

Automatic Annotation of Legal References (Allegationes) in the Liber Extra's Ordinary Gloss

Andrea Esuli^{1,*}, Vincenzo Roberto Imperia² and Giovanni Puccetti¹

¹Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", via G. Moruzzi 1 – 56124 Pisa PI, Italy

²Università degli Studi di Palermo - Dipartimento di Giurisprudenza, via Maqueda, 172 – 90134, Palermo PA, Italy

Abstract

The study of normative corpora of the past is a key activity in the fields of Religious Studies and Legal History. The development of intelligent software tools that support this activity is of paramount importance to support the digital transformation of the community. We present an interdisciplinary activity that lead to an accurate automatic annotation of legal references in the Liber Extra's Ordinary Gloss. An index of legal references as been derived from the annotations enabling the creation of novel navigation and data analysis tools. The contribution of this work is twofold: the actual index is already by itself valuable resource for the discipline, and we detail the process that lead to its production, showing that an effective result can be delivered by a small team with limited resources. Both the index and the code are made publicly available.

Keywords

Legal references, Information Extraction, Conditional Random Fields, Dataset

1. Introduction

ITSERR [1] (Italian Strengthening of the ESFRI RI RESILIENCE) is a interdisciplinary and distributed Research Infrastructure for Religious Studies. The development of innovative AI-based tools that support of the digital transformation of the community is one of the many goal of the project. In this context we present a contribution to the "GNORM" software, which aims to provide users with a set of tools and functionalities to facilitate research on large normative *corpora* of the past. One case of study is to enable the consultation of the *Corpus Iuris Canonici* [2] text and its *Glossa Ordinaria* (Ordinary gloss) [3], enabling users to access the network embodied by the visual model in which legal texts (manuscripts and printed books) were structured, as described in Section 1.1.

1.1. *Glossa and Allegationes*, from Medieval Books to Computer Science

Before examining the specific case study presented here and evaluating the concrete applications in the field of legal history resulting from the development of automatic annotation techniques for legal allegations, it is necessary to clarify the meaning of terms such as gloss, ordinary gloss and allegations, and in particular to consider the nature of allegations in the intellectual context of medieval jurists.

The term *glossa* (gloss) refers to "a brief annotation composed and written to explain a text and addressing either its terminology and its exterior trappings or its animating spirit and its underlying principles" [4]. From the late 11th century and for several centuries thereafter, the gloss served as the principal paratextual tool through which law masters in law schools, later Universities, explained and commented on Roman and Canon law compilations, enabling students and practitioners to engage

IRCDL 2025: 21st Conference on Information and Research science Connecting to Digital and Library science, February 20-21 2025, Udine, Italy

*Corresponding author.

† Contributions: Esuli and Puccetti wrote the code, ran the experiments, and wrote Section 2. Imperia made the manual annotation, and wrote Sections 1 and 3. All the authors wrote abstract, conclusions, and revised the final version of the paper.

✉ andrea.esuli@isti.cnr.it (A. Esuli); vincenzoroberto.imperia@unipa.it (V.R. Imperia); giovanni.puccetti@isti.cnr.it (G. Puccetti)

🌐 <https://esuli.it/> (A. Esuli); <https://gpucce.github.io/> (G. Puccetti)

🆔 0000-0002-5725-4322 (A. Esuli); 0000-0001-5029-181X (V.R. Imperia); 0000-0003-1866-5951 (G. Puccetti)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

with technically complex texts. The term *glossa ordinaria* (ordinary gloss) refers to apparatus of glosses compiled by some eminent jurists, renowned for their exceptional quality and thoroughness. Their widespread acceptance ensured their consistent reproduction alongside the normative reference text, establishing them as the *ordinary apparatus* [5]. The phenomenon of normative texts accompanied by ordinary gloss is evident both in the manuscript production of the late Middle Ages, the earliest printed works of the late 15th century, continuing into the significant editorial initiatives of the 16th century.

Manuscripts and printed texts retained the standard page layout essentially unchanged, with the normative text and the ordinary gloss arranged according to a model that has been defined, in its essential features, as the “agora template” [7], as shown in Figure 1. This spatial configuration, with the normative text at the center framed by the glosses, mirrored the dialogical nature of the legal interpretation method: the authoritative text was accompanied by commentary that explained, questioned or contradicted it. It can be affirmed that it is the very content of the apparatus that qualifies them as “hypertexts” [8], since they demand the active participation of the reader. The margins of the normative text contain various comments, each serving a specific function, but collectively acting as tools to help the user understand the main text, making it more accessible and consultable. Particular attention will be paid here to one of these components, the *allegations*, defined by Hermann Kantorowicz [9] as references to *auctoritates* that must justify every assertion or, at least, deal with those that at first glance appear to contradict it. These references are, in the vast majority of cases, citations of legal norms, presented in a highly abbreviated form, using abbreviations and numbers that correspond to the conventional criteria for identifying the legal source containing the norm. These references, initially sporadic in the earliest layers of glosses, gradually increased in number until they became quantitatively predominant, alongside the refinement of interpretive techniques developed by jurists. Many recent work [10, 11, 12, 13] reaffirmed and clarified that allegations are not merely citations: the use of a reference constitutes an appeal to an authority that not only lends greater solidity to the legal argument, but forms its very foundation. The use of legal references were carried out on legitimate grounds, based, on the one hand, on applicable and valid legal norms and, on the other, on the consistent use of legal categories shared within a common cultural and intellectual context [11].

These conceptual premises are closely related and provide the backdrop for considerations concerning more practical aspects. The first pertains to the citation style of *allegations*, which presupposed a normative text that was now stable, fixed, and unalterable. This allowed the reader to pinpoint the referenced passage with certainty [13]. The second concerns the nature and purpose of these references. The primary aim was to create a network in which the various *sedes materiae* were organised and linked, thus enabling the reader-user to navigate vast normative compilations [8].

Moreover, the use of legal references could take on an even more incisive argumentative function if, in addition to a principle or rule derived from the normative text, allegations *pro* and *contra* were included. This technique quickly evolved into a distinct literary genre, with many works constructed



Figure 1: A page from the *Liber Extra's Ordinary Gloss* [6]. The box of text in the upper center of the page is the normative text from *Liber Extra*. The surrounding text are the glosses that also contain the legal references.

according to these specific criteria [14].

The formal and substantive nature of the *allegationes* in the works of the jurists underscores the need to exploit the potentials of computer science, and specifically machine learning, to develop accurate tools capable of preserving their peculiarities in future digital editions of medieval legal texts [13].

2. Annotation of legal references

The complex and stratified genesis of the ordinary apparatus to normative compilations, closely linked to the concept of the authoritative text itself in the Middle Ages, prevented the preparation of critical editions of them in the modern sense. Consultation of them still requires recourse to manuscripts or early modern printed editions [11]. This is also the case for the *Corpus iuris canonici*. For this reason, the *Glossa Ordinaria* (Ordinary Gloss) to the *Decretales Gregorii IX*, best known as the *Liber Extra*, was chosen for the design and development of an automatic annotation system for the legal allegations. Promulgated by the Pope with the bull *Rex Pacificus*, the legal collection is subdivided into 5 books, 185 titles, 1971 chapters, with a total of 9872 lemmas. The gloss of each lemma contains a variable number of legal references. From the text of the 1582 *Editio Romana* [3], which is the common reference edition, a complete digital transposition was made by Edward A. Reno III, as part of the project “The Digital Decretals” [15]. The files containing the digital text can be found on the project’s website, together with specific references to the sections of the apparatus not included in the transcription work, as well as the interventions made to standardise elements of spelling, abbreviation, punctuation and numbering in the text.

Figure 2 illustrates the process that lead to an accurate automatic annotation of the whole *Liber Extra*’s Ordinary Gloss based on the annotation of a small subset by a human expert, followed by the creation of an index of all the annotation in which every legal reference, pointing to a specific norm, is linked to a lemma, chapter, and title, of the *Liber Extra*. The next Sections will detail the process.

2.1. The annotation schema

An annotation is a span of text marked in the text to identify some relevant information of a specific type. The annotation schema we defined identifies the four types of entities:

- Annotation of type *glossed lemma*, “Lemma glossato”, indicates the specific terms that are glossed. Legal references are included in the gloss text.
- Annotations of type *legal reference*, “Allegazione normativa” annotate the references to legal norms or regulations.
- The last two annotations types are *title*, “Titolo”, and *chapter*, “Capitolo”, which together with the lemma precisely identify the position of the legal reference in the *Liber Extra*.

The annotations of these entities thus enables building a link between the *Liber Extra* and the legal norms and regulations that are crucial for the interpretative framework of the book. The annotation data also include the exact character position of the beginning and the end of the annotated text in the digital text.

2.2. Expert annotation

An expert of the domain performed the annotation of the legal references on a subset of the book. The expert annotated 12 titles out of 185, making a total of 4578 annotations. The titles have been randomly sampled across the whole book to better cover the different content of each section of the book, i.e., following Reno’s notation [15]: 1.02, 1.11, 1.33, 2.02, 2.23, 3.02, 3.26, 4.17, 4.19, 5.01, 5.03, 5.23. The expert focused only on annotating the legal references, as the annotations of the other three types have been made successively in a completely automatic way, as described in Section 2.3.

The tool used by the expert to perform the manual annotation is INCEpTION [16]. INCEpTION is a popular annotation platform designed for collaborative and efficient text annotation. It supports a wide

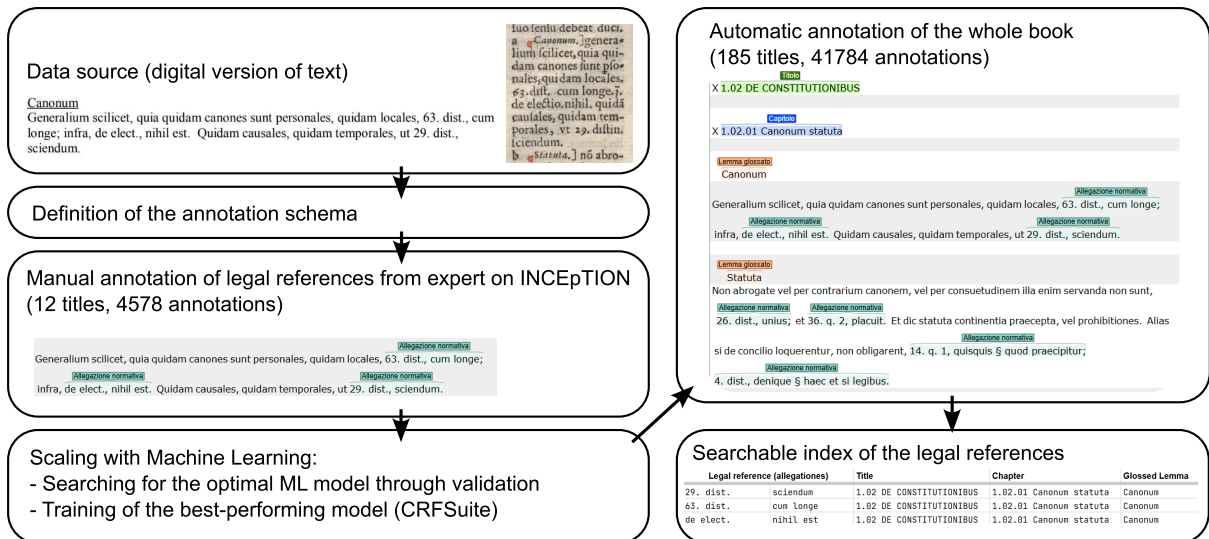


Figure 2: An illustration of the process that led to construction of the index of all the legal references in the *Liber Extra’s Glossa Ordinaria*.

range of tasks, including named entity recognition, relation extraction, and classification. INCEpTION helps the manual annotation with a dedicated GUI design, and also by providing automated suggestions for spans of text that are evaluated as potential new annotations, which the expert can validate with a single mouse click. The automated suggestions are based on the definition of a recommender, i.e., a machine-learning algorithm that trains a model by continuously observing the annotations made by the expert. The automatic annotation model trained by INCEpTION is a valuable help for the annotator, yet it is not of sufficient accuracy to be used to perform a complete automatic annotation of the rest of the book. We thus used the 12 annotated titles as a training set for a proper batch training process of more accurate models, as detailed in Section 2.3.

2.3. Training a model for the automatic annotation of legal references

The code to replicate the automatic annotation is published with an open source license [17]. We focus here only on the legal references, as the annotation of the other types of entities has been solved using different methods, as detailed in Section 2.3.1.

The automatic annotation process is modeled as a word labeling problem. We adopted the BILOU labeling schema [18] in which each word can be assigned to one of a set of labels, either if is part of a legal reference or not (O label). For words that are part of a legal reference, different labels are used if the word is at *beginning* of the annotation (B label), it is the *last* (L label), it *inside* the sequence of words that form the legal reference (I label), or the legal reference is composed of a *unique* word (U label).

We tested two approaches to train the automatic annotation model: using a “traditional” statistical machine learning algorithm, i.e., Conditional Random Fields [19] (CRFs), and fine-tuning a transformer-based Large Language Model (LLM) [20]. A preliminary test of few-shot prompting [21] of a generative LLM reported a very low accuracy and was discarded.

CRFs are probabilistic graphical models that are used to model sequential or structured data by defining conditional probabilities of a set of target variables given a set of observed variables, allowing for the incorporation of context and interdependencies among the variables in the target space. In natural language processing (NLP), CRFs are often used for tasks like named entity recognition or part-of-speech tagging, where the output labels (e.g., tags) are interdependent. We tested two configurations for the CRFs graph: a *simple* configuration in which the observed variables are all the word bigrams in a context window of three words before and after the one to be annotated, and a *rich* configuration that considered bigrams and trigrams in a context windows of size seven.

For the fine tuning of LLMs we tested two models: the original BERT [22], as a baseline, and Latin

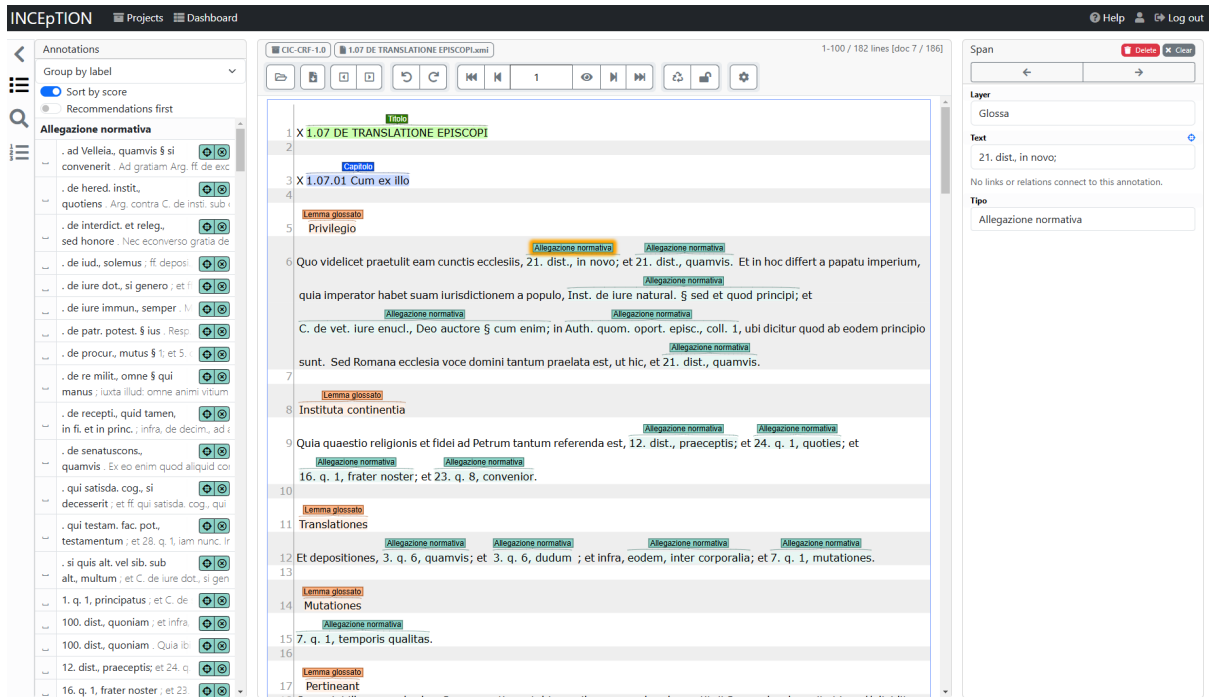


Figure 3: The annotation interface in INCEption. The document been annotated is shown in the center of the page. Text highlighted in different colors indicate the annotated entities of different type: orange for the “Lemma glossato” (glossed lemma), green for the “Allegazione normativa” (legal reference). The left-side bar lists all the annotations in the document, the right-size bar show the details of the currently selected annotation.

BERT [23], a state-of-the-art model for Latin. The fine tuning of LLMs was made using the Transformers python package [24], training for 10 epochs, using a learning rate of $2 \cdot 10^{-5}$, a weight decay value of 0.01, and a batch size of 32. Sequences longer than 512 token were split in multiple sequences.

The comparison of the methods is based on a 10-fold cross validation on the expert-annotated data. The annotated data is split into ten parts. The splits are kept the same across all the tested methods. For every fold one tenth of the annotated data is considered to be the test data and the remaining nine tenth are the training data. The process repeats for all the folds, collecting for each tested method an automatic annotation of all data. The accuracy of the automatic annotation is determined by comparing it with the annotations from the expert, using a token-and-blank evaluation model [25].

Table 1

Results of cross-validation experiments (10 folds) on the expert annotated data. “Training Time” reports the time require to train an automatic annotation model on the entire set of expert-annotated data.

Method	Accuracy	Training Time	Hardware	Model size
CRFs basic	0.891	7m	CPU (20 cores)	0.5MB
CRFs rich	0.978	21m	CPU (20 cores)	1.1MB
BERT [22]	0.837	13m	GPU (4 A40)	411.8MB
Latin BERT [23]	0.924	13m	GPU (4 A40)	423.3MB

Results in Table 1 show that CRFs in the rich configuration performed with close to perfect accuracy. The more complex graph of the rich configuration required three times the training time of the basic configuration, yet the training time is still short and the improvement is worth the additional cost.

The comparison between BERT and Latin BERT shows the impact of the main training language to a task. The tokenizer of BERT obviously struggles with Latin words. For example the five-words expression “Quae est radix omnium malorum” is tokenized by BERT in 19 tokens, whereas Latin BERT produces exactly five tokens. This allowed Latin BERT to better identify the relevant elements of the language that are related to the expression of legal references, whereas BERT struggled with very short

sub-word tokens, which are evidently less statistically related to the concept of legal reference.

The comparison of CRFs with LLMs highlights that in this annotation task CRFs have many advantage points. The obvious one is that CRFs get the best accuracy. Not less important is that CRFs require less computational resources. The training time of CRFs is based on using a desktop with a single Intel i9 CPU, while the LMMs time are based on a dedicated server with 4 A-40 NVIDIA GPUs, costing roughly ten times the desktop computer. Similarly the sizes of the final models shows a clear advantage for CRFs. The differences in training time are not very relevant, considering the differences in hardware, and the fact that even the longest training time is relatively quick. CRFs makes possible to train an annotation model on personal hardware, enabling the adoption of machine learning to small research groups with limited computational resources. More experiments on the fine tuning of LLMs may have reduced the gap with CRFs, yet the high accuracy obtained by CRFs satisfied our goals, thus not justifying the additional computational costs.

2.3.1. Automatic annotation of other entities

The annotation of chapters and titles has been made using a regular expression, exploiting the specific format used in The Digital Decretals [15]. For the annotation of the glossed lemmas, we had only a ordered list of the glossed lemmas, leaving us to identify their position in the text. We solved this with a search algorithm that considered the constraints imposed by the list. For example, the word “Omnipotens” occurs 31 times in the whole text, but the only instance as a glossed lemma occurs after the glossed lemma “Incomprehensibilis” and before “Ineffabilis”. Solving all the constraints for all the glossed lemmas allowed us to find the exact position for all of them.

2.4. Building the index of legal references

The automatic annotation of the whole text identified 41784 legal references [26], each linked to a glossed lemma, a chapters, a title, and a book part (among the five parts of the Ordinary Gloss relative to the five books of the *Liber Extra*). The automatic annotation thus had a multiplier effect of 9.1 with respect to the number of annotation from the expert.

We split each legal references in two parts, the “title” or “section” that contains the referenced norm, and the string that identifies the norm. For example, the legal reference “22. q. 4, incommutabilis” points to Gratian’s *Decretum, Causa 22, Quaestio 4*, and its chapter (or canon) “incommutabilis” (according to today’s citation standard: C.22 q.4 c.9). This allowed us to identify a total of 1795 referenced titles or sections, from various collections, i.e., the *Corpus Iuris Canonici* and the *Corpus Iuris Civilis*.

The index enables the researchers to perform the activities presented in Section 1. In addition to the use as a browsing resource, the index itself can be the subject of analysis. For example, Table 2 lists the ten titles or sections with most references in the Ordinary Gloss, considering the whole *Liber Extra* and each of its books separately. This is a simple example of a data analysis enabled by the index.

Table 2

The ten sections or titles with most references in the Ordinary Gloss, considering the whole *Liber Extra*, or each book of the *Liber Extra* separately. Numbers indicate the number of references.

Liber Extra	Book 1	Book 2	Book 3	Book 4	Book 5
de elect. 1217	de elect. 739	de appell. 499	de praeben. 225	de spons. 200	de sent. excom. 405
de appell. 930	de offi. deleg. 366	de testib. 363	12. q. 2 181	de despon. impub. 100	de simon. 223
de offi. deleg. 721	de rescript. 361	de offi. deleg. 219	de iure patron. 177	de eo qui dux. 63	de accusat. 172
de sent. excom. 689	de appell. 296	de elect. 207	de elect. 164	qui fil. sint legit. 59	11. q. 3 150
de rescript. 585	de praeben. 158	de praescrip. 153	16. q. 1 162	de eo qui cog. cons. 55	de homic. 127
de testib. 545	ff. de procur. 121	de restit. spol. 145	de decim. 160	de cond. appos. 48	de privileg. 116
de praeben. 439	de re iudic. 104	de iureiur. 135	de conuer. coniug. 108	de frig. et malef. 45	1. q. 1 114
12. q. 2 388	ff. de recepti. 103	de re iudic. 132	16. q. 7 108	de divort. 37	de haeret. 107
de simon. 379	de sent. excom. 91	2. q. 6 129	de censib. 103	27. q. 2 36	50. dist. 102
11. q. 3 359	de renunciat. 91	de probat. 124	de concess. praeben. 95	de cons. et affin. 34	de elect. 87

3. Qualitative analysis

A qualitative evaluation of the results of the automatic annotation process is particularly relevant when considering the degree of correctness and accuracy in relation to the amount of data processed.

Regarding the correctness of the annotation, the system's ability to distinguish legal references from the rest of the text, which is usually written in a discursive and argumentative style, according to the explanatory function inherent in the gloss, is impressive. Concerning the accuracy of the annotation, the system has demonstrated, in most of the cases, its ability to delimit the exact scope of legal references. There are a few exceptions when the text of the allegation is unusually long or contains overly specific references. There are no particular issues in identifying allegations that refer to the *Liber Extra* or to the three *partes* of Gratian's *Decretum*, as each is clearly defined in its citation form. The same holds true for the allegations to the Code or the Institutes of the *Corpus Iuris Civilis*, which are usually clearly preceded by their respective symbols (C. or Inst.).

The automatic annotation has a recurring imprecision when identifying references to the Digest, preceded by the abbreviation "ff.", which is erroneously omitted in many cases. In this case the automatic annotation model shown a preference for shorter annotations, considering that in most cases both versions of the annotations, with or without "ff.", are potentially correct yet referring to different sources. The specific nature of this error made it easy to solve with a simple automatic post processing of the annotations, checking for any eventual "ff." preceding them and adding it to the annotation, as we found no cases in which the "ff." expression preceding an annotation was to be excluded. The published version of the index [26] is thus not affected by this issue.

Quantitatively insignificant are the cases in which the system detects text passages as legal allegations when they are not. Particularly interesting are the cases where the system, due to the syntactic structure of the text, has detected actual allegations that are "non-legal", such as references to Gospel passages. In conclusion, the large amount of correct annotations and the limited effort required to correct the erroneous ones, which will be the next step of the project's activities, confirms the validity and potential of the approach to annotation we presented.

4. Conclusions

In the field of Legal History, the possibility of using an automatic annotation system for legal allegations is proving to be a valuable tool. The development of coordinated and interconnected databases of normative texts and their commentary apparatus, as well as the creation of digital editions of fundamental works of medieval jurists, are just a few examples of potential applications. Moreover, the creation of specific reference ontologies could enable the training of increasingly complex retrieval systems, capable of extracting and sorting a vast amount of data that would otherwise be unmanageable, making larger projects practically infeasible.

In the case of study presented in this work, we have been able to annotate a large amount of text exploiting the annotation a human expert on a tenth of the corpus and then a very effective machine learning setup to have a complete, accurate annotation of the whole text. The outcome of this research activity is twofold: we produced a valuable resource, the index, that will contribute to the GNORM software and support studies on the Ordinary gloss and the *Liber extra*; and we have proved an effective low-resource pipeline that can be replicated for similar activities in the field of Religious Studies, Legal History, and many related disciplines in the humanities.

Acknowledgments

This work was supported by project "Italian Strengthening of ESFRI RI RESILIENCE" (ITSERR) funded by the European Union under the NextGenerationEU funding scheme (CUP:B53C22001770006).

References

- [1] ITSERR (Italian Strengthening of the ESFRI RI RESILIENCE), 2024. URL: <https://itserr.it>.
- [2] K. Pennington, *Corpus iuris canonici*, in: J. Otaduy, A. Viana, J. S. Rueda (Eds.), *Diccionario general de derecho canónico*, Thomson Reuters Aranzadi, 2012, pp. 757–765.
- [3] Bernard of Parma, *Glossa Ordinaria to Decretals Gregory IX*, in: *Decretales D. Gregorii Papae IX. suae integritati una cum glossis restituta. Cum privilegio Gregorii XIII. Pont. Max. et aliorum Principum*, Romae, In Aedibus Populi Romani, 1582.
- [4] M. Bellomo, *The common legal past of Europe, 1000–1800*, 1995.
- [5] G. Dolezalek, *Glosses and the juridical genre “apparatus glossarum” in the middle ages*, *Rivista Internazionale di Diritto Comune* 32 (2021) 9–54.
- [6] *Decretales D. Gregorii Papae IX. suae integritati una cum glossis restituta. Cum privilegio Gregorii XIII. Pont. Max. et aliorum Principum*, Romae, In Aedibus Populi Romani, 1582, available at *UCLA Library Digital Collections*, 2024. URL: <https://digital.library.ucla.edu/catalog/ark:/21198/zz0014rx7w?cv=35>.
- [7] A. M. Hespanha, *Form and content in early modern legal books*, *Rechtsgeschichte-Legal History* 12 (2008) 12–50.
- [8] G. Speciale, *Apparatus: ipertesto vivo e aperto*, *Ius Commune. Zeitschrift für Europäische Rechtsgeschichte* 28 (2001) 47–59.
- [9] H. U. Kantorowicz, *Die allegationen im späteren mittelalter*, *Archiv für Urkundenforschung* (1935) 15–29.
- [10] D. Quaglioni, *Licet allegare poetas: formanti letterari del diritto fra medioevo ed età moderna, Poesia e diritto nel Due e Trecento italiano.-(Memoria del tempo; 65)* (2019) 209–219.
- [11] S. Menzinger, *Reflections on the connection between author and text in medieval juridical production*, *Historia et ius* 11 (2017).
- [12] S. Menzinger, *The past, the others, himself: The open dialogue of a medieval legal author with his text*, in: *Sicut dicit: Editing Ancient and Medieval Commentaries on Authoritative Texts*, Thurnout, 2019, pp. 273–299.
- [13] S. Menzinger, *Interazione tra testo e ‘citazione’ nella dottrina giuridica civilistica: secoli XII e XIII*, in: *Juristische Glossierungstechniken als Mittel rechtswissenschaftlicher Rationalisierungen*, Erich Schmidt Verlag, 2022, pp. 15–26.
- [14] P. Weimar, *Argumenta brocardica*, *Studia Gratiana* 14 (1967) 89–123.
- [15] E. Reno, *The Digital Decretals*, <https://www.digitaldecretals.com/>, 2024. [Online; accessed 1-November-2024].
- [16] J.-C. Klie, M. Bugert, B. Boullosa, R. E. de Castilho, I. Gurevych, *The inception platform: Machine-assisted and knowledge-oriented interactive annotation*, in: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, 2018, pp. 5–9. URL: <http://tubiblio.ulb.tu-darmstadt.de/106270/>, event Title: The 27th International Conference on Computational Linguistics (COLING 2018).
- [17] A. Esuli, G. Puccetti, *A Machine Learning pipeline to automatically annotate legal references (allegationes) in the Liber Extra’s Ordinary Gloss*, 2024. URL: https://github.com/aesuli/CIC_annotation. doi:10.5281/zenodo.14381817.
- [18] L. Ratnov, D. Roth, *Design challenges and misconceptions in named entity recognition*, in: *Proceedings of the thirteenth conference on computational natural language learning (CoNLL-2009)*, 2009, pp. 147–155.
- [19] C. Sutton, A. McCallum, et al., *An introduction to conditional random fields*, *Foundations and Trends® in Machine Learning* 4 (2012) 267–373.
- [20] A. Vaswani, *Attention is all you need*, *Advances in Neural Information Processing Systems* (2017).
- [21] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., *Language models are unsupervised multitask learners*, *OpenAI blog* 1 (2019) 9.
- [22] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, *BERT: Pre-training of deep bidirectional transformers for language understanding*, 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, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: <https://aclanthology.org/N19-1423>. doi:10.18653/v1/N19-1423.

- [23] D. Bamman, P. J. Burns, Latin bert: A contextual language model for classical philology, arXiv preprint arXiv:2009.10053 (2020).
- [24] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, 2020, pp. 38–45.
- [25] A. Esuli, F. Sebastiani, Evaluating information extraction, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2010, pp. 100–111.
- [26] A. Esuli, V. R. Imperia, G. Puccetti, Automatic Annotation of the Legal References in the Liber Extra's Ordinary Gloss (1.0) [Data set], 2024. doi:10.5281/zenodo.14381709.