# Link prediction in court judgments

**Lexum**

Lexum (lexum.com) is a company specialized in legal informatics (legal tech). Lexum is a spin-off from the University of Montreal. The company is an undisputed leader in the development of information retrieval tools for the law - statutes, regulations and decisions of courts and tribunals. Lexum develops CanLII (canlii.org) - Canada's most widely used and trusted search engine for legal information, known as Google for legal information.

Lexum has 35 employees and is located at 4200 Boul. St-Laurent (corner Rachel).

**Project description**

Court decisions represent unstructured textual data from which a certain amount of metadata must be extracted in order to index them and make them useful for research. Also, the quotation relationships between documents - decision that quotes another decision, or law, etc. - is of great value for legal research. However, this citation information is present in the form of free text and must be identified and extracted by heuristic algorithms and more recently with deep learning approaches.

Our project aims to provide, through a deep learning approach which documents should be cited by another document, based on the document text as well as on a graph of the citation network that we have extracted from the judgments.

Ultimately, one can imagine that the developed approaches will allow us to give a legal qualification to any factual description, like this one:

> *"I got a rejection letter today for my visa application (for a 6 week tourist visitor trip) to Canada from India. The reasons given, however, are very vague. None of the checkboxes for documents are ticked but it says under "Other reasons" (ticked):*
>
> *"Given economic conditions, employment prospects, considering your travel history, economic establishment and family ties, I am not satisfied that you would respect the terms of your admission as a temporary resident in Canada".*
>
> *This is quite unclear and surprising to me because I had applied to visit my cousin on holiday. I am still employed in India and submitted my passport (showing my travel history), flight bookings (for this trip), last 6 payslips, 4 months bank statements and tax returns, along with other required documents. Moreover, I had sent visa invitation letter, passport, work permit details of my cousin.*
>
> *Has anyone else received a similar rejection letter? Any suggestions on how to re-apply and address these rejection reasons would be very helpful."*

From this text, and based on knowledge acquired through learning from court decisions, the initial prototype of our solution is able to determine that this is a situation governed by ***section 72 of the Immigration and Refugee Protection Act***.

Essentially, this is a link prediction project adapted to legal information.

**Data**

Our database contains more than 2 million judgments. In these 2 million judgments, we extracted over 14 million citations.

**Tags**

Since it is inconceivable to pass directly the characters forming a judgment to a neural network, we must transform this document into a vector of constant size, an embedding. We are currently using Doc2Vec to generate document embeddings, but we could explore other options (like generating our own embeddings). The encoding of a citation graph is also necessary.

**Calculation Resources**

P2 and P3 AWS instances will be made available to perform the various experiments.

**Main responsibilities of the intern**

The intern will be responsible for some of the experiments with the goal to improve the success rate of our predictions. He/she will also be responsible for some academic research (find scientific articles about our problem and implement them).

**Feasibility of the project**

Based on the time we spend experimenting with various approaches, we are convinced of the feasibility of the goals that the intern will have to meet.

**Skills required**

Ideally Python and PyTorch (but mostly ability to learn it quickly).

**Spoken languages:** French or English

**Contact:** Marc-André Morissette, CTO <morissm@lexum.com>