Visualizing the Results of Shepards CiteFinder

Annie Jacobs
College of Computing
Georgia Institute of Technology
801 Atlantic Avenue
Atlanta GA 30332
+1 234 567 8910
imani@cc.gatech.edu

Kevin McDonald
College of Computing
Georgia Institute of Technology
801 Atlantic Avenue
Atlanta GA 30332
+1 234 567 8910
tribe@cc.gatech.edu

ABSTRACT
In this paper, we describe the process we undertook to design an interactive visualization for a data set created by the Shepard's CiteFinder legal search engine.

Keywords
Information visualization, document retrieval

INTRODUCTION
CiteFinder is a powerful tool for legal research. This information retrieval system utilizes a comparison and weighting algorithm to retrieve statutory and common law authorities. An attorney inputs between two and five case or statute citations and the engine produces a list of the cases and the statutes that are either cited by, or cite to, one, or more, of the inputs. The resulting cases on the list are ranked in order of relevance. The larger the number of the inputs that a result set element references to, or is referenced by, the more relevant that element is considered to be and, thus, the higher it is on the list. The assumption underlying this definition of relevancy is that it is in the relationships, or areas of intersection, between the inputs that the most useful material for the given question of concern is likely to be found.

All elements in the result set are preceded by small octagonal symbols that, according to their color, are meant to provide information about the contents and the status of that element. If the result set item was a case, this information would include topics such as whether it is from the same jurisdiction as one of the input cases and whether it has been upheld, overruled, or modified by a subsequent decision. As no legend appears on the display, the attorney using CiteFinder must already be familiar with the meaning of the symbols.

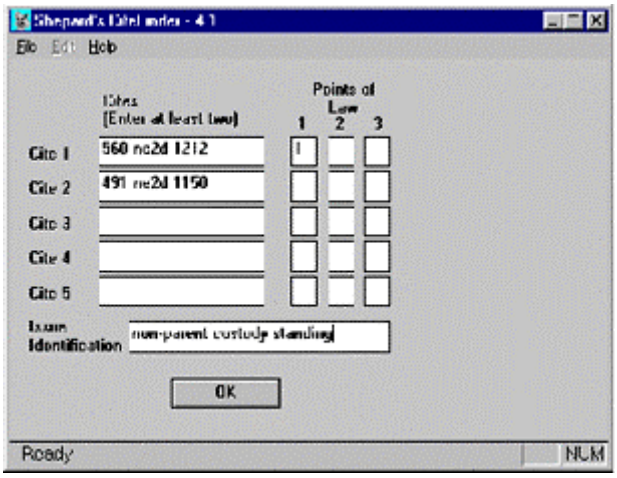

Fig. 1. CiteFinder Citation Input Screen - An attorney inputs between two and five case or statute citations

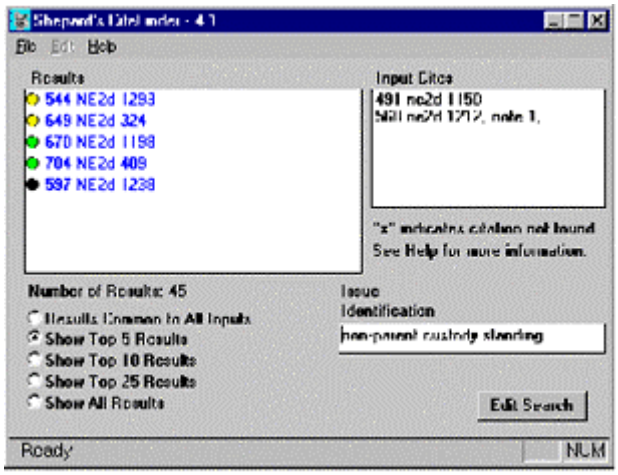

Fig. 2. CiteFinder's current method for displaying a result set. The engine produces a list of the cases and the statutes that are either cited by, or cite to, one, or more, of the inputs. The resulting cases on the list are ranked in order of relevance.

Rank	Citation	Relevancy	Issue Discussed
1	544 N.E.2d 1293	relevant	Custody petition standing (stepmother)
2	649 N.E.2d 324	relevant	Custody petition standing (adoptive parents)
3	670 N.E.2d 1198	relevant	Custody petition standing (stepmother)
4	597 N.E.2d 1238	relevant	Custody petition standing (grandparents)
5	510 N.E.2d 555	relevant	Custody petition standing (aunt/uncle)

Fig. 3. Another view of the relevancy list generated by CiteFinder

The ranked list that CiteFinder returns only barely begins to satisfy the needs of an attorney who is writing a brief or other document or preparing for trial. The list does give the lawyer the citations for cases and statutes that are highly likely to be relevant to the issue at hand but it says nothing about the internal structures of those cases and statutes.

The tutorial for CiteFinder states that the best way to find additional relevant authority is to examine the relationships between cases and statutes that are already known to be on point. The algorithm used to produce the result set and display it in ordered fashion may, indeed, look at those relationships but the ranked list is of very little use in sharing that information with the lawyer. The lawyer must have access to this relational information if he or she is to know not only what authorities a court has used in its' discussion of a point of law but precisely how the court has invoked them. Employment of case and statutory authority includes questions such as what scope and weight did the court assign to that authority and how did the court depict the connections between points of law and supporting authorities?

GOALS

At the inception of our project, we hoped to solve some of those shortcomings through the design of an information visualization that would bring forth the relationships between result set elements. Our visualization achieves this goal by revealing the deep structure of each case returned in the result set. The visualization uses multiple views to show how the cases and statutes that a decision contains, and upon which it relies to support itself, are related both to each other and to the specific substantive points of law that the attorney had indicated interest in. The attorney uses this structural information to craft an argument about how the case is, or is not, relevant to the issue at bar. Designing that argument is a highly flexible and creative process. Consequently, the visualization is not meant to deliver pre-formatted instructions for the application of the case material to the matter under consideration. Rather, the purpose of the visualization is to furnish the attorney with the information that he/she needs to shape and perform that application.

THE DATA SET AND THE TASK

The data set we used resulted from a search on the phrase "*death penalty"* in the database of decisions issued by the Supreme Court of Oklahoma. The search returned seven cases with a wide variety of related issues. One case dealt with the payment of a lawyer's fee for representing an indigent defendant in a capital case; another dealt with the issue of whether or not an indigent defendant can require the jurisdiction to pay the expenses of an out-of-state witness; and yet another set centered on the adoption of a minor child over the protests of an imprisoned parent whose death sentence was commuted to several life-terms. Annie Jacobs, a member of both the Georgia Bar and this project team, read through each of the cases in the result set and highlighted every case law and statutory reference contained therein. Next, working with each case individually, she drew a graphical representation of the connections between that case and its references. These representations became the basis for a significant part of the visualization.

The first graph she drew had the primary case at the center. She placed the references in a web-like arrangement around this focal point and drew lines from them to the center. Ovals represented statutes and rectangles depicted cases. Within each shape, she included a brief statement as to the contents of that statute or case. To the line that connected a shape to the center, she attached a sentence or two as to why the court had cited to that case.

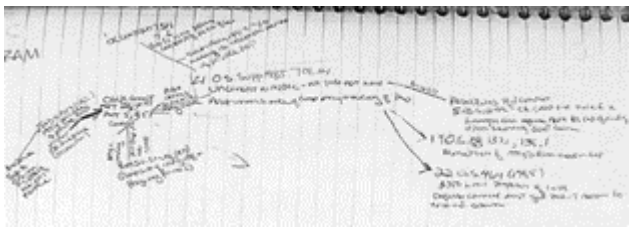

Fig. 4. A web diagram of a case related to the Oklahoma death penalty.

As she progressed, she used a slightly different arrangement for the remaining cases. For instance, she placed the primary case at the top of the graph rather than at the center. The ovals, rectangles, and their accompanying commentaries then flowed downwards.

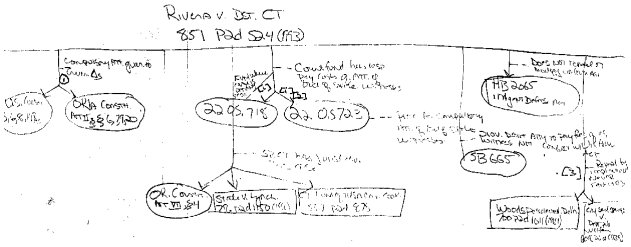

Fig. 5. The top-down approach resulted in an arrangement more compatible with the way an attorney would engage in legal analysis.

There was no conscious decision to switch styles. It was only at the end, when comparing the sketches that she realized what she had done. This led to some interesting conclusions. While she found the web diagrams to be more aesthetically pleasing, the top-down arrangement felt more compatible with the *way that an attorney would engage in legal analysis.*

Lawyers tend to burrow down, deeper and deeper, from the surface case. As they do so, they encounter an ever more intricate network of citations to cases and statutes. An efficient way in which it is possible to move through these interweaving paths is to proceed along one branch at a time. Sometimes, an attorney must return to the center trunk of the main case before choosing another branch. At other times, the attorney is led from the end of one branch directly to the end of another.

The visualization tool was designed to enable this type of search. A researcher should be able to examine a single branch while also maintaining a sense of the placement of that branch within the entire network of citations and reasoning. It will not always be necessary to have focus and overview simultaneously but, at times, such a combination will be invaluable. The visualization must also do more than simply show the existence of connections between references and between a reference and the main case. It must provide access to the content of those connections.

ITERATIVE DESIGN I – PAPER SKETCHES

Throughout the project's course, we followed an iterative design process that progressed from low to high fidelity prototypes. Initially, we explored some novel visualization techniques (based on the literature presented in class). Through a series of paper sketches, we slowly narrowed the focus of the visualization. Some early designs utilized a combination of the *fisheye* lens [1] with a *hyperbolic* browser [2] representing the cases in the result set. The entire hierarchy of cases would be contained within a circular region. In the initial view, icons rather than text would denote case names. By applying a lens, the user could get more detail from the icons, including the name of the case and a synopsis. In these early designs, we included sliders that would give the researcher the opportunity to customize the visualization. Similar to what we saw in Spotfire, an attorney could determine the number of cases shown on the visualization field by using sliders representing different aspects of each case (i.e. case name, point of law, jurisdiction).

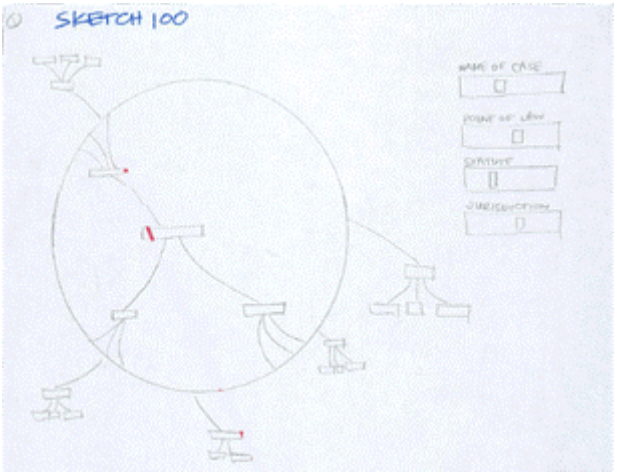

Fig. 6. Fisheye + hyperbolic browser view – researchers could use the categorical sliders to determine the number of cases visible on the screen.

Another early concept employed a *tiled* view as an alternative to the fisheye lens. The tiled and modified hyperbolic views provided focus and context without the distortion associated with the fisheye.

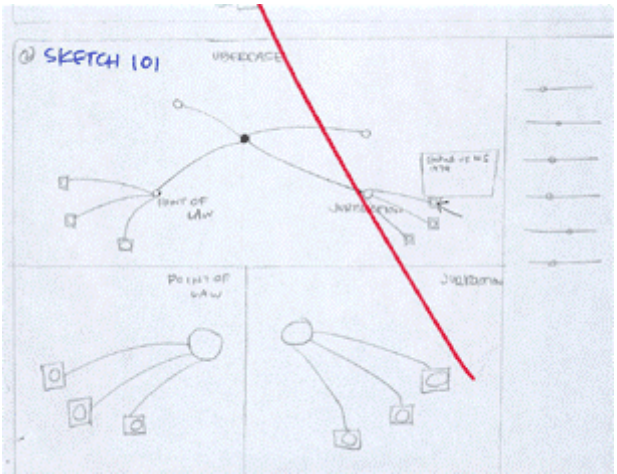

Fig. 7. The tiles + hyperbolic browser view provided focus and context without the distortion of the fisheye lens.

While the hyperbolic browser was certainly interesting from a visualization point of view, we were not convinced that it would be of much assistance to the researcher. With this in mind, we moved away from the hyperbolic visualization and towards the top-down approach that Annie had mapped during her early work with the data set.

Subsequently, we replaced the hyperbolic hierarchy with a flow diagram (or case map) that was reflective of her

original drawings. In this design, we introduced a brushing functionality meant to provide the researcher more detail about each member of the result set.

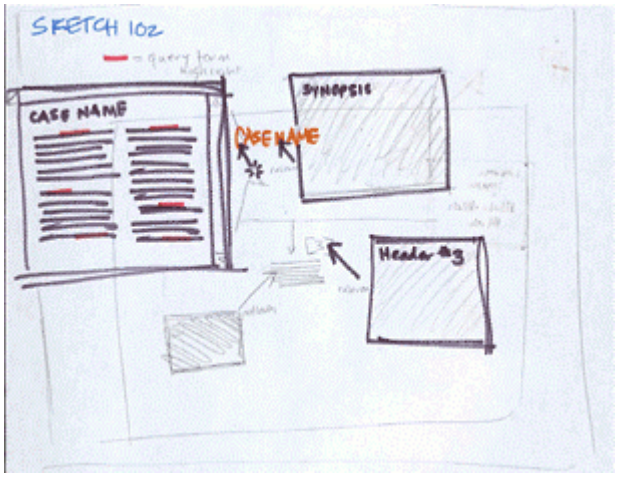

Fig. 8. This design reflects our return to the case map layout found in Annie's original drawings. Brushing provides the researcher with increasing levels of depth about each case.

ITERATIVE DESIGN II – STATIC PROTOTYPES

In our review of the sketches, we always kept the lawyer's *task* foremost in our minds. As we progressed from paper sketches to static prototypes (completed in Illustrator), we decided to focus our attention on two visualizations intrinsically related to the researcher's task. One was the case map diagram mentioned above while the other was a set of overlapping circles representing the points of law. This second visualization surfaced as the most pivotal during our early design efforts.

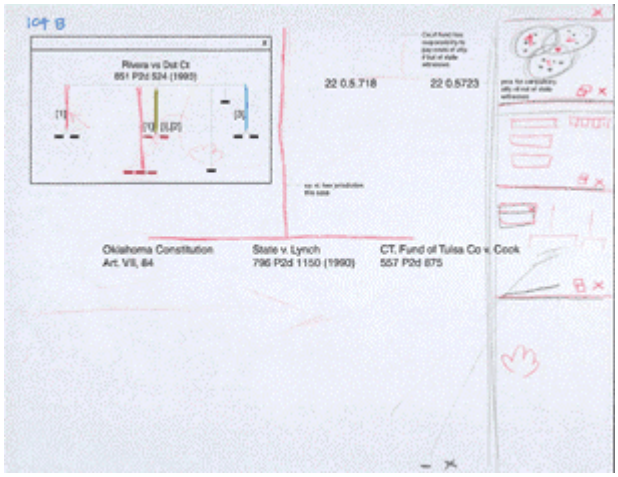

Fig. 9. The case map view emerged as one of two important visualizations during the progression from paper sketches to electronic prototypes.

In Figure 9, the mini case map (top left) provides overview, while the large visualization (center) provides detail. The right hand menu allows the researcher to change the view to other visualizations (i.e. the points of law view).

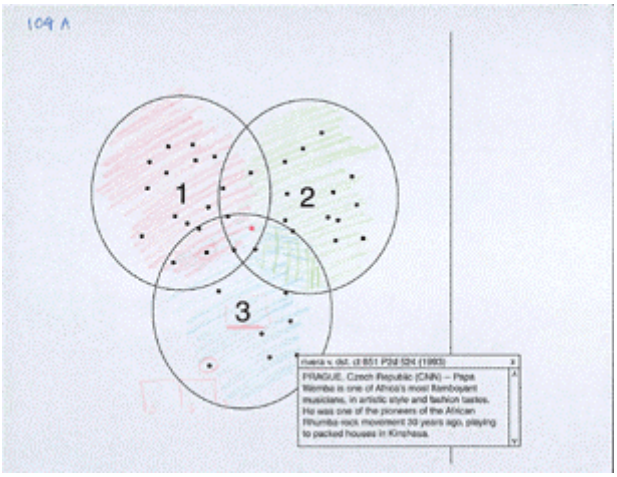

Fig. 10. Concentric circles represent the points of law. The overlapping regions contain cases that contain multiple points of law.

While working on the concentric circle design, we entertained the concept of motion (via a slider tool) to animate the relationships between the points of law.

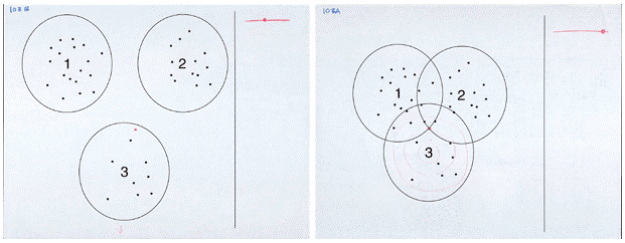

Fig. 11. Experimenting with animation – a researcher uses the slider (top right) to converge the points of law.

Ultimately, we decided the animation was gratuitous and of little import to the researcher. A static view would contain the same relational information while conveying a more professional, less toy-like, feel.

As we iterated through the paper sketches and static prototypes, it became clear that color would play and important role in the visualization. Color would differentiate the points of law (and the associated citations) in the concentric circle visualization. These color assignments would pervade the rest of the views and allow the researcher to recognize the prevailing point of law no matter what the level of detail at which they were viewing the information.

After another run through of the existing CiteFinder tutorial, we included a bull's eye relevancy view as an additional tool for the researcher. CiteFinder returns a list of cases that reference (or are referenced by) the input citations. We agreed it was important to visualize the list in order to allow an "at a glance" capability to determine the most relevant cases. Using Director, we built a small

prototype with simple interactions and dummy data to demonstrate how the visualization might work.

ITERATIVE DESIGN III – DIRECTOR PROTOTYPES

In this first prototype, we offered the attorney **three** views into the data set returned by CiteFinder. One view depicted the relevant *points of law* and which cases demonstrated those points. Three circles with unique colors represented each point of law. Each case (represented by a small icon) was placed in the circular region of the point of law it referenced. If the case referenced multiple points, it was located within the area of intersection of the appropriate circles.

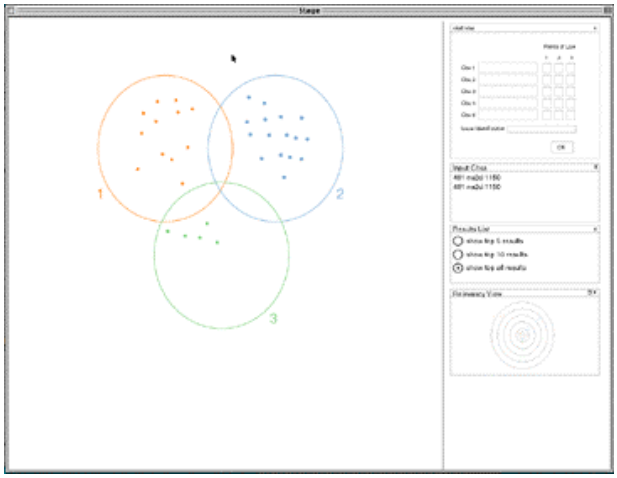

Fig. 12. Points of Law view

The second view was the *relevancy target*, which visualized the relevant and irrelevant cases in a target.

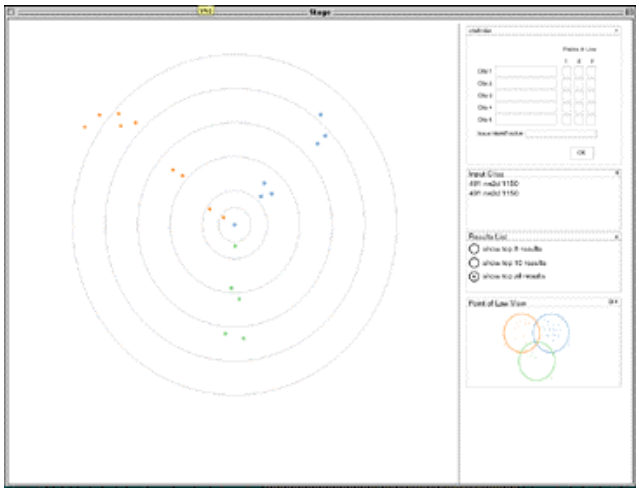

Fig. 13. Relevancy Target view

Finally, the *details* view allowed the attorney to delve deeper into specific cases of interest.

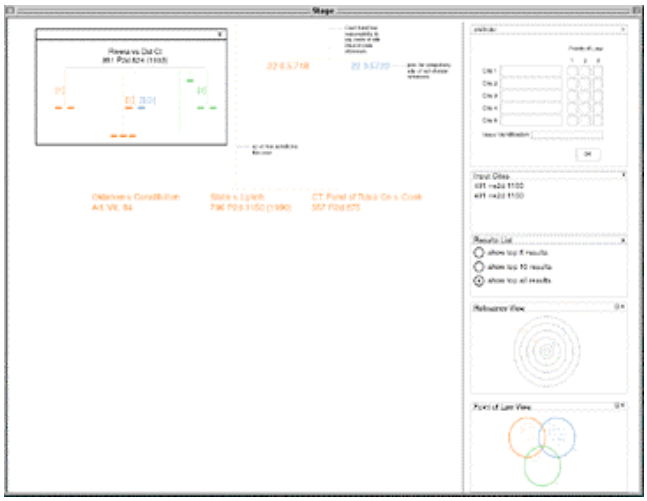

Fig. 14. Details view

In each of the views we provided a brushing functionality that enabled the attorney to see some initial information (i.e. a synopsis of the case, the subject matter of the point of law represented by a number) about the case without having to leave the current view. The user could also switch between the views using a series of scaled windows located in a right hand toolbar. Additionally, actions undertaken in the current view (i.e. selecting case in *points of law*) were reflected in the scaled representations.

DESIGN ERRORS

As we had anticipated, some early design decisions that seemed successful turned out to be the opposite. While the individual case maps for each member of the result set appeared to work well; the view with the interlocking circles representing the points of law did not.

Not long into the design process, we became aware that the lack of uniformity in the underlying database would be an issue. What is point of law #1 in one case may, very well, have a different number(s) in the other case or cases in which it is present. We thought that we had adequately addressed this matter by means of a brush-over of the numeral that identifies a circle. The brush-over would reveal the substance of that point. Having solved the problem, we believed, we forgot about the arrangement of the database and concentrated on other aspects of the visualization.

In the course of putting the real data into our design, we looked again at the input citation box and its' point of law fields. Suddenly, it became clear that we had not solved the uniformity problem after all. The points of law fields are labeled 1,2,3. If our circles are also labeled with numbers, how is the system to know, for example, whether circle #1 should contain point of law #1 from the second or the fifth input citation? We realized that the lack of uniformity created a situation in which it was not possible to show both

the relationships between cases and the distinct points of law with a single image.

How, we wondered, had we come to make the design error that we did? Upon reflection, we saw that, in our desire to visualize the relationships between points of law across all of the cases in the result set, we had forgotten about the absence of uniformity in the database and the consequences thereof for our design. We also surmised that one reason why CiteFinder displayed its' results as a simple ranked list was that, in doing so, the problem that we encountered was avoided.

ITERATIVE DESIGN IV – THE FINAL PROTOTYPE

In our final design, we solved this problem by creating separate sets of "point of law buckets" for each result set case. The buckets show the contents of each result set case arranged by point of law. For the input citations, the buckets also show references to those inputs. These, too, are ordered according to point of law. We kept the case maps. The maps show the same point of law contents but with the added information of how the points are related to one another in the overall structure of the case.

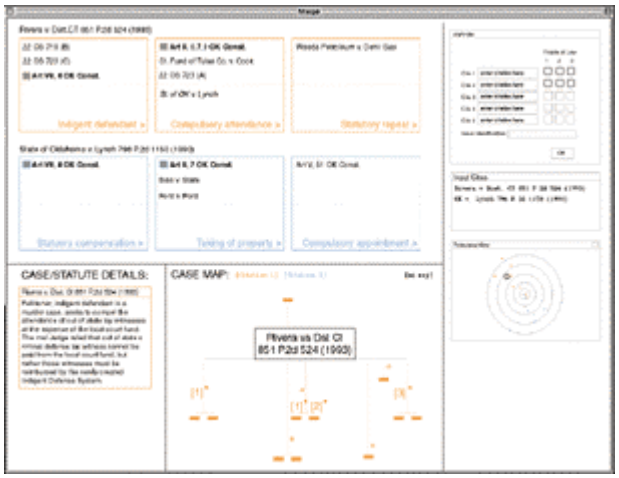

Fig. 15. The Point of Law Buckets - the buckets show the contents of each result set case arranged by point of law

Different cases are denoted by different colors. In order to take maximum advantage of the fact that the CiteFinder search engine returns both cases/statutes that cite to one (or more) of the inputs and those cases/statutes that are cited by one (or more) of the inputs, we designed the buckets for the input citations so that within the buckets, the cases are arranged by the direction of the citation. If the resulting cases cite *to* the input case, they are placed above a faint boundary line that separates the bucket into two halves. If the resulting cases are cited *by* the input case, they are listed below the boundary line. This representation is reflected in the case map window.

The search engine does not return this bi-directional information for those elements of the result set that are not one of the input citations. For one of these 'ordinary' cases, the search engine is able to provide only information (i.e. other statues and cases) regarding what that case points to. Consequently, the buckets for this result set cases do not contain the dividing line. For both input citations and 'ordinary' cases in the result set, an adjacent square icon and bold text denote elements that are shared by two or more of the result set cases.

The bull's eye was adjusted to reflect the combined weights of all of the points of law.

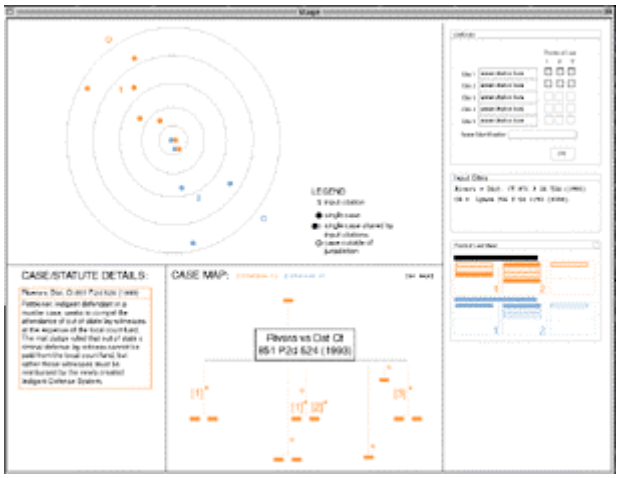

Fig. 16. The Relevancy View

In the center is the case(s) that contains the highest number of points of law across the broadest range of input citations. The next rings of cases represent those that have, first, fewer points of law from all input cases and, second, fewer points of law from a subset of the input cases. Cases with points of law from a greater number of the input citations will always be closer to the center than those that contain points of law from a fewer number of the input citations. The bulls-eye works where the interlocking circles do not because the former does not identify the individual points of law and, thus, does not have to worry about confounding point of law number with the underlying substance of that point. The result set cases for which the point of law buckets is currently displayed in the main window are highlighted in the bulls-eye.

A result set for a typical legal research query could easily contain a large number of items. Fifteen to twenty elements would not be at all unusual. The default view of the result set will, however, be to show the point of law buckets for only two cases at a time. This design choice is driven by the nature of the analytic task and not by constraints such as limited screen real estate.

When examining a result set case in the point of law view, a lawyer would need to read the contents of each bucket very carefully. He or she would also need to compare the contents of the different buckets within a single case and between cases. The lawyer is looking for more than the authorities (cases and statutes) that the court cites in its' discussion of a particular point of law. The lawyer must have information on both the scope and the weight that the court gives to these authorities, how it understands the relationships, if any, between them, and how it portrays the connections between the different points of law. The presence, or absence, of the same authority(ies) across the points of law of a single case and of multiple cases can shed the needed light on these matters.

Given these informational requirements, it might seem that displaying as many result set cases as possible in the point of law bucket view would be desirable. Showing more than two cases at a time is not a wise option for the following reasons: 1) Legal citations, particularly to statutes, are complex. They contain title, section, sub-section, and paragraph numbers and letters. Many are almost identical and, when presented in close succession, it can be difficult to disambiguate one from the others even when the type is of a large size. 2) The distinctions between the ways that the same authority is used in multiple separate cases, whether decided by the same or different courts, are frequently subtle. Teasing out these points of divergence demands a great deal of in-depth, focused attention that is directed to pairs of cases at a time. Attempting to address three or more cases simultaneously in this manner fosters confusion more than clarity.

The attorney using our visualization system will be able to scroll down and see the bucket displays for the remaining cases in the result set. As the next display appears, the one at the top of the screen will roll off. The attorney will also have the option of shifting to a view that will show the buckets for all of the result set cases at once. This view, which will replace all of the other elements on the screen, is intended for quick overview and orientation purposes. It is not designed for analysis for the reasons just discussed.

In the final design, we returned to the tiled window view to allow for Tufte's *macro/micro* view [3]. We also included the notion that behaviors in one window are reflected in another. The tiled view was included to support the task mentioned above – the attorney's need to delve deeper into a case while still maintaining some context of the big picture.

In returning to the tiled design, we modified the details view we proposed in our first Director prototype. One will recall that we chose to enlarge a portion of the case map to show focus, while keeping a small version of the map in the top left corner of the visualization to show context. In the final version, we provide detail for the researcher by allowing them to click on a case name and have the full text of the case appear in the upper tile. They can read through the text while still maintaining a sense of the case's place within the hierarchy. The case in focus is also highlighted in the case map window as well as on the smaller points of law and bulls eye windows in the right-hand menu.

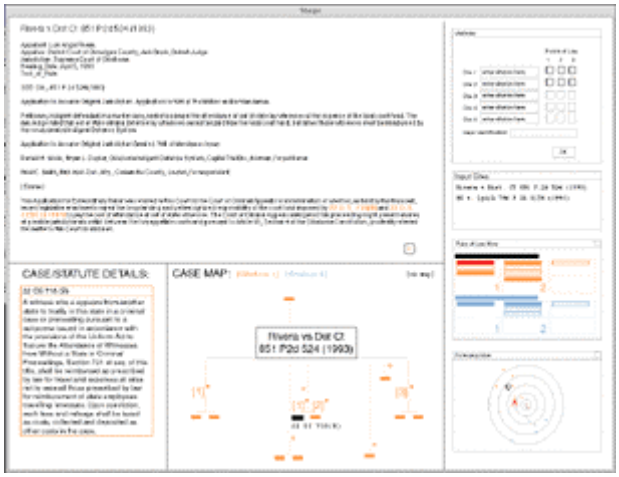

Fig. 17. The Case Details View

LIMITATIONS

From a visualization point of view, we were constrained by the nature of the legal researcher's task. Many of the novel techniques we studied in class simply did not apply, in a direct sense, to our problem domain. Nevertheless, there are examples of many of the important underlying principles of information visualization that were covered – most notably the use of the subject area and the needs of the target user to inform and direct the design decisions.

REFERENCES

1. Sarkar, M. & Brown, M. H., "Graphical fisheye views", *Communications of the ACM*, Vol. 37, No. 12, pp. 73-84, Dec. 1994.
2. Lamping, J. & Rao R., "The Hyperbolic Browser: A Focus + Context Technique for Visualizing Large Hierarchies", *Journal of Visual Languages and Computing*, Vol. 7, No. 1, pp. 33-55, March 1996.
3. Tufte, E. (1990). "Micro/Macro Readings", *Envisioning Information* (pp. 37-51). Cheshire, CT: Graphics Press