

The Data Science of Law

A Rules-based Approach to Measuring Prescriptivity in Canadian Regulations

Wolfgang Alschner, 11 March 2019

Faculté de droit | Faculty of Law

uOttawa.ca



uOttawa

Prelude: Of words and numbers

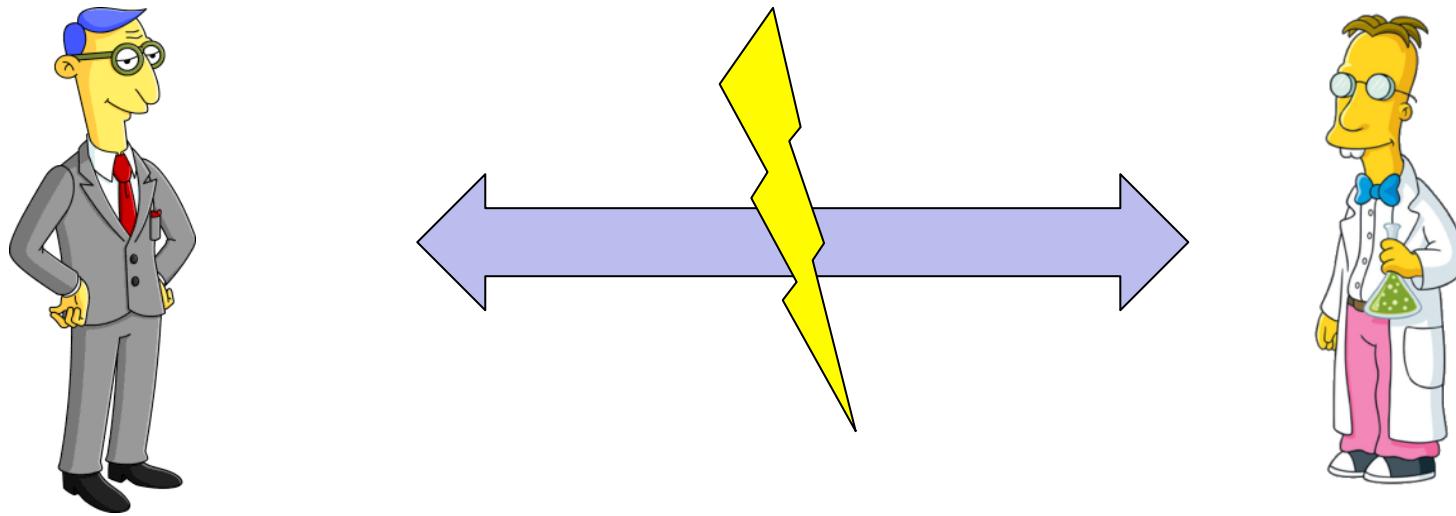


"To succeed in the profession of the law, you must seek to cultivate command of language. **Words are the lawyer's tools of trade.**

... When I was young, I did not think much in words. At Oxford, I studied **Mathematics**. No need for words there. The tools I used then were **numbers, letters and symbols**. Lifeless things without meaning or sound – the **necessary tools of the scientist but not of the lawyer.**"

Lord Denning, "The Discipline of Law", 1979

Law vs Science



Law as Data

Traditional: Structured Data

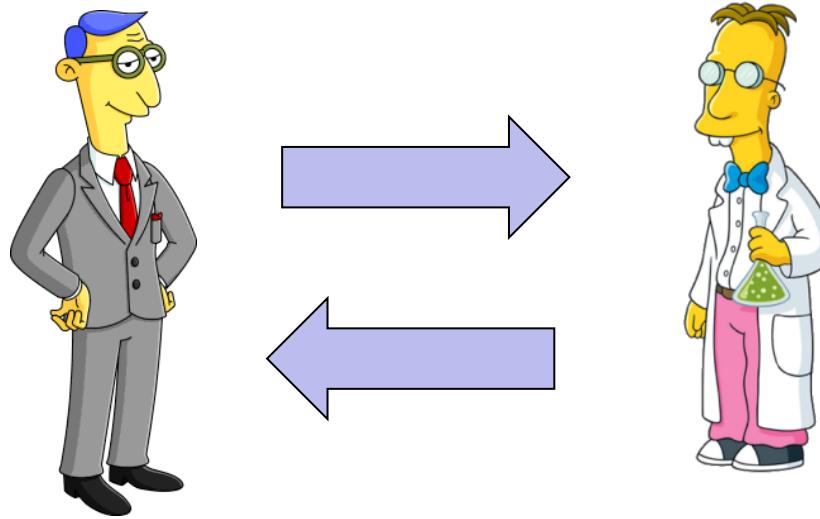
	A	B	C	D	E	F	G	H	I
1	Item #	Description	Vendor	Category	Size	Unit	Starting Qty	Starting Value	Wk 1 Qty
2	492229	TURKEY SLICED .5 OZ	Ben E Keith	2 - FROZEN FOOD	0	0	0.00	\$ -	0.00
3	662371	DRESSING CAESAR CREAMY	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
4	779243	MARGARINE LIQUID OLEO	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
5	815306	LID PLAS SOUFFLE CLEAR	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
6	860055	LID PLAS 16SL SLOTTED	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
7	860060	CUP FOAM 16OZ 16J16	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
8	774704	PAPRIKA	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
9	664005	Mustard Prepared	Ben E Keith	4 - GROCERY	512	fl oz	0.00	\$ -	1.00
10	750100	CHEESE PARMESAN SHRED	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
11	250025	EGG FRESH SHELL MED USDA AA	Ben E Keith	1 - PRODUCE	0	0	0.00	\$ -	1.00
12	686034	VINEGAR APPL CIDER 40GRAIN	Ben E Keith	4 - GROCERY	0	0	0.00	\$ -	0.00
13	29078	LIME 12 CT	Ben E Keith	1 - PRODUCE	12	ct	0.00	\$ -	2.00

Law as Data

New: Unstructured Data



A Data Science Approach to Law



Agenda

1. A new toolkit for lawyers
2. Enabling technologies
3. Legal data analysis
4. Case study: Semantic Analysis of Regulations
 - Dataset
 - Method
 - Results
 - Limitations

Rethinking Legal Analysis



<complementary>



Traditional Legal Techniques

Detailed Reading

Interpretation

Small Scale

Legal Data Science

Distant Reading

Computation

Large Scale

Legal Data Science -> scalable legal analysis

Agenda

1. A new toolkit for lawyers
2. Enabling technologies

Technological Drivers

- 1) Faster:** Increase of computational power
- 2) More:** Big data
- 3) Better:** Artificial Intelligence

1. Faster: Moore's Law

1 The accelerating pace of change ...



2 ... and exponential growth in computing power ...

Computer technology, shown here climbing dramatically by powers of 10, is now progressing more each hour than it did in its entire first 90 years

COMPUTER RANKINGS

By calculations per second per \$1,000



Analytical engine
Never fully built, Charles Babbage's invention was designed to solve computational and logical problems



Colossus

The electronic computer, with 1,500 vacuum tubes, helped the British crack German codes during WW II



UNIVAC I

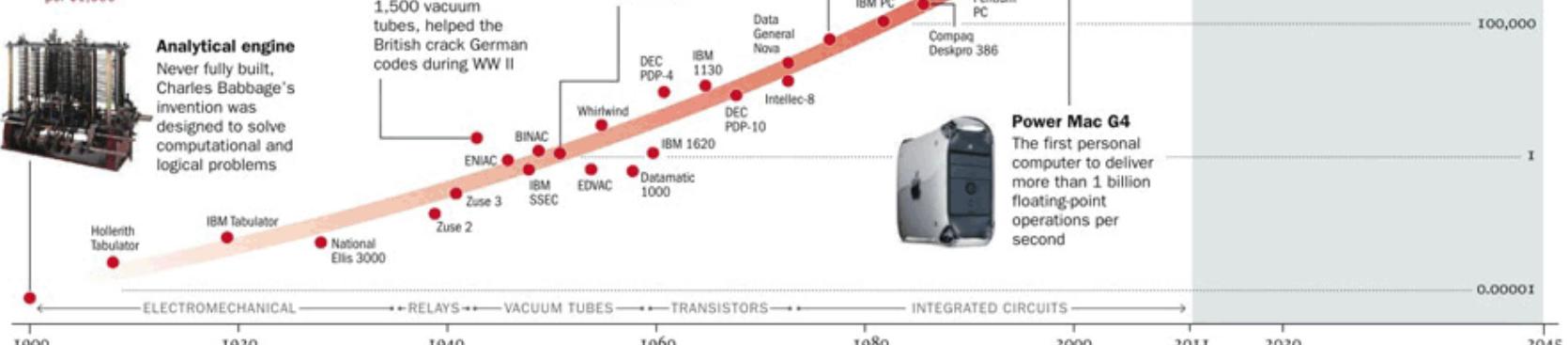
The first commercially marketed computer, used to tabulate the U.S. Census, occupied 943 cu. ft.

Whirlwind
BINAC
ENIAC
Zuse 3
IBM SSEC
EDVAC
Datamatic 1000
IBM 1620
DEC PDP-4
IBM 1130
DEC PDP-10
Intellic-8

IBM PC
Data General Nova
Compaq Deskpro 386
Pentium II PC
Pentium PC
Dell Dimension 8400
Mac Pro
Nvidia Tesla GPU & PC

Power Mac G4

The first personal computer to deliver more than 1 billion floating-point operations per second

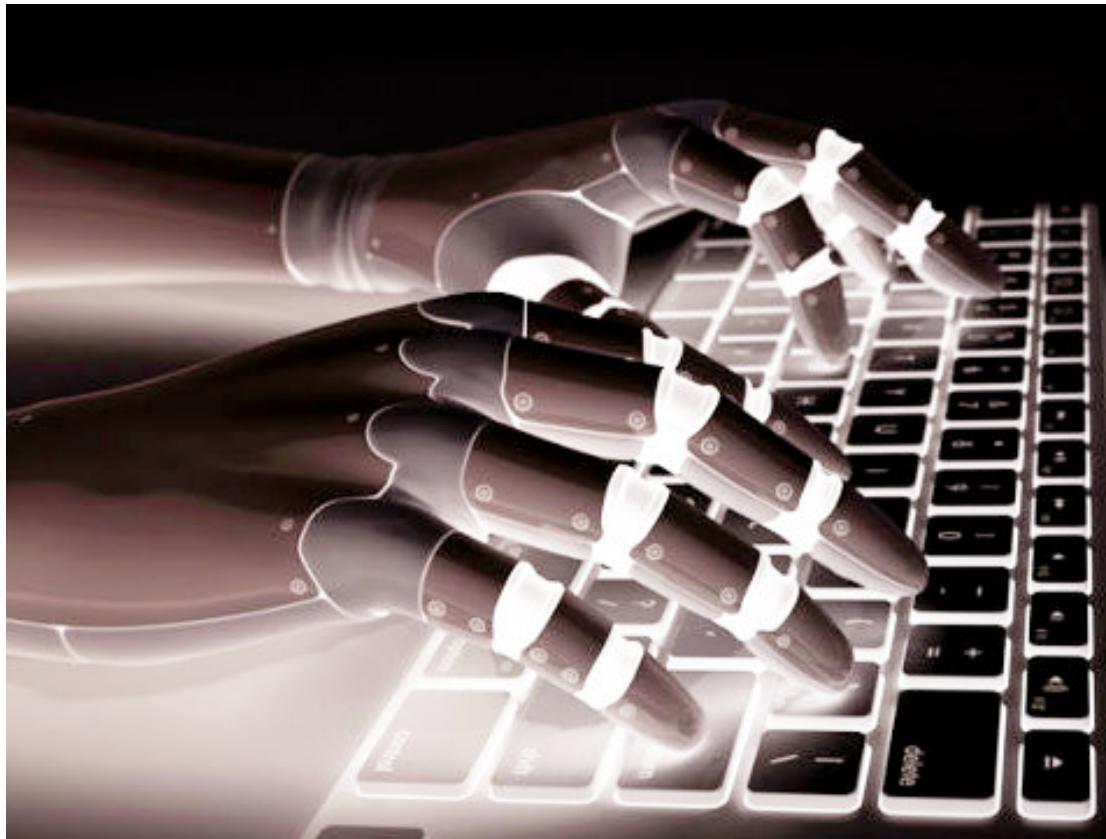


3 ... will lead to the Singularity

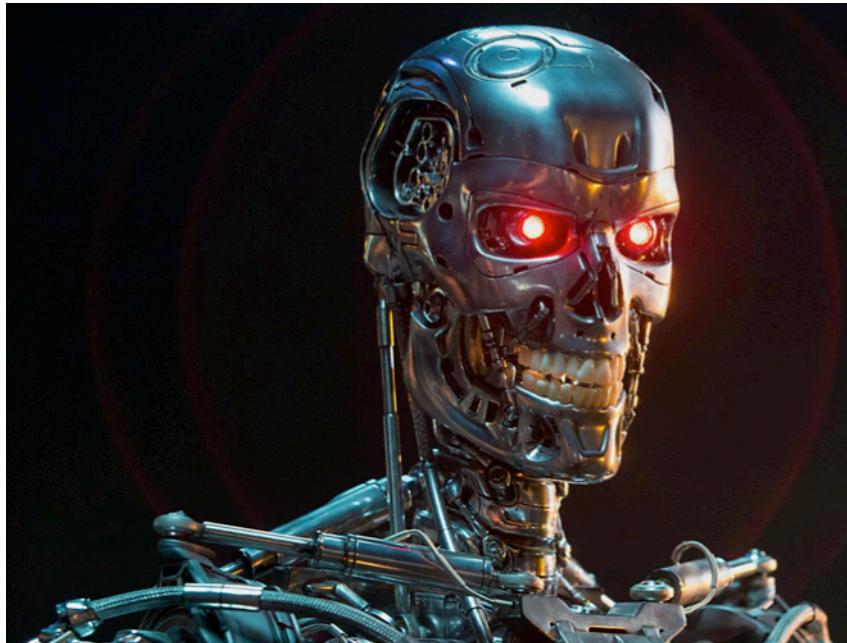
2. More: Big Data



3. Better: Artificial Intelligence



Artificial Intelligence



No Terminators yet...

Today's AI

- is *specific* rather than *general*
- mimics human *output*, not human *reasoning*

Artificial Intelligence

**Intelligence
derived from
*rules***



**Intelligence
derived from
*data***

NETFLIX

Agenda

1. A new toolkit for lawyers
2. Enabling technologies
3. Legal data analysis

Data Analysis of Legal Texts



Data Analysis of Legal Texts

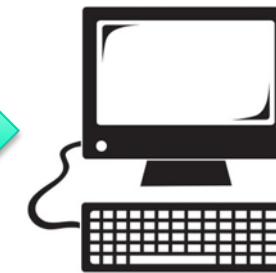
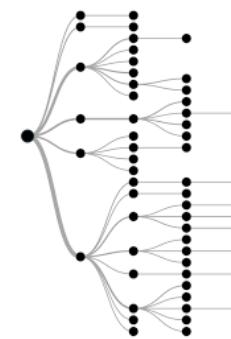


Data Analysis of Legal Texts



Canada's Parliament

printed with permission granted by the office of the Prime Minister <http://www.pmc.gc.ca>

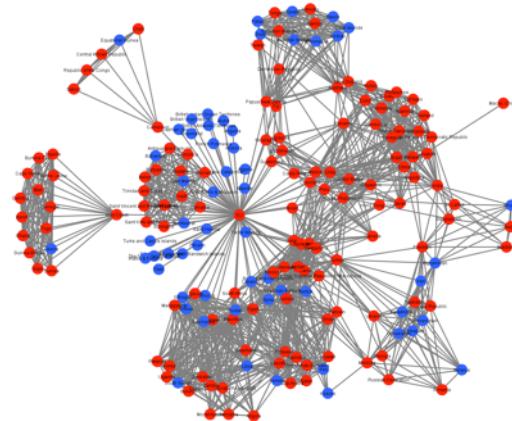


Data Analysis of Legal Texts

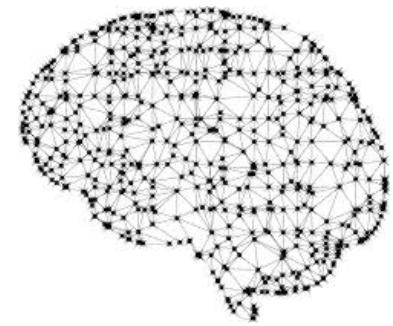
Natural Language Processing

may
referred
canada
person
year appeal
subsectionaircraft information
including **repealed** operator
certificate
made unless party time described
one area used applicant surface part
product land apply **regulations** part
amount goods issued service board air
material act date within operation company producer
case required day following good
period permit flight per notice use order
document purposes application plan
control provided equipment
pursuant accordance designated number
officer section set system
minister production court **subsection**
shall
paragraph

Network Analysis



Machine Learning



Data Analysis of Legal Texts

Natural Language Processing

A word cloud visualization of legal terms. The words are colored by category: pink for 'may', orange for 'canada', green for 'person', blue for 'subsection', red for 'repealed', purple for 'operator', grey for 'shall', and black for other terms like 'aircraft', 'information', 'appeal', 'year', 'section', 'subsection', 'aeronautics', 'regulations', 'service', 'board', 'air', 'company', 'producer', 'following', 'good', 'operation', 'day', 'within', 'flight', 'per', 'notice', 'use', 'order', 'application', 'value', 'purposes', 'plan', 'means', 'document', 'control', 'provided', 'number', 'respect', 'means', 'equipment', 'canadian', 'control', 'officer', 'accordance', 'section', 'minister', 'court', 'production', 'paragraph', 'subsubsection', 'designated', 'set', 'schedule', 'system'.

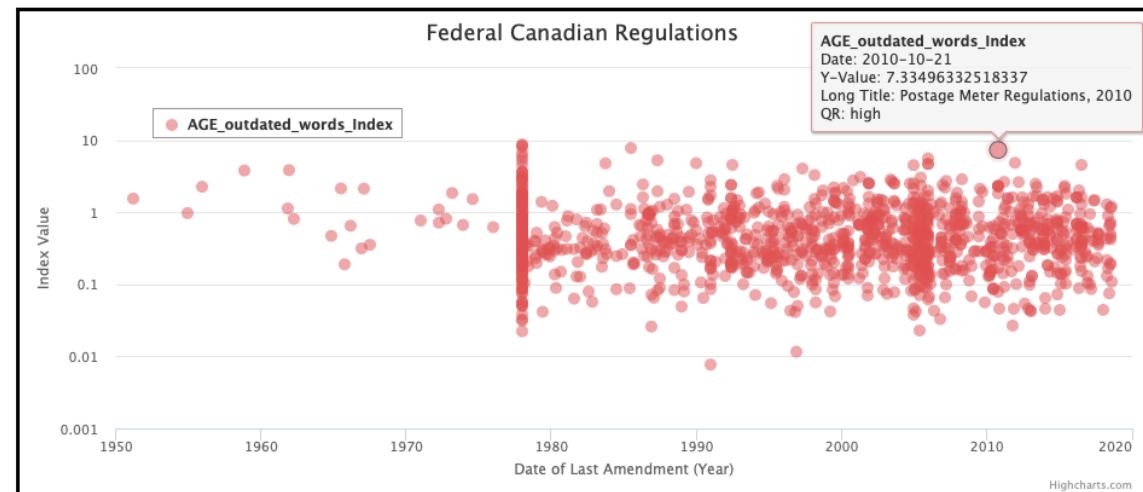
- Trace words and terms
- Measure text characteristics
- Identify content of corpus
- Cluster texts
- Detect similarity patterns

Data Analysis of Legal Texts

Natural Language Processing

may
referred
canada
person
year appeal
subsection aircraft information
including repealed operator
made unless party time described
one area used applicant surface
land apply regulations part
product one issued service board air
amount goods required day within
material act operation company producer
date following good
case flight per notice use order
period application plan
document purposes means
canadian control provided equipment
pursuant accordance number schedule
officer section designated system
minister production court set
shall subsection paragraph

- Trace words and terms
- Measure text characteristics
→ e.g. tracking words related to "outdated" technology in regulations

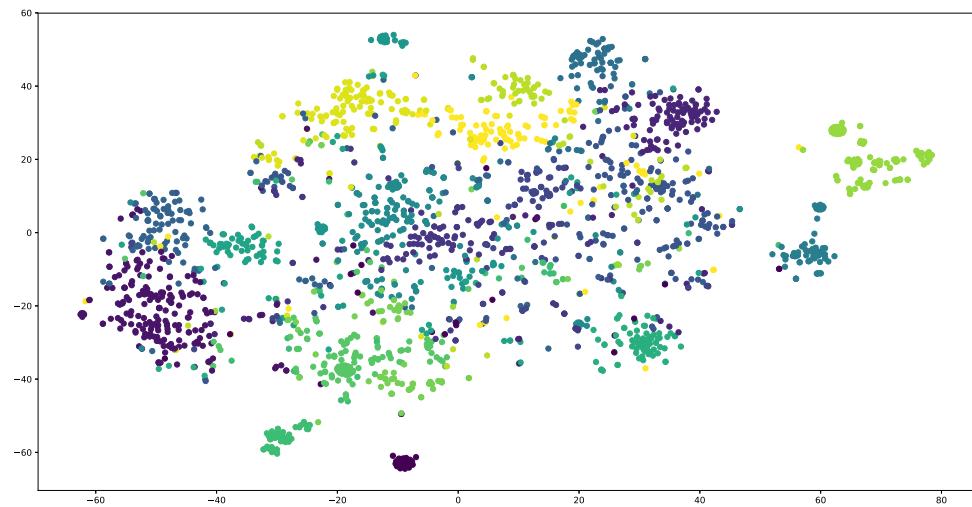


Data Analysis of Legal Texts

Natural Language Processing



- Cluster texts
- Detect similarity patterns
→ e.g. group similar regulations together based on their shared words



Data Analysis of Legal Texts

Natural Language Processing



A large stack of legal documents represented by horizontal lines, decreasing in height from left to right.

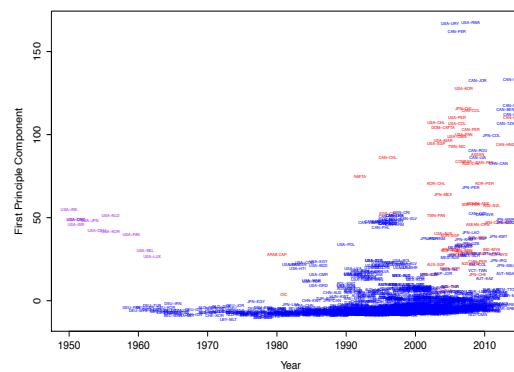
may
referred
canada
person
year appeal
subsectionaircraft information
including repealed operator
certificate made unless party time described
product one area used applicant surface part
amount goods issued service board air
material act date required day within company producer
case permit operation following good
period flight per notice use order
document purposes application plan
canadian control provided equipment
pursuant accordance designated number
officer section set system
minister production court subsection
shall paragraph

① Text to data



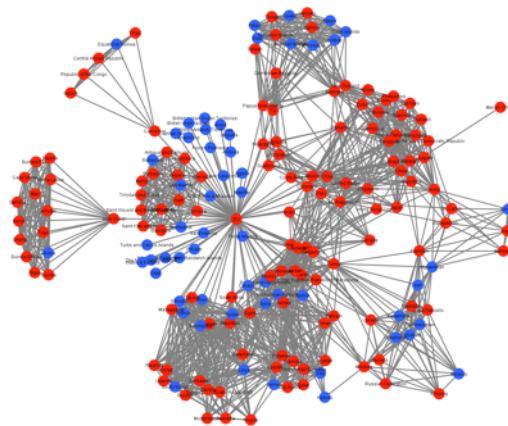
0	1	1	1	0	1
1	0	1	1	1	0
0	0	0	1	1	1
1	0	1	1	1	0
0	0	1	0	1	0
1	0	1	1	1	0
0	0	0	1	1	1
1	1	1	1	1	0

② Statistical Data Analysis



Data Analysis of Legal Texts

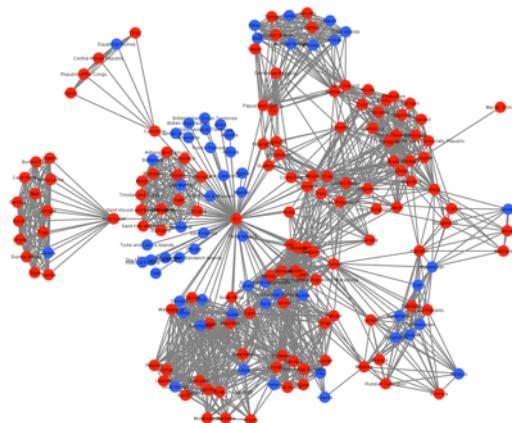
Network Analysis



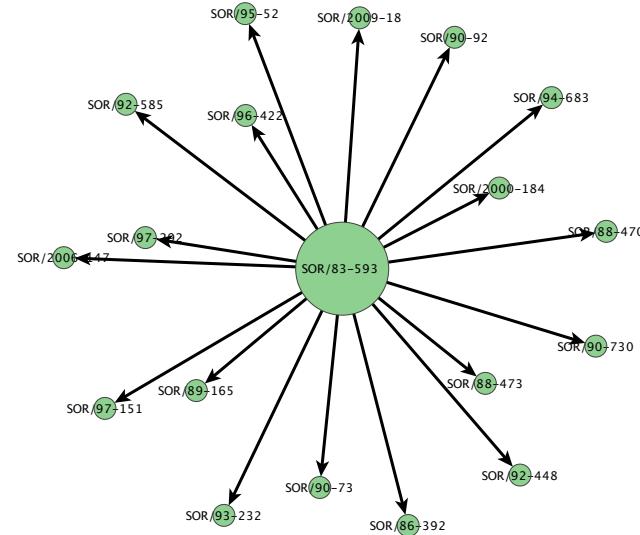
- Understand links between texts
- Reveal network structures
- Measure importance of nodes
- Track diffusion patterns

Data Analysis of Legal Texts

Network Analysis



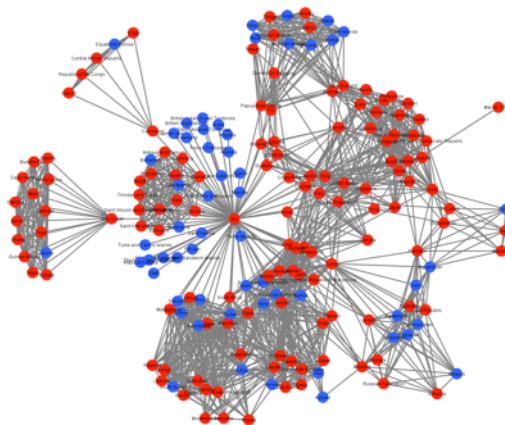
- Understand links between texts
→ e.g. *cross-references between regulations*



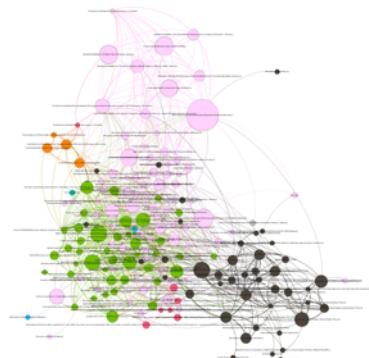
Data Analysis of Legal Texts

① Link Extraction

Network Analysis



② Network Creation



③ Network Measures

- Centrality
- Distribution
- ...

Data Analysis of Legal Texts

Machine Learning

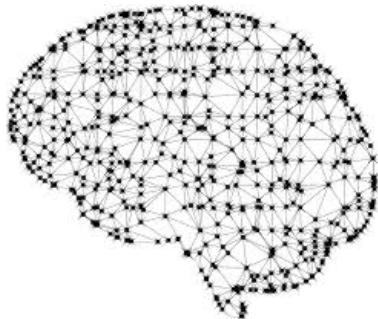


- Classification of text
- Revealing of latent traits

Data Analysis of Legal Texts

- Classification of text
→ e.g. grouping norms into categories

Machine Learning



Carriages and Strollers Regulations

(6) *Child restraint system — stroller*

Every seating unit of a stroller must be equipped with a child restraint system that meets the following requirements:

- (a) it must be permanently attached to the frame or upholstery of the stroller [rule-based regulation];
- (b) its anchorages must not break or become detached from their attachment point when tested in accordance with Schedule 4 [outcome-based regulation];

Agenda

1. A new toolkit for lawyers
2. Enabling technologies
3. Legal data analysis
4. Case study: Semantic Analysis of Regulations
 - Dataset
 - Method
 - Results
 - Limitations

Project Overview

Empirically investigating four aspects of Federal Canadian regulations by leveraging legal data science:

Prescriptivity: How binding are regulations?

Flexibility: How responsive are regulations to change?

Complexity: How easily understandable are regulations?

Age: What is the average age of regulations?

Regulations Dataset

Regulation count: 3196

- CRCs: 571
- SORs: 2625
- Repealed: 427

Corpus:

- Words: 7'049'156
- Average length: 2206 words



Source:

ftp://205.193.86.89/Consolidation_Regs_1.2.0/EN

Data Example in XML Format

```
-<Regulation regulation-type="SOR" xml:lang="en">
- <Identification>
  <InstrumentNumber>SOR/2002-227</InstrumentNumber>
+ <RegistrationDate></RegistrationDate>
+ <LastAmendedDate></LastAmendedDate>
+ <ConsolidationDate></ConsolidationDate>
+ <LastModifiedDate></LastModifiedDate>
- <EnablingAuthority>
  <XRefExternal reference-type="act">IMMIGRATION AND REFUGEE PROTECTION ACT</XRefExternal>
  <XRefExternal reference-type="act">FINANCIAL ADMINISTRATION ACT</XRefExternal>
</EnablingAuthority>
<LongTitle>Immigration and Refugee Protection Regulations</LongTitle>
+ <RegulationMakerOrder></RegulationMakerOrder>
</Identification>
+ <Order></Order>
-<Body>
+ <Heading level="1"></Heading>
+ <Heading level="2"></Heading>
+ <Section></Section>
-<Section>
- <MarginalNote>
  Definition of
  <DefinedTermEn>agent</DefinedTermEn>
  — section 148 of the Act
</MarginalNote>
<Label>1.1</Label>
-<Subsection>
  <Label>(1)</Label>
  - <Text>
    For the purposes of section 148 of the Act, agent includes any person — whether or not an independent contractor
    — who provides services as a representative of a vehicle owner, operator or charterer.
  </Text>
-<Subsection>
-<Section>
- <MarginalNote>
  Definition of
  <DefinedTermEn>agent</DefinedTermEn>
  — paragraph 148(1)(d) of the Act
```



Meta Data
describes text



Text Data
structured full text

Methodology

Recap: 2 ways to make algorithms “intelligent”

1. Intelligence derived from *rules*
2. Intelligence derived from *data*

How to measure prescriptivity in regulations?

Rules-based intelligence:

- Reverse engineering legal drafting process
- Legal drafting is partially rules-based
 - These rules leave traces in legal documents

Reverse engineering legal drafting process

Interpretation Act (R.S.C., 1985, c. I-21)

Section 11

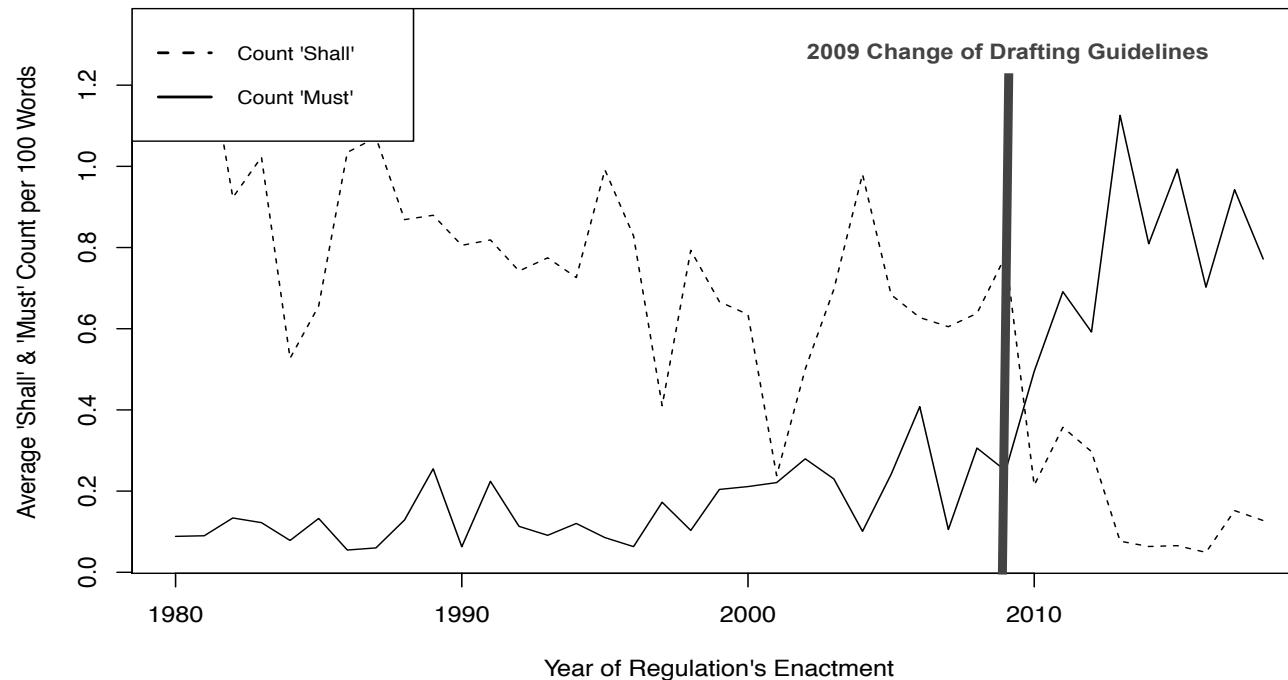
Imperative and Permissive Construction

The expression “shall” is to be construed as imperative and the expression “may” as permissive.

Reverse engineering legal drafting process

2009 - Change of drafting guidelines

“shall” is replaced by “must”



Reverse engineering legal drafting process

Conceptualization: obligations
permissions

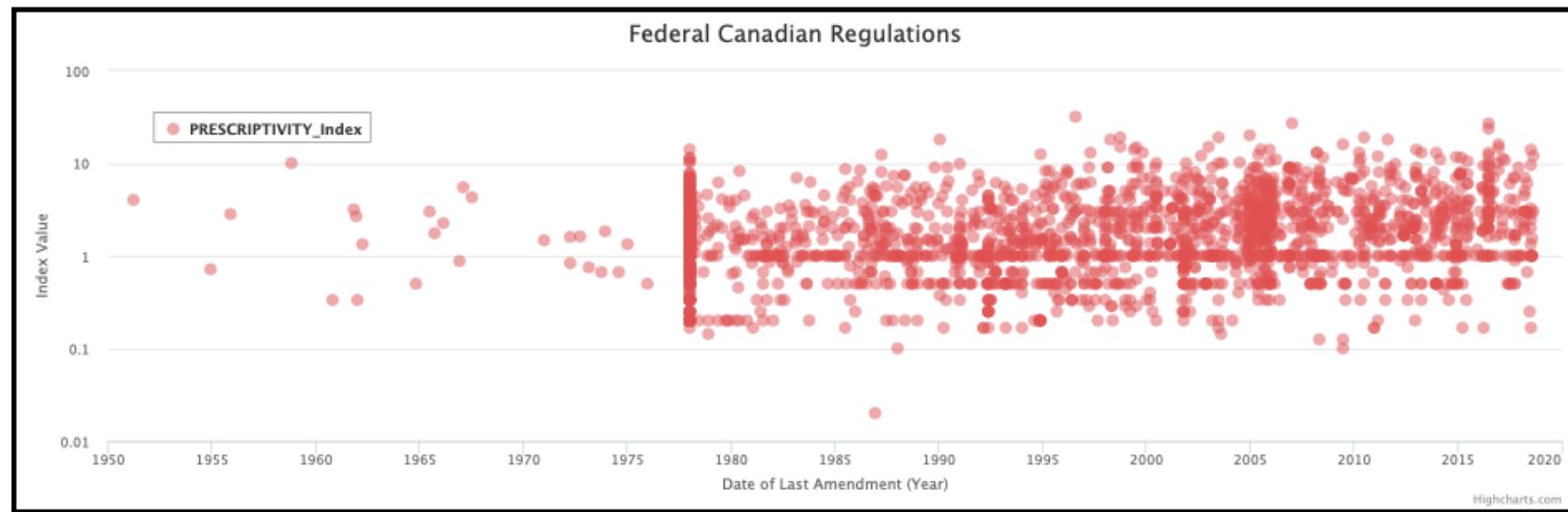
must
shall
required
prohibited
is to be
cannot be
may not
may only

(may)(?! not| only)
not required
not be required
not prohibited
entitled to
has the right to

Website

- Data, code and results are freely available at:
<https://github.com/WAlschner/Semantic-Analysis-Regulations>
- Website to explore results interactively:
<https://regulations.herokuapp.com/>

Federal Canadian Regulations



Reg ID	Regulation Name	Enabling Authority	Quantile Rank	PRESCRIPTIVITY_Index
SOR/96-334	Transportation Information Regulations	CANADA TRANSPORTATION ACT	high	31.75
SOR/2006-347	2-Butoxyethanol Regulations	CANADIAN ENVIRONMENTAL PROTECTION ACT	high	27
SOR/2016-152	Cribs, Cradles and Bassinets Regulations	CANADA CONSUMER PRODUCTS SAFETY ACT	high	26.833333333333
SOR/2016-167	Carriages and Strollers Regulations	CANADA CONSUMER PRODUCTS SAFETY ACT	high	23.5
SOR/98-202	Conditions of Transferring Firearms and Other Weapons Regulations	FIREARMS ACT	high	20
SOR/98-467	Order Declaring an Amnesty Period	CRIMINAL CODE	high	19
SOR/2010-127	Northern Canada Vessel Traffic Services Zone Regulations	CANADA SHIPPING ACT, 1985	high	19
SOR/2003-212	Canadian Nuclear Safety Commission Cost Recovery Fees Regulations	NUCLEAR SAFETY AND CONTROL BOARD REGULATIONS	high	19
SOR/90-97	Coal Mining Occupational Health and Safety Regulations	CANADA LABOUR CODE	high	18.0416666666667
SOR/98-213	Firearms Records Regulations	FIREARMS ACT	high	18

Prescriptivity Findings

- High prescriptivity regulations:
 - high-risk products or processes
 - vulnerable consumer groups (e.g. children)
 - transportation
- Low prescriptivity regulations:
 - casual goods
 - administrative processes

Limitations

Metric does not account for:

- Different addressees of obligations and permissions (regulator vs regulated)
- Different types of prescriptions (conduct vs outcome)
- Difference between prescriptive drafting and actual administrative burden imposed

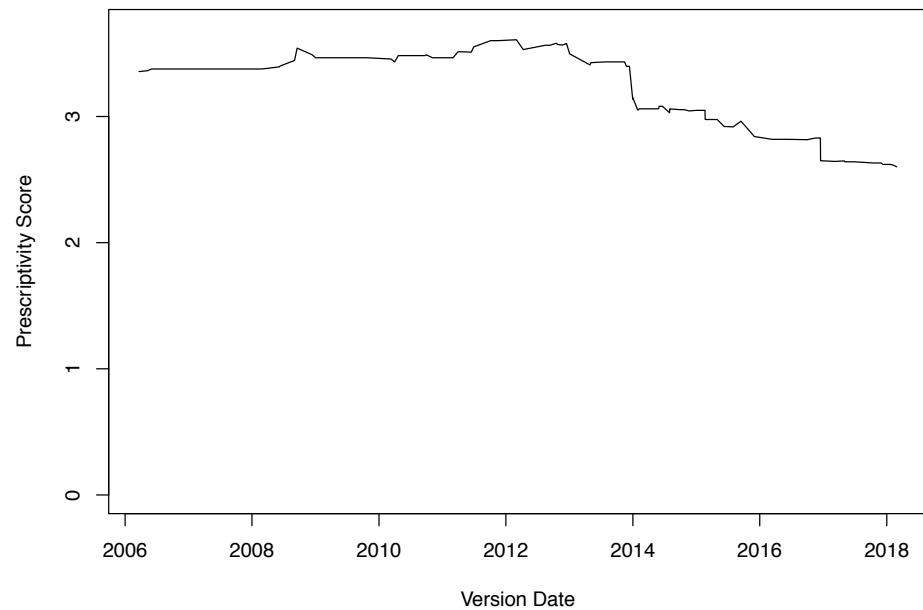
Not generalizable:

- Drafting conventions are jurisdiction-specific
- Different approach in French drafting

Future Work

- Possible Applications:
 - Track regulatory trends
 - Identifying outlier regulations in comparable groups that may indicate overly high or low prescriptivity
 - Compare prescriptivity scores of same regulation across different times
 - Interactive drafting aid

Prescriptivity in Immigration and Refugee Protection Regulations



Merci!!

Get in touch: wolfgang.alschner@uottawa.ca