Design and Evaluation of the Tightly Coupled Perceptual-Cognitive Tasks in Knowledge Domain Visualization

Marie B. Synnestvedt

University of Pennsylvania School of Medicine
Center for Clinical Epidemiology and Biostatistics
107 Blockley Hall
423 Guardian Drive, Philadelphia, PA 19104-6021
Email: synnestv@mail.med.upenn.edu

Chaomei Chen

College of Information Science and Technology
Drexel University
3141 Chestnut St., Philadelphia, PA 19104-2875
Email: chaomei.chen@cis.drexel.edu

Abstract

Knowledge Domain Visualization (KDViz) aims to reveal the dynamics of how knowledge domains evolve over time. It is the ultimate role for information visualization to facilitate a wide variety of user tasks associated with the study of such dynamic, large-scale, transient, and self-organized phenomena. Since the evolution of knowledge domain is necessarily influenced by an array of factors, the design and empirical evaluation of a KDViz system is particularly challenging. In addition, KDViz typically involves a complex and lengthy sequence of data transformation and computational modeling that must be simplified for effective handling by the user interface design. In this article we present a perceptual-cognitive task taxonomy for designing and evaluating KDViz systems. The design implications are illustrated by the development and some refinements of CiteSpace, a Java application that enables users to study how networks of research fronts develop over time. Cognitive tasks such as finding pivotal intellectual points in an evolving knowledge domain are translated into tasks for visual search of objects with visually salient features. The design rationale is explained in terms of the taxonomy, including the use of Gestalt principles, graph drawing aesthetics, and pre-attentive visual search. The implications of the taxonomy on conducting empirical evaluations are illustrated by a case study using evaluation techniques focusing on usability issues. Three levels of user tasks are considered by the taxonomy: 1) the lowest level: manipulating the user interface; 2) the medium level: manipulating the visualization; and 3) the highest level: tracking the growth of a knowledge domain.

1 Introduction

Knowledge Domain Visualization (KDViz) is a type of information visualization method that aims to reveal the dynamics of how knowledge domains evolve over time (Chen & Paul, 2001). Progressive Knowledge Domain visualization is a KDViz technique under development that is specifically concerned with temporal patterns of intellectual turning points as a knowledge domain advances (Chen, 2004b). While Information Visualization in general is moving out of research laboratories with a growing number of commercial products and infiltration of main stream applications, Information Visualization interfaces are still a novelty for many users. Systematic and practical empirical evaluations of information visualizations still fall behind the increasing number of design options (Chen & Czerwinski, 2000). The first information visualization meta-analysis was conducted by Chen and Yu (2000). A more recent assessment of the current evaluation practices in the field classifies evaluative studies into four groups (Plaisant, 2004):

1. Controlled experiments comparing design elements such as widgets.
2. Usability evaluation of a tool.
3. Controlled experiments comparing two or more tools, the most common type of study.
4. Case studies of tools in realistic settings, the least common type of study.

The reports of Information Visualization usability studies and controlled experiments have been able to demonstrate faster task completion and reduced error rates as measured in laboratory settings but empirical studies could be improved. Empirical evaluations have generally included only simple tasks such as locate and identify tasks, while tasks requiring users to compare, associate, distinguish, rank, cluster, correlate or categorize are rarely included in study designs. Reporting of comparative studies could be improved by reporting results per task instead of reporting overall performance for a combined set of tasks. Task selection remains an ad-hoc process that would be aided by the development of task taxonomies (Plaisant, 2004).

2 Theoretical Foundations for Evaluation of KDViz

Good representations are not arbitrary but take into account the properties of the visual system (Ware, 2000) and good empirical studies are built on solid theoretical foundations (Chen, 2004a). Ware (2000) maintains that "Visualization is all about mapping data into a form that enables humans to perceive important patterns", and "Good representations are not arbitrary but take into account the properties of the visual system". The techniques for making "good information visualization" involve the use of design guidelines based on tested scientific theories.

Gestalt Laws are among the most widely known principles of pattern perception (Koffka, 1912; Ware, 2000). Ware translates Gestalt Laws into a set of design principles for information displays:

- *Spatial Proximity*: Things that are close together are perceptually grouped together.
- *Spatial Concentration*: Regions of similar element density are grouped together perceptually.
- *Similarity*: Elements of similar appearances tend to be grouped together.
- *Continuity*: Smooth and continuous connections between elements are easier to perceive than abrupt changes in direction. More fundamentally, continuity assumes connectedness, which can be a more powerful grouping principle than proximity, color, size, or shape.
- *Symmetry*: Symmetrically arranged elements are perceived as forming a visual whole much more strongly than elements with lesser symmetry.
- *Closure*: A closed contour tends to be seen as an object.
- *Relative Size*: Smaller components of a pattern tend to be perceived as objects.
- *Figure and ground*: Object like figures are perceived as being in the foreground, the ground is whatever lies behind the figure.

Many information visualization designs have been influenced by these principles. Spatial proximity, for example, has become one of the most widely known and tried design rationales, especially in the force-directed graph drawing paradigm (Eades, 1984; Fruchterman, 1991). The main processes in graph and visualization comprehension are established by Bertin's influential research on graph and visualization comprehension task analysis (from Trafton, Marshall, Mintz, & Trickett, 2002).

1. Encode visual elements of the display: For example, identifying lines and axes. This stage is influenced by pre-attentive processes and is affected by the discriminability of shapes.
2. Translate the elements into patterns: For example, notice that one bar is taller than another or the slope of a line. This stage is affected by distortions of perception and limitations of working memory.
3. Map the patterns to the labels to interpret the specific relationships communicated by the graph. For example, determine the value of a bar graph.

Cleveland and McGill (1984) proposed the theory of graphical perception (the visual decoding of information encoded on graphs) as a guideline for graph construction. Ten elementary perceptual tasks were ordered from most to least accurate with the ordering determined by how accurately people perform those tasks based on the theory and study of psychophysics and the authors own reasoning and experimentation. Cleveland and McGill demonstrate how they were able to, guided by their theory, identify the weaknesses in graphical design and improve the design subsequently. They advocated graph design that uses elementary perceptual tasks as high in the order as possible, thereby eliciting judgments that are as accurate as possible and maximizing the viewer's ability to detect patterns and organize quantitative information. Cleveland and McGill's original list of elementary perceptual tasks did not include color hue and texture. Recent studies by Kosara, Healey, Interrante, Laidlaw, and Ware (2003) have identified optimal color sequences for representing categories and form.

3 Evaluation of Knowledge Domain Visualization

The unit of analysis in Knowledge domain visualization is a dynamic, self-organized, and emergent complex intellectual system that underlies a topical theme, a field of study, a discipline, or the entirety of science. Several existing knowledge visualization systems such as SemNet focus on the graphical representations of three types of components (Chen & Paul, 2001):

1. The identity of individual elements in a large knowledge base
2. The relative position of an element within a network context
3. Explicit relationships between elements

Within these components, there are perceptual and cognitive tasks which might be involved in decoding a knowledge domain visualization that can be connected to elementary perceptual tasks (Cleveland & McGill, 1984) and Gestalt principles (Ware, 2000). Using a document co-citation approach to knowledge domain visualization and the CiteSpace interface (Figure 1) as an example, a series of sub-tasks can be identified:

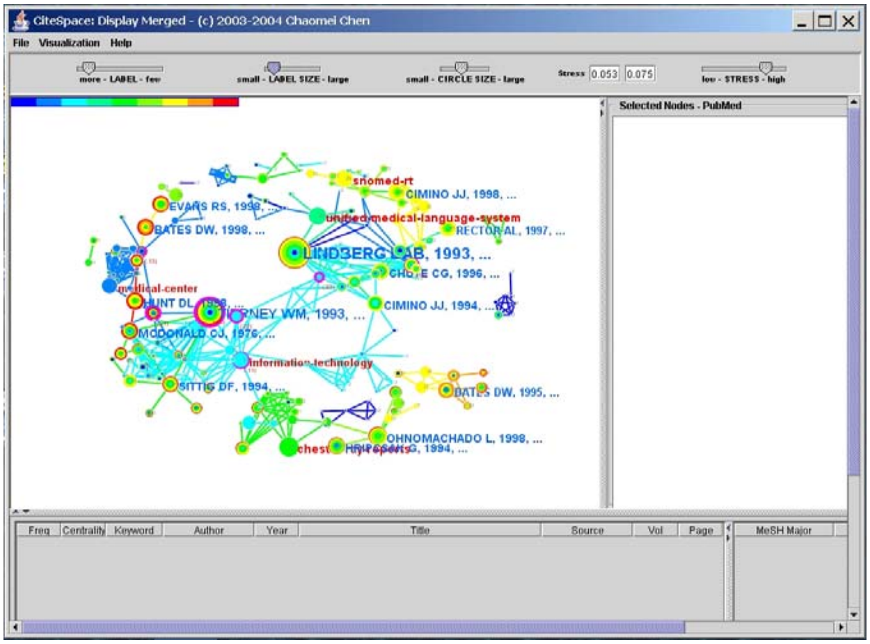

Figure 1: CiteSpace interface.

1. Information regarding significant individual documents within a large document base:
   - How often has the document been cited, relative to others? In the CiteSpace design metaphor, frequency of citation is represented by size of a circle, which involves a comparison of area in terms of elementary perceptual tasks (Cleveland & McGill, 1984).
   - What position does the document hold during the overall time period and when did it become initially important? These two questions involve a perceptual task of assessment of color, and a cognitive task of switching attention between a color scale and a graph. This task could be switched to a more accurate perceptual task if the graph is represented time along a common scale. In addition to the node-and-link view, CiteSpace now provides a TimeZone view, in which entities are shown in parallel vertical strips corresponding to their publication time.

2. The relative position of a document within a network context:
   - Is a document related to other documents (representing a common theme)? This involves two cognitive tasks from a Gestalt perspective – both assessments of continuity of connections and spatial proximity. The connections between documents will be easier to recognize if the co-citation network has been pruned to observe the graph drawing aesthetic of minimized edge crossings, and the graph laid out in way that clusters have some distance between them.

3. Explicit relationships between documents:
   - Is a document more frequently co-cited with one document relative to others? In the CiteSpace design metaphor, the frequency of co-citation has been represented by the width of connecting lines between two documents. This is analogous to an elementary perceptual task of comparing length.

When moving from focusing on assessing individual elements (documents) to assessments of clusters of documents, two tasks involved are evaluating the persistence of clusters over time, and the semantic meaning associated with clusters. Evaluating the persistence of clusters over time again involves the perceptual task of assessment of color and a cognitive task of switching attention between a color scale and a graph, as well as the cognitive task of trying to spatially manipulate the clusters along a mental model of a time-scale. The semantic meaning of clusters in the example is impossible to evaluate without prior knowledge of the domain being visualized, which ideally underlines a fundamental challenge to evaluating a KDViz system in general because of the very nature of what we want to achieve with KDViz systems. In order to convey the structural and dynamic characteristics of a knowledge domain in a concise and meaningful way, it is not only inevitable but also necessary for a comprehensive evaluation to take any knowledge gains resulted from the use of a KDViz system into account. Our pilot study is also expected to make methodological contributions to the evaluation of KDViz systems.

4 Methods

The primary goal of this study was to evaluate the visualization in a way that would better inform the design process during KDViz system development. A secondary goal was to obtain first-hand experiences that may contribute to the development of a comprehensive KDViz evaluation framework. The problem with usability testing based on information retrieval tasks (locate, identify) is that the testing reports success or failure of a task but not why the user failed. Evaluation was needed of tasks requiring users to compare, associate, distinguish, rank, cluster or categorize; and task selection would be aided by development of task taxonomies (Plaisant, 2004). The criteria established for selecting the evaluation methods for this study were low cost methods that would afford evaluation at the level of elementary perceptual and cognitive tasks, be suitable for testing during the design process, and inform the design process. The study consisted of two sub-studies:

- Study I: heuristic evaluation of the system interface for manipulating the visualization.
- Study II: usability testing of the visualization design with a perceptual/cognitive task questionnaire.

4.1 Study I: Heuristic Evaluation of CiteSpace Interface

The heuristic evaluation of the CiteSpace interface (Figure 1) for the interactive graph display component was conducted by three evaluators, using Nielsen's ten heuristics and severity ratings (Nielsen & Molich, 1990) as guidelines. The evaluators had formal knowledge of user interface design principles and usability inspection methods, but were not specialists in heuristic evaluation or knowledge domain visualization. The CiteSpace interface has limited functionality, and the evaluators were given a brief orientation session on the system operations before conducting the evaluation. Evaluators were asked to test each of the menu, widget, and control operations but the total evaluation time did not need to exceed forty-five minutes. This was the first heuristic evaluation of this system, and the effort was focused on finding the most severe problems.

4.2 Study II: Usability Testing of CiteSpace KDViz Graph

The design of the usability test of the KDViz graph incorporated concepts from several standard evaluation methods. From Cognitive-Walk Through came the concept of testing in a way that would determine how intuitive the graphical representations used in the visual design were – i.e., could a user successfully decode the graph without any prior training in the form of a legend for the graph? From ethnography came the method of think-aloud protocols which were implemented in the form of the "show your work" method of explaining reasoning used in solving math problems. A task analysis of KDViz was conducted and the KDViz tasks were translated into a series of comparison tasks, Tasks T1 through T8 (Table 1). The comparison tasks have also been translated into a perceptual-cognitive task taxonomy of KDViz systems that identifies the perceptual and/or cognitive principles on which the design metaphor of CiteSpace has been based (Table 2). The tasks were administered in the form of a questionnaire using a series of marked static graphs which had been stripped of any domain specific content (See Figure 2 for example question/task).

Table 1: Task Analysis of KDViz systems.

KDViz Representation Task	Usability Test Questions	Decoding Required
T1) Explicit relationships between documents	Which words were used together more/less often?	Lines connecting circles = words that were used together. Line thickness = relative number of times words were used together.
T2) Relative position of a document within a network context	Which words are more/less likely to be about a similar topic?	Lines connecting circles = words that were used together.
T3) Significant individual documents within a large document base	Which word was used more/less often overall?	Diameter of a circle = number of times the word was used.
T4) Time of formation of research fronts	Which word began to be used earlier/later in the time period	Colors = years from the beginning (blue) to the end (red) of the time period. Color of the bands in a circle = years the word was used.
T5) Time of formation of research fronts	Which pair of words began to be used together earlier/later in the time period?	Colors = years from the beginning (blue) to the end (red) of the time period. Color of a line = first year words were used together.
T6) Persistence of clusters of document citation over time	Which group of words was used more at the beginning/end of the time period?	Diameter of a circle = number of times the word was used. Colors = years from the beginning (blue) to the end (red) of the time period. Color of the bands in a circle = years the word was used.
T7) Persistence of document citation over time	Which word was used less/more often at the start/end of the time period?	Diameter of a circle = number of times the word was used. Colors = years from the beginning (blue) to the end (red) of the time period. Color of the bands in a circle = years the word was used.
T8) Persistence of document citation over time	Which word's use decreased/increased at the end of the time period?	Diameter of a circle = number of times the word was used. Colors = years from the beginning (blue) to the end (red) of the time period. Color of the bands in a circle = years the word was used.

Table 2: A perceptual-cognitive task taxonomy of CiteSpace KDViz system.

Tasks	Test Questions	Position_Common_Scale	Position_NonAligned_Scale	Length	Direction	Angle	Area	Volume	Curvature	Shading	Color_Saturation	Proximity	Similarity	Continuity	Symmetry	Closure	Relative_Size	Figure_And_Ground
T1	Which words were used together more/less often?			X										X			X	
T2	Which words are more/less likely to be about a similar topic?											X		X				
T3	Which word was used more/less often overall?						X										X	
T4	Which word began to be used earlier/later in the time period		X								X							
T5	Which pair of words began to be used together earlier/later in the time period?										X			X				
T6	Which group of words was used more at the beginning/end of the time period?						X				X	X	X	X			X	
T7	Which word was used less/more often at the start/end of the time period?			X			X				X						X	
T8	Which word's use decreased/increased at the end of the time period?			X			X				X						X	

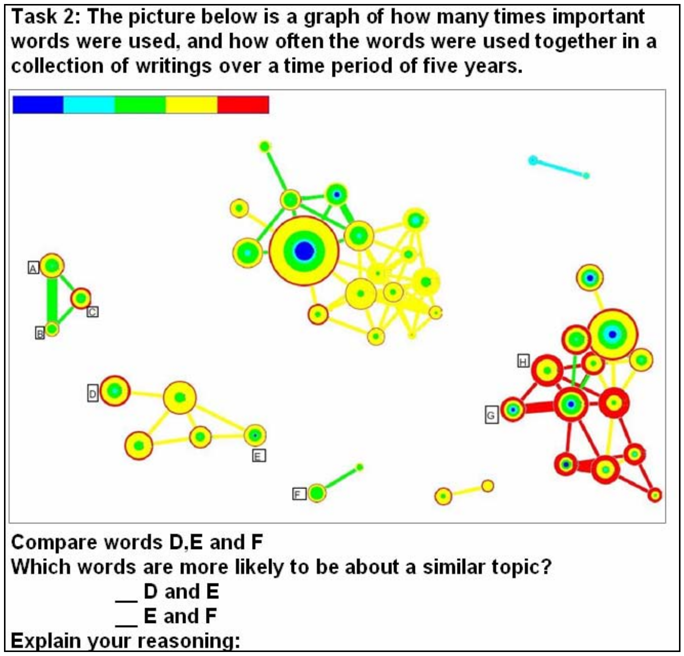

Figure 2: Task 2's instructions and the test question in Study II.

The target end-users for a system such as CiteSpace are medical researchers who are in the process of developing research proposals. The ideal usability test of CiteSpace would test both the usability and utility of the system with these target end-users in a full scale test of the full system. However for purposes of piloting this questionnaire based approach to testing the usability of the KDViz graph, a pure convenience sample was used. Participants included a seventh grade female student, a 50-60 year old male engineer, and a 30-40 year old physician-researcher specializing in epidemiology of infectious diseases. The questionnaires were administered at locations convenient to the participants (home or office), with each session taking twenty-thirty minutes. The participants were first given the questionnaire without any legend accompanying the graphs. After completing the questionnaire, a legend for decoding the graph was provided and the questions reviewed a second time for any changes in answers and reasoning process. The users were then asked if they had any comments about difficulties in interpreting the graph.

5 Results

5.1 Study I: Heuristic Evaluation Results

The heuristic evaluation by three evaluators identified twenty-two unique problems with violations of nine of Nielsen's ten heuristics (Table 3).

Table 3: Results of Heuristic Evaluation by Heuristic.

	Heuristic	Problems Found	Severity
H1	Visibility of system status	8	+++
H2	Match between system and the real world	5	++
H3	User control and freedom	4	+++
H4	Consistency and standards	5	++
H5	Error prevention	3	++++
H6	Recognition rather than recall	4	++
H7	Flexibility and efficiency of use	1	++
H8	Aesthetic and minimalist design	3	+++
H9	Help users recognize, diagnose, and recover from errors	0	
H10	Help and documentation	3	++++

However the missing heuristic (H9 – Help users recognize, diagnose, and recover from errors) would also fit problems coded under other violations. Twelve problems were classified as having a severity rating of 3-4 (Table 4).

Table 4: Results of Heuristic Evaluation by Problems identified.

Problem #	Location	Problem Description	# of Evaluators	Avg. Severity Rating	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
1	Controls	Control information hidden in Help system	2	2						2				
2	Controls	Zoom control doesn't work as expected	1	2		1								
3	Controls	Pan control is backwards	1	3		1		1						
4	Controls	Marquee selection causes graphics freeze	2	4	1				1					
5	Controls	Zoom can cause graph to disappear	1	4					1					
6	Frames	Frame resizing can truncate panels	1	1		1								
7	Graph	Mouse click on graph results in node moving	2	1.7		1	1	1						
8	Graph	Stopping layout from running causes graph to jump up on screen	1	1				1						
9	Graph	No rollover – hard to tell where mouse is on tightly packed graph	1	2	1									
10	Graph	Legend moves out of sight	1	4								1		
12	Help	Help is minimalist	2	3								1		2
11	Help/Legend	Meaning of legend is unclear	3	3.	1					1				1
13	Menus	Choosing visualization/run wipes out any manual rearranging of graph	1	3			1							
14	Menus	File menu should have ellipses to indicate sub-menus	1	1				1						
15	Menus	Redundant error message when canceling File/open	1	2								1		
16	Menus	Aborting File/Open can cause system crash, freezes if tries floppy drive	1	4			1							
17	Multi	No support for error recovery with exception of exiting and starting over	1	3			1							
18	Widget	Labels slider seems inoperative or scaled poorly	2	2.5	2									
19	Widget	"Stress" is undefined, and consequences of adjusting not clear	1	1						1				
20	Widget	Meaning of changing circle size unclear	2	3	2									
21	Widget	Adjustments to stress slider do not take effect until new visualization started	1	3	1									
22	General	Language and tools of system would be foreign to average user	1	2.5		1		1	1		1			

Among the twenty-two identified problems, four problems #4, #5, #10, and #16 have the average severity rating of 4. The nature of these problems is briefly discussed in the following.

Problem #4, that the marquee selection caused graphics freeze, falls in the category of the lack of the visibility of system status. When the marquee selection was initiated, CiteSpace immediately launches a batch search to PubMed on the Internet without notifying this process to the user. If a large number of nodes were selected by this function, the waiting time could be as long as a minute, especially if the connection to the Internet is slow.

Problem #5, that zoom can cause graph to disappear, is similar to a focus+context problem. Currently, CiteSpace only provides a local view. If the user zooms in, the whole display area could land on an area with not objects. A global view would resolve this problem.

Problem #10, that the legend moves out of sight, is in part related to Problem #5. In addition, the legend is currently drawn as part of the graph; its position is affected by zooming. A simple solution is to separate the legend from the content display.

Problem #16, File Open can cause system crash, is due to a known bug in Java SDK. The following instruction was documented on CiteSpace's homepage:"javaw.exe - No Disk". This is a known bug in Java SDK. Ignore it and press "continue" to proceed. The problem of this kind can be expected to be resolved in more recent releases of Java.

Problems with less severity could be also informative on evaluating the design. For example, Problem #18, that label sliders seem inoperative or scaled poorly, is due to the fact that it is often necessary to coordinate both the label threshold slider and the label size slider in order to control the display of labels.

In summary, the relatively low-cost heuristic evaluation identified a wide range of problems. The most severe ones tend to echo the calls for clear conceptual models by Norman in his Invisible Computers (1997).

Table 5: Results of Usability Testing.

Task	Questions	Decoding Required	Errors w/out Legend	Errors w/ Legend	Total Errors
T1	Which words were used together more/less often?	The lines connecting circles show words that were used together. The line thickness shows the relative number of times words were used together.	2	0	2
T2	Which words are more/less likely to be about a similar topic?	The lines connecting circles show words that were used together.	1	0	1
T3	Which word was used more/less often overall?	The diameter of a circle equals the number of times the word was used.	2	0	2
T4	Which word began to be used earlier/later in the time period	The colors represent the years from the beginning (blue) to the end (red) of the time period. The color of the bands in a circle represents the years the word was used.	6	5	11
T5	Which pair of words began to be used together earlier/later in the time period?	The colors represent the years from the beginning (blue) to the end (red) of the time period. The color of a line represents the first year words were used together.	5	2	7
T6	Which group of words was used more at the beginning/end of the time period?	The diameter of a circle equals the number of times the word was used. The colors represent the years from the beginning (blue) to the end (red) of the time period. The color of the bands in a circle represents the years the word was used	4	4	8
T7	Which word was used less/more often at the start/end of the time period?	The diameter of a circle equals the number of times the word was used. The colors represent the years from the beginning (blue) to the end (red) of the time period. The color of the bands in a circle represents the years the word was used	2	1	3
T8	Which word's use decreased/increased at the end of the time period?	The diameter of a circle equals the number of times the word was used. The colors represent the years from the beginning (blue) to the end (red) of the time period. The color of the bands in a circle represents the years the word was used	0	0	0
	Total errors w/out and with legend		22	12	34

5.2 Study II: Usability Testing Results

The usability questionnaires were scored based on two kinds of errors. An error could be made if the user's answer to a question did not match the expected (or "correct") answer, or an error could also be scored if the reasoning the user gave indicated they were not decoding the graph as desired. Points were assigned for both kinds of errors (Table 5, above). Tasks with the lowest errors (T1, T2, T3, and T7) involved primarily comparisons of line width, circle area, or band width as Cleveland and McGill's (1984) theory would predict. In terms of the total errors, the most error-prone task was T4, followed by T6 and T5. These tasks require color-to-time decoding. There were fewer errors during the second pass through the questions using the legend. A common theme among the explanations from the users was errors related to equating colors with topics or number of colors with frequency of word use. The user who had the greatest difficulty with decoding the graph was the physician-researcher, who commented that he expected time to be oriented left-to-right, and did not understand why change over time was only represented within words and not within the connections between words.

6 Discussion and Conclusions

The findings of the first study, heuristic evaluation, with slightly more than half of the identified problems having a severity rating of 3-4 indicate that the CiteSpace system interface should be modified before additional usability testing with the targeted end-users. Several interface design changes were indicated by the evaluation which had not been previously noticed as a problem by the designer, such as putting the color legend in a separate frame so it would not go out of site during zooming and panning. Further evaluation with double specialists having formal knowledge of interface design and visualization design principles would be expected to improve the quality of findings from a heuristic evaluation (Nielsen, 1992).

The usability testing, Study II, was informative in two respects. The findings of relative ease of tasks fit with a-priori expectations based on theoretical perceptual and cognitive principles underlying the design. The most error-prone tasks required color-to-time decoding and a test with a larger sample size would be desirable to investigate underlying factors. In addition, the difficulties the most highly educated user, the physician, had in decoding the graph may be due to prior background knowledge of biostatistical graphs which represent change over time in a left-to-right fashion. This may lead to further studies of the impact of background knowledge on interpreting visualizations, and lead to further design changes.

This pilot study of a questionnaire based approach to evaluating the usability of visualizations was successful, especially when considering the low-cost process lended by heuristic evaluation method. Given the information-cost ratio, we would recommend this method to be used for KDViz systems in general to establish a good understanding of what design features may hinder the comprehension of knowledge trends represented by such systems. Tasks related to pre-attentiveness need to be incorporated into the taxonomy, and the collection of completion time for each task would help empirically evaluate the relative difficulty of tasks related to decoding the visualization.

References

Chen, C. (2004a). *Information visualization: Beyond the horizon (2nd ed.)*. Heidelberg: Springer.

Chen, C. (2004b). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 101*(1), 5303-5310.

Chen, C. & Czerwinski, M. (2000). Empirical evaluation of information visualization: An introduction. *International Journal of Human-Computer Studies*, *53*(5), 631-635.

Chen, C. & Paul, R. J. (2001). Visualizing a knowledge domain's intellectual structure. *IEEE Computer*, *34*(3), 65-71.

Chen, C. & Yue, Y. (2000). Empirical studies of information visualization: A meta-analysis. *International Journal of Human-Computer Studies*, *53*(5), 851-866.

Cleveland, W. S. & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*, 531-554.

Eades, P. (1984). A heuristic for graph drawing. *Congressus Numerantium*, *42*, 149-160.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software - Practice and Experience*, *21*, 1129-1164.

Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt Brace.

Kosara, R., Healey, C.G., Interrante, V., Laidlaw, D.H., & Ware, C. (2003). Thoughts on user studies: Why, how, and when? *IEEE Computer Graphics and Applications*, *23*(4), 20-25.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation, *Proceedings of the SIGCHI conference on Human factors in computing systems*, 373-380.

Nielsen, J. & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 249-256.

Norman, D. A. (1997). *The invisible computer*. Cambridge, MA: MIT Press.

Plaisant, C. (2004). The challenge of information visualization evaluation. *Proceedings of the working conference on Advanced visual interfaces*, 109-116.

Trafton, J. G., Marshall, S. P., Mintz, F, & Trickett, S.B. (2002). Extracting explicit and implicit information from complex visualizations. *Diagrams*, 206-220.

Ware, C. (2000). *Information visualization: Perception for design*. San Francisco: Morgan Kaufmann Publishers.