LDA python code

<https://shuyo.wordpress.com/2011/05/18/latent-dirichlet-allocation-in-python/>

Options:

-h, --help show this help message and exit

-f FILENAME corpus filename

-c CORPUS using range of Brown corpus' files(start:end)

--alpha=ALPHA parameter alpha

--beta=BETA parameter beta

-k K number of topics

-i ITERATION iteration count

-s smart initialize of parameters

--stopwords exclude stop words

--seed=SEED random seed

--df=DF threshold of document freaquency to cut words

note: download **'corpora/wordnet'** . import nltk and then did nltk.download().

cmd

e.g. python lda.py -c 0:20 -k 10 --alpha=0.5 --beta=0.5 -i 50

e.g. python lda.py -f ap.txt -k 10 --alpha=0.5 --beta=0.5 -i 50