**Storyboard for Module 1.3 – Basic Natural Language Processing with Python**

**Last revised: 26th June 2019**

**Roles:**

* Curriculum design:
* Story writer/resource finder:
* Jupyter notebook coder:
* Tester/rewriter:

**Time scale:**

* Target date for completion of Modules 1.1 to 1.4 is end of July 2019

**Working environment:**

* MS Notebooks + Slack ?

**Module 1 theme: Basic Python concepts for corpus processing (3 + 1 weeks in total)**

Objectives of Module 1: (i) Understand the basic concepts of Python programming using Jupyter Notebooks in the Microsoft Azure environment (‘*Azure Notebooks*’). This will include variables, collection types, loops, control structures and functions. Along the way students will learn to (ii) Write efficient regular expressions to solve text-based extraction tasks such as sentence segmentation, part-of-speech tagging and building a simple ELIZA-like chatbot; (iii) Apply the edit distance algorithm to text sequence problems; (iv) Work with corpus data to calculate statistics using loops, dictionaries and counting; (v) Consolidate understanding of commonly used evaluation metrics such as accuracy, precision, recall and F-score.

**Module 1.3 (week 2) – Basic Natural Language Processing with Python**

Number in brackets show estimated time to complete (needs testing). **Text in bold** highlights points of self-assessment or formal assessment. **Text in blue** corresponds to topics covered in the lectures.

Module 1.3 Week 2 (2 hours total)

**Short pre-module quiz** (<-- activate knowledge of regular expressions from the lecture and reading from files from Module 1.2)

Regular expressions

1. Searching text using basic regular expressions
2. Practice writing a regular expression for identifying dates
3. Substituting text
4. Build a simple Eliza chatbot
5. **Practice Quiz:** regular expressions (5 multiple choice questions that the students can use for formative assessment)

Functions

1. Writing functions (<-- e.g. using a recursive edit distance as the exemplar?)
2. Introduction to spaCy (<-- as a library of functions)
3. Practice using basic library functions (<-- e.g. using the spaCy to segment a sentence, tokenize, stem, lemmatize, do sentence boundary detection of a text and print out the results)
4. **Practice Quiz:** functions (5 multiple choice questions that the students can use for formative assessment)

**Quiz 2:** Basic Python assessment (3 questions to be handed in on Week 4)