

Toronto Resources on Supervised Consumption Services Offer the Most Comprehensive Information for the Public*

Analysis of Multi-Level Government Content using Topic Modelling

Finn Korol

20 April 2023

This paper evaluates content on government websites regarding supervised consumption sites using topic modelling. Website content was used to train three models optimized using quantitative metrics. Topics were analyzed based on the number of topics discussed and their contents. In this work, a general lack of information was identified to be problematic, however at the City of Toronto level, the most comprehensive resources were available. I recommend that government websites, particularly at the Ontario level, provide more comprehensive information to promote informed decision-making and reduce drug-related harms.

Table of contents

1	Introduction	2
2	Data	3
2.1	Data Origins and Methodology	3
2.2	Data visualisation	3
3	Model	6
4	Results	8
5	Discussion	9
5.1	Findings	9
5.2	Weaknesses	9

*Code and data are available at: https://github.com/korolodf/drug_gov_sent.

5.3 Next steps	10
6 Appendix	11
References	13

1 Introduction

There has been a growing awareness of the public health crisis associated with the opioid epidemic in Canada. One approach adopted by Canada has been to mitigate harm caused by controlled substance abuse. Through harm-reduction, Canada acknowledges that drug-abuse happens, but by making it safer through the development of supports, there are many practical benefits (Canada 2023). One of the most focused upon branches of harm reduction is the establishment of supervised consumption sites (SCS), which provide a medically supervised environment for individuals to use drugs through inhalation, injection or oral administration. These sites have been shown to reduce the risk of overdose, decrease the transmission of infectious diseases, connect individuals with healthcare and social services and lessen the damage caused by drug-related organized crime. Moreover, SCS make communities more safe by providing a secure place for the disposal of drug paraphernalia, such as needles. Despite their proven effectiveness, SCS remain a controversial and politicized topic in many jurisdictions (Ziegler, Wray, and Luginaah 2019).

Government websites are a primary source of information for citizens seeking information about SCS, as they are government regulated. However, it is unclear whether these websites provide unbiased information, and the depth of the information provided across levels. This paper aims to evaluate the information offered on these government websites regarding SCS using natural language processing techniques, primarily through topic modelling.

In Canada, it is the federal government which funds the provinces to provide healthcare services. Applications to open a SCS are made by individuals who feel that the community is in need, as well as by municipalities like the City of Toronto. Due to the Canadian laws against the distribution of drug paraphernalia, it is required that these sites receive an exemption to all applicable laws, which is granted on an application-basis (Canada 2021; Toronto 2022). From there, the provinces are responsible for allocating federal funding to these sites. Due to the varying influence of each level of government on the creation and facilitation of safe injection sites, content from governments at the municipal, provincial and federal levels will be studied. These will be the Governments of Canada and Ontario, and the City of Toronto.

The use of natural language processing in analyzing text data has become increasingly popular in recent years due to its ability to process text data, and has been expanding to the realm of social sciences as a method to rapidly analyze and label written material (Storopoli 2019). Topic modelling, and sentiment analysis, in particular are widely used methods for identifying the underlying topics and emotional tone of text. Data was collected from easily accessible

websites published by governments and grouped by level. To begin this analysis, the contents and associated sentiments of the data at each level were analyzed. This was followed by topic modelling to identify the number of topics identified per collected level, concluding this analysis with an evaluation of these topics. Next, this paper moves on to discuss findings, limitations and next steps. This paper aims to shed light on the potential deficits that may exist in the information presented to the public related to SCS.

This data investigation was done using (R Core Team 2020). Packages used were here (Müller 2020), tidytext (Silge and Robinson 2016), wordcloud (Fellows 2018), syuzhet (Jockers 2015) as part of sentimentr (Rinker 2021), knitr (Xie 2023), topicmodels (Grün and Hornik 2011), tm (Feinerer, Hornik, and Meyer 2008), ldatuning (Nikita 2020), rvest (Wickham 2022), xml2 (Wickham, Hester, and Ooms 2021), pdftools (Ooms 2023) and tidyverse (Wickham et al. 2019).

2 Data

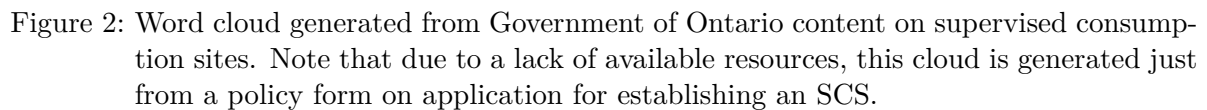
2.1 Data Origins and Methodology

An emphasis was put on availability in the collection of this data, as this paper aims to investigate the varying resources by government level of information made accessible to the public. Searches for websites were made on Google, as well as on each level of government’s own websites using searches “safe injection (site)” and “supervised consumption (site)”. From the City of Toronto, these sites included sites titled (*Toronto Public Health submits application to continue providing supervised injection and treatment services*) (Toronto 2018), (*Supervised Consumption Services*) (Toronto 2022), and (*Consumption & Treatment Services (CTS) Inspection Program*) (Toronto 2023). From the Ontario Government, this included the sole document titled (*Consumption and Treatment Services: Application Guide*) (Ontario 2018). From the Government of Canada, this included a website called (*Supervised consumption explained: types of sites and services*) (Canada 2023) and a legal website called (*Subsection 56(1) class exemption in relation to urgent public health need sites in the provinces and territories*) (Canada 2023).

These sources were web scraped of all data, which was then cleaned of symbols and punctuation using natural expressions. Data collected from the same level of government was grouped together to prepare for natural language processing.

2.2 Data visualisation

Without removing stop words, three word clouds were made — one for content from each level of government. City of Toronto data is visualized in Figure 1, Government of Ontario’s is in Figure 2 and Government of Canada’s is in Figure 3.



To determine the optimal number of topics for each document, the Cao et al. (2009) and Deveaud, SanJuan, and Bellot (2014) metrics were used in tandem on models ranging from 2-20 topics each to evaluate performance (view Appendix). A metric-based model selection approach was useful here due to the limited data so as not to lose training data to a test/train split. Topic number was chosen based on which models were identified to have the strongest scores of both density and coherence, respectively. This metric-based method revealed that 6 topics were ideal for the Toronto model and 2 topics were ideal for both the Ontario and Canada models. Gibbs sampling was selected for these models. The resulting topics were extracted using terms and were used to understand the most relevant themes in each document.

Models are described by top words across topics in Table 2 (Toronto model), Table 3 (Ontario model) and Table 4 (Canada model).

Table 2: Top 10 words identified across the Toronto model

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
overdose	pmtuesday	services	supervised	last	call
drug	supervision	health	community	note	drugs
toronto	city	consumption	shelter	queen	risk
street	closemonday	injection	pmthursday	staff	use
service	pmfriday	overdoses	iphare	prevention	hours
streetphone	pmwednesday	edundas	note	monday	action
people	hygienic	pmfriday	park	harm	benefits
public	these	preobtained	pmparkdale	locations	data
will	trained	shows	related	opioid	eastsherbourne
centre	dundas	stqueen	sunday	street	program

Table 3: Top 10 words identified across the Ontario model

Topic 1	Topic 2
consumption	services
community	cts
clients	will
local	the
care	applicants
requirements	health
service	must
application	treatment
include	ministry
may	provided

Table 4: Top 10 words identified across the Canada model

Topic 1	Topic 2
health	exemption
services	illegal
minister	substances
substance	sites
drug	uphns
provincial	territorial
consumption	the
holder	checking
use	overdose
uphns	public

4 Results

From the insights provided by the models, topics were analyzed. In the Toronto model, topics were highly varied seeming to focus on supervised consumption services and related topics, such as supervision, street phones, and Toronto’s drug center. Other topics related to health services and referrals had an emphasis on community support and outreach. Moreover, topics discussed social services and reducing drug use, with a focus on outreach to specific populations, such as those living on the street. There were also key words in topics centered on injection practices, overdoses, and statistics related to these issues. There is a topic related to reducing harm and overdose prevention, with a focus on individuals’ lives and benefits. Finally, topics seemed to discuss the geographic location of safe injection sites and related services.

At the provincial level, Topic 1 appears to be related to healthcare services and treatment options, including local requirements for clients, whereas Topic 2 seems to be related to community services and support programs for clients, which may include funding and security measures.

At the federal level, Topic 1 seemingly discusses information related to territorial and provincial exemptions for illegal substances, drug checking sites, and staff within public health services. Topic 2 seems to include discussions related to public health services, substance use, and drug holders, and at the federal level includes.

5 Discussion

5.1 Findings

At the Toronto level, the focus appears to be on providing a range of services to individuals using drugs, including supervised consumption sites, health services and referrals, social services and outreach, harm reduction, and overdose prevention. The emphasis is on addressing the needs of specific populations and reducing drug use through outreach and support programs. In contrast, at the provincial level, the focus appears to be on healthcare services and treatment options, as well as funding and security measures. At the federal level, there is more discussion of public health services, substance use, and exemptions for illegal substances, as well as drug checking sites and staff within public health services.

These differences in focus and emphasis have significant implications for individuals in need of information related to supervised consumption sites and related services. Depending on their location, individuals may have access to different types of services and support, and may face different requirements for accessing these services. It is important for individuals and advocates to be aware of these differences in order to better navigate the system and advocate for their needs. However, these differences also highlight the need for a more coordinated and comprehensive approach to addressing drug use and related issues at all levels of government.

5.2 Weaknesses

A significant weakness in this analysis, both in terms of the breadth of information to be modeled and the strengths of the models was the lack of data available. Where the Ontario and Canada models both only contained 2 topics (the minimum number allowed using the metric-based approach to selection), it is likely that there was not varied enough information to form comprehensive topic models in the data used. This weakness also points to the fact that there is not enough government information readily available for the public on SCS services. Considering the fact that SCS locations are subject to losing funding on short notice (Ziegler, Wray, and Luginaah 2019), it is of concern that there is not more information to be accessed on the rapidly evolving landscape of harm reduction. It was shocking, particularly at the Ontario level, that no simple web-hosted information was available and a policy form had to be resorted to in order to include the province's data in this modelling approach.

Another weakness is inherent to the methodology used. Topic modelling is an unsupervised machine learning technique, and so the analysis of topics by a human based upon limited keywords is subject to much misinterpretation. Even strong models trained on ample data are typically unable to distinguish between varied contexts for written information (i.e. written for policy, as news items, or as a fact-sheet).

5.3 Next steps

Future work should apply topic models in a predictive capacity on new data to observe the changing discussion around SCS. It would be useful to examine the impact of different approaches, such as the Toronto model versus other models, on outcomes such as client engagement, treatment success rates, and community perceptions. In addition, by training models on future data might reveal the changing focus on topics related to SCS. Topic modelling could even be leveraged to predict the closure of sites to help their users to prepare or voice their concerns.

Another useful development in natural language processing would be to develop a lexicon specific to Canadian government information sources on public services to be applied. In so doing, sentiment analysis could be applied directly to tokens related to safe injection and consumption practices, providing these services, legal exemptions to drug laws at SCS sites, the advocacy for safe access to controlled substances, and harm reduction in general. While these techniques are certainly not fool-proof, as in this work, they form a strong basis for analyzing the discourse around important issues.

6 Appendix

Table 5, Table 6 and Table 7 show the metrics calculated at each number of topics for the Toronto, Ontario and Canada models, respectively. Note that the Cao et al. (2009) is optimized by minimizing the value, and Deveaud, SanJuan, and Bellot (2014) by maximizing.

Table 5: Metric-based model optimization of the Toronto Model

topics	CaoJuan2009	Deveaud2014
20	0.1755674	0.7715197
19	0.1785334	0.7796329
18	0.1752744	0.7794694
17	0.1669828	0.8572167
16	0.1700758	0.8593863
15	0.1601185	0.8755447
14	0.1449512	0.9385367
13	0.1441014	0.9778194
12	0.1483344	0.9987254
11	0.1508316	1.0668794
10	0.1347947	1.1135168
9	0.1361260	1.1569931
8	0.1268668	1.2107909
7	0.0929003	1.3046493
6	0.0984893	1.3367652
5	0.1050836	1.3733984
4	0.1078580	1.4098031
3	0.1091201	1.4575311
2	0.1090383	1.5356177

Table 6: Metric-based model optimization of the Ontario Model

topics	CaoJuan2009	Deveaud2014
20	0.1998366	0.6500101
19	0.1950242	0.6738425
18	0.1865549	0.7032067
17	0.1746502	0.7461683
16	0.1786575	0.7512475
15	0.1611234	0.8144619
14	0.1657940	0.8152898
13	0.1600138	0.8594299
12	0.1475022	0.9009598

topics	CaoJuan2009	Deveaud2014
11	0.1303379	0.9747933
10	0.1275075	1.0168412
9	0.1145948	1.1006002
8	0.1155441	1.1455266
7	0.1013386	1.2219972
6	0.0921233	1.3228473
5	0.0934574	1.3991338
4	0.0740100	1.4963244
3	0.0655188	1.5673877
2	0.0523067	1.6654106

Table 7: Metric-based model optimization of the Canada Model

topics	CaoJuan2009	Deveaud2014
20	0.2043437	0.6557433
19	0.1880086	0.6936370
18	0.1728755	0.7478458
17	0.1855159	0.7323248
16	0.1659653	0.7803579
15	0.1736817	0.8094403
14	0.1717787	0.8213332
13	0.1666347	0.8860136
12	0.1405035	0.9473498
11	0.1483070	0.9516155
10	0.1473401	1.0155421
9	0.1263029	1.1882482
8	0.1148571	1.1567647
7	0.1115937	1.2232419
6	0.1150523	1.2645895
5	0.1241868	1.3499766
4	0.1357102	1.3846018
3	0.1203308	1.4895360
2	0.1094343	1.6159443

References

- Canada, Health. 2021. “Government of Canada.” *Canada.ca.* / Gouvernement du Canada. <https://www.canada.ca/en/health-canada/services/substance-use/supervised-consumption-sites/apply/how-to-apply.html>.
- . 2023. “Government of Canada.” *Canada.ca.* / Gouvernement du Canada. <https://www.canada.ca/en/health-canada/services/substance-use/supervised-consumption-sites/expained.html>.
- Cao, Juan, Tian Xia, Jintao Li, Yongdong Zhang, and Sheng Tang. 2009. “A Density-Based Method for Adaptive LDA Model Selection.” *Neurocomputing* 72 (7-9): 1775–81. <https://doi.org/10.1016/j.neucom.2008.06.011>.
- Deveaud, Romain, Eric SanJuan, and Patrice Bellot. 2014. “Accurate and Effective Latent Concept Modeling for Ad Hoc Information Retrieval.” *Document Numérique* 17 (1): 61–84. <https://doi.org/10.3166/dn.17.1.61-84>.
- Feinerer, Ingo, Kurt Hornik, and David Meyer. 2008. “Text Mining Infrastructure in r.” *Journal of Statistical Software* 25 (5): 1–54. <https://doi.org/10.18637/jss.v025.i05>.
- Fellows, Ian. 2018. *Wordcloud: Word Clouds*. <https://CRAN.R-project.org/package=wordcloud>.
- Grün, Bettina, and Kurt Hornik. 2011. “topicmodels: An R Package for Fitting Topic Models.” *Journal of Statistical Software* 40 (13): 1–30. <https://doi.org/10.18637/jss.v040.i13>.
- Jockers, Matthew L. 2015. *Syuzhet: Extract Sentiment and Plot Arcs from Text*. <https://github.com/mjockers/syuzhet>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Nikita, Murzintcev. 2020. *Ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*. <https://CRAN.R-project.org/package=ldatuning>.
- Ontario, Government of. 2018. “Consumption and Treatment Services: Application Guide.” *Health.gov.on.ca*. Ministry of Health; Long Term Care. https://health.gov.on.ca/en/pro/programs/opioids/docs/CTS_application_guide_en.pdf.
- Ooms, Jeroen. 2023. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rinker, Tyler W. 2021. *sentimentr: Calculate Text Polarity Sentiment*. Buffalo, New York. <https://github.com/trinker/sentimentr>.
- Silge, Julia, and David Robinson. 2016. “Tidyttext: Text Mining and Analysis Using Tidy Data Principles in r.” *JOSS* 1 (3). <https://doi.org/10.21105/joss.00037>.
- Storopoli, José Eduardo. 2019. “Topic Modeling: How and Why to Use in Management Research.” *Revista Ibero-Americana de Estrategia* 18 (3): 316–38. <https://doi.org/10.5585/ijsm.v18i3.14561>.
- Toronto, City of. 2018. “Toronto Public Health Submits Application to Continue Providing Supervised Injection and Treatment Services.” *Toronto.ca*. <https://www.toronto.ca/news/toronto-public-health-submits-application-to-continue-providing-supervised-injection->

- and-treatment-services/.
- . 2022. “Supervised Consumption Services.” *Toronto.ca*. City of Toronto. <https://www.toronto.ca/community-people/health-wellness-care/health-programs-advice/supervised-injection-services/>.
- . 2023. “Consumption & Treatment Services (CTS) Inspection Program.” *Toronto.ca*. <https://www.toronto.ca/community-people/health-wellness-care/health-inspections-monitoring/consumption-treatment-services-cts-inspection-program/>.
- Wickham, Hadley. 2022. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://CRAN.R-project.org/package=rvest>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2021. *Xml2: Parse XML*. <https://CRAN.R-project.org/package=xml2>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Ziegler, Bianca R, Alexander JD Wray, and Isaac Luginaah. 2019. “The Ever-Changing Narrative: Supervised Injection Site Policy Making in Ontario, Canada.” *International Journal of Drug Policy* 74: 98–111. <https://doi.org/10.1016/j.drugpo.2019.09.006>.