Introducing Intelligence to the Semantic Analysis of Canadian Maritime Case Law: Case Based Reasoning Approach

Bola Abimbola[1] [0000-0001-9554-8551] Qing Tan[2] [0000-0002-6447-52133], José R. Villar[3] [0000-0001-6024-9527]

Abstract: The use of machine learning and semantic analysis in case law is the new trend in modern society. Case Based Reasoning tools are being used to analyze texts in courts to make and predict judicial decisions which are designed to base the outcomes of current court proceedings from past and or learning from the mistakes to make better decisions. Because of the accuracy and speed of this technology, researchers in the justice system have introduced Machine Learning to optimize the Case-Based Researching approach. This paper presents a study aimed to critically analyze semantic analysis in the context of machine learning and proposes a case-based reasoning information retrieval system. It will explore how CBR-IR is being used to improve legal case law information retrieval. The study covers the importance of semantic analysis. The study will discuss limitations and recommendations for improvement and future research. The study recommends that it is necessary to conduct further research in semantic analysis and how they can be used to improve information retrieval of Canadian maritime case law.

[1]Bolanle Abimbola, University of Oviedo, U0285018@uniovi.es
[2] Qing Tan, Athabasca University, qingt@athabascau.ca
[3] José R. Villar, University of Oviedo, villarjose@uniovi.es

1 Introduction

Intelligence Information retrieval refers to the process of accessing case documents that are related to a given case. The process is important as it helps in linking related information and thus ensuring ideal solutions are met in every case. There are several approaches that are used in retrieving these data. For instance, Information extraction which uses natural language processing, machine learning techniques, rule-based, and case-based approach. However, to deliver the much-needed value, it is critical to ensure that all case and court processes are efficient, effective, and equitable. These approaches are used to retrieve the previous case with respect to the current case. Machine learning, semantic learning, and computers favor state legislatures and all other courts in the land by allowing all the stakeholders to design and develop equitable, effective, and efficient legal systems to serve justice for all [1]. Therefore, the purpose of this study is to identify and evaluate how machine learning and semantic learning integration into the legal domain is done by focusing on the analysis of legal texts and their impact on the legal system.

2 *Problem Statement*

One of the major problems for courts in the modern environment is the inability to optimally analyze quantitative legal data. Inadequate data acquisitions by courts today are a major problem that hinders decision making and litigation of cases. With limited information, courts are unable to have consistency between inputs and outputs. Today, courts are unable to normalize inputs and outputs, implying that they end up making uninformed decisions. Inconsistencies between inputs and outputs can be linked to the approaches that are used to gather information. Legal institutions rely on existing data or newly generated data to make predictions. Therefore, it is imperative for courts to ensure that these sources are reliable and gathered from people who are affected and who will benefit from the designed predictions. The presence of different types of data can pose a challenge in data analysis and generation of reliable predictions. The vast amount of data requires courts to constantly rely on similarities and differences in legal data, which can significantly aid generate optimal predictions. Because modern courts operate in a highly dynamic and competitive environment, they have to address the needs of all populations, which prompts them to adopt approaches that can help improve customization [11]. Also, these changes lead to changes in ML algorithms and thus, establishment of input and output inconsistencies.

Legal reasoning (LR) is a key component in legal practice, LR are methods that lawyers use to apply laws to facts to answer legal questions, it is mainly based on data gathered from previous cases. It is safe to imply that the verdict of cases can be made easier or difficult depending on the available legal data [20]. In the research [20], the authors argue that the effectiveness and efficiency of legal processing depends on how data is stored and retrieved. In this context, analysis of legal cases has been a challenge in the past because of the inability to store and retrieve data as needed. In the past, courts relied on hard copies and manual retravel of data, which was time consuming, predictable, and exposed to errors.

In most countries today, legal data is retrieved manually or by the use of syntax [12]. Using these methods to analyze cases have numerous problems, which inhibits the abilities of judges to analyze and identify cases. The use of manual methods and syntax methods are associated with generation of inadequate and less useful information. The problem of effective and efficient data retrieval and storage is significant in the legal domain because it is primarily on information. Besides, information gathered in the legal domain is important because it defines the survivability of society. Because of the inefficiency and ineffectiveness, the justice system is unable to provide justice where it is deserved. The soundness of judicial verdict depends on the quality of information gathered by legal experts. Another dimension of this problem is that judicial verdicts are based on the information gathered from past cases. As time passes, the volume of the information in past cases increases, making it a challenge for legal experts to analyze the data.

3 Related work

The current literature shows that machine learning, information technology, and semantic learning have been used in the court system to enhance the quality of services provided [9]. For example, machine learning and semantic learning allow text creation, storage, and retrieval, which have become more accessible and an essential part of the legal process [2]. Aside from the hearing capacity, judges need to create composed rulings, judgment, and purposes behind the choices they ceaselessly make [10]. After the typewriter era, judges were forced to write decisions in longhand, and secretaries would then type the same out in typescript [8].

Additionally, machine learning and semantic learning have improved access to the law. In most countries and jurisdictions, the applicable law is found in different sources, including statutes, law reports, and customary laws [3]. In the past legal documents were kept as paper copies stored in filing cabinets and folders, but now it is possible access digital copies over the internet, it is easy to access and use the existing information to make critical decisions.

The use of ML to aid in semantic analysis of texts, phrases, and language of law data has helped in quantitative analysis of legal data [17]. The progress in artificial intelligence, language processing, and machine learning are linked to rapid evolution in algorithms and data-based practices [13]. An evolution in algorithms has helped courts to make informed predictions through provision of important insights and knowledge about [14]. Because of the need to offer accessibility to court cases and decisions, and the need to promote equality, machine learning is used to find similarities and differences in litigation judgements, which improves court outcomes [15].

4 Machine learning and sentimental analysis

Semantic is a branch of linguistics that focuses on exploring the meaning of sentences and words. Understanding words and sentences help understand language. Semantic analysis helps in the comprehension of forms and how they interact with each other. In machine learning, semantic analysis is used to determine the significance of syntax in a program. It is used to verify whether software declarations are correct. It is used to determine if a code in a program is accurate.

Machine learning and semantic analysis are becoming an essential part of the modern court system. One of the easier parts of electronic court innovation is utilizing an advanced camera or scanner to take exhibit and show it on the screen [4]. This underlines how machine learning and semantic learning have revolutionized the modern legal and court systems [5]. The use of advanced cameras and scanners in the legal system shows that machine learning and semantic analysis to retrieve Canadian maritime law cases are improving and changing the Canadian legal and court system. However, future studies should focus on the challenges of adequate application and use of machine learning, information technology, and semantic learning in the court system as the court system moves towards efficiency and effectiveness.

According to [18], employment of semantic retrieval technology will enhance information quality retrieval. The work suggests a technological framework that will help users retrieve information by providing them with relevant case-based and semantic retrieval documents. Users will key in their query terms in their normal language and the system conduct analysis on them. [19] argues that comparing ideological query representation against the conceptual representations database aid in choosing the match within a close range. [19] further contends that users can use natural language or a relevant document or Boolean query to search-related documents.

The suggested sets of concepts for geographical-related information retrieval included a mixture of quantitative and qualitative geometric data, including sparse coordinate and topological relations information representing the geographical place's footprints. Geographical categories classified places and then connected to non-geographical cases grouped by conceptual ranking. According to [20], the goal was to combine Euclidean and hierarchical distance analysis to establish a geographical distance evaluation. Another proposal was on query advancement approach that employed advanced dependent topologies to bring queries nearer to document collection characteristics and user's preferences [20]. It aimed at linking each concepts' case and class with a feature vector to modify these ideas to the terminology and document collection utilized. The concepts and their related feature vectors processed results after identifying user's query on the search engine [20].

According to [21], concepts established in the Spirit web project were employed to support document retrieval that was geographically suitable to users' request, where query advancement used domain and spatial concepts. A concept network deduced from concepts from the original query words served as a knowledge base for modifying query advancement [20]. The conceptual query advancement quality relied on the quality of concept network. The purpose of a concept network was to match original query words which resulted in the development of other concepts and queries terms.

While most concepts categories used the WordNet as a controlled vocabulary to advance the query, the approach suggested by [20] combines the advantages of statistical methods and concept use. For query advancement, field concepts were utilized as controlled terminology. The primary presupposition is that users create a query concurrently illustrating an issue they are trying to address. The CBR approach uses other related cases to address an information retrieval request [22]. A solution is acquired by providing several links related to the user's query. A case-based approach contains a set of data that defines and provide information about other data. The model by [20] offers various intelligent query advancement and processing benefits.

5 Case-based Reasoning Information Retrieval System (CBR-IR)

Traditional IR systems recovered information without defining any user's specific field of interest. As such, the system provided massive data that was not important to the user. [23] demonstrates how to employ concepts effectively included in various multi-disciplinary fields comprising of different terminologies aiming goal to improve the browsing outcome quality for extensive search systems.

One strength of Case-Based Reasoning (CBR) systems is the ability to reason about a problem case and perform highly intelligent problem-solving, such as the generation of legal arguments or detailed operational plans [26].

Query searching using concepts is a promising and unique approach in the retrieval process. Users do not need to know the documents implementation; their focus is on the conceptual searching level. Domain concept is helpful for query advancement by increasing the input terms with the appropriate domain ontologies. WordNet adds meronyms, homonyms, and synonyms to the index terms, making the indexing stage effective [28]. There exist two problems in utilizing the concept–based approach. The first problem is using keywords to extract semantic concepts. Its main problem is determining relevant concepts that identify documents and determine the language used in user queries. It is essential to avoid matching and connecting inappropriate concepts and disregard appropriate concepts. The second problem is document indexing. Field ontology is established, and property and concept relationship in the professional field is described.

Nevertheless, query expansion contains some built-in dangers. Thesaurus has been used in information retrieval to identify the linguistic entities and synonymous expressions semantically the same. A query drift can occur due to query ambiguity providing information that is irrelevant to the user. For instance, the term windows could mean Microsoft Windows OS (operating system) or the actual house windows. The system should employ domain concepts to solve the problem. Not every tokenized term should be set for expansion. Query expansion process should replace the terms in the domain concepts with the original user terms and their related domain concepts.

A concept-based method utilizes concepts from a specific domain and CBR approach with various metadata containing relevant documents defines a case. A case bases act as a piece of document information to examine the query process and retrieve information from appropriate documents in the digital library [24]. It aimed at improving concept-based information retrieval by integrating

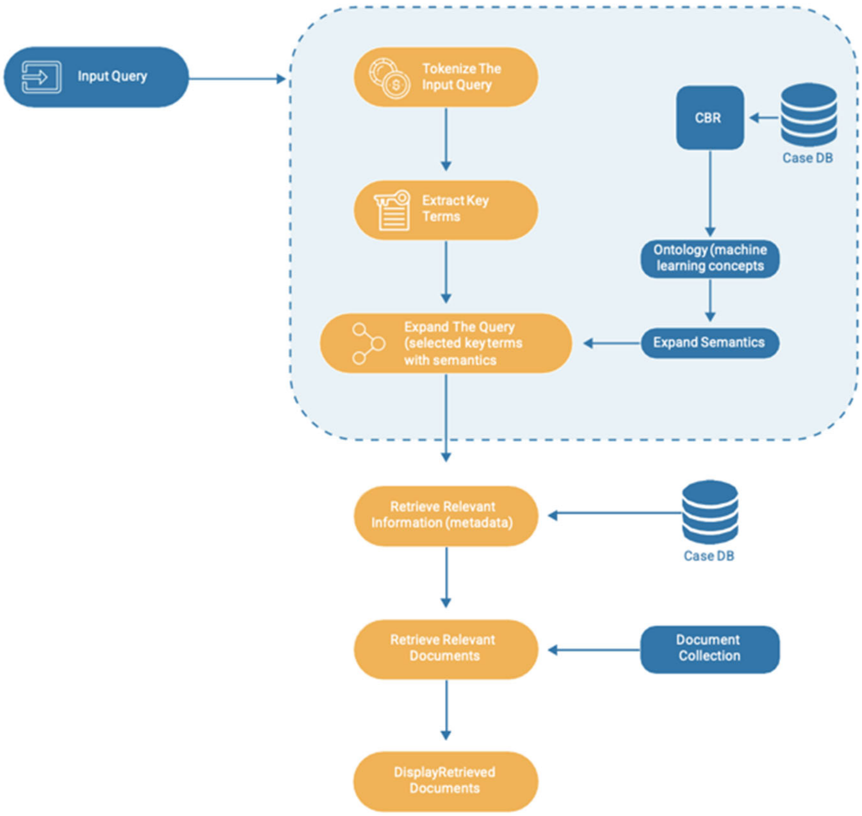

Fig. 1: The conceptual system diagram of the CBR-IR System

domain ontology, case-based reasoning process, and traditional information retrieval process [24]. The model proposed by [24] employs concepts to expand queries and integrates textual and case-based closeness to recover a set of information for relevant documents to give users several document recommendations options. The steps in figure 1 are as follows:

Step one: matching a new case using other cases in the case dB – The CBR-IR system performs a CBR analysis of inputs. It then uses the analyzed results to generate a text-based document retrieval system. [28]

Step two: recover the closest matching case from the past cases' library - Defines categories and classes for a ranking system in this step. This is used to generate a standard query of these texts' top N terms or top D pairs of terms. The CBR-IR module checks large number of cases in the case dB. Full texts of the court's opinions and cases selected from the CBR module's Case dB. CBR system used a similarity measure that is based on a generalized weighted distance metric.

Step three: reuse the recovered case to address the existing issue - the system sorts the retrieved documents into an initial ordering of cases relevant to the problem case. The categorization is done on the bases of "on-point". The model of relevance and on-pointiness used are then used in CBR-IR style system settings. The cases selected are passed through the preprocessing steps, it is presented to the "on-point" classifier to retrieve similar previous cases from the CBR-IR. [28]

Step four: re-evaluate and modify the suggested solution if needed - The on-pointiness model sorts the unique cases. The sorting generates a partial ordering in which Case A is more "on-point" than Case B (if the set of applicable dimensions A shares with the problem case are more than those of case B and the problem case. Maximal cases in this ordering are called most on-point cases. The result of sorting these instances is virtualized in a "claim lattice." (See Figure 2 for an example.)

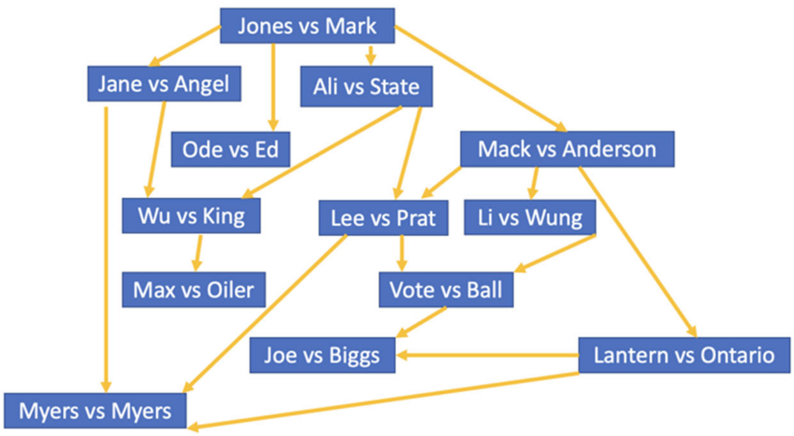

Fig. 2 Claim lattice for Jones vs Mark

Step five: rename the final result as a new case - These CBR analysis results are represented in a "claim lattice" that uses a hybrid CBR-IR system to select unique cases from the claim lattice. For instance, the most on-point cases and maximal cases in the on-point ordering. The texts associated with the unique cases are passed to a modified version of the relevance feedback (RF) mechanism of the “INQUERY IR” system [27],

Apache Lucene provides all the relevant interfaces for query, index engines and analyses the overall text format content, such as txt, pdf, docx, and web page data. The functions can be easily embedded into various applications to implement full text retrieval functions. Comparing Lucene to other text retrieval system, it provides faster speed and better performance repository and data retrieval without degrading the system performance and provides steady indexing across database in data center and files in various formats. [30] Lucene core data is encapsulated into classes by defining a file index format, and then program processes to form a complete information search engine. Lucene's has three main parts: External Interface, Index Core and Infrastructure Package. Index Core is the direct focus when manipulating file indexes [29].

Case similarity evaluation is done for the subject and author attributes [25]. Statistical IR methods used the Apache Lucene to measure Title attribute content. In Lucene, the Boolean technique and vector space model were used to determine the relevancy of a particular document to the user’s request.

6 Conclusion

Semantic analysis is an important area of study today because courts today strive to achieve equality and optimal decision making by adopting technologies. The ability to understand information that courts gather is vital because it helps make effective predictions and improve performance. Semantic analysis can be used to address technical challenges that are faced in machine learning. Semantic analysis can be used to address problems associated with limited data, presence of different data types, and inconsistencies in AI predictions and real-life solutions.

The research has demonstrated that the already developed domain-specific ontology with Machine Learning can be efficient for query advancement. Many researchers have used semantic retrieval technology using concepts. It helps solve problems that lack semantics in traditional retrieval technology. Using concepts in ontology enhances search results. Expanding a query aims at minimizing document or query mismatch by adding related phrases and terms to the relevant documents set.

This research has introduced Machine Learning to the semantic analysis module in the proposed Case-Base Reasoning Information Retrieval (CBR-IR) system. The future research will further study and develop the CBR-IR by implementing the machine learning techniques with semantic analysis in the information retrieval system.

7 Acknowledgement

This research has been funded by the SUDOE Interreg Program -grant INUNDATIO-, by the Spanish Ministry of Economics and Industry, grant PID2020-112726RB-I00, by the Spanish Research Agency (AEI, Spain) under grant agreement RED2018-102312-T (IA-Biomed), and by the Ministry of Science and Innovation under CERVERA Excellence Network project CER-20211003 (IBERUS) and Missions Science and Innovation project MIG-20211008 (INMERBOT). Also, by Principado de Asturias, grant SV-PA-21-AYUD/2021/50994.

8 References

1. M. Venkateswarlu Naik, "Building a Legal Expert System for Legal Reasoning in Specific Domain-A Survey," International Journal of Computer Science and Machine learning , vol. 4, no. 5, pp. 175–184, Oct. 2012.
2. T. L. Mosweu and O. Mosweu, "Electronic Court Records Management Systems: A Review of Literature in Selected African Countries," Mousaion: South African Journal of Information Studies, vol. 36, no. 4, Oct. 2019.
3. S. J. Pratiwi, S. Steven, and A. D. P. Permatasari, "The Application of e-Court as an Effort to Modernize the Justice Administration in Indonesia: Challenges & Problems," Indonesian Journal of Advocacy and Legal Services, vol. 2, no. 1, pp. 39–56, Mar. 2020.
4. N. Imamnazarovna, "Aspects Of Legal Regulation Of Electronic Document And Electronic Document Circulation In Business," The American Journal of Political Science Law and Criminology, vol. 02, no. 11, pp. 8–14, Nov. 2020.
5. M. Procopiuck, "Machine learning and time of judgment in specialized courts: What is the impact of changing from physical to electronic processing?," Government Information Quarterly, vol. 35, no. 3, pp. 491–501, Sep. 2018.
6. H. M. Levitt, Z. Morrill, K. M. Collins, and J. L. Rizo, "The methodological integrity of critical qualitative research: Principles to support design and research review.," Journal of Counseling Psychology, vol. 68, no. 3, pp. 357–370, Apr. 2021.
7. M. Epp and C. C. Otnes, "High-Quality Qualitative Research: Getting Into Gear," Journal of Service Research, p. 109467052096144, Sep. 2020.
8. P. E. Mészáros, "The evolution of electronic administration and its practice in judicial proceedings," Pravni vjesnik, vol. 34, no. 3–4, Dec. 2018.
9. Mohammad Ikbal Hasan and Badsha Mia, "Initiation of Virtual Court System during COVID-19 Pandemic and E-Judiciary: Challenges and Way Forward," Daengku: Journal of Humanities and Social Sciences Innovation, vol. 1, no. 1, pp. 8–17, Mar. 2021.
10. L. Gupta and S. Gadiwala, "Coping with the Coronavirus Disease-2019 pandemic: A giant leap towards digital transformation in academic research," Indian Journal of Rheumatology, vol. 0, no. 0, p. 0, 2021.

11. K. Tae & S. So. Machine-learning-based deep semantic analysis approach for forecasting new technology convergence. Technological Forecasting and Social Change. 157. 120095. (2020).
12. M. Medvedeva,. M, Vols, & M. Wieling. (2019). Using machine learning to make decisions of the European Court of Human Rights. Artificial Intelligence and Law. (2019).
13. X. Huang. C, Zanni-Merk., & B. Cremilleux. Enhancing Deep Learning with Semantics: an application to manufacturing time series analysis. Procedia Computer Science, 159. (2019). 437-446.
14. H. Sarker. (2021) Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160. (2021).
15. E. P., Ebietomere, & G. O., Ekuobese. A Semantic Retrieval System for Case Law. Applied Computer Systems, 24(1), (2019). pp. 38-48.
16. J. Cao., M. Wang., Y. Li., & Q Zhang. Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment. PLoS ONE 14(4): e0215136. (2019).
17. G. Babacar., Z. Dezheng., & W. Aziguli. (2021). "Improvement of Support Vector Machine Algorithm in Big Data Background", Mathematical Problems in Engineering 2021, https://doi.org/10.1155/2021/5594899.
18. J. Baoxian, "Application of Intelligent Information Retrieval for Big Data Oriented Brain Science," Advances in Engineering Research, vol. 66, 2018.
19. D. Joby, "Expedient Information retrieval system for web pages using the natural language modeling," June 2020, vol. 2, no. 2, pp. 100–110, 2020.
20. L. Afuan, A. Ashari, and Y. Suyanto, "A study: query expansion methods in information retrieval," J. Phys. Conf. Ser., vol. 1367, p. 012001, 2019.
21. R. Tehseen, "Semantic Information retrieval: A survey," J. Inf. Technol. Softw. Eng., vol. 08, no. 04, 2018.
22. K.-S. Lin, "A case-based reasoning system for interior design using a new cosine similarity retrieval algorithm," J. Inf. Telecommun., vol. 4, no. 1, pp. 91–104, 2020.
23. J. On-Piu Chan, "Digital transformation in the era of big data and cloud computing," Int. J. Intell. Inf. Syst., vol. 9, no. 3, p. 16, 2020.
24. Anaissi, M. Goyal, D. R. Catchpoole, A. Braytee, and P. J. Kennedy, "Case-based retrieval framework for gene expression data," Cancer Inform., vol. 14, pp. 21–31, 2017.
25. M. U. Devi and G. M. Gandhi, "Scalable information retrieval system in semantic web by query expansion and ontological-based LSA ranking similarity measurement," Int. J. Adv. Intell. Paradig., vol. 17, no. 1/2, p. 44, 2020.
26. Janet L. Kolodner. Case-Based Reasoning. Morgan Kaufmann, 1993.
27. Jody J. Daniels and Edwina L. Rissland "A Case-Based Approach to Intelligent Information Retrieval," SIGIR'95, Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, Washington, USA, July 9-13, 1995 (Special Issue of the SIGIR Forum).
28. Hélène Feuillâtre ,Vincent Auffret, Miguel Castro, Florent Lalys, Hervé Le Breton, Mireille Garreau and Pascal Haigron, "Similarity measures and attribute selection for case-based reasoning in transcatheter aortic valve implantation,". PLOS ONE September 3, 2020. https://doi.org/10.1371/journal.pone.0238463
29. Wu Jian. Design and implementation of campus network search engine based on Lucene [J]. Journal of Hunan Institute of engineering, 2012,22
30. Zheng Youzhuo, Fu Yu; Zhang Ruifeng; Hao Shuqing; Wen Yi. Research on Lucene Based Full-Text Query Search Service for Smart Distribution System