

ABSTRACT

The increasing digitization of judicial systems has resulted in the rapid growth of legal information in the form of judicial metadata, court records, and legal judgments. While these resources provide valuable insights into judicial institutions and legal decision-making, much of the available information remains embedded within unstructured textual documents. Consequently, the effective representation, organization, retrieval, and analysis of judicial information have emerged as important challenges in legal informatics and judicial analytics. Recent advances in semantic technologies, knowledge graphs, Large Language Models (LLMs), and representation learning provide new opportunities to transform such information into machine-interpretable forms that support intelligent analysis and knowledge discovery. Within this context, this thesis investigates computational approaches for analysing two major components of the judicial ecosystem: judicial personnel and judicial documents.

The research begins by investigating the semantic representation and graph-based analysis of judicial metadata. To address the challenge of transforming unstructured judicial information into structured knowledge, a citizen-focused Judicial Information Ontology (JIO) is developed using metadata associated with judges of the Supreme Court of India. The ontology provides a formal representation of judicial personnel information and supports semantic querying, reasoning, and knowledge discovery through ontology-based technologies. To further explore structured representations of judicial information, the thesis investigates the automatic construction and analysis of judicial knowledge graphs derived from textual data. A reliability-aware framework is proposed that integrates LLM-based triple extraction, relation-level reliability estimation, knowledge graph construction, multiplex network modelling, community detection, graph representation learning, and metadata prediction. The framework introduces a Relation Reliability Weighted Projection (RRWP) formulation that explicitly incorporates extraction reliability into graph construction. Experimental results demonstrate that extraction uncertainty influences graph topology, community organization, and representation learning outcomes, highlighting the importance of reliability-aware graph construction for downstream analytical tasks.

Attention then shifts to judicial case documents, with a focus on understanding relationships among legal judgments. A comparative framework is developed to analyse judicial similarity from both citation-based and semantic perspectives. Using a corpus of Indian Supreme Court

judgments, the study investigates how citation overlap and textual representations capture different dimensions of relatedness among legal cases. The findings reveal substantial differences between citation-based and semantic similarity patterns, suggesting that these approaches provide complementary views of legal relationships and can jointly support legal information retrieval and precedent discovery.

Since semantic representations are influenced by the quality and form of the underlying textual content, the thesis further investigates the applicability of modern language models for the abstractive summarization of judicial documents. A large-scale empirical evaluation involving multiple state-of-the-art LLMs and ensemble variants is conducted on judicial judgments. Since manually annotated reference summaries are generally unavailable for large legal corpora, a fully automated reference-free evaluation framework is proposed. The framework combines lexical overlap, semantic alignment, and qualitative assessment measures to evaluate relevance, coherence, consistency, fluency, and semantic preservation. The results reveal substantial variation in summarization behaviour across models and demonstrate that recent long-context and reasoning-oriented models achieve the most balanced performance. The study also shows that summarization influences the semantic representation of legal documents and affects their subsequent organization within embedding spaces.

The effects of representation become particularly evident when judicial documents are organized into larger collections. To examine this relationship, the final study investigates the use of alternative semantic representations for the unsupervised clustering of judicial documents. A multi-view representation learning framework is proposed in which multiple LLM-generated summaries are treated as complementary semantic views of the same collection of legal judgments. These representations are transformed into vector embeddings using different sentence embedding models and subsequently analysed using multiple clustering algorithms. Cluster quality is evaluated using internal validation metrics together with neighbourhood-based semantic coherence analysis. The results indicate that summary-driven representations can effectively preserve semantically discriminative information while reducing textual noise, leading to more coherent clustering structures. The findings further demonstrate that the choice of summarization model, embedding technique, and clustering algorithm significantly influences cluster organization and semantic consistency.

A central finding emerging from this thesis is that the usefulness of judicial information depends fundamentally on how that information is represented. The studies demonstrate that

alternative representations, ranging from ontological concepts and knowledge graphs to citation networks, summaries, and semantic embeddings, reveal different structural and semantic characteristics of judicial data. Consequently, representation should not be viewed merely as a preprocessing step but as a critical component that shapes retrieval, reasoning, similarity assessment, clustering, and knowledge discovery. By integrating semantic technologies, LLMs, graph learning, and representation learning within a unified analytical framework, this thesis advances the development of intelligent methods for understanding and organizing judicial information. The resulting methodologies provide both practical and methodological foundations for future judicial information systems, legal analytics platforms, and data-driven approaches to legal knowledge discovery.

Keywords: Judicial Analytics, Ontology Engineering, Knowledge Graphs, Graph Representation Learning, Large Language Models, Legal Document Summarization, Semantic Similarity, Multi-view Representation Learning, Semantic Embeddings, Document Clustering