Hybrid Classification of Audit Court Decisions using Online Context-Driven Neural Networks

Leticia Decker[1], Monica Palmirani[2], Salvatore Sapienza[2] and Vittoria Pistone[2]

[1]*University of Bologna and Scuola Normale Superiore di Pisa*
[2]*University of Bologna*

Abstract

This paper proposes an experiment made on a portion of decisions of the Italian Audit Court (Corte dei Conti) with the aim to classify the typologies of the amounts included in the text for permitting better searchability by the judges according to the principle "follow the money". The experiment is based on the decisions modelled into the Akoma Ntoso format, one of the most popular XML standards used in LegalXML domain for representing legal knowledge. Akoma Ntoso permits semantics annotation to facilitate Natural Language Processing (NLP) and Machine Learning (ML) approaches fostering legal metadata and document structure. Therefore, it makes possible to promote cross-document analyses with massive amounts of data and correlate events and detect anomalies in different legal processes. We introduce a methodology to automate the pattern recognition, information extraction, and document qualification through semantics annotation to improve information retrieval by legal operators within the judicial system and outside. The methodology recognises patterns of interest (PoI) using regular expressions. A supervised ML classifier uses the semantics context of PoI to infer the attribute values of its annotation. The extracted PoI context is used to generate a dataset organized in a tree format in which each level represents an abstraction level of the PoI classification. The annotation is a tagging of the recognised pattern in the text associated with its description as attribute value. The strategy learns the vocabulary of the PoI classification linking different levels of a conceptual abstraction. As use-case, PoI consists of monetary amounts associated with their respective origin (attribute), facilitating the identification of common patterns in the rulings of Audit Court, within the same Region/Court and among them.

Keywords

LaTeX class, Natural Language Processing, Legal Informatics, Machine Learning, Semantic Annotation, Legal Ontology,

1. Introduction

In the 2022 Rule of Law report [1] the European Commission examines the state of the art of the anti-corruption policies in each member state of the EU landscape[1]. In this report, there is a specific note to Italy to reinforce the policies and the instruments for fighting corruption,

FOIS 2023 - KM4LAW Workshop, Canada
leticia.deckerdesousa@sns.it (L. Decker); monica.palmirani@unibo.it (M. Palmirani); salvatore.sapienza@unibo.it (S. Sapienza); vittoria.pistone@unibo.it (V. Pistone)

[1] "The anti-corruption frameworks, focusing on the effectiveness of national anti-corruption policies and assessing different key areas of action taken by Member States to prevent and fight corruption. Effective anti-corruption frameworks as well as transparency and integrity in the exercise of state power, strengthen legal systems and citizen and businesses' trust in public authorities."

especially in public administration[2]. This paper wants to classify the amounts included within the judgments of the Italian Audit Court for detecting the meaning under the legal point of view of these valuable parts of the decisions. Using the classification we could create a data analytics dashboard in the future, supporting the work of the judges to apply in a homogeneous way similar cases.

Digitization is a primordial process to improve the efficiency of data analysis, extracting information, facilitating their access, and automating their processing through machine learning (ML) approaches [2]. In particular, one of the issues of the management of legal information is the extraction of semantic knowledge to establish topic-driven legal ontologies [3].

This paper contributes to the areas of knowledge mining and information extraction, which are laborious and error-prone tasks and can be approached through semantic annotations [4] for improving the searchability of decisions especially using the meaning of the amounts. Such computer-understandable information is based on descriptions of resources through annotations in different modes as metadata or text markup [5]. Applied to jurisprudence, semantic annotation identifies legal concepts, (e.g. type damage, agents connected), in textual documents, combined with the temporal parameters (e.g., period of time connected with the damage) and jurisdiction (e.g., regional districts) [6] [7].

This paper investigates the use of Online Learning classifiers for the legal classification of quantities in the rulings by the Italian Audit Court, on the basis of a legal taxonomy built on procedural and substantive public law. The methodology used is based on hybrid AI [8]. Following this introduction, Section 2 presents the methodology used in this study, while Section 3 identifies related works. Some key concepts and the background of the study are presented in Section 4. Section 5 discusses the advantages of XML-enriched documents and the use of taxonomic trees for representing knowledge. Section 6 introduces the technical background of Online Context-Driven Neural Networks. Section 7 details the hybrid methodology of the case study with regard to data structuring and annotations, while analysing the dataset and metrics used in the experiments. Section 8 presents the results and their legal significance. Section 9 concludes the paper by addressing its limitations and setting directions for further research.

2. Methodology

We use hybrid AI for merging semantic analysis (e.g., taxonomy, metadata from Akoma Ntoso (AKN)) and ML approach in order to optimize the accuracy, minimise the supervision effort of the dataset by the legal experts, and include elements of explicability for the final legal interpretation. We used a dataset of decisions of the Italian Audit Court in the interval 2000-2015. We have selected about 500 decisions concerning the Emilia-Romagna case-law[1]

The methodology is the following: 1) all the decisions between 2000-2015 are transformed in Akoma Ntoso for reusing the descriptiveness of the portions of the document (e.g., decision, motivation, facts, temporal metadata, jurisdiction, entities); 2) three legal experts (one senior

[2]See the Report of the Ministry of Interior of Italy https://www.interno.gov.it/sites/default/files/2022-03/report_reati_corruttivi_02.2022.pdf; and the European Court of Auditors statistics where Italy has the larger number of staff in Europe devoted to the Audit Court https://op.europa.eu/webpub/eca/book-state-audit/en/#A-9

[1]The dataset and the code are available at https://gitlab.com/CIRSFID/ai4justice-classifier. The selection has been mainly based on temporal elements to ensure the generalisability of the results in a significant timespan

researcher, one young researcher, one lawyer) define a taxonomy of the different typologies of the amounts according to administrative law. This taxonomy is based on three axies: (i) the role of the amount in the final decision (e.g., damage or costs for the trial), (ii) the type of harm (e.g., image, revenue), (iii) the source of payment (e.g., salary, pension). The taxonomy is serialized in RDF-Turtle; 3) the paragraphs including the amounts are extracted using NER technique. This step provides inputs for improving the taxonomy with real data (bottom approach). 4) once the taxonomy is accurate enough, it is used by the ML tool for creating a model of classification (62 documents, 239 samples); 5) a new dataset from Emilia-Romagna is used for the evaluation (33 documents, 52 samples dataset); 6) two legal experts (lawyer, young reserch) manually classify this second dataset; 7) the comparison between the ML classification and the manual one produces the evaluation of the indexes; 8) legal analysis of the comparison results is done by legal experts for discovering hidden information for improving the taxonomy or the model.

3. Related Works

Pattern mining commonly produces a large set of recurrent patterns that may be classified according to the semantic context in order to be interpreted properly. Information Extraction (IE) and Semantic Annotation (SA) often appear together in this type of text-processing problem. The medical domain [9], instructional texts [10], and social media analysis [11] make use of IE and SA to solve different problems in free-text data.

Semantic annotation is commonly applied to legal document processing but rarely this step is used as input for ML elaboration. A paper develops a solution to the IE automation focused on structured knowledge, improving semantic search and ontology generation on textual databases [3]. The extraction system is based on NLP techniques as Part-Of-Speech tagging and Syntactic Parsing. The information is generalised in features to identify its linguistic context. A Support Vector Machine classifier generates a model to the SA automation using the new data. The system is tested using legal texts, focusing on the identification of hypernym relations and definitions. A work presents the GaiusT tool as a SA framework for textual data that generates elements of requirements modelling, as actors, rights, and obligations [12]. The problem is approached through analysis and annotation of legal documents in prescriptive natural language. GaiusT and its later versions are web-based systems that extract processes in a semi-automate way. It is a multi-phase framework that pre-processes the text and supports SA using existing linguistic tools preferably [13, 6]. Still on the same topic, a paper uses Cerno to Italian legal documents. Cerno is a lightweight framework with multilingual SA support applied to regulations in different domains [14]. Another paper presents a system to enrich legal texts with SA in order to index and recover legal documents. SALEM (Semantic Annotation for LEgal Management) solution is a software developed for SA automation of (Italian) law texts, using NLP to classify law paragraphs, and extract relevant text fragments that differentiate semantically types of regulatory content [15]. Legilocal, a software solution to help local authorities to improve the French and European administrative acts, is presented by [16]. Legilocal centralises the content management and the actor interactions, combining content management, networking, SA, and search techniques from the drafting to acts publication. A paper approaches the semantic recognition of logical parts of Vietnamese legal documents based on a Conditional Random

Fields implementation – a statistic ML method [17]. The solution is based on the identification of linguistic features of words and part of speech, and semantic characteristics of triggers and ontologies in a Vietnamese Business Law dataset. The solution obtains 78.12% at precision and 68.72% at recall measure. A recent work makes use of Akoma Ntoso-marked documents and a hybrid pipeline to find correlations and patterns among derogations in European Union Law [8].

4. Preliminary Concepts

In this section, we describe the main concepts to approach the presented problem – (i) Akoma Ntoso markup language, (ii) the processing of Akoma Ntoso documents, (iii) the taxonomy of the Audit Courts.

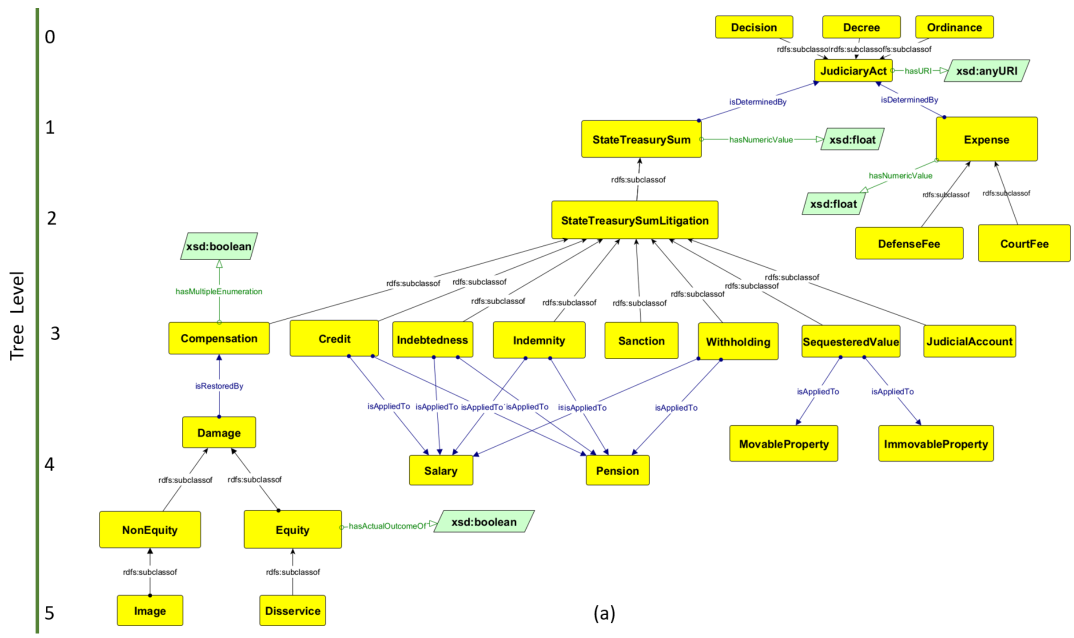
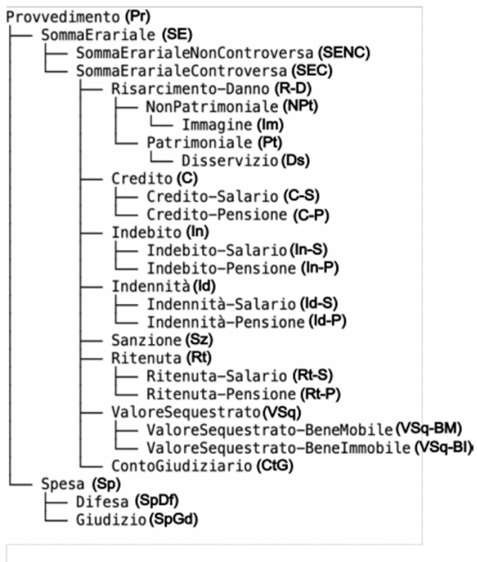


Figure 1: In (a), it is shown the <quantity> taxonomy in graph, in which the nodes are legal concepts that are linked according with different relations shown in the figure. The classes of the taxonomy have been translated to English to ease the reading of the paper. The original classes (in Italian) are displayed in Figure B below

4.1. Akoma Ntoso Description Markup

The Akoma Ntoso (Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontology) is a XML-standard of LegalDocML Technical Committee of OASIS (Organization for the Advancement of Structured Information Standards), promoted by the United Nations Department for Economic and Social Affairs (UNDESA) for



(b)

Figure 2: In (b), the taxonomy is mapped into a semantics tree, linking legal concepts according with their abstraction levels, generating a class-subclass relationship between parent-child nodes. Higher is the tree level, less abstract is the legal concept. In (b), each class is associated with its abbreviation that is used over the text.

modelling the legal documents. Akoma Ntoso is formed by a set of XML descriptions of parliamentary, legislative, and judiciary documents, defining legal documents in structured format, being based on 3-level structure: (i) text, (ii) structure, and (iii) metadata [18]. In (i), the original document provides the text to its XML version; in (ii), each type of legal document is mapped in a textual skeleton with mandatory and optional parts, keeping the trade-off between correctness and flexibility; and in (iii), the set of information is derived from the text to simplify the data analysis. The metadata is a preprocessing step that organise the legal information extracted from the document since the NLP is an expensive task [19]. Each Akoma Ntoso representation is associated with just one original expression of a single work, according with the Functional Requirements for Bibliographic Records (FRBR) model, and each document category is associated with on just structural format even if in different expression types. This permits to assign to each provision of the decision a specific identifier that is used in the ML step for inferring the position in the document of the classified information, and so to know other important related metadata information (e.g., time parameters).

4.2. Audit Court Decisions in Akoma Ntoso

Legal documents can be transcribed to Akoma Ntoso manifestations, being structured in (i) introductory part, (ii) main body content, and (iii) conclusions. In (i), the document presents the identification data as the document type, number,and issuing authority. Following, in (ii), it is presented the content lies organized on background of the case (<background>), argumentation of the judge (<motivation>), decision of the judge (<decision>). At the end, in (iii), the document is concluded through the closing formulas, dates, and signatures [18]. Akoma Ntoso offers also several inline elements useful for creating rules (e.g., person-role-amount).

Italian Court of Audit decisions are formalised using the type of document of AKN called 'judgement' that is structured in: (i) <meta>, (ii) <header>, (iii) <judgementBody>, and (iv) <conclusions> [20]. In (i), metadata contain information about the type of the Court, the jurisdiction with the region to which the court belongs (because the Italian Audit Court's jurisdiction is based on a regional criterion), the temporal parameters, and any other legal metadata useful for classifying the decision (e.g., keywords). In (ii), introductory information is displayed, such as the date, the number of the decision, the parties, the lawyers, the judges and the introductory formulae. In (iii), the presentation of the parties' claims, the factual background, the motivation of the ruling, and the decision. The decision puts an end to the case by identifying the unsuccessful party, issuing sanctions, and determining the judicial expenses. Therefore, it is separated from the rest of the text and it is introduced by the "P.Q.M." (Per Questi Motivi, i.e., For These Reasons) formula. In (iv), a closing paragraph contains the signature(s) of the judge(s) and the issuing date of the ruling. A particular role is played by the normative citations and of other relevant case-law. The function of these citations are annotated in the metadata (e.g., citation in support, in overrule, etc.).

4.3. Taxonomy of Amounts in Italian Audit Court Decisions

The taxonomy of the typologies of amounts of the Italian Audit Court has the aim to describe the role that each amount plays in the final decision of the judge according to public law. We used four criteria: a) the role played in the procedural law of the "Code of Accounting Justice"; b) the type of injury; c) the provenance of the sum; d) the accumulation of the amounts. The heterogeneous typologies are in this taxonomy simplified with the goal to offer to the judge a new instrument of analysis of the case-law and also to perform, in the future, network analysis of the relationships between different case-law. The first used criterion is based on the teleological role of the amounts involved in the litigation: expense (Spesa) or State treasury sum (SommaErariale). All the nomenclature used (in Italian) is based on the Code of Accounting Justice[3]. The "State Treasury Sum" (SommaErariale) top-class represents the typologies of each income or payment in the treasury coffers, including the values that are only hypothetical because they were determined as a precautionary measure. The "Expense" (Spesa) top-class represents the disbursement approved by the judge and charged to all parties or, more frequently, to the losing party. The subclasses "Defense", which means defense costs, (Difesa), and "CourtFee" (Giudizio), which means the costs for the offices, specify the nature of the disbursement by indicating the service that justifies it. Taking into account the competence for matters attributed to the Audit Court by art. 103 of the Constitution and of the various phases of the accounting process, it possible to give the sub-classes name (Compensation, Credit, Indebtedness, Indemnity, Sanction, Withholding, Sequestered Value, JudicialAccount) representing the sum according to the object of the dispute. The sum, for example, is classified

[3]See also http://www.normattiva.it/eli/stato/DECRETO_LEGISLATIVO/2016/08/26/174/CONSOLIDATED

as a "Credit" when it expresses the economic value of a credit right which is the object of the dispute in the pension proceedings. The typology based on the source of the amount is classified here using “Salary” (Salario), “Pension” (Pensione), “MovableProperty” (BeneMobile), “ImmovableProperty” (BeneImmobile). It offers a second criterion for a further specification of the state treasury sum, which, in addition to acquiring characteristics from the transit in the process, affects the components of such movable and immovable property in the case of the sanction and the salary and pension in the case of withholding. This relationship is managed using *isAppliedTo*. Another important class is “Compensation” (Risarcimento) which introduces the third criterion of analysis: the typology of injury. We have "Damage" (Danno) class, which is connected with the top-class with the relationship *isRestoredBy*. It expresses the economic value of the injury suffered by the injured party, which the sum qualified as "Compensation" intends to restore. The "Equity" and "Non-Equity" sub-classes, in compliance with the provisions of the Italian Civil Code, characterize the damage on depending on the type of interests harmed by the damage (e.g., property rights, personality rights, etc.). In the light of the type of interests harmed by the damage, the jurisprudence then identified the damage from disservice and damage to the image corresponding to the subclasses "Disservice" and "Image". It must be emphasized that the latter classes, of jurisprudential derivation, are more subject to modifications and interpretation. The boolean *hasMultipleEnumeration* expresses the fourth criterion: the possibility that the compensation is single or cumulative or that the amount is representative of a single compensation item or is the sum of several compensation items. By way of example, the provision could indicate the amount of 100 as total compensation in which 30 is justified by the compensation for the damage to the image and 70 for the damage from the disservice.

5. Tagging Patterns in AKN Documents

In AKN context, tagging patterns evolve two steps: (1) to extract patterns of interest (PoI) from textual data using regular expressions, and (2) identify the values of the PoI attributes. In (2), attributes are identified through data mining using semantic contexts of PoI as input to ML approaches. Figure 1 shows the possible values of the PoI attributes organized into a taxonomic tree.

5.1. Identifying Patterns

In this work, PoI identifies monetary quantities (<quantity>) cited in a legal sentence, being defined as

```
{"?:valute": ["(?:£||Euro|euro|EURO)  (?:\\.)?" ],
 "?:value" : ["\\d+\\.?"],
 "quantity": ["{{valute}}(?:\\s*)?{{value}}
             +((?:\\,\\d+))?"],
 "patterns": ["{{quantity}}"]}
```

The <quantity> tag is expressed by a regular expression using JavaScript Object Notation (JSON). The monetary values can be found in euro (€, 'euro', 'EURO', or 'Euro') or Italian lira (₤). It can be followed or not by a dot ('.'), and by an undefined sequence of spaces ($? : \\s*$) before achieving the numerical value. In sequence, there is at least one numeric digit ($\\d+$), followed or not by a comma (','), and at least another numerical digit ($? : \\, \\d+$). The $?$: operator is a group constructor, and a symbol followed by an interrogation mark (?) informs the possibility of reading the symbol, but not the obligation. Some <quantity> examples are €$20.561, 83$, ₤$.82.967, 92$, and euro $888, 10$. In the tagged text, e.g. €$215, 10$ will appear as

```
<quantity class="Difesa"> 215,10</quantity>,
```

in which 'class' is the attribute of <quantity> that identify the money origin. The attribute 'class' can be associated with any of the possible classes.

5.2. Taxonomic Tree

The taxonomic tree (TT) provides an abstraction control mechanism to classify PoI using its semantic context. Figure 1-a shows the <quantity> taxonomy in a graph, in which the nodes (legal concepts) are linked with different relationships. To make possible intermediary classification systems, the taxonomic graph was mapped to a taxonomic tree (Fig. 2-b) in which the parent-child nodes are linked using class-subclass relationship. With this modelling, an element of a child node is also an element of its parent, except for when the relationship is not different. Suppose a set of concepts organised using the abstract data type tree to define the PoI taxonomy, in which the closer a concept is to the root, the more abstract its definition, and the closer to the leaves, the more concrete its meaning. In this sense, the root is the most generalized TT class, and a leaf is the most specific one of its branches. Each TT concept defines a possible class of the PoI taxonomy. A data sample classified as belonging to the root class means that the sample is poorer in semantic context information than a sample classified as belonging to a leaf class. A tree branch is defined by the way between the root and a leaf, in which the abstraction level of the classes decreases on the way. Data samples are mapped into the set of possible classes $C = \{c_1, c_2, \ldots, c_i, \ldots, c_n\}$, in which c_i is a TT node (and a legal concept) and $n = 29$ in this work. As we can see, there are two special kinds of nodes: the root, and the leaves. The root, the unique node characterised by has no parent; the leaves, the nodes that have no children. TT is composed by a set of legal concepts associated with the <quantity> PoI. Each tree level m represents an abstraction level in which a sample can be classified, having a set of possible classes C_m, in which $m = 0, \ldots, L$, and L is the maximum level of the tree (in this work, $L = 5$). C_m is a subset of C and defines the possible classes of a classification system m associated to the tree level m. For example, C_2, the second level classification system, has the {'JudiciaryAct', 'StateTreasurySum', StateTreasurySumNonLitigation, 'StateTreasurySumLitigation', 'Expense', 'DefenseFee', 'CourtFee'} possible classes, as we can see in Fig. 1. For this example, each sample classified using the C_2 classification system will be associated with one class of C_2. In this work, we have 6 classification systems, $\{C_0, C_1, C_2, C_3, C_4, C_5\}$ with 1, 3, 7, 15, 27, and 29 possible classes respectively. Each legal concept can be seen as a discrete sequence of semantic abstraction classes, from it until the root, e.g. a data sample classified as 'Image' is also classified as 'NonEquity', 'Compensation-Damage', 'StateTreasurySumLitigation', 'StateTreasurySum', and 'JudiciaryAct' since the 'Image' class is contained in the 'NonEquity' class that is contained in the 'Compensation-Damage' class, and so on. The multi-classification system is modelled to be a legal support tool. A <quantity> sample is classified in the most specific-concept classification system (C_5), orbiting in all the legal concepts defined by its branch, the path formed between the root and its most specific classification, that can be a leaf but not necessarily. The hierarchical concepts provide an abstraction scheme that will be used by a decision-making support tool for judges and legal analysts, helping them to analyse the <quantity> origins, enabling them to identify types of money flows. In this sense, if someone needs to understand a legal process in a more general view, the abstraction level of the information can be increased using a more general schema moving towards C_0, and if it is necessary to be more specific, the abstraction level can be decreased, floating towards C_5.

6. Online Context-driven Neural Networks

Context-driven Neural Networks is a fully-supervised one-layer Neural Network (NN) constructed incrementally from an online data stream [21]. Its processing units are neurons, used to implement a NLP classification of PoI extracted from textual data. The network architecture results from an incremental construction from the scratch. Consider that the data stream $(\mathbf{x}, y)^{[h]}$, $h = 1, ...$ Inputs $\mathbf{x} = [x_1, x_2, \ldots, x_j, \ldots, x_k]$ and output y are a word vector with variable size and a class. To simplify the notation, let's assume that the word vectors have the same length k. Figure 3 shows the NN architecture in which the input layer receives $\mathbf{x}^{[h]}$. The neural layer is a set of dictionaries $T = \{T^1, \ldots, T^i, \ldots, T^c\}$. Each T^i is a dictionary of term frequencies, in which the frequency calculus varies according to the used aggregation functions A^i, being the same for all neurons in a given NN. Each $(\mathbf{x})^{[h]}$ is a vector in which each element x_j is a content word extracted from the neighborhood of the associated <quantity> sample. Content words are word classes that have semantic meaning, i.e., nouns, adjectives, adverbs, and lexical verbs. In this work, we focused on the semantic context formed by nouns.

In Fig. 4, the neuron T^i receives $\mathbf{x}$ and compares each x_j with the terms of the dictionary T^i until it finds a coincidence. If x_j is in T^i, w_j^i is used in the aggregation function A^i. The comparison between x_j and the T^i terms is crisp, therefore the result can be "0" if there is no match, and "1" if x_j is in T^i. The comparison output is weighted by the frequency value w_i^j in A^i, generating the o^i aggregated output.

The o^i output is the expression of the membership level of $\mathbf{x}$ in T^i. A^f is an aggregation function that choose the best T^i candidate, o^{i*}, to be the $\mathbf{x}$ class, $\hat{C}^{i*}$. A^f is $S([o^1, \ldots, o^c])$, in which S is the max operator.

6.1. Aggregation Functions

The aggregation functions A^i are the same for a NN setting with the exception of A^f that is the S-norm in all configurations.

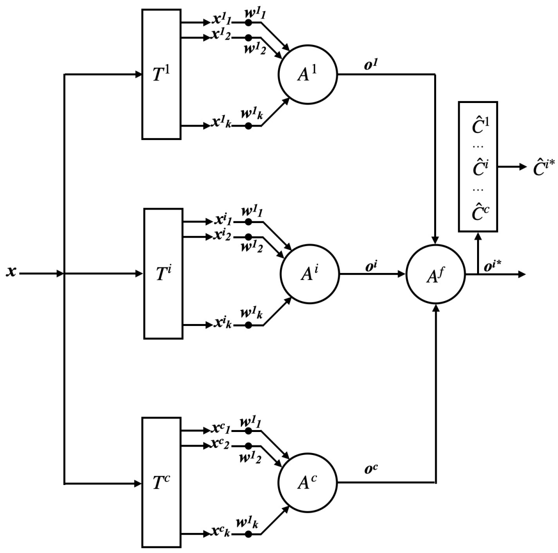


Figure 3: Online Context-driven Neural network to the classification of PoI extracted from XML documents.

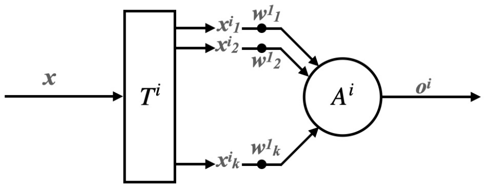


Figure 4: Neuron Model

6.1.1. Term Frequency–Inverse Document Frequency

The naive version of $TFIDF$ metric is presented [22]. The approach is based on the calculus of TF and IDF in a collection of documents. In this work, a document is a T^i dictionary, and its collection $T = \{T^1, \ldots, T^c\}$, in which each T^i is the dictionary of the class i.

The TF ranks the item j based on its importance $TF_{i,j}$ on the document i, i.e., the importance of x_j in T^i

$$TF_{i,j} = \frac{n_{i,j}}{N_i} \tag{1}$$

being $n_{i,j}$ the number of occurrences of the term j in the document i; and N_i the total number of items in i. The IDF counts the number of relevant documents in the collection related to the term j,

$$IDF_{j,T} = log\left[\frac{c}{1+|i \in T : j \in i|}\right] \tag{2}$$

in which T is the collection of documents with $|T| = c$, the number of known classes, and $1 + |j \in T : j \in i|$ indicates the number of documents in which the term j appears at least once. $TFIDF$ is the result of the multiplication of the terms. The higher the $TF_{i,j}$ and the $IDF_{i,D}$, the higher the importance of the term j in T, i.e., the term j must be frequent in a small number of documents i in T to be ranked as important [23].

In this work, it is used 3 variations of the presented IDF formula

$$IDF_{j,T} = log\left[\frac{c}{|i \in T : j \in i|}\right], \tag{3}$$

$$IDF_{j,T} = log\left[\frac{c}{|i \in T : j \in i|} + 1\right], \tag{4}$$

and

$$IDF_{j,T} = log\left[\frac{c}{1+|i \in T : j \in i|} + 1\right], \tag{5}$$

being referenced as tfidf2, tfidf3, and tfidf4 in Fig. 7- 6. Each word j in T^i has an associated TFIDF value that is used as weight w^i_j in A^i, in which o^i is the average value of w^i_j of x_j in T^i.

6.1.2. P-Norm

P-norm is defined as

$$||\mathbf{w}||_p = \left(\sum_{j=0}^{k} |w_j|^p\right)^{\frac{1}{p}} \tag{6}$$

in which $p = \{1, 2, 3, 4\}$ in this work, being referenced as 1-norm, 2-norm, 3-norm, and 4-norm in Fig. 7- 6. Each word j in T^i has an associated frequency as $TF_{i,j}$ that is used as weight w^i_j in A^i= P-Norm.

6.1.3. S-norm

S-norms (S) are commutative, associative and monotone operators. The used S-norm is the maximum operator,

$$S_{max}(\mathbf{w}) = \max_{j=0,\dots,k} w_j, \tag{7}$$

being referenced as max in Fig. 7- 6. Each word j in T^i has an associated frequency as $TF_{i,j}$ that is used as weight w^i_j in A^i = S-norm.

6.1.4. Online Context-driven Neural Network Algorithm

In lines 1-3, the accuracy vector acc is initialized, the level tree m and the aggregation function f are read, and the neural network nn is created. In lines 4-13, there is the $x^{[h]}$ stream classification: in the line 9, nn receives $x^{[h]}$ and returns the estimated class $\hat{C}^*$. The nn is updated with the new information $(\mathbf{x}, y)^{[h]}$ in the line 10. The online accuracy acc is calculated in line 11 according to Eq. (8).

Online Learning: Context-Driven NN Classifier

1: //Initialization
2: $acc = [0]$; Read m, f;
3: nn = createNN(f, m) ;
4: **For** $h = 1, \ldots$
5:     Read $\mathbf{x}^{[h]}$
6:     **If** $h = 1$ **then**
7:         initialize(nn, x_1);
8:     **Else** //$h = 2, \ldots$
9:         $\hat{C}^*$ = nn($x^{[h]}$)
10:         update_NN(nn, $x^{[h]}$, $y^{[h]}$)
11:         $acc(new)$= update_Acc($acc(old)$, $\hat{C}^*$, $y^{[h]}$); Eq.(8)
12:     **end if**
13: **end for**

7. Dataset Analysis and Metrics

This section describes the data gathering process, the annotation, the datasets and the metrics used to report results.

7.1. About the dataset

The dataset is composed of information extracted from legal rulings of the Italian Court of Audit from 2000-2015 of the Emilia-Romagna region with 584 samples in total, 327 judgments. Rulings are presented as Akoma Ntoso-formatted documents, from which are extracted monetary values from the textual data of their 'Decision' parts. The dataset is organized into the columns <filename>, <quantity>, <Decision_Part>, <Semantic_Context>, and <Class_Expert>, in which:

- <filename>: the name of the file of the legal sentence,
- <quantity>: the extracted monetary quantity,
- <Decision_Part>: the text from which the <quantity> sample was extracted,
- <Semantic_Context>: a list of words, a maximum of 20 words apart from <quantity>, extracted from the <Decision_Part> text, characterizing semantically <quantity>, and
- <Class_Expert>: one of the classes described in the taxonomy that is associated with <quantity>.

Considering the $(\mathbf{x}, y)^{[h]}$ stream, <Semantic_Context> is $\mathbf{x}$ and <Class_Expert> is y.

7.2. Annotation

One legal expert annotated the 584 samples by identifying the correct class at the lowest possible level of abstraction, i.e., by preferring a more specific class over a broader one, when possible. Such possibility is offered solely by the context, as detailed by the decision. Following the methodology used in other works, e.g., [24, 25], the annotator did not follow criteria (e.g., case law, interpretation by scholars) when labelling the instances other than a) the textual wording of the decision, and b) the description of the classes.

7.3. Dataset Analysis

The data distribution is shown in Fig. 5-a. The y-axes shows the class name and the x-axes, the number of samples. The partial data distribution (PDD) considers the amount of samples of each class and, on the other hand, the total data distribution (TDD) takes into account PDD added to the recursive sum of all its children samples. Figure 5-b shows the distribution of the quantity of samples through the set of classes, informing how many classes have the amount of samples into a given range. For example, it is considered ranges with widths of 10 samples, i.e., from 0 to 10, from 11 to 20, and so on. As it is possible to see, the problem is an unbalancing classification in which 18 from 29 classes have until 10 samples. In Fig. 5-c, it is shown how the size of the semantic-context list is distributed through the dataset, approaching roughly a Gaussian distribution.

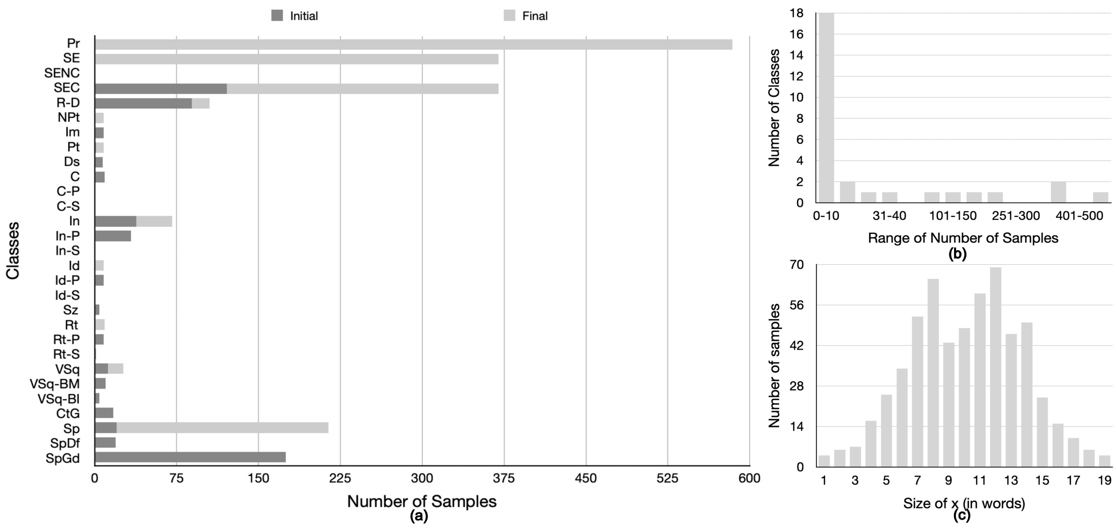


Figure 5: In (a), it is shown the sample distribution by classes. PDD considers just the samples classified by the class itself. TDD includes PDD added to the recursive sum of all its descendent samples. In (b), the samples are presented as range, in which 0-10 means that the amount of samples can varies from 0 to 10 samples maximum and so on, showing how many classes are in each range. In (c), there is the distribution of the samples by the size of the input **x** in words.

The data distribution makes it difficult to classify data with high accuracy because it is available a highly unbalanced small data. To improve the classification quality, it is proposed a multi-classification scheme in which each tree level represents a possible classification problem. Based on the fact that an element of a subclass is also an element of its super class, the classification system is composed by 6 classification problems with 1, 3, 7, 15, 27 and 29-classes, in which it is available 584 samples distributed over the possible classes in each one of the problems.

7.4. Metrics

Online classification accuracy, $Acc \in [0, 1]$, is obtained from

$$Acc(new) = \frac{h-1}{h}\,Acc(old) + \frac{1}{h}\,\tau, \tag{8}$$

in which $\tau = 1$ if $\hat{C}^{[h]} = C^{[h]}$ (right estimation); $\hat{C}^{[h]}$ and $C^{[h]}$ are the estimate and actual classes, i.e., the automatic classification and the expert classification respectively. Otherwise, $\tau = 0$ (wrong class estimation) [26] [23]. The execution time is given in seconds using a MacBook Air(M1, 2020), macOS Version 12.6, chip Apple M1, 16GB of memory.

8. Experimental Results

8.1. Technical Evaluation

We evaluated the Context-driven NN considering 584 <quantity> samples. We considered 9 different aggregation functions, 2 algorithm versions, and 6 classification systems, i.e., 108 combinations of hyperparameters. Each combination was repeated 20 times, being executed 2160 simulations in total. Table 1 shows the values of the hyperparameters used in the simulations. In the aggregation functions, f, tfidf and p-norm have 4 variations each one.

Table 1
Hyperparameters in evaluation

Initial Setting	
f	$\{tfidf(4), max, p-norm(4)\}$
Algo. version	$\{online, offline\}$
Class. System	$\{1, 3, 7, 15, 27, 29\}$

As a side effect of the Pigeonhole principle, increasing the number of possible classes, decreases the method accuracy as shown in Fig. 6-(a) and Fig. 7-(a,b). The probability $Pr(c)$ to choose the correct class by chance is $\frac{1}{c}$, in which c is the number of possible classes, and for the 1, 3, 7, 15, 27, and 29-classes problems, the probability will be 1, 0.33, 0.14, 0.06, 0.037, and 0.034 respectively, and their $\frac{Acc}{Pr(c)}$ rate is in the order of 1, 2.51, 5.57, 10.67, 14.95, and 16.51. As we can see, when increasing the number of possible classes, the impact of chance on the classification becomes smaller. Since the taxonomy is structured as classes and sub-classes, the probability of misclassification among the classes of the same branch is bigger than between different branches. As we have data extracted from free-style text, the juridical terms used to describe a <quantity> sample can vary greatly, contaminating **x** with words with no semantic importance to identify the correct class or with words that are useful to identify other possible classes. Table 2 abstracts the results of the simulations considering our best performance using tfidf3 with 99.5% of confidence. The accuracy of the offline algorithm version (94.18%, 61.53%) is slightly higher than the online one (89.87%, 57.02%). It is easy to understand the results if we see Fig. 6-b which describes the behaviour of the accuracy on-the-fly for the online NN version. In the beginning of the learning process, the accuracy is unstable, tending to converge to the method accuracy over time. The offline NN has a training phase which includes the beginning of the learning process. In this way, it is expected that the testing phase has better accuracy than the training phase in this case.

Table 2
Performance of Algorithms with 99.5% of confidence considering TFIDF3 Aggr. Function.

Algorithm	Class	$Acc(\%)$	Time (s)
Online	3	89.87 ± 1.04	0.49 ± 0.01
Offline	3	94.18 ± 1.31	0.45 ± 0.01
Online	29	57.02 ± 2.11	0.49 ± 0.01
Offline	29	61.53 ± 2.79	0.44 ± 0.01

If we compare Fig. 6-a and Fig. 7-(a,b), it is possible to see that the confidence interval, expressed by the grey area around the lines, is slightly narrower in Fig. 6-a than in Fig. 7-(a,b) since the online version uses the samples used by the training and testing phases together. Figure 6-b shows the online accuracy of 3-classes and 29-classes problems, providing the asymptotic limits to the accuracy of all classification systems. Figures 6-7 present a similar behaviour of the accuracy of the classification systems, being unquestionable the superiority of the results related to the TFIDF aggregation functions family, the unique approach exclusively NLP.

8.2. Legal Evaluation

Following the methodology, we analysed the most frequent errors from a legal point to check whether the taxonomy was tailored to the text under scrutiny or whether adaptations were needed. The analysis was carried out on each experiment with similar results. For reasons of brevity, the following detailed assessment only concerns the experiment with the lowest accuracy. Fig. 8 shows the 7 most frequent errors for the 29-class experiment and their relative number of misclassification, which account for more than 50% of the total number of errors. In the most frequent error, the algorithm classifies "Expense" instead of "CourtFee". Therefore, it is unable to go from a parent class to a child class. The reasons underlying this error may be linked to the absence of clear linguistic formulae that identify the costs in Court in their general meaning with respect to the costs referred to in the decision. Reading the misclassified cases, two problems are detected: 1) quantities and judicial expenses are mentioned consecutively in the decision, thus the linguistic formulae influence the classification negatively; 2) the formula

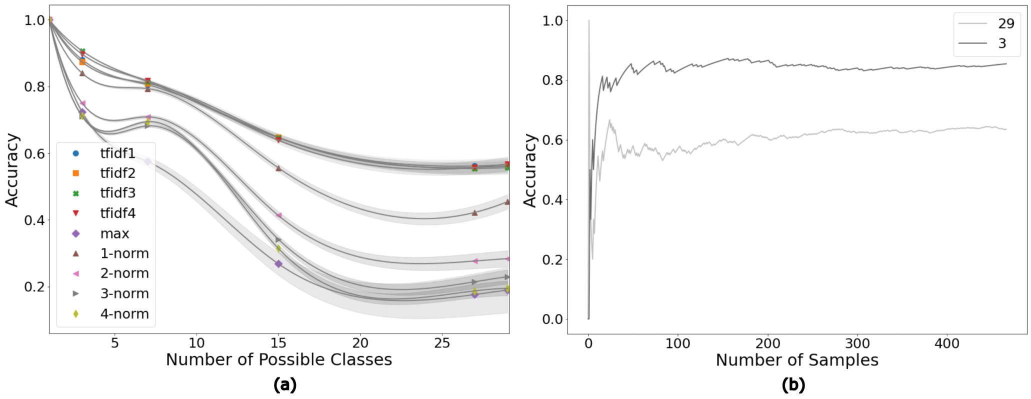


Figure 6: In (a), it is shown the accuracy of the 6 possible classification systems of the online NN classifier. In (b), the convergence of the online accuracy of the 3-classes and 29-classes systems.

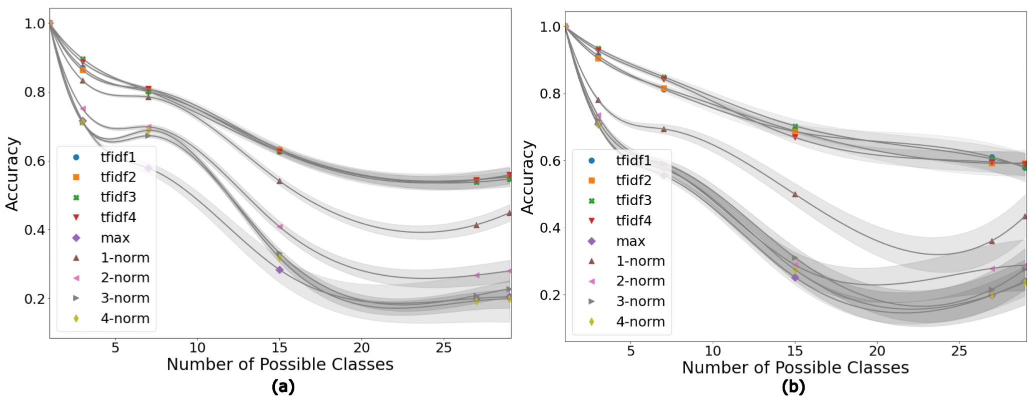


Figure 7: In (a, b), it is shown the accuracy of the 6 possible classification systems of offline NN classifier. In (a), the accuracy of the training phase; in (d), the accuracy of its testing phase.

"spese compensate" ("compensated expenses") significantly affects the classifier. The taxonomy could be refined - in the future - by adding a boolean specification on the compensated expenses. With regards to the second and third most common errors, it can be observed that the classifiers chose a "child" ("Equity") node rather than a "parent". One possible explanation of this recurrent mistake is that the classifiers associate some linguistic patterns to the "child" class that are hidden from the annotator, who opted for a more abstract term. However, it has to be noted that misclassifications occur on subclasses that have some degree of connection to the right class. The same kind of misclassification can be observed for the fourth, fifth and seventh most frequent cases, but these are more likely to be justified by the lack of specific data. The sixth case is similar to the first one, and can be explained by the presence of the word "pensionistico" ("about retirement/pension") in the sentence, which might trigger the wrong case. However, the context did not allow the annotator to define the quantity as related to the pensions.

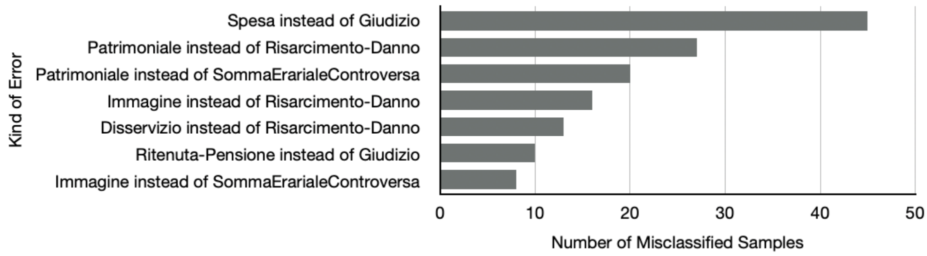


Figure 8: The histogram shows the 7 most frequent errors for the 29-class experiment and their relative number of misclassification. Taken together, these account for more than 50% of the total number of errors.

9. Final remarks

In this study, Online Context-Driven Neural Networks were tested in classifying quantities from decisions of the Audit Court, also in comparison with offline algorithms, in combination with TF-IDF aggregation functions. From the legal evaluation of the experiment, it emerges that the class "CourtFee" is particularly problematic due to its frequency (5th most represented class overall, 1st as initial class) and the lack of a clearly-identifiable linguistic pattern to be associated with this class. This is a rather expected result given the correlation between these two factors. In further studies, a refinement of the ontology could identify some implicit instances of this class. Similarly, expanding the taxonomy with regard to other forms of judicial expenses, including the compensated ones, could be helpful in increasing the mapping of the quantities to be labelled. Finally, a cross-validation that includes other regions and other courts could be a viable and valuable avenue for further research. In any case the proposed technique is quite promising for classifying the amount using a taxonomy and so to improve the categorization of the different decisions also on the basis of the role plaid by the quantities in relation of the type of damage. In particular it is possible to make complex query like "give me all the decisions where the quantity of the damage is > X, concerning 'Disservice' AND where the CourtFee is >Y". Additionally the Online Context-Driven Neural Networks permits to adequate the ML model dynamically according with the evolving of the language of the decisions in relation of the jurisdiction (e.g., different regional language) or due to the historical period (e.g., different terminology) and the legislative changes.

Acknowledgements

This work was supported by the ER4JUSTICE - NEXT GENERATION initiative (CUP E44I20000270002) for the efficiency of the judicial system, funded by Regione Emilia-Romagna and coordinated by Fondazione CRUI.

References

[1] E. COMMISSION, 2022 Rule of Law Report, COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS 2022 Rule of Law Report The rule of law situation in the European Union, Technical Report, EUROPEAN COMMISSION, 2022. (See also https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52022DC0500).

[2] K. D. Ashley, Artificial intelligence and legal analytics: new tools for law practice in the digital age, Cambridge University Press, 2017.

[3] G. Boella, L. Di Caro, A. Ruggeri, L. Robaldo, Learning from syntax generalizations for automatic semantic annotation, Journal of Intel. Info. Sys. 43 (2014) 231–246.

[4] N. Kiyavitskaya, N. Zeni, L. Mich, J. R. Cordy, J. Mylopoulos, Text mining through semi automatic semantic annotation, in: U. Reimer, D. Karagiannis (Eds.), Practical Aspects of Knowledge Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 143–154.

[5] E. Oren, K. Möller, S. Scerri, S. Handschuh, M. Sintek, What are semantic annotations, 2006.

[6] N. Zeni, L. Mich, J. Mylopoulos, Gaiust 2.0: Evolution of a framework for annotating legal documents, in: Metadata and Semantic Research, MTSR 2016, volume 672 of *Communications in Computer and Information Science*, 2016, pp. 43–54. 10th International Conference on Metadata and Semantics Research (MTSR), Gottingen, Germany, nov 22-25, 2016.

[7] M. Palmirani, G. Governatori, A. Rotolo, S. Tabet, H. Boley, A. Paschke, LegalRuleML: XML-Based Rules and Norms., RuleML America 7018 (2011) 298–312.

[8] M. Palmirani, D. Liga, Derogations analysis of european legislation through hybrid ai approach, in: Electronic Government and the Information Systems Perspective: 11th International Conference, EGOVIS 2022, Vienna, Austria, August 22–24, 2022, Proceedings, Springer, 2022, pp. 123–137.

[9] A. Esuli, D. Marcheggiani, F. Sebastiani, An enhanced crfs-based system for information extraction from radiology reports, Journal of Biomedical Informatics 46 (2013) 425–435.

[10] N. Nabizadeh, D. Kolossa, M. Heckmann, Myfixit: An annotated dataset, annotation tool, and baseline methods for information extraction from repair manuals, in: Proceedings of the 12th Int. Conf. on Language Resources and Evaluation (LREC 2020), 2020, pp. 2120–2128. Marseille, France, may 11-16, 2020.

[11] M. A. Tocoglu, O. Ozturkmenoglu, A. Alpkocak, Emotion analysis from turkish tweets using deep neural networks, IEEE Access 7 (2019) 183061–183069.

[12] N. Zeni, N. Kiyavitskaya, L. Mich, J. R. Cordy, J. Mylopoulos, Gaiust: supporting the extraction of rights and obligations for regulatory compliance, Requirements Engineering 20 (2015) 1–22.

[13] N. Zeni, L. Mich, J. Mylopoulos, J. R. Cordy, Applying gaiust for extracting requirements from legal documents, in: 2013 6th Int. Workshop on Requirements Engineering and Law (RELAW), 2013, pp. 65–68. Rio de Janeiro, Brazil, jun 16, 2013.

[14] N. Zeni, N. Kiyavitskaya, J. R. Cordy, L. Mich, J. Mylopoulos, Annotating regulations using cerno: an application to italian documents extended abstract, in: ARES 2008: Proceedings

of the 3rd Int. Conf. on Availability, Security and Reliability, 2008, pp. 1437+. Barcelona, Spain, mar 04-07, 2008.

[15] R. Bartolini, A. Lenci, S. Montemagni, V. Pirrelli, C. Soria, Automatic classification and analysis of provisions in italian legal texts: A case study, in: R. Meersman, Z. Tari, A. Corsaro, P. Herrero, M. Perez, M. Radenkovic, V. Robles, C. Santoro, A. Albani, K. Turowski, M. Jarrar, A. Gangemi, E. Duval, P. Spyns, A. Palinginis (Eds.), On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops, Proceedings, volume 3292 of *Lecture Notes in Computer Science*, 2004, pp. 593–604. On the Move Confederated International Workshop and Conference, Agia Napa, Cyprus, oct 25-29, 2004.

[16] F. Amardeilh, D. Bourcier, H. Cherfi, C. H. Dubail, A. Garnier, S. Guillemin-Lanne, N. Mimouni, A. Nazarenko, E. Paul, S. Salotti, M. Seizou, S. Szulman, H. Zargayouna, The legilocal project: the local law simply shared, in: Legal Knowledge and Information Systems, volume 259 of *Frontiers in Artificial Intelligence and Applications*, 2013, pp. 11–14. 26th Int. Conf. on Legal Knowledge and Information Systems (Jurix), Bologna, ITALY, dec 11-13, 2013.

[17] N. T. Son, H. B. Quoc, N. T. P. Duyen, N. L. Minh, Recognizing logical parts in vietnamese legal texts using conditional random fields, in: B. Thi, A. Drogoul, P. Kuonen, C. Tran, A. Tran, N. ThaiNghe (Eds.), 2015 IEEE RIVF Int. Conf. on Computing & Communication Technologies, Innovation, and Vision for the Future (RIVF), 2015, pp. 1–6. Can Tho, Vietnam, jan 25-28, 2015.

[18] M. Palmirani, Legislative XML: Principles and Technical Tools, ERMES Servizi Editoriali Integrati S.r.l., 2012, pp. 53–80.

[19] F. Giubaldo, Uno strumento di supporto all'analisi e visualizzazione di dati strutturati nel contesto giuridico, Master's thesis, University of Bologna, 2018.

[20] M. Ceci, M. Palmirani, Ontology framework for judgment modelling, in: AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents: International Workshop AICOL-III, Held as Part of the 25th IVR Congress, Frankfurt am Main, Germany, August 15-16, 2011. Revised Selected Papers 3, Springer, 2012, pp. 116–130.

[21] L. Decker, D. Leite, F. Viola, D. Bonacorsi, Comparison of evolving granular classifiers applied to anomaly detection for predictive maintenance in computing centers, in: IEEE Conf. on Evolving and Adaptive Intel. Sys. (EAIS), IEEE, Bari, Italy, 2020, p. 8p.

[22] F. Minarini, L. Decker, Time-series anomaly detection applied to log-based diagnostic system using unsupervised machine learning approach, in: Proc. of the 27th Conf. of Open Innovations Association (FRUCT), FRUCT Oy, Trento, Italy, 2020, pp. 343–348. doi:10.5281/zenodo.4026500.

[23] L. Decker, D. Leite, F. Minarini, S. R. Tisbeni, D. Bonacorsi, Unsupervised learning and online anomaly detection: An on-condition log-based maintenance system, Int. Journal of Embedded and Real-Time Comm. Syst. 13 (2022) 1–16.

[24] F. Sovrano, M. Palmirani, B. Distefano, S. Sapienza, F. Vitali, A dataset for evaluating legal question answering on private international law, in: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law, 2021, pp. 230–234.

[25] M. Palmirani, G. Bincoletto, V. Leone, S. Sapienza, F. Sovrano, Pronto ontology refinement through open knowledge extraction, in: Legal Knowledge and Information Systems, IOS

Press, 2019, pp. 205–210.
[26] L. Decker, D. Leite, D. Bonacorsi, Explainable log parsing and online interval granular classification from streams of words, in: 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, Padua, Italy, 2022, pp. 1–8. doi:10.1109/FUZZ-IEEE55066.2022.9882710.