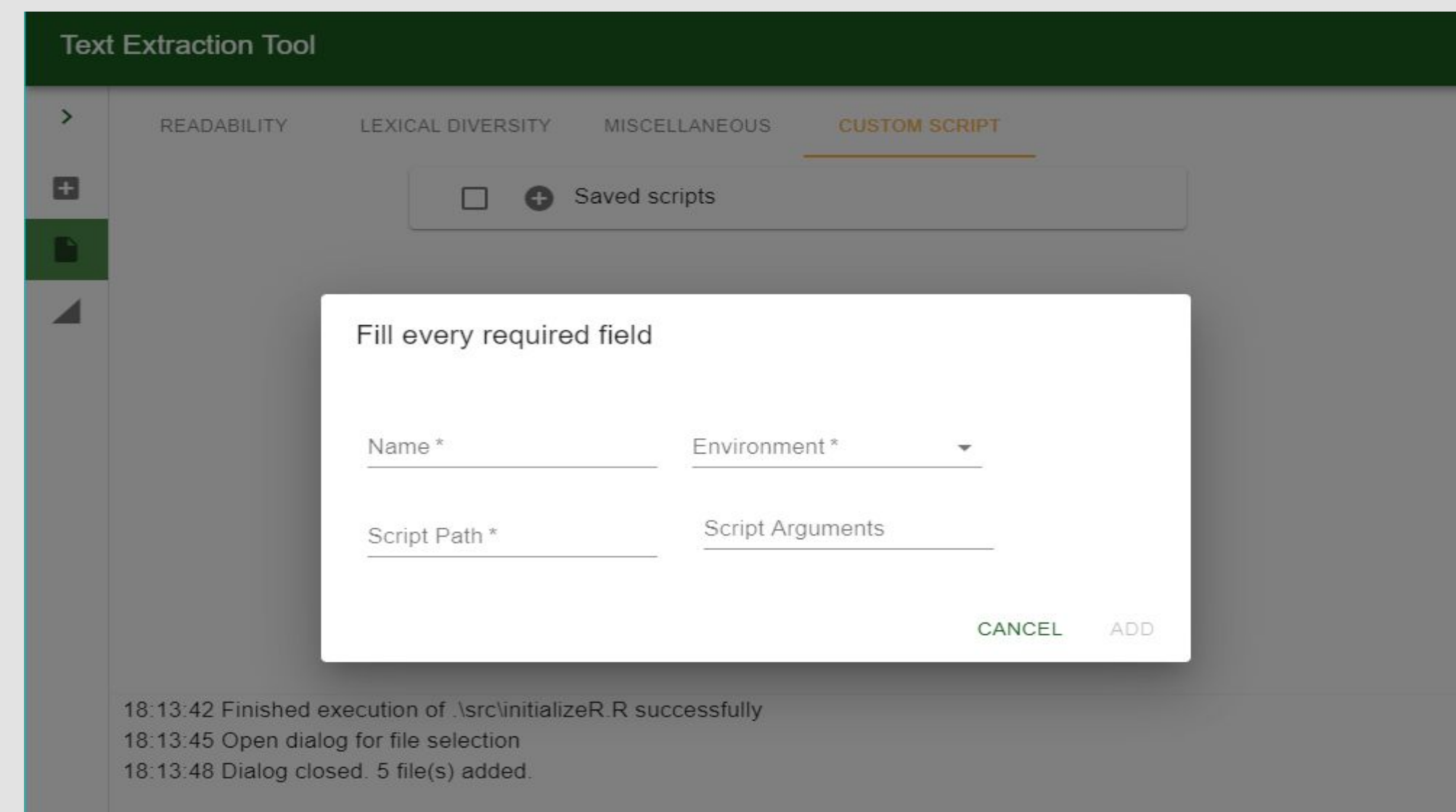




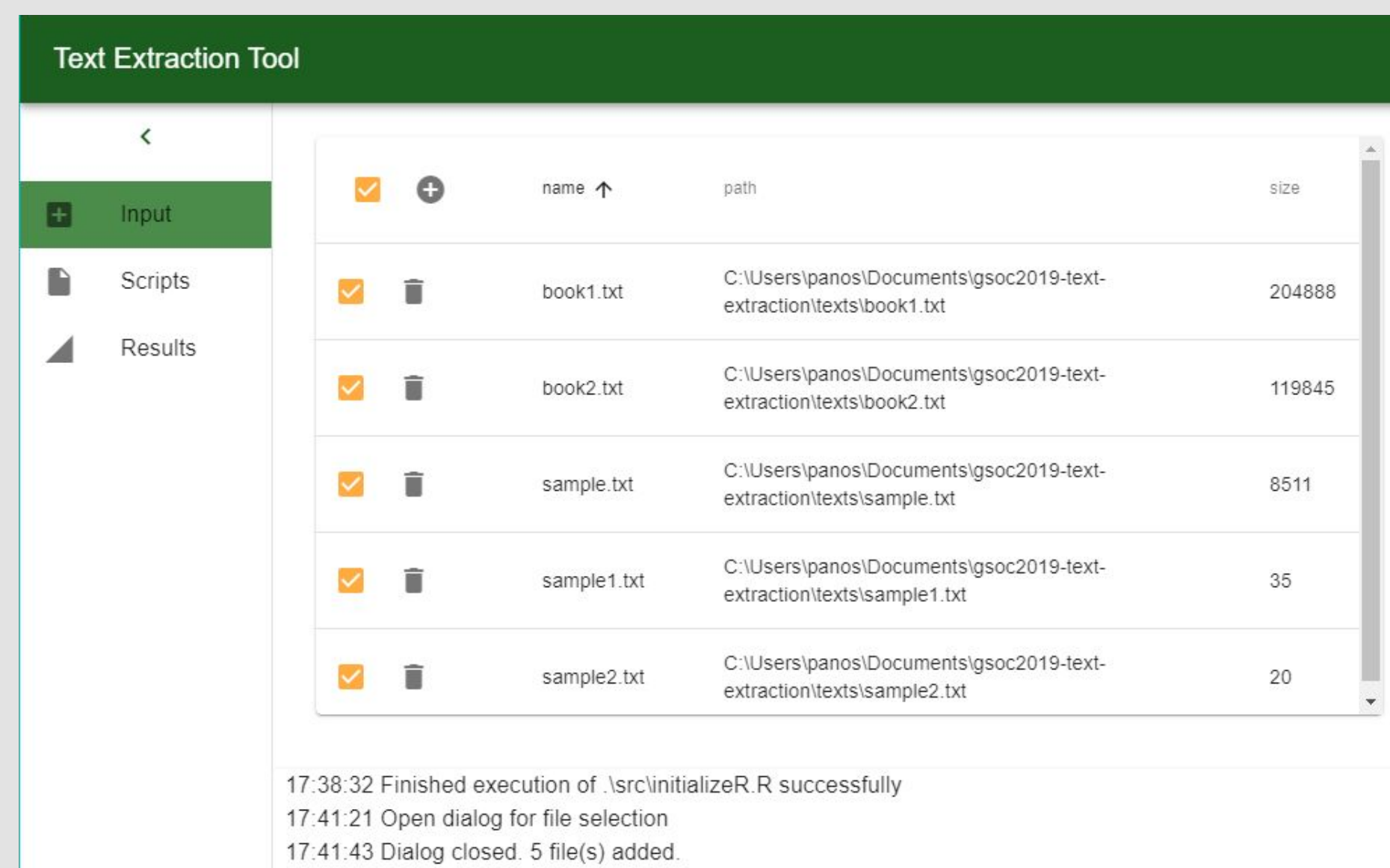
### Abstract

This project aims at developing a tool for extracting quantitative text profiles from large collections of texts. Existing solutions and tools are hard to use and require high technical knowledge, often absent in researchers of Digital Humanities (DH). We developed an open-source working prototype that can also be used in combination with other tools to create an automated pipeline of text analysis.



### Features and Technical Information

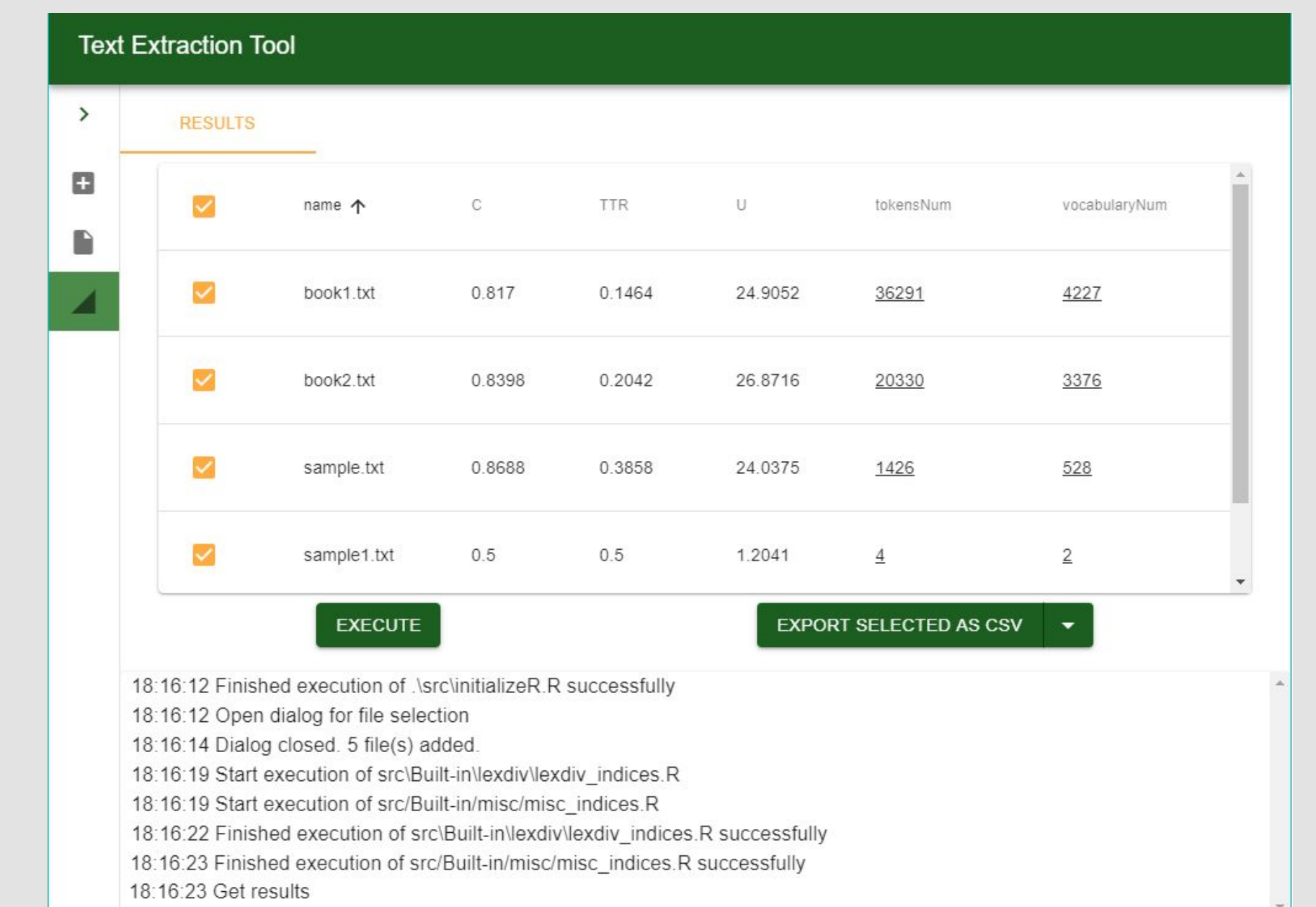
- Xtralingua implements quantitative text indices from R packages (Quanteda & koRpus) and the standalone software QUITA.
- Intuitive GUI allowing researchers to produce complex quantitative text profiles with only a few clicks.
- Support of 61 different quantitative text indices for text readability, lexical diversity and specialized measurements based on quantitative linguistics theory.



### Graphical User Interface

Simple GUI to be able to navigate and operate the application. It consists of three main tabs...

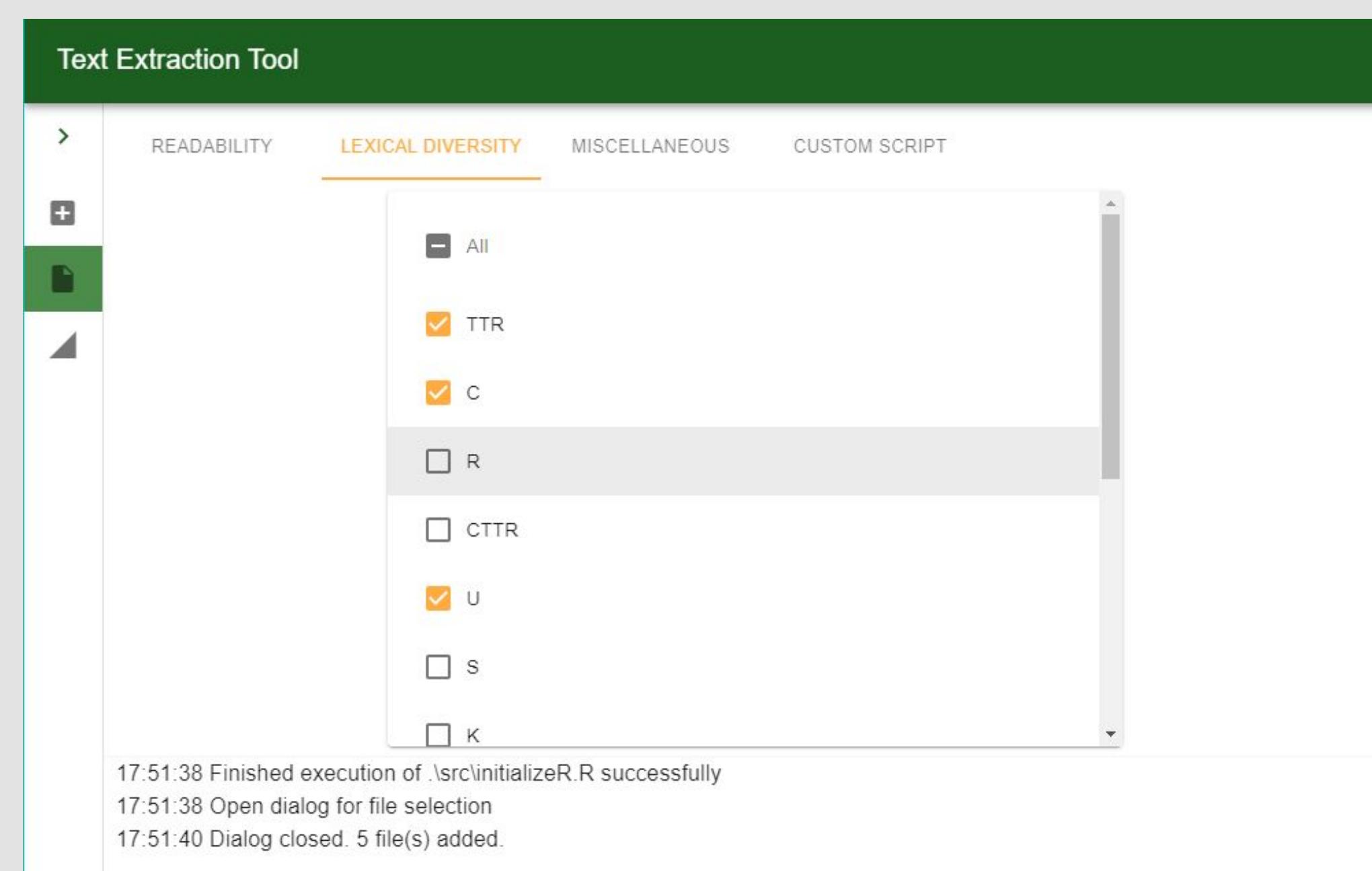
- *Input Tab*: Selection of text files to process.
- *Scripts Tab*: Selection of the quantitative text indices to be calculated.
- *Results Tab*: Inspection of the document-feature matrix and data export (CSV or JSON).



### Methodology

Development of an integrated tool for quantitative text analysis, with the following characteristics:

- User friendly Graphical User Interface (GUI).
- Calculation of many readability & lexical diversity indices.
- Integration of existing open-source text analysis packages under a single platform.
- Flexibility and scalability.
- Open-source code.



### DOWNLOAD NOW - BE PART OF OUR COMMUNITY

<https://github.com/hocrt/Xtralingua-2.0>

HELPDESK [info@hellenicOCRteam.gr](mailto:info@hellenicOCRteam.gr)



### Acknowledgements

This project was developed for the *Hellenic OCR Team*, a scientific crowdsourcing initiative for the processing and analysis of parliamentary data, during the Google Summer of Code 2019 program.