
Rozha: Supporting and Simplifying Multilingual Natural Language Processing

Ian Goodale, UT-Austin

What is Rozha?

- A Python package that simplifies many NLP pipelines
 - Includes functions to perform a wide variety of NLP processes using over 70 languages, from stopwords removal to sentiment analysis and many more, in addition to visualizations of the analyzed texts
 - Allows users to choose from NLTK, spaCy, and Stanza for many of the processes it can perform, allowing for easy comparison of the output from each library
 - Released on [GitHub](#) and [PyPI](#) under the GNU General Public License, allowing users to use and contribute to the tool with minimal limitations.
-

Why create Rozha?

- To serve researchers and other users by streamlining NLP processes
 - To make NLP more accessible to students and beginners interested in exploring the field
 - To promote NLP in languages other than English by making it easy for users of varying experience levels to work with languages other than English
-

What can it do?

- The package contains three modules for processing, analyzing, and visualizing text
 - In addition to NLP processes, the visualize module allows users to generate graphs of their processed NLP data
 - A full breakdown of the functions can be viewed here:
<https://github.com/ian-nai/Rozha/blob/main/Functions.md>
 - Sample code demonstrating the package can be run in the following Jupyter notebook:
 - <https://colab.research.google.com/drive/108FvBVuNIIMneNzdNLYSVbu91D2dKcGs?usp=sharing>
-

Questions?

Please let me know in the conference Slack, or by writing to me at ian.goodale@austin.utexas.edu

I am also active on GitHub at <https://github.com/ian-nai/>
