

Topic Modeling: Mallet Installation

[Table of contents](#)

Windows Instructions	1
JAVA installation.....	1
Environment variable.....	1
Reboot	2
Mac Instructions.....	3
JAVA installation.....	3
File path with blanks	3
References	3

Windows Instructions

The project page of [MALLET](#) can be found here. You can also [download MALLET here](#).

JAVA installation

You will also need to install [Java developer's kit](#) as well.

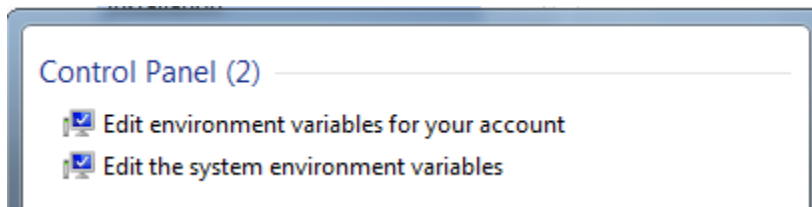
Environment variable

MALLET installation also requires modifying an environment variable (essentially, setting up a short-cut so that your computer always knows where to find the MALLET program).

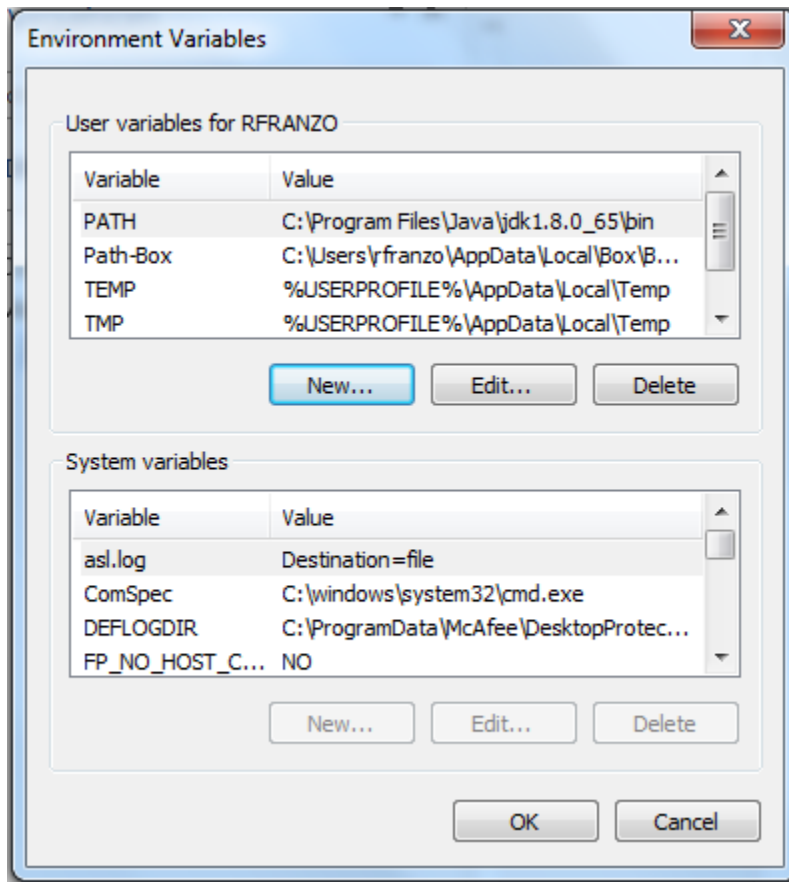
THE NLP SUITE SCRIPT topic modeling_mallet_main.py automatically sets up the required MALLET_HOME environment variable.

So... no need for you to setup the environment variable. But, if you really want, below is a description of how to setup manually the required MALLET_HOME environment variable.

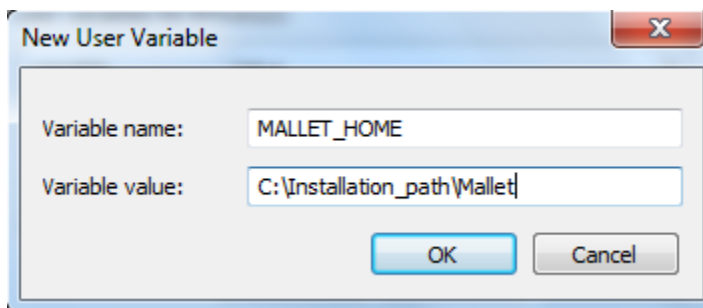
To modify environment variable, goto Start and enter environment in the search bar.



Click on the first entry and click on NEW.



Enter MALLET_HOME under Variable name and the full Mallet installation directory path (e.g., C:\Installation_path\Mallet) under Variable value. Then click OK. And OK again in the previous form.



Reboot

REBOOT THE MACHINE FOR THE ENVIRONMENTAL VARIABLES TO TAKE EFFECT

Mac Instructions

Comparing to Windows, the installation of MALLET on MacOS is similar yet easier.

The project page of [MALLET](#) can be found here. You can also [download MALLET here](#). You will also need to install [Java developer's kit](#) as well.

JAVA installation

You will also need to install [Java developer's kit](#) as well.

File path with blanks

Wherever you install Mallet, remember that Mallet will bomb if the full installation file path contains blanks. Thus, C:\Program files\Mallet will cause Mallet to fail. But not C:\Mallet Installation\Mallet.

The same is true for the directory where the TXT corpus files are stored. You do not need to have these files under the Mallet subdirectory. But wherever you store your TXT files, do not leave any blanks in the full directory path of your corpus files. Thus, C:\My Text Corpus\ will cause Mallet to fail. But not C:\MyTextCorpus\ or C:\My Text Corpus\

References

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2002. "Latent Dirichlet Allocation." *Advances in Neural Information Processing Systems*, 14.

Griffiths, T. L. and M. Steyvers. 2004. "Finding scientific topics". *Proceedings of the National Academy of Science*, 101, 5228-5235.

Sievert, Carson and Kenneth Shirley. 2014. "LDAvis: A Method for Visualizing and Interpreting Topics." Pp. 63–70 in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. Baltimore, Maryland, USA: Association for Computational Linguistics.

TIPS NLP Topic modeling Mallet.pdf

TIPS NLP Topic modeling Gensim.pdf

TIPS NLP Java download install run.pdf