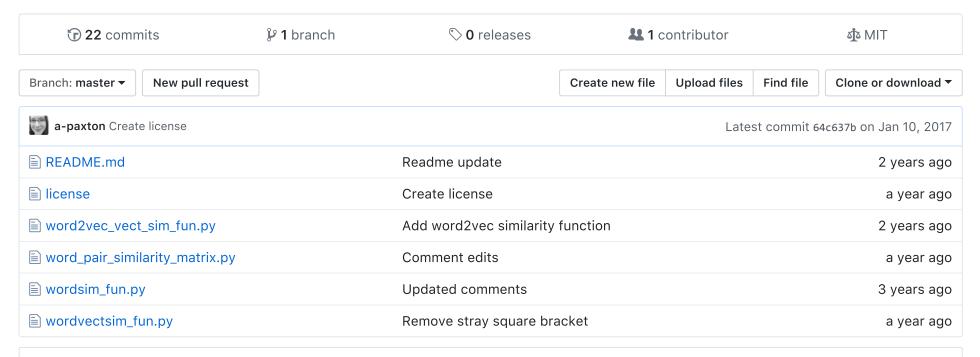
## a-paxton / Gensim-LSI-Word-Similarities

Generate word-word similarities from Gensim's latent semantic indexing (Python)

#word2vec #gensim #language #text-analysis #latent-semantic-indexing #latent-semantic-analysis



**README.md** 

## **Gensim-LSI-Word-Similarities**

Two simple little functions to create word-word similarities from Gensim's latent semantic indexing in Python. Both functions produce an inverted cosine similarity score (0 = low, 1 = high) between two words in a Gensim-generated LSA/LSI space across the total number of dimensions specified in the creation of the model (i.e., *num\_topics* from *gensim.models.LsiModel*).

Both require Gensim, Pandas, and SciPy.

## Includes four functions:

- wordsim: Create cosine-derived similarity score (from 0-1) between individual words. Input:
  - word1 (string or string variable)
  - word2 (string or string variable)
  - target\_dictionary (Gensim-created LSI dictionary)
  - target\_lsi\_model (Gensim-created LSI model)
- wordvectsim: Same as wordvect but created to calculate similarity scores (from 0-1) for word pairs in a 2-dimensional word vector (e.g., using numpy.apply\_along\_axis). Input:
  - word\_vector2d (2D string vector or 2D string vector variable)
  - target\_dictionary (Gensim-created LSI dictionary)
  - target\_Isi\_model (Gensim-created LSI model)
- Two additional functions/series of functions added (detailed documentation available in each function and will be added here soon):
  - word2vec\_vect\_sim\_fun: similarity score function for gensim's word2vec
  - o word\_pair\_similarity\_matrix: word-word similarity matrix function for gensim's LSI (LSA) model