

Building and Analyzing the Brazilian Legal Knowledge Graph

Rilder S. Pires^{1,*†}, Henrique Santos^{2†}, Ricardo Guedes^{1†}, João A. Monteiro Neto^{1†}, Carlos Caminha^{1†} and Vasco Furtado^{1†}

¹Universidade de Fortaleza, Fortaleza CE, Brazil

²Rensselaer Polytechnic Institute, Troy NY, USA

Abstract

Artificial Intelligence has proven to be effective in streamlining processes in several domains. The Brazilian judiciary, specifically, has a very large number of cases, above the work capacity of the courts, generating urgency in the creation of methods that mainly support the access and manipulation of unstructured data. This paper presents the construction of a Knowledge Graph of the Brazilian Legislation using Semantic Web standards that allows an understanding of how Brazilian laws interact with each other. The Knowledge Graph was quantitatively evaluated using complex network analysis and it was found to be useful to support experts in understanding the Brazilian legislation by detecting special nodes, namely the “bridge-like nodes”, that play an important role in the structure of the graph.

1. Introduction

There is a growing demand for the application of Artificial Intelligence (AI) resources in all branches of human activity. Many of the tasks performed by people can be performed by machines with greater speed and reliability. In this context, Knowledge graphs (KGs) have been successfully applied in many domains as a way to encode knowledge for supporting downstream applications [1]. They represent knowledge in the form of triples *subject - predicate - object* and are constructed based on domain sources, including authoritative documents, regulations, and domain practitioners.

These tools can be effectively applied in the legal domain, where accessing more structured data about legal documents, especially laws, and understanding their connections and relationships can provide useful information not only for legal scholars investigating how legal topics evolve but, also, for the emerging Legal Tech Industry. A well-established source that accurately represents the law, its topics, and relationships, could provide the path to more effective and reliable legal knowledge which could be used to support applications ranging from automatic classification of legal documents to deployment of predictive decision-making support.


Joint Proceedings of ISWC2022 Workshops: the International Workshop on Artificial Intelligence Technologies for Legal Documents (AI4LEGAL) and the International Workshop on Knowledge Graph Summarization (KGSUM) (2022)


*Corresponding author.

†These authors contributed equally.

✉ rilder@unifor.br (R. S. Pires); oliveh@rpi.edu (H. Santos); ricardobmg@edu.unifor.br (R. Guedes); joaoneto@unifor.br (J. A. M. Neto); caminha@unifor.br (C. Caminha); vasco@unifor.br (V. Furtado)

ORCID 0000-0003-4873-5308 (R. S. Pires); 0000-0002-2110-6416 (H. Santos)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

In addition, there are, several approaches for encoding legislation and laws, including as KGs, mainly in Europe [2, 3]. In this sense, using AI to streamline processes in the Courts of Justice is a promising application [4], especially, in Brazil, where the Judiciary has a very large number of cases that overwhelm its current capacity [5] and strengthen the notion of “slow justice”.

In this paper, we present the construction of the Knowledge Graph for Brazilian Federal Legislation through the examination of relationships between laws, including citations, changes, and repeals. For this, we leveraged the LexML data portal, which serves as a source of semi-structured Brazilian legal data [6]. To represent the data as a KG, we explore methods and Web Semantic standards, including established legal vocabularies, to model and build an RDF-based Brazilian Legal Knowledge Graph where nodes represent legal documents and edges the relationships between them. Finally, we use graph theory to extract information from the KG that allowed us to build a profile of the structure of Brazilian legislation. In addition, we test the hypothesis that special nodes, namely the “bridge-like nodes” (nodes that belong to the neighborhood of at least two nodes associated with Brazilian Legal Codes), can be successfully used to identify legal documents that play an important role in the legislation.

2. Related Work

Semantic Web and Data Science applications for the legal domain have been very common in the last 15 years [7, 8, 9, 10, 11, 12, 13]. In general, these applications are focused on the problem of transforming a representation of legal information in natural language into representations of structured data, such as creating ontologies [14], including investigating design patterns to create these knowledge graphs in the legal domain [10]. By using the Resource Description Framework (RDF) as a means of representing legal information, Ebenhoch outlines the challenges of describing legal resources and highlights that the main approach to enriching legal data is to enhance it with metadata [9].

In recent years, systems for processing legal domain content received a lot of attention. In the same measure, efforts have been done on the creation of knowledge graphs for the legal domain. An example of such efforts is the EU-funded project Lynx [15] which has shown many use cases in the legal domain. Despite that, Saias and Quaresma describe the problem of lack of semantics in legal information retrieval systems and propose an ontology to enrich legal data based on semantics in the Portuguese legal system [16]. Although it is possible to observe the proposition of a series of semantic representations of the legal domain in the literature, we identified a gap to be filled with the proposition of a Knowledge Graph of Brazilian Legislation.

3. Legal Data Sourcing / Knowledge Graph Construction

We sourced data by extracting information from the LexML portal [6], a website that specializes in legal and legislative information in Brazil. The portal uses Uniform Resource Names (URNs) [17] to uniquely identify each legal document. URNs are persistent and unambiguous identifications and we leverage them to unify, organize and facilitate access to descriptive information about the legislation. This allowed our approach to remain comprehensive since the URN can be unequivocally and internationally recognized [18].

Legislation type	Quantity
Ordinary Laws	15741
Decrees	164066
Complementary Laws	189
Decree-Laws	12547
Delegated Laws	13
Provisional Measures AE32	6189
Provisional Measures PE32	1080
Constitutions	5
Legislative Decrees	14577
Total	214407

Table 1
Legislation extracted from LexML portal

The data extraction was performed using Python’s *Requests module* [19], which allowed us to send HTTP/1.1 requests to the LexML portal and get back information about the federal legislation. Table 1 shows the list of legislation types extracted from the LexML portal and the corresponding quantity in which legislation appears for each type.

The database created using this approach is composed of information about legislation, such as type, name, date of approval, date of publication, and summary. Besides that, it contains information about relationships between elements of the legislation such as changes, citations, regulations, and repeals. This amount of collected information allowed us to create a *Legal Knowledge Graph* capable of representing the different elements of the legislation and the different relationships between them.

In Figure 1, we show a complete visualization of the Legal Knowledge Graph. This graph has a total of 214,407 nodes, that correspond to legislation presented in Table 1, and a total of 142,734 edges where 28,451 represent changes, 17,452 citations, 17,452 regulations and 93,905 repeals. We highlight that different legislation types (nodes) can interact through different types of relationships.

3.1. Employing Semantic Web Standards

Although KGs are not defined by the technology they adopt, Semantic Web standards have proven useful when used to represent knowledge in a graph structure, including provenance [20], named-entity disambiguation [21], and domain knowledge representation [1]. Another important advantage is making the content easier to use by downstream applications, allowing access to it through the use of standardized query languages [22] and vocabularies [23, 24].

Knowledge graphs often employ ontologies as a way to abstract graph content, supporting graph cohesion by providing standardized terminology for nodes and edges. Ontologies help during automatic KG construction (supporting link prediction) as well as to enable graph querying. We used the *European Legislation Identifier (ELI)*¹ ontology, which provides a set of fundamental legal concepts and is reasonably capable of being reused in many legal sources.

¹[https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012XG1026(01))

We also developed routines for the automatic generation of legal resource metadata in RDF from the legislation database presented in the previous section. Listing 1 shows an example containing Federal Decree-Law No. 2848 of 1940 and Law No. 14110 of 2020 that changes it.

```

1 <#LOF#14.110#2020-12-18> a eli:LegalResource ;
2   eli:date_publication "2020-12-18"^^xsd:date ;
3   eli:in_force "2020-12-21"^^xsd:date ;
4   eli:changes <#DLF#2.848#1940-12-7> ;
5   eli:description "Altera o art. 339 do Decreto-Lei n. 2.848, de 7 de
6     dezembro de 1940 (Codigo Penal), para dar nova redacao ao crime de
7     denunciao caluniosa" ;
8   eli:type_document "Lei Ordinaria Federal" ;
9   eli:number "14.110" .
10 <#DLF#2.848#1940-12-7> a eli:LegalResource ;
11   eli:date_publication "1940-12-7"^^xsd:date ;
12   eli:in_force "1940-12-7"^^xsd:date ;
13   eli:changed\_by <#LOF#14.110#2020-12-18> ;
14   eli:description "Codigo Penal" ;
15   eli:type_document "Decreto-Lei Federal" ;
16   eli:number "2.848" .

```

Listing 1: Example of Brazilian legal resources in RDF using the ELI ontology

The ELI ontology provides the high-level `LegalResource` class and the `type_document` predicate that allows the characterization of the legal resource according to a domain-specific ontology. Other predicates, including `changes`, `amends`, and `repeals`, help to build the graph by allowing the proper identification of connections between laws. We provide an online version of the Legal Knowledge Graph using this ontology at <https://github.com/hansidm/br-legal-kg>.

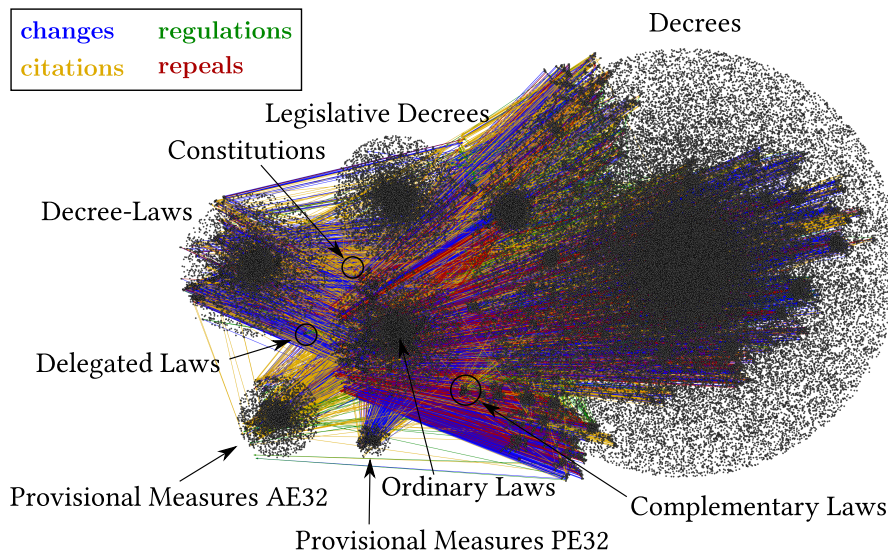


Figure 1: Graphical visualization of the Legal Knowledge Graph. The nodes correspond to the legislation presented in Table 1. The edges represent the different relationships between the legislation according to the colors presented in the upper right corner of the figure.

4. Complex Network Analysis

The approach described in the previous sections has the advantage of allowing the application of complex network techniques to explore Brazilian legislation. Here, we perform a direct simplification of the constructed KG by defining a network where nodes represent unique legislation and edges represent any type of relationship described in the previous section. Then, we use graph theory to extract topological information from the network that allows us to build a profile of the structure of Brazilian legislation. This process resulted in a *directed unweighted graph* with $N_v = 214,407$ nodes and $N_e = 141,568$ edges. It is important to note that the fraction $N_e/N_v \approx 0.660$ is too low, which is an indication of the graph's edge density. This suggests that the network is not a connected graph and is likely split into many small fragments. To explore this property, we perform a *connected components analysis* of the graph. By doing this, we find a total of 98,710 *weakly connected components*, where the largest of them has 110,466 (51.5%) nodes and 95,484 (44.5%) of them are unconnected nodes. The remaining 8,457 nodes are divided into the other 3,225 components. This indicates that a number of legal documents do not relate with which other or are independent.

To quantify the global importance of a given node, it is often convenient to look at some node-centrality measures. In this case, the simple and natural choice for measuring importance in our context is certainly *in-degree centrality* [25]. In Figure 2, we show the rank plot for the in-degree centrality of the nodes in the network described before. As we can see, this centrality is a good measure of importance, since the nodes with the highest in-degree are the Legal Codes of Brazilian Legislation. In the inset of this figure, we show that the in-degree distribution follows a power-law with exponent 2.65 which indicates a scale-free behavior, common in

Legislation	Code name	k_{in}
Decree-Law No. 5452 of 01/05/1943	Consolidation of Labor Laws	313
Law No. 5172 of 25/10/1966	National Tax Code	257
Decree-Law No. 2848 of 07/12/1940	Penal Code	128
Decree-Law No. 3689 of 03/10/1941	Penal Procedure Code	78
Law No. 5869 of 11/01/1973	Civil Procedure Code (1973)	78
Law No. 4737 of 15/07/1965	Electoral Code	73
Law No. 10406 of 10/01/2002	Civil Code	54
Law No. 9503 of 23/09/1997	Brazilian Traffic Code	50
Law No. 4117 of 27/08/1962	Brazilian Telecommunications Code	36
Law No. 8078 of 11/09/1990	Consumer Protection Code	31
Law No. 7565 of 19/12/1986	Brazilian Aeronautics Code	24
Decree-Law No. 227 of 28/02/1967	Mine Code	22
Law No. 12651 of 25/05/2012	Forest Code	18
Decree-Law No. 1002 of 21/10/1969	Military Penal Procedure Code	13
Decree-Law No. 1001 of 21/10/1969	Military Penal Code	9
Law No. 13105 of 16/03/2015	Civil Procedure Code (2015)	8
Decree No. 24643 of 10/07/1934	Water Code	3

Table 2
Relation of Brazilian Legal Codes.

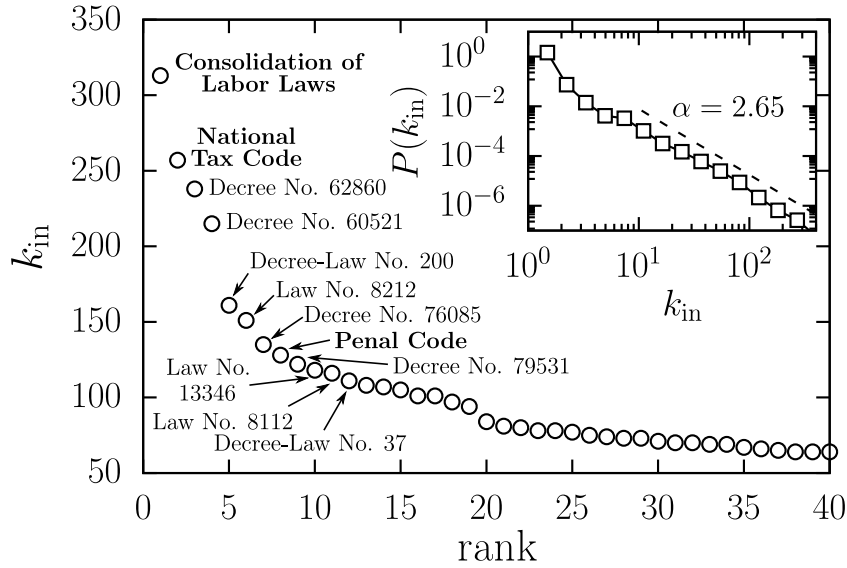


Figure 2: Rank plot for in-degree centrality for the simplification of the Legal Knowledge Graph. In the inset, the in-degree distribution shows a power-law behavior with an exponent of 2.65.

several phenomena [26, 27, 28, 13, 29].

In Table 2, we show some Brazilian legal codes² and the corresponding in-degree k^{in} of the nodes associated with them. In this table, the laws are sorted in descending order of k^{in} . As we can see, the highest in-degree node is associated with the “Consolidation of Labor Laws”, followed by the “National Tax Code” and the “Penal Code”. Surprisingly, there are some Legal Codes that, despite their importance, have a very low in-degree. Nonetheless, in-degree still is a good proxy for the relevance of a Legal Code in the same way that it is for usual nodes. This fact just indicates that having a high in-degree is sufficient but not a necessary condition for a node to be important.

4.1. Detecting Bridge-Like Nodes

Brazilian law practitioners usually have a qualitative notion of what specific laws are of interest, without specific metrics that would reveal them from a computational perspective. We have worked with domain practitioners to identify elements that would characterize a legal document as of interest. In doing so, we define the *bridge-like* nodes as the nodes that belong to the neighborhood of at least two nodes associated with Brazilian Legal Codes. We hypothesize that if a node is *bridge-like*, then this node represents a legal document that is of interest according to the domain practitioners.

Proceeding in that way, we detect 55 bridge-like nodes connected to 16 Brazilian Legal Codes shown in Table 2. There is only one code that has no bridge-like nodes connected to it, namely the “Water Code”. A convenient representation of the relationship between the bridge-like nodes

²The list of Brazilian Legal Codes used for this analysis can be found at: <http://www4.planalto.gov.br/legislacao/portal-legis/legislacao-1/codigos-1>.

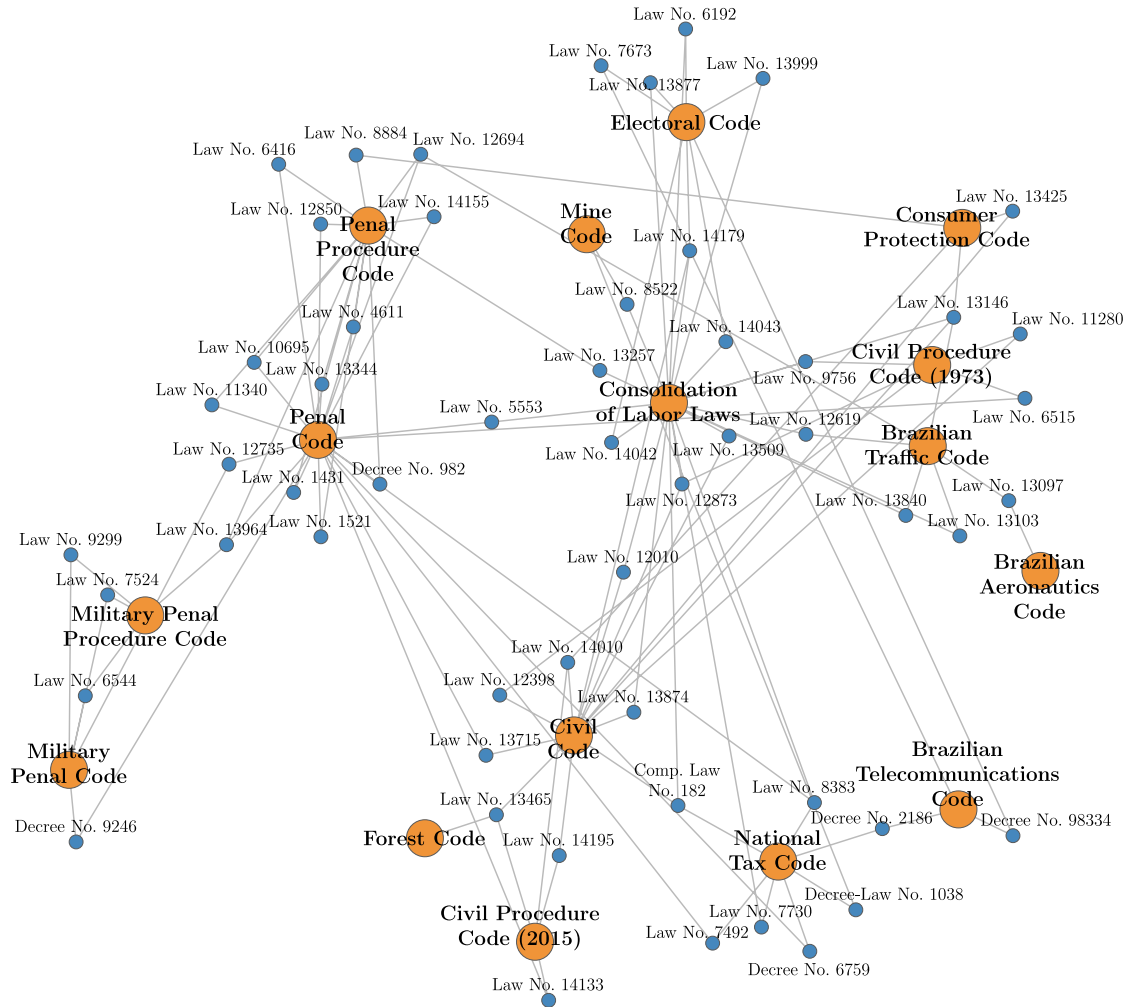


Figure 3: The subgraph defined by the Brazilian Legal Codes, the bridge-like nodes, and the edges between them. The orange circles represent the Brazilian Legal Codes shown in Table 2, the blue circles represent the bridge-like nodes and the gray lines represent the connections between different nodes.

and the Brazilian legal codes can be done by drawing the *subgraph* defined by the Brazilian Legal Codes, the bridge-like nodes, and the edges between them. In Figure 3, we show this subgraph. There, the Brazilian Legal Codes are represented by the big orange circles, and the bridge-like nodes are represented by the blue circles. The subgraph defined by Brazilian Legal Codes and bridge-like nodes has a total of 76 connected nodes and is a fraction of the largest component of the network.

In Table 3, we show the five bridge-like nodes with the highest k_{in} . The first bridge-like node corresponds to “Law No. 8383 of 30/12/1991”. This law changes the income tax legislation and establishes the Tax Reference Unit, a tax correction factor created in a period of quite high inflation in Brazil. The second node corresponds to “Law No. 8884 of 11/06/1994”. It provides for the prevention and repression of infractions against the economic order and transforms

Legislation	k_{in}
Law No. 8383 of 30/12/1991	45
Law No. 8884 of 11/06/1994	21
Law No. 13146 of 06/07/2015	18
Law No. 11340 of 07/08/2006	16
Law No. 7730 of 31/01/1989	15

Table 3
Top five Bridge-Like nodes by k_{in} .

the Administrative Council for Economic Defense (Cade) into an Autarchy, which is a type of indirect public administration entity with administrative and financial autonomy. The third node corresponds to “Law No. 13146 of 06/07/2015”. It establishes the Brazilian Inclusion Law of Persons with Disabilities. An important Brazilian law that later became known as the “Statute of Persons with Disabilities”. These three examples show that, despite the bridge-like nodes having a low in-degree, they are of great importance to Brazilian Legislation.

5. Conclusion

We presented an approach for constructing a Knowledge Graph for Brazilian Legislation. Legal data from *LexML portal* has been extracted and integrated into a database that serves as input to a method that represents this data as linked data in RDF using Semantic Web standards. Then, we used graph theory to extract information from the Knowledge Graph that allowed us to make characterizations about the structure of Brazilian legislation, essentially building a graph profile. We found indications of low edge density in the graph, which was confirmed by performing a connected component analysis. We also calculated the *in-degree centrality* for the network nodes and observed that it is a good global-importance measure of a given node in our network since the nodes with the highest degree coincide with the Legal Codes of Brazilian Legislation.

This approach, also, allowed us to detect “bridge-like nodes” (nodes that belong to the neighborhood of at least two nodes associated with Brazilian Legal Codes). These nodes have been shown to play an important role in the structure of the network since they connect different important parts of Brazilian Legislation. We believe that a careful analysis of these bridge-like nodes would allow a more detailed understanding of the inner structure of the network. It could also bring insights into the processes that lead to the formation of the Legal Knowledge Graph. The information generated by the bridge-like nodes approach can be very useful for legal scholars to understand the real impact of laws as it promotes a new way to identify to what legal dimensions one particular law establishes connections.

For future work, in terms of graph applications, we intend to carry out a deeper analysis of the Knowledge Graph in order to better understand its underlying topological structure and characterize temporal correlations between laws that can help us to understand the evolution of Brazilian Legislation. In terms of Semantic Web Standards, we intend to refine the legal modeling in the graph by exploring the formal definition of the extensions needed for further

characterization of legal data in the form of an ontology. We also intend to explore the inner structure of the legal documents which could shed light on the relations between parts of documents. As LexML is of a high level, we can refine the representation by providing contextualized terminology for the Brazilian scenario.

5.0.1. Acknowledgments.

We gratefully acknowledge CNPq, CAPES, FUNCAP, and the Edson Queiroz Foundation for financial support.

References

- [1] B. Abu-Salih, Domain-specific knowledge graphs: A survey, *Journal of Network and Computer Applications* 185 (2021) 103076.
- [2] J. Breuker, P. Casanovas, M. C. A. Klein, E. Francesconi, The Flood, the Channels and the Dykes: Managing Legal Information in a Globalized and Digital World, *Law, Ontologies and the Semantic Web* (2009) 3–18.
- [3] E. Filtz, S. Kirrane, A. Polleres, The linked legal data landscape: linking legal data across different countries, *Artificial Intelligence and Law* (2021).
- [4] L. F. Salomão, Artificial Intelligence: Technology Applied to Conflict Resolution in the Brazilian Judiciary, Research report, 2021. URL: https://ciapj.fgv.br/sites/ciapj.fgv.br/files/report_ai_ciapj.pdf.
- [5] Conselho Nacional de Justiça, GovRisk The International Governance & Risk Institute, UK-Brazil Cooperation: improving efficiency and performance in Brazil’s Judiciary, 2016/17 (2017).
- [6] LexML Brasil: Rede de Informação Legislativa e Jurídica, <https://www.lexml.gov.br>, ????
- [7] C. Biagioli, E. Francesconi, A. Passerini, S. Montemagni, C. Soria, Automatic semantics extraction in law documents, in: *Proceedings of the 10th international conference on Artificial intelligence and law*, 2005, pp. 133–140.
- [8] A. Boer, R. Hoekstra, R. Winkels, T. v. Engers, F. Willaert, Proposal for a dutch legal xml standard, in: *International Conference on Electronic Government*, Springer, 2002, pp. 142–149.
- [9] M. P. Ebenhoch, Legal knowledge representation using the resource description framework (rdf), in: *12th International Workshop on Database and Expert Systems Applications*, IEEE, 2001, pp. 369–373.
- [10] A. Gangemi, Design patterns for legal ontology construction, *Design Patterns for Legal Ontology Construction* (2007) 1000–1021.
- [11] E. Filtz, Building and processing a knowledge-graph for legal data, in: *European Semantic Web Conference*, Springer, 2017, pp. 184–194.
- [12] F. Amato, A. Mazzeo, A. Penta, A. Picariello, Building rdf ontologies from semi-structured legal documents, in: *2008 International Conference on Complex, Intelligent and Software Intensive Systems*, IEEE, 2008, pp. 997–1002.
- [13] J. L. B. de Araújo, J. A. M. Neto, F. Siqueira, C. M. Santos, R. Vasconcelos, E. A. Oliveira,

- C. Caminha, V. Furtado, Characteristics of effective paths in brazilian legal processes (2021).
- [14] G. Lame, Using nlp techniques to identify legal ontology components: concepts and relations, in: *Law and the Semantic Web*, Springer, 2005, pp. 169–184.
- [15] J. M. Schneider, G. Rehm, E. Montiel-Ponsoda, V. Rodríguez-Doncel, P. Martín-Chozas, M. Navas-Loro, M. Kaltenböck, A. Revenko, S. Karampatakis, C. Sageder, et al., Lynx: A knowledge-based ai service platform for content processing, enrichment and analysis for the legal domain, *Information Systems* 106 (2022) 101966.
- [16] J. Saias, P. Quaresma, Semantic enrichment of a web legal information retrieval system, in: *JURIX*, 2002, pp. 11–20.
- [17] URIs, URLs, and URNs: Clarifications and Recommendations 1.0, W3C Note, W3C, 2001. <https://www.w3.org/TR/2001/NOTE-uri-clarification-20010921/>.
- [18] Rede de Informação Legislativa e Jurídica, <https://projeto.lexml.gov.br/documentacao/Parte-2-LexML-URN.pdf>, ????
- [19] Requests: HTTP for Humans, <https://docs.python-requests.org/en/latest/>, ????
- [20] L. F. Sikos, D. Philp, Provenance-Aware Knowledge Representation: A Survey of Data Models and Contextualized Knowledge Graphs, *Data Science and Engineering* 5 (2020) 293–316.
- [21] I. O. Mulang', K. Singh, C. Prabhu, A. Nadgeri, J. Hoffart, J. Lehmann, Evaluating the Impact of Knowledge Graph Context on Entity Disambiguation Models, in: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20*, Association for Computing Machinery, New York, NY, USA, 2020, pp. 2157–2160.
- [22] SPARQL 1.1 Overview, W3C Recommendation, W3C, 2013. <https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>.
- [23] R. Guha, D. Brickley, RDF Schema 1.1, W3C Recommendation, W3C, 2014. <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>.
- [24] OWL 2 Web Ontology Language Document Overview (Second Edition), Technical Report, W3C, 2012. <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/>.
- [25] M. Newman, *Networks: An Introduction*, Oxford University Press, London, 2010.
- [26] C. Caminha, V. Furtado, T. H. Pequeno, C. Ponte, H. P. Melo, E. A. Oliveira, J. S. Andrade Jr, Human mobility in large cities as a proxy for crime, *PloS one* 12 (2017) e0171609.
- [27] C. Caminha, V. Furtado, V. Pinheiro, C. Ponte, Graph mining for the detection of overcrowding and waste of resources in public transport, *Journal of Internet Services and Applications* 9 (2018) 1–11.
- [28] C. Ponte, H. P. M. Melo, C. Caminha, J. S. Andrade Jr, V. Furtado, Traveling heterogeneity in public transportation, *EPJ Data Science* 7 (2018) 1–10.
- [29] R. Albert, Scale-free networks in cell biology, *Journal of cell science* 118 (2005) 4947–4957.