Making Way for Intelligence in Case Space

Philip Greenspun
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts
philg@mit.edu

Marc Lauritsen
Harvard Law School
Cambridge, Massachusetts
laurit@hulaw1.harvard.edu

Abstract

Courts should and eventually will adopt modern information systems that allow all transactions presently occurring on paper to be conducted and reviewed electronically. The widespread deployment and interconnection of such systems will pose dramatic opportunities and challenges for artificial intelligence and law. This paper describes an early effort to build a distributed system for computer-mediated litigation, and explores some of the issues involved in the intelligent use of such a system.

1. Introduction

"Paperless" litigation is still a curiosity in the United States, seen largely in connection with major cases like those involving the Exxon Valdez oil tanker disaster, the O.J. Simpson double murder prosecution, and the Keating savings and loan scandal. Electronic filing is routinely available in only a few courts, and required only in certain complex matters by special order of a judge. [Asay 1994; Yerton 1994]. It seems reasonable to assume, though, that most paper-based activities in the court system will eventually be replaced by electronic data transactions. [McMillan 1992]. The text and images of filings and the contents of judicial administrative databases (e.g.,

dockets and calendars) will be electronically accessible. Litigants will be able to file documents and access information about their own and other cases digitally and remotely. Internet browsers will enable the public to peruse court dockets and pick up audio and video streams of contemporaneous or historical courtroom activity.

It seems only slightly more speculative to assume that documents received and produced by courts will eventually be marked up with codes that identify their contents in terms of abstract categories relevant to the judicial process. For instance, paragraphs in complaints will be tagged according to their character as factual allegations, jurisdictional statements, legal claims or defenses, or requests for other judicial action. For each case in a particular court, a data structure will exist that contains the names and other characteristics of all parties, claims, witnesses, and decision makers, summaries of all formal legal events, and pointers to all documents filed or generated. Conventional querying tools will allow one to ascertain, for example, the average award in product liability actions against asbestos manufacturers in which the plaintiff was 45 years or older and resident in a southern state. For purposes of this article, we call this distributed database "case space".

We have developed a tool for helping litigants compose pleadings and make related moves in a controversy brought to court. We have also built facilities for using this tool and accessing a database of disputes modelled in it over the Internet via the World Wide Web. We will briefly summarize the

workings of this system in Part 3 of this article. In Part 4, we take up several questions about how one might make intelligent mechanical use of the distributed database of judicial activity that would result if such a system were adopted widely.

2. Motivations

Lawyers and judges have been privileged relative to other professions because so much of the raw material of their craft—judicial opinions, statutes, regulations, and secondary material—has long been available online in full text form. Several graduating classes of law students in the United States have now been accustomed to having free round-the-clock access to these enormous textual repositories. Natural language tools for querying the online databases have emerged, and research proceeds in many quarters on next-generation interfaces that will likely combine multiple paradigms for online research. At least in most Western countries, we are rich in legal information and tools for using it.

But this substrate of raw text, only marginally segmented and parameterized, hardly suffices for some of our informational and analytical needs. Having access to the text of hundreds of thousands of judicial opinions is wonderful, but they represent only a small percentage of the cases filed, and even in those cases rarely contain any of the authentic pleadings or evidence. Most are appellate opinions that provide only summaries of the proceedings "below". However central such texts are to our jurisprudence, they provide a woefully incomplete picture of what actually goes on in the courts, and yield to automated analysis only insofar as our computers can understand natural language and apply common sense.

Once businesses, consumers, courts, and lawyers are automated and networked, it becomes natural to want access to the full grist of litigation: pleadings, motions, orders, opinions, transcripts, and documentary evidence such as the text of contracts, engineering drawings in product liability cases, and maps and photographs of crime scenes. The likely demand for and increasing technological feasibility of achieving this form of access opens up the prospect of an alternative to the full-text substrate described above. A good substrate that we can build with current technology is one that

- does not depend on natural language understanding, common sense reasoning, or laborious formalization of facts,

- can present multiple views of disputes simultaneously (plaintiff, defendant, judge, clerk, public),

- reduces the administrative burdens of a lawsuit, saving time for both parties and the judge,

- incrementally builds a substantial database of formalized legal argument, and

- uses the Internet/Web infrastructure to make the database available to legal and computer science researchers and to open the courts to the public.

We believe that it will be upon such a substrate that machine learning programs will learn the law, probabilistic reasoning systems will make reasonable predictions, and new ideas for computer-aided law will grow.

3. KTA: A System for Computer-Aided Litigation over the Internet

KTA (Knowledge Theorist for Attorneys) is a Common Lisp program that captures structured legal argument from litigants and judges. KTA includes productivity tools for those directly involved in a lawsuit and a World Wide Web interface so that the public can browse court dockets.

Briefly, our design goals were to

- create a unified database for all information in a lawsuit,

- protect confidential and privileged information,

- eliminate all physical transmission and storage of documents,

- supply a unified user interface for each class of user (e.g., judge, litigant),

- capture legal argument in a finer grained form than current pleading documents,

• have a strong enough model of procedure to explain cases to lay people, and

• achieve isomorphism with current paper practices.

A. Database

At the core of KTA is a distributed object database. As shown in Figure 1, the views of the database presented to parties, the judge, the court clerk, and the public may be quite different. In particular, the clerk's view is the database of record and contains filed, immutable pleadings, motions, and documents. A party's private view may contain draft pleadings as well as argument and authority that he has not yet been forced by the rules of procedure to reveal to the court or his opponents. A judge's private view may contain draft opinions. The public's view via the Web obscures documents that have been filed under seal and contains additional computer-generated explanations of jargon and court procedure.

KTA's object database is distributed in the sense that objects are stored across multiple networked computers. It is also distributed in the sense that the instance variables of a single object may be stored across multiple computers. For example, in the complaint that starts a lawsuit, a plaintiff need only claim "breach of warranty." The lawyer who files the complaint may have already looked up case law to find authorities that give the elements of breach of warranty. He may have already associated facts asserted by the client with each of those elements. He may have already associated pieces of evidence brought to him by the client with each of those facts. Figure 2 shows an example of a claim for intentional infliction of emotional distress, where some material is kept on the plaintiff's machine, shielded from other parties until the plaintiff chooses to expose it.

B. Capturing legal argument in fine-grained form, or "If you have strong syntax, maybe you don't need semantics"

A plaintiff does not file a complaint in KTA as a monolithic block of ASCII text. It is broken up into allegation objects and claim objects. The plaintiff is encouraged but not required to pick a claim from a standard set so that the system automatically picks up a canonical title such as "breach of contract". The

claim object has a slot for arbitrary text, perhaps setting forth which of the allegations lead to this claim, and a slot for a text prayer, e.g., asking the court for $10,000 in damages. KTA makes no attempt to interpret either of these text strings. To facilitate automated processing, the claim object also contains a slot called PRAYER-DOLLAR-AMOUNT that the plaintiff is encouraged to fill in with the number, e.g., 10,000. This enables an expected value calculation to be performed on a lawsuit without needing to wade into text prayers and pull out the numbers with natural language smarts.

The claim contains slots called AUTHORITY and RESPONSES. If the user picked the claim from an on-line horn book or practice guide, then the authority for that claim is preserved. This doesn't go to the court, but may prove useful a year later when a new associate takes up the case by providing hypertext access to, for example, the statute that the filing attorney thought relevant.

RESPONSES is a list of responses to the claim by opposing parties and the court. When a defendant files an answer, it is in the form of response objects linked back to their counterparts in the complaint, plus new allegations and affirmative defenses. The defendant is encouraged but not required to categorize his response as an "admission", "partial admission", "partial denial", "denial" or "insufficient information". He can choose the "other" category, but can very likely save time by choosing one of the predefined categories because KTA automatically brings in appropriate boilerplate.

The overriding theme here is that structure is captured when possible, in a manner that is painless for the participants, but KTA never limits what a party can ask of a court. Anything that can be pleaded on unstructured paper can be pleaded in KTA.

When a judge wants to review the status of a case, KTA produces an easily understood hypertext outline. Allegations or claims that have been admitted or previously decided are shown in special colors so that the judge can concentrate on parts of the case that remain in dispute. (Figure 3 illustrates what a judge may see mid-way through a lawsuit.)

KTA's attempt to capture legal argument does not stop with the parties. The judge is encouraged to rule

in a structured manner, allegation by allegation and claim by claim. This enables one to ask a computer to "show me all the breach of contract claims that Judge Smith denied in the last two years."

C. World Wide Web interface

Although it is theoretically possible to represent KTA's rich set of linked objects with SGML, it is certainly impractical to do so with HTML, a simple SGML document type definition that is the lingua franca of the World Wide Web. [Berners-Lee 1992]. KTA makes no use of SGML or HTML in its internal database, but is capable of walking through its network of Common Lisp Object System objects and reformatting all the information into HTML files ready for installation on a Web server. Some of the data is even improved in the process. For example, KTA maintains a small dictionary of legal terms (a subset of Black's Law Dictionary) and walks through strings of English text as it writes them into HTML files. When KTA finds a word that is in the dictionary, it links that word to its definition, something that might be very useful to a lay person checking out a local court's docket.

With the KTA Web interface, court documents that are supposed to be public become truly public, available to anyone in the world with an Internet connection, 24 hours a day.

D. Software Status

Much more work remains to be done on KTA to make it a practical system for courts and attorneys. High priority items would be the following:

- mating a reliable disk-based object database to the current memory-based object database, and

- implementing classes for modelling complex legal argument such as that found in summary judgment motions.

KTA is free software that will be available to anyone on the Internet under the same conditions as tools such as GNU Emacs. It is currently limited to running on Apple Macintosh computers due to its use of the Macintosh Common Lisp window system and user interface extensions. There are a few files of machine-dependent code, some of which are fairly

complex (e.g., for drag-and-drop editing).

4. Discussion

A. Do we need "AI"?

Most of what arguably needs to be done in the judicial information processing arena—and has been prototyped with KTA—can be achieved with thoroughly conventional computer science techniques. There is a vast range of improvements that could be made by applying well understood and non-experimental technologies to our court system. The difficulties there seem overwhelmingly sociological, political, and organizational ones.

To be sure, there will be enormous engineering challenges in building, maintaining, and querying massive distributed databases of judicial information. Standards will have to be developed. Policy will need to be articulated and executed at many levels of federal and state government. But these challenges are comparable to those involved in other large transactional systems like airline and hotel reservation networks.

Our view is that artificial intelligence is of little importance to the basic task of "electrifying" the courts, but will play an increasingly central role in the legal world that results after that process has occurred. We need to begin thinking systematically about the knowledge representations and processes needed to perform these activities. A few main realms of effort come to mind.

B. Enforcing the rules

A natural first context for deploying knowledge-based technology is that of developing clerical/administrative tools to monitor and enforce compliance with procedural rules. While some rules can be straightforwardly modelled in procedural code—such as having to file an answer within a certain number of days after the service of a complaint absent court approval, others involve complexities (such as context dependence and defeasance) that may be best handled with declarative knowledge representation techniques. The early work by Jeff Meldman [1978] on the use of Petri nets to model civil procedure suggests the surprising complications one can encounter. One more recent and very elegant approach is Thomas

Gordon's normative formalization and computational implementation of civil pleading [1993].

C. Following the rules

Another obvious place for AI is in advisory systems for litigants, attorneys, and judges. Such systems ought to be able to advise parties what moves are possible or recommended in given contexts. A good model of procedure is needed before a system like KTA can generate good explanations for lay people.

Companies and individuals will be interested in programs that monitor the net and notify users when claims are asserted or post-complaint filings are made. A firm or government agency that frequently finds itself in litigation can be greatly assisted both in terms of efficiency and quality of effort.

Case-based reasoning models will likely play a role here. Strategic reasoning will also be called for as multiple intelligent players try to "game" the system.

D. Making the rules

Court and legislatures contemplating proposed reforms will want tools that access the distributed case database as an information refinery in aid of legislative policy judgments. ("If we shorten the statute of limitations on mail fraud claims, how is that likely to affect the number of filings?")

Similarly, parties to disputes will often find themselves better off enacting "private legislation" through forms of computer-aided negotiation. Tools like KTA can assist people in narrowing the scope of disagreement, assessing probabilities of success, and identifying opportunities for win-win solutions.

E. Calling the shots

Perhaps the most ambitious efforts in this area will be those that attempt to exploit systems like KTA to aid judges in making decisions and writing opinions. Karl Branting [1993] has described an issue-oriented approach to judicial document assembly that involves the structured entry of case information and subsequent rule-based guidance of a judge through decision making and justification. One can imagine decision-support tools that make such use of large databases of similar electronic case records.

5. Conclusion

Even an elaborately developed and painstakingly implemented electronic data system for court information will fail to capture the full particularity of human controversies that make up the business of the judicial system. Without robust natural language and common sense abilities, no computer system will yield results regularly comparable to those of competent jurists in participating in and making sense of most judicial activity. But a great deal of the transactions taking place in court can and should be supported by more intelligent information technology. If the rule-governed and formulaic aspects of judicial activity can be crystallized out of the muddle of its human context and subjected to intelligent distribution and analysis, gains in rationality and efficiency will be achieved that might ultimately yield better justice.

References

Alan Asay. Toward Paperless Utah Courts. 1994. MIT Journal of Computer-Aided Litigation, http://kta.org/mjcal.

Berners-Lee, T.J., Cailliau, R., Groff, J-F, Pollermann, B., CERN 1992. "World-Wide Web: The Information Universe", in "Electronic Networking: Research, Applications and Policy", Vol. 2, No. 1, pp. 52-58. Meckler Publishing, Westport.

L. Karl Branting. An Issue-Oriented Approach to Judicial Document Assembly. In Proceedings of the Fourth International Conference on Artificial Intelligence and Law, pp. 228-235. ACM, 1993.

Thomas F. Gordon. The Pleadings Game. In Proceedings of the Fourth International Conference on Artificial Intelligence and Law, pp. 10-19. ACM, 1993.

Jeffrey A. Meldman. A Petri-Net Representation of Civil Procedure. IDEA 19(2). 1978.

Jim McMillan. Judicial EDI: The Need for National Standards. The Court Manager. Winter 1992, pp. 17-22.

Stewart Yerton. "CLAD" for Litigation. The American Lawyer. October, 1994, pp. 111-112.

Plaintiff

Draft Motion for Summary Judgment
　Authorities
　Evidence

Defendant

Dynamic Trial/Settlment Model

Draft Protective Order Request

Judge

Draft Denial of Motion to Dismiss
　Presumptions of Fact

Internet

Court Clerk

π: Complaint
　Δ: Answer
π: Interrogatories
　Δ: Answer
π: Request for Documents
　Δ: Response
Δ: Interrogatories
　π: Response
　　Δ: Motion to Compel
　　　π: Opposition Memo
　　　J: Allowed
Δ: Motion to Dismiss
　π: Opposition Memo
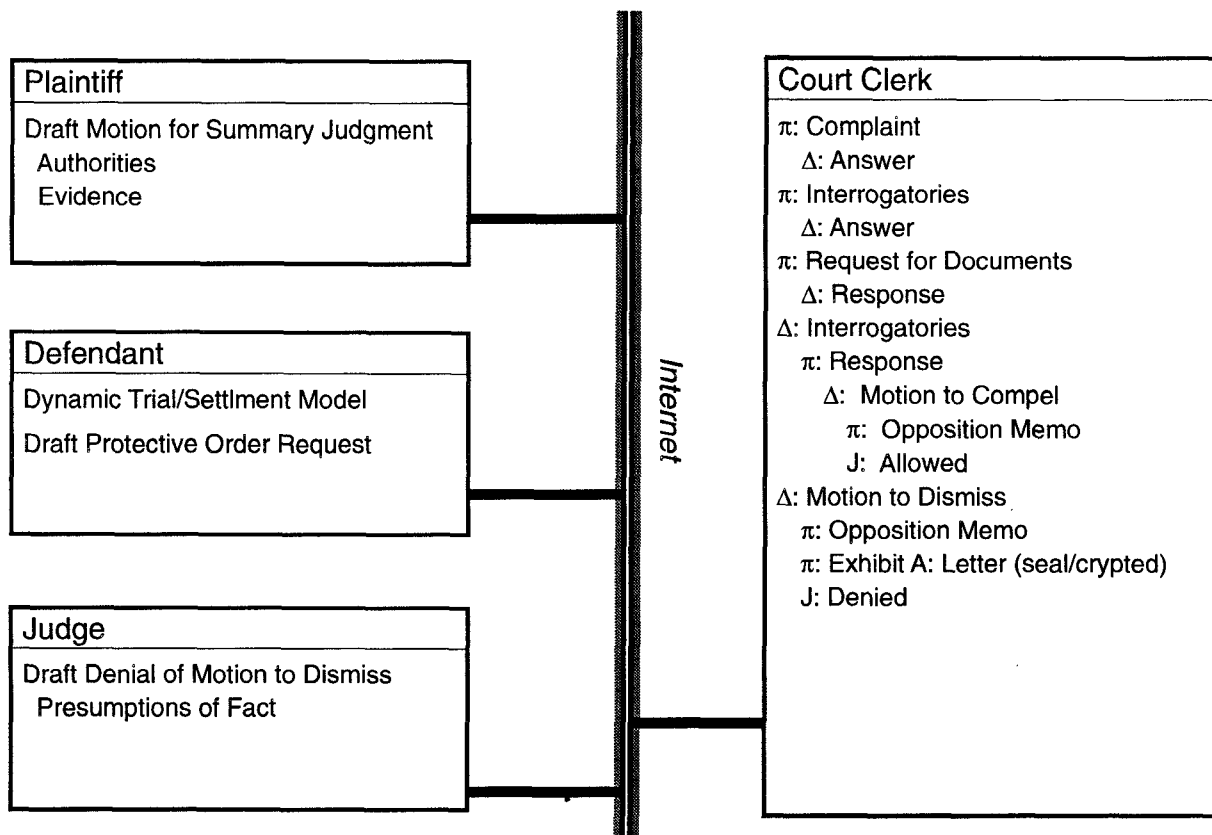　π: Exhibit A: Letter (seal/crypted)
　J: Denied

Figure 1: KTA is a system for structuring and storing legal discourse. Each box above represents the portion of a distributed database that is stored on a lawsuit participant's computer. The court clerk's computer holds the primary record of the case, i.e., all the information that has been filed, most of which is publicly accessible (in this case, there is only one letter, an exhibit for a motion to dismiss, that is sealed and protected via encryption). Note that a π in front of a "document" indicates that it was filed by the plaintiff, a Δ by the defendant, a J by the judge.

The plaintiff's local database contains only a draft motion for summary judgment, currently hidden from all other parties in this action.

A dynamic trial/settlement model on the defendant's machine reflects his concern about the ultimate cost of this case. This model is updated when events in the litigation change probabilities. For example, if an motion to exclude evidence succeeds, that may make it difficult for the plaintiff to prove a critical element of a claim. The probability of loss on that claim will be reduced, thus reducing the expected cost of the case. Because over 90% of cases settle before trial, this kind of information can be very valuable in a complex case.

The judge in this case has retained her draft denial of the defendant's motion to dismiss because she may want to use the same presumptions of fact in hearing a summary judgment motion. KTA assists judges by outlining cases and could quickly show which claims cannot succeed given certain assumptions about the facts.

Plaintiff's View   *Public View*

Claim—Intentional Infliction of Emotional Distress

Text: "Defendant Dr. Joe Smith D.D.S. is ...
Prayer: "Plaintiff respectfully asks the court...
Prayer-Amount: $100,000
 	
————————Private————————
 	
Element-Authority: <u>Prosser</u>
Element1: Defendant Act, Extreme/Outrageous
  Evidence Supporting: <u>letter 3/12/93</u>
  Evidence Supporting: π testimony
  Evidence Supporting: π wife testimony
  Authority: <u>Figueiredo-Torres v. Nickel</u>
Element2: Intent
  Evidence Supporting: ?
Element3: Causation
  Evidence Supporting: Dr. Yee testimony
Element4: Severe Emotional Distress
  Evidence Supporting: W-2 forms

Claim—Intentional Infliction of Emotional Distress
Text: "Defendant Dr. Joe Smith D.D.S.committed the tort of intentional infliction of emotion distress by initiating an affair with plaintiff's wife while plaintiff was a patient." Prayer: "Plaintiff respectfully asks the Court for $100,000 in compensatory damages plus reasonable attorney's fees and costs and whatever additional damages the Court deems appropriate." Prayer-Amount: $100,000 Authority: NIL Responses: <u>Denied by Δ</u>

Figure 2: The plaintiff's view of his breach of his intentional infliction of emotional distress claim is very different from the public's at this early stage in litigation. American civil procedure requires a plaintiff only to state a claim by name and hence that is all that shows up in the public view: "intentional infliction of emotional distress." Note that in the Plaintiff's private view, his attorney has sketched in some notes about what the elements of this claim are according to a legal authority and what pieces of evidence exist to support those elements.

KTA's distributed object database keeps private information on the plaintiff's private machine until it becomes public on the plaintiff's initiative. Thus, the letter of 3/12/93 that supports the element of an outrageous defendant act, is available as a hypertext link (note the underlining) but will be supplied from the local disk. The Figueiredo case (in which a psychiatrist had an affair with a patient's wife) is also available as a link, to be supplied either from a local cache, the World Wide Web, or a commercial service such as Lexis.

Note that in the public view of the case, no authorities have been cited by either side to support or attack the fit of the facts of this case to the claim of intentional infliction of emotional distress. That is presumably because no motions for summary judgment have been filed. In any case, the defendant apparently denied the claim; the full text of the denial would be available with a mouse click.
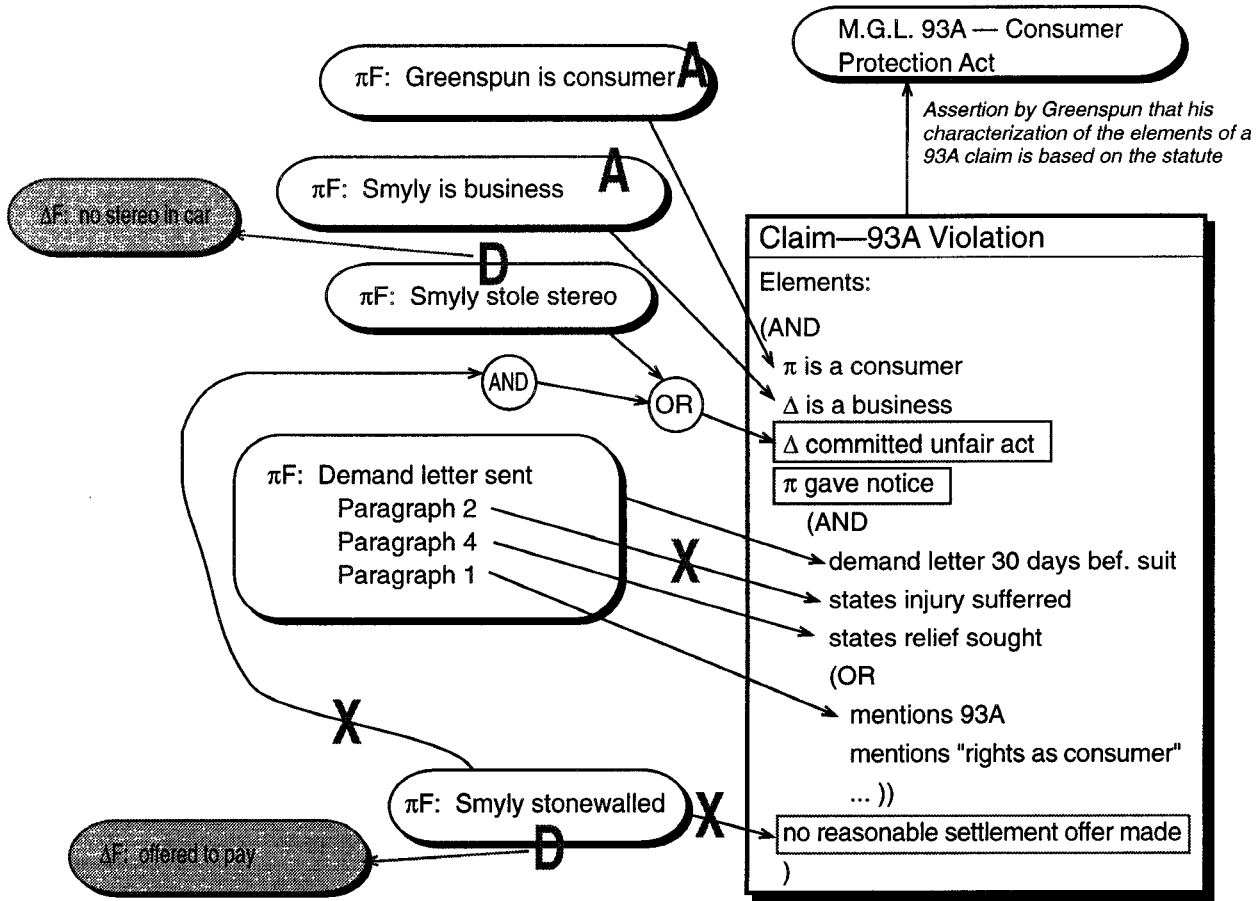
Figure 3: What a Massachusetts judge might see mid-way through a lawsuit. Greenspun is the plaintiff. He alleges that Smyly Dodge stole his car stereo while servicing his car. The ovals marked "πF" are facts that Greenspun asserts support all the elements of a Consumer Protection Act claim. While it might be valuable to present one side's case to a judge in just this schematic form, it is even better to see the other side's counterarguments superimposed in a different color (alas impossible in this proceedings).

Big A's indicate that Smyly admits that Greenspun is a consumer and that they are a business. The big D's and supporting defendant facts ("ΔF") are part of Smyly's argument but another critical part is show by the big X's. What is being disputed with the topmost X is not that the demand letter contained a 2nd paragraph, but whether or not the text of that paragraph stated the injury suffered sufficiently precisely to satisfy the statute. Becuase of this X, KTA has marked the element "π gave notice" under dispute by surrounding it with a red rectangle (in practice, this is done on-screen by displaying the text in red).

(Note: this figure is loosely based on a real lawsuit, Greenspun v. Smyly Autos, but does not accurately represent the legal arguments actually made. The original documents are available on the Web at http://smyly.com/smyly)