This directory contains the scripts, data files, and output files from

analyzing the chat corpus by performs topic modeling utilizing LDA

(Latent Dirichlet Allocation).

Here we are used the LDA implementation from GitHub PyMallet at:

https://github.com/mimno/PyMallet

The LDA code is based on their lda\_reference.py code written in Python

The PyMallet project has an MIT License see below.

The two subdirectories contain:

1) PyMallet\_LDA\_Scripts - contains the scripts and necessary data files to

analyze chat corpusby performs topic modeling utilizing LDA

(Latent Dirichlet Allocation) transformation.

2) Output - topic files resulting from the LDA transformations.

Two types of text (.txt) files are included:

a) "raw\_" text (.txt) file listing topics with each word scored

(e.g., raw\_PyMallet\_LDA\_wholeChatsFile\_15topoics\_8words.txt)

b) "PyMallet\_LDA\_" text (.txt) file containing only the text for the

specified number of topics with the specified number of words

per topic (e.g., PyMallet\_LDA\_wholeChatsFile\_15topoics\_8words.txt)

MIT License

Copyright (c) 2019 mimno

Permission is hereby granted, free of charge, to any person obtaining a copy

of this software and associated documentation files (the "Software"), to deal

in the Software without restriction, including without limitation the rights

to use, copy, modify, merge, publish, distribute, sublicense, and/or sell

copies of the Software, and to permit persons to whom the Software is

furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all

copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR

IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,

FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE

AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER

LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,

OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE

SOFTWARE.