# SUBMISSION 2 PLANNING SECTION

# Weekly Planning Diary: 24/02/2020

## Theory-based Challenges

* Research **graph-based approaches** to semantic representation.
* Consolidate understanding of **vector-based representations**.
* **Compare** and **contrast** both approaches; deduce which is preferred.

## Practical Challenges (Improve my NLP General Knowledge)

* Build **a Password classifier**, then update my how-to guide with steps for making this using Streamlit.
* Build an **Automated** **Text Summariser** using Gensim and/or Sumy, then **update my how-to guide** with steps to reproduce it.

## Week 7

This is a pivotal week in the project; all research must be finished by Sunday night so that Week 8 can be dedicated totally to updating my blog and how-to guide. I’m conscious that the QCS Formal happens next week which knocks out nearly all of Thursday and Friday so my workload this week is heavier than usual. However, I’m more comfortable with my plan for this week than week 6 which felt like a massive struggle because I still hadn’t grasped the concept of formalising text properly. My confidence has been renewed this week because I’ve started properly adopting the zen-like mentality which was suggested after my suggestion 1 feedback; everything will come together sooner or later and everyone else is likely finding their research as difficult as mine.

Part of my plan is still to improve my general knowledge about NLP is no harm because I’m conscious that my employer will likely place me in an NLP role over summer so it’s worth my while understanding the overall field. John encourages developing that which is of lasting value to others; my employer’s product is its people therefore my enriched skillset will benefit clients this summer and will reduce the amount of training I will need to undergo.

### Researching Graph-based approaches to semantic representation

* I only discovered this concept towards the end of week 6. Therefore, I intend to meet this deliverable by reading the following papers and summarising the key points in each.
* My systematic approach for finding these papers will involve reviewing the agendas for large NLP conferences until I find someone who is giving a talk on it. I will then look at their list of publications and find other papers to read. Those papers are likely to cite other relevant papers so eventually, I should be able to build up a picture of what progress has been made.
* If I need any additional information, I will reach out to a researcher who has contributed to this field and ask for their guidance.

### Designing the Password Classifier

* Determining the strength of a password is an NLP task which doesn’t relate to the overall objective of my research. However, I feel that it’s important to continue building up a repository of code to show that I am continuously upskilling.
* The level of interactivity in this app is still quite high because my plan is for the user to be able to randomly generate or else manually write a password, then have its strength classified.
* This requires the Streamlit framework. I plan to include an article about this library in my troubleshooting page.

### Designing the Automated Text Summariser

* This NLP task at present isn’t centred around semantics because the ML process for picking the most important sentences uses the metric of how similar sentences are to each other when selecting the most important ones. Neither Gensim nor Sumy actually has contextual awareness and can give a reason based in common sense for why it selected the sentences it did.
* The level of interactivity in this app is high because the user can enter whatever body of text they want, select one of two summarisers and view the results. Note that Gensim and Sumy usually produce different results
* This requires the Streamlit framework. I plan to include an article about this library in my troubleshooting page.

## Week 8

This week doesn’t feel as learning intensive as last week as most of my time will be spent reviewing my progress and re-wording it to entice new readers. During my research, I found Medium blogs to be insightful during my 3 weeks of research. If I can produce a blog which **balances** **explaining** what I’ve found with **being concise**, I will have made something which will retain the attention of the reader. I’m approaching this blog as an attempt to publicise this field of research because at the moment, blogs like Lingthusiam don’t really approach representing text as an ML problem.

### Planning the Blog

* Consider the target audience: students around my age who are enthusiastic about either machine learning or linguistics. My objective therefore should be to think like Tom Scott if he were to write a blog as opposed to make a video. People in my target demographic typically watch him and can take away valuable understanding about the topic at hand.
* The bulk of the blog should be finished on Wednesday night. It’s unlikely that I’ll get any work done on Thursday due to personal reasons however, Friday’s 2 hours of work will be totally dedicated to proofreading it.

### Planning the Videoconference

* Ask open-ended questions as semantic representation is an area which is unlikely to have been explored in industry before. Given how her PhD was awarded for designing a text-weighting system for graph-based representations, it’s highly likely that she can recommend other experts and their papers.
* John advised me to be conscious that experts tend to resort to jargon if they are faced with an area of their specialism which they have not seen before.

### A Last-minute change of plan!

* As luck would have it, Dr. Shanavas emailed me 2 hours before the planned videoconference asking to reschedule for Thursday at 12pm. I had already made personal commitments relating to the Queen’s Computing Society Formal and thus, will have to include her commentary in my final submission.
* I feel that this is a huge setback because it removes one of the USPs of my blog (personal insight from a figure of authority) but I’ll just have to include something else that’s less personal but just as insightful.
* Therefore, I plan to talk about my vision for the end of the project instead. I also noticed that there is an application similar to my Everest in John’s github repository, so I plan to expand my learning log with talk about that.

Planning for the future

Despite this week not going at all as planned, I will get back on track next week by conducting the videoconference and beginning to apply all the NLP programming theory I’ve acquired in order to start work on a supervised model which will produce a visual simulation of the meaning behind a sentence.