Artificial Intelligence at Lexum
Our Data
Our Data

CANLII

Provides online access to legal information from the 14 Canadian jurisdictions.

→ Most consulted legal information website in Canada
  – 1M visits, 9M searches, 10M pages / month

→ 2.4M judicial/administrative decisions

→ 40K legislative documents
Our Data

THE CITATION NETWORK

We recognize legal citations in the body of documents to insert hypertext links.

→ 22M links

→ 13M pointing to case law
  – 371K decisions cited at least one time
  – 34.5 citations / decisions on average

→ 8M pointing to legislation
  – 24K legislations cited at least one time
  – 545K sections cited at least one time
  – 24.5 citations / legislation on average

1. Access to Information Act, RSC 1985, c A-1
   Consolidated Statutes of Canada — Canada (Federal)
   cited by 1845 documents

2. Privacy Act, RSC 1985, c P-21
   Consolidated Statutes of Canada — Canada (Federal)
   cited by 1156 documents

   cited by 420 documents

   cited by 134 documents
Our Data

POTENTIAL

→ Google uses hyperlinks to provide semantic hints about the cited page
  
  → Topic
  
  → Authority

→ With enough links, the certainty of these hints becomes statistically stronger

→ Hyperlinks turn database size from a liability into an asset

→ Authoritative documents (i.e. “obvious results”) are often heavily linked to
The text around legal citations provide context / additional information

We have a map of the Canadian law!
Our Setup
Our Setup

**LEXUM LAB**

Internal team of 2.5 software developers dedicated to AI projects

→ Lexum provides the data and workforce
→ Industry partner at IVADO for training and networking with peers
→ Weekly coaching by an AI expert from MILA
→ AI algorithms from MILA and DIRO
Facts2Law
**THE IDEA**

Given a statement of facts, identify the most relevant primary law.

→ Premised on not just matching “words” but on actually understanding the semantic

→ Uses the citation network

→ Made possible by the latest deep learning techniques
THE COMPETITION

→ Casetext CARA A.I.
  - “See what’s missing from your draft or opposing counsel’s motions”

→ Ross EVA
  - “You will never again miss a critical case, spend hours scouring fact scenarios or looking for specific cases”
Facts2Law

OUR APPROACH

→ Group documents into clusters
  - Civil law / common law
  - Jurisdictions
  - Judicial history

→ Tweak learning parameters on a representative document
  - Fit Intercept
  - Elastic Net Parameter
  - Maximum # of iterations
  - Aggregation Depth
  - ...
Facts2Law

OUR APPROACH

→ Apply the resulting learning method on the cluster
  - Get baseline results
→ Run the learning method on a neural network
Learning to Rank
Learning to Rank

THE CHALLENGE

→ Classic search algorithms based on keywords alone are too “noisy” for broad queries
  – Often, the obvious result(s) get lost in the noise
  – The problem grows with the size of the database

→ Creating and tuning ranking algorithms is hard
  – Highly dependant on the domain (reuse is not always possible)
    • Legal research, retail, jobs, intranets, web search, etc. all require different algorithms
  – Requires a lot of effort
    • For manually tuning
    • For evaluating
Learning to Rank

THE IDEA

→ Instead of writing our own search algorithm, can we let the machine figure it out?

→ ML can take on tons of new signals
  – “Bag of words” Okapi BM25 scores not just on text body, but on titles, references, etc.
  – The citation network to provide results that are semantically meaningful
  – User history to identify domains of interest
  – User location to identify relevant jurisdiction

→ Can we get users to evaluate the ranking?
  – They won’t bother to rate individual results, but…
  – What if results “engagement” was a good enough metric?
Learning to Rank

Query context
(keywords, user history, user location, etc.)

Document context
(text body, content from citation network, citation count, court, jurisdiction, age, etc.)

Machine Learning Algorithm

Engagement Metrics
(clicks, reading time, printed, etc.)
Learning to Rank

Query context
(keywords, user history, user location, etc.)

Document context
(text body, content from citation network, citation count, court, jurisdiction, age, etc.)

Machine Learning Algorithm

Engagement Metrics Score
(clicks, reading time, printed, etc.)
Learning to Rank

Query context
(keywords, user history, user location, etc.)

Document context
(text body, content from citation network, citation count, court, jurisdiction, age, etc.)

Machine Learning Algorithm

Score

Used by Bloomberg, Lexis-Nexis, WalMart, etc.
Learning to Rank

**OPPORTUNITIES**

- Handles non-linearities
- Parameter tuning via clickstream
- Can handle almost infinite complexity gracefully

**CHALLENGES**

- Engagement is sensitive to all sorts of user biases
  - First results benefit from very large biases
- LTR only learns to tune the signals
  - It does NOT have an understanding of the query or documents
  - Wholly dependent on the quality of the incoming signals
- Requires lots of IT resources
Learning to Rank

Apache Solr + Lexum
5 Current Pilot Projects
Current Pilot Projects

COMMITMENTS BY LEXUM

→ Processing the client data with our AI-powered engines

→ Providing an online visualization and validation interface for results

→ Providing all validated results in a structured format.

→ Processing data in a secure environment and signing an NDA if necessary

Lexum’s Facts2Law Engine Pilot Project

Lexum is currently looking for partner organizations interested in testing its Facts2Law engine. Lexum’s Facts2Law engine is a product of Lexum’s Lab Artificial Intelligence (AI) initiative that uses deep learning to predict the most relevant Canadian primary legal information when provided with a statement of facts. It has been trained on the millions of documents available on the Canadian Legal Information Institute (CanLII) website. It is a versatile solution that can be used to obtain various outputs:

- Identify relevant decisions and applicable legislative provisions for any statement of facts related to a legal issue.
- Identify cases and legislative provisions potentially missing from briefs or opinions.
- Guide self-represented litigants near the initial stages of legal research.
- Cluster legal documents based on common characteristics.

If your organization deals with substantial volume of legal data and is interested in testing the potential of AI to enhance its use, we would like to hear from you.

Lexum Commits to:

- Processing your data with its AI-powered Facts2Law engine.
- Providing your organization with an online visualization and validation interface for results.
- Providing all validated results to your organization in a structured format.
- Processing your data in a secure environment and entering into a Non-Disclosure Agreement (NDA) if necessary.

Requirements:

- Your organization provides a dataset composed of hundreds to thousands of legal documents in text format, MS-Word, WordPerfect, or generated PDF (not scanned).
- The processing of these documents by Lexum can be completed off-site (in the cloud).
- Your organization shall help us validate the results to assess the success of the pilot.

Contact Us to Get Started with your Pilot:

Pierre-Paul Lemay
Director of Business Development
514.318.2098
lemayp@lexum.com
Current Pilot Projects

REQUIREMENTS FROM CLIENTS

→ Provide a dataset made of of 100’s to 1,000’s of legal documents

→ Processing can be completed off-site (in the cloud).

→ Help us validate the results to assess the success of the pilot.
AI @ Lexum?

Leveraging our map of the law and how users navigate it!