EMPIRICAL LEGAL ANALYSIS

An Interactive Qualifying Project Report

submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

Zachary A. Kleinfeld

and

James A. Roumeliotis

Date: May 2, 2007

Professor Kent J. Rissmiller

1. Content Analysis
2. Computers
3. Law

Table of Contents

Abstract

Content analysis is used to study interpersonal communications. This project attempts to test the feasibility of adapting computers to content analysis, using opinions of the Supreme Court of the United States as a data set . Using textual search algorithms in several experiments, we contrast traditional human analysis with computerized techniques, and highlight their strengths and weaknesses. We conclude that computerized methods have potential to make analysis more efficient, but are not easily adapted to answering qualitative questions.

Executive Summary

As computer technology improves, advancements propagate to all fields of study. In this project, we attempt to determine how basic computerized techniques might contribute to the analysis of opinions issued by the Supreme Court of the United States.;

As human beings, our minds have the capability to understand and comprehend all types of documents, media, questions, and problems,. This ability to form opinions and determine relevant information is used by social scientists to systematically study interpersonal communications with techniques known as content analysis. For years, content analysis has been used in case studies to find trends or patterns and to draw conclusions about various media, messages, and communications. More recently, computers have become more involved to aid in automating portions of content analysis experiments. With current technology, there have been great strides (in many different fields) that could help analyze documents in numbers that were previously too tedious and inefficient for humans to interpret.

To test the feasibility of adapting this technology to content analysis tasks, we chose a data set of such an insurmountable size: opinions of the Supreme Court of the United States. The Supreme Court has amassed a tremendously large data set in which content analysis (or, at least, the computerized kind) has yet to provide insight. We thus set out to determine what potential computerized content analysis techniques might have in legal scholarship.

Some fields have already used computerized content analysis techniques, yet most of the documents and media analyzed tend to have rigid structures making it easier to compare and contrast to each other. Opinions of the Supreme Court are all written with a similar fashion, but each individual opinion is different, with different facts, concepts, context and writing style. The largest complication was to determine if computers could be used to help find any trends or patterns among such a disparate data set. While the question seemed fair, the task at hand had multiple obstacles to cross.

In order to tackle this problem, a tool was developed to analyze the data sets. TEAL (Tool for Empirical Analysis of Law) was developed for use in this project to read in each individual opinion of the data set and derive information dependant on the experiment being tested. The tool was developed to be modular to allow easy modifications throughout the project for each individual experiment. Aside from the use of the TEAL application, the setup of a data set was just as important. With the use of the Cornell Law School's Supreme Court Collection, approximately 5,000 opinions were sorted through. From those opinions, experiments sampled between 5,000 and 2,000 opinions to draw conclusions from. With the data sets established and TEAL ready for testing, a series of experiments were created to test the capabilities of content analysis without human intervention. Each experiment created interesting results that were then themselves analyzed to determine the successfulness of the content analysis.

A total of five experiments were used to show the capabilities of computer content analysis. Different methods of content analysis were applied through letter-pairing, "dictionary" lists (a content list of relevant words), and categorization. Each experiment was designed to work with TEAL and carefully probe opinions of the Supreme Court, and to build on the results of the previous experiment. The first experiment attempted to explain the correlation of minority citations over time. We revealed that citations of the minority actually grew over time. While the results consisted of many small variations, there was a steady increase over the century we observed. We were unable to correlate the variations with the sitting Chief Justice, but speculated that part of the explanation was due to the change in entire Court's makeup. In the second experiment, two categories of data sets were used; one containing "normal" opinions and the other containing more important opinions. The important opinions were selected using a culmination of a few resources. The objective of this experiment was to determine if the minority was cited more often either in important opinions or "normal" opinions, and if this metric could then be used to categorize other opinions. Continuing along with this experiment, the third one used the same sample set but TEAL was used to measure other metrics. The application was used to determine if one sample set contained more case citations than the other. The length of each opinion was mapped directly to the number of case citations as well to create a frequency metric based on citations per word count.

Categorization was used again but involved more complex content searches in Experiment 4. Rather than determining the sample set prior to the experiment, TEAL was used to configure nearly 1,800 opinions into one of five categories. From there, each opinion was analyzed for the number of case citations and word count to see if any particular category featured more than the others. The experiments exemplified the variety that each and every opinion had as well as the usefulness of TEAL. To conclude all of the experiments, the Federalist Papers were used as a simple example to show how authorship could be distinguished using computer content analysis. While the authorship of the Federalist Papers had already been determined, the applications capabilities to produce the same results was tested. With the use of a letter-pairing algorithm, the Federalist Papers experiment sorted through the twelve articles with disputed authorship to determine proper ownership of the articles.

While not all results produced insights into the Supreme Court, the experiments provided interesting findings suggesting further research could provide to be useful. Content analysis through the use of computers seemed to prove useful in aspects under categorization, trends and patterns, and even simple metrics. Searching for simple repetition of words was fast and efficient and depending on what was being searched could actually prove to be useful in case studies. The categorization of opinions worked rather well when limited to content. When an opinion is classified under a concept or context differences, it became more difficult to identify the different categories. With further research and time set for understanding the similarities between Supreme Court opinions, better “dictionaries” could have been created and would have amounted to more conclusive results. As far as studying the content of opinions, TEAL was capable of working efficiently. Once more complicated experiments involved different contexts and concepts, results revealed that further research was needed to obtain a complete benefit from computer content analysis.

1 Introduction

At the heart of almost all social science research lays one concept: analyzing media in an objective manor. For centuries, scholars have sought to quantify human action, communications, and interactions with the least bias possible. This task has proved problematic, as analyzing communications and interactions, even when expressed in a concrete medium, requires interpretation.

It is difficult to produce and analyze data through interpretation without opening a door for error to be introduced. When experts make interpretative decisions, they frame their interpretations with their own unconscious biases and prejudices. If a group or committee is charged to interpret a source, the members often can reduce the effect of their individual biases, but are still subject to their collective world-view. The disciple of “content analysis” has sought to find the best of such methodologies in order to interpret a source document as data, while introducing minimum bias.

In this project, we investigate a new technique in content analysis: the use of computers. We implement and experiment with several analysis algorithms, some based on previous work and some of our own devising. In order to test the limits of these computer techniques, we use source documents that require highly sophisticated interpretation: opinions of the United State Supreme Court.

From our experiments, we determine the effectiveness of different computer content-analysis methods, casting an eye towards the scholarly goal of objectivity, but also observing differences in time efficiency, required effort, overall accuracy, and reproducibility. From these results, we draw conclusions concerning the usefulness, practicality, and potential of computers in content analysis as a whole.

2 Background

2.1 Content Analysis

2.1.1 History

Webster's Dictionary of the English Language defines "content analysis" as:

> *analysis of the manifest and latent content of a body of communicated material (as a book or film) through a classification, tabulation, and evaluation of its key symbols and themes in order to ascertain its meaning and probable effect.*

It has listed this term since 1961, but the idea content analysis goes back much further. Records of quantitative analysis go back all the way to the eighteenth-century. While scholars believe the idea of analysis had been perceived earlier, the first well documented case involved a collection of hymns originating from Sweden (Krippendorff, 5). The *Songs of Zion* were analyzed by the Swedish state church to see if the songs contained any kind of "dangerous" ideas. By the turn of the century, publications analyzing the structure of content arose in Germany. In the United States, quantitative analysis blossomed with the increase of newsprint. Inquiries about the informative value of newspapers (and other similar journalistic analyses) began to spark the start of what the world now sees as content analysis. Klaus Krippendorff's book Content Analysis: An Introduction to Its Methodology describes three factors that led to the growth of content analysis.

> *First, the new and more powerful electronic media of communication could no longer be treated as an extension of the newspapers. Second, the period following the economic crisis brought numerous social and political problems to which the new mass media were thought to be causal. Third, the emergence of empirical methods of inquiry in the social sciences.* (Krippendorff, 15)

Krippendorff explains how common people became well aware of modern political and social issues. This is in part due to the spread of communication through the media. Once average citizens were capable of acquiring many types of information, it was only a matter of time until comparisons and analysis became mainstream.

As the years progressed with war and other significant worldwide events, content analysis grew at a faster rate. Different media provided different methods of analysis and the term "content analysis" changed to encompass the new methods and media. In the late 1950's, interest in mechanical translation and information systems grew dramatically. The introduction of the computer allowed "large volumes of written documents to be analyzed" in a short amount of time (Krippendorff, 19). The computer was held as a natural ally in content analysis.

2.1.2 Approaches

The world of content analysis consists of varying approaches. People study documents to obtain certain information about a particular event or object or data set. The question for any aspiring analyst is how to begin his research and what steps should be taken. The topic is vast and can cover any sort of field ranging from medicine to pop culture. The first crucial step is to understand what the objective is; there must be some sort of goal that the research should accomplish. A practical approach is to find a subject matter that is covered by a large number of documents that could potentially be used as data for the analysis. Topics with a small number of relevant documents are unsuited for most content analysis approaches.

Primary sources are not necessarily required; data can be generated through a plethora of methods. For instance, data could be created through a survey of opinions, agreement of a focus group or speeches of individuals. Krippendorff outlines various types of sampling schemes that are dependent on the results one wishes to acquire. Random sampling assumes no prior knowledge about the topic and is derived from numerous amounts of documents, speeches, newspapers, and even individuals. The units that will actually be included in the sampling are chosen through some randomized structure (be it dice, roulette wheel, or random number table) to try to prevent any sort of bias based on grouping or other patterns. Systematic sampling is the process of selecting every k^{th} unit from a data list for sampling. An issue with the systematic approach is that since k is a constant it will create biased samples if there are similar occurrences in the data every k^{th} interval (i.e. seasonal changes). Krippendorff explains how "systematic sampling is favored when data stem from regularly appearing publications" and continues to explain about string-like order and its relevance (Krippendorff, 67). Other types of sampling, such as Stratified and Cluster, focus on distinct subcategories and branches and break the samples into groups of elements. The results gained by analyzing the data are tied closely to the types of samples chosen to work with. For further discussion of sampling, see Krippendorff.

With large amounts of data processed, the results need to be organized in some fashion. The metrics that should be calculated have a direct correlation with the objective of the research. Some basic metrics would be as simple as the number of words per document. This would be known as a simple Mathematical Metric. Krippendorff illustrates other metrics such as ordinal, interval, and ratio metrics. Ordinal metrics are used for recording comparisons between units. Useful comparisons include: greater and less than, cause and effects, conditions of, contained within, and refinements of. If the objective of a search is to see if a phrase is a positive statement, that type of metric would be known as an ordinal metric. Interval and ratio metrics represent quantitative differences between the sample units. Interval relates to expressible differences of distances, similarities, or associations such as time elapsed, distance traveled, movements of attitudes and so on. Ratio metrics consists of relative relations compared to an absolute sample. The proportion of length in comparison to the relative sample would be a valid ratio metric. For the most part, "Ordinal scales (chains with an ordinal metric) are probably the most common in the social sciences" but interval metrics also "provide the traditional backbone of empirical research in social sciences" (Krippendorff, 96-97). While these are not all of the metrics in existence they are some of the more favorable within social science research.

2.1.3 Subjects

Content analysis has an immense range of topics and subjects. The range includes different purposes and reasons for content analysis as well as different subject materials and media. The actual types of data and sampling that can be used consist of many different formats from newsprint to propaganda (Riffe).

In the first well documented case of content analysis, church hymns were studied for patterns of "evil." The analytical approach spread quickly once newspapers and other various mediums of information were available. Scholars questioned the validity of the news that was broadcast throughout cities in a particular case. But newspaper and other formats that present the news to the public were not the only possible sources for sampling. Wars of the 20th century were promoted through propaganda. Propaganda presents its own ideas since the purpose is to persuade people for one reason or another. Countries would study and examine propaganda of their enemies to look for weaknesses or clues to help gain an advantage. As time progressed and more formats were available, the sampling for content analysis expanded as well. Sampling was not restricted to hard print and reading material either. With audio recording, speeches have been known to be examined extensively for additional meanings. Paintings and even music were interpreted for comparison reasons. Studies have linked the connection between pieces of art and the times that they were created to get a better understanding about the artist and what s/he was trying to convey. With so many different formats and media to examine and study, and with the number of topics that surround the world it is no wonder that content analysis has only grown in time.

The different types of media and topic selection were not the only differences between content analysis cases. The reason for the research and the goal of the analysis was just as important. Relating back to the church hymns, scholars thought that some of the hymns contained evil passages and messages. The purpose was to examine those documents and determine whether or not they were acceptable for society. While most of content analysis consists of pattern recognition, its uses extend far beyond identifying documents as being positive or negative as viewed by society. Content analysis has been used and can be used to answer more in-depth questions, including authorship. For example, The Federalists Papers were written in 1787 and 1788 and consisted of 85 essays that were written by Alexander Hamilton, James Madison, and John Jay. All of the essays were signed "Publius" and while it was understood that the papers were written by only three people, no one knew for sure who wrote which paper. With the use of pattern recognition of content analysis, the Federalist Papers were assigned authorship. Content analysis can determine useful statistics and find similarities between media that can prove useful to help sort out similar problems. The variety ranging from topic selection, media, and reasons/purposes for content analysis is vast.

2.2 Computers and Content Analysis

2.2.1 Disadvantages

When computers are applied to problems of content analysis, several hurdles immediately present themselves. It is difficult to create a computer program capable of parsing an English sentence, much less understanding its meaning (Stone, 7).The scaled judgments humans are often called upon to make in traditional analysis ("On a scale from 1 to 10, how would you rate…") cannot be easily made by software. While techniques exist to 'teach' computers to comparatively rate inputs, the judgments are crude and not nearly as abstract as those a human reviewer can make. Programs are quite capable of making quantitative judgments, but the capability to make qualitative decisions remains an open problem in computer science. To illustrate, consider a World-War II propaganda slogan: "Loose Lips Sink Ships." A human can immediately consider this phrase's meaning and implications; that talking about things you shouldn't, that talking to the wrong people, that not keeping secrets hidden could lead to the loss of military assets. By contrast, a good computerized algorithm may be able to determine that lips, particularly those that are loose, may cause misfortune. The problem of a computer incorporating context in order to decipher 'deeper' layers of understanding is one without a solution at present.

2.2.2 Advantages

On the other hand, the application of computers does present several benefits. Any scientific inquiry should be systematic and objective, and computers practically guarantee both. In an example human content analysis, researchers may not identify each occurrence of a characteristic. Errors are easily made, and humans naturally vary in their opinions. By contrast, any number of computers running the same deterministic program will make the same judgments about the same documents. They are only biased insofar as humans make their criteria, and will not change those criteria based on external context. Experiments, therefore, are readily reproducible.

Because of advances in digital storage and processing power, modern computers allow collection of data to be much faster than with conventional methods. Even the largest documents can be analyzed in a very short amount of time, and large collections of documents can be automatically processed in series with minimal human effort (and minimal chance for unintentional interference) (West, 15).

2.3 Techniques and Applications

Applications have always sought to capitalize on computers' strengths while minimizing reliance on their weaknesses. Some methods do so by looking for a specific linguistic feature; a particular word, phrase, fragment, or combination thereof. Others seek to parse English sentences into machine-readable data.

2.3.1 Simple Metrics

The most obvious and most prevalent method of applying computers to content analysis is by exploiting simple metrics. Using simple search, the frequency of a linguistic feature can be easily determined. Counting is a task computers are especially suited to, and a simple count is sufficient for drawing some limited conclusions. For example, given a selection of topical newspaper articles, the usage of a specific word to describe the topic can be tracked over time. In political propaganda, the use of specific phrases or slogans can be counted to determine their importance relative to each other. Computerized counting is also very useful in authorship experiments. For instance, by counting the number of specific word-fragments and comparing the results to those obtained from known works, the authorship of a disputed work can be determined with reasonable accuracy (see Experiment 5).

2.3.2 Simple Relations

Another technique in the application of computers is that of simple relations. Similar to simple metrics, it uses a search to find occurrences of a specific linguistic feature. However, unlike simple metrics, the search then continues; another language feature is sought within a given distance. The effect is to find relations between features. For example, given a set of newspaper articles, the number of times an idea is used to justify a policy decision can be measured over time.

2.3.3 Tagging

Expanding on simple relations are methods of tagging. Instead of looking at linguistic features, each word in a document is parsed and replaced by computer-readable data given by a dictionary. By then applying one of the simpler approaches, more complex data can be gathered, like the number of words expressing disapproval, or the number of words expressing disapproval in proximity to a specific idea. To illustrate, a program can be given a list of words expressing disapproval (which can be data-mined from a common dictionary) and can learn to recognize a specific simple idea (most easily by combining a dictionary with simple relations). It can then determine how often these two tags are in proximity with each other.

This method has had the most success in previous forays into computerized content analysis. Most notably, the General Inquirer system, a computer system created by Phillip J. Stone, implements this technique, and has been applied to subject ranging from journalistic bias to clinical psychology (Stone).

2.3.4 Neural Networks and Other Methods

Parsing and understanding natural language remains an open problem in computer science, but a problem that has some leads. The application of a trainable neural network has potential to greatly increase the accuracy of authorship experiments as well as to facilitate the recognition of ideas expressed by a given text. “Expert” systems with modern machine learning techniques have potential to destruct natural language into parts more easily understandable by computer systems. Details of these experimental systems remain beyond the scope of this project.

2.4 *Empirical Legal Studies*

The union of mathematics and the law is fertile ground for academic study. However, the size of this union is somewhat limited. The law is grounded in nuance, and often resists conversion to purely empirical measures. There are, however, several topics in law that naturally are expressed in numbers.

2.4.1 Math and Law

2.4.1.1 Sentencing and Damages

Perhaps the most obvious and most useful empirics in law are in the topics of criminal sentencing and damages. Since both are expressed in numbers already, no effort is needed to abstract data into numerical form. Studies are easily done by statistically analyzing these existing numbers to show trends, variations, and find anomalies. By comparing punitive and compensatory damage amounts in successful sexual harassment suits, for instance, one might try to determine the effect of different variables (for example, the severity of the alleged harassment) on the case's outcome (Sharkey). The statistical methods applied typically do not vary radically from those used in any other study with existing statistical data.

2.4.1.2 Policy Decisions

Policy decisions of legislatures often necessitate the modification of existing statutes in order to accomplish a goal. When laws are amended, it may affect more than simply the intended problem. Empirical legal analysis can be use to gauge the effect of a decision on complaints brought under the newly modified law. Frequency and outcome of cases brought under the statute can be easily determined and compared to results previous to a change (Miller). The effects of changes in tax policy, real estate law, and workers compensation have all been analyzed in this manner. Similar statistical methods are often applied to policy changes in criminal laws. Studies using mathematical methodologies seek to determine if, for instance, harsher punishments reduce crime rates.

2.4.2 Computers Analysis of Law

Analyses that are mathematical in nature are easily translated into computer code. Most of modern statistical analysis is computer-aided. However, due to the nuances of law, very few results are calculated without significant human intervention. In a typical study, cases, opinions, or briefs are reduced to their component parts and somehow quantified before applying a computer to the analysis. For instance, the number of writs of certiorari issued in a year can be counted by a human before being inputted as a data point. For more complicated analyses, humans are integral to the process. To the best of this author's research, no program has been written to recognize ideas in a text, to autonomously generate data on a case's outcome, or to otherwise draw empirical meaning from written human language in a legal text.

3 Methodology

3.1 Development of TEAL

In order to properly test the successfulness of computers in content analysis of Supreme Court cases, we developed a tool capable of running several different content analysis experiments (using simple metrics, simple relations, and some very basic tagging).The idea behind TEAL (Tool for Empirical Analysis of Law) was to create a modular program that could be refined for the specifics of each experiment and return results that could be comparable to some standard. With the use of techniques such as regular expressions, “dictionaries”, and other various pattern recognition methods, TEAL was created specifically for content analysis of law.

Multiple methods were used to search through Supreme Court cases and various texts. A pattern recognition method known as regular expressions was used to establish basic searches for matching sets of strings. A regular expression has a specific syntax which changes depending on the programming language. This syntax is used to compile a string to describe or match a set of strings found in a document or any form of computer text. In essence, a regular expression is a pattern recognition structure that the user can control. The benefits range from being able to search large documents for complex strings to incredible speed and processing power as well as great accuracy. Regular expressions are great for finding preset words or sentences. When it comes to opinions and other means of content analysis, regular expressions are not sufficiently powerful by themselves. More complex algorithms and problem solving methods must be introduced into the scheme of regular expressions to reap the benefit of computer content analysis.

While regular expressions can search for just about any type of string, “dictionary” methods can be used to search for concepts that can be expressed more than a single way. When reading a document, there are often sentences that create emotion or are not explicit in their meaning. It is hard for a computer to just scan a document and interpret the meaning of all the text. Certain aspects need to be analyzed to understand the full meaning of the document. The human mind has the ability to infer meaning based on concepts and context, but the processing power of a computer is based on logic, lack of emotion, and is with limited context. The importance of a dictionary is to study key words that may be found in sentences that present a certain mood. Some words in the English language provide information that will move the reader in one emotional direction or another. A method of content analysis includes the research of specific words or phrases that create positive or negative connotations. With the use of regular expressions and the research of specific dictionary terms that offer feelings, a document can be scanned and reported as a positive or negative paper. It is a complicated effort to understand the meanings of every individual word, but the idea combined with computers and regular expressions allows opinions and feelings in documents to be more accurately understood.

For the development of our tool, the Java language was used for the convenience of portability. The idea of developing a tool that various users could access anywhere to aid in research required an application that could be used on any platform. The structure of the code follows a modular approach that allows expansion and future updates to happen with ease. The application has been split into several areas including user interface, file/directory navigation, and of course a content analyzer. The class structure allows these separate areas to be easily modified over time. More classes and structure can be implemented in the future with the way the tool has been developed.

3.2 Apply TEAL

With the development of TEAL, the next step was to apply the application to our collection of sources. In order to test the successfulness of computer content analysis in empirical law and legal analysis, multiple experiments were designed to test the integrity of TEAL. The experiments tested a number of different ideas dealing with Supreme Court cases and dealt with different size samples. There were a total of three major concepts to test with each one having many subparts that would require different algorithms and techniques to acquire useful results.

The hypotheses covered three main content analysis paradigms; numerical analysis, categorization, and authorship. In the first hypothesis, citations of the minority in the majority opinion were enumerated and analyzed. One test to understand this hypothesis was to test whether or not certain Chief Justices would seek to form a consensus on the decision to help minimize the adversity between justices. Following from our results in this experiment, we sought to use the same metric to categorize important decisions. To further test the content analysis capabilities of computers, ideological or political splits to promote the minority opinion were also examined. Rule-based and dictionary based methods were used to attempt to categorize opinions by result. Lastly, we returned to statistical-based metrics and relations to attempt to replicate authorship analysis of the Federalist Papers.

Due to the complexity of some of the hypotheses, TEAL would be able to determine the ability of content analysis in the empirical law and legal analysis field.

3.3 Examine Results

Once the various hypotheses were applied to TEAL, the results of the many experiments were analyzed. The data collected from all of the samples provided a useful way to determine the efficiency of TEAL, and by extension, the capabilities of computer content analysis as applied to Supreme Court opinions.

Each experiment required a unique approach that TEAL had to be adapted to. Algorithms ranging from the “dictionary” idea to letter-pairing present valuable and indicative methods for allowing a computer to interpret documents. The results from these experiments represent the use of these various methods. While not all of the experiments yield a successful result, an interesting trend, or an important insight into the Supreme Court, each one helps to determine the capabilities of TEAL. The results consist of statistics based on the information found from scanning and analyzing numerous samples. These statistics, and their precision, vary depending on the method used to carry out the experiment but the results are consistent in each sample set, and any degradations in accuracy are noted, and are useful in assessing computerized techniques as a whole.

4 Experiments

With TEAL developed to a level sufficient to be applied to Supreme Court cases, we sought to complete 5 experiments in order to draw conclusions on computerized content analysis as a whole.

4.1 Experiment 1: Correlating Citations of the Minority with Chief Justice

4.1.1 Introduction

Eras of the Supreme Court of the United States are often divided up according to the Chief Justice who presided over them. (The Warren Court under Chief Justice Earl Warren or the Marshall Court under Chief Justice John Marshall, for instance). But the Court consists of nine justices; eight associate justices and one Chief. The Chief Justice has no more voting power than any other Justice, but is considered the most senior member of the Court, chairing and speaking first in conferences, and assigning the writing of opinions when his vote is in the majority.

These privileges may seem small and ultimately useless in shaping the Supreme Court's decisions, but we hypothesize the opposite is true. First, by chairing meetings, a clever Chief Justice has a chance to frame discussion; to build support for his ideas, to make compromises. Second, A Chief Justice who knows the ideology and writing style of the associates can greatly influence the reasoning behind a decision. He can broaden an opinion by a assigning it to a Justice (or himself) who will write it broadly, or minimize the effect of an opinion by assigning it to a Justice who favors a more limited scope.
In this manner, and through discussion with the associates, a Chief Justice, we hypothesize, is able to influence the opinions produced by the entire court. In this experiment, we try to determine how far this influence extends; do changes in the Chief Justice noticeably affect the discourse presented in opinions? Do some Chief Justices seek to form consensuses more often? Do some Chief Justices instead produce more adversarial opinions, arguing with the minority and talking about their arguments?

To determine what effect a Chief Justice has on the adversarial tone of majority opinions, we will count the number of citations to the minority, and correlate these numbers with the Chief Justice. We expect some Chiefs to preside over large number of such citations, while others to have significantly fewer.

4.1.2 Methodology

Applying TEAL to this problem is somewhat complicated. While we can use simple metrics to count citations of the minority, we need to generate a “dictionary” with all keywords reasonably used to refer to a minority opinion: phrases like “the dissent” or even the names of justices.

1. Obtain sample documents. In this case, we’ve sampled almost every Supreme Court Opinion held on Cornell Law School’s Supreme Court Collection (http://www.law.cornell.edu/supct/index.html) dated from 1870 to 2007. We’ve excepted around 200 opinions that do not contain relevant metadata, specifically, the date of the opinion. From this, we’ve further excepted opinions of very small size (approximately three sentences or less) in order to eliminate orders and other irrelevant texts. Our data set for this experiment was ultimately comprised of 4000 documents.
2. Generate Dictionary. Key words and phrases that identify a citation of the dissent were compiled into a formatted list, called a dictionary. To do so, five decisions that cited the minority were identified, and phrases used to do so were added to the dictionary. Most phrases are obvious (“the dissent” for instance) but care must be taken so as to not oversimplify; for example, searching for “minority” would wrongly identify strings in Civil Rights cases.
3. Run TEAL with sample documents and search dictionary. After running through all opinions, output should be easily compiled into citations per year.
4. Cross Reference with Chief Justice. By partitioning the data into categories of Chief Justices, we can quickly see what effect, if any, they have on our generated statistics.

4.1.3 Results

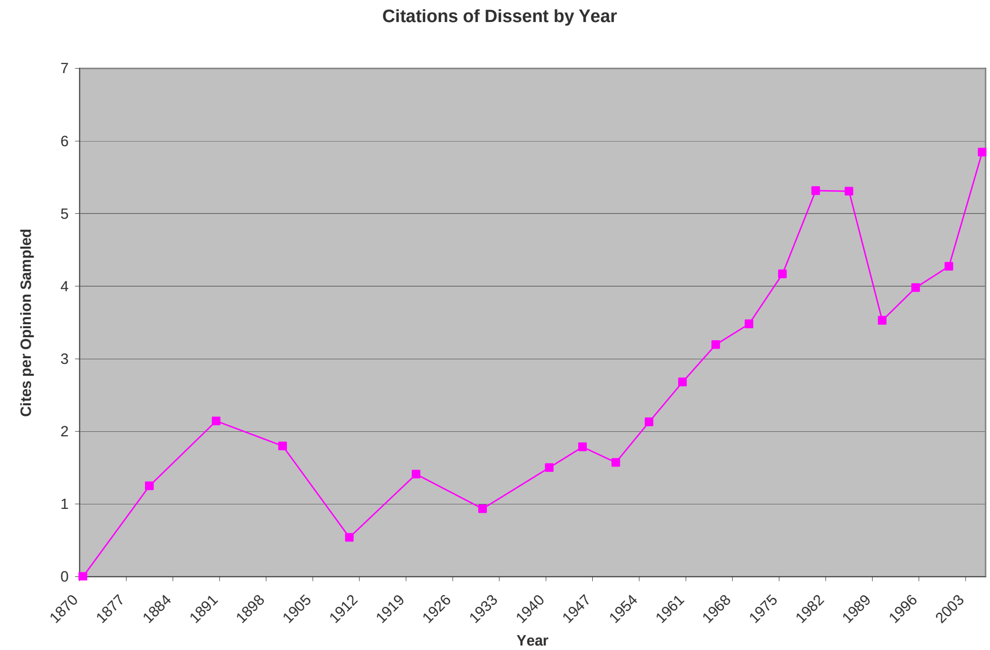
By organizing opinions into decades, TEAL and some math can quickly produce normalized frequency data, correctly for unevenness in sample size. These variations exist for several reasons; the modern Supreme Court produces more opinions per year than they did in 1900, and the 2000 decade is not over at the time of this writing.

Frequency of Citations by Decade (Citations per Opinion Sampled)

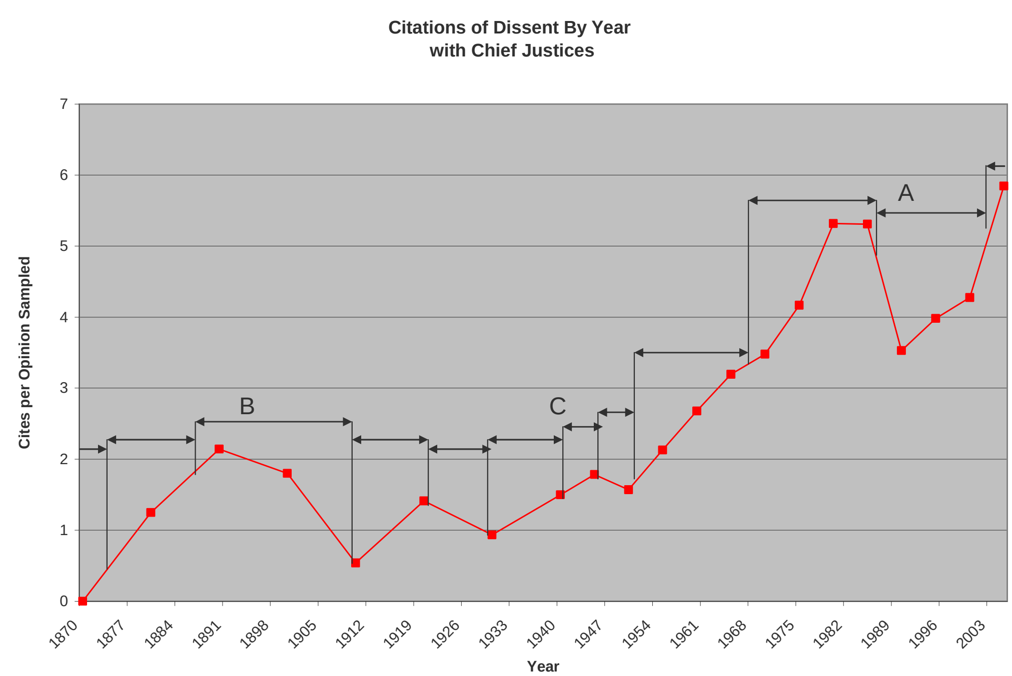
Decade	Frequency
2000	6.65
1990	2.65
1980	5.35
1970	3.86
1960	3.08
1950	1.82
1940	1.68
1930	0.94
1920	1.41
1910	0.54
1900	1.80
1890	2.14
1880	1.25
1870	0.00

One result is immediately obvious. The present decade has far greater citations per opinion than any other. Likewise, the 1980s seem especially fertile for citations. By contrast, in the 1870s, not a single opinion in our sample cited the minority outside of the formal acknowledgment of its existence.

By plotting the frequency by decade, half-decade and year, we found that indeed there is variation between years, and a visible overall trend.



On examination, while there are some major fluctuations (between the 2000s and 1990s), there does seem to be an overall trend towards greater discussion of the minority. The frequency does seem to be somewhat independent of the Chief Justice.



Of most potential interest are changes in Chief Justice (marked by vertical lines) just prior to radical changes in the number of citations. Keeping in mind that Justices can only effect the data points inside their terms (not any previous trend movements), no overarching trend is immediately clear. But inspecting these boundaries, a few observations do stick out.

First (marked ‘A’), the elevation of Chief Justice William H. Rehnquist proceeds a large drop in frequency. Only once Rehnquist has left the court does the frequency return to its previous height.

Second, it can be noticed that no Chief Justice presided over both a large increase and decrease in citations (the term of Chief Justice Fuller, marked ‘B’ for instance, only decreased). Either the number of citations drops when the Justice is appointed, or they increase; never one, then the other. This could be an argument in favor of a correlation, but, on its own, is not persuasive without further evidence.

Third, it is clear that in some time periods, at least, changes in Chief Justice have no effect. Around mark ‘C’, there are several Chief Justices appointed, and no significant variation in cite frequency.

4.1.4 Conclusion

This data set and its resulting metrics unfortunately do not prove nor disprove our hypothesis clearly. Unlike our work with the Federalist Papers (see Experiment 5), the evidence does not support a clearly binary conclusion. On one hand, there are data points that are compelling: Rehnquist sees a huge drop in citations just after his elevation, while a couple other terms seems to precede large increases in our results. In addition, the fact that no Chief presides over both an increase and a decrease would seem to suggest some correlation.

On the other hand, there are several changes in Chief Justice that do not affect the results significantly. Furthermore, correlation is not causation; the dramatic drop in citations upon Rehnquist’s elevation could be due to the court’s makeup as a whole, to the political positions of the President, to the relative divisiveness of politics at the time, or half a dozen other factors. Likewise, the regularity results over a Chief Justice’s term may not be significant at all; it could be due to chance and relatively short terms: Justices serve an average of about ten years; there may not be enough time for two trends in citations to develop.

Overall, while this hypothesis has not been falsified, very little evidence has been generated to prove that the Chief Justice uses his leadership role to shape the court in a manner detectable with this metric. Alternate explanations for the variations present in the result may prove far more compelling. For example, the relative divisiveness of a Presidential administration may correlate better than Chief Justice. Grover Cleveland, controversial with unions and overwhelmed with economic depressions, sees an increase in the 1890s. Lyndon B. Johnson, known for political strong-arming and with controversial positions on things like civil rights, presided over another increase in the 1960s. William Jefferson Clinton, with a very high approval rating and two appointments to the Supreme Court, saw a dramatic drop in citations over the 1990s, while George W. Bush sees a large increase after his first term in office. This explanation has no more evidence than our original hypothesis, but fits the data at least as well.

Given the overall lack of correlation, and given the ability for alternate hypotheses to explain the results at least as well, these results do not provide sufficient evidence to prove our hypothesis, even with a couple dramatic correlations. Further research in the area could attack the problem slightly differently; extending the data set back past 1870, or directly categorizing the decision by the overall makeup of the Supreme Court. We feel it is likely that expanding the number of factors observed (correlating Presidents and Associate Justices) this hypothesis will not be supported further, but alternate explanations would be noted. For instance, it is possible that some associate justices may prefer arguing with the minority view. Looking at the trend in view of the opinion’s author, or in view of the court’s entire makeup, may indicate this factor correlates more strongly.

4.1.5 Performance of TEAL

While our hypothesis was not supported, TEAL proved to be mostly up to the task it was given. We encountered some difficultly managing the data set and marking opinions with the appropriate date, but ultimately were able to do so automatically.

Building a dictionary with a low false-negative rate proved easy, but minimizing the false-positive rate proved difficult; our search strings quickly became large and difficult to manage. More specialized algorithms, such as those used in concept mapping and natural language parsing would be more appropriate and reduce these difficulties somewhat, be remain beyond the scope of this project.

In the end, however, our hypothesis was not supported because of the contents of the data and because of our experiment design decisions, not because of any insufficient in computerized content analysis.

4.2 Experiment 2: Opinions that are "Landmark" or "Politically Important" Promote the Citation of the Minority

4.2.1 Introduction

Opinions written by the Supreme Court sometimes have a significant impact on how future decisions are made. These important opinions often concern ideologically important topics, and often bring public attention to the court. For instance, *Baker V. Carr* would be important, affecting every following case that dealt with apportionment of voting districts. By contrast, *Nix v. Hedden* would not be important, holding simply that a tomato is a vegetable for the purposes of tax law. The question at hand is whether or not the Minority is cited more in these ideological and politically important opinions than in others. We hypothesis that there will be a distinct difference in citation frequency between opinions that are "Landmark" and opinions that are not.

In a debate between two ideological parties, especially when the conflict is close and neither side changes their views, the idea of disproving the minority may be used to help promote the ideas of the majority. In opinions that are seen as important, this become even more likely; to prevent lower courts from considering arguments put forth by the minority, it seems logical that Majority Justices might explain why the minority is mistaken. Landmark opinions, especially those that are based on new concepts (take, for example, cases defining what nonjusticiable political questions are) would be difficult to write without explaining why the previous reasoning (perhaps still held by the minority) is wrong. There are a plethora of reasons one might assume that major opinions may be more adversarial than minor ones, and we expect this will show up in the frequency of citations.

In order to test if this hypothesis is correct or not, the data set was split into two major groups: one with important ground-breaking opinions, and the other with standard opinions. If the minority is cited more frequently in the landmark opinions, the results of the analysis should show this, as the mean frequency in one group should be significantly different than the mean frequency in the other.

4.2.2 Methodology

The “dictionary” list successfully used in Experiment 1 proved to be useful once again in this test. The list was composed of popular minority citation words that were scanned for in each opinion. The more complicated part of this experiment was to distinguish which opinions cause a greater impact in the empirical legal field, and thus, are landmark.

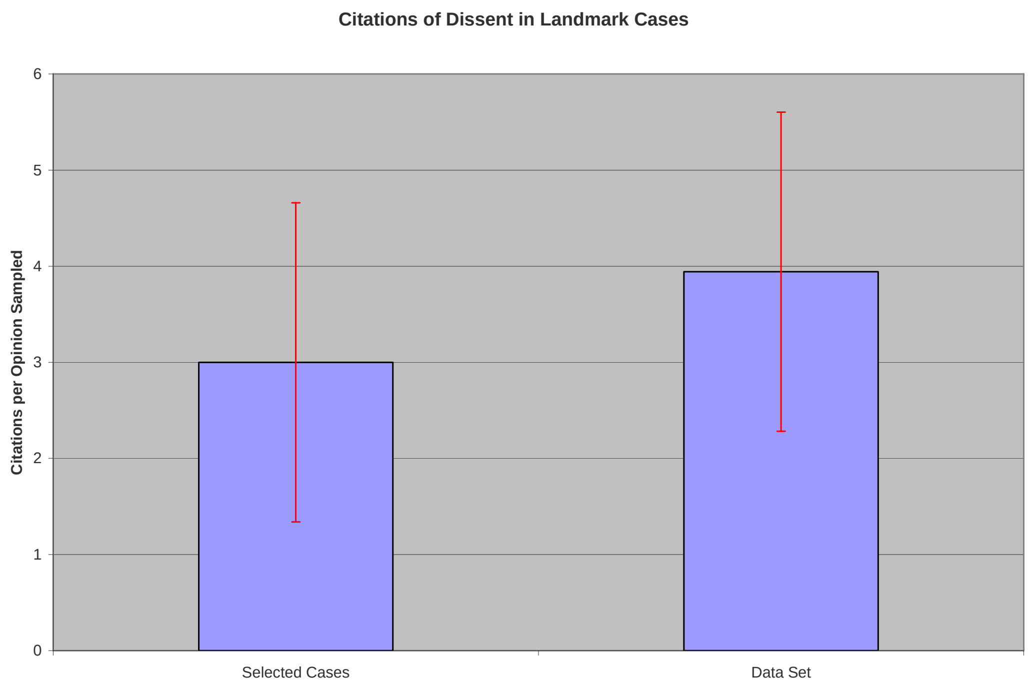
1. The first step is to gather sample documents. All of the opinions collected are sampled from almost every Supreme Court Opinion held on Cornell Law School’s Supreme Court Collection (http://www.law.cornell.edu/supct/index.html) dated from 1870 to 2007. Opinions without meta-data and very short documents were excepted, as in Experiment 1. The complication in this experiment was to isolate the ideological opinions from the rest. In order to do so, two sources, constructed very differently, were consulted. First, the community-built encyclopedia, Wikipedia, was consulted to gather a list of the landmark decisions made by the Supreme Court according to the judgment of a community of both experts and laymen. Secondly, the law-professor-written “Cornell Law School List of Historic Decisions” was obtained. Since compiling such a list is very subjective, any one list has its own biases; to minimize these, only opinions that were found in *both* the Wikipedia list and the Cornell list were selected as “Landmark”. This resulted in a category contain 81 important decisions (see Appendix A).

2. Create the “dictionary” list. This list contains key words and phrases that identify a citation of the minority. To do so, five decisions that cited the minority were identified, and phrases used to do so were added to the dictionary. Most phrases are obvious (“the dissent” for instance) but care must be taken so as to not oversimplify; for example, searching for “minority” would wrongly identify strings in Civil Rights cases. For this step, the “dictionary” from Experiment 1 was used with minimal modification.

3. Run TEAL with sample documents and search dictionary. TEAL is run independently on both the selected sample set of important opinions as well as the sample set of standard opinions.
4. Compare the results of the standard opinions to the ideological sample set results to see if the minority was cited more times in either case.

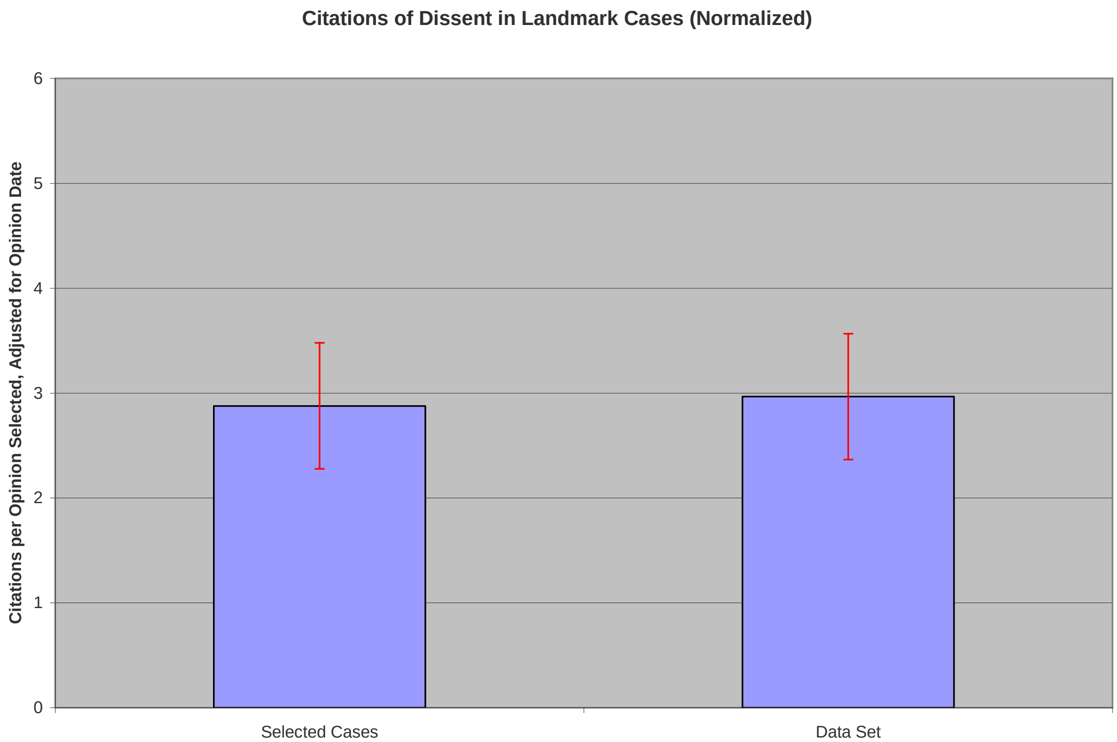
4.2.3 Results

After partitioning "politically important" cases from those not considered landmark, we compare the relative frequency of references to the dissent. By examining our data set and results from Experiment 1, we can establish a standard deviation from mean frequency, so we can compare the results from each category.



Initially, it appears our hypothesis, that our selected cases would have a significantly higher frequency, is wrong. In fact, it appears that important cases may contain less conflict (perhaps because of efforts to put the Court's entire weight behind important decisions). However, this may be misleading. The red lines, marking the standard deviation of the data set, easily include the resulting frequencies for both sets; the differences may simply be random fluctuations.

Furthermore, by partitioning the data set into two categories, the mean decision date has been skewed, and is not the same for the selected cases and the remaining unselected opinions. The data set as a whole centers around 1960, about 20 years later than the center of our sampled time period, but the sampled cases cluster around 1940. Since we've already determined that citations vary by year (see Experiment one), is may be prudent to normalize our results with the expected frequencies for an opinion's date of issue.



When adjusted for discrepancies in mean data-set date, there is practically no difference in frequency between the two categories.

4.2.4 Conclusion

The unadjusted difference in frequency between selected cases and the remaining data set exists, but is well within the entire dataset's standard deviation. When adjusted for variation relative to the year of the decisions, the frequencies are strikingly close; within 0.1 cites per decision. The average standard deviation *for each year* is far larger than this variation, at 0.6 cites per decision.

Thus, using our judgment of what cases are “politically important”, and “landmark”, there is no evidence to suggest any difference in frequency of conflict (or deference) in the Court using this metric. On the contrary, this result suggests that, using this metric, there is no significant difference between opinions that reshape the legal landscape, and opinions that apply to specific sets of facts.

Clearly, there do exist differences in the cases; experts have identified them as more important than others. Compelling future research might attempt to determine what, if any, computerized methods can differentiate between landmark opinions and opinions that are not as important. Based on our research, we suggest letter-pair analysis or artificial neural-network algorithms may be up to the task, but it is possible that there is no measurable difference in the written opinions themselves. The key to landmark decisions may lie in their context, and be undeterminable by a context-free reader like a computer.

4.2.5 Performance of TEAL

In this experiment, no significant differences between landmark cases and “run-of-of-the-mill” cases are detected. In reality, differences are evident to a human reader with any training in law, history, or politics. It is not that TEAL is wrong per se; the conclusion that “politically important” decisions do not cite the minority more or less so than other opinions is probably true. We cannot, however, extend that conclusion to speak towards the actual amount of discourse in the Court. To assert that opinions that change the legal landscape, say, *Roe V. Wade*, are no more divisive to the court than a “regular” opinion defies logic. As such, this experiment constitutes a failure for the system, for our experiment design, and perhaps most importantly, for our choice of metric.

It does, however, serve as a successful data point in finding the limitations of computerized content analysis, and of empirical experiment design.

4.3 *Experiment 3: Opinions that are “Landmark” Tend to Contain More Citations and Writing*

4.3.1 Introduction

Categorization is a task that is very useful in content analysis, and a task the human brain is distinctly suited to. In every-day life, a person’s brain is constantly placing things into categories. Data from all five senses is processed to categorize certain inputs as certain objects. Even in the dark or from a distance, humans can quickly categorize a seen object as a car or not a car. A heard noise can quickly be identified Not only is this done quickly, but it can be done robustly; even with a minimum of data (a poor view, a drawing, or photograph) an object can be identified with near perfect accuracy.

By contrast, computers seem relatively poor at categorization. Teaching a computer to differentiate between seen objects is difficult. While computers are technically faster than a human brain, they seem unable to accomplish categorization of complex information at reasonable speed with reasonable accuracy. The problem seems to be that few complex data sets can be categorized based on a single observation, or sets of observations of a single metric (Hawkins).

In Experiment 2, we sought to differentiate between two categories of cases based on a very specific metric. We were unable to do so. In this experiment, we sought to revisit this categorization problem with two different measures; the number of case citations and the number of words in the majority opinion. We hypothesize that “landmark” decisions cite more case law and are longer. This hypothesis seems logical; opinions likely to guide future decisions would naturally warrant more attention from their writers, and more attention would likely increase the amount of legal research (measurable by case citations) and writing (measurable by length) that goes in to the opinion.

Barring application of complex techniques such as neural networks or introducing context from other opinions, we believe this is our best chance to differentiate between categories. If successful, given any opinion, we should be able to predict which category is should exist in with a high degree of accuracy.

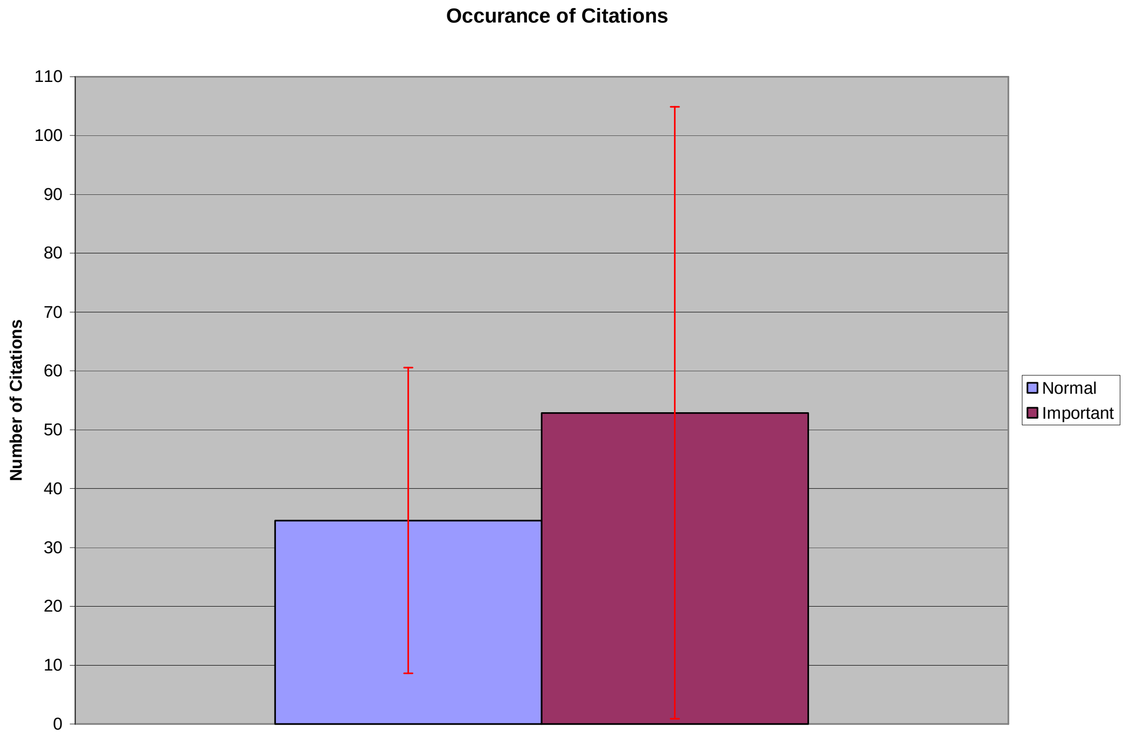
4.3.2 Methodology

The list of important opinions used in Experiment 2 was recycled for this experiment for the sake of consistency. The Dictionary, containing search strings, was not reused, as the metric in this test is found differently. While a citation of the dissent can take many forms, a legal citation takes only a few.

1. The first step is to gather and categorize sample documents. All of the opinions collected are sampled from almost every Supreme Court opinion held on Cornell Law School's Supreme Court Collection dated from 1870 to 2007. Opinions without meta-data and very short documents were excepted, as in Experiment 1. As in Experiment 2, a complication in this experiment was to isolate the ideological opinions from the rest. These categories were taken directly from this previous experiment, for expediency and to minimize variation in results between the two tests due to inconsistent data sets.
2. Create the "dictionary" list for case citations. This list contains all key words and phrases that identify a legal citation. Unlike previous experiments, false positive and negatives are nearly impossible in this test; there is a finite (and more importantly, small) number of ways a legal citation can be made. Looking for "v.", "in re" and "ex parte" flags effectively every citation. From there, eliminating false positives (such as the opinion's own name) is also straightforward.
3. Run TEAL with sample documents and search dictionary. TEAL is run independently on both the selected sample set of important opinions as well as the sample set of standard opinions.
4. TEAL is also rerun in a mode that produces a word count, to determine the length of the opinion.

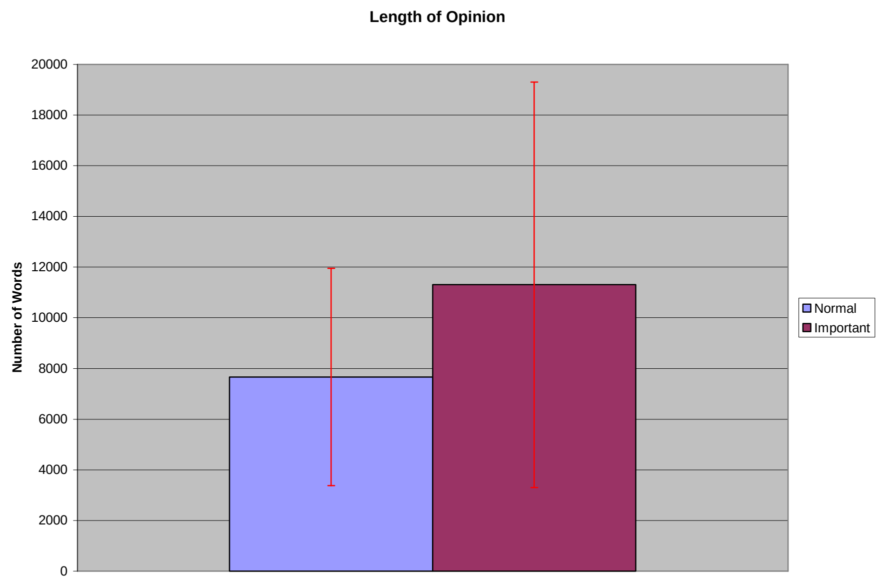
4.3.3 Results

With the data set separated into important and regular opinions, we used TEAL to scan through all of the opinions. First, the number of case citations was determined. We found significant variation inside each category, and some variation between the two.



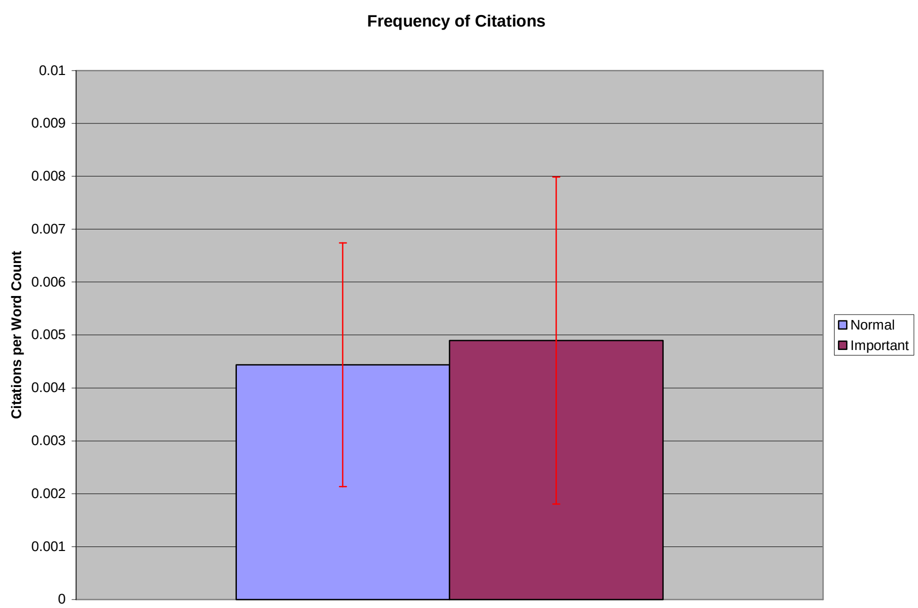
As is visible, there is a higher mean occurrence of case citations in the more important opinions. However, the standard deviation indicates huge variation within each set, perhaps invalidating this conclusion. The important opinions ranged from having two citations all the way up to over one hundred, with more listed in footnotes. The standard deviation for the standard opinions indicates a greater cluster around the mean, but the variation is still almost as significant.

From these results, the next process was to look at the total length of the opinions overall and see if the more important cases contained a higher count (as we expect).



The Length of Opinion chart shows a similar result to the previous chart. The standard deviation for the important opinions has a large spread (although it is less than the previous metric). There still seems to be a twenty-five percent increase in the average length of the opinions for the more important ones, but there still is truly massive variation within each category.

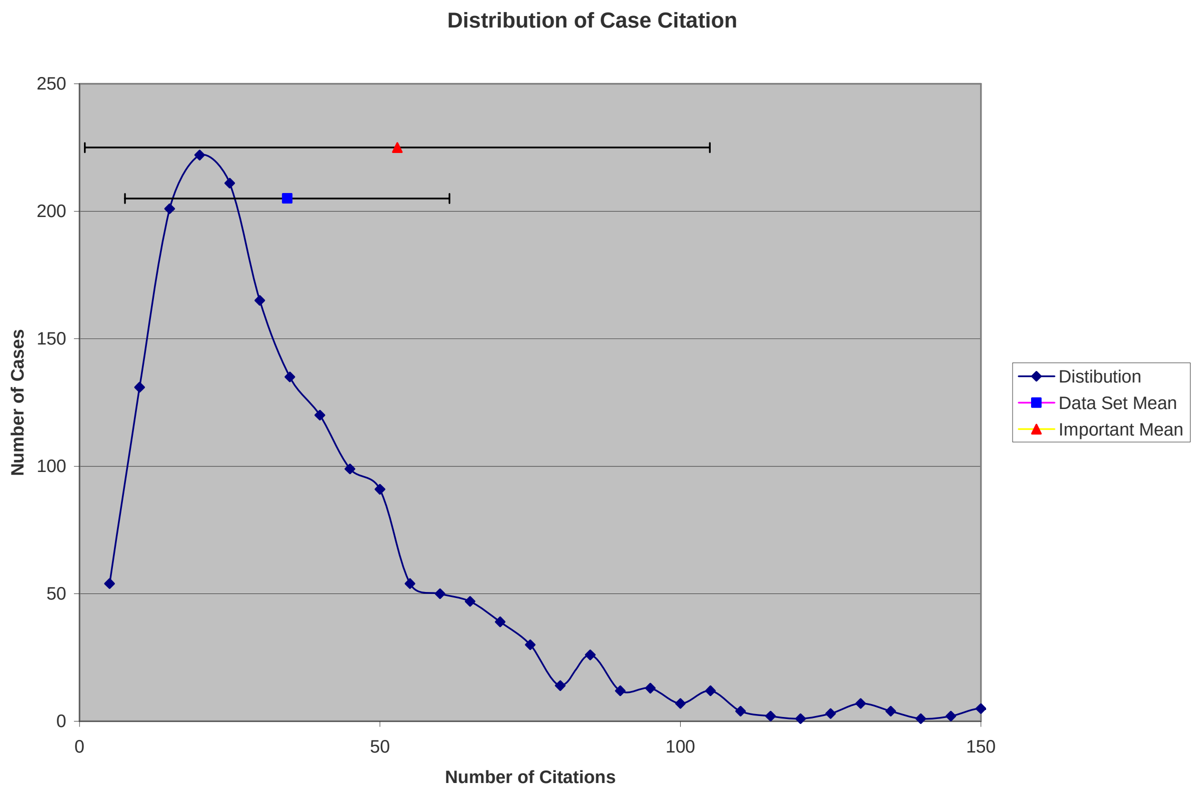
With the word count assessed and the occurrence of citations accounted for, the next best thing was to calculate the frequency of the opinions. This was done simply by taking the number of citations per word and averaging the data set together.



This new chart provides a further point of interest in that the difference is miniscule compared to the other charts. The standard deviations are not as steep as before, yet even next to these smaller variations, the difference between the important opinions and the normal opinions is far less than a single standard deviation.

4.3.4 Conclusion

Ultimately, these results are inconclusive. While there are differences between the sets visible with these metrics, we assert they are mostly superficial. While the mean number of citations and mean length is greater in the important opinions, the standard deviation may be so large as to prevent this difference to be useful. Statistical methods prove helpful in deciding just how real these differences are. For example, let us examine case citations exclusively. In the dataset as a whole, the number of citations does not follow a normal distribution.



While the vast majority of cases still are within a standard deviation of the mean (marked in blue and red), the skew prevents us from applying the “Empirical Rule” to determine the indicativeness of our averages. Instead, we must apply Chebyshev's inequality, where we are only guaranteed that 50% of the values are within 1.4 standard deviations. Looking back at the charts in our results, this means we're only guaranteed to have about *half* of our data exist within the standard deviation bars. In view of this, and the fact that our standard deviations *already* overlap, it is difficult to assert that this metric has a significant correlation with an opinion's category. If an arbitrary opinion was categorized based on its number of citations, the odds of it being categorized correctly are effectively the same as chance.

Likewise, using word count suffers from the same problem. The standard deviations overlap, and the means are not radically different. Furthermore, when calculating the frequency of citations (as seen in Results) the differences become even less significant.

Therefore, while there appears to be superficial difference, we are once again unable to effectively differentiate between important and other decisions using statistical means. If an arbitrary opinion is categorized based on number of citations, opinion length, or any statistical measure derived from them, the odds of it being correctly categorized are indistinguishable from chance.

As in Experiment 2, future research could attack this problem from a different angle. Research has suggested that Neural Networks, a simulation of the highly categorizing biological brain, are promising for this sort of problem. Likewise, more complex analysis, perhaps based on a combination of the statistical methods we have tried separately, may be able to raise the odds of correct categorization over 1/2.

4.3.5 Performance of TEAL

While TEAL functioned perfectly and proved especially suited for an analysis with such binary results, the rules provided proved to be inadequate. By designing our experiment to target length and case citations, we necessarily prevented TEAL from producing any useful result. As in the previous experiment, this test shows that empirical analysis only is useful so far as useful empirical measurements actually exist (and can be identified in the experiment design).

4.4 Experiment 4: Number of Case Citations and Case Results

4.4.1 Introduction

In experiment 3, we sought to use statistical measures to categorize an opinion into one of two expert delimitated groups. In this experiment, we expand on this method, in two parts.

First, we seek to categorize opinions not based on their context (important vs. not important) but on their content. Specifically, we set up rules to automatically, without human intervention, determine the result of a case. This result enables us to determine the effectiveness of using “dictionaries,” (specially selected lists of words and phrases), to group opinions based on hard and fast rules, rather than statistical probabilities. For instance, if the opinion contains the phrase “The judgment of (some court) is Affirmed” but not the equivalent phrase for reversal, it belongs in the affirmed category. Or, if the opinion ends with “It is so ordered,” but no phrases indicating reversal or affirmation, the case belongs in a category with other orders too complex to grasp via simple categorization. We hypothesis that these simple rules will sort opinions with very few or no errors; when writing opinions, the Supreme Court traditionally follows certain simple conventions that can be differentiated quickly and automatically.

Second, we then seek to determine what differences, in terms of the statistical metrics of Experiments 2 and 3, exist between these computer generated categories. This serves to determine what, if any, empirical differences exist between the opinions other than the key phrases TEAL previously used to assign categories. Specifically, we seek to measure the amount of effort put into writing an opinion by quantifying the length and the number of case citations in the majority opinion. We hypothesize, based on our results from previous experiments, that, while categorization will be successful, no real statistical difference will be found between opinions that reverse lower courts and opinions that affirm them. Furthermore, we expect that opinions with indeterminate results, or those too complex to easily categorize (for example, *Bush v. Gore, 531 U.S. 98 (2000)*, reaches several decisions on different topics, and ends with “It is so ordered.”) will not be significantly longer or possess more citations that those with “simple” results.

While one would expect such opinions to contain greater amounts of reasoning in order to instruct lower courts, we expect that on average, this will not be visible statistically.

4.4.2 Methodology

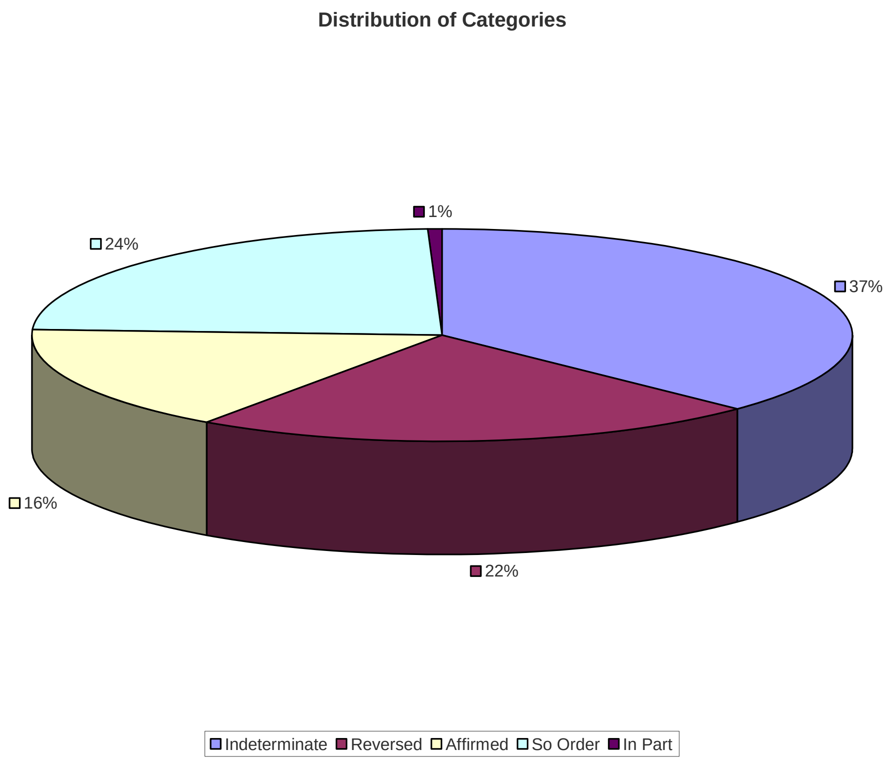
The methodology in this experiment differs significantly from previous tests. Instead of testing one hypothesis, it is in two parts: first categorization and then data collection.

1. The first step is to gather sample documents. As in previous experiments, all of the opinions collected are sampled from almost every Supreme Court opinion held on Cornell Law School's Supreme Court Collection dated from 1870 to 2007. Opinions without meta-data and very short documents were excepted, as in Experiment 1. Because we will later wish to examine the documents individually to determine our error rate, we then randomly sample (by dice roll) from this population. Ultimately, our sample is 2000 documents.
2. Create rules and dictionary for automatic categorization. Dictionaries were made by randomly sampling 40 opinions, reading them, and adding distinct features of sentences discussing the case's outcome to the appropriate list.
3. We then set our rules to group opinions into these categories:
    a. If phrases from only the "Judgment Affirmed" dictionary are found in the document, the result is categorized as "Affirmed."
    b. If phrases from only the "Judgment Reversed" dictionary are found, the result is "Reversed."
    c. If phrases from only the "So Ordered" dictionary are found, the result is not clear, and is put in the category titled "So Ordered".
    d. If phrases from more than one dictionary are found, the case complex, and contains parts that are reversed, parts that are affirmed and possibly further orders. It is categorized as a decision "In Part."
    e. If no phrases from any dictionary are found, the opinion does not contain a clear result. It is categorized as "Indeterminate."

4. Create the “dictionary” list for case citations. This list contains all key words and phrases that identify a legal citation. The dictionary from the previous experiment was reused for this purpose.
5. Run TEAL with sample documents and search dictionary. For efficacy, it is possible to run all three tests, categorization, citation numbers, and word count, all in parallel.

4.4.3 Results

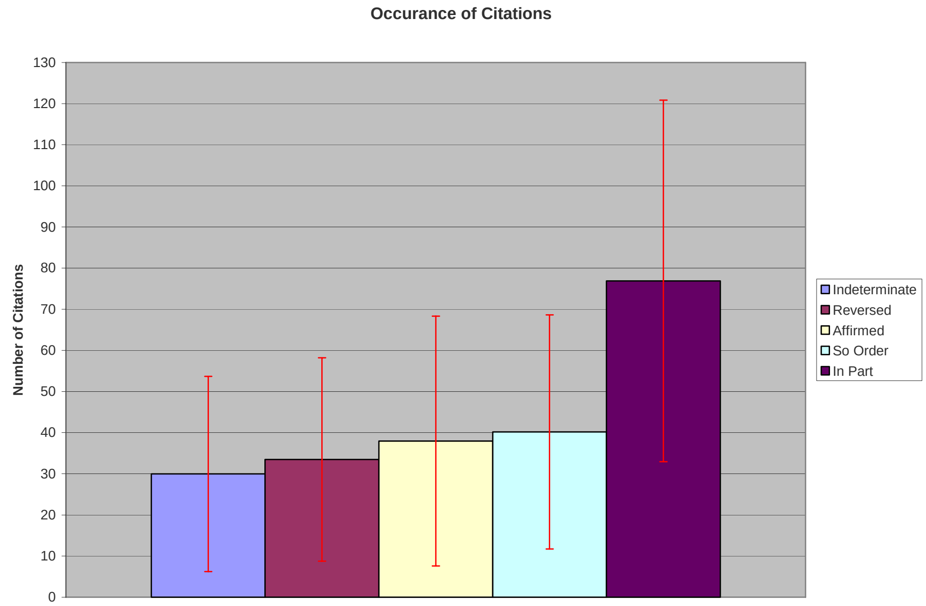
For the first portion of the experiment, the opinions were broken into five separate categories (mentioned above) based on their results.



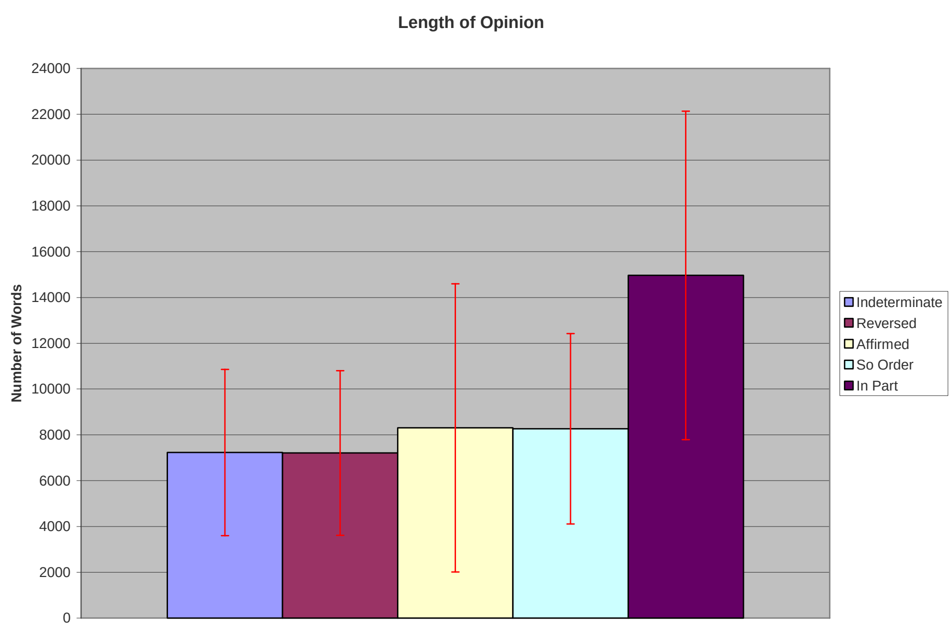
The pie chart shows the percentage distribution of the opinions. The most overwhelming category by far is the indeterminate opinions; opinions which were unable to be fit (at least with our dictionaries) into any of the other categories. While it may be possible for an expert to categorize some of these opinions, creating a computer to do so would require extensive research into fully understanding all of the different phrasings that go into the decision of an opinion. A total of thirty-eight percent of the opinions were determined as being affirmed or reversed. Out of that percentage, nearly forty-two percent of those opinions were affirmed. Nearly a quarter of all the opinions in the sample set did not contain affirmation or reversal language, but did contain instructions to lower courts and orders relating to the disposition of the case. These were placed under the categorization of “So Ordered” (because of their tendency to end with the sentence “It is so ordered). Only a small number of the opinions mentioned explicitly affirmed and reversed parts of the lower court’s decision. Overall, it seems that the split in determined opinions (Affirmed, Reversed, and So Order) are fairly even.

An important point here is the rate of error. How many opinions are misidentified? By reading opinions in each category and by ‘stepping’ through the identification process, we determined that error is very low. Theoretically, zero opinions in the Affirmed, Revered, So Ordered, or In Part categories are wrongly placed, simply by the fact that they must contain language designed to bring about this outcome. While it is possible that an opinion might not have the full effect of its outcome (for example, a decision could be reversed and the case remanded, but with instructions that eventually bring about the same result) the outcome at the Supreme Court level is still technically correct. However, the indeterminate category is not as clear cut. While inspection of the source documents suggests most are cases with complex results, it is possible that some opinions in this category have a simple result, but for whatever reason do not express it clearly. Still, any of these errors seem relatively harmless, as, if they exist, will not affect the relationships between other categories.

The next step was to compare the different categories in terms of citation occurrences and the average length of the opinions. Both the occurrence of citations and the length of opinions charts seem to illustrate a common trend.



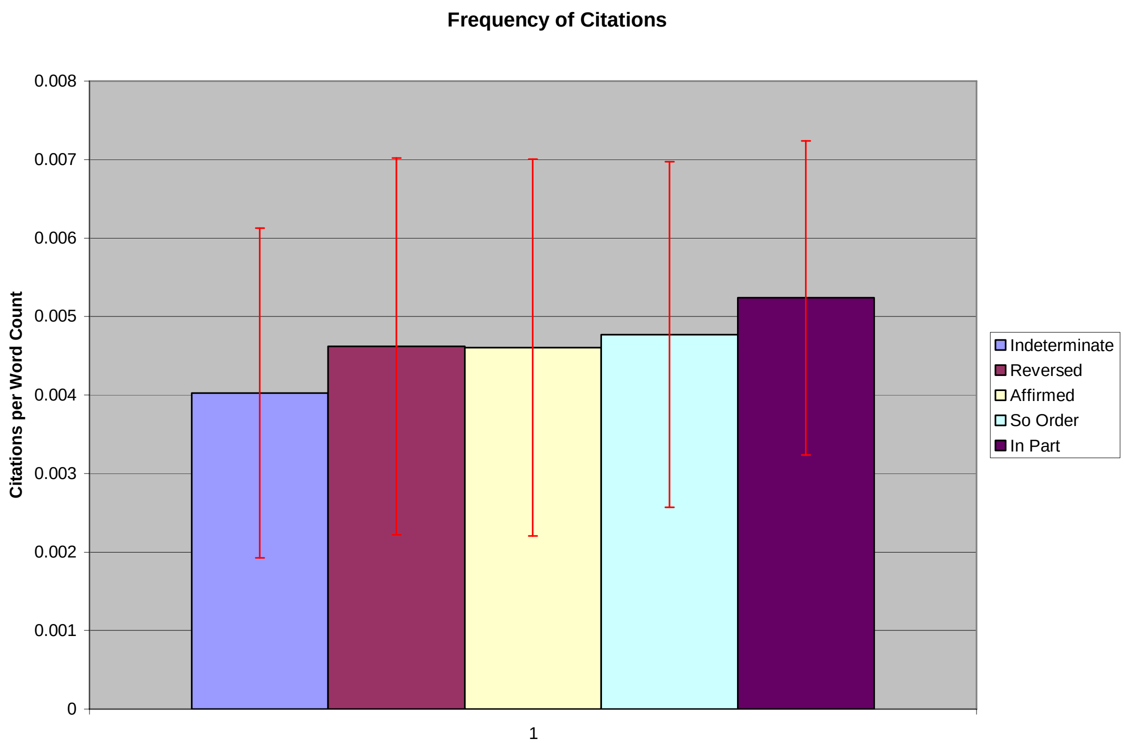
In the occurrence of citations chart, nearly all of the categories have an average of less than forty citations per opinion. The interesting aspect is that of the "In Part" category. This category tracks all of the opinions that feature an explicit affirmation and a reversal of different parts in the decision. The question of why this appears to happen might be explained using the length of opinion chart.



The length of the opinions follows the same trend as the previous chart. The average of most of the categories is around or under 8000 words with the exception of the

"In Part" grouping once again. The interesting fact is that the "In Part" category has an average of nearly 15000 words. This is in line with the previous metric, showing 80 citations per "In Part" opinion against a set-mean of 40. While standard deviations are high, they do not encompass each other's means, and suggest that this variation is indeed statistically significant. This would seem to show that opinions where decisions are in-depth enough to both affirm and reverse, the opinion tends to have much longer opinions and contain many more citations than most opinions.

The final chart shows the frequency of citations per length of opinion throughout all of the categories.



All of the categories have a fairly even frequency ratio hovering around 0.0045 citations per word count. Even thought the "In Part" category featured many more citations than any other category, the mean length of the opinions in that category was equally great. The frequency results suggest that over the entire data set, case citations are fairly uniformly distributed through decisions.

4.4.4 Conclusion

This experiment first underlines a key point in categorization. While simple statistical methods are not sufficient to differentiate between most of these categories (as in the previous two experiments), the application of simple rules to find distinct features of documents works just fine, with low error. However, the distinction must also be made that the authors wrote these features into the documents intentionally; these rules are searching for features that are explicitly added, rather than unintentionally caused. Future research (and, in fact, the next experiment) might look to determine if *complex* statistical methods can determine these features as well, or features that are less intentionally placed.

Secondly, this experiment shows an interesting statistical variation between categories. As we had predicted (and in line with other experiments) there is no significant difference in opinion length and case cautions between Reversed and Affirmed decisions. Likewise, cases with indeterminate results and cases that result in orders are approximately the same. However, our results suggest that the few cases that explicitly affirm in part and reverse in part have almost twice the number of citations and words in the majority opinion. This number is more than a standard deviation away from the mean of the other categories; suggesting that this variation is significant. It may be that reasoning and legal support is needed both for the parts that are affirmed and for the parts that are reversed, causing this increased length.

4.4.5 Performance of TEAL

The application of simple rules to categorize documents worked exceedingly well, and did so quickly and with a minimum of effort. Through this experiment, TEAL was capable of organizing varying opinions using regularly found patterns. This experiment with TEAL proved that more complex organizational skills are possible (and maybe even better). The separation of five categories using multiple regular expressions allowed TEAL to scan each opinion and tag it with a marker to place in one of the categories. It is intuitive that perhaps written documents are simply unsuited to some binary classifications, and that more continuous scale, such as those used in traditional content analysis, are more apt.

4.5 The Federalist Papers

4.5.1 Introduction

Categorization based on a qualitative interpretation of simple statistics has proven difficult. In previous experiments, we were able to categorize based on a features explicitly added to the text, but not based on the importance of concepts contained in them. This experiment serves as a middle case: we use statistical measures to try to determine the authorship of a set of documents; a feature intrinsic in the documents, but not explicitly written.

The Constitution was sent to the states for ratification in late September 1787. Soon after, articles and letters written by those who opposed the proposed federal system emerged in opposition. The Federalist Papers are a series of 85 articles written by Alexander Hamilton, James Madison, and John Jay, and published in October 1787. The original documents of the Federalist Papers were signed by the pseudonym "Publius" with the goal of influencing votes in favor of ratification and to direct public opinion away from the anti-federalists.

Before their deaths, Jay, Hamilton and Madison released lists of which papers they had written, but both Hamilton and Madison claimed to have penned Federalists 49-58 and 62 and 63. Being such an important set of documents in American history, the unknown authorship of these twelve of the papers and the claims of authorship from both Hamilton and Madison, has served as a key test case for algorithms designed to recognize patterns and determine authorship.

Since out of the eighty-five papers, fifty-one have been accepted as Hamilton's work, fourteen as Madison's, and three as Jay's work, historians had concrete evidence of the different writing styles. "The authorship-attribution study that has become the standard of the field is Mosteller and Wallace's examination of the Federalist papers" (West, 53). One method introduced by Mosteller and Wallace was the use of word discrimination. Certain words were found to be used by Madison while similar words presented a different spelling in Hamilton's works. Preliminary work and study "found that Hamilton used 'while,' whereas Madison used 'whilst'" (West, 53). By finding other key words (such as "any," "from," "may," "upon," "can," "his,", "do," "there," "on," and “every.”) the twelve papers were able to be distinguished using this method of manual content analysis. Probability distribution and frequency of word usage were mapped to appropriately distinguish the authorship, and suggest with a high probability that Madison penned all 12 disputed works, as result that agrees with the opinion of most previous scholarship (West, 55). The problem with this method was that it required extensive research, human analysis, and judgment of the papers themselves to decide which words were used by which authors. While the method worked, it took a great deal of time and presented a great deal of work for those wanting to reproduce its results.

An improved method proposed by Bennett encouraged the use of microcomputer computation. This method allowed authorship to be distinguished based on the frequency in which pairs of letters would be scanned and searched for. Bennett illustrates that “There are [26 x 26 =] 676 possible letter pairs, starting with aa and ending with zz” (West, 59), far too many to count by hand. However, with the use of computers and algorithms such as regular expressions, letter pairs can be efficiently counted and matched for comparison. Thus, Tankard applies Bennett’s letter pair methodology against the Federalist Papers to search for the highest-frequency letter pairs, determining that th, he, on, er, in, en, re, ti, an, and of appear slightly more often in Madison’s writing, and in the writing of the unknown author. Interestingly, many of these latter pairs are substrings of words Mosteller and Wallace found to be statistically interesting in their previous work. With this methodology, however, researchers would not need to spend time physically analyzing documents but instead allow a computer to search for patterns more efficiently.

In an attempt to reproduce Bennett’s and Tankard’s work, we have taken several samples of the Federalist Papers and pre-determined ten highest-frequency letter pairs in Madison’s work. As Tankard, we hope to find that the frequency of these pairs is different in Madison’s work than it is in Hamilton’s.

4.5.2 Methodology

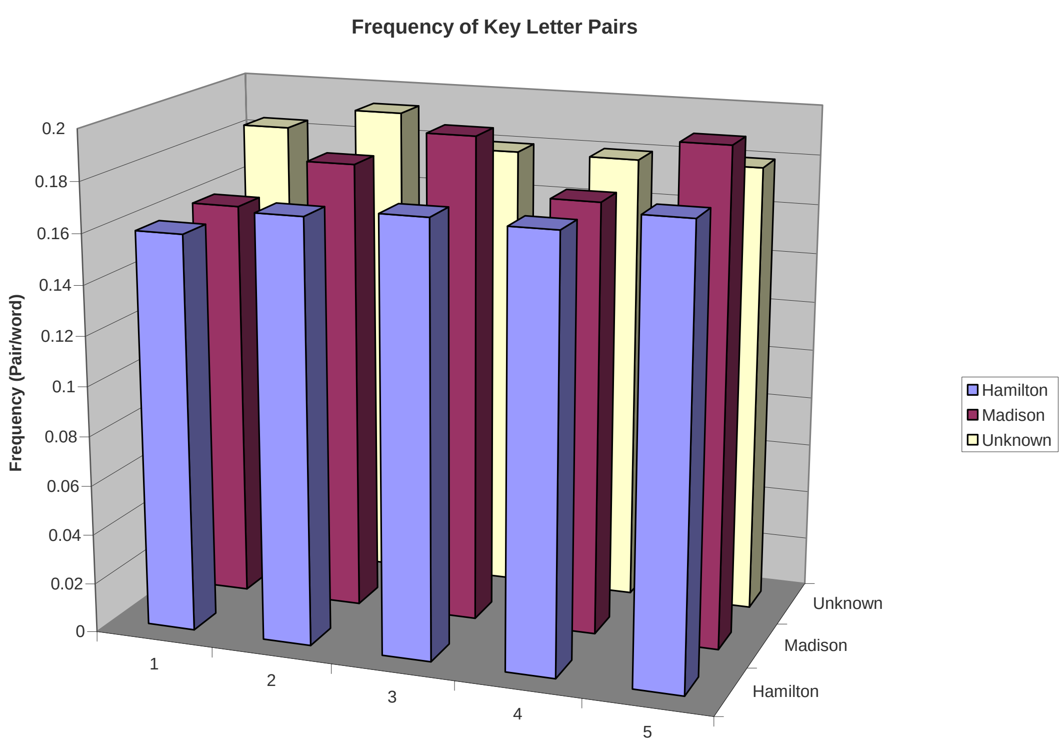
Applying TEAL to this problem takes is an exercise in metrics extended with a more complex algorithm. We must determine what we wish to count, in which documents, and then how we must transform the data in order to obtain useful results.

1. Select sample documents. The Federalist Papers, in their unedited entirety, have been compiled and are available from a verity of sources in the public domain. For the purposes of our experiments, five papers written by Hamilton, five by Madison, and five unknowns were selected completely at random from all non-Jay Federalist Papers. Given the assumption that all 12 disputed papers are written by the same author (a safe assumption given Mosteller, Wallace, and Tankard's results) it is not necessary to know the specific papers selected, and in this experiment, we do not, instead assigning them, randomly, the identifiers 1 through 5.
2. Create search string. Looking to Tankard's research, we pre-determined 10 string pairs that appears with high frequency in Madison's work (th, he, on, er, in, en, re, ti, an, and of). Using TEAL, we then created a regular expression that reluctantly matched all of these pairs. It's worth noting that TEAL uses a regular expression library that implements the Knuth-Morris-Pratt algorithm, and can not only match without human intervention, but also in the optimally efficient linear-time.
3. Run TEAL with sample documents and search string. After a runtime of mere seconds, results (both the sum of all found letter pairs and total number of words) are obtained.
4. Transform data. Each sample document has a different length. In order to normalize the frequency, we modify Tankard's method (key pairs per possible pairs) to the (we believe) more meaningful key pairs per word. This change gives us the probability of encountering a key pair in each word, which is a slight compromise between the word-based approach of Mosteller and Wallace and the pair based approach of Bennet.

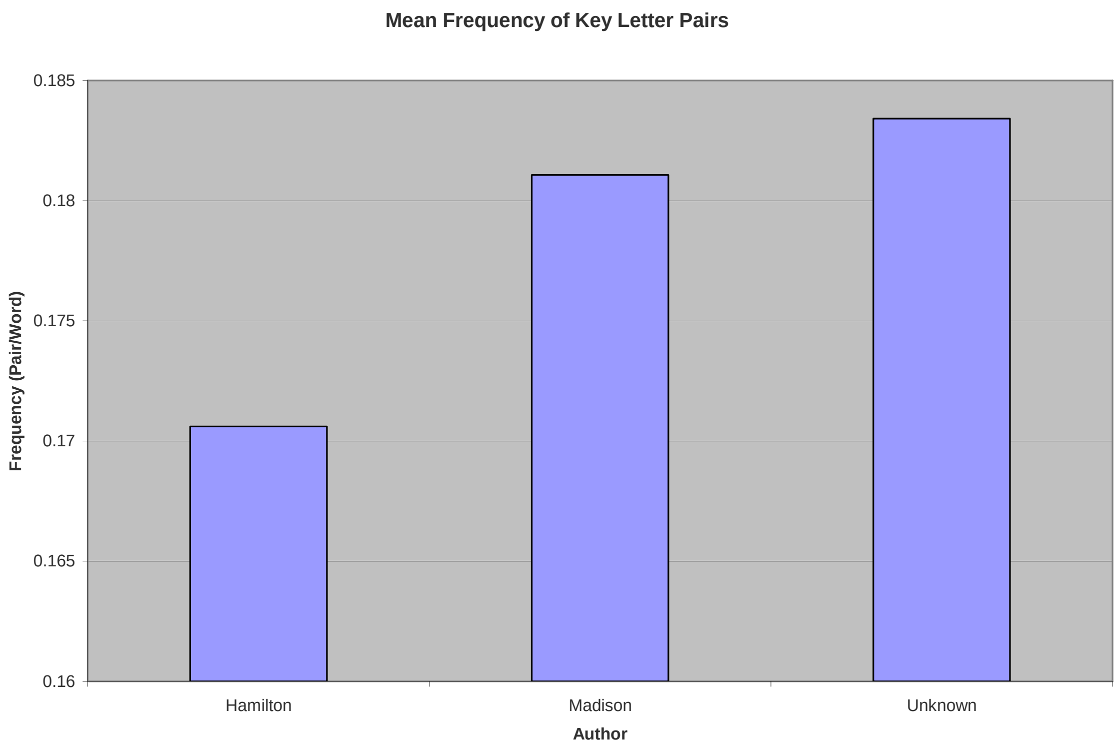
4.5.3 Results

	Frequency of Key Letter Pairs		
Author	Hamilton	Madison	Unknown
1	0.159289176	0.160518	0.184068
2	0.169660679	0.180556	0.192921
3	0.172888889	0.194664	0.180054
4	0.171710526	0.172472	0.180051
5	0.179487179	0.197183	0.180001
Mean	0.17060729	0.181079	0.183419
Variance	5.35678E-05	0.000235	3.13E-05
Std Dev.	0.007319	0.015338	0.005592

After a determining the frequency of the key pairs, we plot the likelihood a key pair will be encountered in a word.



Individually, the results are not obvious; individual documents vary both amongst authors and within. However, statistical analysis reveals that the variance is miniscule and the standard deviation within authors is, at most, .01 pairs per word. By contrast, the means of Hamilton and Madison are more than one standard deviations apart, suggesting a statistically significant difference between the two authors. Plotting the mean frequencies next to the disputed papers, there is a clear candidate for authorship.



Clearly, the probability of encountering a key pair in Madison's work is significantly higher than in Hamilton's. The unknown author has a probability close to (and indeed, within 0.002 pairs/word) Madison's, but more than one standard deviation away from Hamilton.

4.5.4 Conclusion

As in Tankard's work that we seek to reproduce, these results suggest that the disputed papers were written by James Madison. This result also agrees with Mosteller and Wallace, and with the vast majority of scholarly opinion (West, 54). Interestingly, these results are not as definitive as Tankard's (who found a far lower standard deviation). This suggests one of three things: the frequency of key letter pairs may not be as definitive as Tankard's pair per possible pair calculation suggests, or our pair per word calculation is not as effective at determining authorship as Tankard's, or that analyzing the remaining papers would significantly reduce the deviation.

4.5.5 Performance of TEAL

Interesting observations were also made in terms of TEAL. From the time the dataset had been obtained to the time results were analyzed, only three hours had elapsed, with the majority of time spent on constructing a suitable regular expression. In our opinion, this constitutes a success for the tool and a case study in the usefulness of simple metrics. Furthermore, it is a successful example of categorization based on statistical measures.

5 Results

Analyzing the results from our five experiments, several initial conclusions regarding the usefulness of empirical, computerized content analysis (at least when used on legal opinions) may be drawn.

First, as is evident by Experiment 1, empirical methodologies do have the potential to reveal interesting trends on the Supreme Court. While we found no uniform correlation, we were able to successfully link the sitting Chief Justice with the number of references to dissents under him. Had there existed a more strong correlation, our tools and methodology would have shown it. We found variations in the results do exist from year to year, and we were able to automatically enumerate them with little error, and quickly compare these results with a variable that might (but seems not to) explain it.

Had we wanted to reproduce this experiment without computer aid, we would be required to count manually (so as to filter out extraneous words not actually part of the opinion). Thus, we would need far more manpower or would be required to sample a smaller portion of the entire data set. Assuming we could still use a computer to manage the statistics generated (as is commonplace) we would still need to manually categorize each data point into its appropriate year and Chief Justice. This procedure would clearly take far more effort and more time, and would introduce many opportunities for human error.

Second, we revealed several things about categorization. While humans accomplish this task highly efficiently, and based on several different metrics at once, more care must be taken when computers are used. Categories based on concepts or contexts have proved difficult to differentiate between using statistical or numerical methods. To illustrate: in Experiment 3, we chose our metrics to differentiate between cases that are “important” (containing certain concepts and surrounded by a certain context) and cases that are not. This was not successful; the metrics were insufficient, and while our results accurately represented the documents, they did not serve to advance categorization based on the importance one iota. By contrast, categories based purely on content are more suited to computer analysis; in Experiment 4, categories were based on the result expressed in the opinion itself. This measurement was based only on the text contained in the opinion, not on the opinion of experts, on the reactions of society, or on any external factor. These categories were easy to differentiate, and opinions were sorted with very little error. Likewise, categorization based on year (which is written explicitly in the decision) and the voting spread (which, except in complex cases, is also explicit) are easy to handle, and complete far faster than a human could ever hope to read through even a tenth of the data set. Experiment 5 proved an interesting middle case; it used simple textual search, but with a complex algorithm based on key letter-pairs. Where previous attempts to use statistics failed, the Federalist papers were sufficiently differentiated by using more complexity.

Third, we have scouted the limits of basic textual search and simple metrics. To convert textual data into numeric results, we applied variations of simple metrics to enumerate specific features of the documents. After completing Experiment 2, we quickly found the limits of using simple text search to search for any complex metric. It is very difficult to generate a regular expression that will match all the ways an opinion author is likely to reference the minority's reasoning. Simple textual search is simply not sophisticated enough to do this. Our solution is somewhat stop-gap: we replaced a single search string with a "dictionary" of overlapping strings, including all combinations and permutations of words that could be used to refer to the desired feature, while excluding words that could not. While it proved sufficient for our purposes, this technique is somewhat cumbersome in comparison to other, simpler, computerized techniques. It requires an expert (or at least a person having read several opinions) to assemble the list, and is unlikely to clearly indicate false-negatives, should this list not be exhaustive.

Experiments 2 and 3 also highlight the limits of empirical metrics in the analysis of text documents as a whole. TEAL worked flawlessly in both tests, but, in both cases, failed to reveal anything of use. The key is that empirical numbers only reveal information about a data set when the numbers actually *mean something*. It seems, for example, that the length of the opinion has almost no correlation with the contents of the decision (In Part results being the exception). Thus, looking to the opinion length to determine anything to do with the contents makes little sense. Experiment design remains at least as important in computerized analysis as it is in more traditional studies.

6 Conclusion

We set out to determine the effectiveness of computerized content analysis, and trials on this data set have proven useful to this end. Compared to traditional content analysis, our experiments show it does have several advantages in addition to those mentioned previously. Each test is very efficient. After counting the number of references to the dissent in Experiment 1, we could cross-reference this data with other metrics in minutes. When our hypothesis that the Chief Justice was the primary causation for variation was not proven, we could have formed a new hypothesis and tested it more or less immediately, without designed an entirely new experiment. The man-hours required to generate data are far less; over the course of this entire project, we read approximately 100 opinions of the Supreme Court of the United States, but gathered data from over 20 times that. Traditionally, researchers or their assistants would be forced to read every single one of the opinions they wished to analyze, increasing man-hours, and likely decreasing the size of the sample. Results can be manipulated and analyzed proportionally as fast. Generating graphs, either traditional bar and scatter charts, or more sophisticated webs or relational charts, can be done seamlessly and fast enough to cycle through many possible charts, until a useful display is found that effectively shows correlations, or lack thereof.

However, we also determined that there are situations where computerized content analysis is just not effective. Answering questions that are conceptual or contextual is nearly impossible, at least when compared to a human solution. Computers can read a document, and search for known phrases. But, (barring an AI solution) unlike humans, they will not adapt to new phraseology, or learn new language from previous opinions. Likewise, computers are unlikely to “read into” a document. A clever (or exceptionally poor) writer obfuscating his meaning will easily outwit the most sophisticated computer systems. This is not as big an issue for humans; people can generally gain ‘meaning’ from an unusual sentence, and can recognize similar concepts presented in different phraseologies.

Overall, the most important point highlighted by this project is that of experiment design. In order to use computerized content analysis effectively, researches must use it only for tasks it is suited for: tasks that involve specific features of content, and not abstract concepts or views of context. Effort must be made to avoid asking questions a computer program cannot effectively answer.

7 Further Research

As this project is a look into the effectiveness of computerized methodologies, possibilities for future research are perhaps the most important byproducts of our research and experimentation.

First, as we noted, tests are efficient, and the generation of visual results is fairly fast. Techniques already exist in financial and business environments to quickly manipulate and visualize data such as that generated by computerized content analysis. Adapting techniques such as “On Line Analytical Processing” (a multidimensional approach to quickly generating answers to financial queries) to legal opinions might show correlations we have missed.

Second, experiments with algorithms, metrics, and data sets that lie outside the scope of this project may make for interesting and valuable tests. Computationally attacking text documents is a technique not often utilized in some subjects, and could reveal trends presently unknown. Likewise, the application of algorithms such as neural networks stands to improve understanding not only of the data-set, but of cognition and computer science in general. Some algorithms meant to improve computer understanding of, for instance, natural language, remain open problems in computer science.

Third, a key point made herein is that computers are not helpful answering questions of concepts. We propose that future research in this area could be extremely fertile. Building a system to differentiate concepts perhaps based on a complex set of dictionaries or rules (known as an ‘expert system’) would be very instructive. Many approaches to this problem remain untested, and the problem remains largely unsolved.

8 Bibliography

Bennett, W. R. (1976). Introduction to computer applications for non-science students. Englewood Cliffs, NJ: Prentice-Hall.

Cornell Law School. Supreme Court Collection. April 2, 2007 <http://www.law.cornell.edu/supct/index.html/>.

Hawkins, Jeff. (2004) On Intelligence. Times Books, New York

Krippendorff, Klaus. *Content analysis : an introduction to its methodology*. Beverly Hills, CA: Sage Publications, 1980.

Miller, Edward Alan. *Federal Oversight, State Policy Making, and the Courts: An Empirical Analysis of Nursing Facility Litigation Under the Boren Amendment*. Malden, MA : Blackwell Publishing, 2006

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem. Journal of the American Statistical Association, 58, 275– 309.

Mosteller, F., & Wallace, D. L. (1984). Applied Bayesian and classical inference: The case of “The Federalist” papers (2 nd. ed). New York: Springer-Verlag.

Rehnquist, William H. (2004). The Supreme Court. Alfred A. Knopf, New York

Riffe, Daniel, Stephen Lacy, and Frederick G. Fico. *Analyzing media messages : using quantitative content analysis in research.* Mahwah, N.J.: Lawrence Erlbaum, 2005.

Sharkey, Catherine M. *Dissecting Damages: An Empirical Exploration of Sexual Harassment Awards*. Malden, MA : Blackwell Publishing, 2006

Stone, Phillip J. *The general inquirer; a computer approach to content analysis.* Cambridge, MA: M.I.T. Press, 1966.

Tankard, J. (1986, February). The literary detective. Byte, pp. 231– 238.

West, Mark D. *Applications of computer content analysis.* Westport, CT : Ablex Pub., 2001.

West, Mark D.(Editor). *Applications of Computer Content Analysis.* Westport, CT, USA: Greenwood Publishing Group, Incorporated, 2001. p 59. http://site.ebrary.com/lib/wpi/Doc?id=10020872&ppg=65

9 Appendix A – Landmark Cases

In our second experiment, we compiled a set of landmark cases by cross-referencing the list made by experts of Cornell Law School's Supreme Court collection with those cases recognized as important by Wikipedia's community. This served to filter out both cases without legal merit and cases without any public awareness.

The list of cases we identified as landmark, in alphabetical order, is below.

Abington School Dist. v. Schempp
Adarand Constructors
Agostini v. Felton
Barnes v. Glen Theatre
Bolling v. Sharpe
Bowers v. Hardwick
Boy Scouts of America v. Dale
Brandenburg v. Ohio
Breard v Greene
Brown v. Bd. of Educ. of Topeka
Chaplinsky v. New Hampshire
Church of Lukumi Babalu Aye v. City of Hialeah
City of Boerne v. Flores
Clinton v. Jones
Cohen v. California
Craig v. Boren
Cruzan v. Dir.
Daubert v. Merrell Dow Pharmaceuticals
Erie v. Pap's A. M.
Escobedo v. Illinois
Feist Publications v. Rural Telephone Service
Fletcher v. Peck
Frontiero v. Richardson
Furman v. Georgia
Gates v. Collier
Gideon v. Wainwright
Goldberg v. Kelly
Gonzales v. Oregon
Gregg v. Georgia
Griswold v. Connecticut
Grutter v. Bollinger
Heart of Atlanta Motel
Hurley v. Iris American Gay Group of Boston
Hustler Magazine v. Falwell
Jones v. Alfred H. Mayer Co.

Jurek v. Texas
Katz v. United States
Katzenbach v. McClung
Korematsu v. United States
Lawrence v. Texas
Lee v. Weisman
Lemon v. Kurtzman
Lochner v. New York
Loving v. Virginia
Mapp v. Ohio
Marbury v. Madison
Martin v. Hunter's Lessee
McCulloch v. Maryland
Miller v. California
Miranda v. Arizona
Missouri v. Holland
New York Times v. Sullivan
New York Times v. United States
Planned Parenthood of Se. Pa. v. Casey
Plessy v. Ferguson
Printz v. United States
Proffitt v. Florida
Regents of the Univ. of Cal. v. Bakke
Reno v. ACLU
Roberts v. Louisiana
Roe v. Wade
Romer v. Evans
Roper v. Simmons
Rosenberger v. University of Virginia
Roth v. United States
San Antonio Independent School District v. Rodriguez
Schenck v. United States
Scott v. Sandford
Sony Corp. of America v. Universal City Studios
South Dakota v. Dole
Texas v. Johnson
Tinker v. Des Moines Independent Community School District
U.S. Term Limits v. Thornton
United States v. Lopez
United States v. Nixon
United States v. Virginia
Vacco v. Quill
Vernonia School District 47J v. Acton
Washington v. Glucksberg
Wisconsin v. Yoder
Woodson v. North Carolina

10 Appendix B – Example Code

For the interested reader, example code from TEAL is included below. Included first is the Content_Reader, which defines the methods needed to read from any textual file format. When circumstances require reading from other file formats, this class is extended to do so.

Second, Txt_Reader reads specifically from text files, and was used to accomplish our data-input for all of our experiments.

Third, TEAL-WordCount is a demonstration implementation of simple word counting using regular expressions.

10.1 Content_Reader

```
/** Content_Reader
 * Package: teal_reader
 * April 30th, 2007
 *
 * The Content_Reader abstract class features methods
 * for opening a directory as well as opening files.
 */
package teal_reader;
import java.io.*;


/**An abstract class for the TEAL computerized content analyser's input.
 * @author Zack A. Kleinfeld (zkleinfeld@gmail.com)
 * @author James A. Roumeliotis (bubbs@wpi.edu)
 * @version 0.2
 *
 */
public abstract class Content_Reader {
	/**BUFFER_SIZE defines the max buffer for use in reading files, in number of
bytes.
	 * Logical numbers are recomended.
	 */
	public static final int BUFFER_SIZE = 1000000;

	/**Sets the directory in which all files to be processed are located.
	 * @param inDIR Directory to open
	 * @return An array containing all files (and directories) located in the opened
directory
	 */
	public File[] openDIR(File inDIR) {

		File[] children = inDIR.listFiles();
		if (children == null) {
			//if there are no files in the directory, we have made a mistake.
			throw new NullPointerException();
			//FIXME: A custom exception class would be nice.
		} else {
			return children;
		}
	}

	/**Open's the indicated file and reads it to a string. Line Breaks and other
	 * special characters should remain intact.
```

```java
	 * @param fileList array of files containing the specified file to open.
	 * @param which Specifies the file in fileList to read to a string.
	 * @return a String containing the entire specified file.
	 */
	public abstract String openFile(File[] fileList, int which) throws IOException;

}
```

10.2 Txt_Reader

```
/** Txt_Reader
 * Package: teal_reader
 * April 30th, 2007
 *
 * The Txt_Reader class reads in the files individually
 * and converts each file into a string of characters.
 */
package teal_reader;
import java.io.*;

/**A class to read in normal txt files.
 * @author Zack A. Kleinfeld (zkleinfeld@gmail.com)
 * @author James A. Roumeliotis (bubbs@wpi.edu)
 * @version 0.1
 *
 */
public class Txt_Reader extends Content_Reader {

	/* (non-Javadoc)
	 * @see teal_reader.Content_Reader#openFile(java.io.File[], int)
	 */
	@Override
	public String openFile(File[] fileList, int which) throws IOException{
		// determine if the object is a file
		if (fileList[which].isFile()) {
			// create a Buffer reader
			BufferedReader in = new BufferedReader(new
FileReader(fileList[which]));
			StringBuffer buffer = new StringBuffer(BUFFER_SIZE);
			char[] cstring = new char[BUFFER_SIZE];

			// read in each character of the file
			while ((in.read(cstring)!=-1)) {
				buffer.append(cstring);		// append each
character to the buffer
			}

			in.close();		// close the buffer
			buffer.setLength(buffer.indexOf("\u0000")+1);
			String result = buffer.toString();		// convert to string
			result.trim();
			// return the file as a string
			return result;
		} else {
			return null;
		}
	}
}
```

10.3 TEAL_WordCount

```java
/** TEAL_WordCount
 * Package: teal_demo
 * April 30th, 2007
 *
 * The TEAL_WordCount is a demo class used to count all
 * of the words found in any number of files found in a
 * user specified directory.
 */
package teal_demo;
import teal_reader.*;
import teal_processor.*;
import java.io.*;
import java.util.regex.*;

/**Demo that shows how one would go about counting the number
 * of the occurances of a word.
 * @author Zack A. Kleinfeld (zkleinfeld@gmail.com)
 * @author James A. Roumeliotis (bubbs@wpi.edu)
 * @version 0.4
 */
public class TEAL_WordCount {

	public static void main(String[] args) {
		// create an imput reader and buffer
		InputStreamReader isr = new InputStreamReader(System.in);
		BufferedReader stdin = new BufferedReader(isr);

		// initialize cmd strings
		String cmdDir = null;

		// ask user to enter directory
		System.out.print( "Please enter a directory: " );
		try {
			cmdDir = stdin.readLine();
		} catch (IOException ioe) {
			System.out.println("IO error: Invalid directory input");
		  System.exit(1);
		}
		// create a text reader and open the directory
		Txt_Reader in = new Txt_Reader();
		File[] dir = in.openDIR(new File(cmdDir));

		// use regex to match  word count
		Pattern wordCount = Pattern.compile("\\W*\\W", Pattern.CASE_INSENSITIVE);

		// initilize variables for each file being read
		String doc;
		int totalWords =0;

		// calculate simple metrics for each file in the directory
		for(int i=0; i<dir.length; i++) {
			try {
				doc = in.openFile(dir, i);
			} catch (IOException e) {
				e.printStackTrace();
				doc = null;
			}
			if (doc != null) {
				totalWords = Simple_Meter.countPattern(doc, wordCount);

				// print out the metrics
				System.out.println(dir[i] + ": " + totalWords);
			}
		}
	}
}
```

11 Appendix C – Example Dictionary

Dictionaries can take one of two forums in our experiments: either a list of regular expressions or a list of words and phrases. When expressed as the former, they can be a list of regular expressions. This gives the tool the ability to find complex phrases, and simultaneously to categorize opinions based on which expression is found in the text. By increasing the size and complexity of the dictionary, we were able to enumerate arbitrarily complex textual features.

For example, in Experiment 4, we found we could begin to categorize opinions (based on result) with some very simple expressions. Expressions on the first line indicate reversal, while expressions on the second indicate affirmation. The expression on the third line suggests neither (and so is subject to other rules not present in the dictionary).

(<I>Reversed and remanded.</I>)|(<I>Reversed.</I>)|(The judgment .? is reversed)*
(<I>Affirmed.</I>)|(The judgment .? is affirmed)*
<I>It is so ordered.</I>

A dictionary alone is usually not sufficient to generate complex metrics, but can provide an effective starting point for enumeration and categorization.