Cognitive Assistants for Document-Related Tasks in Law and Government

L. Karl Branting
The MITRE Corporation
7515 Colshire Dr
McLean, VA, 22102, USA

Abstract

The legal relationship between government and citizens is mediated by documents. This paper identifies four classes of cognitive assistants having the potential to improve the experience of citizens and government officials in using and understanding government documents: self-filling forms; error-detecting forms; proactive information search; and deductive document synthesis.

Introduction

Many of the legal relationships between governments and citizens are mediated by documents. Government agencies often use *forms* as the primary vehicle by which citizens can request information, services, and other benefits. *Informational documents* are often the primary source of information for citizens about how to complete forms, comply with legal requirements, or obtain legal benefits. The inability of citizens to understand or correctly fill out forms or to accurately understand informational documents is a major source of frustration, errors, and inefficiency in compliance and access to benefits.

A significant contributor to this inability is the *ad hoc* character of the process by which most forms and informational documents are drafted. As depicted in Figure 1, the connection between the goals of citizens and organizations and structure and content of forms and informational documents is in the mind of the drafter, but not explicitly modeled in a fashion that could permit automated assistance to either drafters or citizens.

A complete solution to the *ad hoc* nature of government documents would require formalization of rule-makers' goals, citizens' goals, and the logical structure of both rules and documents. While there is an extensive literature on the logical structure of legal rules,[1] there is much less work on the structure of legal documents, *e.g.*, (Igari, Shimazu, and Ochimizu 2012; Branting, Lester, and Callaway 1997), and still less on formal models of the relationships between rule-makers' and citizens' goals. Until such inclusive models exist, the opportunity for cognitive assistants must be limited to individual components of these relationships.

This paper explores four areas in which cognitive assistants could improve the experience of citizens and government officials in using and understanding government documents: self-filling forms; error-detecting forms; proactive information search; and deductive document synthesis.

Self-Filling Forms

It is not unusual for citizens to fill out similar or identical forms on multiple successive occasions, *e.g.*, periodic requests for benefits, license renewals, tax returns, *etc.*, often with many values invariant across multiple occasions. Moreover, there is often redundancy among separate fields of documents such that the value of the contents of one field may be predictable based on the contents of other fields. Under these circumstances, supervised concept learning can be used to prepopulate fields based on values in other fields and on previous values.

An early prototype of this approach was used in a cognitive assistant for routine form completion described in Hermens and Schlimmer's 1994 "Machine-learning apprentice for the completion of repetitive forms" (Hermens and Schlimmer 1994). The system maintained a separate classification model for each field based on values of surrounding fields. Two alternate supervised concept learning algorithms were tested for prediction: COBWEB and ID4. Each time the user filled in a field in a form, the system's prediction module updated the default values for fields that the user had not yet edited. In the system's evaluation, performance of the system was compared with three baselines: no-learning, most-commonly-used-values, and most-recently-used values. Hermens and Schlimmer found that the best-performing algorithm, ID4, reduced the number of keystrokes required for form completions by 87% as compared to the no-learning method.

The training data for Hermen and Schlimmer's system consisted of actions of all users. The models were therefore suited for auto-filling values that could be equally applicable to all users, such as "academic year" or "summer" depending on the month and whether the string in the name field appeared in the list of faculty members. However, this approach could be extended to repetitious completion of forms by individual users.

This approach, in which field values are predicted by an inductive model trained on a global data set of previous con-

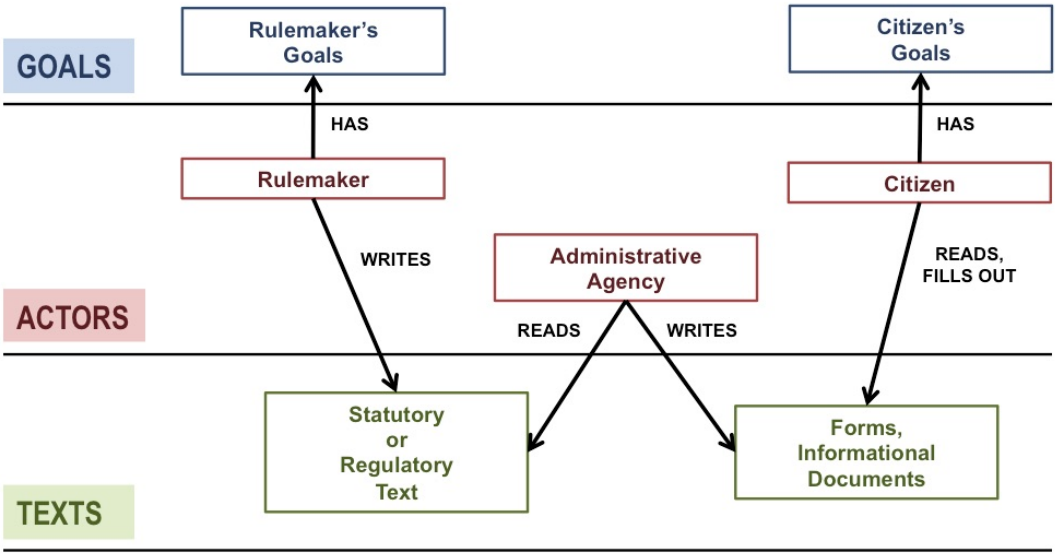

Figure 1: The current process whereby the goals of regulators and those of citizens are linked through multiple ad hoc connections.

text/field pairs, is applicable to a wide range of routine government forms. As additional users complete a given form, predictive accuracy could improve, permitting each citizen to benefit from the actions of previous users. While constraints or rules can be explicitly programmed into form fields, this learning approach obviates that programming step and provides the capability of adapting to changes in the behavior of the user population.

Error-Detecting Forms

Pro se litigants are citizens who represent themselves in court without the assistance of a lawyer (Goldschmidt et al. 1998). Pro se litigants typically make many errors when completing the forms required to initiate and pursue legal cases. In bankruptcy court, for example, auditing errors by pro se litigants consume a significant quantity of judicial resources (O'Brien 1997). However, observations of this auditing process suggest that a small number of error types account for most of this auditing effort and that errors are therefore typically quite predictable. Moreover, the auditing process itself can be viewed as, in effect, an annotation process for pro se documents. Specifically, US Federal courts use an electronic case management system, CM/ECF,[2] that records each document submitted to a court as part of a docket entry that includes metadata about the case filing together with auditing decisions and comments. This provides a rich training set not just for suggesting possible field values, but more importantly for detecting errors and providing information about the nature of errors and possible corrections.

Docket entry and auditing information from CM/ECF constitutes a training set for learning the concept "incorrect document field," *i.e.*, learning to predict whether a field's content is erroneous based on the nature of the document (as specified in the meta-data) and the values of the other fields. Even as simple a model as nearest-neighbor might provide acceptable classification accuracy, and for fields classified as incorrect, the audit error message for the most similar erroneous field could provide appropriate feedback to the user by describing both the error and the steps needed to correct the error.

As shown in Figure 2, an audit-based error-detecting form approach could be incorporated into the current submit/audit pipeline with minimal disruption and could significantly reduce error rates, increase compliance, and improve citizen satisfaction with interactions with the court.

Proactive Interfaces

Current web interfaces to government services typically perform no user modeling. However, recent research on proactive information search illustrates how interactions with documents can be improved by dynamically offering assistance to the user based on a user model.

Schwartz et al. (Schwartz, Berger, and Hernandez 2015) have developed a Microsoft Word plug-in that uses Latent Dirichlet Allocation (LDA) to identify and suggest prior authorities on the same topic as the paragraph that the user is typing. The system is an intelligent citation assistant that constantly updates its topic model of the document that the user is writing and proactively searches for documents whose topic-model similarity indicates relevance to the user's topic.

Similarly, *Proactive Legal Information Retrieval and Filtering*[3] is a system under development by Brian Carver and Yi Zhang at the University of California at Berkeley that infers models of the users' interests from the query histories and uses this model to anticipate the user's next queries.

Both projects illustrate how user modeling through

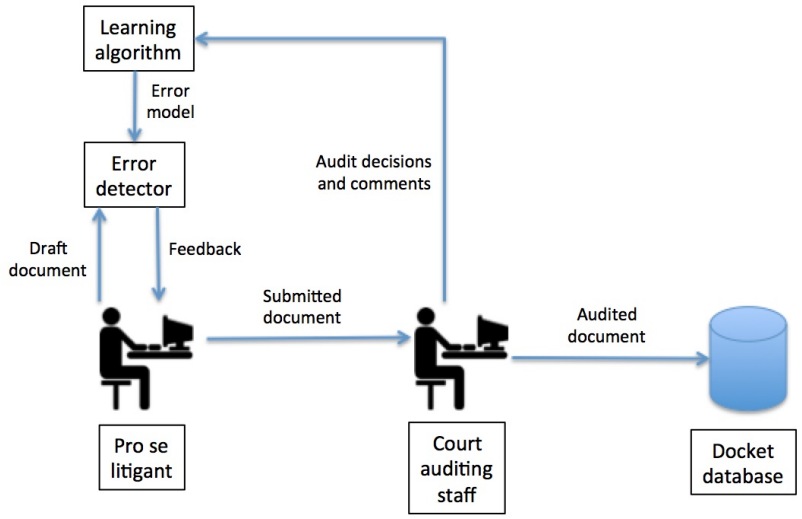

Figure 2: An audit-based error-detecting cognitive assistant.

topic or goal detection can permit a system to anticipate, cache, and suggest potential resources for achieving users' document-oriented goals.

Deductive Document Synthesis

Citizens, court officers, and other government staff often must draft documents to accomplish various legal goals. For example, as discussed above, pro se litigants must draft petitions and other legal documents to assert or defend against legal claims. Petitions in tax cases are often highly schematized and require only form-filling, but more complex legal cases, such as dissolution of marriage or obtaining a protection or restraining order, are typically less stereotypical in structure and require more than simply filling in forms. Government personnel must draft forms that optimize the ability of citizens to obtain information and benefits to which they are entitled. This section discusses deductive synthesis scenarios for pro se litigants and for court officers. For an overview of techniques and systems for legal document drafting systems, see (Lauritsen 2007).

Deductive Document Drafting for Pro Se Litigants

Pro se litigants often require help both with understanding the nature of their legal rights and with producing a document that correctly asserts those rights and supports the assertion. If a given area of law can be axiomatized with rules that include elicitation operators and performative utterances consisting of inserting text into a document, then a back-chaining inference strategy can construct a document as a side-effect of proving that the user is entitled to the benefits being sought. This approach was implemented in an advisory system for protection-order applicants prototyped at the Idaho Supreme Court and deployed in three Idaho counties and several Idaho domestic violence advocacy offices (Branting 2000). This approach is very general, with a single axiomatization supporting both a TurboTax-like interview process and document generation. The challenges of manually formalizing legal statutes make this approach feasible only for relatively simple domains, like domestic violence protection, or domains in which the formalization effort is amortized over a very large number of users, as in tax form completion.

Discourse-Based Document Synthesis

A more knowledge-intensive approach to drafting legal and other government documents uses an explicit model of the discourse structure of classes of documents to guide creation of new documents. Discourse structure consists of the relationships among statements in a multi-sentential text that are responsible for the text's coherence. The roots of this approach are in speech-act theory, which is the study of illocutionary content of discourse, *e.g.,* the goals that speakers seek to accomplish through their discourse (Grice 1975; Searle 1969). A discourse model consists of two elements: (1) an illocutionary goal structure that expresses the goal dependencies among the relevant legal predicates, and (2) the connection between performative text segments and the illocutionary goals that they achieve and a rhetorical structure that expresses the stylistic and discourse conventions of the document's genre. The illocutionary and rhetorical structures of a document together constitute the document's discourse structure (Branting et al. 1999). A prototype discourse-based system was constructed for show-cause orders, a relatively routine judicial document that is nevertheless too complex and variable to be amenable to simpler template-based approaches (Branting, Lester, and Callaway 1998). The architecture for this prototype is shown in Figure 3.

The deductive document synthesis techniques described in this section can produce complex and idiomatic legal and other government documents, but they depend on a detailed discourse model of the particular legal domain. Thus, they differ from the first three techniques which depend on

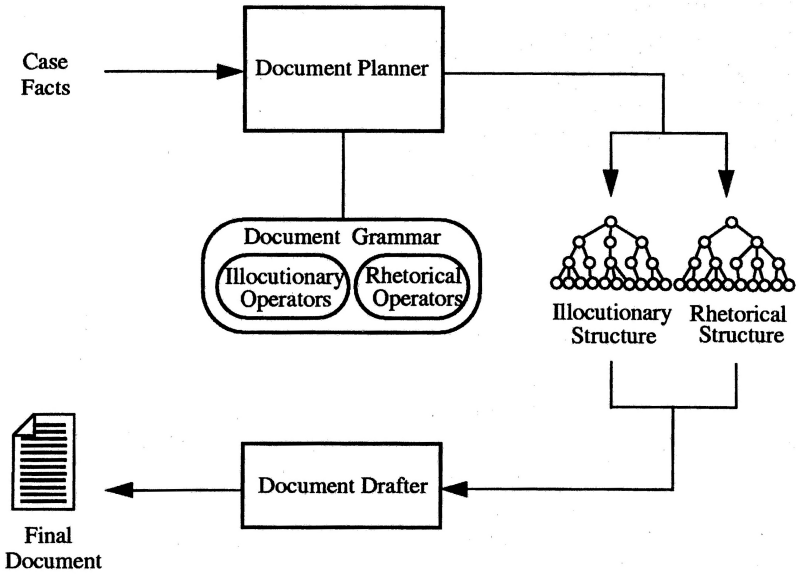

Figure 3: The architecture for discourse-based deductive document synthesis described in (Branting, Lester, and Callaway 1998).

machine learning algorithms without requiring an explicit, manual formulation of the operative legal rules.

Summary

This abstract has briefly described four classes of cognitive assistants that could improve the experience of citizens and government officials in using and understanding government documents: self-filling forms; error-detecting forms; proactive information search; and deductive document synthesis. Each of these classes of cognitive assistants has the potential to significantly improve access to justice and delivery of information, services, and other benefits to citizens by improving the ability of citizens to understand and correctly fill out forms and to comprehend informational documents.

Acknowledgment

This work was funded under contract number CECOM W15P7T-09-C-F600. The MITRE Corporation is a not-for-profit Federally Funded Research and Development Center chartered in the public interest. This document is approved for Public Release, Distribution Unlimited, Case Number 15-2786.

Notes

[1]Research on legal rule representation is published in the International Conference on Artificial Intelligence and Law, organized biennially since 1987 by the International Association for Artificial Intelligence and Law (http://www.iaail.org), the Springer Artificial Intelligence and Law Journal (http://rd.springer.com/journal/10506), and the annual JURIX conference on the Foundation for Legal Knowledge-Based Systems (http://jurix.nl).

[2]http://www.uscourts.gov/courtrecords/electronic-filing-cmecf

[3]http://citris-uc.org/ctrs-groups/project/proactive-legal-information-retrieval-and-filtering/

References

Branting, K.; Callaway, C.; Mott, B.; and Lester, J. 1999. Integrating discourse and domain knowledge for document drafting. In *Proceedings of the Seventh International Conference on Artificial Intelligence & Law (ICAIL-99)*.

Branting, K.; Lester, J.; and Callaway, C. 1997. Automated drafting of self-explaining documents. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law (ICAIL-97)*, 72–81. University of Melbourne, Melbourne, Australia: ACM Press.

Branting, K.; Lester, J.; and Callaway, C. 1998. Automating judicial document drafting: a discourse-based approach. *Artificial Intelligence and Law* 6:111–149.

Branting, K. 2000. An advisory system for pro se protection order applicants. *International Review of Law, Computers & Technology* 14(3).

Goldschmidt, J.; Mahoney, B.; Solomon, H.; and Green, J. 1998. *Meeting the Challenge of Pro Se Litigation*. American Judicature Society.

Grice, H. 1975. Logic and conversation. In Cole, P., and Morgan, J., eds., *Syntax and Semantics 2: Speech Acts*. New York, N.Y.: Academic Press. 41–58.

Hermens, L. A., and Schlimmer, J. C. 1994. A Machine-Learning Apprentice for the Completion of Repetitive Forms. *IEEE Expert: Intelligent Systems and Their Applications* 9(1).

Igari, H.; Shimazu, A.; and Ochimizu, K. 2012. Document structure analysis with syntactic model and parsers: Appli-

cation to legal judgments. In Okumura, M.; Bekki, D.; and Satoh, K., eds., *New Frontiers in Artificial Intelligence*, volume 7258 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 126–140.

Lauritsen, M. 2007. Current frontiers in legal drafting systems. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*. Palo Alto, California: ACM Press.

O'Brien, K. 1997. Pro se services in the U.S. bankruptcy courts: A national survey. Technical report, Ninth Circuit Judicial Council Subcommittee on Non-Prisoner Pro Se Litigation.

Schwartz, T.; Berger, M.; and Hernandez, J. 2015. A legal citation recommendation engine using topic modeling and semantic similarity. In *Law and Big Data Workshop, Fifteenth International Conference on Artificial Intelligence and Law*.

Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.