

WE1S “topic_modeling” module

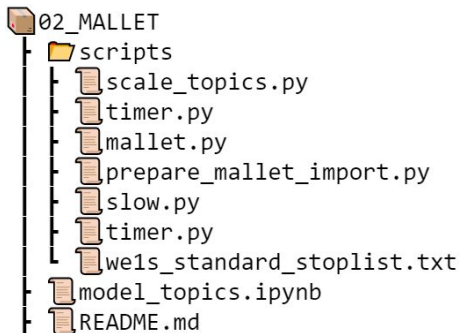
WE1S’s topic_modeling” module is a key component in the WE1S Workspace (see [S-2](#) on Workspace). The module performs topic modelling with [MALLET](#) on a collection of texts by providing an easy-to-configure interface for the two basic steps of importing texts into MALLET and training topics. (MALLET is a widely used, Java-based toolkit for LDA topic modeling.¹)

Optionally, the module can also perform a preprocessing step on texts before the import, and a post-processing step (scaling) after training topics.

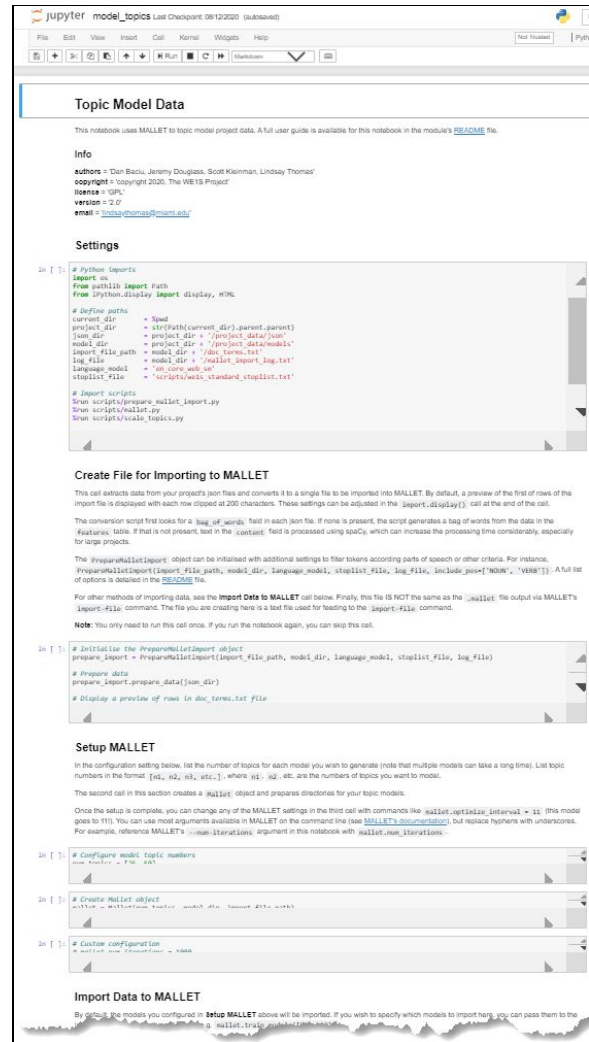
After modelling is complete, there is also an option to create a scaled data file for use in the topic model visualisation interfaces that can subsequently be generated through other modules in the WE1S Workspace--e.g., `dfr_browser`, `topic_bubbles`, `pyldavis`, and others. (See our [Tools for Visualizing Topic Models.](#))

The Jupyter notebook in this module is `model_topics.ipynb`. It includes a step-by-step use guide, one cell in the notebook at a time.

Module structure



¹ See our [M-2](#) card for a quick overview of topic modeling. See Ted Underwood, “[Topic Modeling Made Just Simple Enough](#)” for an explanation of LDA (Latent Dirichlet Allocation) for a non-technical audience.



Screenshot from `model_topics.ipynb` ([larger](#))

Further Information:

- * [MALLET topic modeling toolkit](#) by Andrew McCallum (Common Public License)
- * [M-2](#) (on topic modeling)
- * Ted Underwood, “[Topic Modeling Made Just Simple Enough](#)” (2012)

Jupyter notebook in this module:

- * `model_topics.ipynb`

Code source: [TBD] ([MIT License](#))