

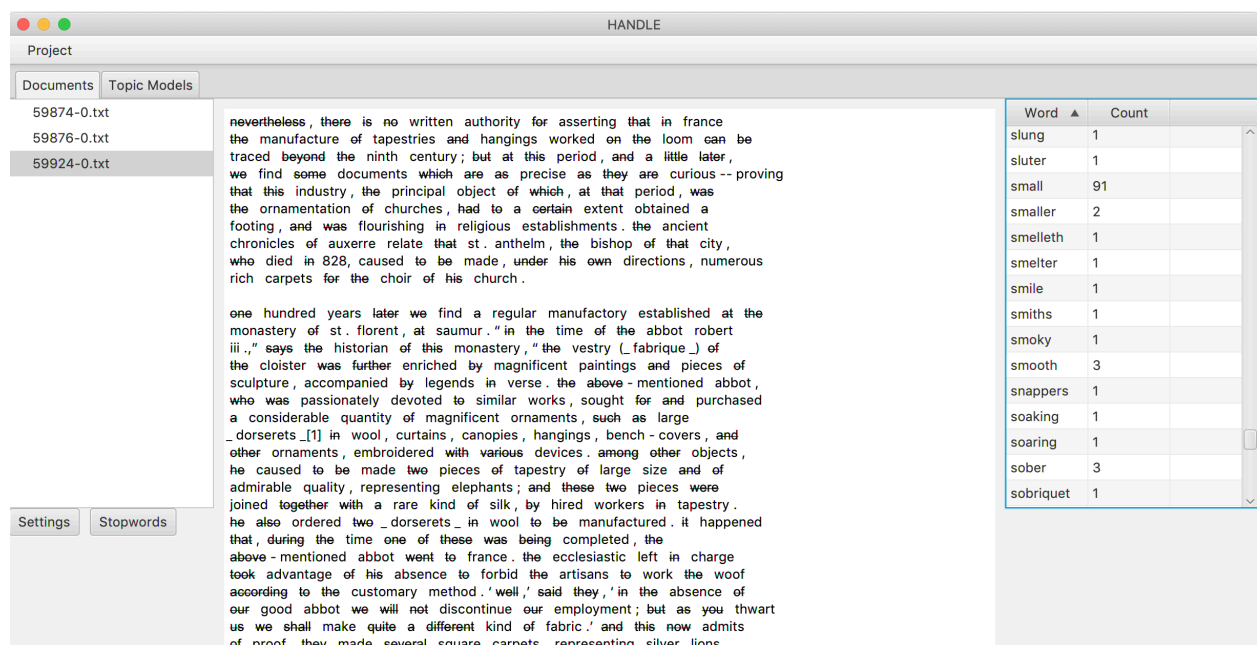
HANDLE

Handle is a tool for easy topic modeling. It's aimed at students and scholars who want to experiment with topic modeling, but who don't want to work with the more technically complex interfaces of other topic modeling programs. It emphasizes ease of experimentation.

If you're familiar with MALLET, HANDLE is a friendly interface for MALLET. It's MALLET at its core, but it is not affiliated with or supported by the MALLET developers.

This guide provides a basic overview of how to use HANDLE.

Importing Documents



The document import window is where we bring data into HANDLE. This has 3 frames, from left to right: the **document list**, the **preview window**, and the **word count table**.

Document List

A list of all documents for your topic model. Drag and drop files from any folder into this window to bring them into HANDLE. HANDLE supports dragging and dropping multiple files. You can also drag and drop multiple times, so that, for example, you can bring in files from many different folders. Importing isn't a one-shot process.

HANDLE supports plain text files (ASCII or UTF-8) and Microsoft Word (.doc and .docx).

NOTE: HANDLE does not check for duplicate files. If you drag and drop the same file 20 times, it will appear 20 times in your corpus.

Click on a window in the document list to show it in the Preview window (see “Preview Window” below).

To remove a document, right click on it and select “Remove” from the popup menu that will appear.

Beneath the document list is the “Stopwords” button. Click here to edit the stopwords list.

Preview Window

The Preview window shows a document as MALLET will see it. If you choose to lowercase all text (the default), all text will be lowercase. Stopwords will appear struck out.

Please note that spacing is preserved here for easy reading, but space doesn’t matter to MALLET.

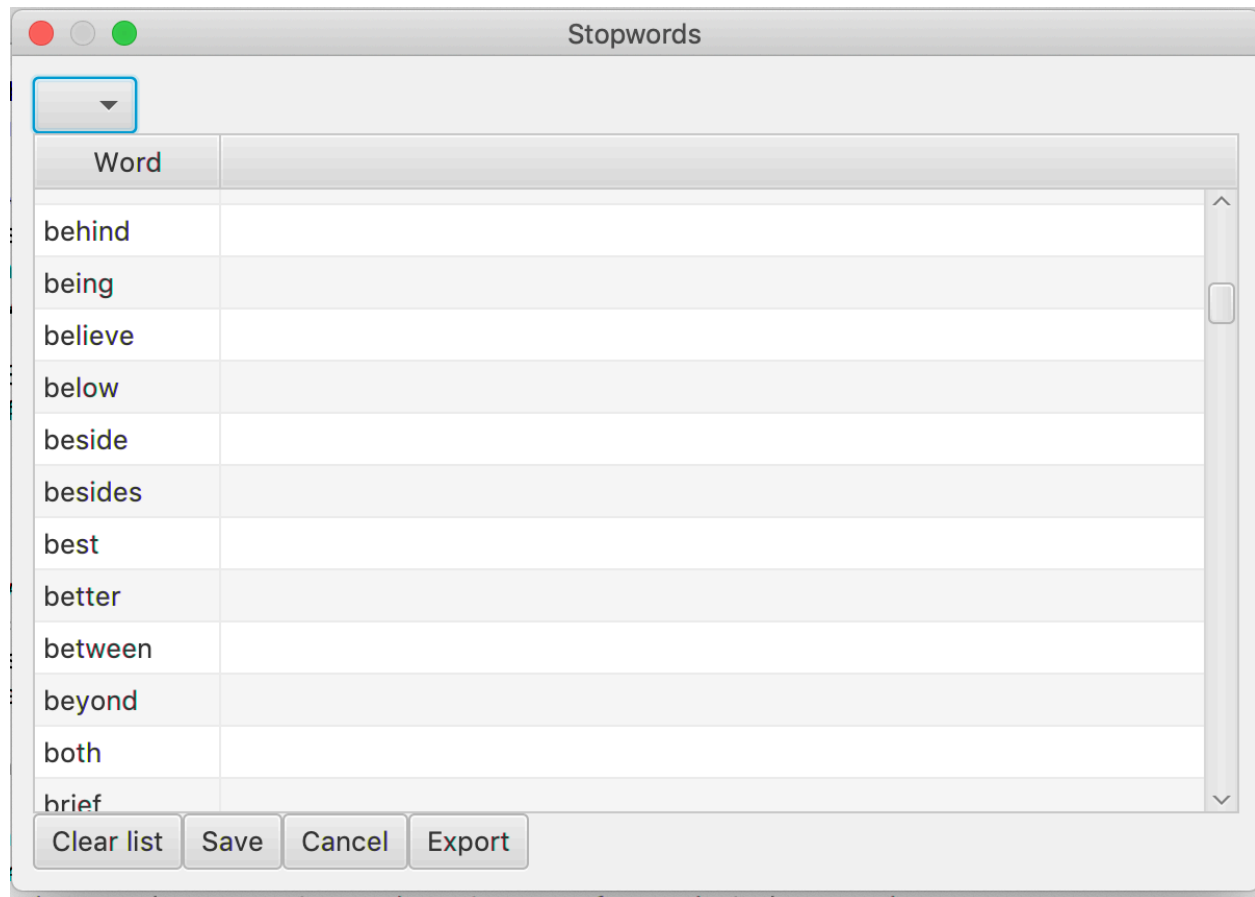
Word Count

The Word Count table shows the count for words as they appear in the selected document. This can help you discover words that should be stopwords.

To sort the words in the word count table, click on the appropriate column header.

To add a word as a stopword, right-click on it and choose “Add as stopword.”

Stopwords List



This shows a list of all stopwords in the project. To add more stopwords, drag and drop a plain text file of stopwords into this window. To remove a word from the stopword list (that is, to make MALLET count it), right click on it and click “remove” from the popup menu that shows.

If you dragged and dropped a stopwords file into this window, removing a word from the stopwords list will not make any changes to the file that contains the stopword list. The stopwords file is copied into HANDLE and then HANDLE basically forgets about the file. This also means that if you edit the file, it won’t affect the stopwords list, either.

You must click “Save” before exiting this window. Clicking “Cancel” will undo any changes you made to the stopwords list.

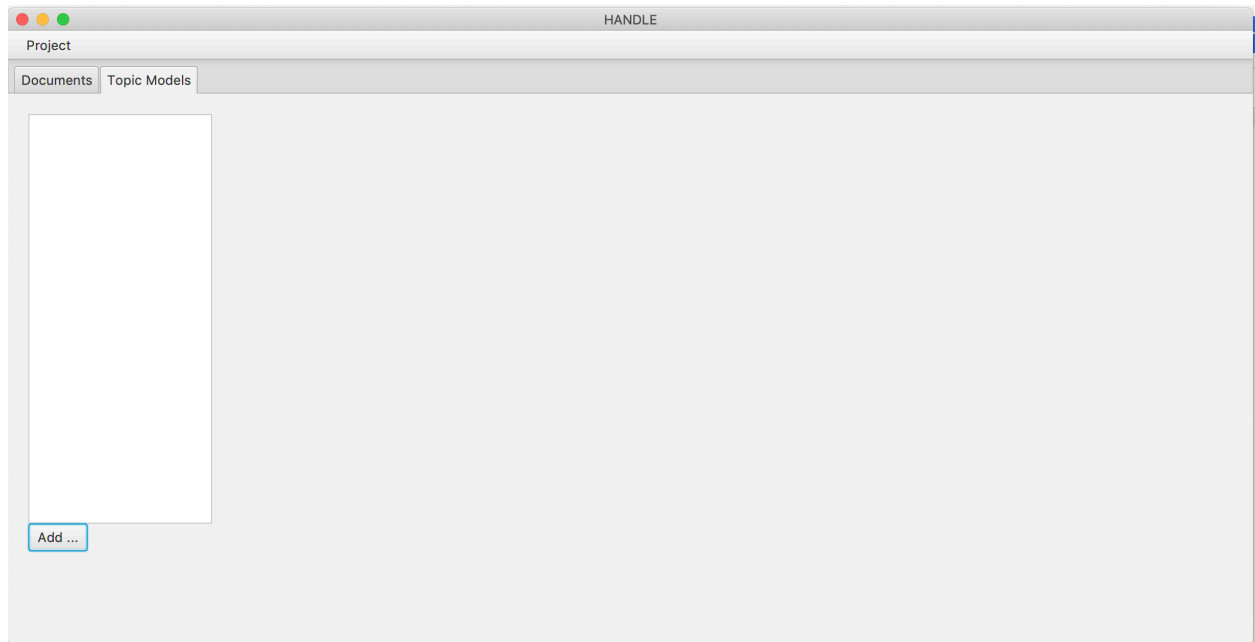
- Remove a single stopword by right-clicking on it.

- Clear the entire list by clicking “Clear list.”
- To export all the stopwords you used into a file, click “Export.”

Topic Models

Creating a topic model

First, click the “Topic Models” tab. You’ll see the following window:



Click the “Add” button. The “New topic model” window will pop up.

The screenshot shows a window titled "New topic model" with a tab labeled "Round 3". Below the tab is a "Topic Count" field set to "20". A section titled "Advanced ..." is expanded, revealing several settings: "alpha" is 5.0, "beta" is 0.01, "random s..." is -1, "iterati..." is 100, "optimize inter..." is 50, "Burn-in" is 200, and "Threads" is 8. At the bottom of the settings section is a "Run" button and a long, empty progress bar. The rest of the window is a large, empty gray area.

Set the settings or leave the defaults. Then click “Run.” A progress bar will appear. Then, you’ll be able to see the topics that were generated once it’s complete.

Now, you have a choice: do you want to keep the topic model you generated, or do you want to try again with different settings? If you want to keep the topic model, you **must** click the “Save” button that replaced the “Run/Stop” button when the topic model finished. If you just close the window, the topic model will be lost.

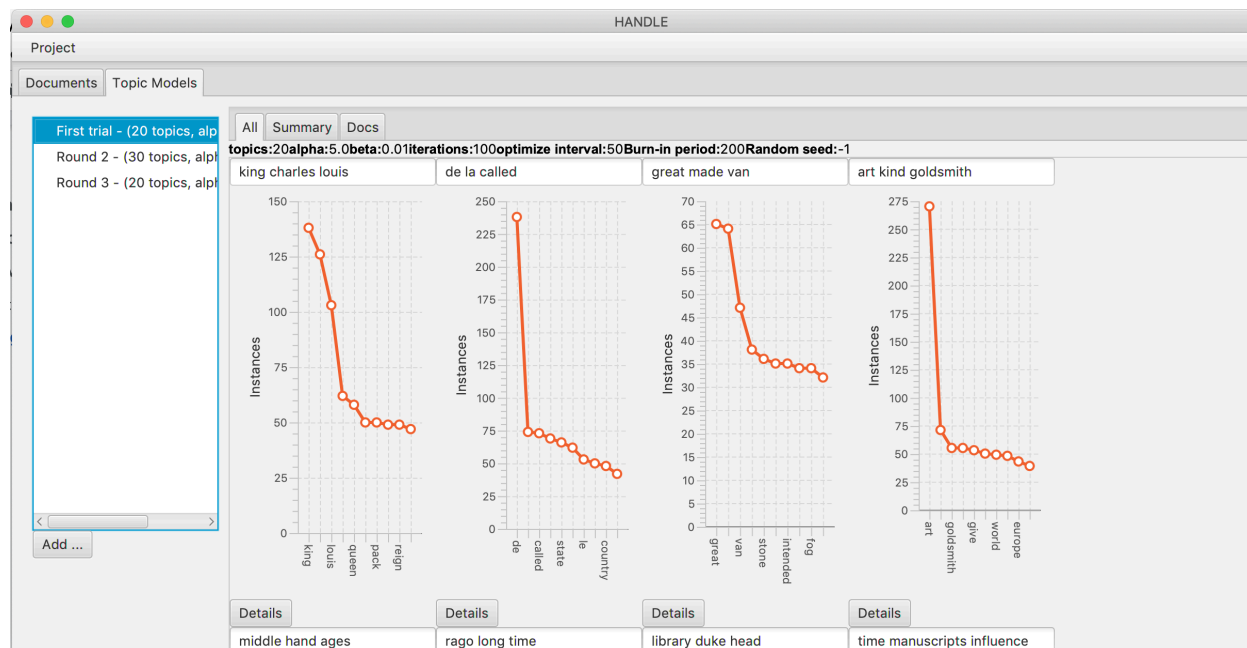


If you're an experienced MALLET user, the graphs that show are the word lists that MALLET produces every few iterations. Think of it like a topic-phrase report that is redrawn every 50 iterations or so.

Viewing topic models

Now, you're ready to work with your topic model. The topic model will show up in the list in the left-hand side of the window. To name a topic model, double-click on it.

The three tabs at the top of the screen offer different options for visualizing and working with the data.



All

"All" is a summary of all topics in the topic set. This is the view in the new topic model window. You can rename a topic here by editing the value in the box.

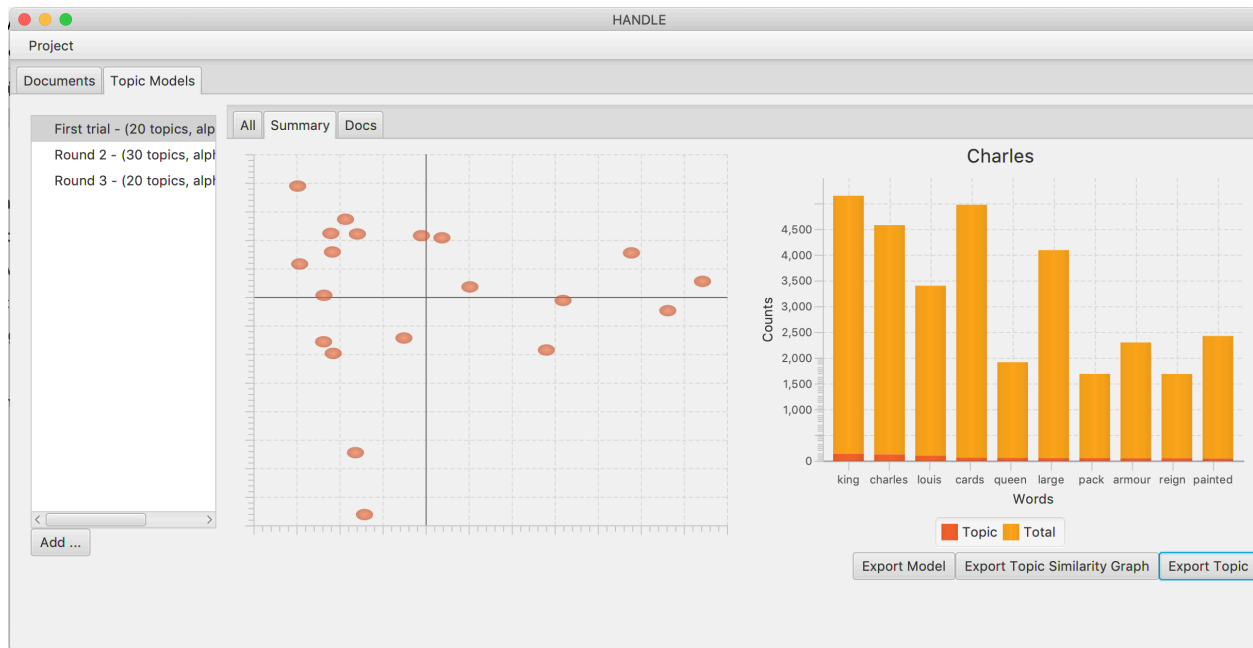
Click the "Details" button to see a simple report of how each document mentions the topic.

If you're a MALLET user, this is like the xml-topic-phrase-report option.

Summary

"Summary" is an LDAviz-inspired graph that shows how similar the topics are. Hover over a dot to see the top 3 words in it. Click on the dot to see a report of the top 10 words in that topic, and

a comparison of the number of times each word appears in that topic vs. the number of times it appears in the whole corpus.

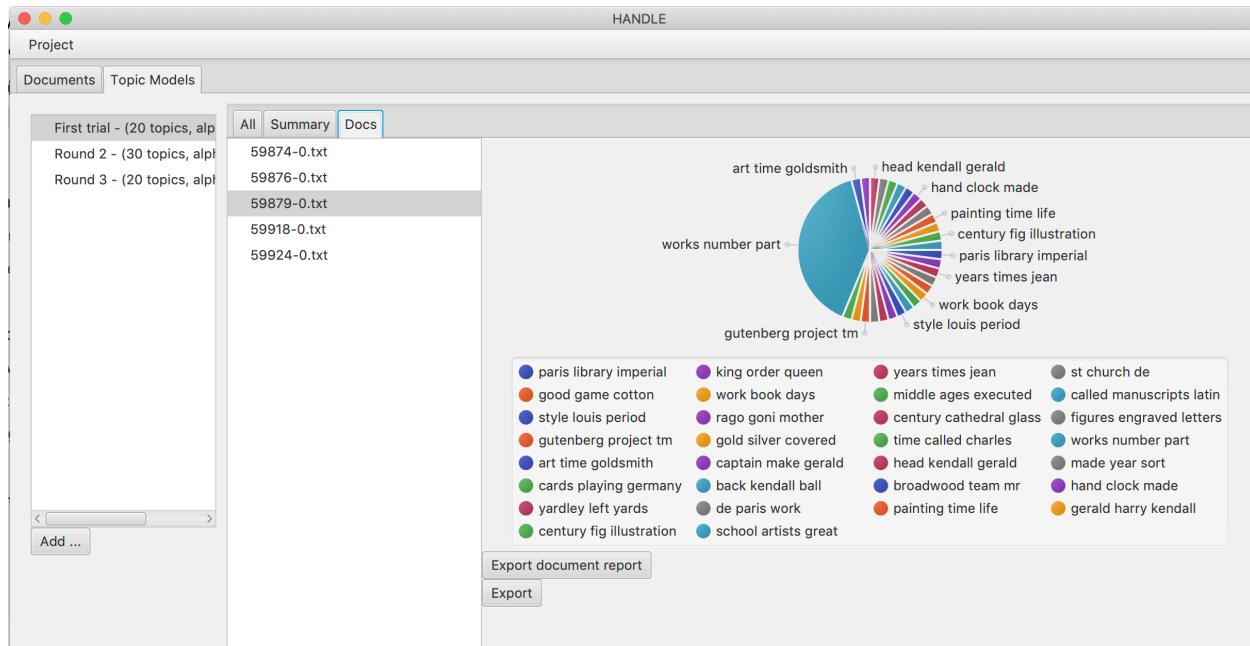


This pane also has three buttons:

- **Export model:** Save the model to a separate file to re-use later.
- **Export Topic Similarity Graph:** export a picture of the topic similarity graph to the left
- **Export Topic:** export a picture of the bar graph with a list of topics above

Docs Tab

This window allows looking at single documents.



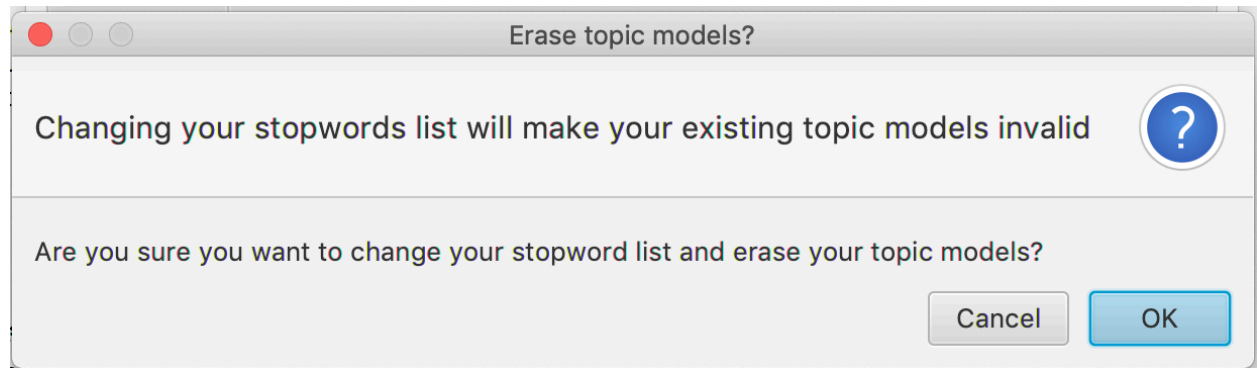
On the left is a list of all the documents. Clicking on a document shows a document composition graph: a pie chart of how that document is made up of each topic. For example, in the graph above, the biggest topic in the selected document is “works number part.”

Buttons:

- **Export document report:** save an Excel spreadsheet (.xlsx) with a report on the composition of each document, for experimentation in Excel, LibreOffice, or the like. This is similar to MALLET’s document-topics report.
- **Export:** save an image file of the selected document topic

Adding documents and changing stopwords after generating topic models

Once you create your first topic model, HANDLE will warn you if you try to add new documents or change the stopwords list. This is because you are modifying the parameters used to generate your topic models. You will get a warning like the one below:



Clicking "OK" will erase all the topic models, but Cancel will undo the previous action. Do this carefully because you cannot undo the previous action.