Ontology-Driven Parliamentary Analytics: Analysing Political Debates on COVID-19 Impact in Canada

Sabrina Azzi[1][0000-0003-2448-9819] and Stéphane Gagnon[2][0000-0001-7732-9849]

[1] University of West of Scotland, Scotland, UK
sabrina.azzi@uws.ac.uk
[2] Université du Québec en Outaouais, Gatineau, Canada
stephane.gagnon@uqo.ca

Abstract. Parliamentary debates are usually published in Parliament's websites to allows citizens to be informed on the latest national debates, proposals and decisions**.** To enhance citizen experience and engagement, functionalities such as debates annotation and question answering are necessary. Annotating text requires semantic content and ontologies are known for their ability to describe a common vocabulary for a domain and can be a solid base for annotation and question answering. We report on an ongoing study to enhance parliamentary analytics using an ontology and knowledge graph to sharpen annotations and facilitate their query by end-users. As a salient case, a sample of debates are collected on the COVID-19 impact in Canada, as its complexity shows the relevance of using advanced knowledge representation techniques. We focused on the development of a new "Impact of COVID-19 in Canada Ontology" (ICCO) that provides contextualized semantic information on impact in numerous policy areas, as this ontology is entirely built from Canadian parliamentary debates. It has been evaluated and validated by experts. Our conclusion underscores the importance of integrating ontology-driven parliamentary analytics within the broader context of digital transformation in legislative institutions, and the need for new platforms supporting free and open Digital Humanities.

Keywords: Ontology, COVID-19, Parliamentary debate.

1 Introduction

In the latest three years, COVID-19 accelerated the phenomenon of digital transformation and become a pure necessity while it was before a just technological opportunity. Parliaments and governments known about their resistance and fear in adopting new technologies [1] have showed a great interest in incorporating innovative technologies in their processes and services towards electronic automation [2-5]. Moreover, with opening their non-confidential data online, enhancing citizens experience and engaging them is necessary to improve the quality of public policy-making and public services provisioning [6]. Websites are one of the means to promote parliamentary

public engagement and being increasingly used by citizens makes it incumbent on parliaments to implement modern technologies for remaining relevant and friendly [7].

House of Commons of Canada (HoC) [8] recently invested in developing new technology functionalities in the context of legislative intelligent (LegisIntel) solutions project [3]. LegisIntel refers to Artificial Intelligence (AI) and analytical tools adopted by parliaments to enhance citizen experience in monitoring interrelations among diverse contents of parliamentary proceedings [3]. Text annotation that consists of reading and assigning labels to parts of the text is one of the solutions being studied within this project.

The present work is part of this project and consists of the first phase focused on developing an ontology to, among other, annotate text automatically, i.e., parliamentary debates. In fact, Canadian parliamentary debates are manually annotated, and annotation experts work daily on reviewing each new debate and annotate. This process is expensive, time consuming and usually prone to errors resulting from several factors such as user familiarity, personal motivation, etc. Annotating text requires semantic content and ontologies are known for their ability to describe a common vocabulary for a domain and define the meaning of the concepts and the relations between them. An ontology is defined as 'a formal and explicit specification of a shared conceptualization' [9]. Ontologies are also used for sharing, reusing, and integrating knowledge. Ontology development has been very active during the last five years in many sectors including electronic government and parliament due to the increasingly need of semantic, homogeneity, reasoning, etc.

COVID-19 is one of the effervescent topics in parliamentary discussions in the latest three years as it exponential spread has not only impacted public health [10] but also economy [11], labour market [12], education [13], etc. Using 'COVID-19 impact' as a keyword for search, search engine returned 955 long interventions published between 23 November 2021 and 31 March 2023 (see **Fig 1**).

After discussions with information management team within the HoC, an ontology that covers the impact of COVID-19 in Canada is the best domain to start with to relieve annotation experts.

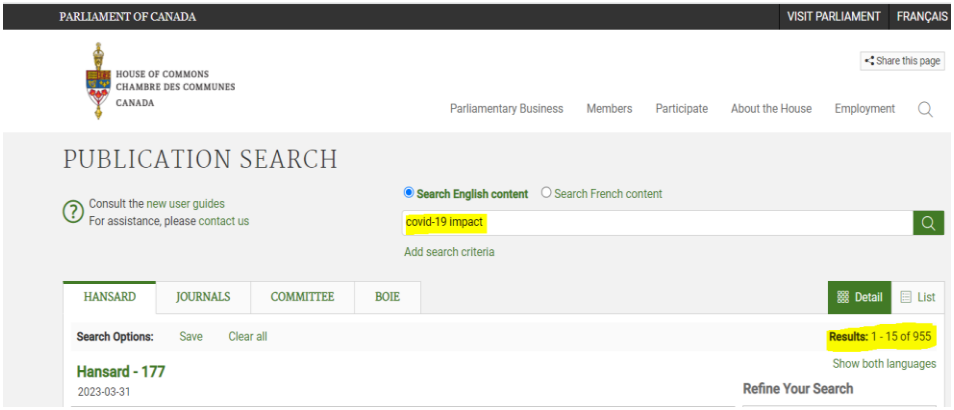

Fig 1. Number of interventions obtained by the House of Commons of Canada website when searching the impact of COVID-19 in Canada.

This paper describes how we created the 'Impact of COVID19 in Canada Ontology' (ICCO). The next section outlines a related work. Materials and methods are presented in section 2. Our ICCO development started with defining the domain and scope of ontology (phase 1). We used Canadian parliamentary debates as a primary knowledge source (phase 2) to build preliminary ICCO (phase 3). Parliamentary debate are formal discussions of proposals that the members of Parliament (MP) take in turns to speak. These debates are then transcribed, edited, corrected, and build up a Hansard. The preliminary ICCO has been subsequently refined through the reuse of the elements of other ontologies (phase 4). Finally, we evaluated ICCO using parliamentary debates and by involving domain experts (phase 5). Results and discussion are presented in section 4 and conclusion is section 5.

2 Related work

Ontology-driven annotation of parliamentary debates depends heavily on reliable corpora and language features that can be accurately linked to ontology concepts and instances. For this reason, while ontologies have not yet been an extensive focus in this area, there has been important efforts by several Natural Language Processing (NLP) and computational linguistics researchers that pave the way for the present research. They can be classified in three groups depending on what language features they focus on: corpus integration, linguistic annotations, and argumentation mining.

At the corpus level, European teams have been among the most active in developing parliamentary corpora, especially with multilingual features and models [14]. A major example is ParlaMint, a multilingual annotated corpus available through the CLARIN.SI repository [15]. National parliamentary traditions can influence corpus contents and organization, and as such more targeted studies can allow more diverse feature extractions for AI applications. Many countries, again in Europe, have developed such corpora that enable advanced features, including Czech Republic [16], Denmark [17], Finland [18], Germany [19], Norway [20], Portugal [21], and Slovenia [22]. These corpus integration efforts include national language disambiguation and evolution analysis, while ensuring policy debates are contextualized in time and context.

At the linguistic level, annotations of parliamentary debates have offered the opportunity to use of linguistic metadata, which can be extracted reliably using NLP and deep learning methods, linking entities to knowledge graphs [23]. Typical annotations include speaker identification, debate context, keywords, and especially relations to various entities such as laws and policies. Unification of these features within a single corpus can help greatly reduce the semantic annotation task in later phases and provide a partial gold standard for classification tasks given the highly reliable curation process of these metadata. While AI applications are still limited in parliaments, there some countries have focused on semantic relation formalization and relationships extraction, while ensuring their reliable reuse, such as efforts in Chile [24-26], Greece [27-29], United Kingdom [30-32], and United States [33-35]. These studies rely on semantic technologies of varying maturity given they span a long period and have taken distinct approaches. But they share the common goal of simplifying annotation schemas,

ensuring entities are accurately related to one another, facilitate government interoperability and legal-policy transparency, monitor legislative productivity and voting patterns, and used schemas to enrich debate analytics and search features.

At the discourse level, interactions between parliamentary speakers can be modelled through argument mining [36]. This approach involves developing typical patterns of speaker interactions, and identifying frequent associations among various linguistic features, whether words, entities, noun phrases, relationships, identities, and policy-related contexts. Taken within a flow of interactions, they can reveal how arguments take form, and how they influence legislative outcomes. Notable studies, relaying on varying degrees on computational linguistics, have been carried out in parliaments of Estonia [37, 38], the European Union [27, 39, 40], India [41], Ireland [42], Latvia [43], and Sweden [44]. Models of parliamentary arguments are increasingly recognizable, with potential to enrich ontologies with discourse patterns.

The present study builds upon recent advances at all three levels of parliamentary analytics. First, corpus development for our study relies on a rigorously, professionally curated bilingual and translated (English and French) corpus from the Canadian Hansard. We are therefore given assurances that the corpus and its keywords are a true gold standard with which our ontology annotations can be compared. Second, linguistic features of the debate samples rely again on existing relationships embedded in debate metadata. This allows us to draw COVID19-related debate contexts and ensure that its impacts are clearly identified as per the policy context of Canada. Third, as we plan pursuing discourse analysis in later stages, our debates will be annotated further using more complex semantic relationships within our new ontology, such as linking policy actions to public health outcomes, and therefore representing predefined discourse patterns as per the latest research on argument mining. As well, given a successful next stage in our research, we will focus our attention on measuring the use of "evidence-based policy making", i.e., the extent to which parliamentarians rely on complex fact-checking and knowledge understanding to ensure reliable decisions to fight the pandemic.

3 Materials and Methods

Many methodologies for building ontologies were proposed [45-49]. In [50], a new methodology was proposed and covers all the necessary phases to build an ontology of quality with providing guidance for the knowledge engineer. Unlike other methods that do not fully adhere to the Institute of Electrical and Electronics Engineers (IEEE) standards [51], this methodology supports a rigorous process of structuring the ontological concepts and facilitating modularization. This methodology has been previously used in building an ontology of quality that was evaluated by experts and also published [52].

The process of ICCO development is composed of five phases as illustrated in **Fig 2**. During the process, we followed well-established Open Biological and Biomedical Ontology (OBO) Foundry principles [53] and the ARCHitecture for ONTological Elaborating (ARCHONTE) method to create preliminary ICCO.

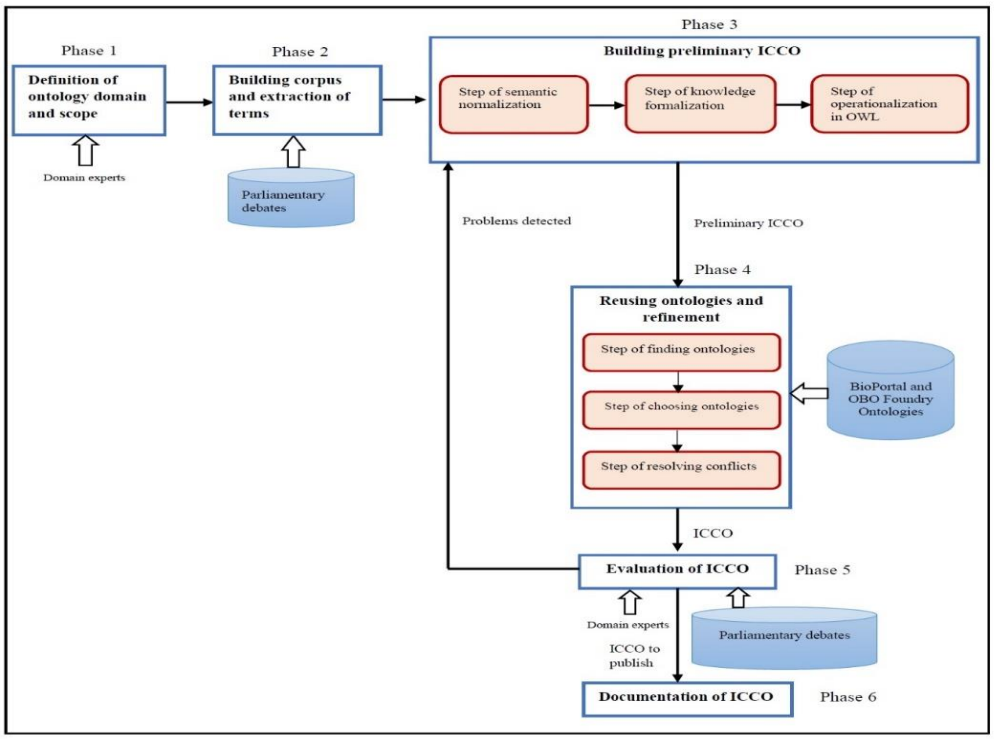

Fig 2. ICCO development process.

3.1 Phase 1: Definition of ontology domain and scope

This phase consists of establishing the domain of interest of the ontology to model and its scope. It is done by consulting domain experts within the HoC. First, two fundamental questions must be answered:

1. What is the domain that the ontology will cover?
2. What is main purpose of developing the ontology?

Second, functional requirements for the ontology or questions that an ontology should be able to answer are defined. In order to support this, competency questions (CQs) established by domain experts can be used. Indeed, these CQs are presented in the form of informal queries that a particular community of users believes that the ontology should answer.

We defined the domain of ontology and scope after discussions with information management experts within the HoC. The impact of COVID-19 in Canada, which is a very of effervescent topic in parliamentary debates in the latest three years, was identified. The following CQs were established by information management experts withing the HoC:

1) CQ1- What is COVID-19?

2) CQ2- What are the different sectors impacted by COVID-19 in Canada?
3) CQ3- What is the impact of COVID19 on Canada's economy?
4) CQ4- What is the impact of COVID-19 on Canadian mental health?
5) CQ5- What is the impact of COVID-19 on Canadian public health?
6) CQ6- What is the impact of COVID-19 on Canada's healthcare system?
7) CQ7- What is the impact of COVID-19 on Canada's financial system (banking…)?
8) CQ8- What is the impact of COVID-19 on Canada's education?
9) CQ9- What is the impact of COVID-19 on Canada's real estates?
10) CQ10- What is the impact of COVID-19 on Canada's labour market?
11) CQ11- What is the impact of COVID-19 on Canada's immigration?
12) CQ12- What is the impact of COVID-19 on Canada's tourism?
13) CQ13- What is the impact of COVID-19 on Canada's legislation?
14) CQ14- What is the impact of COVID-19 on Canada's media?
15) CQ15- What is the impact of COVID-19 on Canada's public safety?
16) CQ16- What is the impact of COVID-19 on Canada's federal-provincial relations?

3.2 Phase 2: Building corpus and extraction of terms

In this phase, we extracted and identified relevant terms to the ontology domain defined in phase 1, from a corpus of knowledge. The corpus of knowledge is constituted of publicly available parliamentary debates in Hansard [54] latter is a verified and accurate record rather than verbatim transcript. Redundancies are removed, recognizable mistakes are corrected, and interjections omitted unless responded to by the principal speaker. Our corpus of knowledge is composed of 168 members of Parliament interventions from December 2021 to March 2022.

We used Text2Onto [55] tool that offers good results to extract terms from the debates. Text2Onto has been developed to support the acquisition of ontologies from textual documents. It offers three extraction algorithms calculating the following measures: Relative Term Frequency (RTF), Term Frequency Inverted Document Frequency (TFIDF), Entropy, and the C-value/NC-value. TFIDF gives the best results because it evaluates how relevant a term is to a document in a collection of documents.

3.3 Phase 3: Building preliminary ICCO

This phase was dedicated to the building of a preliminary ICCO (p-ICCO) by following the ARCHONTE methodology. The strong point of ARCHONTE is the recommendation of semantic commitment in the design process. In other words, explaining the meaning of each of the ontology concepts, the similarities and differences that the concept has with those close to it using natural language. The ontology structure is similar to a tree structure, which facilitates the determination of the meaning that a concept has according to its relative position. ARCHONTE is composed of three steps: 1) semantic

normalization; 2) knowledge formalization; and 3) toward a computational ontology. They are elaborated below.

Step of semantic normalization

The objective of this step is to reach a semantic agreement about the meaning of the concept labels. We applied differential principles on the set of candidate concepts chosen previously. According to the differential paradigm proposed by Bachimont [56], the meaning of a concept (or a node, since an ontology is structured as a tree) is determined by its closest neighbors: parent and its siblings, *i.e*., its position in the ontological hierarchy based on terminological structure. Similarities and differences of each concept concerning its parent and its siblings are expressed in natural language. Four differential principles are distinguished: 1) similarity with each parent; 2) similarity with each sibling; 3) difference with each sibling; 4) difference with each parent. The result of this step is a so-called differential ontology. We applied these semantic principles to create a hierarchy of concepts and relationships organized according to their similarities and differences. For example: *primary school, secondary school* and *university* are sibling concepts because of their similarity (education), and at the same time, they are different concepts because represent different level of education.

Step of knowledge formalization

The objective of this step is to link each concept to a set of objects that allows defining new concepts and relations through set operations. During this step, we specified the arity and domains of the relations, in correspondence with the intended models. We added some logical axioms to constrain the domains of the relations. The result of this step was an ontology that covered differentials concepts, relations, instances, and axioms, such as subclass ones, for the impact of COVID-19 in Canada domain.

Step of operationalization

The objective of this step is implementing the ontology in an operational language of knowledge representation such as OWL. We used Protégé [57], we operationalized p-ICCO in OWL language because it perfectly meets our needs in terms of expressiveness and manageability. At this last step of the ARCHONTE method, we obtained the final version of p-ICCO.

3.4 Phase 4: Reusing ontologies and refinement

In this phase, we added and detailed concepts in p-ICCO. For this purpose, we looked for other relevant ontologies. These ontologies had to be evaluated for the suitability of their content coverage and the depth of knowledge they represent. Unfortunately, no ontology responded to the requirements of ICCO except DOID [58] that covers human diseases. We used Wikipedia to enrich concepts with definitions and synonyms.

3.5 Phase 5: Evaluation of ICCO

Ontology evaluation is the process of assessing the quality of an ontology concerning a set of evaluation criteria and its adequacy for being used in a specific context for a specific goal. A classification of ontology evaluation was proposed by Brank and his colleagues in [59], and it consists of four categories: 1) approaches that use the target ontology in an application and evaluate the application results [60]; 2) approaches that compare the target ontology to a gold standard [61]; 3) those that use data sources about a specific domain [62]; and 4) approaches that recommend a manual assessment by domain experts according to a set of criteria [63].

Internal consistency and adherence to standard ontology practices

We used Pellet reasoner integrated with PROTÉGÉ to check the internal consistency of ICCO, and no errors were detected. We manually verified the full adherence of ICCO to OBO Foundry guidelines and Cimino's desiderata [64].

Evaluation using dependent domain sources

We used new 126 member of Parliament interventions from April 2022 to June 2022. The goal of this step was to allow us to verify each concept from an intervention that may refer to the domain, whether it is addressed in the ontology or not. The concepts concerned by the annotation process include adjectives, adverbs, proper nouns, nouns, verbs, verbal phrases, nominal phrases, adjectival phrases, and participatory phrases. At the end of this step, 39 new concepts were added to ICCO. This step was manual.

We also evaluated ICCO by using an annotator developed within our team. The annotator was connected to ICCO and for each intervention, it highlights concepts related to the impact of COVID-19 in Canada by providing definitions. All needed concepts were annotated.

Evaluation by domain experts

The evaluation is done by two domain experts, from information and management team within the HoC, to assess how well the ontology meets a set of predefined requirements. They recommended to add more relations between financial system and economy and axioms. For example: *deficit*, *government spending*, and *inflation* that had *financial system* as parent should also have *economy* as parent.

3.6 Phase 6: Documentation of ICCO

We annotated every concept in ICCO and used OWLDoc to document the ontology. Every phase of the building process is described in this work.

4 Results and discussion

ICCO contains 501 classes, 41 object properties, 4069 axioms, 56 individuals (instances) and 19 annotation properties. **Fig 3** shows an overview of ICCO. **Fig 4**

illustrates the main classes of ICCO demonstrating the hierarchy of class tourism. The object properties are presented in **Fig 5**. For example, the object property *provide* asserts the relation between a vaccine and an immunity *vaccine **provide** immunity.*

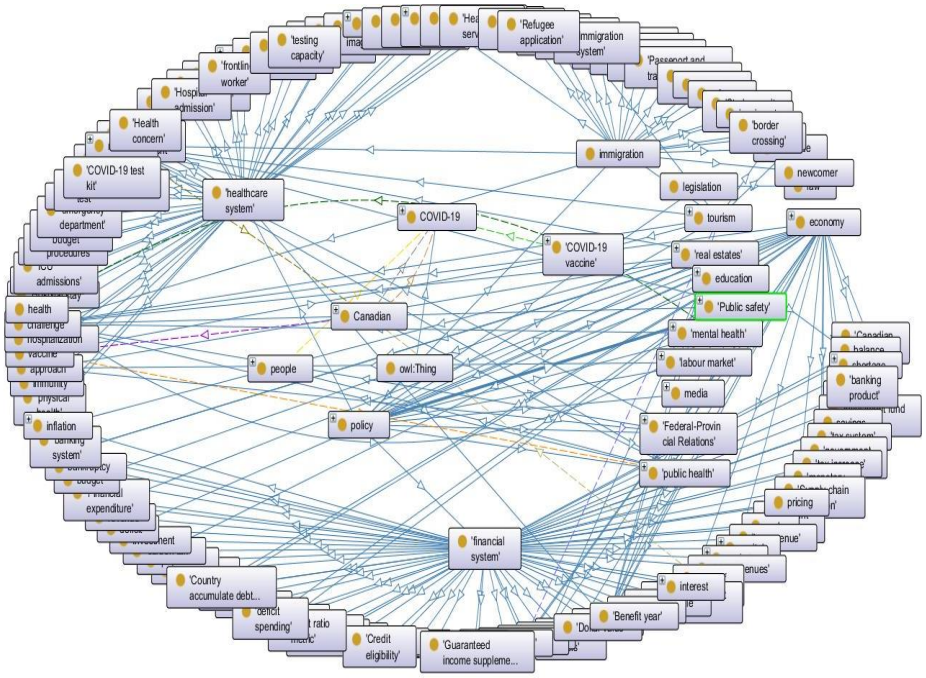

Fig 3. ICCO overview.

The resulted ICCO covers the impact of COVID on education, healthcare system, public health, public safety, mental health, real estates, labor market, tourism, media, legislation, federal-provincial relations, immigration, economy, and financial system.

The phase 3 was the most consuming during the process of ICCO development. This phase required semantic commitment in analyzing each concept regarding its siblings and parents. The phase 4 was the quickest one as no ontology responded to ICCO requirements. Therefore, we did not benefit from ontology reuse which would have enrich ICCO with concepts and definitions, synonyms, etc., and increase reusability and interoperability with existing ontologies.

For the interoperability and reusability purposes, all ICCO concepts were annotated with standard identifiers formatted as recommended by OBO Foundry, synonyms, and definitions collected as well from Wikipedia, where such information is available.

The main limitation of ICCO is the lack of modular structure. A modular ontology design offers numerous benefits, such as easy reuse of a single component in other ontologies, good reasoning, facilitating ontology understanding, reduced complexity, to name a few. While there were attempts to define how to select “the best modular ontology”[65], currently there are no tools supporting the creation of modular ontologies. This limitation applies to the majority of ontologies publicly available.

Fig 4. ICCO main classes and tourism hierarchy.

Fig 5. ICCO object properties

5 Conclusion

This study has presented a new ontology specifically targeted for integration within an ontology-driven parliamentary debate annotation platform. The new "Impact of COVID-19 in Canada Ontology" (ICCO) provides contextualized semantic information on impact in numerous policy areas, as this ontology is entirely built from Canadian parliamentary debates. We began with a review of the latest advances in the field of legislative debate analytics, classifying studies at the levels of corpus integration, linguistic annotation, and argument mining. We followed with an outline of the materials and methods of our study, namely the project lifecycle for ontology design, implementation, and evaluation. Findings from this effort were then discussed within the context of parliamentary and legislative intelligence.

The potential impact of ontology-driven intelligence can be positioned within the broader context of digital transformation in parliaments [66], where AI offers promising opportunities, enabling seamless integration across parliamentary and government processes. Ontologies and knowledge graphs are essential assets in formalizing artificial intelligence capabilities in various operational contexts. Given that parliaments are primarily discursive environments, they present less constraints in data, information, and knowledge modelling. These solutions can be improved by integrating specialized ontologies to provide contextualized policy-specific analytics,

such as the one presented here on COVID19 impact in Canada. Their systematic reuse, and continuous enrichment, can also serve to rapidly enhance user experience in legislative intelligence applications.

References

1. Meijer, A., *E-governance innovation: Barriers and strategies.* Government Information Quarterly, 2015. **32**(2): p. 198-206.
2. Leventis, S., V. Anastasiou, and F. Fitsilis. *Application of enterprise integration patterns for the digital transformation of parliamentary control*. in *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*. 2020.
3. Gagnon, S. and S. Azzi. *Semantic Annotation of Parliamentary Debates and Legislative Intelligence Enhancing Citizen Experience*. 2022. Cham: Springer International Publishing.
4. Al-Mushayt, O.S., *Automating E-government services with artificial intelligence.* IEEE Access, 2019. **7**: p. 146821-146829.
5. Engstrom, D.F., et al., *Government by algorithm: Artificial intelligence in federal administrative agencies.* NYU School of Law, Public Law Research Paper, 2020(20-54).
6. Purwanto, A., A. Zuiderwijk, and M. Janssen, *Citizen Engagement With Open Government Data: A Systematic Literature Review of Drivers and Inhibitors.* International Journal of Electronic Government Research (IJEGR), 2020. **16**(3): p. 1-25.
7. Serra-Silva, S., *How parliaments engage with citizens? Online public engagement: a comparative analysis of Parliamentary websites.* The Journal of Legislative Studies, 2022. **28**(4): p. 489-512.
8. *House of Commons of Canada*. 2023 [cited 2023 12 April]; Available from: https://www.ourcommons.ca/en.
9. Studer, R., V.R. Benjamins, and D. Fensel, *Knowledge engineering: Principles and methods.* Data & knowledge engineering, 1998. **25**(1-2): p. 161-197.
10. Levitt, E.E., et al., *Public health guideline compliance and perceived government effectiveness during the COVID-19 pandemic in Canada: Findings from a longitudinal cohort study.* The Lancet Regional Health-Americas, 2022. **9**: p. 100185.
11. Beland, L.P., et al., *The short-term economic consequences of COVID-19: Occupation tasks and mental health in Canada.* Canadian Journal of Economics/Revue canadienne d'économique, 2022. **55**: p. 214-247.
12. Lemieux, T., et al., *Initial impacts of the COVID-19 pandemic on the Canadian labour market.* Canadian Public Policy, 2020. **46**(S1): p. S55-S65.
13. Buckner, E., Y. Zhang, and G.L. Blanco, *The impact of COVID-19 on international student enrolments in North America: Comparing Canada and the United States.* Higher Education Quarterly, 2022. **76**(2): p. 328-342.

14. Gartner, R., *Using structured text corpora in Parliamentary Metadata Language for the analysis of legislative proceedings.* Digital Humanities Quarterly, 2018. **12**(2).
15. Erjavec, T., et al., *The ParlaMint corpora of parliamentary proceedings.* Language Resources and Evaluation, 2022.
16. Kratochvíl, J., P. Polák, and O. Bojar. *Large corpus of czech parliament plenary hearings*. in *12th International Conference on Language Resources and Evaluation, LREC 2020*. 2020. European Language Resources Association (ELRA).
17. Kirkedal, A., M. Stepanovic, and B. Plank. *FT SPEECH: Danish parliament speech corpus*. in *21st Annual Conference of the International Speech Communication Association, INTERSPEECH 2020*. 2020. International Speech Communication Association.
18. Mansikkaniemi, A., P. Smit, and M. Kurimo. *Automatic construction of the Finnish parliament speech corpus*. in *18th Annual Conference of the International Speech Communication Association, INTERSPEECH 2017*. 2017. International Speech Communication Association.
19. Blätte, A. and A. Blessing. *The Germaparl corpus of parliamentary protocols*. in *11th International Conference on Language Resources and Evaluation, LREC 2018*. 2019. European Language Resources Association (ELRA).
20. Lapponi, E., et al., *The Talk of Norway: a richly annotated corpus of the Norwegian parliament, 1998–2016.* Language Resources and Evaluation, 2018. **52**(3): p. 873-893.
21. Almeida, P., M. Marques-Pita, and J. Gonçalves-Sá, *PTPARL-D: An annotated corpus of forty-four years of Portuguese parliamentary debates.* Corpora, 2021. **16**(3): p. 337-348.
22. Fišer, D., N. Ljubešić, and T. Erjavec, *Parlameter - A corpus of contemporary slovene parliamentary proceedings.* Prispevki za Novejso Zgodovino, 2019. **59**(1): p. 70-98.
23. Lu, F., P. Cong, and X. Huang, *Utilizing Textual Information in Knowledge Graph Embedding: A Survey of Methods and Applications.* IEEE Access, 2020. **8**: p. 92072-92088.
24. Cifuentes-Silva, F., et al., *Describing the nature of legislation through roll call voting in the chilean national congress, a linked dataset description.* Semantic web, 2021.
25. Cifuentes-Silva, F., D. Fernández-Álvarez, and J.E. Labra-Gayo, *National Budget as Linked Open Data: New Tools for Supporting the Sustainability of Public Finances.* Sustainability, 2020. **12**(11): p. 4551.
26. Cifuentes-Silva, F. and J.E. Labra Gayo, *Legislative document content extraction based on semantic web technologies: A use case about processing the history of the law*, in *ESWC 2019*. 2019, Springer Verlag. p. 558-573.
27. Loukis, E., et al. *Argumentation systems and ontologies for enhancing public participation in the legislation process*. in *EGOV*. 2007. na.

28. Loukis, E. and A.L. Xenakis. *A methodology for ontology-based knowledge-level interoperability among parliaments*. in *15th Americas Conference on Information Systems 2009, AMCIS 2009*. 2009.
29. Figgou, L. and E. Andreouli, *Appeals to Consensus and Partisan Politics in Parliamentary Discourse on the Pandemic.* Political Psychology, 2022.
30. Nanni, F., et al. *Semantifying the UK Hansard (1918--2018)*. 2019. IEEE Press.
31. Abercrombie, G. and R. Batista-Navarro. *Semantic Change in the Language of UK Parliamentary Debates*. 2019. Association for Computational Linguistics.
32. Coole, M., P. Rayson, and J. Mariani. *Unfinished Business: Construction and Maintenance of a Semantically Tagged Historical Parliamentary Corpus, UK Hansard from 1803 to the present day*. 2020. European Language Resources Association.
33. Ballard, A.O., *Bill Text and Agenda Control in the US Congress.* Journal of Politics, 2022. **84**(1): p. 335-350.
34. Atkinson, M.L., R. Mousavi, and J.H. Windett, *Detecting Diverse Perspectives: Using Text Analytics to Reveal Sex Differences in Congressional Debate About Defense.* Political Research Quarterly, 2022.
35. Xing, Z., S. Hillygus, and L. Carin. *Evaluating U.S. Electoral representation with a joint statistical model of congressional roll-calls, legislative text, and voter registration data*. in *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2017*. 2017. Association for Computing Machinery.
36. Lawrence, J. and C. Reed, *Argument mining: A survey.* Computational Linguistics, 2019. **45**(4): p. 765-818.
37. Koit, M. *How are the members of a parliament arguing? Analysis of an argument corpus*. in *13th International Conference on Agents and Artificial Intelligence, ICAART 2021*. 2021. SciTePress.
38. Koit, M. *Annotating arguments in a parliamentary corpus: An experience*. in *12th International Conference on Knowledge Discovery and Information Retrieval, KDIR 2020 - Part of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2020*. 2020. SciTePress.
39. Defrancq, B., *The European Parliament as a discourse community: Its role in comparable analyses of data drawn from parallel interpreting corpora.* Interpreters Newsletter, 2018(23): p. 115-132.
40. Calzada-Pérez, M., *Researching the european parliament with corpus-assisted discourse studies from the micro- and macro-levels of text to the macro-context.* Revista Espanola de Linguistica Aplicada, 2018. **30**(2): p. 465-490.
41. Sen, A., et al. *Studying the discourse on economic policies in India using mass media, social media, and the parliamentary question hour data*. in *2019 ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2019*. 2019. Association for Computing Machinery, Inc.

42. Burroughs, E., *The Discourse of Controlling "illegal Immigration" in Irish Parliamentary Texts.* Journal of Language and Politics, 2015. **14**(4): p. 479-500.
43. Skulte, I. and N. Kozlovs. *Discourse on safety/security in the parliamentary corpus of latvian saeima*. in *5th Conference Digital Humanities in the Nordic Countries, DHN 2020*. 2020. CEUR-WS.
44. Eide, S.R. *The swedish poligraph: A semantic graph for argument mining of swedish parliamentary data*. in *6th Workshop on Argument Mining, ArgMining 2019, collocated with ACL 2019*. 2019. Association for Computational Linguistics (ACL).
45. Gruninger, M. and M.S. Fox. *The design and evaluation of ontologies for enterprise engineering*. in *Workshop on Implemented Ontologies, European Conference on Artificial Intelligence (ECAI)*. 1994.
46. Uschold, M. and M. King, *Towards a methodology for building ontologies*. 1995: Citeseer.
47. Noy, N.F. and D.L. McGuinness, *Ontology development 101: A guide to creating your first ontology*. 2001, Stanford knowledge systems laboratory technical report KSL-01-05 and ….
48. Staab, S., et al., *Knowledge processes and ontologies.* IEEE Intelligent systems, 2001. **16**(1): p. 26-34.
49. Bachimont, B., A. Isaac, and R. Troncy. *Semantic commitment for designing ontologies: a proposal*. in *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web: 13th International Conference, EKAW 2002 Sigüenza, Spain, October 1–4, 2002 Proceedings 13*. 2002. Springer.
50. Azzi, S., *Nouvelle méthodologie de construction d'ontologies médicales: cas d'étude: diagnostic de la pneumonie*. 2021, Université du Québec en Outaouais. p. 135.
51. Schultz, D.J., *IEEE standard for developing software life cycle processes.* IEEE Std, 1997: p. 1074-1997.
52. Azzi, S., W. Michalowski, and M. Iglewski, *Developing a pneumonia diagnosis ontology from multiple knowledge sources.* Health Informatics Journal, 2022. **28**(2): p. 14604582221083850.
53. *Open Biological and Biomedical Ontology Foundry*. 2023 [cited 3 April 2023; Available from: https://obofoundry.org/principles/fp-000-summary.html.
54. *Hansard*. 2023 [cited 2023 19 April]; Available from: https://www.ourcommons.ca/PublicationSearch/en/?PubType=37.
55. Cimiano, P. and J. Völker. *text2onto*. in *International conference on application of natural language to information systems*. 2005. Springer.
56. Bachimont, B., A. Isaac, and R. Troncy. *Semantic commitment for designing ontologies: a proposal*. in *International Conference on Knowledge Engineering and Knowledge Management*. 2002. Springer.
57. *PROTÉGÉ*. 2023 [cited 2023 20 April].

58. *Human Disease Ontology*. 2023 [cited 2023; Available from: https://bioportal.bioontology.org/ontologies/DOID.
59. Brank, J., M. Grobelnik, and D. Mladenic. *A survey of ontology evaluation techniques*. in *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*. 2005. Citeseer Ljubljana, Slovenia.
60. Porzel, R. and R. Malaka. *A task-based approach for ontology evaluation*. in *ECAI Workshop on Ontology Learning and Population, Valencia, Spain*. 2004. Citeseer.
61. Maedche, A. and S. Staab. *Measuring similarity between ontologies*. in *International Conference on Knowledge Engineering and Knowledge Management*. 2002. Springer.
62. Brewster, C., et al., *Data driven ontology evaluation.* 2004.
63. Lozano-Tello, A. and A. Gómez-Pérez, *Ontometric: A method to choose the appropriate ontology.* Journal of Database Management (JDM), 2004. **15**(2): p. 1-18.
64. Cimino, J.J., *Desiderata for controlled medical vocabularies in the twenty-first century.* Methods Inf Med, 1998. **37**(4-5): p. 394-403.
65. Kumar, S. and N. Baliyan, *Quality Evaluation of Ontologies*, in *Semantic Web-Based Systems*. 2018, Springer. p. 19-50.
66. Koryzis, D., et al., *Parltech: Transformation framework for the digital parliament.* Big Data and Cognitive Computing, 2021. **5**(1).