

# Project Proposal

## Story Generator

Norbert Logiewa  
nl253

### Project Description

Story Generator will be a system allowing to generate prose from an initial chunk of text. The system will ask the user for input and will try to complete the story using the input as the initial state. It will attempt to produce text in a style particular to stories i.e. it will ensure consistency in the use of past tense, atmosphere and use of language. The ultimate aim is that it generates text that resembles prose written by a human author.

### Technique

The mechanism for generating stories will rely on Markov Chains. The system will compute conditional probabilities of some word  $\alpha$  occurring based on the preceding  $n$  words from a corpus of text. The text (the corpus) will be provided to the system in advance and will be taken from Project Gutenberg which makes available to the public digital versions of writings for which copyright has expired.  $n$  will be a tweakable parameter which will allow so see, how the quality of text changes in relation to the amount of "lookbehind". The system will attempt to demonstrate how much can be achieved in terms of computational creativity with a limited amount of training data and a relatively simple algorithm.

### Supporting Software

The project will be written in Python and use the NLTK library which originated in the University of Pennsylvania. NLTK is released under the Apache License 2.0. The reasoning behind it is as follows: Python is concise, allows for quick prototyping, is very legible and makes common tasks (e.g. splitting strings into words) very easy. It also gives access to the said library.

### Evaluation

In terms of evaluation the system will try to quantify the quality of its outcomes by considering:

- linguistic diversity (creative use of language, richness of vocabulary)
- local coherence (if the language used, phrases make sense, are they syntactically correct)
- global coherence (if the story makes sense as a whole)
- interestingness (how much the story draws the reader in)
- structure (presence of an introduction, a conclusion, paragraphs, sentences, correct punctuation etc.)