

Providing Actionable Insights for Jurisprudence Researchers

Jamie Schram¹, Christian Dirschl², Jessica Kent¹, Quentin Reul¹, Vincent Henderson³
and Harry Sabnani¹

¹ Wolters Kluwer R&D U.S. LP, Riverwoods, USA

² Wolters Kluwer Deutschland GmbH, Munich, Germany

³ Wolters Kluwer France, Paris, France

{Firstname.Lastname}@wolterskluwer.com

Abstract. Attorneys and legal practitioners spend inordinate amounts of time reading case-law documents, trying to find relevant, precedential or exemplary decisions that support particular patterns of claims made in adjudicatory matters on behalf of their clients. To ameliorate this ubiquitous problem, we have crafted a Legal and Regulatory domain-specific ontology that works in tandem with our enterprise upper ontology. Up until recently, attorneys have relied and trusted books (including digital books) over more modern ways of consuming information. Since the start of the COVID-19 pandemic, there has been an erosion of the belief in the need for information to be delivered in book form [2]. As legal professionals are used to retrieving granular information in their personal search engine of choice (most start with Google), they come to expect similar capabilities from their legal search engine, which requires extracting more domain-specific insights from jurisprudence when using research products in their daily activities. Specifically, the metadata that is traditionally represented in our content is descriptive of, and specifically focused on, the document as the canonical subject. Rather than focusing on the document, we instead employ an abstract, information-centric representation of the specific semantic units that are realized in the text of jurisprudence documents (such as claims made by the litigants, facts of the case, etc.). We will describe some of the challenges we have faced, and lessons learned, in moving from a more “traditional” document-based mantra of enrichment to more domain-specific semantics.

Keywords: Legal Ontology, Knowledge Extraction, Semantic Search.

1 Problem description

Wolters Kluwer is a global company which provides “professional information, software solutions, and services” [1]. Our expert solutions use a combination of domain knowledge, advanced technology, and services for a variety of fields including legal and regulatory, medical, tax and accounting, and governance, risk, and compliance. A large segment of our Wolters Kluwer customer base in the legal and regulatory field

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[12] is tasked with a complex research challenge: jurisprudence (case-law) research. Attorneys and legal practitioners spend inordinate amounts of time reading case-law documents, trying to find relevant, precedential or exemplary decisions that support particular patterns of claims made in adjudicatory matters on behalf of their clients. As long as the main business model was to bill by hours, this did not have a negative effect on revenue. Due to changed requirements coming from the customer, who prefer more and more fixed price mandates (which can also be more easily compared with offerings from competitors), this effort has all of a sudden a major effect and optimization of processes and tasks is playing a critical role in business success.

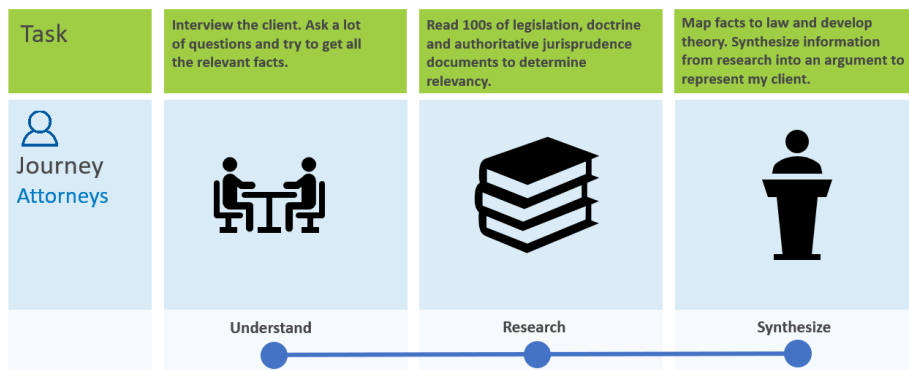


Fig. 1. Journey of the Attorney in constructing a legal argument on behalf of a client

During the Research portion of the journey, relevance, accuracy and speed is critical. “Traditional” searches present a number of unique challenges for the attorney:

- Searching by keywords returns a majority of irrelevant jurisprudence where the judge’s decision has nothing to do with the subject. **The lawyer must read the document to know if it’s relevant.**
- Searching jurisprudence by article of law returns mostly irrelevant documents where cited articles are not the basis for the decision. **The lawyer needs to read each document to determine whether the legal ground is pertinent.**
- A traditional search may sort cases by date; however, **the most recent case may not really be the most relevant.**

2 Solution description

For lawyers who need to craft a legal strategy, Wolters Kluwer aims to deliver a legal intelligence solution that takes any description of a legal matter as input to match to relevant jurisprudence documents based on extracted insights (e.g. claims, legal arguments, etc.), so that the lawyer can select effective arguments. Unlike current full-text search engines, our solution uses jurisprudence semantics to organize information and provide quantitative data to support the goals of the lawyer. This upholds our

company strategy to build expert solutions that actively contribute to the professional goals of users [11]. Note that this endeavor extends and builds upon the work that was done in Wolters Kluwer Germany to create a knowledge graph-based search engine designed for German court case data [3]. The knowledge graph-based search was aimed at surpassing traditional keyword matching by identifying legal concepts in search queries and map these directly with legal concepts in the documents. This approach had its limitations in that a legal concept in the query could be found in different parts of the document and depending on the semantics of this document section, the retrieved document could be relevant or not. And this is where the notion of the semantic unit comes into play. A more NLP based focus in the LYNX project [10] seems to be a complementary effort to our work.

Using a combination of an enterprise ontology for Wolters Kluwer based on industry-standard technologies such as RDF [7] and the Web Ontology Language [OWL] [6], a Knowledge Graph [5] and Machine-Learning techniques, we provide a solution that can identify and understand the relationship between various semantic units intrinsic to case-law documents (i.e., facts, claims, legal grounds, decisions, etc.). This then allows our customers to provide a natural language description of a client's factual circumstances giving rise to a legal risk or conflict, and to receive insights into case-law that most closely meets their specific situation, and more specifically to the outcomes of those related case-law documents, thus assisting the professional in determining his strategy for the course of action to adopt relative to the specific merits of the matter at hand.

So the solution at hand enables and enhances both use cases: searching via legal concepts and searching via a natural-language based case description.

2.1 Modeling jurisprudence semantics using the enterprise ontology

We have created an extension to our Wolters Kluwer enterprise ontology to define concepts covering the Legal & Regulatory domain and the relationships between them [4].

Wolters Kluwer solutions must have similar capabilities to Google and other information retrieval sources as professionals (legal and otherwise) as well as the general public are accustomed to fast, accurate information from their personal search engine of choice (most start with Google). To provide these capabilities, our solutions require the extraction of more domain-specific insights during their use of the research products.

Traditionally the metadata represented in our content is descriptive of, and specifically focused on, the document as the canonical subject. This traditional approach is insufficient for the expected capabilities of modern-day jurisprudence research. Rather than focusing on the document, we instead employ an information-centric representation of the specific "semantic units" that are realized in the text of jurisprudence documents.

For example, in order to formulate an effective legal argument on behalf of a client, an attorney researching jurisprudence must be able to ascertain and understand the discrete claims that are made by the litigants, the facts of the case, and any legal

grounds that are used to establish legal precedent for the arguments made in the case. The researcher must then clearly understand the relationships between these legal aspects as the foundation for the judge's reasoning behind the decision rendered in the case, and subsequently any remedies that are ordered in association with that decision. For an illustration of the semantic units used in a determination, see figure 2.



Fig. 2. Examples of Semantic Units in Jurisprudence

As noted above, attorneys are increasingly looking for fast, intelligent solutions that match the patterns unique to their specific judicial matter to relevant case-law throughout the corpus of a given legal domain and jurisdiction. The driving factor to accelerate legal research for the attorney is the fact that as mentioned above billing has changed.

Creating a model to represent jurisprudence semantics in this manner leads to many challenges, such as logistics and scale. Specifically, we need to provide a model that can represent jurisprudence across multiple countries, and accounts for text in multiple languages. Every judicial model in Europe (and throughout the world, for that matter) shares commonalities, but also has its own unique attributes. Our approach for the design and management of the ontology aims to strike a balance between a very loose semantic model, which provides great flexibility and extensibility (based primarily on making use of SKOS [9] for controlled vocabularies, and making use of SHACL [8] as a mechanism to enforce business-specific constraints) and is easier for general business users to understand, and a more semantically precise ontological model which tends to be more practical for developers to work with.

2.2 Initial Approach

At the outset of our program, we approached the problem with respect to the enterprise ontology extension from a somewhat “traditional” standpoint. Specifically, we considered the discrete semantic units, such as claims, facts, legal grounds, etc. to be represented as fragments of text within documents, where those fragments were specifically “typed” making use of pre-defined SKOS controlled vocabularies. This approach allowed us to use the enterprise ontology largely “out-of-the-box”, with only minimal extension.

Using a combination of manual and ML-based techniques, we were able to recognize relevant sentences in jurisprudence documents and map them to this model.

While this approach provided a great deal of flexibility in the modeling, and was relatively easy for Subject Matter Experts to understand, it also had a number of limiting aspects:

- The model was “document-centric”;
 - Representing our jurisprudence semantic units (i.e., claims, legal grounds, facts, etc.) as simple typed fragments within the boundaries of documents makes it more difficult to recognize and map the same logical unit realized in other documents, where the wording and phrasing may be completely different, or even represented in a different language.
- The model provided indirect and imprecise semantics (i.e., defining semantics via SKOS v. OWL);
 - Rather than working with semantic units that employ first-class semantic typing, (i.e., this is a Claim, this is a Fact, etc.) we indirectly modeled those semantics in terms of fragment types (i.e., this is a fragment in a document, which happens to be typed as a Claim, etc.)
- The model was not easily understood, and perhaps more importantly, not easily actionable by the developers who were building the expert solution for the reasons noted in the previous points.

2.3 Revised Approach

We soon realized that while the original modeling paradigm had some advantages, we needed something that was more robust and more semantically precise. As a result, we pivoted to an approach that defined jurisprudence semantics as a set of domain-specific extension classes in a Legal & Regulatory Jurisprudence ontology extension. This allowed us to define jurisprudence semantic units such as Facts, Claims, Legal Grounds, etc. as “first class” semantic units, rather than having them represented solely as typed fragments buried in document text.

These ontology extensions represent semantic units (instances of OWL Classes) for claims, legal grounds, facts, etc., which are realized in jurisprudence document text, but we desire to manage as normalized, document-agnostic abstract objects. For example, we recognized that the same normalized object (Fact, Claim, etc.) may be realized in multiple jurisprudence documents, but the language in the text may be completely different. This shift in modeling allowed us to represent those objects

completely independent of the documents from which they may happen to appear. This allowed us to establish semantic relationships between these normalized semantic units, based on applied business logic and subject matter expert insight, which then helps to drive the application to provide actionable insights for the attorney, such as profiles of what sets of jurisprudence content, regardless of phrasing or terminology, most closely match the unique factual situation of their clients, and insights into how arguments based on that jurisprudence have fared, i.e., how often did the plaintiff win such a case, etc.

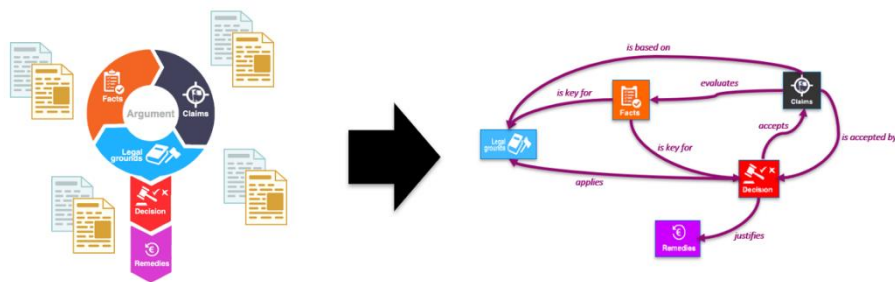


Fig. 3. Jurisprudence Semantics Represented as Document-Independent Objects with Interconnected Semantic Relationships

3 Lessons Learned

Semantics are crucial to driving new, innovative methods of search

- Depending on what capabilities you want to provide, your semantic modeling choices need to facilitate them;
- Find the sweet spot – strike balance between flexibility and reuse of the core ontology with formal semantics of ontology extensions.

The revised approach has helped us to better achieve our vision:

- More precise, normalized jurisprudence semantics allows us to drive an expert solution that can provide more value than traditional research products;
- Provide more than just document results, e.g.:
 - Quantitative analysis
 - Actionable insights

This also enables the creation of APIs from the ontology that could be used by developers to generate RDF triples programmatically.

4 References

1. About Us. (n.d.), Wolters Kluwer Homepage, <https://www.wolterskluwer.com/en/about-us>, last accessed 2020/11/19
2. Bonacum, L. (2020, June 5). 2020 Wolters Kluwer future ready lawyer: Performance drivers and change in the legal sector, Wolters Kluwer Homepage, <https://www.wolterskluwer.com/en/news/2020-wolters-kluwer-future-ready-lawyer-performance-drivers-and-change-in-the-legal-sector>, last accessed 2020/11/19
3. Crotti Junior, A., Orlandi, F., Graux, D., Hossari, M., O'Sullivan, D., Hartz, C., Dirschl, C., Knowledge Graph-based Legal Search over German Court Cases. To be published in: A. Harth et al. (Eds.): ESWC 2020 Satellite Events, LNCS 12124, 2020
4. Crotti, A., Jr., Orlandi, F., O'Sullivan, D., Dirschl, C., & Reul, Q., Using Mapping Languages for Building Legal Knowledge Graphs from XML Files. In: Proceedings of the Contextualized Knowledge Graphs (CKG) Workshop co-located with the 18th International Semantic Web Conference (ISWC 2019)
5. Dirschl, C., Kent, J., Schram, J., Reul, Q., Enabling Digital Business Transformation through an enterprise Knowledge Graph. To be published in: A. Harth et al. (Eds.): ESWC 2020 Satellite Events, LNCS 12124, 2020
6. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2012, December 11). OWL 2 Web Ontology Language Primer (Second Edition), <https://www.w3.org/TR/owl2-primer/>, last accessed 2020/11/19
7. Klyne, G., Carroll, J. J., & McBride, B. (2004, February 10). Resource Description Framework (RDF): Concepts and Abstract Syntax, <https://www.w3.org/TR/rdf-concepts/>, last accessed 2020/11/19
8. Knublauch, H., & Kontokostas, D. (2017, July 20). Shapes Constraint Language (SHACL), <https://www.w3.org/TR/shacl/>, last accessed 2020/11/19
9. Miles, A., & Bechhofer, S. (2009, August 18). SKOS Simple Knowledge Organization System Reference, <https://www.w3.org/TR/skos-reference/>, last accessed 2020/11/19
10. Moreno-Schneider, J., Rehm, G., Montiel-Ponsoda, E., Rodriguez-Doncel, V., Revenko, A., Karampatakis, S., Khvalchik, M., Sageder, C., Gracia, J., Maganza, F., Orchestrating NLP Services for the Legal Domain. In: N. Calzolari et al. (Eds.): Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020), p. 2325-2333, Marseille, France, 2020. European Language Resources Association (ELRA)
11. Our Strategy. (n.d.), Wolters Kluwer Homepage, <https://www.wolterskluwer.com/en/about-us/strategy>, last accessed 2020/11/19
12. Wolters Kluwer Homepage. (n.d.), <https://www.wolterskluwer.com/en>, last accessed 2020/11/19