

Augmented Reading and Similar Case Matching: from Legal Domain Experts' Modus Operandi to a Computational Pipeline

Rachele Mignone^{1,3,4}, Ivan Spada^{1,3,4}, Chiara Bonfanti¹, Michele Colombino¹, Giorgia Iacobellis¹, Laurentiu Jr Marius Zaharia¹, Marianna Molinari², Susanna Marta², Ilaria Angela Amantea¹, Emilio Sulis¹, Luigi Di Caro¹ and Guido Boella¹

¹Computer Science Department - University of Turin, Via Pessinetto 12, 10149, Torino, Italy

²Law Department - University of Turin, Lungo Dora Siena 100/A, 10154, Torino, Italy

Abstract

The increasing backlog in processing legal cases in the Courts pushes toward the implementation of solutions that can reduce the workload and thus the time to complete cases. Searching for similar judgments is a relevant task that enables consultation of related legal proceedings. This paper proposes to automate the search process by developing a pipeline inspired by the *modus operandi* of legal domain experts. The pipeline includes an augmented reading of judgments for the purpose of semantically analyzing texts taking into account the context given by their classification, decisions and citations. The results are interpretable by legal domain experts, while the similarity case-matching output is underpinned by the information extracted from the relevant documents. The paper addresses a case study based on Italian national laws.

Keywords

Augmented reading, Similar Case Matching, Explainability, Information extraction, Legal case study

1. Introduction

The length of trials and the disposal of the backlog are two of the main challenges of many juridical systems. This is mainly due to the disproportion between the number of cases and the number of judges and resources assigned to juridical offices.

³Corresponding author.

⁴These authors contributed equally.

Knowledge Management and process mining for Law (KM4Law), 9th Joint Ontology Workshops (JOWO 2023), co-located with FOIS 2023, 19-20 July, 2023, Sherbrooke, Québec, Canada

✉ rachele.mignone@unito.it (R. Mignone); ivan.spada@unito.it (I. Spada); chiara.bonfanti@edu.unito.it (C. Bonfanti); michele.colombino@edu.unito.it (M. Colombino); giorgia.iacobellis@edu.unito.it (G. Iacobellis); laurentiu.zaharia@edu.unito.it (L. J. M. Zaharia); marianna.molinari@unito.it (M. Molinari); susanna.marta@unito.it (S. Marta); ilariaangela.amantea@unito.it (I. A. Amantea); emilio.sulis@unito.it (E. Sulis); luigi.dicaro@unito.it (L. D. Caro); guido.boella@unito.it (G. Boella)

ORCID 0009-0009-2699-8730 (R. Mignone); 0009-0002-0459-1189 (I. Spada); 0009-0007-8015-7786 (C. Bonfanti); 0009-0007-3248-1661 (M. Colombino); 0009-0003-1730-7711 (G. Iacobellis); 0009-0002-3559-8367 (L. J. M. Zaharia); 0009-0003-1832-8135 (M. Molinari); 0009-0003-4014-261X (S. Marta); 0000-0003-1329-1858 (I. A. Amantea); 0000-0003-1746-3733 (E. Sulis); 0000-0002-7570-637X (L. D. Caro); 0000-0001-8804-3379 (G. Boella)



© 2021 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 order to lighten the workload of the magistrate, our proposal is to shorten reading and researching times by providing a methodology to facilitate the reading of legal documents by efficiently extracting and exploiting the knowledge embedded in the judgments.

Nevertheless, the full exploitation of juridical documents via efficient automatic information extraction allows us to perform Similar Case Matching and further speed up the magistrates' research process [1] as we prove by investigating a case study based on Italian legal judgments.

This work describes a pipeline for the augmented reading of legal documents which results in a tool for the detection of the judgment's relevant content, in order to facilitate the use of juridical databases, extracting information that goes far beyond the document's metadata and keywords, allowing for a more in-depth understanding of the document and for advanced functionalities such as pairwise similarity comparison of juridical documents.

After a brief legal (Section 2) and technical (Section 3) contextualization, we present the specific case study by thoroughly analyzing the domain experts' *modus operandi* when reading and comparing legal documents (Section 4). Subsequently, we propose a pipeline for the augmented reading of legal judgments (Section 5) followed by its evaluation in terms of time efficiency and Similar Case Matching (Section 6).

2. Legal Background

The idea of creating a computerized database of case law had been around since the 1960s [2]. Its implementation has gradually been refined in parallel with the development of information technology. The creation of such a system involved the goal of combining information technology and law to facilitate the knowledge and use of the latter. Nowadays a database of a legal management system can be functional to the objectives of greater productivity, the raising of quality, transparency, and dialogue with the territory.

In particular, the case-law databases, from a means of knowledge and information, are a tool to organize judicial offices and implement their efficiency. The concept embraces the traditional collections of judgments, which are used as research and support tools; but it can also be referred to those statistical surveys aimed at recognizing the flows of litigation, e.g. the analysis of controversies, having a given object (and sub-object), defined by a specific office; the outcome of the judgment; to what extent appeals are successful or rejected (in relation to the matter of the proceedings).

Definitions We introduce some definitions for the understanding of the technical details contained in the article.

- **Legal judgment:** authoritative decision given by a court or tribunal, including a decree, order, decision, or writ of execution.
- **Judicial grade/ Instance:** reference degree of the issuing judicial office (Court of First Instance or Court of Second Instance - Appeal).
- **Section (i.e. Sezione):** a distinct group of Judges, specialized in a given area of law (Labour Law, Family Law, Criminal Law, etc), within a larger body of them, which constitutes the Judicial Office taken into account.

- **Legal institute:** set of rules dealing with a particular legal relationship, or facts.
- **Topic:** the matter of a legal institute or of a section of a legal institute.
- **Citations:** normative and case-law references.
- **Normative:** Rules, especially rules of behavior.
- **Case law:** Judicial decisions resolved by courts using the concrete facts of a case.

3. Related Work

Reading and comparing textual documents are cross-domain tasks. The task of Similar Case Matching can be tackled with different technologies. Many early contributions focus on a text-based similarity measure (i.e. TF-IDF-based similarity measures [3]) which, alone, fail to convey the semantic meaning of a legal text. Other techniques rely on the extraction of citations within the documents to compute a similarity score. In particular, these citation-based approaches can be further divided into two sub-categories: mere citation analysis [4] and citation-graph analysis [5] which allows the use of graph-based deep learning techniques such as GNNs [6]. Nevertheless, text-based and graph-based techniques have been combined, achieving sub-optimal performances [7].

Many contributions refer to the analysis of scientific papers, which share a lot of characteristics with legal documents. Firstly, both classes of documents can be described as semi-structured: the structure of first- and second-instance judgments, as well as the IMRaD format of scientific papers, suggest a standard that is often disregarded or misinterpreted. For this reason, both domains present difficulties in dealing with discourse changes within the document. They are conceptually segmented into ordered paragraphs describing the information (preamble, introduction, specific case, reasons and procedures) leading to the decision in a legal judgment or to the paper’s conclusions. Moreover, they contain citations that interconnect legislation and case-law or scientific papers as appropriate. In particular, scientific papers have interconnections that determine their contextualization in the literature. In the case of the smart citation index *scite* [8], papers were analyzed from multiple perspectives in order to determine the interconnection between documents and the context. In particular, their work focuses on extracting citations and determining the semantic context given by how and where they are cited in the document. In addition, it is pointed out that there may be different types of citations that need to be treated differently in defining the context around citations and that may change in importance and meaning over time.

Another important aspect is the transparency and interpretability of the results. The pipeline framework [9] focuses on the extraction of feature sentences for the purpose of providing interpretability to Similar Case Matching. Extracting them through Deep Learning technologies can make the output difficult to interpret for non-AI experts.

4. Legal Judgments Reading and Comparison: a Case Study

In order to elaborate our methodology, we focus on a case study based on Italian judgments of first- and second-instance followed by generalization proposals to extend our pipeline to other legal systems and to documents written in other languages.

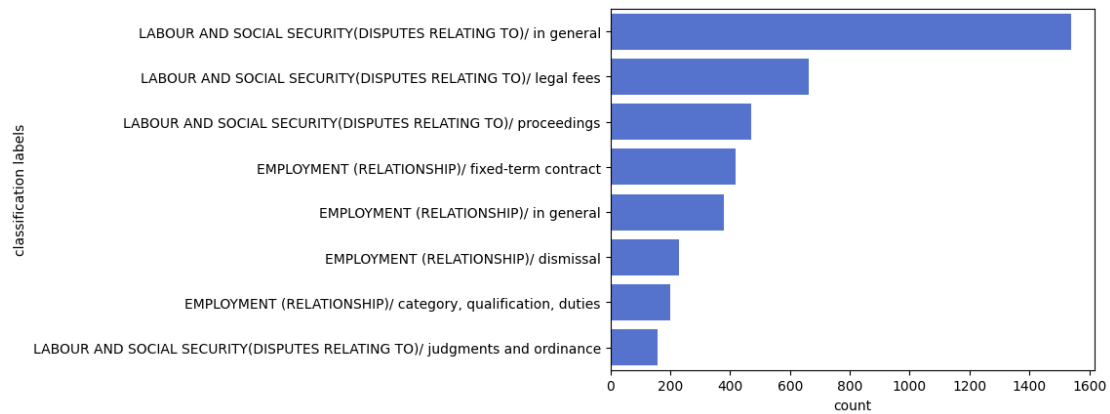


Figure 1: Judgments distribution within the 8 most common labels in the dataset.

4.1. Data

The dataset used for these experiments was retrieved from a public online platform [10] for the consultation of Italian legal documents. The collection is a sample downloaded from the platform that contains legal judgments (i.e. *Sentenze*) of the first and second instance from the labor section (i.e. *Sezione Lavoro*), each entry is described with the following attributes: filename, Court of jurisdiction, section (i.e. *Sezione*), labels (i.e. *Voci*), judgment identifier (Code and Year), NRG (Code and Year, with which each case is associated) and type (first- or second-instance). *Sezione* and *Voci* describe the area to which the legal judgment belongs, the former describing the macro-area and the latter the sub-area in which the related legal case lies.

A total of 5059 legal judgments were extracted, labeled with the 8 most populated *Voci* (Figure 1) and having the decisions paragraph well defined and separated from the facts paragraph of the document. Though the data sample is not totally random, the possible bias introduced doesn't compromise the outcome of the experiments to the best of our knowledge. It has been observed that in first- and second-instance legal judgments the structure of the document is not standard. Moreover, facts and decisions paragraphs may be combined in a single paragraph of the judgments file and therefore these were discarded in order to treat the two parts of the document separately properly.

4.2. Modus Operandi Analysis

In order to draw inspiration from the legal domain experts' *modus operandi* for reading and comparing legal judgments, we proceeded by conducting a process analysis performed by two lawyers. They read two judgments together and commented aloud on them by referring to relevant and irrelevant information in the document to determine whether the two judgments are similar or not similar. This step aimed to intercept the sequence of operations, reasoning, and weights that are given to the components of legal judgments in order to bring out the domain know-how and methodologies in the workload. Our assumption (and proposal) is that the process of reading and comparing can be generalized into an algorithmic form, in this way

it can be automated to provide daily support. The manual process took about 1 hour, and we expect that the automated system can be faster, bringing organizational benefits.

Reading of the judgment Having specified the context-based similarity criterion, the lawyers began reading one judgment at a time. The focus initially fell on *Sezione*, matter, and Court grade (first-instance, second-instance (or Appeal), or Supreme Court). Afterward, they read the parts of the legal judgment paying attention to the presence of institutions; careful attention should be paid to the fact that judgments are often anonymized in respect of privacy. Thereafter, they carefully read all paragraphs of the document (e.g. facts, decisions, PQM - i.e. an Italian acronym used to introduce the decision, etc.). The first part of the facts and decisions paragraphs was of particular relevance in defining, in a coarse-grained manner, the domain of the case on which the legal procedure was applied. Lawyers pointed out that the legislation and case law cited would require extensive knowledge that in some cases may require consulting the rules and maxims for a greater understanding.

Comparing two judgments Once the accurate reading of the judgments was completed, they proceeded by comparing the information found in the judgments so as to define whether or not they were similar. In particular, it was noted how *Sezione* and *Voci* are of paramount importance in constrainingly defining the context within which the judgments fall. If *Sezione* and matter are different, then the judgments are by their nature different from each other. Subsequently, the focus was on the first part of facts, which define the specific scope of the case, and decisions, which define the legal procedure applied to the specific situation described in the facts.

The discussion with the lawyers revealed that, depending on the search criteria, it is useful to analyze the facts, if one wants to look for similar situations, and/or the decisions if one wants to give more importance to the legislation and case law cited. Since the second-instance judgments contain information related to the first-instance, which would influence the comparison without introducing additional relevant information, it was preferred to consider only the first part of the decisions that contains information of greater importance to define the legislation and case law applied that allow defining the topic and field.

5. Methodology

This section describes the steps that led the study from the analysis of the *modus operandi* of legal domain experts to the contextualization of legal judgments and the calculation of the Similarity Score. The output is accompanied by auxiliary information useful for understanding the system's results.

5.1. Augmented Reading and Comparison Pipeline

The *modus operandi* of domain experts led us to the implementation of a pipeline (Figure 2) with the purpose of translating the process of reading and comparing judgments into a generalized algorithmic form. Taking into consideration the scalability of our solution, we propose a

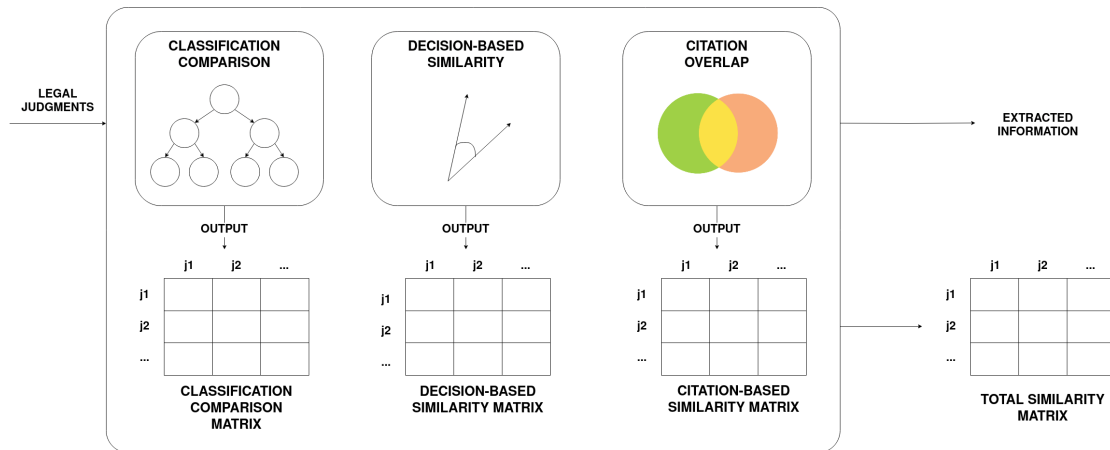


Figure 2: Flowchart describing the pipeline for comparing legal judgments.

modular pipeline, whose steps are somewhat independent and can be computed separately, storing partial results in JSON or CSV format. This methodology allows a single calculation that results in a final matrix of similarity scores that allows the data to be queried with linear complexity in the number n of legal judgments in order to retrieve the m most similar documents given a judgment.

The pipeline consists of three modules (as shown in Table 1): classification comparison, decision-based similarity and citation overlap calculation, followed by the calculation of the similarity score according to the outputs of the previous steps. Given all the sentences in the dataset, each step produces in output a matrix that will contribute to the calculation of the last matrix containing the total similarity scores.

The target of the system includes experts in the legal domain such as magistrates and lawyers. Since computer systems are often difficult to understand, it is necessary to provide a way to make the results explainable and allow critical reasoning with respect to the effectiveness of the output. To do this, information is extracted during data processing that is useful in calculating individual scores in modules, which upon completion of reading and comparing judgments allows for an explanation of why the machine returned the results. Since trust in the computational system applied in the legal domain is fundamental, it is very important to be able to critically evaluate the results to come up against accountability and the effects on the parties in the legal case.

5.2. Classification Comparison

As revealed by the performed process analysis, the most important features in evaluating the relatedness of two judgments are their *Sezione* and *Voci* which determine their classification within each Court since they carry important information about the documents' topic. In developing our work, these labels were available within the dataset.

Expanding this pipeline to a more generic scenario this information can be extracted through a process of automatic legal document classification. Domain experts claim that two judgments

module	extracted information	output
classification comparison	For each judgment, Sezione and Voci with which they are labeled.	Classification Comparison Matrix where for each pair of judgments, we have 1 if the two documents are classified in the same way and 0 otherwise.
decision-based similarity	For each judgment, the portion of the decisions paragraph analyzed in the module.	Decision-based Similarity Matrix where for each pair of legal judgments we have a decisionsSimilarity score between -1 and 1.
citation overlap	For each pair of judgments, legislative and case-law citations in the overlap.	Citation-based Similarity Matrix where for each pair of legal judgments we have a citationOverlap score between 0 and 1

Table 1

The information extracted from each module of the pipeline as well as their output.

having different *Sezione* or *Voce* are to be considered different and cannot be compared further. Conversely, once it is established that two documents cover the same topic, it is possible to continue investigating their level of similarity through the analysis of their content. This first step of the pipeline acts as a filter and consists of comparing the labels *Sezione* and *Voci*. If they are equal we proceed with the calculation of the similarity scores based on the document semantics performed in the next steps otherwise, the two legal judgments are different. A legal judgment can have multiple *Voci*, to be considered equal they must have at least one entry in common.

Given two judgments $j1$ and $j2$, the classification comparison value was calculated as follows:

$$classificationComparison = \begin{cases} 1 & \text{if } Sezione1 == Sezione2 \text{ and } len(Voci1 \cap Voci2) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $(Sezione1, Voci1)$ and $(Sezione2, Voci2)$ are the *Sezioni* and *Voci* of legal judgments $j1$ and $j2$, respectively.

5.3. Decision-based Similarity

The second step is the computation of the judgments' lexical similarity. This value is obtained by comparing the first portion of the documents' decisions paragraph, which, as suggested by domain experts, contains the most information regarding the legal case. The text was preprocessed through a pipeline that includes 1) conversion to lowercase, 2) removal of special characters, 3) removal of URLs and HTML tags, 4) conversion of word numbers to their numeric form, 5) removal of stopwords and 6) lemmatization through Morph-it! [11]. Each text was then converted into embeddings using Italian Legal BERT [12] in order to perform a pairwise comparison through cosine similarity.

The choice for this kind of approach was dictated by the fact that very few Courts provide a classification taxonomy with a granularity finer than the documents' *Voci*, making it difficult to find a subtopic label at this stage.

Although this approach allows us to find a deeper level of similarity, it reduces the overall interpretability of the pipeline results since it relies on deep learning techniques.

Given two judgments $j1$ and $j2$, the decision-based similarity score was calculated as follows:

$$decisionSimilarity = cosineSimilarity(textEmbedding(j1), textEmbedding(j2)) \quad (2)$$

5.4. Citation Overlap

Subsequently, we focused on the extraction of citations within the decisions' part of the judgment as they can be used to define the juridical context in which the judgments are located. In this perspective, the document can be characterized by the citations it contains; hence documents with a higher citation overlap will have a higher similarity level as they appear in the same juridical proximity. In other words, as in topic modeling the words describe the topic of the document, in the case of legal judgments the citations define the context and domain of application of the legislation and procedures applied to the decisions.

Citations are extracted following the idea that two judgments dealing with similar decisions cite similar legislation and case-law. Citations in documents were extracted through the use of Linkoln [13], which allows, given a text, to extract legislative and case-law citations using a regex-based methodology. The extraction provides a list of citations described with text, context, identifiers, reference type (legislative or case-law), authority (Court of first- and second-instance or Supreme Court), document type (legal judgment, decree-law, legislation, etc.), normative link (where available), etc. The proximity between citation-base contexts was calculated through the overlap between citations within legal judgments as follows.

Given two judgments $j1$ and $j2$:

$$citationOverlap = \frac{|citations(j1) \cap citations(j2)|}{|citations(j1) \cup citations(j2)|} \quad (3)$$

5.5. Total Similarity Scores Calculation

The aforementioned modules can work in isolation and independently of each other, each providing in output a square matrix of order n , where n equals the number of judgments in the dataset. The matrices that contribute to the calculation of the Total Similarity Matrix (TSM) are as follows: Classification Comparison Matrix (CCM), Decision-based Similarity Matrix (DSM), and Citation-based Similarity Matrix (CSM). Each cell in the TSM matrix is populated with a value ≤ 2 calculated as below.

Given two judgments ($j1, j2$) and the equations 1, 2 and 3:

$$similarityScore = classificationComparison * (\alpha * decisionSimilarity + \beta * citationOverlap) \quad (4)$$

where α and β represent the weights to be assigned to the last two pipeline components.

In matrix form on total judgments, we can rewrite the calculation as below:

$$TSM = CCM * (\alpha * DSM + \beta * CSM) \quad (5)$$

where α and β represent the weights to be assigned to the last two pipeline components.

The α and β multipliers allow for different weighting of similarity in decision-based context and citation-based context. This is because, within a legal case decisions, legislation, and case-law can take on different impacts in setting the context. When comparing two judgments, legal domain experts attach different importance to the components just mentioned and in some ways, it can be subjective in nature. α and β have a default value of 1, the first approach was to keep this value in the similarity calculation, thus considering the two pipeline components equally.

Given a legal judgment, making a matrix with total similarity scores allows one to access the matrix with linear complexity and select the m judgments that are most similar from the perspective of the context described by decisions, legislation and case-law. In this way, reading the judgments and extracting the information occurs only once showing a potential advantage over reproducing the same operation manually.

5.6. Output Interpretability

For each sentence pair, the pipeline outputs a total similarity score and auxiliary information useful for understanding the result. The information extracted and maintained during the execution of the individual modules is broken down as in Table 1. Given two judgments $j1$ and $j2$, the extracted information supporting the score is composed by combining the information extracted from the individual modules. This kind of interpretable output format allows for a more transparent similarity measure, easy to motivate from a legal perspective.

Since we cannot provide an interpretable output for the decision-based similarity module, the numeric score is substituted with the text used to compute it, so that a legal actor can easily evaluate the accuracy of the comparison.

6. Evaluation

Due to the nature of the task and the domain of the documents, it was necessary to resort to domain experts in order to obtain an evaluation of the resulting similarity scores. The experts were presented with a set of triplets of judgments with shape $(\star, +, -)$ as evaluated by the pipeline where

- \star indicates the judgment taken into consideration, selected randomly from the analyzed data
- $+$ indicates a judgment evaluated as similar to \star
- $-$ indicates a judgment evaluated as different from \star

They were then asked to binary rate the outputs as accurate or inaccurate and the time taken for the reading of each judgment.

Inter-Annotator Agreement The Inter-Annotator Agreement was evaluated using Cohen's Kappa [14], obtaining $\kappa = 0.769$. The same metric was used to compute the average agreement

between the annotators and the pipeline output, resulting in $\kappa = 0.885$. It is worth noting that the final evaluation was influenced by the subjective nature of Similar Case Matching.

Time The average reading time (minutes per page) for legal judgments is calculated among the annotators and separated by types ($\text{time}_* = 3$, $\text{time}_+ = 3$, $\text{time}_- = 1.7$). The time starts with the reading of the document in order to define whether two sentences are similar or different and ends with the closing of the document. The reading does not have to be complete. Notably, these measurements do not include the search for judgments that legal domain experts would have had to do in the absence of this system.

On the other hand, through the computational pipeline analysis, in a few seconds a judgment can be analyzed and compared to every other judgment in the dataset, in order to obtain a similarity score. Moreover, it is important to acknowledge that once the judgment is analyzed by the system, all the partial results are stored and accessible in $O(n)$.

Further considerations From the evaluation process it emerged that judgments redacted by the same judge in a close time interval tend to be very similar to each other. This can be motivated by personal stylistic choices such as chosen lexicon and personal interpretation of the cited documents.

7. Future Work

In this section, possible future developments of this work are presented by topic.

7.1. Classification Comparison

From the perspective of generalizing the pipeline to different judgment sets, its first step should not rely solely on the labels present in the dataset. A more generic approach can be found by automatically classifying the documents and storing this information together with the judgments' metadata. Given the absence of a standard of classification among courts, this task presents different challenges and requires an alignment of the different classification taxonomies.

7.2. Decision-based Similarity

The extraction of the lexical embeddings as presented in this work can be improved in different ways. The first aspect to take into account is the extraction of the most significant part of the judgment on which to focus to obtain said embeddings. While it is true that the most relevant information is to be found in the first part of the decision paragraph, there has to be a more thorough analysis of the length of the portion to be selected. Furthermore, since juridical documents are unstructured by nature, the decision paragraph is not always present and can often be merged with other *Sezioni* such as *Fatti*. For this reason, it is necessary to find a method of identifying such a paragraph that does not rely solely on the formatting of the document. Looking at the module from another perspective, more investigation could be done on the model to be used to create said embeddings.

7.3. Citation Overlap

In order to define a more precise and meaningful citation overlap score, this measure could be weighed in different ways.

Citation stance A first suggestion would be to refine the calculation of overlap between citations by weighing each citation according to the stance of the context in which it appears. This is because a sentence may cite another document to accommodate its meaning or to distance itself from it. From this point of view, two judgments having the same citation but with different stances would indeed be in the same legal context but would have a different connotation that would allow a different nuance of meaning to be highlighted.

Naively, this differentiation can be tackled through the computation of the polarity of the sentences in which the citations appear. Subsequently, said results can be further refined by a more precise stance detection analysis. Nevertheless, given the domain of the documents and the fact that the target of the analysis is known and always present in the sentences taken into consideration, the two tasks are in some way overlapping and the polarity of a sentence can coincide with its stance in regards to the citation.

Citation relevance The score computation could be further perfected by assigning a different degree of importance to each kind of citation. Although all references to other legal documents contribute in some way to defining the document's context, some (such as Supreme Court decisions) have higher relevance than others.

Through a deeper investigation of the legal landscape, it would be possible to configure these weights so that each citation could be given the importance that it actually has, ulteriorly refining the overlap score.

On the same idea, some documents could be often referenced without a direct correlation to the judgment's topic. In this scenario, it would be appropriate to consider a relevance measure to lower their contribution to the overall overlap measure. In particular, always-cited and background citations do not attribute additional information for defining the citation-based context, so it may be relevant to ignore them.

Improved extraction of caselaw citations The extraction of the Supreme Court's rulings can be explored even to a greater extent since it can be useful to map *principles of law* stated and renewed. In the Italian legal system, it is considered a principle of law a generalization of the interpretation and application of the rule to a concrete case, given by the function of nomophilacy given by the Supreme Court (i.e. Corte di Cassazione). As discovered through the analysis conducted by our legal domain experts, there is a one-to-one correlation between principles of law and Supreme Court's citations in the explicit form. With the proper pre-processing and structure of the data, the aid of proper algorithms, the pattern found by the regex used in our experiments can represent an accurate extraction methodology of both Supreme Court citations and *principles of law*.

Outcome extraction An additional step in our pipeline would be the extraction of the judgment's outcome which would allow for a better understanding of the judgment as a whole.

With legal research in mind, being able to automatically extract the output of a judgment would enable more sophisticated features such as filtering the documents given their final decision.

Time evaluation It will be interesting to study the manual task again with the support of the developed computer system, so it may be possible to evaluate the usefulness, effectiveness and time gained. In addition, feedback given by domain experts about the actual potential of the search engine support is important.

8. Conclusions

The algorithmic reproduction of legal domain experts' *modus operandi* can be implemented and of great interest for the optimization of legal proceedings, as it was shown to be far more effective and time-efficient than the current procedure.

The modularity of the pipeline allows for a scalable solution, which allows for a thorough analysis and comparison of legal documents in seconds and an indexing in $O(n)$. By integrating such computation in an online platform for the consultation of legal documents, it would be possible to increase the productivity of magistrates by providing them with an interpretable and transparent resource targeted to legal domain experts. Consequently, by facilitating the reading and the retrieval of relevant information within the text, this approach would contribute to the disposal of the backlog inside Courts.

Furthermore, the reproduction of an existing procedure for reading and comparing legal documents, allows for the collection of relevant data, which can be used by legal domain experts to evaluate the pipeline output.

The dataset used for these experiments was retrieved from a public online platform aimed at consulting Italian legal documents [10]. Nevertheless, the proposed pipeline can be extended to any measure by any legal system, by using a legal dataset retrieved from Eur-Lex [15], an online portal that provides the official and most complete access to EU legal documents, in all 24 official EU languages. Its use may allow the identification of judicial similarities even between judgments written in different languages, facilitating the drafting motivation-phase of rulings by Member States Courts, related, for example, to preliminary reference to CJEU (i.e. Court Of Justice Of The European Union).

Furthermore, the availability of the same text in multiple languages would allow optimal results by computing every module of the pipeline using the language in which the technologies used have the best performances.

Acknowledgments

The research work has been funded in Next Generation UPP⁵ project supported from the European Union, National Operational Program Governance and Institutional Capacity 2014-2020, European Social Fund and European Regional Development Fund. The Next Generation UPP project is part of the "Unitary project for the dissemination of the Office for Trial and the

⁵<https://www.nextgenerationupp.unito.it/home>

implementation of innovative operating models in the judicial offices for the disposal of the backlog”, promoted by the Italian Ministry of Justice and implemented in synergy with the interventions envisaged by the National Recovery and Resilience Plan (NRRP) in support to the justice reform.

References

- [1] Marco Ciccarelli, Le banche dati di giurisprudenza e l’ufficio per il processo, accessed 02.05.2023. URL: <https://www.questionegiustizia.it/articolo/le-banche-dati-di-giurisprudenza-e-l-ufficio-per-il-processo>.
- [2] W. G. Harrington, A brief history of computer-assisted legal research, *Law. Libr. J.* 77 (1984) 543.
- [3] S. Kumar, P. K. Reddy, V. B. Reddy, M. Suri, Finding similar legal judgements under common law system, in: *Databases in Networked Information Systems: 8th International Workshop, DNIS 2013, Aizu-Wakamatsu, Japan, March 25-27, 2013. Proceedings 8*, Springer, 2013, pp. 103–116.
- [4] G. Wiggers, S. Verberne, G.-J. Zwenne, Citation metrics for legal information retrieval: Scholars and practitioners intertwined?, *Legal Information Management* 22 (2022) 88–103. doi:10.1017/S1472669622000160.
- [5] S. Paul, P. Goyal, S. Ghosh, Lesicin: A heterogeneous graph-based approach for automatic legal statute identification from indian legal documents, in: *AAAI Conference on Artificial Intelligence*, 2021.
- [6] J. S. Dhani, R. Bhatt, B. Ganesan, P. Sirohi, V. Bhatnagar, Similar cases recommendation using legal knowledge graphs, *arXiv preprint arXiv:2107.04771* (2021).
- [7] P. Bhattacharya, K. Ghosh, A. Pal, S. Ghosh, Methods for computing legal document similarity: A comparative study, *arXiv preprint arXiv:2004.12307* (2020).
- [8] A. Cohan, W. Ammar, M. van Zuylen, F. Cady, Structural scaffolds for citation intent classification in scientific publications, in: J. Burstein, C. Doran, T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics, 2019, pp. 3586–3596. URL: <https://doi.org/10.18653/v1/n19-1361>. doi:10.18653/v1/n19-1361.
- [9] N. Lin, H. Liu, J. Fang, D. Zhou, A. Yang, An interpretability framework for similar case matching, *ArXiv abs/2304.01622* (2023).
- [10] Leggi d’italia p.a., accessed 27.04.2023. URL: https://pa.leggiditalia.it/#mode=home,__m=site.
- [11] E. Zanchetta, M. Baroni, Morph-it, A free corpus-based morphological resource for the Italian language. *Corpus Linguistics* 1 (2005) 2005.
- [12] D. Licari, G. Comandè, ITALIAN-LEGAL-BERT: A Pre-trained Transformer Language Model for Italian Law, in: D. Symeonidou, R. Yu, D. Ceolin, M. Poveda-Villalón, D. Audrito, L. D. Caro, F. Grasso, R. Nai, E. Sulis, F. J. Ekaputra, O. Kutz, N. Troquard (Eds.), *Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowl-*

edge Management, volume 3256 of *CEUR Workshop Proceedings*, CEUR, Bozen-Bolzano, Italy, 2022. URL: <https://ceur-ws.org/Vol-3256/#km4law3>, iSSN: 1613-0073.

[13] Lincoln, accessed 27.04.2023. URL: <http://ittig.github.io/Linkln/>.

[14] J. Cohen, A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* 20 (1960) 37 – 46.

[15] eur-lex.europa.eu, accessed 26.05.2023. URL: <https://eur-lex.europa.eu/homepage.html>.