Text

Description automatically generated

* Representation/grouping/training of word vectors
* GloVe:
  + <https://paperswithcode.com/method/glove>
  + Paper: <https://aclanthology.org/D14-1162.pdf>
  + Implementation: <https://nlp.stanford.edu/projects/glove/>
* FastText:
  + <https://paperswithcode.com/paper/efficient-estimation-of-word-representations>
  + Paper: <https://arxiv.org/pdf/1607.04606v2.pdf>
  + Implementation: <https://github.com/facebookresearch/fastText>
* Full process of visualization of glove’s vectors: <https://towardsdatascience.com/visualizing-word-embedding-with-pca-and-t-sne-961a692509f5>
* Walkthrough with word2vec: <https://www.kaggle.com/code/jeffd23/visualizing-word-vectors-with-t-sne/notebook>
  + <https://code.google.com/archive/p/word2vec/>

README.md  
  
Dependencies: (specified in requirements.txt- if you want to run this locally, you want to have these (the only one that seemed to matter in my testing was genism 3.8.3, because accessing the vocabulary of words works differently in 4.x)

1. Show glove script (in the zip folder)
   1. Setup.sh
   2. Creates “vectors.txt” > this is the trained word vector
   3. Move this to the same directory as “Word-Embedding-master”
   4. Run “train\_model.py” (this is in a different spot because training vectors takes a while)
      1. Creates a “glove2wordmodel.sav” using pickle
   5. Streamlit app.py > views that glove2wordmodel.sav file