Step 2 – Test and Form Research Topic

* Usage/use cases of the gothic dataset and its original notebooks
  + Most common words (not very interesting or insightful)
    - Most of the common words meaningless
  + Build NN models based on raw texts
    - Predict / generate words
* My research needs / use cases
  + Test for their existing features of project / research areas recommendations
  + From a collection of papers / raw texts, discover cutting edge (new) topics that are worth digging / researching
    - If individual author:
      * Language pattern(s) shift along the author’s timeline?
      * Most common words / phrases / imageries that are significantly different from other authors?
        + Needs comparison dataset
        + Needs a lib of literary words / phrases / imageries (meaningful)
      * Sentiment analysis etc.
      * …?
    - If a corpus of a genre (e.g. Gothic):
      * Language pattern(s) shift along the historical timeline?
        + Emerging trends?
      * Most common words words / phrases / imageries that are significantly different from other genres?
        + Needs comparison dataset
        + Needs a lib / pre-trained model of literary words / phrases / imageries (meaningful)
      * Continue writing for unfinished work using NLP DL models
        + Or suggestions for new writing patterns in the contemporary era
    - If a corpus of literature / papers (realm of meta-analysis):
      * [TODO] A good paper on this approach!! <https://www.sciencedirect.com/science/article/pii/S2590159122000231>
      * Most common tags (if available)
      * Extract abstracts of these papers => raw-text dataset =>
        + Most common (research) topics

But how to determine? Should differentiate with just “tags”

[TODO] Would probably need a topic modelling / topic classification: <https://monkeylearn.com/blog/introduction-to-topic-modeling/>

* + - * Extract “future work” sections of these papers => raw-text dataset => make sense of the dataset
        + This strategy is good, but may not be useful to implement because usually the num of papers at this stage is small enough for human to process and should be processed by human too.