School of Computing  
CA326 Year 3 Project Proposal Form

**SECTION A**

Project Title Automated Document Summarization using Natural Language Processing

Student 1 Name Niall Egan ID Number 19378906

Student 2 Name John Mccormack ID Number 19517396

Student 3 Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ID Number \_\_\_\_\_\_\_\_\_\_\_

*(A third team member is exceptional and requires detailed justification.)*

Staff Member Consulted for supervision Ray Walshe

Project Description (2 pages max)

**Introduction**

The area our project will be covering is Natural Language Processing. Specifically, it will be based around automatically processing pdfs and summarizing the information.

**Outline**

Our software will automatically summarize documents. Using Natural Language Processing to find the what the document is about and then summarize the important information.

**Background**

The initial idea came from a lecturer’s project ideas. The initial idea was to visualize the search tree for search algorithms like Breadth-First Search. However, after discussion with our supervisor, we agreed this idea was too simple. Through his guidance and our own research, we started looking more into Natural Language Processing. NLP seemed like a much more interesting topic to cover. Visualization didn’t seem like it would be the best use of NLP’s strengths, so we decided to move to the idea of document summarization.

**Goals**

We believe our idea could make the browsing of technical documentation significantly easier. It could provide summaries for documents so anyone browsing the documentation could tell at a glance if something is relevant to their search.

**Programming language(s) and tools**

Language: Python 3

Libraries: PyPDF, Natural Language Toolkit, Gemsim and Word Cloud

Operating System: Windows 10

IDE: PyCharm

**Breakdown of work**

**Student 1 – Niall**

I will work on implementing the Natural Language Processing. This will involve reading in text, processing it into tokens and handling the algorithm that does the auto-summarization

**Student 2 – John**

I will work on implementing the software’s interfaces. This will involve creating a User interface that will except input from a user and cleanly present the summary and other data back. It will also involve creating an interface that will pass the summaries to other software.