# Lexos

# Summary

Lexos is a suite of tools designed to facilitate the computational analysis of literary and historical texts. It offers an integrated workflow in which the pre-processing ("scrubbing"), analysis, and visualization steps can be accomplished in a single, web-based environment. Scrubbing features include handling punctuation, stop words, markup tags, and character consolidations, as well as document segmentation, culling, and n-gram tokenization. Analytical tools include basic document statistics, hierarchical and k-means cluster analysis, rolling window analysis, cosine similarity ranking, and z-score analysis. Visualizations include word and bubble clouds, comparative "multiclouds" (which can be used to analyze data from [MALLET](http://mallet.cs.umass.edu/)-produced topic models). Analytical tools produce line, PCA, Voronoi cell, and dendrogram graphs. Each of the tools has export functionality.

Lexos is aimed at entry-level users as well as advanced scholars using small to medium-sized text corpora. It places particular emphasis on the processing of ancient and non-standard languages, as well as non-Western languages that do not use the Roman alphabet.

Lexos is produced by the [Lexomics Research Group](http://lexomics.wheatoncollege.edu). An online version of Lexos v3.0 is available at <http://lexos.wheatoncollege.edu/>.

# Release history

This repo reflects ongoing development since our Summer 2016: Lexos v3.1.

Earlier versions are available at <https://github.com/WheatonCS/Lexos/releases>.

# Installation

Installation instructions for Lexos v3.1 are available in the project Wiki.

* [Windows Install Guide](https://github.com/WheatonCS/Lexos/wiki/Windows-Install-Guide)
* [macOS Install Guide](https://github.com/WheatonCS/Lexos/wiki/macOS-Install-Guide)
* [Linux Install Guide](https://github.com/WheatonCS/Lexos/wiki/Linux-Install-Guide)

## System Architecture (in brief)

Lexos v3.1 is written primarily in Python 2.7.11 (as distributed in [Anaconda](https://www.continuum.io/downloads) 4.1.1) using the [Flask](http://flask.pocoo.org/) microframework, based on Werkzeug and Jinja2.

Lexos v3.2 (the current master branch of this repository) is now being developed in Python 3.

The front end is designed using [jQuery](https://jquery.com/) and the [Bootstrap 3](http://getbootstrap.com/) framework, with a few functions derived from [jQuery UI](https://jqueryui.com/) and [DataTables](https://datatables.net/). We increasingly incorporate the wiz from [D3.js](http://d3js.org/) in our visualizations and the power in the [scikit-learn](http://scikit-learn.org/stable/) modules for text and statistical processing.

The directions for setting up the development environment for testing (using localhost:5000) on your local machine are stored in the 0\_InstallGuides directory.

## Dependencies

Lexos requires the following Python packages:

chardet, flask, gensim, matplotlib, natsort, numpy, pandas, pip, scikit-learn, scipy, pip, requests

On MacOS, the PDF Viewer extension needs to be enabled in the Chrome browser.

Lexos works on Chrome and Firefox. Other browsers are not supported, and some features may not function.

## License information

See the file LICENSE for information on the terms & conditions for usage and a DISCLAIMER OF ALL WARRANTIES.

## Citation information:

Kleinman, S., LeBlanc, M.D., Drout, M. and Zhang, C. (2016). Lexos. v3.0. <https://github.com/WheatonCS/Lexos/>. <doi:10.5281/zenodo.56751>.