Atitit nlp 词频统计法

s = SnowNLP([[u'这篇', u'文章'],

[u'那篇', u'论文'],

[u'这个']])print(s.tf) #TF意思是词频(Term Frequency)print(s.idf) #IDF意思是逆文本频率指数(Inverse Document Frequency)

print('文本相似：')

print(s.sim(['文章']))*# [0.3756070762985226, 0, 0]*

print(s.sim(['文章','真']))*# [0.7731414846187967, 0, 0*