

# Final Year Project Proposal

Braddy Yeoh  
17357376  
braddy.yeoh@ucdconnect.ie

## Title of Final Year Project:

Automated Literature Review

## Aims and Objective:

The goal of the project is to find relevant literature on any topic (e.g., Covid-19 vaccines) automatically. This will require combining the crawling of the citation graph of publications and using document similarity and basic natural language processing (NLP) techniques to determine the relevance of different papers. The project will use simple machine learning techniques to combine the different features from NLP and graph search. There will also be considerable software development involved, as the process of finding the relevant literature is likely to be iterative -- The system recommends some papers, the user may indicate whether they are relevant/irrelevant and then the system updates its recommendations and so on.

## Background and Justification:

As we become more connected than ever, there is a vast amount of information available, and manually going through each available article, link or blog becomes tiresome. It also may be the case that the article is irrelevant. This project will resemble something like Google's PageRank (PR). We allow the user to grade the relevance of the recommended paper, and the system will update its recommendations.

When I was researching a topic area of a project I am interested in, I noticed that some citations contained useful or relevant information that I was looking for. The reason for this project is further justified by the fact that the same cited paper appeared in Google search further down. This project serves to be very useful as it automates the searching of relevant information when carrying out research which saves a lot of time.

## Methodology

---

### Mandatory:

1. The building of a document crawler.
2. The building of an NLP model to determine relevance of different papers.

3. The building of a document similarity algorithm to determine relevance of different papers.
4. A user interface that allows user to insert a paper, displays the recommended papers and indicates the relevance of the recommended article.

---

### Discretionary:

1. A graphical graph that shows how the documents are connected.

---

### Exceptional:

1. Deployment of this project to a serverless full stack application.

## Data Sources:

The data source will consist of the papers crawled.

## Related Work:

2. [Academia - Web Crawlers](#)
3. [Academia - Web Crawlers](#)
4. [Rsearch Gate - Web crawler research methodology](#)
5. [Arxiv - A Search/Crawl Framework for Automatically Acquiring Scientific Documents](#)
6. [PromptCloud - Web Crawling in Scientific Research Bigger Breakthroughs](#)
7. [iopscience - Summary of web crawler technology research](#)
8. [SSRN - Obtaining Data from the Internet: A Guide to Data Crawling in Management Research](#)
9. [Hindawi - An Efficient Approach for Web Indexing of Big Data through Hyperlinks in Web Crawling](#)
10. [Hindawi - An Improved Focused Crawler: Using Web Page Classification and Link Priority Evaluation](#)
11. [GitHub - 100 must read NLP papers](#)
12. [Papers With Code - Natural Language Processing](#)
13. [Towards Data Science - BERT Explained: State of the art language model for NLP](#)
14. [https://en.wikipedia.org/wiki/BERT\\_\(language\\_model\)](https://en.wikipedia.org/wiki/BERT_(language_model))

## Bibliography:

The sources I have used to research this problem are the following:

1. [Wikipedia Vehicle Routing Problem.](#)
2. [Wikipedia - Combinatorial Optimization](#)

3. [Wikipedia - NP Complete Problems](#)
4. [Research Gate - Learning Fine Grained Search Space Pruning and Heuristics for Combinatorial Optimization](#)
5. [Cornell - Learning Combinatorial Optimization Algorithms over Graphs](#)
6. [Medium - Machine Learning and Combinatorial Optimization Problems](#)
7. [Papers With Code - Combinatorial Optimization](#)
8. [Cornell - Reinforcement Learning for Solving the Vehicle Routing Problem](#)
9. [Cornell - Attention, Learn to Solve Routing Problems!](#)
10. [Research Gate - A Survey on the Vehicle Routing Problem and its Variants](#)
11. [Research Gate - Vehicle Routing Problem with Time Windows](#)
12. [Science Direct - Approaches to Solve the Vehicle Routing Problem in the Valuables Delivery Domain](#)
13. [Elsevier - An Optimization Model for the Vehicle Routing Problem in Multi-product Frozen Food Delivery](#)
14. [Cornell - A Deep Reinforcement Learning Algorithm Using Dynamic Attention Model for Vehicle Routing Problems](#)
15. [Cornell - Scalability of using Restricted Boltzmann Machines for Combinatorial Optimization](#)
16. [Hindawi - Machine Learning-Based Parameter Tuned Genetic Algorithm for Energy Minimizing Vehicle Routing Problem](#)
17. [Semantic Scholar - Pointer Networks](#)
18. [Cornell - Neural Combinatorial Optimization with Reinforcement Learning](#)
19. [Academia - Vehicle Routing Problems](#)