

Dublin City University
School of Computing
CA4009: Search Technologies
Continuous Assessment: Research and Development Project

Total CA Weighting: 15%

Assessment of this CA consists of **two** components:

Presentation: 20% of total CA mark

It is planned to hold these in Week 9/10. A timetable will be distributed.

Written report: 80% of total CA mark

Submission Deadline: Friday 1st December 2023 (Week 12). Please note that a penalty will be applied for late submission without good reason.

Reports should be submitted electronically via the loop page for CA4009.

1 Aims

The aim of this assignment is to provide you with an opportunity to examine some of the topics covered in CA4009 *Search Technologies* in greater detail.

In completing the assignment you should demonstrate effective use of material taken from the module lecture notes and reading beyond these. Your reading can be from any source including books, journals and proceedings of research conferences from the library and material from the WWW. The DCU library gives online access to most of the major science publishers' journals and conference publications via IReL database links¹.

The analysis of the techniques in your report should seek to identify aspects which you consider to be their strengths and weaknesses, and, if possible, you should develop ideas for extensions to these existing methods which you believe may be worthy of investigation.

2 Management Arrangements

This is a group assignment to be carried out in groups of **four**, please let me know if you would prefer to work in a smaller group or on your own.

Please upload details of your group members and proposed project title to loop by 5.00pm Friday 3rd November 2023.

¹<https://www.dcu.ie/library/z-library-databases>. For off campus access sign in using your DCU login details when requested.

3 The Assignment

For the assignment you are required to create a proposal for the development of a new search application. To complete the assignment you need to perform an analysis and develop a technical specification including an implementation plan and an evaluation plan for your proposed application. This should be a novel application bringing together material from topics studied in the *CA4009 Search Technologies* module.

Your proposed system should be based on *existing* established techniques and current *ongoing* research in search and information retrieval and related applications. Thus it should be possible to build and evaluate a laboratory prototype of your system based on your specification document and to do this now, not at some point in the future. Essentially, your document should be complete and clear enough to enable someone else taking CA4009 to use your document to develop and evaluate your proposed system.

There are many possibilities for new search applications based on the technologies covered in the module. Some ideas are:

- Personalization and Adaptation in Information Retrieval
- Exploiting information gathered from users in Information Retrieval
- Multimedia Question-Answering
- Video Summarization
- Multimedia Enterprise Search
- Social Network Search
- Novel use of Markup and Metadata for Information Retrieval
- Mobile Search
- Personal Information Archive Search
- Application of Large Language Models (LMMs) in information retrieval

or any other relevant topic of your choice.

Credit will be given for proposals exploring emerging technologies which pose new search challenges or offering potential solutions to problems relating to these technologies, and for inclusion and discussion of references to latest research findings in search technologies, e.g. making reference to articles, papers and books published in the last two years.

If you have an idea for a new search application and are not sure if it is suitable please feel free to contact me via email outlining your idea, and I will let you know if it is a suitable topic for your project.

4 Assessment

The assessment will consist of two elements: a short oral presentation and a written report.

4.1 Oral Presentation

You should prepare a short presentation summarising the details of your proposed search application covering all the topics in the written report. Presentations will be allocated 5 minutes for each group with an additional couple of minutes for discussion and feedback. It is planned to hold presentations in week 9/10.

Presentations will be assessed based on your proposal and against the detailed criteria below for the final report, but they are also an opportunity for you to receive feedback and ask questions about your proposal. You can then incorporate this feedback into your final project report.

One member of each team upload a copy of your slides to loop after your presentation before the report deadline of Friday 1st December 2023.

4.2 Written Report

Your written submission should be in the form of a functional specification including an implementation plan including an evaluation plan, for a research project to investigate your proposed system. Your report **must** include the following sections:

Front page:

- Your names, student ID numbers, email addresses, programme, module code, and date of submission.
- The following disclaimer:

A report submitted to Dublin City University, School of Computing for module *CA4009: Search Technologies*, 2023/2024.

I understand that the University regards breaches of academic integrity and plagiarism as grave and serious.

I have read and understood the DCU Academic Integrity and Plagiarism Policy. I accept the penalties that may be imposed should I engage in practice or practices that breach this policy.

I have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations, paraphrasing, discussion of ideas from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the sources cited are identified in the assignment references.

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work. My submission makes no use of any generative AI methods in its creation.

By signing this form or by submitting this material online I confirm that this assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study. By signing this form or by submitting material for assessment online I confirm that I have read and understood DCU Academic Integrity and Plagiarism Policy (available at: <http://www.dcu.ie/registry/examinations/index.shtml>)

Name(s):

Date:

Main report:

Your report **must** follow the following structure:

- A concise meaningful **Title** describing the topic to be addressed in your report.
- Your names, student ID numbers, email addresses, programme, module code, and date of submission.
- A short (200 words maximum) **Abstract** overviewing the contents of your report. It should be possible for someone with a background in search technologies to read and understand this abstract without needing to make reference to the report itself, and to have a clear idea from reading the abstract of the topic explored in your report and the purpose and functionality of your proposed system.
- The main body of the report should include numbered sections (and where you feel appropriate sub-sections) as follows:
 - An **Introduction** giving an overview of the background, motivation and objectives of your proposed system. Include details of what need or opportunity the application will address, what the system will do, how it will do it and what functional or technical constraints or limitations your system will have.
 - A **User Analysis** considering one or more operational scenarios for your system, and the characteristics of the users who you expect to be using the system. For example, will users of the system have particular knowledge of the topic area which they are searching, will they be using your system for a specialised task, or are they members of the general public with no specialised knowledge. Your analysis should also consider the types of information needs and search queries (broad or narrow, expert or general) that you would expect from the user, and what this means in terms of the types of queries which your system must be designed to be able to process. You should also include an example of the flow of a search operation from the stage of the user information need to the user interacting with and interpreting retrieved results.
 - A **Scientific Functional Description** of the system. Give details of the **overall architecture** and the **algorithms** to be used with citations to where the methods you are using were first published or are described or how they relate to your proposed methods if they are new, so that it is clear what components are contained in your system and how they are combined into the overall system. Include an analysis of the pros and cons of your design in terms of the suitability of the technologies for both the overall system and individual components. In this you should identify limitations of what they can do, or assumptions you are making about their operational behaviour that you are making in your design, e.g. about the data, about the algorithms, about the system hardware, about the user, about the user's tasks.
 - **Evaluation** plan. Evaluating the retrieval behaviour of search systems is a key part of testing their effectiveness to address the purpose for which they are being developed, and inclusion of a suitable evaluation plan is a crucial part of this research and development plan of a search system and this report. Your evaluation plan should include details of your overall evaluation strategy, including the overall objective, data that you propose to use, the collection of test suitable search queries, criteria for determining content relevance, how relevance data will be collected, and the metrics which will be used for evaluation, explaining why they are appropriate to measure the effectiveness of your system.
 - **Concluding Section** giving a brief overview summary of the contents of your report, and listing next steps which could be taken towards developing your proposed system.

- A list of the **References** used in your research giving **full** publication details. You should cite your references at relevant points throughout your report. For each one you should aim to include the following details as appropriate in the reference list: author(s), title, year of publication, conference/journal, pages, volume no, geographical venue of a conference publication. Publications should be listed either in alphabetic order of first author or your first citation of them in your report.

You can see how references are included in scientific papers by consulting publications from the major publishers of computing research such as the ACM, IEEE and Springer. Many examples of papers from these organisations are easily available online through the DCU library via IReL, as introduced earlier, via the “IReL” link. Or similar papers can be found online by using a search engine to look for topics in search and information retrieval.

Simply giving the Web address from which a document was downloaded or from which some software is available is NOT sufficient. You need to give full details of the document or software that is available for download from this address. Also you should give the date on which you accessed this web address to view the item that you are citing as evidence of the content of source the you are referencing when you consulted it.

Your report should be presented so as to make the material as easily accessible to the reader as possible. Thus you should use an appropriate combination of presentation formats, e.g. short sections of prose, bullet points of key points, clearly labelled diagrams, equations, pseudo code, and possibly other formats.

Your report should include no more than 3000 words, excluding the abstract and reference list. The number of pages will depend on the number and size of diagrams, lists, etc, but should typically be no longer than 15 - 18 pages.

You should NOT include any code as part of your submission. This is an analysis, design and specification assignment, not an implementation task.

Unacknowledged copying of text and diagrams is plagiarism. The disclaimer at the head of your report states that you are aware of the need to acknowledge material included in your report which is not original.

Marks will be deducted for failure to adhere to the structure described above and not to include all the sections and information required.

Marking Criteria

The following can be said of a perfect submission:

“You have written an excellent and well structured report. The writing and other presentations in the report are clear and concise. You have provided an excellent description of the motivation for your system and its design, with one or more examples of a situation in which it could support a user in addressing their information need. You have included an informative analysis of the expected users of your system in terms of their subject knowledge, search experience and ability to engage with your system. You show a good awareness of the state-of-the-art in search technologies, and your system includes interesting novel features which cannot be found in existing literature. You have provided a clear description of how your system could be implemented with currently available search and information retrieval technologies, and included a clear evaluation plan that would enable the effectiveness of your system to be tested. You have provided full reference details of your sources and you show clear evidence of reading and analysis beyond material provided in lectures. You have shown excellent judgment in your choice of relevant material from your sources. You have compared and contrasted the material from your reading, and commented thoughtfully on aspects which you feel to be strengths and weaknesses of the techniques used.”

The following gives the criteria for assessment against this ideal:

Grade I ($\geq 70\%$) : You have written a very good and well structured report. You have clearly described the motivation and specification of your system which includes some elements of novelty. You have provided a clear analysis of expected users, and their engagement with the system. Your analysis and design are well integrated, and include up to date material beyond that covered in lectures. You have described a complete implementation of your system based on existing information retrieval technologies and a full evaluation plan. You have compared and contrasted techniques clearly, and displayed insights into their relative merits.

Grade II(i) (60 – 69%) : Your report is good but the structure may not be smooth or some points may not be well explained. You have shown good evidence of limited background reading which may not represent the latest findings in search technologies. You have described a full implementation of your system with an adequate evaluation plan. You have examined the users of your system, and made a good basic attempt to critically analyze the methods covered in your report.

Grade II(ii) (50 – 59%) : Your report is lacking in some areas or the flow of ideas may not be completely logical. You have shown evidence of some background reading. The user analysis, implementation and evaluation plan are reasonable, but may not be complete. There is limited analysis of material or the description may be unclear.

Grade III (40 – 49%) : You have written a complete report, but it has some weaknesses in its structure. You show very little sign of background reading and your report is limited and difficult to follow. The user analysis, implementation and evaluation plan have significant weaknesses.

Grade Fail ($< 40\%$) : Your report is incomplete. There is no evidence of background reading. Your presentation is poor and difficult to follow, and significant required elements are inadequate or missing.