users to qualitatively analyze text and audiovisual materials. Each currently maintained tool fulfills different needs across the qualitative community, including Taguette. Previously maintained qualitative include Aquad (G. L. Huber & Leo Gürtler, n.d.), RQDA (Huang Ronggui, 2018), and the Coding Analysis Toolkit (CAT) (Texifter, 2010).

Acknowledgments

We thank Dr. Sarah DeMott, whose work triggered the creation of Taguette. We would also like to thank our contributors on GitLab and our translators on Transifex, and the qualitative analysis community for their warm welcome and feedback.

In addition, we have recently started an OpenCollective to support the development of Taguette, with the initial goal to cover the cost of a dedicated server for our hosted service. We are grateful to the backers for their kind donations to the project.

References

Bayer, M. (2012). SQLAlchemy. In A. Brown & G. Wilson (Eds.), *The architecture of open source applications volume II: Structure, scale, and a few more fearless hacks*. aosabook.org. http://aosabook.org/en/sqlalchemy.html

Curtain, C. (n.d.). *QualCoder*. https://github.com/ccbogel/QualCoder

Elin Waring, Dan Sholler, Jenny Draper, & Beth Duckles. (n.d.). *QCoder*. rOpenSci Labs. https://github.com/ropenscilabs/qcoder

Facebook Inc and contributors. (n.d.). *Tornado web framework*. https://www.tornadoweb.org/

G. L. Huber, & Leo Gürtler. (n.d.). *AQUAD*. https://www.aquad.de/E_Uebersicht.html

Goyal, K., & contributors, C. (n.d.). *Calibre*. https://calibre-ebook.com/

Hipp, R. D. (2000). *SQLite*. https://www.sqlite.org/

Huang Ronggui. (2018). *RQDA*. https://rqda.r-forge.r-project.org/

Knowledge Bank. (2018). *The CAQDAS. Software for Qualitative Analysis*. https://www.mvorganizing.org/the-caqdas-software-for-qualitative-analysis/

McNamara, C., & contributors. (n.d.). *wvWare*. http://wvware.sourceforge.net/

Python Software Foundation. (2021). *Python programming language*. http://www.python.org/

Rinker, T., Goodrich, B., & Kurkiewicz, D. (n.d.). *Qdap: Bridging the Gap Between Qualitative Data and Quantitative Analysis*. https://CRAN.R-project.org/package=qdap

Taguette zotero library. (n.d.). https://www.zotero.org/groups/4373578/taguette

Texifter. (2010). *Coding Analysis Toolkit*. http://cat.texifter.com/