In this lesson you will first learn what *topic modeling* is and why you might want to employ it in your research. You will then learn how to install and work with the MALLET *natural language processing* toolkit to do so. MALLET involves modifying an *environment variable* (essentially, setting up a short-cut so that your computer always knows where to find the MALLET program) and working with the *command line* (ie, by typing in commands manually, rather than clicking on icons or menus). We will run the topic modeller on some example files, and look at the kinds of outputs that MALLET installed. This will give us a good idea of how it can be used on a corpus of texts to identify topics found in the documents without reading them individually.

First Paragraph from

Shawn Graham, Scott Weingart and Ian Milligan, "Getting Started with Topic Modeling and MALLET," *Programming Historian* (02 September 2012), http://programminghistorian.org/lessons/topic-modeling-and-mallet