

Structured Casenotes:

How Publishers Can Add Value To Public Domain Legal Materials On the World Wide Web

L. Thorne McCarty
Computer Science Department
and
Faculty of Law
Rutgers University
New Brunswick, NJ 08903 USA
`mccarty@cs.rutgers.edu`

1 Introduction

The days of proprietary legal databases are rapidly ending, as primary legal materials become available on the World Wide Web. In the United States, for example, the decisions of the Supreme Court are now available on line all the way back to 1893. The decisions of the U.S. Courts of Appeal are also available, dating back several years in most cases. State materials are not as complete, but many state courts have recently matched the federal courts in putting their decisions on line. For statutory materials, the record is even stronger. There are multiple sources for the United States Code and the Code of Federal Regulations, and complete statutory databases are now accessible on line in at least 41 states.

How can legal publishers adapt to this new environment? Most analysts agree that legal publishers must now add value to public domain legal materials (statutes, cases, etc.) in the form of secondary legal analysis. Although publishers have always done this, the added value has traditionally taken the form of additional texts: treatises, annotations, practitioner's guides, etc. A good example would be the West Headnote and Key Number System, which

dates back more than a century. A *headnote* is a summary of a single point of law in a particular case, and a *key number* is an assignment of that headnote to a particular location in a massive classification scheme for American law. In principle, this system should allow us to locate cases that discuss similar legal issues, and it should be as useful in an electronic environment as it was in a manual environment. In practice, however, since the headnote is just a piece of free-form text, the only way to retrieve it is by means of a key-word search, and this tends to be very imprecise.

In this paper, I will argue that publishers should be adding value to public domain legal materials, not in the form of additional texts, but in the form of *computational structures*, using recent advances in Knowledge Representation (KR) and Natural Language (NL) techniques. I will illustrate my point with an example: *structured casenotes*. A structured casenote is a computational summary of the procedural history of a case along with the substantive legal conclusions articulated at each stage of the process. It would play the same role in the legal information systems of the 21st century that West Headnotes and Key Numbers have played in the 20th century.

To see the difference in these two approaches, I will focus most of the discussion on a single case: *Quality King Distributors, Inc., v. L'Anza Research International, Inc.*, 118 S.Ct. 1125 (1998), which was decided by the U.S. Supreme Court on March 9, 1998. In Section 2, I will describe the traditional editorial enhancements to this case, written by West case editors and classifiers. In Section 3, I will describe what a structured casenote would look like for the same case. In Section 4, I will discuss the research that still needs to be done to make this vision a practical reality.

2 Traditional Headnotes and Casenotes

There are three opinions in *Quality King v. L'Anza Research*, two of which are available free of charge on the World Wide Web. The Supreme Court opinion is available at:

<http://supct.law.cornell.edu/supct/html/96-1470.ZS.html>

and the opinion of the Court of Appeals for the Ninth Circuit is available at:

<http://www.vcilp.org/Fed-Ct/Circuit/9th/opinions/9556447.htm>

The original District Court opinion was not officially published, but it is available, for a fee, through WESTLAW, with the citation 1995 WL 908331.

If you retrieve the Supreme Court opinion or the Ninth Circuit opinion from WESTLAW, however, you also receive the editorial enhancements that have been written by West case editors and classifiers. These include a synopsis of the case, and a series of headnotes with their associated key numbers.¹ For example, here is the synopsis for the Ninth Circuit opinion:

Manufacturer of copyrighted hair care products brought action against importer that acquired products outside United States and sold them within United States. The United States District Court for the Central District of California, J. Spencer Letts, J., entered injunction in favor of manufacturer, and importer appealed. The Court of Appeals, D.W. Nelson, Circuit Judge, held that: (1) first sale doctrine did not bar action; (2) copyrighted goods that manufacturer sold to British distributor were “acquired outside the United States” for purposes of the statute prohibiting importation of United States copyrighted goods acquired outside United States without authorization of copyright owner; (3) manufacturer did not violate trade embargo against Libya, so as to subject it to unclean hands defense; and (4) injunction was not unduly burdensome.

Affirmed.

This is followed by ten headnotes that are linked to particular paragraphs in the text of the opinion. For example, here is the headnote most closely linked to the first holding of the synopsis:

[4] First sale doctrine did not bar action involving exportation and later unauthorized importation of United States copyrighted goods originally manufactured in United States. 17 U.S.C.A. §§109(a), 602(a).

This headnote has the string **99k38.5** attached to it. **99** is the *Digest Topic Number*, and it refers to the topic “Copyrights and Intellectual Property”. **38.5** is the actual *Key Number*, and it assigns the headnote to the following

¹All synopses and headnotes reproduced in this paper are Copyright 1998 by West Group.

location in the classification hierarchy under “Copyrights and Intellectual Property”:

I. Copyrights

(B) Scope

k35. Scope of Exclusive Rights; Limitations

k38.5. Effect of transfer of particular copy or phonorecords.

The other headnotes for this opinion are classified in Digest Topic Number 170B, “Federal Courts”; Number 114, “Customs Duties”; Number 150, “Equity”; and Number 402, “War and National Emergency”.

As it turned out, the Supreme Court reversed the decision of the Ninth Circuit Court of Appeals on the copyright issue. Here is the synopsis written by the West case editors:

Manufacturer of copyrighted hair care products brought action against importer that acquired products outside United States and sold them within United States. The United States District Court for the Central District of California, J. Spencer Letts, J., 1995 WL 908331, entered injunction in favor of manufacturer. Importer appealed. The United States Court of Appeals for the Ninth Circuit, 98 F.3d 1109, affirmed. Importer filed petition for writ of certiorari. The Supreme Court, Justice Stevens, held that first sale doctrine, under which owner of particular copy is entitled, without authority of copyright owner, to sell or otherwise dispose of possession of that copy, is applicable to imported copies.

Reversed.

Justice Ginsburg filed concurring opinion.

Notice that the language of the Ninth Circuit synopsis has simply been repeated here, up to the point of the first appeal.

Let us now look at the headnotes for the Supreme court opinion. There are only three, all classified under **99k38.5**. (This is because the copyright issue was the only point upon which the Supreme Court had granted certiorari.) Here are the first two:

[1] First sale doctrine, under which owner of particular copy is entitled, without authority of copyright owner, to sell or otherwise dispose of possession of that copy, is applicable to imported copies. 17 U.S.C.A. §§106(3), 109(a), 602(a).

[2] Under first sale doctrine, lawful owners of hair care products bearing copyrighted labels did not engage in copyright infringement by importing and reselling products without manufacturer's authority. 17 U.S.C.A. §§106(3), 109(a), 602(a).

Notice the subtle syntactic difference between these two headnotes. Headnote [1] is written in the present tense, and it describes a general principle of law. This is known as an *abstract* headnote. Headnote [2] is written in the past tense, and it describes how a general legal principle (in this case, the “first sale” doctrine) was applied to the specific facts of the case. This is known as a *concrete* headnote. In fact, the syntactic difference between these two headnotes is not an accident, since tense is used systematically by West case editors to encode the distinction between the two styles!

I hope these excerpts from *Quality King v. L'Anza Research* have conveyed to the reader the extraordinary amount of effort devoted to editorial enhancements at West. West handles an input of several hundred cases per day, and every published opinion must be processed by a case editor, who writes the synopsis and the headnotes, and a classifier, who assigns the key numbers. Although some cases are short, and may have only a couple of headnotes, it is not uncommon for the longer cases to have 30 or 40 headnotes. There are almost 100,000 distinct key numbers, and the published version of the classification hierarchy consumes over 1500 pages of text in a double-column format. It is amazing that West is able to keep this system up to date.

With all this effort, however, remember that the output of this process is just more unstructured text, and it is not clear that this is what we want in today's computerized environment. In some ways, the syntactic encoding of the distinction between abstract and concrete headnotes is symptomatic of the problem. Assuming that such a distinction is important, shouldn't we store it explicitly in the headnote database, instead of encoding it implicitly by means of the tense structure of natural language? Explicitly stored, it could be used efficiently by a search algorithm; embedded in the syntax of a sentence, it can only be perceived by the human mind.

This is the essential idea of structured casenotes: Encode the important information about a case in an explicit data structure, so that it will be readily accessible to our search algorithms. In the following section, I will explore this idea in greater detail.

3 Structured Casenotes

What would a structured casenote for *Quality King v. L'Anza Research* look like? The basic idea is to model the traditional “brief” that students are taught to write in their first year of law school. Instead of writing this brief in English, however, we will write it in a computable form using recent advances in Knowledge Representation (KR) and Natural Language (NL) techniques.

The traditional case brief focuses on the procedural context first: Who is suing whom, and for what? What is the plaintiff’s legal theory? What facts does the plaintiff allege to support this theory? How does the defendant respond? How does the trial court dispose of the case? What is the basis of the appeal? What issues of law are presented to the appellate court? How does the appellate court resolve these issues, and with what justification? (For a criminal case, these questions would be modified accordingly.) These are highly stylized questions, of course. To ask them and answer them computationally, we need a representation of the rules of civil procedure at some reasonable level of abstraction in some appropriate Knowledge Representation Language. (The plan is to use a version of my *Language for Legal Discourse* for this purpose, although I will not discuss the technical details in the present paper.) We also need a Natural Language Grammar with coverage of the procedural expressions that occur in the synopsis of a typical case. The objective is to be able to represent the meaning of a sentence such as:

Quality King appeals the district court’s order specifying issues existing without substantial controversy and granting summary judgment in favor of L’Anza.

in a computable form. (I will say more about the research that is still needed to reach this objective in Section 4 below.)

Within this procedural framework, we would represent the substantive issues at stake in the decision. This is more complicated, since the territory is so vast, potentially encompassing the entire legal system. We can get started on this project, however, by focusing on particular areas of the law where there is enough interest (and enough money!) to support the writing of treatises, casebooks, hornbooks, etc.

Take copyright law, for example. The West headnotes are not very informative about the nature of the dispute in *Quality King v. L'Anza Research*, but there is a passage in the Ninth Circuit opinion that is crystal clear. I will reproduce it here in full to illustrate what a structured casenote should be able to represent:

Quality King asserts that the district court erred in holding that L'anza's claim under §602(a) of the 1976 Copyright Act is not barred by the first sale doctrine of §109(a). Section 602(a) provides in relevant part, that:

Importation into the United States, without the authority of the owner of copyright under this title, of copies or phonorecords of a work that have been acquired outside the United States is an infringement of the exclusive right to distribute copies or phonorecords under section 106, actionable under section 501. 17 U.S.C. §602(a).

Section 602(a), thus, gives the copyright owner a distribution right under §106. Section 106 provides in relevant part, that:

Subject to sections 107 through 120, the owner of copyright under this title has the exclusive rights to do and to authorize any of the following:

...

(3) to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease or lending;

17 U.S.C. §106(3).

The first sale doctrine, however, limits the rights embodied in §106(3), providing that:

Notwithstanding the provisions of section 106(3), the owner of a particular copy or phonorecord lawfully made under this title, or any person authorized by such owner, is entitled, without the authority of the copyright owner, to sell or otherwise dispose of the possession of that copy or phonorecord. 17 U.S.C. §109(a).

From the text of the 1976 Copyright Act, it is unclear whether §602(a) creates a right that is distinct from §106(3) and therefore is not limited by §109(a), or alternatively, whether §602(a) is merely an extension of §106(3) and therefore is limited by §109(a).

In other words, the main issue in the case boils down to a choice between the two alternatives stated in this last sentence. The Ninth Circuit chose the first alternative, while the Supreme Court reversed and chose the second.

A structured casenote would represent these alternatives by representing the various possible relationships between the rules of §§106(3), 109(a) and 602(a). This would require us to represent the content of the rules themselves, of course, but that is part of the task of writing a treatise on copyright law. We can assume that the various objects (e.g., “copies” and “phonorecords”), relations (e.g., “owner”) and actions (e.g., “import” and “distribute”) have been encoded in our knowledge representation language. We could then represent the possible outcomes in the case by applying §§106(3), 109(a) and 602(a) with alternative scope constraints. The first alternative would be represented by applying §602(a) directly, while the second alternative would be represented by incorporating §106(3) with its §109(a) limitation into the language of §602(a). Moreover, close attention to the procedural context of the case will help us to do this correctly. In the District Court, for example, the plaintiff (L’Anza Research) initially filed a complaint based on §602(a), and the defendant (Quality King) responded by raising an affirmative defense based on §§106(3) and 109(a). Thus the issue was framed in this way from the very beginning.

Now imagine that we have succeeded in writing a structured casenote for *Quality King v. L’Anza Research*. There are other cases that raise similar issues: *Sebastian Int’l Inc. v. Consumer Contacts, Ltd.*, 847 F.2d 1093 (3d Cir.1988) and *Columbia Broadcasting System, Inc. v. Scorpio Music Distributors, Inc.*, 569 F.Supp. 47 (E.D.Pa.1983), aff’d, 738 F.2d 421 (3d Cir.1984) were two cases discussed in the Ninth Circuit opinion, for example.

The structured casenotes for these opinions would have subtle differences, either in the operative facts, or in the justification of the decision, or both. We would thus have a basis for recognizing overall similarities among these cases, as well as a basis for drawing sharp distinctions.

We could, if we want, map our structured casenotes to traditional West key numbers. The main issue in *Quality King v. L'Anza Research* would presumably be classified as **99k38.5**: “Effect of transfer of particular copy or phonorecords”. But notice that the casenote itself is much more informative than the key number. Although key numbers would be useful in establishing linkages back to legacy databases, the classification system inherent in our representation of copyright law is all that we would need to search through a database of structured casenotes.

4 Current Research

Most lawyers would agree that a database of structured casenotes, as described in the previous section, would be much more useful than our current database of headnotes and key numbers. But they would also raise a serious objection to the proposal: Is it possible to build such a database at the appropriate scale?

I agree that this is the main obstacle, and it is the subject of most of my current research. There is one key point to keep in mind, however. Although it will surely be expensive to create a database of structured casenotes, it is also very expensive to maintain the current database of headnotes and key numbers. So the question should be: What is the most effective way to utilize the time and talents of a group of people such as the West case editors and classifiers? If we can provide these people with a good set of software tools to make their work more automatic, it may turn out that the marginal returns on a system of structured casenotes will be higher than the marginal returns on the current system.

As a first step towards answering this question, I am studying the feasibility of extracting the procedural history of a case from the raw text of a judicial opinion. (Recall that the procedural history provides the main organizational framework for a structured casenote.) There are several possibilities:

1. The most ambitious approach is to use Natural Language (NL) tech-

niques to parse the text of the opinion itself, extracting and analyzing the language that addresses the procedural questions. This is a difficult computational task, but recent advances in statistical/lexical parsing suggest that it is worth trying. As a foundation for this approach, we need not only a good NL grammar for the procedural expressions that tend to occur in judicial opinions, but also a good Knowledge Representation (KR) system for the rules of civil procedure.

2. It is unlikely that fully automated natural language processing will be able to solve all of our problems, and I am also studying semi-automated systems, in which a case editor is provided with automated tools to assist in writing a casenote. Two approaches seem promising:
 - (a) If the unrestricted text of a judicial opinion is too complex for our parser, it may still be possible to describe the case in “controlled English”. In this approach, the user communicates in a simplified NL grammar, which the system can understand, but which is still rich enough to capture the essential information about the case. Experiments with natural language interfaces have shown that people adapt quickly to the idiosyncracies of a restricted grammar, as if they are speaking to someone who doesn’t quite understand their language. A key ingredient of this approach, though, is the representation of the rules of civil procedure. If the KR system is not adequate, the case editor will not be able to describe the case, even in a simplified form of English.
 - (b) The final approach is to use a KR-based menu system, or a KR-based graphical system. The basic idea here is that the procedural possibilities are often tightly constrained, and thus it may be sufficient in many cases just to provide a set of choices to the human user. Again, the critical ingredient in this approach is the representation of the procedural possibilities, and this is a fundamental KR problem.

In practice, I would expect some combination of these approaches to provide the best solution.

If this work on procedural history is successful, we will then have a way to focus on the main substantive issues in a case. At this point, we will have

to choose one or more substantive areas of law (e.g., copyrights, secured transactions, corporate reorganizations, etc.) and apply the same NL and KR techniques there.

I will report on the progress of this work in subsequent papers.

CITATION: This paper was originally presented at the *Second French-American Conference on Law and Computing*, held in Nice, France, in June of 1998.