

Mining Information from Legal Sentences in Klondike

Andrea Gatti^{1,*†}, Viviana Mascardi^{1†} and Domenico Pellegrini^{2†}

¹University of Genova, DIBRIS, Italy

²Ministry of Justice, Tribunale di Genova, Italy

Abstract

The “Next Generation UPP” (NGUPP) Italian project aims at devising and implementing new collaborative schemes between universities and judicial offices for improving the efficiency and performance of justice in north-west Italy. The University of Genova was assigned different tasks in NGUPP, including the semantic analysis of sentences and the study of potential bias on a massive scale for legal and analytical purposes. Klondike (for Knowledge Driven Extraction engine), with its ability to determine the type of divorce sentence, extract the master data of the parties and children, extract and visualize statistical information, and anonymize the original document, contributed to achievement of the NGUPP semantic analysis objective.

Keywords

NGUPP, Klondike, NLP4Law, semantic analysis, divorce sentence, frame semantics

1. Introduction and Motivation

The “Next Generation UPP” (NGUPP) Italian project¹, led by the University of Torino, aims at devising and implementing new collaborative schemes between universities and judicial offices for improving the efficiency and performance of justice in north-west Italy. The project, funded under the “PON Governance e Capacità Istituzionale 2014-2020” programme, started in April 2022 and is heading towards its conclusion, in September 2023. The total budget is 8.5 million euro, shared among twelve partners: the universities of Bergamo, Brescia, Genova, Insubria, Milano Bicocca, Milano, Pavia, Piemonte Orientale Amedeo Avogadro, Torino; the polytechnics of Milano and Torino; and the University School for Advanced Studies IUSS Pavia.

The final objective of NGUPP is to allow judicial offices in north-west Italy to adopt a more efficient approach to manage disputes in order to reduce backlog and duration of civil sentences. From a knowledge transfer point of view, NGUPP aims to ensure that all graduates in legal disciplines acquire IT and management skills, and that administrative staff is provided with continuous training.

NL4AI 2023: Seventh Workshop on Natural Language for Artificial Intelligence, November 6-7th, 2023, Rome, Italy [1]

*Corresponding author.

†These authors contributed equally.

✉ andrea.gatti@edu.unige.it (A. Gatti); viviana.mascardi@unige.it (V. Mascardi);

domenico.pellegrini@giustizia.it (D. Pellegrini)

🆔 0009-0003-0992-4058 (A. Gatti); 0000-0002-2261-9926 (V. Mascardi)



© 2023 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)

¹<https://www.nextgenerationupp.unito.it/>, accessed on September 2023.

One of the tasks of the University of Genova in NGUPP was to create a tool for the semantic analysis of sentences and to study potential bias on a massive scale for legal and analytical purposes. KlondikE, for **K**nowledge **D**riven **E**xtraction engine, is the tool based on Natural Language Processing (NLP) that we realized for that aim.

The exploitation of NLP in the legal domain dates back to the seventies of the last century [2, 3, 4, 5], and is still an extremely active research field [6, 7, 8, 9, 10, 11] as also witnessed by funded projects and scientific events.

As an example, the **M**ining and **R**easoning with **L**egal texts project² funded by the European Union's Horizon 2020 research and innovation programme closed at the end of 2019, while the **AI4Lawyers**³, awarded by the European Commission Directorate-General for Justice and Consumers, closed in March 2022 and two out of its five deliverables were related with NLP. The special issue *NLP for legal texts* of the *AI and Law* journal was published in 2019 [12], the fifth edition of the Natural Legal Language Processing Workshop will take place at the end of 2023⁴, and the first Workshop on Machine learning, Law and Society is expected to run in Torino in a few days⁵. It comes with no surprise that the success of applying NLP techniques to the legal domain also raises ethical concerns, as shown by papers [13] and projects⁶.

Following the FrEX approach for extracting semantic frames from PDF files of expropriation sentences [14], also KlondikE is inspired by Frame Semantics [15] and exploits NLP to identify the main actors and their role in divorce cases. Given that the FrameNet portal⁷ provides no frame for the "Divorce" concept, we designed our own frame as follows, assuming that the type Person should at least be characterized by first name, family name, date and place of birth, and gender. The design of the *Divorce_Frame* was driven by magistrates from the Tribunal of Genova and takes their real needs into account. For example, knowing if there are underage children and if domestic violence was reported is fundamental to properly approach the case. As a side effect of being able to extract the values for filling the *Divorce_Frame*, KlondikE can also identify those values in the document for anonymization purposes and collect them for statistical analysis.

Divorce_Frame

- Core Entities
 - Husband (type Person)
 - Wife (type Person)
 - Children (type Person)
 - Underage children (type boolean)
 - Domestic Violence (type boolean)
 - Involvement of Social Services (type boolean)

²<https://www.mirelproject.eu/index.html>, accessed on September 2023.

³<https://ai4lawyers.eu/>, accessed on September 2023.

⁴<https://nllpw.org/workshop/>, accessed on September 2023.

⁵<https://sites.google.com/view/ws-ethics-ecml23/>, accessed on September 2023.

⁶<https://www.matthes.in.tum.de/pages/ztm206067g3q/NLawP-Natural-Language-Processing-and-Legal-Tech>, accessed on September 2023.

⁷FrameNet is a lexical database containing over 1,200 semantic frames, 13,000 lexical units, more than 200,000 manually annotated sentences, <http://berkeleyfn.framenetbr.ufjf.br/>, accessed on September, 2023.

- Share of Extraordinary Expenses (type integer in 0-100)
- Non-Core Entities
 - Lawyer (type Person)
 - Judge (type Person)

The software architecture of Klondike is general enough to make the tool suitable for processing any kind of document once properly customized, and indeed some of its modules have been designed and implemented ad hoc for dealing with the divorce domain, with sentences written in Italian.

Although the adoption of generative AI approaches (GPT-3.5, GPT-4) on divorce sentences gave promising results, with the GPT-4 APIs performing as well as Klondike, GPT-3.5 Turbo and GPT-4 APIs are not open source and data are processed on the OpenAI servers. The guidelines for acquiring and reusing software in Italian public administrations, delivered in May 2019, state that software developed for public administrations must be, by default, open source⁸. While, according to the most permissive Open Source licenses, it is legal to use closed source libraries in an open source project, this raises many practical and economic implications which suggest to avoid this practice. Also, although the transmission of data over the network for being processed at the OpenAI servers is guaranteed to be secure and compliant with privacy requirements⁹, computers in tribunals have many limitations on their access to the Internet.

For the two reasons above Klondike exploits technologies that are fully open source and do not require any access to the network. This gives Klondike the full control over analyzed data, in compliance with current general regulations in terms of data use and storage and with the more restrictive regulations holding for judicial offices. Being open source also allows to properly address ethical concerns and its modular and clean architecture make Klondike easily inspectable, reusable, and extendable by third parties.

The paper is organized in the following way: Section 2 illustrates the Klondike architecture and implementation, experiments are presented in Section 3, and Section 4 concludes.

2. Architecture and Implementation

Given a PDF file, Klondike processes it to reach several objectives, namely:

1. Determine the type of sentence;
2. Extract the master data of the parties and any children;
3. Extract statistical information;
4. Anonymize the document;
5. Produce statistical graphs on the set of sentences considered.

To do this, Klondike features a modular architecture, with one module for each of the presented items as shown in Figure 1. It is implemented as a Python library that counts approximately 900 lines of code.

⁸<https://www.agid.gov.it/it/design-servizi/riuso-open-source/linee-guida-acquisizione-riuso-software-pa>, accessed on September 2023.

⁹<https://openai.com/enterprise-privacy>, accessed on September 2023.

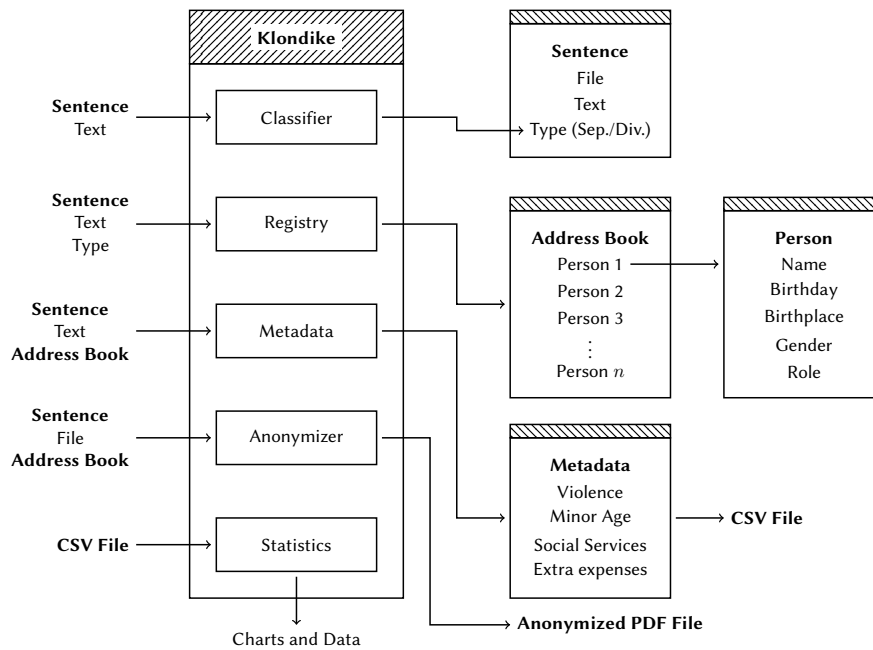


Figure 1: Design of the Klondike Object.

Classifier. The *classifier* module performs a search within the document and classifies the sentence according to its type, namely *separation* or *divorce*. To do this, it takes the text of the sentence as input and performs a simple search in its initial part, looking for one of the predefined types. The type names are fixed, so no further checking is necessary.

At the implementation level, Klondike creates a Python *sentence* object, defined for the project, for each sentence. This module adds the *type* information to this object.

Registry. The *Registry* module constructs a tagged address book of people within the sentence and a list of addresses. To do so, it combines Machine Learning tools and regular expressions to study the form of sentences. The architecture is visible in Fig 2.

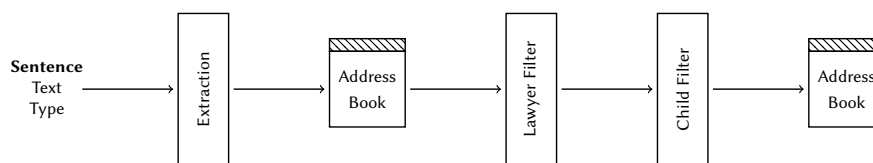


Figure 2: The registry design.

First, using the spaCy¹⁰ library it extracts all strings identifiable as names and using Regex it finds the tax codes. After cleaning the names with a simple filter, a search for name and tax

¹⁰<https://spacy.io>

code matches is performed. For each of the matches, a Person object is created that contains all the information about the person and added to the address book.

Names for which no match was found undergo a second processing step. The module uses regex to find other names that appear in sentences in meaningful places. When the name list is complete we check whether they are present in sentences in which their place and/or date of birth are indicated. If this is the case, a Person object is added to the address book.

If, finally, no match is found in this way either, a Person object is added to the address book with the indication that only the name was found.

A second filtering step is then performed on the rubric to find its role within the sentence. Only regex is used to do this. The possible roles are: (1) Part, (2) Child, (3) Lawyer, (4) Judge and (5) Other.

As a last step, the module searches within the text for all ways and saves them in a list to be anonymised.

Metadata. The *metadata* module looks for information about the sentence for statistical reasons. In particular, it looks for the data that characterize the *Divorce_Frame*:

- the presence of any type of violence;
- the minor age of at least one of the children;
- the involvement of social services;
- the percentage of extra expenses attributed to one of the parties;
- the age of the two parties.

These pieces of information are stored inside a CSV file. For each sentence the module adds a new line and saves it.

Anonymizer. The *Anonymizer* module takes as input the PDF file and the data extracted by the *registry* module and anonymizes them inside the file. It removes the name of the two parties, all the dates and the addresses. It replaces them with an appropriate anonymous placeholder that makes the text still readable and coherent. It produces a new PDF file.

This is done using the PyMuPdf¹¹ library, which allows you to search for strings within the file and replace them.

Statistics. The *statistics* module takes as input the CSV file with the anonymous data and produces different kind of plots.

3. Experiments

To show KlonDikE at work, we devised a different and less tragic conclusion to William Shakespeare's "Romeo and Juliet" masterpiece. After eighteen years of marriage and three children, Romeo and Juliet decide to divorce.

¹¹<https://pymupdf.io>



N. 348623/1341 R.C.
N.....Sent.
N.....Cron.
N.....Rep.
Oggetto:

REPUBBLICA ITALIANA
IN NOME DEL POPOLO ITALIANO

IL TRIBUNALE DI VERONA
PRIMA SEZIONE CIVILE

Composto dai Magistrati:

Domenico Pellegrini
Viviana Mascardi
Andrea Gatti

Presidente
Giudice rel.
Giudice

ha pronunciato la seguente

SENTENZA

nella causa iscritta al n. 348623/1341 promossa da:

ROMEO MONTECCHI (C.F. MNTRMO71H02L7810) elettivamente domiciliato in VERONA VIA ARCHE SCALIGERE 8, 37121, presso lo studio dell'avv. BENVOLIO MONTECCHI (cf MNTBVL65S07L781W) che lo rappresenta e difende in forza di mandato in atti.

PARTE RICORRENTE

Nei confronti di

GIULIETTA CAPULETI (CF CPLGTT75P51L781Z) elettivamente domiciliata in VERONA VIA CAPPELLO 23, 37121 presso lo studio dell'avv. TEBALDO CAPULETI (c.f CPLTLD86M21D150A) che la rappresenta e difende in forza di mandato in atti.

PARTE CONVENUTA

E con l'intervento ex lege del PUBBLICO MINISTERO

CONCLUSIONI DI PARTE RICORRENTE

"Dichiarare la separazione con addebito alla moglie con revoca di qualsiasi assegno per il mantenimento della stessa;

affido condiviso del figlio VALENTINO nato il 15/12/2007;

affido condiviso della figlia GIOVANNA nata il 7/4/2009;

affido condiviso della figlia MARISA nata il 30/1/2012;

Figure 3: The sentence of divorce between Romeo and Giulietta, first page of the PDF document.

The first page of the divorce sentence, written in Italian as all the documents managed in the NGUPP project, is shown in Figure 3 and follows the standard structure of such sentences. It mentions the magistrates in charge for the sentence, the involved parties along with their personal information and their lawyers, and what the first part asks to the second one. The second page, not shown, states that the Tribunal of Verona

1. pronounces the personal separation of Romeo and Giulietta;
2. entrusts the children Valentino, Giovanna and Marisa in a shared form to both parents;
3. assigns the former family home located in Piazza Bra, 1, to Giulietta;
4. declares the maintenance allowance due to Giulietta by Romeo, and that Romeo is required to participate in the amount of 50% of the extraordinary expenses relating to the children.

As explained in the previous sections, Klondike's goal is twofold: first, it looks for the values

```

Sentenza: sentenzaRomeoGiuliettav2.pdf
Tipo: Separazione
Violenza: False
Figli minori: True
Spese straordinarie: 50
Servizi Sociali: False
Persone coinvolte:

```

Name: Montecchi Romeo Birthdate: 1971-06-02 Birthplace: Verona, VR Sex: M Codice Fiscale: MNTRM071H02L7810 Role: part Note:	Name: Capuleti Giulietta Birthdate: 1975-09-11 Birthplace: Verona, VR Sex: F Codice Fiscale: CPLGTT75P51L781Z Role: part Note:	Name: Tebaldo Capuleti Birthdate: 1986-08-21 Birthplace: Cremona, CR Sex: M Codice Fiscale: CPLTLD86M21D150A Role: avv Note:	Name: Valentino Birthdate: 2007-12-15 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:
Name: Giovanna Birthdate: 2009-04-07 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:	Name: Marisa Birthdate: 2012-01-30 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:	Name: Domenico Pellegrini Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:	Name: Viviana Mascardi Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:
Name: Andrea Gatti Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:			

Figure 4: Klondike’s output: completion of the Divorce frame.

that fill the *Divorce_Frame*, and then it looks for these values in the document, to anonymize them.

The result of the first activity is shown in Figure 4: personal details of Giulietta and Romeo like date and place of birth are successfully extracted from their fiscal code (“codice fiscale”, abbreviated as “C.F.”, “CF”, “cf” in the sentence), children are recognized as all being underage at the time of running the system, extraordinary expenses are recognized to be shared in the amount of 50% among the parties, and neither domestic violence nor social services intervention emerges from the sentence.

The anonymization is also accurate, as shown in Figure 5. Not all the details about places are removed (Verona and its postal code are kept visible) but the relevant entities are identified and the occurrence of their names is correctly and consistently replaced with “Parte 1”, “Parte 2”, “Figlio 1”, etc., across the paper.

Although GPT-3.5 and GPT-4 cannot be used in NGUPP for the reasons already discussed in the introduction, we compared Klondike with ChatGPT accessed online that, differently from the GPT-3.5 Turbo and GPT-4 APIs is – at least – free. We fed ChatGPT with the divorce sentence and we asked it to extract the same pieces of information. The result in Figure 6 clearly shows that ChatGPT cannot infer data from the fiscal code, which is instead a precious source of information for Klondike.

On the other hand, ChatGPT anonymization capabilities are almost smart, as shown in Figure 7. The request we made to ChatGPT was to remove all the names of persons and places, GPT-3.5 cannot remove the names of the magistrates, which were instead correctly anonymized in Klondike’s output.

To cope with the need to extract statistical data from sentences, Klondike also generates some charts from the anonymized data. Figure 8 shows the presence of underage children



REPUBBLICA ITALIANA
IN NOME DEL POPOLO ITALIANO

IL TRIBUNALE DI VERONA
PRIMA SEZIONE CIVILE

N. 348623/1341 R.C.
N.....Sent.
N.....Cron.
N.....Rep.
Oggetto:

Composto dai Magistrati:

Giudice 1	Presidente
Giudice 2	Giudice rel.
Giudice 3	Giudice

ha pronunciato la seguente

SENTENZA

nella causa iscritta al n. 348623/1341 promossa da:

Parte 1 (C.F. XXXXXXXXXXXXXXX) elettivamente domiciliato in VERONA ^{Via} ^{anonima} **Parte 1** ^{Via} ^{anonima}, 37121, presso lo studio dell'avv. **Avv 1** **Parte 1** (cf XXXXXXXXXXXXXXX) che lo rappresenta e difende in forza di mandato in atti.

PARTE RICORRENTE

Nei confronti di

Parte 2 (C.F. XXXXXXXXXXXXXXX) elettivamente domiciliata in VERONA ^{Via} ^{anonima}, 37121 presso lo studio dell'avv. **Avv 2** **Parte 2** (cf XXXXXXXXXXXXXXX) che la rappresenta e difende in forza di mandato in atti.

PARTE CONVENUTA

E con l'intervento ex lege del PUBBLICO MINISTERO

CONCLUSIONI DI PARTE RICORRENTE

"Dichiarare la separazione con addebito alla moglie con revoca di qualsiasi assegno per il mantenimento della stessa;

affido condiviso del figlio **Figlio 1** nato il XX/XX/XXXX;
affido condiviso della figlia **Figlio 2** nata il XXXXXXXX;
affido condiviso della figlia **Figlio 3** nata il XX/XX/XXXX;

Figure 5: Klondike's output: anonymization.

and domestic violence, and Figure 9 shows the average age at which the parties divorce, and domestic violence related with age of the parties. The charts are produced from a domain of 100 sentences.

So far, the proposed charts do not show complex data or historical trends. However, they demonstrate the feasibility of a statistical study, also starting from anonymized documents if documents cannot be shared as they are. Given that sentences are time-stamped and that the data we can access, under NDA with the Tribunal of Genova, includes thousand sentences in the last 20 years, analyzing trends over years is indeed possible.

A manual analysis of some of the outputs computed by Klondike has been carried out by one of the authors – and final user of Klondike –, the magistrate Dr. Domenico Pellegrini from the Tribunal of Genova. Due to time constraints that analysis was not systematic, but it was exhaustive enough to suggest that Klondike satisfies the Tribunal needs. Klondike has been

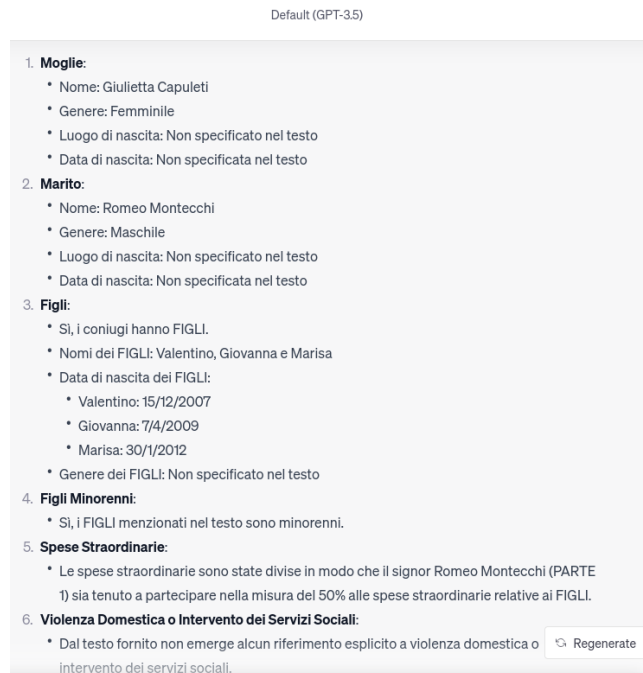


Figure 6: ChatGPT output: completion of the Divorce frame.

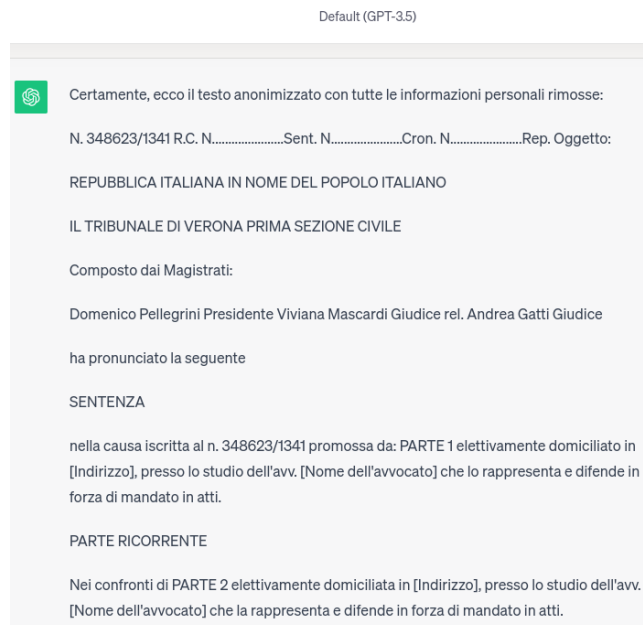


Figure 7: ChatGPT output: anonymization.

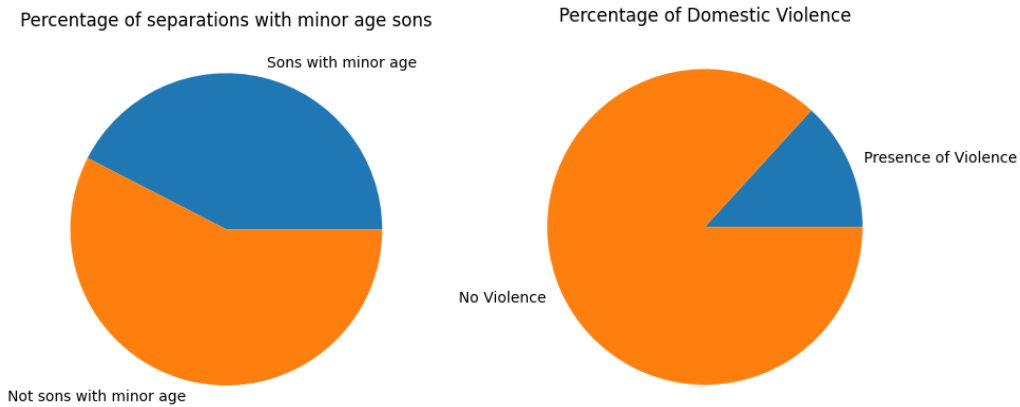


Figure 8: Presence of underage children and domestic violence.

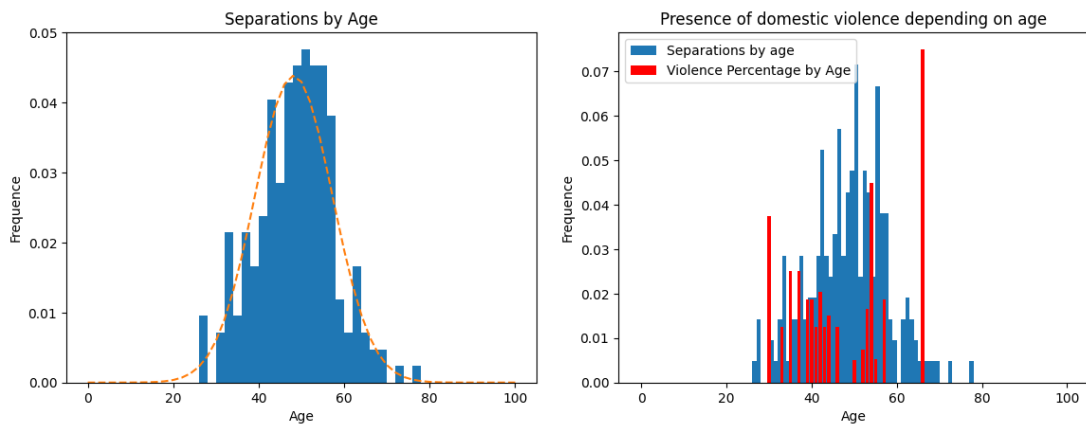


Figure 9: Average age at which the parties divorce, and domestic violence related with age of the parties.

installed in the Tribunal of Genova to allow a deeper evaluation.

4. Conclusions and Future Work

In this paper we presented Klondike, one of the outputs of the NGUPP project, its technical features, and the results of some experiments. Klondike is inspired by frame semantics, and this represents an original approach. Apart from the already mentioned FrEX tool that we developed in 2020, in fact, very few and almost old attempts adopt frame semantics for the legal domain [16, 17, 18].

Klondike is able to extract master data of sentence participants, statistical data, anonymize and produce charts. The tool is thus complete from a functionality point of view. Although GPT-4 performed very well on the same tasks faced by Klondike, it is not acceptable for use in

judicial offices on sensitive sentences possibly involving underage children, and it lacks native support to reading PDF files. Also, GPT-4 ability to produce an output in the required format, namely CSV for statistical analysis and PDF for the anonymized sentence, is very limited.

The application domain of Klondike is, at the moment, restricted to divorce and separation sentences only, but it can be expanded to other kinds of cases; in fact, the only module that should be modified for extracting data from other types of sentences is the one that assigns roles to the persons found. All computations are performed locally and with technologies that can be corrected and refined as needed.

Sentences anonymized using Klondike can be made public without infringing the privacy of the involved parties; anonymized sentences can still be used by other users to extract statistical data or as a benchmark for their own NLP applications in the Italian legal domain. Given that very few benchmarks in this domain exist, and even fewer for the Italian law¹², this would represent a further useful application of Klondike.

Acknowledgments

The authors gratefully acknowledge the support of the “Next Generation UPP: nuovi schemi collaborativi fra Università e uffici giudiziari per il miglioramento dell’efficienza e delle prestazioni della giustizia nell’Italia nord-ovest” project, funded with the contribution of the European Union, Programma Operativo Nazionale Governance e Capacità Istituzionale 2014-2020, Fondo Sociale europeo and Fondo europeo di sviluppo regionale, project code CUP D19J22000240006.

References

- [1] E. Bassignana, D. Brunato, M. Polignano, A. Ramponi, Preface to the Seventh Workshop on Natural Language for Artificial Intelligence (NL4AI), in: Proceedings of the Seventh Workshop on Natural Language for Artificial Intelligence (NL4AI 2023) co-located with 22th International Conference of the Italian Association for Artificial Intelligence (AI* IA 2023), 2023.
- [2] F. Haft, R. Jones, T. Wetter, A natural language based legal expert system for consultation and tutoring – the LEX project, in: Proceedings of the 1st International Conference on Artificial Intelligence and Law, 1987, pp. 75–83.
- [3] M. Lambiris, G. Oberem, Natural language techniques in computer-assisted legal instruction: a comparison of alternative approaches, *J. Legal Educ.* 43 (1993) 60.
- [4] J. A. Meldman, A preliminary study in computer-aided legal analysis, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1975. URL: <http://hdl.handle.net/1721.1/27423>.
- [5] H. Turtle, Text retrieval in the legal world, *Artificial Intelligence and Law* 3 (1995) 5–54.
- [6] M. J. B. II, D. M. Katz, E. M. Detterman, LexNLP: Natural language processing and information extraction for legal and regulatory texts, *CoRR abs/1806.03688* (2018). URL: <http://arxiv.org/abs/1806.03688>. arXiv: 1806.03688.

¹²<https://www.eui.eu/Research/Library/ResearchGuides/Law/Legal-Databases>, accessed on September 2023.

- [7] A. Sleimi, N. Sannier, M. Sabetzadeh, L. C. Briand, J. Dann, Automated extraction of semantic legal metadata using natural language processing, in: G. Ruhe, W. Maalej, D. Amyot (Eds.), 26th IEEE International Requirements Engineering Conference, RE 2018, Banff, AB, Canada, August 20-24, 2018, IEEE Computer Society, 2018, pp. 124–135. URL: <https://doi.org/10.1109/RE.2018.00022>. doi:10.1109/RE.2018.00022.
- [8] G. Ferraro, H. Lam, S. C. Tosatto, F. Olivieri, M. B. Islam, N. van Beest, G. Governatori, Automatic extraction of legal norms: Evaluation of natural language processing tools, in: M. Sakamoto, N. Okazaki, K. Mineshima, K. Satoh (Eds.), New Frontiers in Artificial Intelligence - JSAI-isAI International Workshops, JURISIN, AI-Biz, LENLS, Kansei-AI, Yokohama, Japan, November 10-12, 2019, Revised Selected Papers, volume 12331 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 64–81. URL: https://doi.org/10.1007/978-3-030-58790-1_5. doi:10.1007/978-3-030-58790-1_5.
- [9] F. Fagan, Natural language processing for lawyers and judges, *Mich. L. Rev.* 119 (2020) 1399.
- [10] E. Mumcuoglu, C. E. Öztürk, H. M. Özaktas, A. Koç, Natural language processing in law: Prediction of outcomes in the higher courts of turkey, *Inf. Process. Manag.* 58 (2021) 102684. URL: <https://doi.org/10.1016/j.ipm.2021.102684>. doi:10.1016/j.ipm.2021.102684.
- [11] D. M. Katz, D. Hartung, L. Gerlach, A. Jana, M. J. B. II, Natural language processing in the legal domain, *CoRR abs/2302.12039* (2023). URL: <https://doi.org/10.48550/arXiv.2302.12039>. doi:10.48550/arXiv.2302.12039. arXiv:2302.12039.
- [12] L. Robaldo, S. Villata, A. Wyner, M. Grabmair, Introduction for artificial intelligence and law: special issue "natural language processing for legal texts", *Artif. Intell. Law* 27 (2019) 113–115. URL: <https://doi.org/10.1007/s10506-019-09251-2>. doi:10.1007/s10506-019-09251-2.
- [13] D. Tsarapatsanis, N. Aletras, On the ethical limits of natural language processing on legal text, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021, volume ACL/IJCNLP 2021 of *Findings of ACL*, Association for Computational Linguistics, 2021, pp. 3590–3599. URL: <https://doi.org/10.18653/v1/2021.findings-acl.314>. doi:10.18653/v1/2021.findings-acl.314.
- [14] R. Salvaneschi, D. Muradore, A. Stanchi, V. Mascardi, FrEX: Extracting property expropriation frame entities from real cases, in: P. Basile, V. Basile, D. Croce, E. Cabrio (Eds.), Proceedings of the 4th Workshop on Natural Language for Artificial Intelligence (NL4AI 2020) co-located with the 19th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2020), Anywhere, November 25th-27th, 2020, volume 2735 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 87–103. URL: <https://ceur-ws.org/Vol-2735/paper35.pdf>.
- [15] C. J. Fillmore, Frame semantics and the nature of language, in: *Annals of the New York Academy of Sciences*, volume 280, 1976, pp. 20–32. doi:10.1111/j.1749-6632.1976.tb25467.x.
- [16] G. Venturi, A. Lenci, S. Montemagni, E. M. Vecchi, M. T. Sagri, D. Tiscornia, T. Agnoloni, Towards a FrameNet resource for the legal domain, Number 465 in *CEUR*, 2009, pp. 67–76.
- [17] A. Bertoldi, R. L. de Oliveira Chishman, The limits of using framenet frames to build a legal ontology, in: R. Vieira, G. Guizzardi, S. R. Fiorini (Eds.), Proceedings of Joint IV

Seminar on Ontology Research in Brazil and VI International Workshop on Metamodels, Ontologies and Semantic Technologies, Gramado, Brazil, September 12-14, 2011, volume 776 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2011, pp. 207–212. URL: http://ceur-ws.org/Vol-776/ontobras-most2011_paper26.pdf.

- [18] A. Bertoldi, R. L. de Oliveira Chishman, Developing a frame-based lexicon for the brazilian legal language: The case of the criminal_process frame, in: M. Palmirani, U. Pagallo, P. Casanovas, G. Sartor (Eds.), *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*, volume 7639 of *Lecture Notes in Computer Science*, Springer, 2011, pp. 256–270. URL: https://doi.org/10.1007/978-3-642-35731-2_18. doi:10.1007/978-3-642-35731-2_18.