



A practical application of Artificial Intelligence techniques for legal context analysis

Goal

Legal context analysis: framing a legal document to **explore its semantic**, applying the techniques of domain experts which we can reproduce automatically using lightweight tools.

Looking for a **deeper meaning** allows us to tackle some important and interesting tasks in the legal context

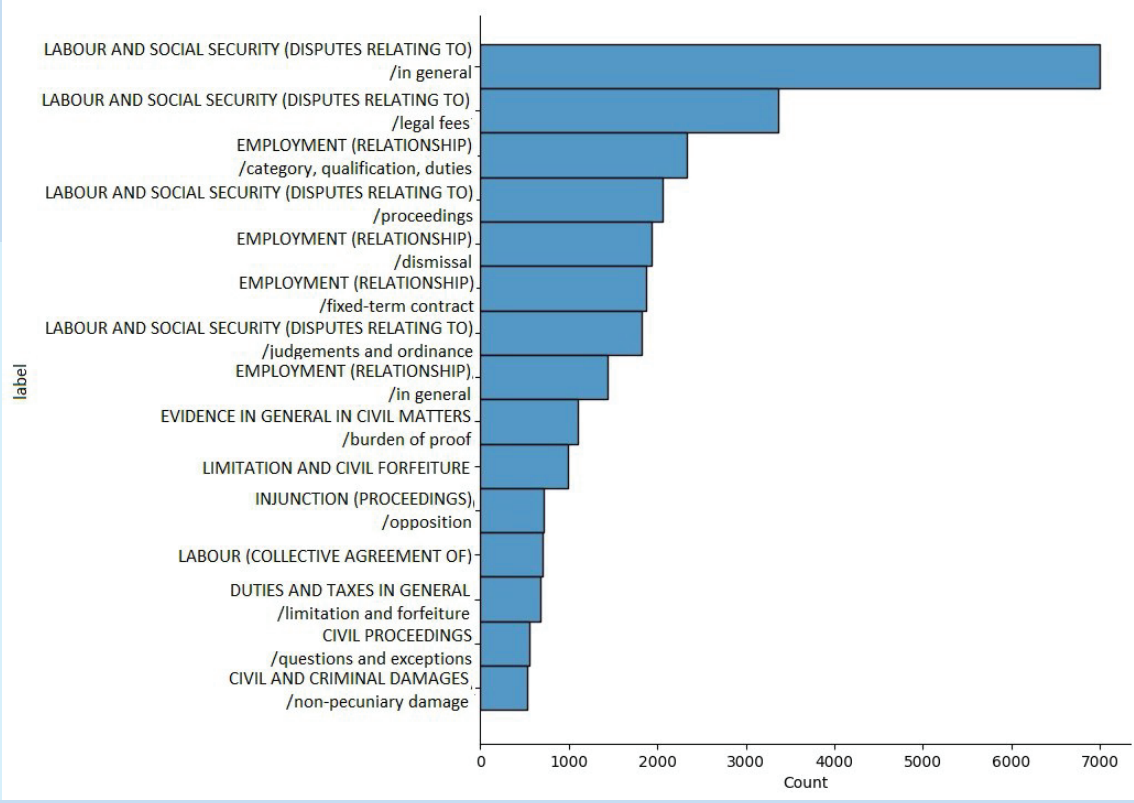


highly unbalanced data different structures

Leggi di Italia corpus (300 labels)

21,562 judgements from Leggi d'Italia web archive

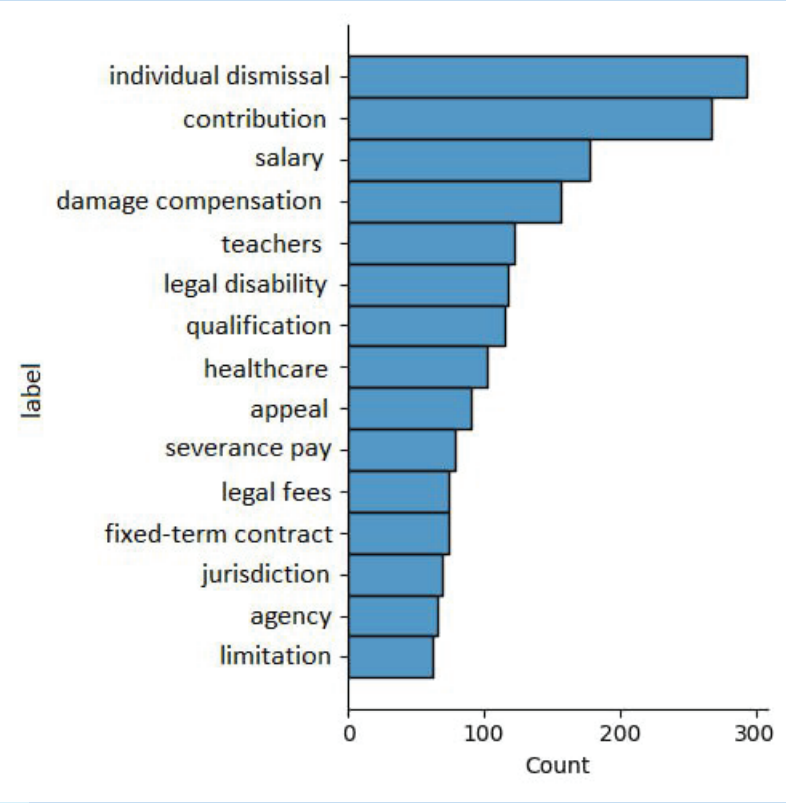
Taxonomic relationship



Turin corpus (309 labels)
27,477 judgments

from the Court of Turin (Labor)
only 4804 are labelled

Linear structure



Augmented reading for understanding juridical context



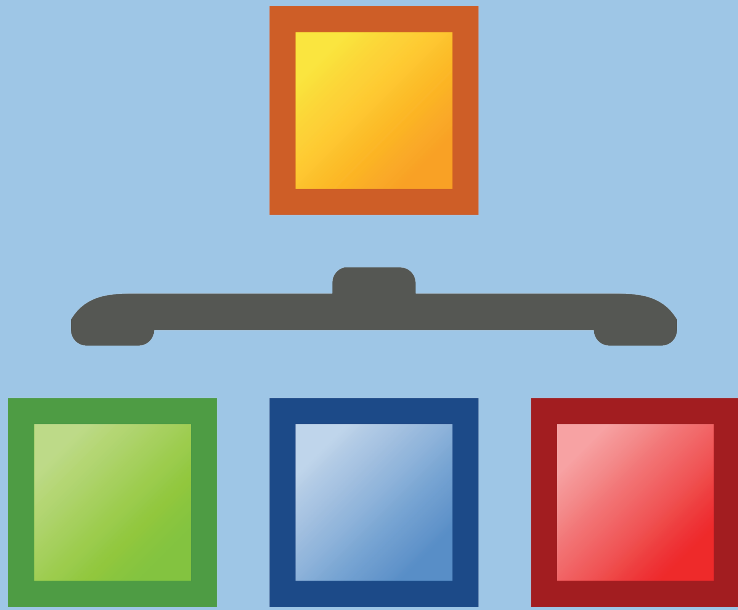
The understanding of a legal document, passes through the extraction and the processing of **semantically relevant information**.

First and second instance judgments follow a **semi-structured schema**. The document is divided into core parts (**header, facts, reasoning, and decisions**).

Taxonomy transfer and classification

There are many benefits to aligning the labels of judgments from two or more courts:

- It is **possible to define a unique and shared hierarchy of labels for each court**
- Improve the quality of the automatic classification of the judgments



Principles of law extraction

Principles of law can be defined as **an interpretation of rules**, in Italy, issued by the Court of Cassation, so they do not constitute a source of law per se. In this way the Supreme Court indicates the correct and exact interpretation of the rule to be applied. They indicate the situation to which a certain rule may be applied, filling the gap between real world scenarios and laws.

In our work we used **regex extraction** as a first naive approach



Table 1: Table explaining the results and validation of the Supreme Court citation regex extraction

Supreme Court citations		Metrics and evaluations					
data	retrieved	relevant	intersection	threshold	precision	recall	f1score
unstructured_data	5	10	4	70%	0.8	0.4	0.533
structured_data	8	13	8	100%	1.0	0.615	0.761

Data: from Turin Set, annotated by our legal experts

Methodology : **regex extraction** and different forms of **pre processing**

Results: the usage of **structured data** betters the **performance**

Authors and Acknowledgements

Ilaria Angela Amantea, Chiara Bonfanti, Michele Colombino, Giorgia Iacobellis, Susanna Marta, Rachele Mignone, Marianna Molinari, Ivan Spada, Laurentiu Jr Marius Zaharia

Supervisors: Emilio Sulis, Luigi Di Caro, Guido Boella

This research has been carried out within the project **NEXT GENERATION UPP**