## data -.mirror ← \$ rsync aleph.gutenberg.org::gutenberg - raw Header A long time ago - PG12345\_raw.txt in a galaxy -PG12346 raw.txt far, far away... Tail -text A long time ago - PG12345 text.txt in a galaxy - PG12346 text.txt far, far away... > nltk.tokenize.sent\_tokenizer > nltk.[...].TreebankWordTokenizer > string.isalpha() tokens > string.lower() long - PG12345\_tokens.txt time - PG12346 tokens.txt ago in counts the 3245 of 1554 - PG12345 counts.txt galaxy 123 - PG12346 counts.txt long 87 far 34