

# NLP with Python 3: Topic visualisation in interactive charts

I am searching for experienced data scientist / machine learning engineer, who can help to build a set of data visualisations for textual corpora. I have 150 raw text files, which are digitized book copies of historical prints (18.–19. century) and I want to:

A.

1. build an interactive scatter plot that projects book into timeline and will render book metadata when hovering over the data point
2. will provide top keywords frequency histogram for each data point, or for the set of data point chosen interactively by the user (eg. user marks the area of the chart)
3. will provide topics extracted for each data point, or for the set of data points chosen by the user

B.

Run an ML algorithm that will identify clusters based on document similarity. Visualise outcomes of such analysis.

C. (to-be-confirmed)

Build a Theme River chart with keywords or topics change/frequency over time.

The outcomes should be presented in HTML page optimised for Chrome browser run on localhost.

For inspiration, here is the similar project for the Task A.: <http://campa-gupta.herokuapp.com/>

This is a high-level brief of the project, details needs to be discussed specifically.

**Project ID:** 19026723

