

**School of Computer Science and Statistics**

**Assessment Submission Form**

|  |  |
| --- | --- |
| **Student Name** | GEORGE CHAVADY |
| **Student ID Number** | 19305272 |
| **Course Title** | M.Sc. COMPUTER SCIENCE – INTELLIGENT SYSTEMS |
| **Module Title** | CS7IS3 – INFORMATION RETRIEVAL AND WEB SEARCH |
| **Lecturer(s)** | GARY MUNNELLY, JOERAN BEEL, OWEN CONLAN |
| **Assessment Title** | Assignment 1 – Lucene and Cranfield |
| **Date Submitted** | 28-2-2020 |
| **Word Count** |  |

**I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at:** [**http://www.tcd.ie/calendar**](http://www.tcd.ie/calendar)

**I have also completed the Online Tutorial on avoiding plagiarism ‘Ready, Steady, Write’, located at** [**http://tcd-ie.libguides.com/plagiarism/ready-steady-write**](http://tcd-ie.libguides.com/plagiarism/ready-steady-write)

**I declare that the assignment being submitted represents my own work and has not been taken from the work of others save where appropriately referenced in the body of the assignment.**

Signed:  Date: 28-2-2020

**Public DNS**: ec2-3-80-189-102.compute-1.amazonaws.com

**On blackboard, submit:**

A copy of your two page report

* Instructions on how to access your AWS instance:
* Don't forget to tell us your instance's IP address
* Any login credentials that we may need:
* Username/Password (if required)
* .pem file (if required)

**Things to do:**

* You need to select appropriate Lucene Analyzers for content processing
  + i.e. tokeniser, stop-word removal, stemming, etc.
* Download and index the Cranfield Collection - http://ir.dcs.gla.ac.uk/resources/test\_collections/cran/
  + a collection of ~1400 documents (short abstracts), 225 queries and graded (1-5) relevance judgements
* Implement and test different scoring approaches in Lucene
  + including at least the Vector Space Model and BM25
* Test your search engine using the queries and relevance judgements provided with the Cranfield Collection
  + Generate Mean Average Precision and Recall scores based upon these relevance judgements
  + These metrics, among many others, should be generated using TREC\_eval
* Demonstrate your functioning search engine
  + Achieved by TA access to the AWS secure install
* Write a short report (Max 2 pages) which describes your implementation. This should describe your implementation, explain your choice of analysers and scoring, report the performance of your search engine etc.
* Format is two columns in 10pt font