**Supervisor: Jian-Yun Nie, Full Professor**

http://rali.iro.umontreal.ca/nie/jian-yun-nie-en/

**Project 1: Information retrieval**

Information retrieval (search engine) aims to retrieve documents that contain relevant information to a search query. Many techniques have been developed, ranging from the traditional techniques based on words, to more recent techniques based on deep learning. While the techniques based on deep learning have shown interesting results, they still have difficulty to outperform the traditional methods in general. The goal of our research in this area is to better understand what deep learning can do and what benefits it can bring to IR. We are experimenting different deep learning models for IR in the research group. For example, we have proposed a deep learning model integrating multi-level matching scores. We are also investigating new document and query representations for IR. These attempts target the specific problem in IR, that different types of query (e.g. broad or specific) would require different matching functions and representations.

This project will allow the student to get familiar with the domain of IR. He/she will be asked to propose a method in concertation with the supervisor and other graduate students, and to implement and test the proposed approach on some test collections.

In addition to deep learning based models, we are also interested in other IR problems such as medical information retrieval, cross-language information retrieval, query suggestion/expansion, social media mining, and so on. Any related project could be of interest.

**Project 2: Question-answering**

The task of question-answering (QA) is to find the right answer(s) to a human question. Question-answering can be investigated in several settings:

- QA from a knowledge graph: it is assumed that the answer exists in a knowledge graph (e.g. Freebase, ConceptNet, etc.), and the goal is to find it out for the given question.
- QA from texts: The task is to find the answer from a large set of texts (e.g. webpages). It requires to first identify a set of documents (or passages) which may contain answers to the question; then a further verification is performed to determine the answer(s) from them.

In our research, we focus on QA of the second type. We aim to incorporate existing knowledge (knowledge graph) into the QA processes. For example, when trying to identify relevant documents or passages, knowledge could be used to infer if a document/passage is related to the question. In the second verification step, knowledge is also required to make more precise selection of answers.

In terms of approach, we will favor deep learning based approaches, although traditional approaches could also be combined. We will target the following key questions:

- How can a knowledge graph be used to identify relevant documents/passages from a large set of texts?
- How can we infer if an entity in a text is the right answer to a question?

This project has a strong relationship with the task of text reading – to understand a text so as to answer questions around the text. A series of representative work on this topic is performed on

the Stanford Question Answering Dataset (SQuAD), on which several systems have produced higher performance than human beings – see https://rajpurkar.github.io/SQuAD-explorer/. While we will not necessarily do the same experiments, the approaches developed on SQuAD will certainly be useful for the second step (verification) of our QA task.

The project could be on any aspect of the QA task: identification of passages, verification of answers, or integration of knowledge. In this project, the student will first do a literature survey on the task (if not familiar with it). Then a method will be defined through discussions with the supervisor and other students in the group. The student will be asked to implement and test the method.