



School of Computing
Year 4 Project Proposal Form

Project Title: **Q&A Extraction from Factual Text**

Student Name: **Traian Svinti**

Student ID: **1443218**

Stream: **CASE4**

Project Supervisor: **Yvette Graham**

Proposal Description

General area covered by project:

The aim of this project is to alleviate the time that students and teachers consume when creating questions based on text where questions may not be available elsewhere. Question generation from factual text can be implemented through Natural Language Processing (NLP). This webapp will take text as input and process this text to extract questions based on the main ideas of the text. This can help teachers to create questions for their students or students to quiz themselves on their knowledge of a text. NLP will be the main area covered by my project.

Outline of the proposed project:

Background

After an abundance of research for a project idea which I was not just interested in, but I believe to be challenging yet executable in the time frame given; I came across a previous project that related to this. I aim to implement this with extra features that will help students and teachers in acquiring knowledge. I believe this to be a suitable project for final year, giving way to many extra features and real world applications.

I would like to implement firstly, the NLP system applying syntactic and semantic analysis (using statistical NLP techniques) to be used and once this is working, can be followed by a UI and other possible tools that can add to this. I believe that this product can be implemented in multiple environments along with a possible question/answer webapp for tests, keeping records of score achieved in a database and this being visible to a teacher all packed in a nice UI. Also another feature that is in scope is a text extractor implementing OCR techniques. This will take input a scan of a textbook, text can be extracted and that could then be uploaded for question extraction.

Achievements

This project aims to achieve an easy to use UI and robust and accurate processing that both students and teachers alike can use. It will provide an interface that can efficiently extract questions and related answers from a piece of text and provide these questions in the webapp to be answered or downloaded so it can be printed for class use.

Justification

- Why - This will be used as a more efficient means of extracting questions from a text, time is valuable to both teachers and students; who could be potentially studying for an exam. Also, being able to access your previous results (if answers were inputted in the app) can help in locating areas that need improvement.
- When - This would be used at any time, but I believe it will be mainly used before a test. Either the teacher is creating the test through this product or a student is studying for a test or if someone wants to examine their general knowledge on a certain piece of information.
- Where - This product will be used in educational environments.
- How - Being able to automatically generate questions from a potentially large piece of text can save time and answering the questions can provide valid feedback to areas that may need further studying.

Programming Languages:

I intend to use a multitude of different languages including, but not limited to (depending on potential new information gained) - Python, Javascript.

Programming Tools / Tech Stack:

I intend to use a multitude of different tools including, but not limited to (depending on potential new information gained) - Django, AWS, HTML5, CSS, Bootstrap, SQLite, NLTK (Natural Language Toolkit), OCR (Optical Character Recognition).

Learning Challenges:

I believe this project will be a challenge due to the fact that I have never looked into the area of NLP and by doing so, I will learn a lot of new techniques which I have not previously been exposed to. As this is a very large area, there are a lot of tools and techniques that can be used to achieve the final goal of this project.

I have also never deployed a full product to a server (AWS) which I intend to do. This will help me with my knowledge of deploys, maintenance and other areas associated with this.

Hardware / Software Platform:

I will be solely using Ubuntu 18.04 LTS (latest version of Ubuntu) for both my development and server operating system. Atom will be the main text editor used for creating and modifying code. Django will incorporate with SQLite database for storing records of users, and their past upload text pieces and questions/answers.