Department	School of Computing, Engineering & the Built Environment
Supervisors	Dr Md Zia Ullah and Prof Dimitra Gkatzia
Project Title	Query Performance Prediction for Sparse and Dense Retrieval Models

PROJECT DESCRIPTION

In the field of information retrieval (IR), query performance prediction (QPP) aims to predict the effectiveness of a system for a given search query without resorting to relevance judgments. QPP has various advantages, such as informing an IR system whether a given search query would be effective or ineffective. Based on the effectiveness of the query, the system can either apply a query reformulation or an adaptive retrieval configuration [1,4] or engage in an interactive session with the user (i.e., conversational IR [5]) to understand the search intent and provide a better search result.

Accurate prediction of query performance is a challenging problem. Existing QPPs are applied over sparse retrieval models (e.g., BM25 or Language model) using the pre-retrieval features based on the collection statistics or the post-retrieval features based on the top-retrieved documents [2, 3]. Due to the advent of the language model (e.g., BERT [6]), dense retrieval models have been proposed and shown to have better retrieval effectiveness. However, QPP on the dense retrieval model has not yet been explored.

This Ph.D. project aims to develop query performance predictors for spare and dense retrieval models. Experiments could be conducted on standard TREC collections (e.g., MS MARCO and TREC Deep learning tracks) to demonstrate the effectiveness of the developed QPPs and compare them with the state-of-the-art approaches. In addition, another goal of this project would be to experiment and evaluate QPPs on question-answering collections in a conversational IR context [5].

Prospective applicants are encouraged to contact the Supervisor before submitting their applications. Applications should make it clear the project you are applying for and the name of the supervisor(s).

Academic qualifications
A first degree (at least a 2.1) ideally in Computer Science or Data Science with a good fundamental knowledge of Information retrieval (IR), Natural language processing (NLP), and Machine learning.

English language requirement
IELTS score must be at least 6.5 (with not less than 6.0 in each of the four components). Other, equivalent qualifications will be accepted. Full details of the University's policy are available online.

Essential attributes:
- Experience of fundamental knowledge of IR, NLP, Language models, and Machine learning.
- Competent in R, Python, Java, PyTorch, Shell scripting
- Knowledge of Information Retrieval and Deep Learning.
- Good written and oral communication skills
- Strong motivation, with evidence of independent research skills relevant to the project
- Good time management

Desirable attributes:
- Experience of IR Tools, such as Terrier IR, Lemur Indri, Lucene.

Indicative Bibliography	1. Mothe J and Ullah MZ, Defining an Optimal Configuration Set for Selective Search Strategy – A Risk-Sensitive Approach, Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM 2021), 2021. 2. Déjean S, Ionescu RT, Mothe J, and Ullah MZ, Forward and backward feature selection for query performance prediction, The 35th ACM/SIGAPP Symposium on Applied Computing (SAC), 2020. 3. Chifu AG, Laporte L, Mothe J, and Ullah MZ, Query Performance Prediction Focused on Summarized Letor Features, The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2018. 4. Deveau R, Mothe J, Ullah MZ, Nie JY, Learning to Adaptively Rank Document Retrieval System Configurations, ACM Transactions of Information Systems (ACM TOIS), 41 pages, pp.3:1-3:41, Volume 37, Issue 1, 2019. 5. Dalton J, Xiong C, Kumar V, and Callan J, CAsT-19: A Dataset for Conversational Information Seeking, Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1985--1988, 2020. 6. Devlin, Jacob and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805, 2018.
Enquiries	For informal enquiries about this PhD project, please contact m.ullah@napier.ac.uk
Web page	https://www.napier.ac.uk/research-and-innovation/research-degrees/application-process