

Internship position at CNAM Paris on Semantic Web approaches for documents graph access

Host ISID team at Centre d'études et de recherche en informatique et communications (CEDRIC), CNAM Paris

Starting date From February 2021

Duration 6 months

Gratification around 600 € / month

*** Candidate profile***

As a minimum requirement, the successful candidate should have:

- A master degree in one or more of the following areas: symbolic AI, semantic web, graph databases, web development.
- Excellent programming skills (Java or Python)
- Excellent command of English
- Experience with semantic web technologies and web development

How to apply

The application should be formatted as ****a single pdf file**** and should include:

- A complete and detailed curriculum vitae
- A cover letter
- a report you wrote (ideally an internship or project report)
- The content of M1 and M2 courses and the corresponding grades
- The contact of two referees and a recommendation letter if possible

The pdf file should be sent to nada.mimouni@cnam.fr

Keywords semantic web, ontologies, graph databases, interactive user interfaces

Supervision Nada Mimouni (Cnam), Fayçal Hamdi (Cnam), Thomas Francart (Sparna)

Context

In various domains, where documents represent the main medium of information and where usually various links exist between these documents, a big challenge is to access an often complex and multidimensional information.

Semantic web technologies are used as an effective solution for modeling and querying these large collections of interrelated texts. However, in the absence of simple tools, the use of these technologies remains restricted to a few knowledgeable users.

The objective of the internship is to design and implement a tool to help discover, access and process textual data available on law digital portals. The application domain of this project is the legal domain, for which information access is of great complexity for both practitioners and ordinary citizens.

An ontological resource to model these collections has been designed and enriched with data extracted from the Legifrance portal and will be used as a starting point for the implementation of the tool.

Objective

More concretely, the recruited intern must address the following points :

- Enrich an existing knowledge base with texts from legal digital portals.
- Propose and implement a semantic approach for storing and indexing textual data into a graph knowledge base.
- Use predefined patterns to generate queries on stored data.
- Integrate the querying process into an interactive user interface. The intern can study if the component “Sparnatural” could be an appropriate solution (<http://sparnatural.eu>).
- Evaluate the usefulness of the results produced compared to those generated by public digital portals.

References

Mimouni, N., Nazarenko, A. and Salotti, S. Answering Complex Queries on Legal Networks: A Direct and a Structured IR Approaches. In *AI Approaches to the Complexity of Legal Systems. AICOL 2015, 2016, 2017, Lecture Notes in Computer Science*, vol 10791. Springer., pages 451-464, 2018.

Mimouni, N., Nazarenko, A., Paul, E. and Salotti, S. Towards Graph-based and Semantic Search In Legal Information Access Systems. In *Twenty-Seventh Annual Conference on Legal Knowledge and Information Systems (JURIX 2014)*, pages 163-168, Krakow, Poland, 2014.

Alschner, Wolfgang, *AI and Legal Analytics* (November 2, 2020). in Florian Martin-Bariteau & Teresa Scassa, eds., *Artificial Intelligence and the Law in Canada* (Toronto: LexisNexis Canada, 2021), Available at SSRN: <https://ssrn.com/abstract=3733957>.

Tarasconi, F., Botros, M., Caserio, M., Sportelli, G., Giacalone, G., Uttini, C., Vignati, L., and Zanetta, F. (2020). *Natural Language Processing Applications in Case-Law Text Publishing. JURIX*.

Chalkidis, I., Kampas, D. Deep learning in law: early adaptation and legal word embeddings trained on large corpora. *Artificial Intelligence and Law* 27, 171–198 (2019). <https://doi.org/10.1007/s10506-018-9238-9>.

Francesconi, Enrico and Küster, Marc and Gratz, Patrick and Thelen, Sebastian. (2015). *The Ontology-Based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services*. 9265. 29-39. [10.1007/978-3-319-22389-6_3](https://doi.org/10.1007/978-3-319-22389-6_3).

Leone V, Di Caro L, Villata S. Taking stock of legal ontologies: a feature-based comparative analysis. *Artificial Intelligence and Law*. 2019 Jun 13:1-29.

Francart T, Dann J, Pappalardo R, Malagon C, Pellegrino M. The European Legislation Identifier. *Knowledge of the Law in the Big Data Age*. 2019 Jul 23;317:137-48.

Bibal, A., Lognoul, M., de Streel, A. et al. Legal requirements on explainability in machine learning. *Artificial Intelligence and Law* (2020). <https://doi.org/10.1007/s10506-020-09270-4>.

B. Walzl, G. Bonczek, E. Scepankova and F. Matthes, Semantic types of legal norms in German laws: classification and analysis using local linear explanations. *Artificial Intelligence and Law*. 2019 27 (1), 43-71.

Agnoloni, Tommaso, Lorenzo Bacci, Ginevra Peruginelli, Marc van Opijnen, Jos van den Oever, Monica Palmirani, Luca Cervone et al. "Linking European Case Law: BO-ECLI Parser, an Open Framework for the Automatic Extraction of Legal Links." In *Jurix*, pp. 113-118. 2017.

Medvedeva, M., Vols, M. and Wieling, M. Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law* 28, 237–266 (2020). <https://doi.org/10.1007/s10506-019-09255-y>.