

Session 3: Position Papers (14:30 – 16:00)

Chair: Dr. Kevin D. Ashley, University of Pittsburgh School of Law

1. Dr. Kevin D. Ashley, “*Emerging AI+Law Approaches to Automating Analysis and Retrieval of ESI in Discovery Proceedings*”
 2. K. Krasnow Waterman, Esq., LawTechIntersect LLC, “*Isomorphic Intermediate Representations Are Needed To Support Cross-Border eDiscovery and Digital Evidence Systems*”
 3. Dr. Simon Attfield, University College London, “*The Loneliness of the Long-Distance Document Reviewer: E-Discovery and Cognitive Ergonomics*”
 4. William P. Butterfield, Esq., Hausfeld LLP, “*Diving Deeper To Catch Bigger Fish*”
 5. David Chaplin, Kroll Ontrack, “*Conceptual Search Technology: Avoid Sanctions, Prevent Privilege Waiver, and Understand Your Data*”
 6. Jorge H. Román, Los Alamos National Laboratory, “*Discovery of Patterns in Digital Records*”
- Moderated discussion

Emerging AI & Law Approaches to Automating Analysis and Retrieval of Electronically Stored Information in Discovery Proceedings

Position Paper

Kevin D. Ashley

Professor of Law and Intelligent
Systems
Senior Scientist, Learning Research
and Development Center
University of Pittsburgh
ashley@pitt.edu

Will Bridewell

Cognitive Systems Laboratory
Center for the Study of Language
and Information
Stanford University
Stanford, CA 94305
willb@csli.stanford.edu

Why care if AI & Law approach to e-Discovery emerging?

- Intuitively, AI & Law ought to have an approach to e-Discovery.
 - DESI I and III held at ICAIL
- In fact, it seems that a distinctive AI & Law approach is emerging.
- Planned special issue of *Journal of Artificial Intelligence and Law*
 - on e-Discovery and emerging AI & Law approaches.

What would an “AI & Law” approach to e-Discovery look like?

- Working definition of core of AI & Law (from aims and scope of *Journal of Artificial Intelligence and Law*):
 - “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.”
- AI & Law approach involves computational models of:
 - legal knowledge, reasoning and decision making as it pertains to
 - document production and analysis of e-documents.
- In e-Discovery context this may map to representing:
 - litigators’ hypotheses (or theories) about relevance and their
 - reasoning as they review and make decisions about documents.

What's so hard about that?

- e-Discovery differs from typical AI & Law problems .
 - AI & Law researchers address information extraction from and automatic classification of legal texts, but with
 - relatively *homogeneous* documents,
 - such as legal opinions dealing with single type of claim and
 - sharing structural features.
 - e.g., Brüninghaus and Ashley, 2005; Daniels and Rissland, 1997; Gonçalves and Quaresma, 2005; Grover et al., 2003; Hachey and Grover, 2006; Jackson et al., 2003; McCarty, 2007; Thompson, 2001; Uyttendale et al., 1998; Weber, 1998.
- The principle difference in e-Discovery:
 - extreme *heterogeneity* of documents produced in litigation
 - not only corporate memoranda and agreements, but
 - full panoply of email and other internet-based communications.
- AI & Law techniques that rely on explicit or implicit structure in comparatively homogenous legal documents may not work.

Focus on litigators' hypotheses (or theories) about relevance

- *Relevance hypothesis*, aka theory of relevance, is:
 - more-or-less abstract description of subject matter that, if found in a document, would make that document relevant. (Hogan et al., 2009)
- Part of senior litigators' "sensemaking",
 - "process of collecting, organizing and creating representations of complex information sets, all centered around some problem they need to understand."
(Bauer et al., 2008)
- By analyzing complaints and document requests (e.g., TREC Legal Track), infer info about relevance hypotheses, e.g.,
 - "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.*"

User modeling / cognitive task analysis re relevance hypotheses

- Bauer et al. (2008), assert that litigators' relevance hypotheses can be elicited in an
 - iterative user modeling procedure,
 - a task analysis “replicating the cognitive sensemaking task of a senior litigator with an automated, computational platform.”

- ...an example of a kind of cognitive task analysis that Buchanan and Headrick (1970) long ago recommended as a prerequisite for progress in applying AI to law and legal reasoning.
 - Buchanan, B. and Headrick, T. (1970) Some Speculation about Artificial Intelligence and Legal Reasoning. *Stanford Law Review*, 23: 40-62.

- User-modeling process in (Bauer, et al., 2008; Hogan, et al. 2009) yields:
 1. “use case”, user’s objectives for production given Request for Production, including objectives to avoid producing too much or too little;
 2. scope of conceptual boundaries of interest (e.g., legal and other concepts relevant to the case);
 3. nuance, or level of specificity of relevant concepts of interest;
 4. linguistic variability—“the variety of ways a concept can be expressed, whether lexically or syntactically.”

Three emerging AI & Law Techniques for e-Discovery

1. Machine learning to extend and apply users' theories of relevance
 - Examples of documents classified according to theory of relevance
 - used to train an automated classifier to identify and classify other documents.
2. Hypothesis ontology to generalize user modeling regarding relevance theories
 - vocabulary of objects (agents) and processes,
 - associated with recurrent areas of interest in e-Discovery such as knowledge transmission in corporate or commercial settings;
 - define[s] the relationships in which these entities can participate.
3. Social network analysis to supplement user models of and to apply relevance theories in document analysis and retrieval.
 - Senders, recipients, and owners of documents identify themselves through email records and documents.
 - Build model of network of knowledge and use resulting structure to draw inferences.
 - By determining how each item moves through social organization, map general flow of knowledge and
 - infer something about relevance of document content based on who has them, who sent them, and who likely read them.

Two Research Challenges re Relevance Hypotheses

1. How well can dynamics of social network over time and time-stamps of communications be used to qualify and refine inferences?
 - E.g., compare inferences from social network that is:
 - Static: if Alice and Bob correspond about nicotine addiction and similar communications between Charlie and Dana, if Bob and Dana correspond frequently they may have discussed nicotine addiction, as well.
 - Dynamic: if A, B, C, and D wrote about nicotine addiction within same or overlapping time frames, then documents between B and D deserve more attention than if A and B wrote to each other in 1981 and Bob and Dana were corresponding in 1995.
2. How well can hypothesis-based retrieval system explain why documents are relevant in terms of users' research hypotheses.
 - Justifications for document's inclusion or exclusion may be complex:
 - E.g., "The Vice President of Marketing communicated frequently with Sara between 1985 and 1989. Sara communicated with a third party, Tom, several times about cigarettes and children. This document between Sara and the Vice President of Marketing mentions children."
 - Explanation schemas need to be integrated with document clustering.

Conclusions

- A distinctive AI & Law approach to e-Discovery is emerging.
 - Focuses on litigators' hypotheses (or theories) about relevance
 - User modeling / cognitive task analysis re relevance hypotheses
 - Emerging techniques relative to relevance hypotheses:
 - Machine learning, Hypothesis ontology, Social network analysis
 - Challenges:
 - Dynamic research challenges re relevance hypotheses
 - Explanation in hypothesis-based retrieval
- Other “AI & Law” approaches?
- Planned special issue of *Journal of Artificial Intelligence and Law*
 - on e-Discovery and emerging AI & Law approaches.
 - Contact ashley@pitt.edu for more information