Hybrid Refining Approach of PrOnto Ontology

Hybrid Refining Approach of PrOnto Ontology

Monica PALMIRANI*, *Giorgia BINCOLETTO, **Valentina LEONE, *Salvatore SAPIENZA, ***Francesco SOVRANO,

*CIRSFID, University of Bologna.
{monica.palmirani, salvatore.sapienza, giorgia.bincoletto2}@unibo.it
** Computer Science Department, University of Turin
{leone}@di.unito.it
***DISI, University of Bologna
{francesco.sovrano}@unibo.it

Abstract. This paper presents a refinement of PrOnto ontology using a validation test based on legal experts' annotation of privacy policies combined with an Open Knowledge Extraction algorithm. Three iterations were performed, and a final test using new privacy policies. The results are 75% of detection of concepts and relationships in the policy texts and an increase of 29% in the accuracy using the new refined version of PrOnto enriched with SKOS-XL lexicon terms and definitions.

Keywords. legal ontology, GDPR, OKE, refinement.

1. Introduction

We have already published several papers about PrOnto ontology [24][26][27][28] which aims to model the concepts and their relationships presented in the GDPR (General Data Protection Regulation EU 2016/679). PrOnto is a core ontology that started with a top-down method, using MeLOn methodology, based on a strong legal informatics analysis of GDPR normative provisions and their interpretations issued by Art. 29 WP (now, European Data Protection Board) through its opinions. PrOnto intends to represent data types and documents, agents and roles, processing purposes, legal bases (Art. 6 GDPR), processing operations, and deontic operations for modelling rights (Chapter 3, Artt.12-23 GDPR) and duties (Chapter 4, Artt. 24-43 GDPR). The goals of PrOnto are: i) supporting legal reasoning and ii) compliance checking by employing defeasible logic theory (i.e., the LegalRuleML standard and the SPINDle engine [25]); iii) helping the legal design visualization based on robust theoretical legal conceptualisation [32]; iv) supporting information retrieval. In the previous papers the validation was carried out by legal experts (e.g., PhD students and researchers) and through SPARQL queries on the basis of some RDF triples. This article presents a different validation process of PrOnto ontology using – following the application of a robust theoretical and foundational top-down methodology – a bottom-up approach, starting from the language adopted in real examples of Privacy Policies. The research investigates: i) if the existing PrOnto classes are sufficiently exhaustive to support NLP tools in detecting GDPR concepts directly from Privacy Policies; ii) if some classes are missing with respect to the pragmatic language forms; iii) if some frequent terminology could be added to the conceptualisation

modelling using e.g., SKOS-XL and so help the Open Knowledge Extraction (OKE) tools to support search engine goals; iv) whether it is possible to create a ML tool that is capable of detecting GDPR concepts in the Privacy Policies and so to classify them with PrOnto creating RDF triples. The paper first examines the used methodology; secondly, it presents the legal analysis of the Privacy Policies chosen for the validation and the related mapping of the linguistic terminology in the PrOnto classes; then, the work describes the ML technique applied to detect the PrOnto concepts from the other Privacy Policies and its results; finally, the conclusion discusses the refinements made to the PrOnto ontology thanks to the validation with the Privacy Policies.

2. Methodology

PrOnto was developed through an interdisciplinary approach called MeLOn (Methodology for building Legal Ontology) and it is explicitly designed in order to minimise the difficulties encountered by the legal operators during the definition of a legal ontology. MeLOn applies a top-down methodology on legal sources. It is strongly based on reusing ontology patterns [15][1] and the results are evaluated using foundational ontology (e.g., DOLCE [11]) and using OntoClean [14] method. Finally, the validation is made by an interdisciplinary group that includes engineers, lawyers, linguists, logicians and ontologists. Hence, the legal knowledge modelling is performed rapidly and accurately while integrating the contributions of different disciplines.

For these reasons, the methodology of this research is based on the following pillars taking inspiration from other research in the legal data analytics [2][3][36]:

1. two legal experts selected ten privacy policies from US-based companies providing products and services to European citizens; so, the GDPR is applied according to its territorial scope (Art. 3);
2. the privacy policies were analysed using the comparative legal method to discover the frequent concepts mentioned in the texts and how they express the legal bases (Art. 6 GDPR), the purposes, the data subject's rights and the duties of the controller, some particular processes like information society services for children (Art. 8 GDPR), profiling and automatic decision-making systems (Art. 22 GDPR), processing of sensitive data (Art. 9 GDPR) and data transfer to third countries (Chapter 5 GDPR);
3. selected portions of text were mapped into the PrOnto ontology with also different linguistic variations, including syntagma. A table summarising the main linguistic expressions related to each PrOnto classes was set up;
4. this mapping was provided to the computer science team that used Open Knowledge Extraction technique starting from the GDPR lexicon, PrOnto ontology and the literal form variants to annotate the Privacy Policies;
5. results were validated by the legal team that returned the feedbacks to the technical team. In the meantime, they also discussed on some possible refinements of PrOnto ontology, following the MeLOn methodology, in order to better model the legal concepts (e.g., they proposed to add classes for Legal Basis);

[1] PrOnto reuses existing ontologies ALLOT [4] FRBR [19], LKIF [6] we use in particular lkif:Agent to model lkif:Organization, lkif:Person and lkif:Role [6], the Publishing Workflow Ontology (PWO) [12], Time-indexed Value in Context (TVC) and Time Interval [31]. Now with this work we include also SKOS-XL [8][5].

6. the steps from 2 to 5 were iterated three times using the results of the algorithm in order to refine the ontology and the software model;
7. finally, new Privacy Policies were selected by the legal experts[2] in order to evaluate the effectiveness of the refined algorithm and ontology.

3. Legal Analysis of the Privacy Policies

We have selected ten Privacy Policies from an equal number of companies. The policies were extracted from the dedicated sections of the companies' websites made available to European visitors. We chose these companies due to their international dimension, their relevance in their market sectors and the diversity of data processing techniques. The Privacy Policies were analysed using the following macro-areas to follow a comparative method: sale of goods, supply of services and sharing economy. We compared for each macro-area the linguistic terms and we distinguished between the legal strict terminologies (e.g., data subject) and the communicative language (e.g., customer or user).

Table 1 - List of the Privacy Policies analysed.

Sale of goods	Amazon	Dell	McDonalds	Nike
Supply of services	American Airlines	TripAdvisor	Hertz	Allianz U.S.
Sharing economy	AirBnb	Uber		

The legal experts have manually reviewed the Privacy Policies to discover the concepts of legal relevance for data protection domain (provisions, legal doctrine, Art. 29 WP/EDPB and case law) that are remarkably recurrent in the text. The interpretation has also kept into account the existing version of PrOnto ontology, in particular to identify the different wording that expresses the same concept recognised through a legal analysis at an equal level of abstraction. Occasionally these terms present different forms as the companies work in different sectors and across multiple jurisdictions. Thus, these forms have been analysed, compared and eventually included in the PrOnto ontology, using techniques like SKOS-XL for adding the different linguistic forms (e.g., `skosxl:leteralForm`). This extension of PrOnto definitely improves the capacity of the OKE tools to detect the correct fragment of text and to isolate the legal concept as well as populating the PrOnto ontology. We also noted that the Privacy Policies tend to use the ordinary, everyday language for reasons of transparency and comprehensibility of the texts. Despite the advantage for the costumer/user, the analysis underlined that certain terminologies are not accurate from a legal perspective. When a manual or NLP-assisted analysis is performed, such legal nuance is more difficult to detect. For instance, the expression "*giving permission*" is a communicative substitute of "*giving consent*" and "*obtain consent*", which implies the freely given, informed, unambiguous and specific nature of the data subject's agreement. This choice is probably made to simplify the expression and highlighting transparency. Another example is the sentence "*you can also update your personal data*" which does not convey the deontic specification of the right to request rectification of personal data (Art. 16 GDPR).

[2] Rover, Parkclick, Springer, Zalando, Louis Vuitton, Burger King, Microsoft-Skype, Lufthansa, Booking, Zurich Insurance.

Moreover, after the analysis it can be argued that some terminologies are misused because the ordinary language in the policy does not reflect the legal sense. As an example, in the sentence "*otherwise anonymized (anonymous data), aggregate or anonymous information*" the type "anonymous data" (Recital 26 GDPR) is not in the scope of the Regulation and it is misled with "anonymised data". This type is a personal data handled to become anonymous by means of a sophisticated technical process (e.g., generalization, aggregation, permutation, etc.). The same issue can be found in the statement "*when collecting or transferring sensitive information, such as credit card details"*: the definition of sensitive or special category of data does not include any kind of financial information (Recital 10, 51, Art. 9 GDPR). In these cases, the PrOnto ontology should steer the technical detection of the legal concepts. Furthermore, we found that certain terminology is borrowed from the computer science domain and goes beyond the legal provisions. For instance, the forms "*to hash*", "*log files*", "*use encryption*" convey a technical meaning that is not used by GDPR requirements as the Regulation has been drafted in a technically neutral way.

4. PrOnto Mapping and New Modules

Following this analysis, we have mapped the synthesis of the different lexicon expressions with the PrOnto classes and this table was the basis for creating mapping between lexicon (terms and definitions) and taxonomy of concepts (classes). This step immediately allowed to detect some missing modules that are described below.

4.1 Legal Basis Module

Under the GDPR, personal data processing (Art. 4.1(2)) is lawful only if motivated by a purpose that must be legitimated by a legal basis (see Art. 6 GDPR on the lawfulness of processing). Therefore, a lawfulness status was needed and was thus added as a Boolean data property of the `PersonalDataProcessing` class. However, from the validation using Privacy Policies, it is extremely important to elicit the `LegalBasis` class because several other implications (rights, obligations, actions) depends to the kind of legal basis (e.g., Art. 22). For this reason, we have modelled the following new module (see Fig. 1 the new classes are displayed in orange).

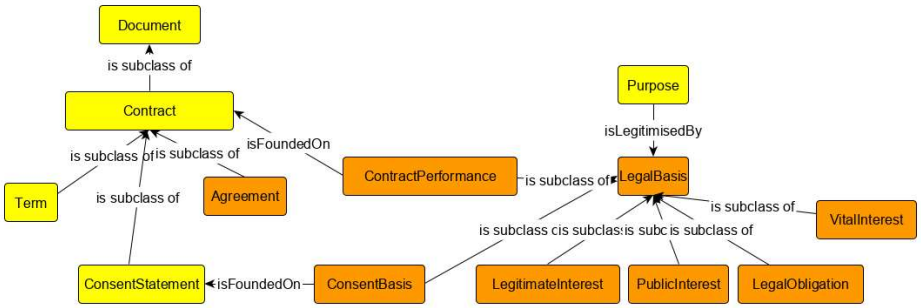

Figure 1 – Legal Basis Module

4.2 Purpose Module

"Archiving" and "Services" are encountered frequently among the purposes of prosessing described in the Privacy Policies and they are added to the `Purpose` module, with also an important kind of service (`InformationSocietyService`) relevant in the child privacy management (Art. 8 GDPR) (See Fig. 2).

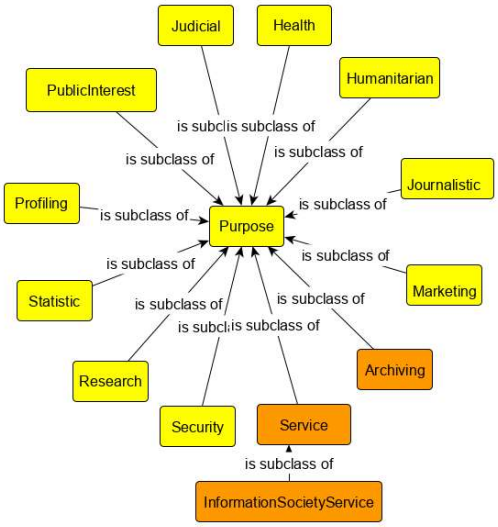

Figure 2 – Purpose Module

4.3 Obligations and Rights Module

The Privacy Policies underlined some rights with the related obligations, like the `ObligationToProvideHumanIntervention` connected with `RightToHaveHumanIntervention` and related with `AutomaticDecisionMaking` that is a new action added to the `Action` module (See Fig. 3).

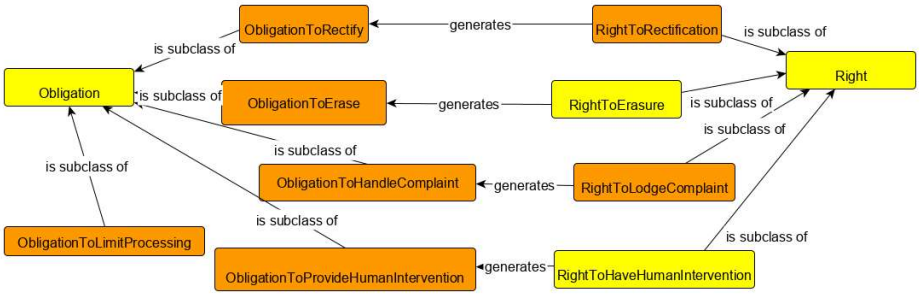

Figure 3 – New Rights and Obligations

5. Open Information Extraction for Legal Domain

We built a software for detecting GDPR concepts from Privacy Policies taking inspiration from the PrOnto ontology and using the tool conceptually based on ClausIE

[9]. ClausIE is a clause-based approach to Open Information Extraction (Open IE), that exploits linguistic knowledge about the grammar of the English language to first detect clauses in an input sentence and to subsequently identify the type of each clause according to the grammatical function of its constituents.
The goal of Open IE is to build information graphs representing natural language text in the form of SVO (Subject, Verb, Object) triples (please note that this is slightly different from building RDF graphs).
The main difference between our tool and ClausIE is that our tool extracts SVO triples by using Spacy's[3] state-of-the-art dependency parser based on Neural Networks.
One of the main issues arising from exploiting such dependency parser might be that it has been trained on common language rather than legislative texts and rhetoric sentences. But we argue that our choice is meaningful and correct since policy text is usually simpler and uses common language to be more understandable.

Identifying SVO triples through dependency parsers based on Neural Networks was used in other relevant works in the past and several problems arise:

i. linguistic variants of the same legal concept inside the agreement/contract text are numerous and they include some overlapping of meaning; Thus, it is hard to understand whether two different words have the same meaning.
ii. Legal provisions sometime are written in passive form in order to make more emphasis on prescriptiveness when addressing the command. This sometimes complicates the extraction of SVO triples.
iii. Legal text has normative references that affect the knowledge extraction.
iv. Legal concepts change over time.
v. frequency is not a good indicator of relevance [34].

The main difference between many classical Open IE techniques and ClausIE is that the latter makes use of the grammatical dependencies extracted through an automatic dependency parser, to identify the SVO triples. ClausIE is able to identify SVO triples, but we need also to correctly associate them to ontology terms and their literal variants provided by the legal expert-team.
Let the GDPR and the Privacy Policies be our corpus C.
In order to perform the automatic text annotation of our corpus with PrOnto concepts, we follow these steps:

1. we firstly identify a list of all the terms (subjects, objects, verbs) in C, by using a simple variant of ClauseIE. The identified terms are said to be possible classes (in the case of subjects and objects) or possible properties (in the case of verbs);
2. we use PrOnto labels of classes and properties, with additional mapping of linguistic and lexical variants;
3. we try to map every possible class/property in C to its closest class/property in PrOnto, by using the same algorithm used in a previous project[4]. This algorithm exploits pre-trained linguistic deep models in order to be able to easily compute a similarity score between two terms.

Ontologies are a formal way to describe classes, relationships and axioms. For this work we focus mainly on classes and properties and their literal forms, without taking into account the other types of knowledge usually coded into an ontology (e.g., Tbox).

[3] https://spacy.io
[4] https://gitlab.com/CIRSFID/un-challange-2019

Furthermore, especially in the case of Privacy Policies, we may expect important concepts not to be distributed uniformly across the whole text. Some important concepts (a.k.a *local*) are usually mentioned only in very specific document areas (e.g., chapters), while others (a.k.a *global*) are scattered throughout the whole text. If the Privacy Policies were marked up using Akoma Ntoso we could use the structure of the XML nodes for better detecting the concepts and properly apply the characteristics *local*/*global*.

6. PrOnto Refinement

The Privacy Policies linguistic analysis with OKE gives some inputs that produce important enhancements in PrOnto ontology.

6.1. Child Class

In the Privacy Policies is frequently mentioned "*child*" that is a particular "data subject" missing in the PrOnto ontology. Initially, we intended to use rules to define child concept because the definition changes for each jurisdiction according to the local implementation of the EU Regulation[5]. However, in light of the important rights and obligations defined in the GDPR for the minors, we decided finally to include a new class in the `Role` module as subclass of `DataSubject`. `Child` class is related with `ParentalResponsabilityHolder` (See Fig. 4).

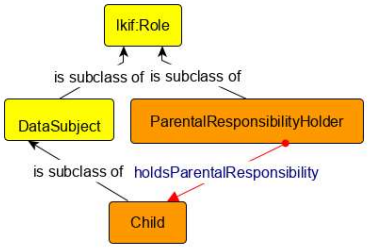

Figure 4: Child class.

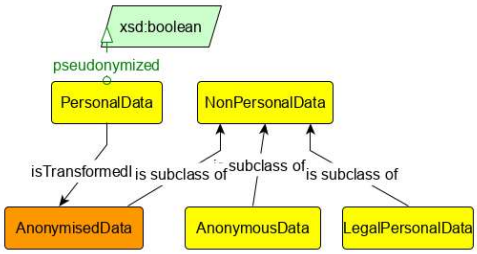

Figure 5: AnonymisedData class.

6.2. Anonymous Data and Anonymized Data Classes

From the Privacy Policies linguistic analysis it emerges that "Anonymised Data"[6] and "Anonymous data" (Recital 26 GDPR)[7] are often misled and confused in the presentation

[5] https://www.betterinternetforkids.eu/web/portal/practice/awareness/detail?articleId=3017751

[6] COM (2019) 250 final "data which were initially personal data, but were later made anonymous. The 'anonymisation' of personal data is different to pseudonymisation (see above), as properly anonymised data cannot be attributed to a specific person, not even by use of additional data and are therefore non-personal data".

[7] Recital 26 GDPR "5. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. 6. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes".

of the data processing. The pragmatic language attempts to simplify the legal terminology and creates a mistake in the conceptualization of those two classes of data that are ontologically different. For this reason, we modelled the relationship `PersonalData isTransformedIn AnonymisedData` in order to clarify the distinction from the legal point of view (See Fig. 5).

6.3. Action Module

The best manner to detect an action is through verbs. However, within OWL ontology, verbs play the role of predicates that connect domain and range. For this reason, the OKE suggested to modify the action's classes with the "ing" form. Some new actions are detected like `Collecting` and `Profiling`. The legal analysis collocates the `Profiling` class as subclass of `AutomatedDecisionMaking` following Art. 22 and the connected Recital 71. In this case, the OKE feedbacks offered a very good input to the legal experts that provided an improvement of the legal ontology by relying on their legal analysis (See Fig. 6).

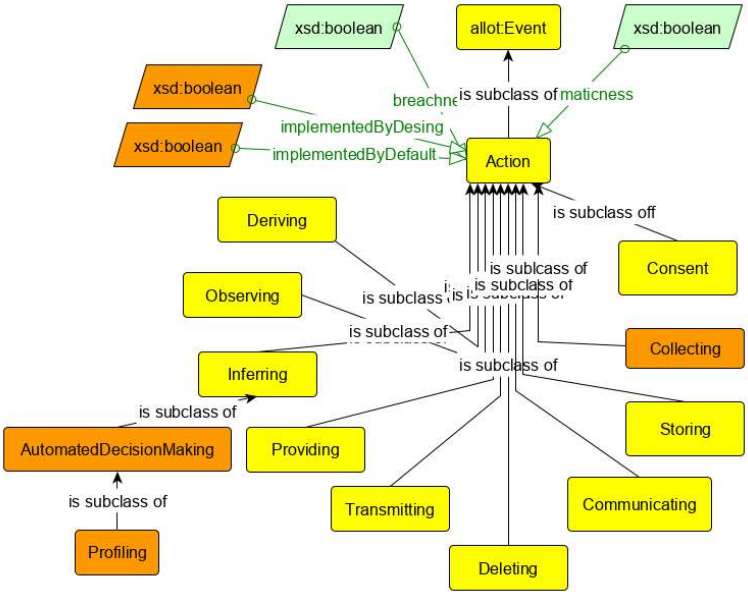

Figure 6: Action Module Refinement

7. Lexicon Modelling in PrOnto

After the validation with OKE, it was evident that it is important to connect the legal concepts to lexical forms. We have chosen to use SKOS and SKOS-XL that is a canonical method for connecting OWL and linguistic variants, using `skosxl:literalform`.

```
    PrOnto:Controller rdf:type owl:Class;
    rdfs:subClassOf PrOnto:Role;
    rdfs:subClassOf skos:Concept.

    PrOnto:DataController rdf:type PrOnto:Controller;
    skosxl:prefLabel PrOnto:controller_1;
    skosxl:altLabel PrOnto:altController_1,PrOnto:altController_2,
    PrOnto:altController_3.

    PrOnto:controller_1 rdf:type skosxl:Label;
    skosxl:literalForm "controller"@en;
    dct: created "2018-05-28"^^xsd:date;
    dct: modified "2019-09-15"^^xsd:date.

    PrOnto:altcontroller_1 rdf:type skosxl:Label;
    skosxl:literalForm "data controller"@en.
    PrOnto:altcontroller_2 rdf:type skosxl:Label;
    skosxl:literalForm "company data controller"@en.
    PrOnto:altController_3 rdf:type skosxl:Label;
    skosxl:literalForm "company that is responsible for your
information"@en.
```

In this manner, it is possible to connect PrOnto Core Ontology with other existing lexicon-controlled vocabulary [18].

8. Related Work

We have at least four main related works to exam in this research.

Privacy ontology. UsablePrivacy and PrivOnto [23] are ontologies oriented to provide linguistic tools in order to define glossaries and taxonomies for the privacy domain, basically starting from the bottom-up annotation of the privacy policies (crowdsourcing annotation). GDPRtEXT [29] lists concepts presented in the GDPR text without really entering the modelling of the norms and the legal axioms (e.g., the actions performed by the processor, the obligations of the controller and the rights of the data subject). GDPRov aims at describing the provenance of the consent and data lifecycle in the light of the Linked Open Data principles such as Fairness and Trust [30]. GConsent is an ontology for modelling the consent action, statement and actors [16]. The SPECIAL Project[8] develops tools for checking compliance in the privacy domain.

Deontic ontology. ODRL provides predicates and classes for managing obligations, permission, prohibitions, but several parts of the deontic logic are missing (e.g., right and penalty classes). LegalRuleML ontology was included in PrOnto.

Lexicon ontology. Controlled vocabularies, thesauri and lexical databases are some examples of linguistic ontologies. They express the terminology concerning a domain of interest by organising terms according to few semantic relations (e.g. hierarchical and associative ones). EUROVOC[9] and IATE[10] are some examples of linguistic ontologies released by the European Union to semantically structure the terminology of documents

[8] https://www.specialprivacy.eu/

[9] https://publications.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc

[10] https://iate.europa.eu/

issued by EU institutions and bodies [33]. However, these resources do not clarify the distinction between legal concepts and their instances [20].
By contrast, the legal domain requires the modelling of legal core concepts, capable to overcome the vagueness of legal jargon that makes the meaning of legal terms subject to interpretation [20]. Thus, the modelling of legal core ontologies is a complex task involving knowledge grounded on legal theory, legal doctrine and legal sociology [10].
Several models have been proposed as natural language interfaces to fill the gap between the high-level ontological concepts and their low-level, context-dependent lexicalisations [22]. In particular, interesting works about SKOS-XL[11][8] and OntoLex [5] [17] are included in this version of PrOnto for combining ontology and linguistic literal forms, in support for NLP and search engines.
Open Knowledge Extraction. Open Information Extraction (Open IE) is to Open Knowledge Extraction (Open KE) as NLP (processing) and NLU (understanding). Open IE is capable to extract information graphs from natural language. Remarkable examples of Open IE tools are ClausIE [9], OpenCeres [21] and Inter-Clause Open IE [1], Open KE builds over Open IE in order to align the identified subjects, predicates and objects (SVOs) to pre-defined ontologies. FRED [13] uses different NLP techniques for processing text and for extracting a raw ontology based on FramNet situations. The challenge of Open KE is that the SVOs alignment requires to understand the meaning of ambiguous and context-dependent terms [35]. The algorithm we designed tackles the Open KE problem by exploiting pre-trained linguistic deep models in order to map information to knowledge.

PrOnto includes an exhaustive strong top-down modelisation reinforced with a bottom-up linguistic approach. This approach guarantees modelisation of institutions of law with a robust theoretical approach not prone to the variants of the language (that can change by country, context, historical period). In the meantime, this work refined the classes (e.g., `Child`), the relationships (e.g., `holdsParentalResponsibility`) and the correlated terminology (e.g., custumer/user) using the OKE.

9. Conclusions and Future Work

We have validated the PrOnto ontology with a sample of Privacy Policies and with a robust legal analysis following the MeLOn methodology, in order to manually check the completeness of the classes and relationships for representing the main content of the policies texts. This exercise detected some new needs in the PrOnto ontology (e.g., the `LegalBasis` module) that originally the team decided to not to include. The legal team detected some inconsistencies in the terminologies between the legislative text and the pragmatic language. For this reason, the legal team produced a map of lexicon variants, then modelled using SKOS-XL. PrOnto and these extensions fill up an Open Knowledge Extraction algorithm to detect concepts in the Privacy Policies. The method was iterated three times and at the end we obtained an increase of 29% in the detection of the concepts respect the first interaction that record an increase of 19%. We are capable to detect the 75% of the concept in the new privacy policies using the new version of PrOnto enriched with SKOS-XL terms. In the future, we intend to perform the same experiment using Consent Statements and also Code of Conducts. This work confirmed the robustness of

[11] https://www.w3.org/TR/skos-reference/skos-xl.html

PrOnto main modules, pattern-oriented and aligned with foundational ones, and in the future this work will be used in order to validate (e.g., with different type of legal documents), refine (e.g., extend with new modules like national customised-US), update (e.g., due to legislative modifications of the GDPR) the PrOnto schema design. This method is also relevant to annotate legal texts with PrOnto and so to create RDF triples for supporting applications (e.g., search engine, legal reasoning, checking compliance).

Acknowledgements

This work was partially supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 690974 "MIREL: MIning and REasoning with Legal texts".

References

[1] Angeli, G., Premkumar, M. J. J., & Manning, C. D., 2015. Leveraging linguistic structure for open domain information extraction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 344-354.

[2] Ashley, K.D., 2017. Artificial intelligence and legal analytics: new tools for law practice in the digital age. Cambridge Univ Press, Cambridge New York Melbourne Delhi Singapore.

[3] Bandeira, J., Bittencourt, I.I., Espinheira, P., Isotani, S., 2016. FOCA: A Methodology for Ontology Evaluation. Eprint ArXiv.

[4] Barabucci, G., Cervone, L., Di Iorio, A., Palmirani, M., Peroni, S., Vitali, F., 2010. Managing semantics in XML vocabularies: an experience in the legal and legislative domain. In Proceedings of Balisage: The markup conference (Vol. 5)

[5] Bosque-Gil, J., Gracia, J., Montiel-Ponsoda E., 2017. Towards a module for lexicography in OntoLex. In Proc. of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets at 1st Language Data and Knowledge conference (LDK 2017), Galway, Ireland, vol. 1899. CEUR-WS, 2017, pp. 74-84.

[6] Breuker, J., Hoekstra, R., Boer, A., van den Berg, K., Sartor, G., Rubino, R., Wyner, A., Bench-Capon, T., Palmirani, M., 2007. OWL Ontology of Basic Legal Concepts (LKIF-Core), Deliverable No. 1.4. IST-2004-027655 ESTRELLA: European project for Standardised Transparent Representations in order to Extend Legal Accessibility.

[7] Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., ... & Sung, Y. H., 2018. Universal sentence encoder. arXiv preprint arXiv:1803.11175.

[8] Declerck, T., Egorova, K., & Schnur, E., 2018. An Integrated Formal Representation for Terminological and Lexical Data included in Classification Schemes. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018).

[9] Del Corro, L., Gemulla, R., 2013. Clausie: clause-based open information extraction. In: Proceedings of the 22nd international conference on World Wide Web. ACM, 2013, pp. 355-366.

[10] Fernández-Barrera, M., & Sartor, G., 2011. The legal theory perspective: doctrinal conceptual systems vs. computational ontologies. In Approaches to Legal Ontologies. Springer, Dordrecht, pp. 15-47.

[11] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L., 2002. Sweetening Ontologies with DOLCE. In International Conference on Knowledge Engineering and Knowledge Management Springer, Berlin, Heidelberg, pp. 166-181.

[12] Gangemi, A., Peroni, S., Shotton, D., Vitali, F., 2017. The Publishing Workflow Ontology (PWO). Semantic Web 8, pp. 703–718. https://doi.org/10.3233/SW-160230

[13] Gangemi, A., Presutti, V., Reforgiato Recupero, D., Nuzzolese, A. G., Draicchio, F., & Mongiovì, M., 2017. Semantic web machine reading with FRED. Semantic Web, 8(6), pp .873-893

[14] Guarino, N., Welty C.A., 2004. An Overview of OntoClean. In Handbook on ontologies. Springer, Berlin, Heidelberg, pp. 151-171.

[15] Hitzler, P., Gangemi, A., Janowicz, K., Krisnadhi, A. (Eds.), 2016. Ontology engineering with ontology design patterns: foundations and applications, Studies on the semantic web. IOS Press, Amsterdam Berlin.

[16] http://openscience.adaptcentre.ie/ontologies/GConsent/docs/ontology

[17] http://www.w3.org/2016/05/ontolex

[18] https://www.w3.org/ns/dpv#data-controller

[19] IFLA Study Group on the Functional Requirements for Bibliographic Records, 1996. Functional Requirements for Bibliographic Records, IFLA Series on Bibliographic Control. De Gruyter Saur.

[20] Liebwald, D., 2012. Law's capacity for vagueness. International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 26(2), pp. 391-423
[21] Lockard, C., Shiralkar, P., & Dong, X. L., 2019. OpenCeres: When Open Information Extraction Meets the Semi-Structured Web. In 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, pp. 3047-3056.
[22] McCrae, J., Spohr, D., Cimiano, P., 2011. Linking Lexical Resources and Ontologies on the Semantic Web with Lemon. In: Antoniou G. et al. (eds) The Semantic Web: Research and Applications. ESWC 2011. Lecture Notes in Computer Science, vol 6643. Springer, Berlin, Heidelberg.
[23] Oltramari, A., Piraviperumal, D., Schaub, F., Wilson, S., Cherivirala, S., Norton, T.B., Russell, N.C., Story, P., Reidenberg, J., Sadeh, N., 2016. Privonto: A semantic framework for the analysis of privacy policies. Semantic Web (1-19).
[24] Palmirani, M., Bincoletto, G., Leone, V., Sapienza, S., Sovrano, F., 2019. PrOnto Ontology Refinement Through Open Knowledge Extraction. Jurix 2019, pp. 205-210
[25] Palmirani, M., Governatori, G., 2018. Modelling Legal Knowledge for GDPR Compliance Checking. JURIX 2018, pp. 101-110.
[26] Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L., 2018. PrOnto: Privacy Ontology for Legal Reasoning. EGOVIS2018, 7th International Conference, EGOVIS 2018, Regensburg, Germany, September 3-5, 2018, Proceedings. LNCS 11032, Springer, pp. 139-152.
[27] Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L., 2018. Legal Ontology for Modelling GDPR Concepts and Norms. JURIX 2018. pp. 91-100.
[28] Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L., 2018. PrOnto: Privacy Ontology for Legal Compliance. In: Proceedings of the 18th European Conference on Digital Government ECDG 2018, Reading UK, Academic Conferences and Publishing International Limited, 2018, pp. 142 – 151.
[29] Pandit, H.J., Fatema, K., O'Sullivan, D., Lewis, D., 2018. GDPRtEXT - GDPR as a Linked Data Resource. In Gangemi A. et al. (eds) The Semantic Web. ESWC 2018. Lecture Notes in Computer Science, vol 10843. Springer, Cham.
[30] Pandit, H.J., Lewis, D., 2017. Modelling Provenance for GDPR Compliance using Linked Open Data Vocabularies. In Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) co-located with the 16th International Semantic Web Conference (ISWC 2017), http://ceur-ws.org/Vol-1951/PrivOn2017_paper_6.pdf.
[31] Peroni, S., Palmirani, M., Vitali, F., 2017. UNDO: The United Nations System Document Ontology, in: d'Amato, C., Fernandez, M., Tamma, V., Lecue, F., Cudré-Mauroux, P., Sequeda, J., Lange, C., Heflin, J. (Eds.), The Semantic Web – ISWC 2017. Springer International Publishing, Cham, pp. 175–183. https://doi.org/10.1007/978-3-319-68204-4_18
[32] Rossi, A., Palmirani, M., 2019. DaPIS: an Ontology-Based Data Protection Icon Set. In G. Peruginelli and S. Faro (Eds.): Knowledge of the Law in the Big Data Age. Frontiers in Artificial Intelligence and Applications. Vol. 317. IOS Press.
[33] Roussey, Catherine, Pinet, F., Kang, M. A., & Corcho, O., 2011. An introduction to ontologies and ontology engineering. In Ontologies in Urban development projects. Springer, London, pp. 9-38.
[34] van Opijnen, M.and Santos, C., 2017. On the concept of relevance in legal information retrieval, Artificial Intelligence and Law 25: 65. https://doi.org/10.1007/s10506-017-9195-8
[35] Welty, C., and Murdock, J. W., 2006. Towards knowledge acquisition from information extraction. In *International Semantic Web Conference*. Springer, Berlin, Heidelberg, pp. 709-722.
[36] Wilson, S., Schaub, F., Liu, F., Sathyendra, K. M., Smullen, D., Zimmeck, S., Rohan Ramanath, Story, P., Liu, F., Sadeh, N., Smith, N. A., 2018. Analyzing Privacy Policies at Scale: From Crowdsourcing to Automated Annotations. ACM Transactions on the Web, 13, 1.

APPENDIX

Technical results and measurement.

First Set of Privacy Policy

PrOnto Version	SKOS support	Found Ontology Concepts	Ontology Concepts	Presence %	Increasing %
8	No	96	139	69,0647482	
8	Yes	101	139	72,6618705	

9	No	119	172	69,18604651	19% *
9	Yes	120	172	69,76744186	20% *

Second Set of Privacy Policy

PrOnto Version	SKOS support	Found Ontology Concepts	Ontology Concepts	Presence %	Increasing %
8	No	106	139	76,25899281	
8	Yes	109	139	78,41726619	
9	No	129	172	75	29% *
9	Yes	129	172	75	29% *

* The increment is respect version 8 of the ontology.