Some Observations on the Nature and Computer-based Representation of Legal Documents and Their Link Data

John Murphy, Robert Steele

University of Sydney, Sydney, Australia

Abstract. Knowledge is recorded and transmitted in the legal domain through well structured documents containing data and meta-data in specific document parts that serve an administrative or procedural purpose. Analysis of this data and its relationships with its containing document parts provides valuable information that can be used by information systems to assist legal practitioners with searching and understanding of legal concepts. This paper analyzes the characteristics of the structures and linkages within legal documents and provides a categorization of the different types of data and explanation of how these types of data interrelate. The paper further demonstrates the identification of data in an example case report document and how the data may be combined to provide information that assists the evaluation of the relevance and importance of legal concepts contained in the document.

Keywords: Information Systems, Knowledge Representation

1. Introduction

Analysis of the nature and use of legal data reveals rich and interesting document structures that have evolved from paper based information systems that communicate legal concepts and enable legal practitioners to understand the complex network of laws contained within large repositories of legal documents. These systems contain both legal concepts and procedural and administrative information within richly structured documents in the form of case reports and legislation supported by secondary documents such as legal thesauri and digests. Legal documents contain data and meta-data, rich document structures and citation links that can be identified and used by computer based information systems to improve the performance of research tasks by legal practitioners. Much knowledge engineering work has been done to classify and encode legal knowledge into ontologies for the purpose of developing systems for decision support [1] and the promotion of systems interoperability [2] with recent development of high level legal ontologies engineered for extension by application developers. This paper analyses the data contained in legal documents and provides categorizations for important characteristics of the data for the purpose of describing how this data can be utilized by legal information systems.

2. Legal Data for Information Systems

Legal data can be broadly conceptualized at two levels:

1. A conceptual or ontological level that includes data that describes the information in documents used for procedural purposes, taxonomies defining shared terminology and the structure of legal documents and the relations between documents.

2. Instance level data that is the content of the concrete documents such as legislation and case reports.

Conceptual level data can be further categorized into the purposes it serves within the legal system. These include:

1. Data that describes the information in documents used for procedural purposes.

2. Data that classifies the concepts of law that are encoded in legislation, argued by legal practitioners and decided by courts.

3. Data representing the structures of documents and relationships between the documents used by the legal system.

The first type of data defines the procedural and administrative information used in legal documents for the functioning of the courts and legislature. This includes data identifying parties and their roles, the different types of organizations and their level of authority, start and end dates of proceedings and the jurisdictional boundaries of a court. An example of this data is the definition of the types of proceedings for a court case. Instances include trial, appeal, or pre-trial hearing. Other examples are the roles of the participants in litigation with instances of plaintiff, defendant, counsel and judge, and the level of the court with instances of lower or superior and a single sitting judge or full bench.

The second type of data serves the purpose of classifying legal concepts and entities of the legal system. This data is in the form of taxonomies such as legal thesauri that record commonly used terms and digests that classify legal concepts into hierarchies. Examples of this data are hierarchical classifications such as "Tort -> Negligence -> Duty of Care" or common legal phrases such as "reasonable care". Reference list are also included in this type of data. Examples include the types of legal documents such as contracts, wills, insurance policies and deeds.

The third type of data is the structures of legal documents and the linkages between them. Documents such as case reports and legislation are composed of parts that serve a particular purpose for transmitting legal knowledge. Legislation has a hierarchical structure of sections, subsections and clauses as well as document parts such as the "Definitions" section that defines important terms used throughout the document. Case reports have parts such as the headnote section that summarizes the important legal issues addressed by the courts and parts of the judgements where reasons for decisions are given. The links between documents and parts of documents are also important conceptual level data. These links exist in the form of containment and collocation relationships for document parts and citation references for linking documents. The meta-data and parts of the document that contribute to the rich structure of the case report are illustrated in figure 1.
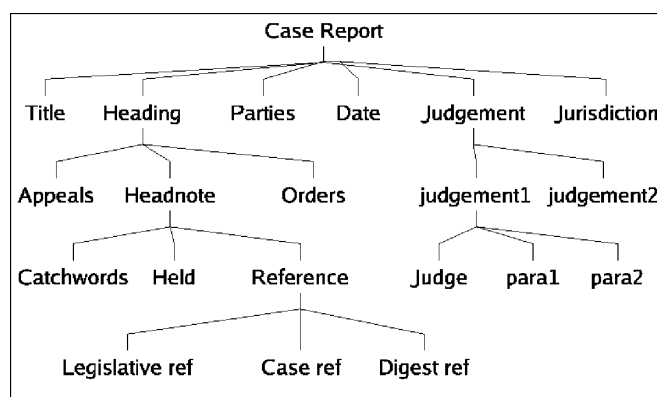

Figure 1: Example Case Report Document

Instance level data consists of content provided by legal institutions in documents including legislation and case reports. There is an interplay between conceptual and instance level data similar to the class and object level within software engineering that provides valuable information to legal practitioners for determining what concepts are important at law. Instance level data when correctly identified and combined with conceptual level data offers strong indicators of the importance of a case.

Legal practitioners understand the common law through an iterative process of searching forward and backward through case citation references to identify cases important in law based on certain characteristic of the case [3]. Information systems can use the instance level and conceptual level data for determining the importance of court cases by assisting the identification of these characteristics. The instance level data can be explicit in the documents as well as historical, obtained by following the linkages between documents that have been created over a period of time. Examples of important types of cases and the data identifying their characteristic include:

- Cases that clarify legislation. The data can be obtained by combining references to statutes through collocation relationships with document structure data such as the "reasons for judgment" document part.
- Cases that construe the meaning of words or phrases in common legal documents. This data can be obtained by combing taxonomies of words and phrases from legal digests such as "interpretation of" with the type of document and the location of the reference within the case report.. Types of documents include wills, insurance policies and contracts.
- Cases that bring a speculative principle of law into an affirmed principle. An indicator of this characteristic is a case that follows an authority in another jurisdiction. The data can be obtained from the citation reference of the cases.
- Cases that apply to a specific class of person. For example, a duty of care applied to doctors or engineers. The characteristic can be identified by combining instance data on occupation from a taxonomy of occupations and the party role data of "defendant" with the conceptual level categorization of terms such as "Duty of care "
- Cases that settle a principle of law that has been in dispute between courts. The data can be obtained from the presence of case citation references to cases with different jurisdictions collocated in the "reasons for decisions" part of the case report.
   A case that deals with an area of law that is quickly evolving because it is likely to reveal trends in the law. The data can be obtained by comparison of the chain of citation references of the evolving legal issues and the dates of the referenced cases.

Legal documents provide valuable data for practitioners both explicitly and in the structures of the documents themselves and their citation references. Data from the conceptual and instance levels can yield information for the identification of legal issues of relevance to a practitioner. Identification and extraction of this data combined with analysis of the characteristics that identify important cases can be used by computer based approaches and legal information systems to assist the identification of cases important to the issues being addressed by legal practitioners.

3. Legal Case Citations

An important concept in law is the authority behind the information contained in legal documents. There are subtleties in the way citations are used that provide indicators of both the authority of the entity making a reference and the authority of the document being referenced. Identifying the entity providing the citation reference and the reasons why the citation reference was provided contributes valuable information about the referenced document. Citation references have been considered by judges and lawyer to have some logical connection to the important legal issues upon which a case is being decided. The types of entities providing citations include:

- Legal advocates: Case reports contain citations provided by the legal advocates acting for the parties in the case. Citations provided by legal advocates generally have a low authority.
- The judge : Most citations in a case report are provided by the judge. The authority of the judge is high but is relative to the authority of the court.
- The court: The level of the court is a factor that influences the authority of all parties providing citations in a case. Appellant courts and full courts lend greater weighting to the authority of citations in their cases.
   Case Reporters: Case reports are composed of the judgments written by the judges and other information the case reporter deems relevant. This extra information can include a summary of the important legal issues and references to sections of legislation and legal digests entries. Court reporters are trained specialist in legal reporting and the references they provide have high of authority.

Examination of the structure of case reports also provides valuable information supporting the authority of a citation reference by indicating the purpose for which the citation was provided. The purpose for citations can be inferred from the document part in which the citation occurs. These parts include:

- The judgment: The judgment contains references for all legal issues addressed by the court whether they were important substantive legal issues or less important matters of procedure.
- Reasons for the Decision: Case Reports include citations provided by judges when deciding a matter. These citations rank highly.
- Cases considered in the decision: Judges sometimes consider a case and specifically apply or not follow the decision. This has a high ranking.
- Affirming or Rejecting a previous decision: Citations provided for the purpose of affirming or rejecting a previous decision have the greatest authority because they are provided by appellate courts and serve the purpose of setting precedent.

Citation data in case report documents provide valuable information that contributes to establishing the authority of a case. Determinations of case authority can be made from the identity of the citation provider and the purpose of the citation that can be derived from its location in the document. Reasons for citations correspond closely to the location of the citation in the case report structure.

4. Legal Paragraphs

Another concept important in the legal system is the encapsulation of legal concepts into legal issues for recording and transmitting legal concepts. Legal issues are the means by which important changes to the law are developed and communicated [4]. Practitioners argue about legal issues and courts decide cases upon the legal issues dealt with in previous court decisions. Analysis of the case report document structure provides information that may assist determination of the relevance of citations references to the legal issues a practitioner is addressing. There are two problems that impact a practitioners ability to determine the relevance of a case that arise due to characteristics of the case report document structure and the use of citations.

The first of these is the problem of multiplicity of legal issues dealt with by a court. A major factor complicating the following of references and making it more difficult to distinguish relevant citation is the many to many relationship between cases and citations. This multiplicity occurs because each case can decide upon many legal issues and cite many other cases in support of those issues.
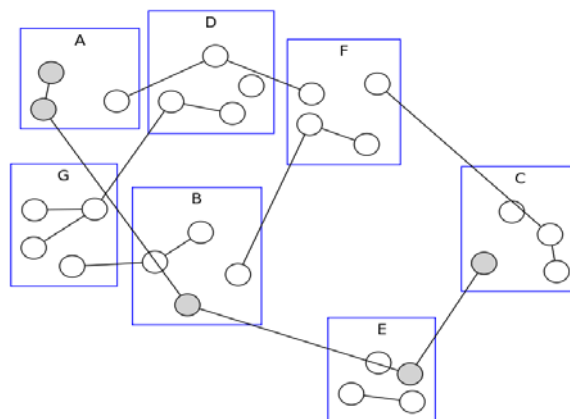

Figure 2. Multiplicity of Legal Issues

Figure 2 illustrates how cases can deal with multiple unconnected legal issues within a single case but a single legal issue can be connected by many cases.

The second problem is the connection between similar legal concepts is not explicit in case report documents. The common usage of citations in legal cases is to cite the case or legal paragraph and not the legal issue. Judgements reference other cases when citing a precedence to support a decision and not a part of the document that specifically deals with the legal issue illustrated in figure 3. Citations to the same case may have no relevance to an issue because the case may be addressing unrelated legal issues.
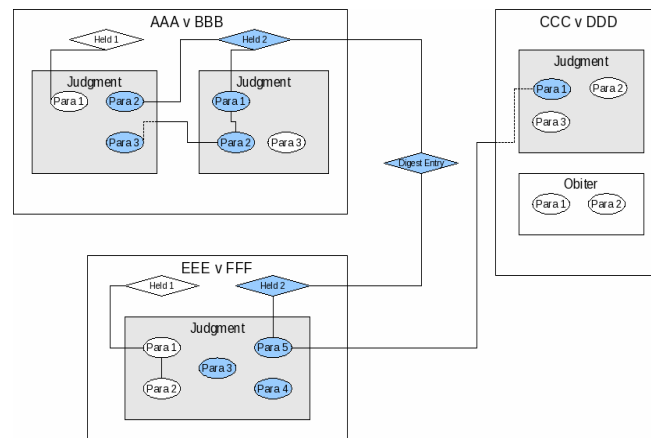
Figure 3. The two types of targets in the citation network

Analysis of the citation references and case report document structures reveal that the network of citations has two types of reference targets. The first type is a container for multiple legal issues. The parts of the case report document that fall into this containment category are the case document as a whole and the judgement parts. The second type of target maps directly to a part of the case report document that expresses information about a single legal issue. These document parts include;

1. the legal paragraph,
2. held part of the head-note sections and
3. references to legal digest entries

The typical use of citation references in case report documents does not permit the identification of related legal issues across multiple case report documents. Knowledge of the citation reference data and its targeted document part within the structure of the case report enables similar legal issues to be indirectly connected across multiple case report documents. This type of information can be utilised by network based information systems.

5. Conclusion

Analysis of the nature and use of data in legal documents reveals rich and interesting information that can be used by legal information systems. Legal documents contain data and meta-data, rich structures and citation links that provide valuable information to assist legal practitioners with the searching and understanding of legal concepts. This paper presents analysis and categorization of the data, document structures and linkages within legal documents and explanations of how the different categories of data interrelate.

6. References

[1] A. Valente, "Types and roles of legal ontologies," in Law and the Semantic Web, ser. LNAI, R. V. Benjamins, P.Casanovas, J. Breuker, and A. Gangemi, *Eds. Spinger Verlag*, Berlin, 2005, no. 3369, pp. 65–76.

[2] J. Breuker, A. Valente, and R. Winkels, "Legal ontologies in knowledge engineering and information management," *Artificial Intelligence and Law*, vol. V12, no. 4, pp. 241–277, Dec. 2004.

[3] Stuart A. Sutton, "The role of attorney mental models of law in case relevance determinations: An exploratory analysis", *Journal of the American Society for Information Science*, vol. 45 no. 3, pp. 186-200, 1994.

[4] C. D. Hafner, "Conceptual organization of case law knowledge bases," in ICAIL '87: *Proceedings of the 1st international conference on Artificial intelligence and law. New York, NY, USA: ACM Press,* 1987, pp. 35–42.