COMP3009J – Information Retrieval
Final Assessment

This assignment is worth **50% of the final grade** for the module.
Due Date: Friday 24 June at 23:59 (i.e. end of Week 18)

Russell and Becker LLP (known as "R&B") are an Irish law firm that were founded in 1916. Since then they have accumulated over 1 million legal documents, including letters, laws, judgments, contracts, affidavits, wills, etc. They have recently completed a process of scanning all of their old paper-based documentation into PDF documents, and they would like an Information Retrieval system to search their document archive.

Their workers are familiar with traditional search engines, and so they would like to be able to use keyword-based queries. These keywords might be names of clients or businesses, document types, or relate to some other legal information need. In the legal business, it is important that they do not overlook an important document when they are doing research. All of their documents are written in English.

Task:

- You are the head of a software engineering team and you have been asked to design an Information Retrieval system for R&B.
- Based on what you have learned in COMP3009J and any outside material you may have read, you are asked to write a report to describe this system.
- Your design should say what the components of the system will be, and you should explain why you have chosen any techniques it includes.
- You should also describe how you will evaluate the system's effectiveness. You will need to show the partners of R&B that the system works well, or they may refuse to pay for the system.

Submission:

- You should submit a single PDF file to Brightspace.
- Your report should be approximately 2,000 words in length (titles and references are not considered part of the word count).

Other Notes:

- This is an *individual* assignment. Anything you submit must be your own work.
- You are not required to cite any information that you have learned from the COMP3009J lectures. However, you should correctly cite any other sources that you rely on.
- Do not copy and paste directly from lecture notes or other sources.
- Do not use direct translations of non-English sources.
- Submitting the work of somebody else is **plagiarism**, which is a serious academic offence. Be familiar with the UCD Plagiarism Policy and the UCD School of Computer Science Plagiarism Policy.
- If you have questions about what is or is not plagiarism, ask.

Grading Guide

The following acts as a guide to show typical examples of different grade levels. Minor grades (i.e. the difference between B+, B, B-) will be awarded as appropriate.

A grades: Excellent solution, which demonstrates a clear grasp of all aspects of the problem. The system design is portrayed clearly, the justifications are excellently reasoned and the evaluation approach is realistic. There is some evidence of knowledge from beyond the core lecture content.
B grades: Very good solution, which demonstrates a clear grasp of the problem. The system design is portrayed clearly, the justifications are well reasoned and there is a reasonable evaluation approach.
C grades: Good solution, which shows that the student has a reasonable understanding of the requirements of the system. An effective Information Retrieval system has been proposed, and suitable evaluation metrics have been identified. At least a basic justification has been given for the most important aspects of the system.
D grades: A system has been proposed that satisfies the basic requirements of an Information Retrieval system. The justifications and explanations may be weak, or the solution may have some substantial flaws in practice.
E grades: Some work has been attempted, but there are serious errors in understanding or relevance.
F grades: Some evidence of work attempted, but little (if any) is of relevance to solving the given problem.
G grades: Wholly unacceptable, little or no evidence of meaningful work attempted.
NM grade: No relevant work attempted.
ABS grade: No submission received.