Find Relevant Cases in All Cases: Your Journey at Doctrine

Nicolas Fiorini
Doctrine
24 rue du Mail, 75002 Paris
nicolas.fiorini@doctrine.fr

ABSTRACT

Domain-specific Information Retrieval (IR) is generally challenging because of the rare datasets or benchmarks, niche vocabularies and more limited literature coverage. Legal IR is no exception and presents other obstacles, reinforcing the need for innovation and, sometimes, paradigm shifts.

Doctrine, one of the largest Legaltech companies in Europe, dedicates an entire data science team to advance on these problems and identify new opportunities. In this presentation, we provide some intuition regarding the specificities of legal IR (e.g., what is relevance?), and we introduce some of the solutions currently used on doctrine.fr.

Particularly, we show how we use named entity recognition in the various forms of contents we host, and how it enhances the search engine. With knowledge extracted from documents, we may built large enough datasets and train learning-to-rank algorithms. This, combined with several specific-domain vocabulary enrichments to increase recall, dramatically improves the search experience for our users.

CCS Concepts/ACM Classifiers

• Information systems~Retrieval models and ranking
• Information systems~Learning to rank
• Information systems~Structured text search
• Information systems~Data mining

Author Keywords: Legal IR; Learning-To-Rank

BIOGRAPHY

Nicolas Fiorini is the Data Chapter Lead at Doctrine, where he is responsible for the strategic vision and skillset of the Data Science Chapter. He did a PhD in Computer Science at the University of Montpellier and Ecole des Mines d'Alès, France, focusing on generic approaches for indexing and clustering documents annotated with knowledge base concepts. He then joined the National Center for Biotechnology Information at the National Institutes of Health to work on PubMed, where he developed the "Best Match" algorithm — a relevance ranking model for biomedical papers.

SIGIR '19, July 21–25, 2019, Paris, France.

ACM ISBN 978-1-4503-6172-9/19/07.
DOI: https://doi.org/10.1145/3331184.3331441