Artif Intell Law (2010) 18:311–320
DOI 10.1007/s10506-010-9098-4

Emerging AI & Law approaches to automating analysis and retrieval of electronically stored information in discovery proceedings

Kevin D. Ashley · Will Bridewell

Published online: 5 October 2010

Abstract This article provides an overview of, and thematic justification for, the special issue of the journal of *Artificial Intelligence and Law* entitled "E-Discovery". In attempting to define a characteristic "AI & Law" approach to e-discovery, and since a central theme of AI & Law involves computationally modeling legal knowledge, reasoning and decision making, we focus on the theme of representing and reasoning with litigators' theories or hypotheses about document relevance through a variety of techniques including machine learning. We also identify two emerging techniques for enabling users' document queries to better express the theories of relevance and connect them to documents: social network analysis and a hypothesis ontology.

Keywords E-discovery · Litigators' relevance hypotheses · User modeling · Machine learning · Social networks · Legal ontologies

1 Introduction

Two of the three workshops on Electronically Stored Information in Discovery (DESI) have been offered in conjunction with International Conferences on Artificial Intelligence (AI) and Law (ICAIL). This fact is consistent with the intuition that the field of AI & Law offers a special or characteristic approach to the

K. D. Ashley (✉)
School of Law, University of Pittsburgh, Pittsburgh, PA 15260, USA
e-mail: Ashley@pitt.edu

W. Bridewell
Cognitive Systems Laboratory, Center for the Study of Language and Information,
Stanford University, Stanford, CA 94305, USA
e-mail: willb@csli.stanford.edu

process of automating the analysis and retrieval of electronically stored information in pretrial discovery proceedings. *Pretrial discovery* is the process by which parties to a law suit request and obtain access to materials in the hands of opponents and others for the purpose of revealing facts and developing evidence. Collecting, exchanging, and analyzing electronically stored information or ESI are increasingly a focus of pretrial discovery, a process referred to as *e-discovery* (Withers 2006).

In this special issue on E-Discovery of the journal of *Artificial Intelligence and Law*, it is appropriate to consider what a characteristically "AI & Law" approach to e-discovery might mean. Of course, any approach that applies AI to address this important practical legal problem could with justification be called an AI & Law approach. (Conrad 2010, this issue, provides suggestions for applying AI to e-discovery.) Nor are AI approaches to this legal problem preferable to other approaches more typical of information retrieval (except, perhaps, to an ICAIL audience).

What makes the question of a characteristic AI & Law approach so interesting is that e-discovery is anything but characteristic of the problems AI & Law researchers have tackled in the past. While research has addressed information extraction from and automatic classification of legal texts, the work has focused on relatively homogeneous documents, such as legal opinions dealing with a single type of claim and sharing structural features (see, for example, Ashley and Brüninghaus 2009; Daniels and Rissland 1997; Gonçalves and Quaresma 2005; Grover et al. 2003; Hachey and Grover 2006; Jackson et al. 2003; McCarty 2007; Thompson 2001; Uyttendaele et al. 1998; Weber 1998). The principal difference is the extreme heterogeneity of documents produced in litigation, including not only corporate memoranda and agreements, but also the full panoply of email and other internet-based communications (Oard et al. 2010, this issue.) As a result, techniques developed in AI & Law that rely on either explicit or implicit structure in comparatively homogenous legal documents may be of little avail in discriminating relevant from irrelevant documents. In addition, the commercial implications of e-discovery are enormous, technical innovation in e-discovery has tended to be driven by private firms rather than academic researchers funded with public research dollars, and there are special issues concerning defining relevance and evaluating performance with which AI & Law researchers would need to grapple. For an overview of the commercial context of e-discovery, see (Conrad 2010, this issue). Oard et al. (2010, this issue) present an extensive discussion of e-discovery issues of relevance, evaluation and pragmatic considerations such as the effects of Optical Character Recognition (OCR).

In defining a characteristic AI & Law approach to e-discovery, one might begin with the working definition of the core field of AI & Law in the Journal's aims and scope as encompassing "theoretical or empirical studies in artificial intelligence (AI), cognitive psychology, jurisprudence, linguistics, or philosophy which address the development of formal or computational models of legal knowledge, reasoning, and decision making." In view of this definition, a special AI & Law approach to e-discovery might well be defined as one that involves such a computational model as it pertains to document production and analysis. More specifically, the computational model would deal with the representation of litigators' hypotheses or theories about the relevance of documents to the legal claims, defenses, and strategies involved in the litigation and their reasoning as they review and make decisions about documents.

2 Relevance hypotheses and sensemaking

In the e-discovery context, a hypothesis or theory of relevance is a more-or-less abstract description of subject matter that, if found in a document, would make that document relevant (Hogan et al. 2009). As described in (Hogan et al. 2010, this issue) a theory of relevance is constructed by senior litigators engaged in "sensemaking", a "process of collecting, organizing and creating representations of complex information sets, all centered around some problem they need to understand" (Bauer et al. 2008 [citation omitted]). According to (Attfield and Blandford 2010, this issue), "sensemaking involves a bi-directional interaction between engagement with data (i.e. bottom-up processing) and continually evolving representations and understanding that account for that data (i.e. top-down processing)." They agree that litigators' "theories or conjectures were central to each line of enquiry, which were theory-led."

In terms of legal practice, litigators' hypotheses about document relevance relate to the formal requests for documents that the litigators filed in the case. These requests, in turn, relate to the details of the plaintiff's complaint that formally states the legal claims with which it commenced a lawsuit or the defendant's answer that formally states its defenses. The parties' litigators prepare and file these statements based upon their best information about the facts that occurred and how the facts connect with relevant legal rules. As new facts, claims, and defenses emerge, the litigators sometimes file modifications. (See (Oard et al. 2010, this issue) for a comprehensive introduction to the legal context of e-discovery.) As a result, according to Ashley and Bridewell (2007), litigators often formulate fairly specific theories about the documents that would be relevant, and they express those criteria in terms of more-or-less specific hypotheses about who communicated what to whom, when, and, to the extent possible, why; some examples (expressed here as natural language statements) might include:

– "There are documents showing that the Vice President of Marketing knew that cigarette advertisements were targeted to children by 1989," or
– "There exist documents to or from employees of a tobacco company or tobacco organization in which a tobacco company officer refers to illegal payments to foreign officials," or
– "There are documents that are communications between Alice and her lawyer Bob between 1985 and 1989," or more generally,
– "There are documents *of a particular kind*, satisfying *particular time constraints*, satisfying *particular social interaction constraints*, that refer to *particular concepts or phrases of interest*."

(Hogan et al. 2010, this issue) provides other examples, such as, "Are promotional offers relevant?" "Is 'buy one get one free' by itself sufficient for relevance?" The case studies reported by Attfield and Blandford (2010, issue) confirm that as litigators review documents, they generate successively more specific theories of relevance and pose more specific questions in order to evaluate their theories about the factual and legal elements of a case. Such specific hypotheses, however, may not be expressible with Boolean queries.

These examples of litigators' relevance hypotheses are based upon the TREC Legal Track (http://trec.nist.gov/call09.html) materials including the University of Illinois's Complex Document Information Processing (CDIP) Test Collection. See (Oard et al. 2010, this issue). The collection comprises some 6.9 million documents released by tobacco companies in connection with a Master Settlement Agreement. As materials for the competition, experts associated with the Sedona Conference (http://www.thesedonaconference.org/)—a continuing forum of jurists, attorneys and technologists addressing issues of complex litigation—have created nine fictitious but realistic legal complaints, and 53 retrieval topics cast as document requests each with a corresponding Boolean query created through a realistic process of negotiation (Baron and Thompson 2007; Oard et al. 2010, this issue). For each topic, the National Institute of Standards and Technology generated judgment pools for relevance assessment based on searches of an expert manual searcher and the participating teams and manual assessments by 35 volunteer assessors (Baron and Thompson 2007; Oard et al. 2010, this issue). The TREC Legal Track competition has also made public another similarly prepared repository of discovery requests and materials in connection with email documents generated by the Enron litigation (Oard et al. 2010, this issue).

This special issue contains two research papers that focus explicitly on modeling e-discovery in terms of litigators' relevance theories. Attfield and Blandford (2010, this issue) report systematic case-study observations of three e-discovery investigations involving real cases, firms, and attorneys. Based on the study, the authors pose several hypotheses concerning the order in which legal investigators encounter documents. The authors explore the ramifications for the design of visualization tools for presenting and relating documents in terms of litigators' theories of relevance.

Hogan et al. (2010, this issue) describe how litigators' relevance hypotheses can be elicited in an automated iterative user-modeling procedure, thus "grounding IR relevance in cognitive task analysis," and resulting in what the authors characterize elsewhere as "a knowledge-based system in the classic AI sense" (Bauer et al. 2008). The iterative cognitive task analysis results in a knowledge representation of litigators' relevance hypotheses as "computational constructs into which the linguistic variations can be expressed and the operations (e.g., Boolean expressions) that enable phrasing of query alternations which capture the meaning being sought in IR computation" (Bauer et al. 2008). Their user-modeling process appears to augment Boolean expressions with relevance information concerning the "use case", (i.e., the user's objectives for production given the Request for Production), the scope of the legal and other concepts that are relevant to the case, the nuance, or level of specificity of the relevant concepts of interest, and linguistic variability in the ways a concept can be expressed (Hogan et al. 2009, 2010, this issue). It is an interesting empirical question whether this mode of representing relevance hypotheses is sufficiently expressive to capture litigators' specific criteria of document relevance; some suggestions concerning the use of social network analysis and a hypothesis ontology are provided below.

Significantly for AI & Law, the user modeling procedure described by (Hogan et al. 2010, this issue) and, to some extent, the case studies of (Attfield and Blandford 2010, this issue) are examples of a kind of cognitive task analysis that Buchanan and Headrick (1970) long ago recommended as a prerequisite for progress in applying Artificial Intelligence to law and legal reasoning. Unfortunately, in the annals of the field, this kind of cognitive task analysis has occurred seldom if at all. It would be a welcome return to first principles, and a contribution to AI & Law, simply to provide a detailed description of how to conduct the modeling process and to represent the relevance knowledge obtained, illustrated with some examples of litigators' relevance hypotheses, and demonstrating empirically the utility of the relevance hypotheses.

3 Machine learning to apply theories of relevance

Machine learning is one AI & Law approach for automating the process of, in effect, capturing, applying, and modifying a human annotator's theory of relevance. Examples of documents that have been classified by the human annotator can be used to train an automated classifier to identify and classify other similar documents. For example the classifier in the SMILE + IBP program was trained (using a variety of machine learning algorithms, such as C4.5, Ripper, RL, nearest neighbor and Naïve Bayes) to identify stereotypical fact patterns that strengthen or weaken a side's claim, but in much more homogeneous textual descriptions of case facts (Ashley and Brüninghaus 2009).

In this special issue, Privault et al. (2010) describe a machine learning system that learns, from human annotations of heterogeneous documents, probabilistic models to automatically categorize new documents. The approach assumes, and seeks to represent probabilistically, hidden ("latent" or "unobserved") factors which may underlie the co-occurrences in categories among a set of observed words and documents. Documents are represented in terms of the words, and frequencies of the words, they contain. The program outputs probabilities of the various categories given the documents, words, and frequencies. The program has been preliminarily evaluated on a subset of the Enron data. A major focus of the paper is on a means of presenting the document set visually to facilitate the manual document tagging from which the probabilistic models are learned. (See also Attfield and Blandford 2010, this issue, for other visual representations of document sets.)

In the user-modeling process described by Hogan et al. (2009), a substantial number of documents are manually assessed as relevant or not; these become training data for an automated classifier in a supervised machine learning process. The classification system is regularly run over all of the documents in the corpus and its effectiveness assessed based on random sampling of classified documents. The data thus generated are used to tune the system and modify the theory of relevance (Hogan et al. 2009). Unfortunately, the authors do not elaborate on the particular machine learning techniques applied. As observed in (Ashley 2007), other areas of research in AI/machine learning may have implications for e-discovery such as: (1) Textual Case-Based Reasoning (TCBR) methods to analyze electronic corporate textual documents and capture implicit structural or ordering information, for instance, using Hidden Markov Models to identify anomalous inspection reports (Mustafaraj et al. 2007). (2) Natural Language Processing (NLP) techniques to recognize subjectivity and sentiment in text to identify textual documents where subjective opinions or emotionally-charged recommendations are made (e.g., Wilson 2008). (For other relevant machine learning techniques, see also Conrad 2010, this issue.)

4 New tools for formulating relevance theories

Representing, either explicitly or implicitly, knowledge associated with users' relevance hypotheses so that an automated system can use it is of primary importance in addressing the key issues in e-discovery of how to formalize queries and how to identify documents satisfying those queries. In the work reported in this issue (Hogan et al. 2010; Privault et al. 2010), eliciting litigators' relevance hypotheses is seen as an iterative user modeling procedure based primarily on refining or inducing queries based on litigators' selections of relevant documents. While in the latter, the knowledge is represented probabilistically, Hogan et al. (2009) state that the knowledge should be "formalized into the system via query development, vocabularies, etc." Since, as suggested above, litigators' theories of relevance are often quite specific, it may be useful to provide them with more expressive tools for formulating exactly what they are looking for. Two such tools are (1) social network analysis and (2) hypothesis ontologies.

4.1 Social network analysis to apply relevance theories

Social networks, graphs representing probable social interactions and relations among a group of people, could enable a system to infer which individuals most likely communicated information or had knowledge relevant to a query (Schwartz and Wood 1993; Agrawal et al. 2003). Since senders, recipients, and owners of documents identify themselves through email records and their contents, one conceivably can build a model of the network of knowledge and, from that structure, infer something about the likely content and targets of the communications. In other words, by determining how each item moves through a social organization, one can map the general flow of knowledge and infer something about the relevance of the content of documents based on who has them, who sent them, and who likely read them.

As an automated technique for supplementing, representing, and reasoning about the information targeted by litigators' relevance hypotheses, social networks could make an important contribution toward using AI techniques to improve legal problem solving. Essentially, the social network reduces ones reliance on individual words and phrases while increasing ones ability to infer relevance based on a history of contact. Given a social network, one can expand the set of responsive documents by including all the documents associated with an author of relevant text or with an individual in contact with that person. For instance, suppose one knows of correspondence between Alice and Bob about nicotine addiction and similar communications between Charlie and Dana. If Bob and Dana are frequent correspondents, then they may have discussed nicotine addiction as well. Changes in the social network over time, as communication links form and break and individuals appear and disappear, directly affect the flow of information and, in turn, the relevance of documents to a query. If the four individuals wrote about nicotine addiction within the same or overlapping time frames, then the documents between Bob and Dana deserve more attention than if Alice and Bob wrote to each other in 1981 and Bob and Dana were corresponding in 1995. A rational consideration of information dynamics could refine the system's decision to include documents via the social network and rule out documents that were created within an unrelated time frame or were communicated between irrelevant parties (See Henseler 2010, this issue).

To some extent, the social networks could be constructed automatically from the documents themselves. As a first approximation to the social structure, a system can use document metadata, such as the *to* and *from* fields in email headers, to build a network of every identifiable individual and create a link only if supported by the metadata or extracted named-entities (Henseler 2010, this issue; Kossinets and Watts 2006; see also, Eichmann and Chin 2007). A document's unstructured text may mention names of people, companies, and other entities that belong in the network, information that could be captured with named entity recognition techniques, which identify meaningful words and phrases that fall into informative social categories (Curran and Clark 2003). In the AI & Law community, for example, Hachey and Grover (2006) present a potential rule-based solution for the narrower problem of recognizing named entities specific to the legal domain such as references to acts, sections, and judgments. Tools from graph mining and statistical relational learning could help to deal with problems of multiple people with the same name, individuals using multiple names and email addresses, and communications that do not imply a strong social link (e.g., interoffice memos). In the user modeling process, litigators would edit this first order social network and supplement it with social role information. Such roles can arise through formal organizational structure (e.g., Chief Counsel) or through the organics of interoffice communication (e.g., social hub), and would aid the system in assessing the importance of a social link. Social network techniques for temporal graph analysis could model dynamic changes in information flow.

4.2 A hypothesis ontology for expressing relevance theories

As proposed by Ashley and Bridewell (2007), a "hypothesis ontology" and an "event-based framework" similar to those developed for bioinformatics (Racunas et al. 2004) could represent, in a broadly applicable form, knowledge associated with users' relevance hypotheses. Based on Racunas's definition, a hypothesis ontology comprises "a vocabulary of objects (agents) and processes", associated with recurrent areas of interest in e-discovery such as knowledge transmission in corporate or commercial settings "and define[s] the relationships in which these entities can participate." For instance, an ontology will include key knowledge transmission concepts, such as "*communicated-with*" and "*during-interval*", that expect arguments of a particular type and invoke an established function. An event-based framework supports specifying hypotheses about "knowledge transmission" events "by naming the agents from the ontology … and the [knowledge transmission] processes… that connect them (Racunas et al. 2004)."

Implemented in a suitable user interface, a hypothesis ontology and event-based framework could enable users to represent their own hypotheses in a semi-formal language that specifies the time and context-dependent relationships among the objects and processes in the corpus (Racunas et al. 2004). Framed in terms from a hypothesis ontology, a rough example of a relevance hypothesis might be, "there exist documents in which Alice *communicated-with* Bob *during-interval* 1976–1978 that *contain-keywords* {tobacco, children, advertising}".

The solicitation of information for building and modifying the hypothesis ontology, event-based framework, and the relevance hypotheses themselves, would be an iterative task of user modeling. Detailed legal ontologies could serve as a basis for indexing schemes required to make legal sense of the electronic documents discovered; such legal ontologies can seed the hypothesis ontologies to be fleshed out from the electronic documents in the corpus and in user modeling (see, e.g., Breuker et al. 2004). The core ontology, including the more general concepts of knowledge transmission in corporate or commercial settings, would be available to serve other users in other cases. The ontology and framework could also support reasoning about communications and documents, especially if supplemented with information about the social network. For example, they would help to address the challenge of explaining to users why retrieved documents are relevant in terms of their research hypotheses using explanation schemas (Thagard 1999; Schank 1986) or techniques for visualizing evidential reasoning in argument diagrams to help litigators collaboratively construct their legal arguments and integrate the documents into them (Tillers 2007).

5 Conclusions

In this article, we have attempted to provide an overview of, and a thematic rationale for, this special issue of the journal of *Artificial Intelligence and Law* devoted to e-discovery. While it is true that any research approach to the problem of e-discovery that includes AI will be duly accepted as an "AI & Law" approach, we have focused on the theme of using various techniques, including machine learning, for eliciting, representing, implementing, and reasoning with litigators' hypotheses or theories of document relevance. In addition, the paper has identified two emerging areas for improving the process of automating analysis and retrieval of electronically stored information in discovery proceedings: (1) social network analysis to supplement and apply user relevance theories, and (2) a hypothesis ontology to provide a language for representing relevance theories. Intuitively, traditional IR techniques may not be precise or expressive enough to identify all documents relevant to a particular factual or legal issue. Machine learning could automate identification of classes of documents relevant to a given issue, ontologies could provide a more expressive query language, and social network analysis could provide relevance clues based on communication patterns rather than just the isolated text of each document.

Since a central theme of AI & Law involves computationally modeling legal knowledge, reasoning and decision making, the techniques reported in this issue illustrate that field's potentially unique contribution to e-discovery. We hope this special issue will inspire future AI & Law approaches toward addressing the challenges of e-discovery.

References

Agrawal R, Rajagopalan S, Srikant R, Xu Y (2003) Mining newsgroups using networks arising from social behavior. In: Proceedings of the twelfth international conference on world wide web, ACM Press, Budapest, Hungary, pp 529–535

Ashley K (2007) Can AI & Law contribute to managing electronically stored information in discovery proceedings? Some points of tangency; position paper; ICAIL 2007 workshop on supporting search and sensemaking for electronically stored information in discovery proceedings (DESI). Stanford. http://www.umiacs.umd.edu/~oard/desi-ws/

Ashley K, Bridewell W (2007) III-CXT-Medium: tracing information flow to improve legal e-discovery. NSF Proposal No. 0803049. Declined

Ashley K, Brüninghaus S (2009) Automatically classifying case texts and predicting outcomes. Artif Intell Law 17(2):125–165

Attfield S, Blandford A (2010) Discovery-led refinement in e-discovery investigations: sensemaking, cognitive ergonomics and system design. Artificial intelligence and law special issue on e-discovery (this issue)

Baron JR, Thompson P (2007) The search problem posed by large heterogeneous data sets in litigation: possible future approaches to research. In: Proceedings of the eleventh international conference on artificial intelligence and law, ACM Press, Stanford, CA, pp 141–147

Bauer R, Jade T, Hedin B, Hogan C (2008) Automated legal sensemaking: the centrality of relevance and intentionality. In: Proceedings 2nd intl. workshop on supporting search & sensemaking for electronically stored information in discovery, London, June. http://eprints.ucl.ac.uk/9131/1/9131.pdf

Breuker J, Valente A, Winkels R (2004) Legal ontologies in knowledge engineering and information management. Artificial intelligence and law, Vol 12(4). Springer, pp 241–277

Buchanan B, Headrick T (1970) Some speculation about artificial intelligence and legal reasoning. Stanford Law Rev 23:40–62

Conrad J (2010) E-Discovery revisited: the need for artificial intelligence beyond information retrieval. Artificial intelligence and law special issue on e-discovery (this issue)

Curran J, Clark S (2003) Language independent NER using a maximum entropy tagger. In: Proceedings of the seventh conference on natural language learning (CoNLL), pp 164–167

Daniels J, Rissland E (1997) Finding legally relevant passages in case opinions. In: Proceedings of the sixth international conference on artificial intelligence and law, ACM Press, Melbourne, Australia, pp 39–46

Eichmann D, Chin S-C (2007) Concepts, semantics, and syntax in E-Discovery. In: ICAIL 2007 workshop on supporting search and sensemaking for electronically stored information in discovery proceedings (DESI Workshop) http://www.umiacs.umd.edu/~oard/desi-ws/

Gonçalves T, Quaresma P (2005) Is linguistic information relevant for the text legal classification problem? In: Proceedings of the tenth international conference on artificial intelligence and law, ACM Press, Bologna, Italy, pp 168–176

Grover C, Hachey B, Hughson I, Korycinski C (2003) Automatic summarisation of legal documents. In: Proceedings of ninth international conference on artificial intelligence and law, ACM Press, Edinburgh, Scotland, pp 243–251

Hachey B, Grover C (2006) Extractive summarization of legal texts. Artif Intell Law 14:305–345

Henseler H (2010) Network-based filtering for large email collections in e-discovery. In: Artificial intelligence and law special issue on e-discovery (this issue)

Hogan C, Bauer R, Brassil D (2009) Human-aided computer cognition for e-discovery. In: Proceedings of 12th international conference on artificial intelligence and law (ICAIL-09), Barcelona, pp 194–201

Hogan C, Bauer R, Brassil D (2010) Automation of legal sensemaking in e-discovery. In: Artificial intelligence and law special issue on e-discovery (this issue)

Jackson P, Al-Kofahi K, Tyrrell A, Vachher A (2003) Information extraction from case law and retrieval of prior cases. Artif Intell 150:239–290

Kossinets G, Watts D (2006) Empirical analysis of an evolving social network. Science 311:88–90

McCarty LT (2007) Deep semantic interpretations of legal texts. In: Proceedings of the eleventh international conference on artificial intelligence and law, ACM Press, Stanford, CA, pp 217–224

Mustafaraj E, Hoof M, Freisleben B (2007) Knowledge extraction and summarization for an application of textual case-based interpretation. In: Weber RO, Richter MM (eds) Proceedings of the 6th ICCBR conference, ICCBR 2007, LNAI 4626, pp 517–531

Oard D, Baron J, Hedin B, Lewis D, Tomlinson S (2010) Evaluation of information retrieval for e-discovery. Artificial intelligence and law special issue on e-discovery (this issue)

Privault C, O'Neill J, Ciriza V, Renders J-M (2010) A new tangible user interface for machine learning document review. In: Artificial intelligence and law special issue on e-discovery (this issue)

Racunas SA, Shah NH, Albert I, Fedoroff NV (2004) HyBrow: a prototype system for computer-aided hypothesis evaluation. Bioinformatics 20:i257–i264

Schank R (1986) Explanation patterns: understanding mechanically and creatively. Lawrence Erlbaum Associates, Hillsdale

Schwartz MF, Wood DCM (1993) Discovering shared interests using graph analysis. Commun ACM 36:78–89

Thagard P (1999) How scientists explain disease. Princeton University Press, Princeton

Thompson P (2001) Automatic categorization of case law. In: Proceedings of the eighth international conference on artificial intelligence and law, ACM Press, St. Louis, MO, pp 70–77

Tillers P (ed) (2007) Special issue on graphic and visual representations of evidence and inference in legal settings. In: Law, Probability, and Risk, vol 6, pp 1–4

Uyttendaele C, Moens M, Dumortier J (1998) SALOMON: automatic abstracting of legal cases for effective access to court decisions. Artif Intell Law 6:59–79

Weber R (1998) Intelligent jurisprudence research, Federal University of Santa Catarina, Florianópolis, Brazil, unpublished doctoral dissertation

Wilson T (2008) Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states. Ph. D. Dissertation, University of Pittsburgh Intelligent Systems Program. http://homepages.inf.ed.ac.uk/twilson/dissertation.html

Withers K (2006) Electronically stored information: The December 2006 amendments to the federal rules of civil procedure. Northwestern Journal of Technology & Intellectual Property, 4, (Property 171) at http://www.law.northwestern.edu/journals/njtip/v4/n2/3