Intended Use Evaluation Approach for Information Visualization

Albert Park

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Computer Science

Dr. Shawn Bohner
Dr. James Arthur
Dr. Denis Gracanin

November 28, 2006
Blacksburg, Virginia

Keywords: Intelligence Analysis, Domain evaluation, Intended Use, Evaluation, Information Visualization, Visualization

Intended Use Evaluation Approach for Information Visualization

Albert Park

ABSTRACT

Information visualization is applied in many fields to gain faster insights with lighter user cognitive loads in analyzing large sets of data. As more products are being introduced each year, how can one select the most effective tool or representation form for the task? There are a number of information visualization evaluation methods currently available. However, these evaluation methods are often limited by the appropriateness of the tool for a given domain since they are not evaluating according to tools' intended use. Current methods conduct evaluations in a laboratory environment with "benchmark" tasks and often with field data sets not aligned with the intended use of the tools. The absence of realistic data sets and routine tests reduces the effectiveness of the evaluation in terms of the appropriateness of the tool for a given domain. Intended use evaluation approach captures the key activities that will use the visual technology to calibrate the evaluation criteria toward these first-order needs. This research thesis presents the results from an investigation into an intended use evaluation approach and its effectiveness of measuring domain specific information visualization tools.

In investigating the evaluation approach, criteria for the intelligence analysis community have been developed for demonstration purposes. While the observations from this research are compelling for the intelligence community, the principles of the evaluation approach should apply to a wider range of visualization technologies. All the design rationale and processes were captured in this thesis. This thesis presents a design process of developing criteria and measuring five intelligence analysis visual analytic tools. The study suggests that in selecting and/or evaluating visual analytic tools, a little up front effort to analyze key activities regarding the domain field will be beneficial. Such analysis can substantially reduce evaluation time and necessary effort throughout a longer period of time.

Acknowledgments

Now at the end of one road, I would like to take the time to thank those who have helped me along the way.

This work would not have been possible without the frequent guidance, insightful comments, and suggestions of Dr. Shawn Bohner.

I would also like to thank Dr. Nicholas Stone, Jim Venuto, Dr. Ben Goertzel, Dr. Michael Miller, Dr. Debbie Duong, Dr. Eunice Santos, and Dr. Edward Fox for the opportunity and guidance throughout the Research and Development Experimental Collaboration Applied Research Center Project conducted with SAIC.

Thanks also to the members of my committee, Dr. James Arthur and Dr. Denis Gracanin.

I thank my mother for her love and support, my father for the encouragement and motivation, my sister for her constant care, and other members of my family for being there.

Also, I would like to thank Danny, Mr. & Ms. Ha, Jung, Yoonbum, and SY for the extraordinary and memorable s and for making Blacksburg Home.

Lastly, I thank the Lord for all wonderfulness in life.

Table of Contents

Table of Tables and Figures

Chapter 1. Introduction

Information visualization helps users with various backgrounds and knowledge to analyze large amounts of data faster to gain relevant insights. Ever since the idea of visualizing information was first introduced, many information visualization tools have appeared in the research community and have been produced in the marketplace. As information visualization technology evolves, perhaps a tool may emerge that will go beyond visualizing data to producing insightful reports in real-time. Until then users must analyze, interpret, and generate their own conclusions/results. Their own conclusions are based on their perspective of the information under observation. Hence, knowing the activities where visualization will be applied offers some leverage on evaluating these tools.

Today, there are many products that are designed to visualize large corpuses of data [1]. With this overwhelming number of visualization tools and methods, how can one select the most effective tool or representation form for the task? Even though there are so many visualization tools out there, no one tool can out-perform every other tool for every single task. At the same time, depending on the domain and/or type of tasks, there are visualization tools that dominate the given area. How do we identify these tools? How do we know which tools are the best to use for a given situation? Even going back to the researching and designing stage, how do we check if a research tool is headed in the right direction? How do we minimize evaluation cost and determine a tool's potential and/or limitation for the job? There are so many questions regarding visualization technologies (tools and methods) and their evaluation. While this thesis will briefly examine each of these questions, it will mainly focus on evaluating information visual analytic tools from an intelligence analyst perspective.

Evaluation and tool selection are closely related. To find the tool best suited for a task, we have to evaluate different tools and measure them. Just as there are many choices for choosing a tool, there are several methods currently in practice, to conduct an evaluation. How do we evaluate more effectively? Current evaluation methods are organized in different ways. Usability testing and controlled experiments remain the backbone of evaluation [2]. Each information visualization tool was developed with a different purpose. Hence, an evaluation should be performed according to the intended use.

Information visualization tools are used in situations anywhere from an aid to help children's education to serving intelligence analysts with national security matters. Even though the range varies greatly, it can be narrowed down to an application's domain since many intentions are to serve a single domain. Identifying the application's domains and their main in-domain tasks will help guide and evaluate information visualization tools according to their intended use.

In this research a proposed visualization technology evaluation approach is investigated that can deliver relatively accurate and valuable information for selecting an effective tool in a timely manner. A combination of different evaluation methods will serve this purpose. Metrics evaluation is the core structure which is used to determine basic needs, deliberation, and features of a tool.

Goals and Scope

The general goal of this thesis is to investigate ways to improve information visualization evaluation, which can deliver relatively accurate and relevant information in a more efficient manner. This work stems from research conducted for the Research and Development Experimental Collaboration Program with SAIC. The goals in this research build on the overall goal to effectively evaluate visualization technologies for the intelligence community. The key objectives are:

1. Reduce the time that is required to evaluate a visualization technology.
2. Ensure that the relative accuracy remains at an acceptable level.
3. Better target domain users' key tasks in visualization technology evaluations.

The general scope of this thesis is stated as following:

1. Focusing largely on information visual analytic tools recognizing that the principles may generalize to the wider population of visualization technologies.
2. Demonstration of the proposed framework within the confines of tools available using realistic data.
3. Selected the intelligence analyst domain for demonstrating the approach.

Problem Statement and Hypotheses

The problem that this thesis investigates is the often protracted amount of time needed to evaluate visualization technologies. The purpose is to examine ways that information visualization tools can be evaluated according to their intended use. While this is by no means new to technology evaluation, it is relatively unique in evaluation of visual analytic tools. This evaluation technique is being demonstrated in the intelligence analysis domain, while recognizing that the concepts may convey to the wider scope of visualization technologies. The concepts and lessons gained may apply especially well to other analytical environments, such as business analytics and scientific visualization, because they share similar activities. In investigating how to evaluate information visualization analytic software, the focus is on exploring key tasks that are performed regularly by analysts.

The **hypothesis** of this research effort is: *if visualization technologies can be evaluated according to their intended use, then a more accurate and timely selection of the appropriate tool can be made that illumines its potential strengths and limitations*. Intended use corresponds to a set of tasks that make up the core activities that are best supported by visualization in the operational environment. The set of tasks will be based on the key activities that the user finds most important and/or most frequently faced.

Summary of Technical Approach

In evaluating any tool, there are a set of activities that are related to preparation (criteria selection, planning, etc.), data collection for the long-list of products, organization, early selection of tools for the short-list, the acquisition of the selected tools, tailoring of the evaluation, conducting the evaluation, validating the results, and ultimately the final selection of relevant tools based on the criteria. This is little difference for the visualization technologies that one would endeavor to evaluate. However, with there being so many different tools available for so many visualization applications, getting to the short-list of tools can be daunting. [3] Hence, the approach taken in this research has been to develop an early conditioning mechanism that would focus the evaluation on tools that have a "best fit" for the visualization activities in the domain where they would be applied. The following list of steps summarizes some of these tasks that one partakes in to evaluate visualization technology:

- Establish a visualization technology evaluation framework based on existing work in the visualization community, to expedite and clarify the evaluation criteria early on.
- Investigate a relevant application domain.
- Identify domain users' key tasks where visualization will provide support and accelerate the completion of the tasks.
- Select existing criteria from the evaluation framework and develop any new criteria that are not well represented for the selected domain.
- Prototype a number of visual analytic tool evaluations, which reflects the evaluation approach.
- Detail the experimental investigation and the results.
- Identify the revealed lessons and key findings from experiments and results.
- Report these findings in this thesis.

Overview of This Thesis

The remainder of this thesis is structured as follows: Chapter 2 describes the background (the strengths and weaknesses of existing evaluation methods) and motivation behind this work. Chapter 3 outlines the existing user tasks taxonomy of information visualization to clarify current practices in the visualization community. Chapter 4 follows up with existing criteria in the visualization community, to expedite the evaluation criteria early on. Chapter 5 investigates the application domain, intelligence analysis, and analyzes analysts' real environment activities and the criteria to measure the activities. Chapter 6 follows up with current information visualization methods and tools, and then evaluates five tools using the criteria developed in Chapter 5. Chapter 7 provides a discussion of the strengths and weaknesses of this approach, revealed lessons, and contributions. Finally, Chapter 8 concludes this thesis and provides a number of possible future works.

Chapter 2. Related Work

In this section, some of the previous efforts in the related field of information visualization evaluation methods are examined. It discusses previous research closely related to this thesis in some detail to give a perspective on how the approach taken compares. Then this section describes how this work builds on and is different from these previous works.

Previous Information Visualization Evaluation Methods

A large number of evaluations have been conducted measuring the effectiveness of information visualization tools using the following methods:

1. **Controlled experiments evaluation -** This is a short-term evaluation, which is being conducted through a number of different controlled experiments. These controlled experiments typically measure the time to perform the task and quality of user solution [4]. In these reports, typically, the study subjects are asked to perform a list of tasks, while the evaluator observes and records the performance time. Independent variables may vary, but they are normally the control aspects of the tools, such as tasks, data and participant classes. The dependent variables usually include accuracy and efficiency measures. In general, accuracy measures include precision, error rate, and the average number of incorrect answers; whereas efficiency measures include the average time to complete the task(s) [2].

   A large number of controlled experiments have been conducted [2] [5]. A few examples of controlled experiment evaluations are: spatial ability and visual navigation evaluation [6], tree visualization system for knowledge discovery evaluation [7], and interface and data architecture for query preview in networked information systems evaluation [8].

   The strength of this method is that it can accurately measure the accuracy, efficiency, and other properties of a tool. Also, this method can be used to check the direction of a research tool to see if it is serving the needs of the developing purpose. In current practice, controlled experiments evaluation helped researchers to compare a newly introduced tool with already existing tools, which are considered to be the standard tools or state of art technology tools [9] [10].

However, there are still needs for improving this evaluation method in areas such as task analysis, usability evaluation and usage analysis [5]. Tasks need to be unified and the results across existing studies need to be synthesized [2] [5]. The absence of realistic data sets and routine tests reduces the effectiveness of the evaluation in terms of the appropriateness of the tool for a given domain. In the intended use evaluation approach, the tasks are calibrated according to the relevant application domain.

2. **Usability evaluation –** Usability evaluations are the two most common evaluations along with controlled evaluations [2]. The approach is also considered as a short-term evaluation, which typically focuses on usability issues concerning the tool interfaces. This type of evaluation identifies problems encountered, provides feedback on those encountered problems, and helps to design a new solution. Usability evaluation methods typically involve a usability survey or a "think aloud" protocol as an evaluator observes participants performing tasks with a tool or just simply using a tool. In these evaluations, the study subjects are important because a lot of outcome depends on the ability of the participants. Usually a background survey is conducted to either screen out participants and/or accommodate into the result. Some examples of usability evaluations are tree visualization systems [11] and usability evaluations of Bifocal Browser [12].

   Specially conditioned usability evaluations, such as attention-limited environments usability evaluation also have been conducted [13]. The strength of usability evaluation is that it values users' opinions. With its unique method, such as “think aloud” and survey, it is easier to spot the user interaction problems compared to some other evaluation methods. The weakness of usability evaluation is that results may be subjective to the background knowledge, experience, and ability of experiment participants. Therefore with usability, a large number of participants are preferred, which can result in a greater expense in money and time. Despite its weakness, evaluating the usability of a tool continues to be considered important.

Often we see people prefer a better usability over better performance. Also, it is not unusual to witness a modest performance and great usability tool out-perform a great performing tool with a poor usability. It is because the user of the tool still remains the overall performing factor. In the intended use evaluation approach, issues about tool interface will be measured through metrics.

3. **Comparing two or more tools –** This is another short-term evaluation usually conducted with a controlled experiment. Although there are evaluations conducted with different methods, such as usability evaluation [11], a controlled experiment is the most common method used. Controlled experiments are the most used method, because results of the experiments are easy to compare. The approach typically compares a number of tools and their respective features and design elements. Most of the time, the evaluation compares the novel technology with the standard or the state of the art technology. Some examples include: comparing SpaceTree to Microsoft explorer and hyperbolic tree browser [9], comparing different three-dimensional information visualization designs [10], comparing text, 2D, and 3D interfaces [14], comparing ExoVis, Orientation Icon with In-Place 3D visualization techniques [15], and comparing TableLens (a.k.a. Eureka), InforZoom (formerly Focus) and Spotfire [16]. The strength of this approach is that the evaluation method can highlight the improvements and the differences of the comparing tools. Also, along the way the approach measures property, accuracy, and efficiency of the tools. The weaknesses of the method are that it requires having the “standard” tool to

compare against. Just as with the controlled experiment evaluation method, this evaluation approach needs to be improved in task analysis, usability evaluation, and usage analysis [5]. The tasks that are being used in the experiment need to be unified [2] [5]. The result of this method informs the performance regarding the routine tasks rather than effectiveness of the evaluation in terms of the appropriateness of the tool for a given domain/job. The intended use evaluation approach will ensure the reliability of the appropriateness of the tools for a given domain/job by selecting controlled experiment tasks based on tool' intended use.

4. **Metrics evaluation** - Metrics evaluation is another short-term evaluation, which measures the effectiveness of a tool. The approach usually evaluates information visualization tools according to the design and evaluation criteria. Some examples of criteria are: expressiveness and effectiveness criteria [17], information content of visual displays criteria [18], variety of other criteria for representation and interaction [12], domain-independent to specific domain criteria [19], high-level variable type criteria [20], and design principle (example) [21]. Some examples of metrics evaluation are: Glass box [22], concept demonstration metric [23], multidimensional information visualization techniques [24], problem-oriented classification of visualization techniques [25], and scenario-based tasks [26]. The whole process of metrics evaluation is typically time efficient, while it still accurately assesses and covers details of the tools. However, there is still a need for improving methods in areas such as synthesizing and unifying taxonomy of users' tasks, including interaction mechanisms evaluation. Also, tasks need to be unified and the results across existing studies need to be synthesized [2] [5]. With the intended use evaluation approach, in order to evaluate interaction mechanisms properly, the criteria were selected based on users' key tasks for a given domain.

5. **Heuristic evaluation –** This short-term evaluation method is similar to usability

evaluation but relies on expert judgment codified in heuristics. Heuristic evaluation here is a method for finding usability problems in a user interface design by having a small set of evaluators examine the interface and judge its compliance with recognized usability principles. It is different from usability evaluation in that it is less formal and intended as a “discount usability engineering” method [27]. The strength of heuristic evaluation is that it values expert point of usability and it effectively and efficiently evaluates interaction mechanism aspects of a tool. Also, like other short-term evaluation, it is time efficient. The weakness of this approach is that results may be subjective. According to J. Nielsen and R. Molich, the result can be varied by an evaluator’s ability to find problems [28]. In the intended use evaluation approach, interactive mechanism is measured through metrics.

6. **Insight-based evaluation –** This is a newly introduced evaluation approach in recent years. An empirical study measuring insight introduced by Saraiya [29] [30] captured the essence of the main purpose in using visualization, to gain insight. An insight-based evaluation method allows evaluators to quantify insight using different characteristics [29]. This evaluation approach seems promising, but only short-term studies have been applied to measure the effectiveness of the evaluation approach. Authors of this evaluation approach themselves explicitly expressed the desire of conducting a longitudinal study over a longer period of time to obtain valuable conclusions.

7. **Field studies/longitudinal studies –** Field studies/longitudinal studies help to discover the effectiveness, problems, and actual use of a tool in a real daily work environment over a longer period of time. A report written by administrative data analysts shows the need for conducting a longitudinal study on information visualization systems. The report indicates the difficulty of actual integration of the visualization system into a daily work practices, and it expressed the need of

redesigning the system [31]. As this example illustrates, the field/longitudinal studies can discover the over-looked matters in short-term studies. Short-term experiments are conducted under controlled environments under a given period of time. Since some analysis and insight can be gained by looking at the visual representation from different perspectives over a long period of time, a short-term evaluation may neglect this aspect of tools.

Often complex data, such as biology data, need a longer period of time to analyze [3]; it may not be accurate to evaluate with short-term evaluations. The main strength of field/longitudinal studies is that it responds to a key short-coming of short-term experiments and provides most accurate information. The short-term studies, such as controlled evaluation, are limited in a sense that tasks are benchmark tasks or predefined tasks; data sets are often non-related to the real work situation. However, this approach can be lengthy and often expensive. *A Field study of Exploratory Learning Strategies* is an example of field/longitudinal study on information visualization tool, Eureka [32]. In the intended use evaluation approach, the strength of field/longitudinal studies has been adopted, while minimizing cost. The approach embraces the strength by analyzing the most important or frequently faced activities in the beginning of the process. While the tasks are calibrated toward intended use, the evaluation format stays as short-term evaluation, which can reduce the cost and time to conduct the evaluation.

8. **Models-** Cognitive architecture to evaluate visualization applications, CAEVA, uses cognitive models to evaluate and design visualization. Models, such as cognitive models, allow users to simulate the use of visualization while they permit researchers to understand the user action regarding attention control, solution strategies, memory failures, memory limitations and learning [33]. The key strength of this evaluation is that it allows researchers to obtain users’ thought processes at a relatively low cost. However, little testing has been conducted and developing a model can be lengthy.

To get a sense of how these approaches compare, the summary of the current evaluation methods is presented below in Table 2.1.

	Duration	Major Strength	Limitation/Weakness
Controlled experiments evaluation	Short-term	- effectively measures efficiency, accuracy, and other properties. - can be used to check the direction of a research tool to see if it is serving the needs.	- hard to assess usability evaluation, task and usage analysis [5]. - tasks must be unified and the results across existing studies need to be synthesized [2] [5]. - absence of realistic data sets and routine tests reduces effectiveness .
Usability evaluation	Short-term	- it values users' opinions. - unique method, such as "think aloud" and survey - easier to spot the problems compared to some other evaluation methods.	- results may be subjective based on the background knowledge, experience, and ability of experiment participants. - requires large number of participants; hence, more money and time.
Comparing two or more tools	Short-term	- it highlights the improvements and the differences of the comparing tools. - effectively measures efficiency, accuracy, and other properties.	- it requires having the "standard" tool
Metrics evaluation	Short-term	- it accurately assesses and covers details of the tools in a timely manner.	- tasks must be unified and the results across existing studies need to be synthesized [2] [5].
Heuristic evaluation	Short-term	- it uses experts' judgments	- results may be subjective [28].
Insight-based evaluation	Short-term	- allows evaluators to quantify insight using different characteristics [29].	- Only a short-term study has been applied to measure the effectiveness of the evaluation approach.
Field studies/longitudinal studies	Long-term	- responds to a key short-coming of short-term experiments - the most accurate assessment.	- often lengthy and expensive.
Models	Long-term	- obtains users' thought processes at a relatively low cost.	- not much testing has been conducted - developing a model can often be lengthy.

Table 2.1 Summary of Current Evaluation Practices

As can be seen in the table above, current evaluation methods are organized in a variety of ways. Usability evaluation and controlled experiments evaluation, the backbone of evaluation [2], can provide insight into the usability, accuracy, efficiency and other properties of a tool in a timely manner. However, the big downside to these evaluations is the uncertainty of actual integration of the visualization tool into the real work setting [31]. The short-term evaluations, including controlled experiments evaluation and usability evaluation, are limited in terms of the appropriateness of the tool. They are conducted in a laboratory environment with "benchmark" tasks and often unrelated field data sets. The absence of realistic data sets and routine tests reduces the effectiveness of the evaluation in terms of the appropriateness of the tool for a given domain. In this evaluation approach, the strength of field studies/longitudinal studies will be embraced, while keeping the approach in timely manner. To establish an evaluation framework, Chapter 3 and 4 illustrates existing user tasks taxonomy of information visualization and evaluation criteria in the visualization community.

Chapter 3. User Tasks of Information Visualization

This section discusses some existing information visualization taxonomies in the visualization community. This process provides the current information visualization tools' capability, expected tasks, performances and usages in the community. Researching the current practices provides the knowledge to properly measure the key activities in this evaluation approach.

The Task by Data Taxonomy

One of the early works by Ben Shneiderman [21], "Task by Data Taxonomy," has repeatedly contributed as the backbone of the user taxonomy and evaluation criteria over the years. "Task by Data Taxonomy" offers a strategy, which is formed with seven data types and seven user data types. The mutually exclusive data types are one-dimensional data, two-dimensional data, three-dimensional data, multi-dimensional data, temporal data, tree data and network data. The seven tasks are created based on Shneiderman's well-known visual information seeking mantra: "Overview first, zoom and filter, then details-on-demand." The following are the seven tasks:

- **Overview –** Users have the ability to overview of the entire data set.
- **Zoom –** Users have the ability to zoom-in on chosen data entry.
- **Filter –** Users have the ability to filter out chosen/uninteresting data.
- **Details-on-demand –** Users have access to view the details of selected data entry or entries.
- **Relate –** Users have the ability to view relationships among data entries.
- **History –** The system keeps a history of actions to support undo, replay, and progressive refinement.
- **Extract –** Users are allowed to save/extract of sub-collections and of the query

parameters in a format that would facilitate with other uses. (Taken from: [21])

These seven tasks are basic but frequently used tasks that have became an essential functionality of information visualization tools. "Task by Data Taxonomy" is closely related in the navigation and retrieving information features of a tool. Some of these taxonomy were included in the intended use approach to evaluate the navigate feature of a system.

Problem-Oriented Classification of Visualization Techniques

Another user task taxonomy that has been noted by many others is "A Problem-Oriented Classification of Visualization Techniques" by Stephen Wehrend and Clayton Lewis [25]. According to the authors, there are 11 operation classes, visualization techniques. The visualization techniques are the following:

- **Identify** – This interaction technique is related to finding the description to be able to understand an object that was not previously known.
- **Locate –** This interaction technique is related to finding desired data entries. This feature can also function like filtering and highlighting.
- **Distinguish** – This task describes when users distinguish between different values of the same variable. For example, users want to recognize which objects have already been identified or interacted with and which objects have not.
- **Categorize –** This task applies when users are categorizing displayed objects into a defined setting.
- **Cluster –** This task applies when users are defining whether the selected data entries are clustered or not.
- **Distribution –** This task applies when users are describing the overall pattern of a selected area.
- **Rank –** This technique is only for scalar and ordinal data. It applies when users

are sorting the data according to their values.

- **Compare between relations –** This task applies when users compare different entities that are used as the basis of comparison.
- **Compare within entities –** This task describes tasks where users are deciding something based on the attributes of similar objects.
- **Associate –** This technique applies when users form a relationship between objects in a display.
- **Correlate –** This technique applies when users recognize a selected object's other attributes, if the object has multiple attributes and is presented in one representation form. (Noted from: [19], [25])

These 11 visualization techniques/user actions are low-level and domain-independent. Because it is domain-independent, it provides a generalized information visualization framework. Some of these taxonomies were included in intended use approach to evaluate navigating and theorizing.

Visual Task Taxonomy

"Visual Task Taxonomy" [34] developed by Michelle Zhou and Steven Feiner, is an extended work of Wehrend and Lewis. "Visual Task Taxonomy" adds a few additional visual tasks, which work as the interface of high-level presentation intents with low-level visual techniques.

- **Associate** - This task corresponds to collecting, connecting, uniting, and attaching of relationships between objects in a display.
- **Background –** This task indicates acknowledging a displayed object's background.
- **Categorize -** This task corresponds to categorizing depending on similarity.

- **Cluster** - This task corresponds to obtaining a group's outline.
- **Compare** - This task corresponds to differentiating or finding an intersecting point of groups or individuals.
- **Correlate** - This technique applies when users recognize a selected object's other attributes, if the object has multiple attributes and is presented in one representation form.
- **Distinguish** - This task describes when users distinguish or isolate between different values of the same variable. For example, users want to recognize which objects have already been identified or interacted with and which objects have not.
- **Emphasize** - This task corresponds any action of modifying, isolating or focusing on an object(s).
- **Generalize** - This task describes when users merge objects.
- **Identify** - This interaction technique is related to finding the description, such as name, portray, individualize, and profile, to be able to understand an object that was not previously known.
- **Locate** - This interaction technique is related to finding desired data entries' position, situating, pinpoint, and outline.
- **Rank** - This technique applies to when users are sorting the data according to their values, such as time.
- **Reveal –** This task corresponds to any action of exposing, itemizing, specifying, or separating objects in a display.
- **Switch –** This task is related to transforming an object.
- **Encode** - This task corresponds signaling action of the following property, such as label, portray, plot, structure, trace, and map of an object(s). (Noted from: [19], [34])

There are a total of 15 visual tasks, some of which have refinements. Refinements are available if the tasks need to be interpreted differently. Each task is described by two parts, which are omitted in this thesis. The two parts are an act and a set of arguments to act on.

E. Morse and M. Lewis Visual Taxonomy

E. Morse and M. Lewis developed a visual taxonomy [19] to evaluate information retrieval (IR) based on "Visual Task Taxonomy" by Zhou and Feiner and "A Problem-Oriented Classification of Visualization Techniques" by Wehrend and Lewis. The main difference from the two is that E. Morse and M. Lewis' work is domain-specific, evaluating IR systems. According to the authors, the full taxonomy contains approximately 50 tasks, but they were not explicitly revealed.

Taxonomy of Tasks for Guiding the Evaluation of Multidimensional Visualizations

This is one of the most recent taxonomies for multidimensional visualizations developed by Eliane R. A. Valiati et al. [35]. The taxonomy focuses on the exploiting process of a large dataset. This work is based on many existing taxonomies [25] [34] [19]. This work integrates the existing taxonomies with new exploratory tasks in a multidimensional dataset.

- **Identify –** This task corresponds to any action of finding, discovering or estimating visually on the following: clusters, correlations, values, dispersion, distribution, patterns thresholds, similarities or differences, data dependency or independency, uncertainty and/or data variation.
- **Determine –** This task corresponds to any action of calculating, defining, or indicating the following: mean, median, variance, standard deviation, amplitude, percentile, sum, proportions, differences, correlation, coefficients, probabilities, and other statistics.

- **Visualize –** This task corresponds to a system's capability to visualize in following areas: dimensions, data items, and domain space.
- **Compare –** This task corresponds to any action of comparing selected data by analyzing dimensions, data items, values, clusters, properties, proportions, locations, distances, and visual characteristics.
- **Infer –** This task corresponds to any action of process of inferring knowledge from information defining hypotheses, rules, probabilities, trends, or characteristics.
- **Configure –** This task corresponds to any support tasks that need analytic tasks.
- **Locate –** This task corresponds to any action of searching and finding of already visualized information, such as: items, values, clusters, distances, proprieties, and visual characteristics. (Taken from: [35])

The taxonomy concentrates on visual techniques for exploratory analysis or for supporting or preceding a more conventional statistical analysis [35]. The exploratory analysis is one of the most frequently faced activities for intelligence analysts [36]. Navigation related tasks have been noted in developing the intended use evaluation approach.

Taxonomy of User's Tasks for Interaction with Multidimensional Visualization Systems

According to Pillat et al. [24], there are seven user tasks: identify, determine, visualize, compare, infer, configure, and locate in evaluating information visualization systems. The first five tasks (identify, determine, compare, infer, and locate) are related to exploring or analyzing the data set and the last two tasks are intermediate level tasks that support the analytical ones.

- **Identify –** This task corresponds to any action of finding, discovering or estimating the following: clusters, correlations, categories, properties, patterns, characteristics, thresholds, similarities/differences, dependencies/independencies, and uncertainties/variations.

- **Determine –** This task corresponds to any action of calculating or defining the following: mean, median, variance, standard deviation, amplitude, percentile, sum, proportions, differences, correlation, coefficients, probabilities, and other statistics.
- **Compare –** This task corresponds to any action of comparing two or more selected data. The following can be compared: dimensions, items, data, values, clusters, proprieties, proportions, positions/location, distances, and graphical primitives.
- **Infer –** This task corresponds to any action of process of analyzing or defining: hypotheses, rules, trends, probabilities, and cause/effect.
- **Locate –** This task corresponds to any action of searching and finding of already visualized information, such as: items, data, values, clusters, proprieties, position/locations, distances, and graphical primitives.
- **Visualize –** This task corresponds to a system's capability to visualize in following areas: number of dimensions, number of items, data, and domain parameters/attribute.
- **Configure –** This task corresponds to a system's capability to normalize, classify, filter, zoom, re-order dimensions, and changing visual characteristics or attributes in representation form. (Taken from: [24])

These previously established user tasks taxonomy of visualization tools provide a guide to a better understanding of currently available information visualization tools and methods, tools' capability, and expected tasks, performances, and use of visualization tools by the community. Knowing these current practices not only provides the knowledge of important and expected tasks in a visual community, but it helps to calibrate and measure the key activities in this evaluation approach. There are a few numbers of essential tasks mentioned by all the information visualization taxonomies. All of the taxonomies that have been mentioned in this thesis contain retrieving specific data entry information and comparing and extracting out the important information. These are essential features in visualization, because the main purpose is to obtain insights from a large amount of data.

Chapter 4. Criteria

This section discusses some existing information visualization evaluation criteria in the visualization community. This process will not only expedite and clarify the evaluation criteria early on, but it will help to guide and provide a basic structure needed in this evaluation approach.

Expressiveness and Effectiveness Criteria

These two sets of criteria by Jock Mackinlay [17], are probably the most recognized information visualization criteria. The expressiveness criteria concern system capability to visualize all the desired information. Effectiveness criteria are related to representing all the desired information; it includes how the information will perceive, aesthetic, and optimization. However, the criteria were developed in 1987. The basic graphic design principle and fundamental structure were adopted into this research's evaluation approach. However, the most of the criteria were discarded, because many were considered to be obsolete for current technology.

Cowley

As a part of developing an information visual analytic tool, Cowley et al. [22] developed process and product metrics. This is a domain specific evaluation method, which is developed to manage complex and iterative process with identifying assumptions, new discoveries with reasoning, and analyzing. Through this process, the evaluation method is to provide feedback on the use of the software, needed improvement, progress assessment, and decision making information. Process metric, which is still in the developing process, includes the following:

- **Solution time –** This applies only to problem-solving software
- **Number of documents read –** Possible number of documents that can be inputted into the system
- **Reading time/document –** Time it takes to upload the inputted data
- **Number of queries –** The number of queries that are attempted to compare
- **Comparison to expert –** Accuracy comparison of process
- **Rate of information growth –** The rate of incoming data or data change
- **Number of system generated hypotheses considered by the analyst –** The number of system recommended hypotheses
- **Number of relevant documents/query –** The number of documents/query required by analysts to generate hypotheses

Product metrics include:

- **Accuracy –** The degree of quality of being near to the absolute truth
- **Quality –** Including confidence in recommendations (Noted from: [22])

Most of these criteria are calibrated for exploratory search purpose, which is commonly faced and a closely related activity for intelligence analysts. Many of the criteria (all except time-related criteria) have been considered and they were recalibrated for intelligence analysis purpose.

Brath

Different information visualizations use a unique visualization method and representation form to convey information to users. Richard Brath [23] proposed a metrics system to evaluate the effectiveness of the 3D information visualizations representation form based on experiences from 60+ 3D visualizations. The author used the following criteria to measure the effectiveness of the representation form:

- **Number of data points** - the number possible data values, which can be represented on one screen
- **Data density** - the number of data points / number of pixels in the display excluding pixels of window borders and menus
- **Complexity –** Since more complex visualization is more difficult to comprehend, the author counted the maximum number of dimensions to measure complexity. He stated the greater the number of dimensions, which are displayed in the visualization, the greater the cognitive complexity for the user.
- **Effectiveness of the mapping –** It measures the effectiveness of the representation by evaluating the effectiveness of the mapping of each data dimension into a visual dimension.
- **Occlusion percentage –** number of data points completely obscured / number of data points
- **Identifiable points percentage –** number of visible data points identifiable in relation to every other visible data point /(number of visible data points * number of visible data points) (Noted from: [23])

Two of the criteria (number of data points and effectiveness of the mapping) are included in the intended use evaluation criteria to evaluate the effectiveness of visual representation forms.

Yand-Pelzez and Flowers

Metrics to evaluate visual displays were developed by Julie Yang-Pelaez and Woodie C. Flowers [18].The authors considered the following to measure the effectiveness of information contents:

- **Information content spanned by the data** – The amount of information spanned by the data is the sum of the information contents for each of the dimensions.
- **Information content of a data display** – ratio of the information contents of the data and the display
- **Information capacity of a display** – the amount of information that can be contained in a display
- **Topological information content** – This information will be determined by an approach by Rashevsky (Taken from: [18] [37])

One of the criteria (information capacity of a display) is included in the intended use evaluation criteria to evaluate the effectiveness of display.

Freitas

Freitas et al. [38] suggested evaluating information visualization techniques by addressing the evaluation of visual representation and interaction mechanisms. The criteria for testing usability of visual representation include:

- **Completeness** – representing all the semantic contents of the data to be displayed
- **Spatial organization** – the overall layout of a visual representation. It associates with the degree of difficulty to locate an information element on the display and awareness of overall information.
- **Information coding** – use of additional symbols besides the mapping of data elements
- **State transition** – the time it takes to redisplay the visual representation after a user action

Criteria for testing interaction mechanisms include:

- **Orientation and Help –** functions to support the user control in level of details
- **Navigation and Querying –** the degree of difficulty to select a data element, changing the user point of view, manipulating geometric representations of data elements, searching a query for specific information, and expanding clustered/hidden data elements
- **Data set reduction –** filtering undesirable information (Noted from: [38])

Completeness and navigation and query criteria were included in the intended use evaluation approach after calibrating for intelligence analysis purpose.

These established criteria were used to evaluate different information visualization techniques and tasks. They provided a guide to a better understanding of evaluating different aspects of a tool. The key activities in intelligence analysis community will be evaluated with similar criteria. Next, chapter presents the of intelligence analyst activities.

Chapter 5. The Intelligence Analyst Activities

This chapter discusses the intelligence analyst activities using information visual analytic tools in isolation from the rest of the system and whole activities. The activities will be based on visualizations' intended use. This consists of sets of tasks of the tool that will purposefully support activities in the operational environment. The set of tasks are selected based on users' analyzing activities using visualization that the user finds most important and/or most frequently faced.

Introduction of Intelligence Analyst Activities

Information visualization is applied in many fields for a faster insight with a less cognitive load for users. Information analysts are beginning to require more powerful tools to exploit massive, heterogeneous, and multi-source global information/data more effectively [39]. According to Gersh et al., in intelligence analysis as in many other domains that insight and hypothesis come from the process of data exploration [36]. In this domain, the process of gaining insight and hypothesis from data can benefit greatly with visualization technology, if the technology corresponds with the purpose well. Therefore, intelligence analysis is well suited for the purpose of this thesis. Within the field of intelligent analysis, or sometimes referred to as Intelligence and Security Informatics, there are a number of different types of work/effort or critical areas that are identified in Chen's work. [40].The one common aim of this area of study is to track and investigate any potential threats to the government or its representatives. The potential threats can be either a person or an event. Those threats can be categorized into two cases. The first case is when the threats are known and analysts are looking for the detailed information or connection to the threat. The second case is when the threats are unknown. In this case, analysts are looking for the general trends, patterns, structures, and outliers that are either suspicious or out of the ordinary.

Indirect observations via previous reports/papers have been applied to obtain intelligence analysts' key tasks. Also, repeatedly mentioned visual tasks in existing works were augmented into the key tasks to support essential information visualization tasks.

Intelligence Analyst Activities

Intelligence analysts' activities are an iterative process consisting of data acquisition, explore/discover, and making decisions as shown in Figure 5.1. Visualization is usually used in the explore/discover process and sometimes also when making decisions. The explore/discover process consists of analyzing data/info and interpreting data/info steps.

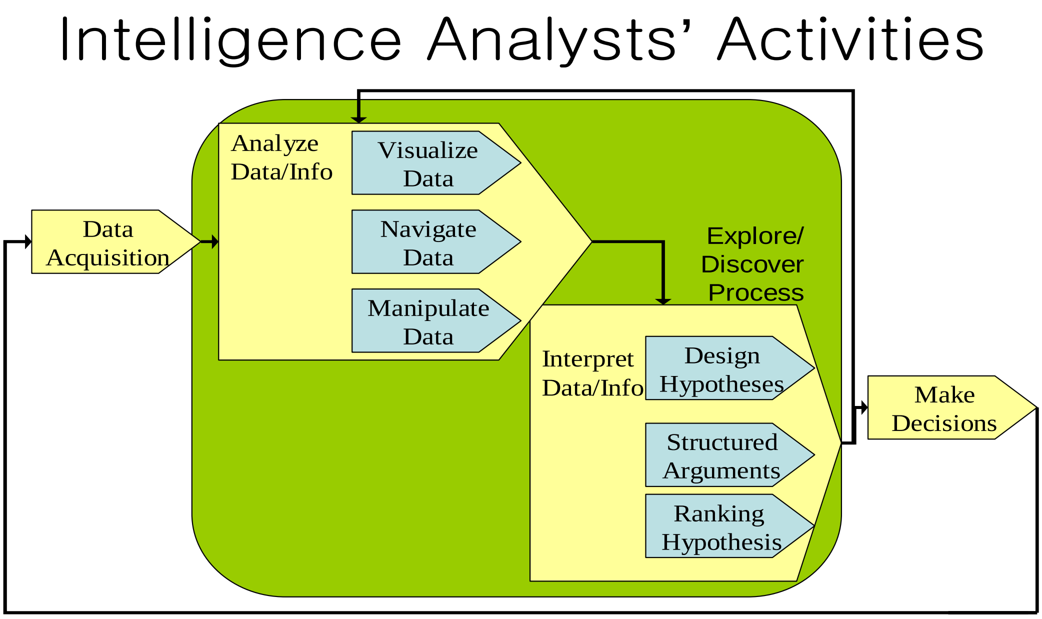


Figure 5.1 The Process of Intelligence Analysts Analyzing a Large Amount of Data.

As shown in Figure 5.1, intelligence analysts first acquire data. The primary use of information visualization in intelligence analysis work is to represent large volumes of data, so that trends, patterns, structures, and outliers within the data can be more readily observed or even perceived. According to Robert E. Horn, a recognized the problem is not a dearth of information but quite the opposite [41]. The proverbial needle in the haystack problem is often what confronts the intelligence analyst. Therefore, the visualization technologies must handle a large amount of documents, often of varied types and forms of media.

Next, intelligence analysts use information visualization is to analyze data as shown in Figure 5.1. In this process, visualization should help analysts in visualizing, navigating/exploring, and manipulating the data in a faster and easier. Due to the large volumes of data, details of the many data can be initially abstracted away from the user's view. The visualizations should provide users with an interaction technique for accessing the details of the data. Not only accessing the details of the data, but a brushing/linking view is also needed to recognize the selected data elements in multiple views. Brushing/linking is an interaction technique used to select a group of items displayed and visually show the connections to other items. If used in conjunction with multiple-linked views, which is common in intelligence analysis, brushing can enable users to understand correlations across multiple dimensions [42] [43] [44].

After intelligence analysts have analyzed data, analysts need to interpret the data. Information visualization allows analysts to observe data in a different perspective, to discover new and/or different insights/hypotheses. A technique such as grouping, which is used to reveal global patterns of relationships among data elements that are otherwise obscured, helps analysts to observe data in a different perspective. Also, highlighting and hiding the desired data are used to attain this task. These interaction techniques provide for a more seamless experience in attaining the key intelligence analysts' tasks [45]. After a general trend(s) has been analyzed, (or if analysts are looking for a specific threat), the analysts need to detect threats. This step can be considered as a brainstorming step. The goal of this step is to recognize any threat(s) or propose one or more hypotheses from the data. Analysts may need to manipulate or alternate data to detect threats/hypotheses, and then gather supporting or contradicting evidence to strengthen or weaken the found threat/hypothesis. If strong threats/hypotheses cannot be structured, analysts need to go back to analyzing data stage (note Figure 5.1).

Once threats have been detected or hypotheses have been recognized, analysts test the threats/hypotheses to make a decision. There are many ways to test threats/hypotheses. A few of the methods are: using animation or simulation, by comparing hypotheses, and by viewing evidence. After the test(s) have been conducted, the analysts rank the threats/hypotheses according to the relevance of the evidence and the credibility to face the crisis more effectively and in a more efficient manner. At times, threats/hypotheses are used as new data for greater threats/hypotheses (note Figure 5.1).

Classification of Analyst Tasks and Criteria

The following criteria were selected and developed once understanding of the type of tasks and goals for these particular intended users, in this case intelligence analysts, have been achieved.

1. Finding general trend(s):

Visualizing data

- **Maximum input size –** This addresses the maximum number of datasets allowed into the system.
- **Maximum output size –** This addresses the maximum number of documents that can be visualized by the system.
- **Time to navigate (getting detailed information) –** This addresses the capabilities for moving within and among views to get the summary of each event (who, when, where, what, how, and to whom) or detailed personal information. This includes the capability to zoom, pan, rotate, reset, tilt, etc.
- **Geospatial and Temporal Representation –** This addresses the system capability for tracking movement according to time.

- **Social Representation –** This addresses the system capability for social analysis.

Explore/Navigate

- **Efficiency –** This addresses the exploring/navigating time to view and access desired information.
- **Effort –** This addresses the availability of additional visual representations (e.g. timeline, animation, table, histogrid, graph, and map), which reduces the analysts' cognitive load and allows to observe in a different perspective.
- **Extra Navigation Feature - Brushing/linking –** This addresses the system capability of enabling analysts to understand correlations across multiple dimensions.

Re-visualizing

- **Multiple views –** This addresses the system capability, which allows analysts to engage in a number of views using the same or different presentation form, for comparison purpose.
- **Theorizing –** This addresses the system capability, such as grouping, highlighting, hiding, etc. that can help analysts' thinking processes of analyzing.

2. Specific threat or second step after finding general trend(s):

Threat Detection/Brainstorming

- **Manipulating/Alternating Data and View–** This addresses the features that allow analysts to manipulate and alternate data and/or view in order to help brainstorming and/or hypothesizing.
- **Gathering supporting evidences –** This addresses the features that help analysts to gather supporting evidences for their hypothesis.

Threat/Hypothesis Testing

- **Animation or simulation –** This addresses the presence or levels of animation or simulation employed in the system to help analysts test the threat/hypothesis.
- **Re-visualizing the hypothesis –** This addresses the presence or levels of features that can re-visualize the hypothesis to help analysts test the threat/hypothesis.
- **Comparing hypotheses (possibly with other analysts' hypotheses) -** This addresses the presence or levels of features that can help compare a number of hypotheses to help analysts test the threat/hypothesis.
- **View evidence -** This addresses the presence or levels of features that can help view evidence of a hypothesis/threat to assist and help analysts recognize the possibility and the level of relevance of the threat/hypothesis.

Ranking the Threats/hypothesis (Threat Analysis)

- **Relevance of hypothesis/threat –** This addresses the presence or levels of features that can help recognize the relevance of hypothesis/threat to assist analysts and to rank the threats/hypothesis.
- **Evidence credibility –** This addresses the presence or levels of features that can help show evidence credibility, which is the number of supporting evidences subtracted by the number of contradicting evidences, on a hypothesis/threat to assist analysts and to rank the threats/hypothesis.

Assumption: all the data have been collected, translated into English, and filtered.

Chapter 6. Test & Result

This section discusses the details about the experimental investigation and the result. Evaluation plays a critical part in many areas. Evaluation is used in developing, application selection, and performance checks. The evaluation presents many facts about subjects that are being studied. In this thesis, evaluation serves to find out information visual analytic tools' suitability for a use/task according to the criteria described in Chapter 5. The criteria were selected and developed after analyzing the type of tasks and goals for intelligence analysts. The evaluation has been applied to a number of information visualization products.

Experiment Design

The main goal of this study is to evaluate a number of information visualization tools based on the criteria developed in Chapter 5 to exercise and investigate the criteria and to verify if it can improve information visualization evaluation. The process is used to ensure whether the evaluation approach can deliver relatively accurate and relevant information in a more efficient manner.

First, to develop criteria that reflect intended usage of a domain field, an investigation on intelligence analysis activities was performed. An indirect observation of intelligence analysts' activities have been conducted via published work regarding intelligence community.

Once the intelligence analysis activities were acknowledged, relevant criteria that can effectively measure some of the activities were carefully chosen from exiting evaluation methods. The rest of the criteria were developed to measure the key intended use/activities.

In this chapter, using the newly gained information from analyzing, five visual analytic tools are being chosen and examined. These tools represent different varieties of information visualization methods and show potential for intelligence analysis use. The evaluations on the five tools were conducted to exercise and investigate the evaluation approach and to understand tools' potential strengths and limitations. For this process, the objectives are:

1. Compare thesis' evaluation approach to existing evaluation methods, to verify if the evaluation approach reduced the time that is required to evaluate a visualization tool.
2. Check if shortened list of candidates selected as a result of the analyzing intelligence activities response properly for intelligence analysis purpose.
3. Check if the results of the evaluation approach are relatively accurate at an acceptable level.
4. Check if the results of the evaluation approach are better targeting the domain users' key tasks than existing evaluation methods.

The entire experiment process is shown in diagram below.

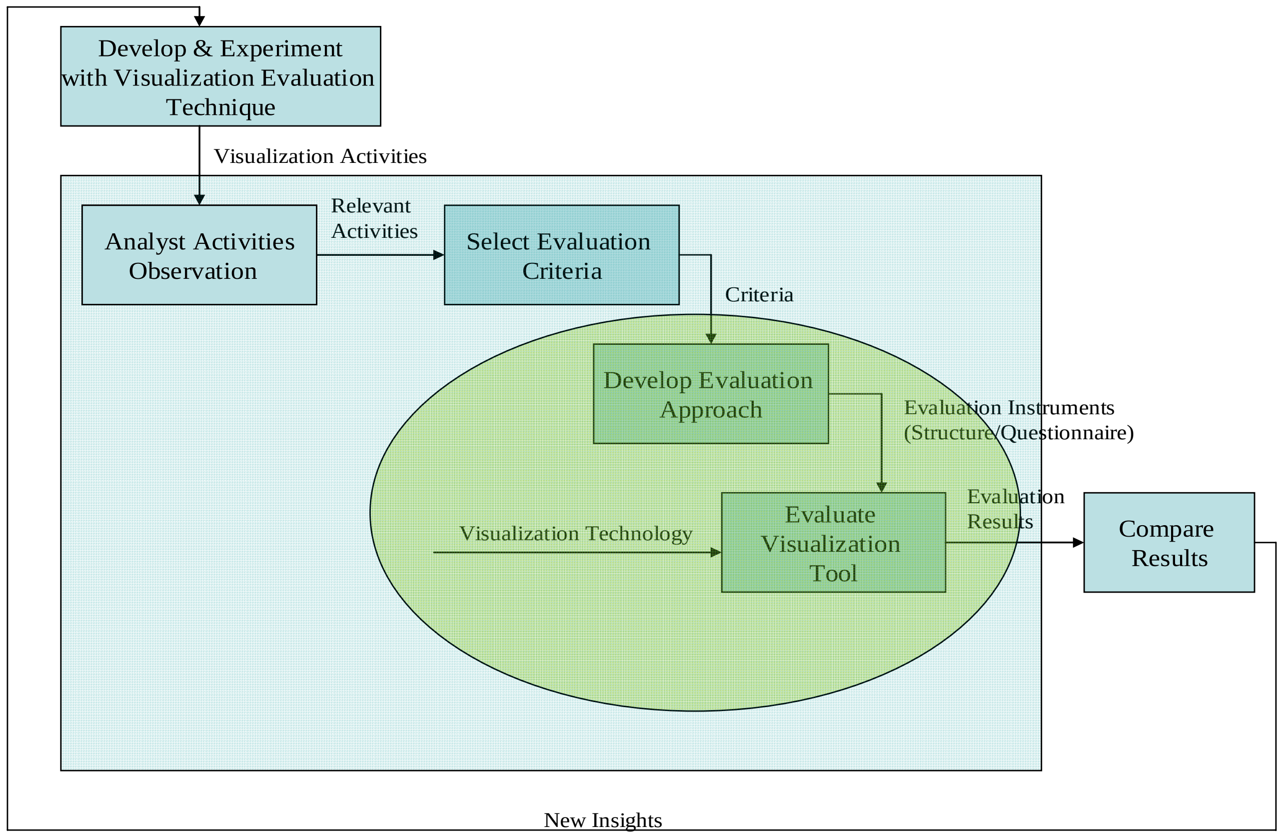


Figure 6.1 Experiment Design

Figure 6.1 depicts the experiment design in demonstrating the intended use evaluation approach. The whole procedure is an iterative process, which consists of three big steps. The three big steps are separated by the big square. The first process represents creating a framework. The first two steps in the big square are preconditioning processes. In these processes, activities are being observed and analyzed and criteria for a domain are being developed. The circled steps represent the actual evaluation processes using the newly developed evaluation approach. The last step represents the validation process.

Information Visualization Methods and Tools

The tools that were selected can well represent different varieties of information visualization methods. According to Card et. al., information visualization methods can be categorized as following: space, interaction, focus + context, document visualization, infosphere, workspace tools and objects [1].

According to the Card et. al., each category is not mutually exclusive and uses its own structure to convert data to visual form, which can be further converted to insights. Space is further categorized as physical data, 1D, 2D, 3D, multiple dimensions, trees and networks. Interaction is categorized as dynamic queries, interactive analysis and overview + detail. Focus + Context is categorized as fisheye views, bifocal lens, visual transfer functions. Document visualization is categorized as 1D, 2D, 3D, and 3D + time. Infosphere, workspace, tools and objects are further categorized as internet and infosphere, information workspaces, and visually enhanced objects. This category is focusing on creating new knowledge using information visualization tools. Infosphere is a term used by the authors to signify all the information in the world outside the users' immediate grasp/environment, such as information on the internet or an online database on a server [1]. Each category and subcategory has a number of tools as shown in Table 6.1.

Visualization Category	Sub-Category	Example tools
Space	Physical Data	Any interactive computer graphics which can represent human structure, mathematical equation, meteorology, and physics
	1D, 2D, 3D	1D: Lifestreams, SeeSoft, Tilebars 2D: Spotfire 3D: Voxel-Man, Theme view, and Narcissus
	Multiple Dimensions > 3	n-Vision VisDB InfoCrystal
	Trees	Treemap SeeSys TennisViewer Clsuterview
	Networks	SemNet HierNet SEENET
Interaction	Dynamic Queries	HomeFinder FilmFinder SDM (Selective Dynamic Manipulation)
	Overview + Detail	SeeSoft Dynamic Queries Sliders LifeLines PDQ (Pruning with Dynamic Queries) Tree-browser visualization tool
Focus + Context	Fisheye Views	Fisheye
	Bifocal Lens	Bifocal
	Visual Transfer Functions	Table Lens Other Focus + Context tools: cone tree, hyperbolic browser
Document Visualization	Text in 1D	Seesoft
	Text in 2D	Kohonen's feature map
	Text in 3D	Galaxies Theme View
	Text in 3D + Time	Galaxy of News
Infosphere, Workspace, Tools, Objects	Internet and Infosphere	Timeline ET-Map Hyper-G Time tube Hotsauce NicheWorks
	Information Workspaces	Cone tree Perspective Wall Starlight
	Visually Enhanced Objects	Document Lens Visible Human

Table 6.1 Information visualization methods and tools according to Card et. al. [1]

Tool Selection

In this process, using the newly gained information from analyzing, five visual analytic tools are being chosen and examined. These tools represent different varieties of information visualization methods and show potential for intelligence analysis use. Each category has a number of commercial or academic tools that are currently accessible. However, the popularity and availability of the tool had the most impact on choosing tools. Note that these tools are developed for a different purpose than intelligence analysis use. The following is the list of evaluated tools:

1. Space: Spotfire [46]
2. Interaction: TimeSearcher 2 [47, 48]
3. Focus + Context: TableLens [49]
4. Document Visualization: In-Spire (Galaxy and Theme view) [50]
5. Infosphere, Workspace, Tools, Objects: HCE3 (Univ. of Maryland) [51]

In the following subsections, each tool is briefly described as well as an explanation for selecting it.

Spotfire

Spotfire is one of the most popular information visualization software packages on the market. It helps users like intelligence analysts to make analytical decisions by spotting general trends, outliers, and unanticipated relationships. This package has a host of different representations including a tabular structure (cells in a spreadsheet), graphical depictions of objects, and key connections between the representations.

Spotfire allows the users to visualize the dataset in five different ways (four conventional 2D graphs, and one 3D graph).

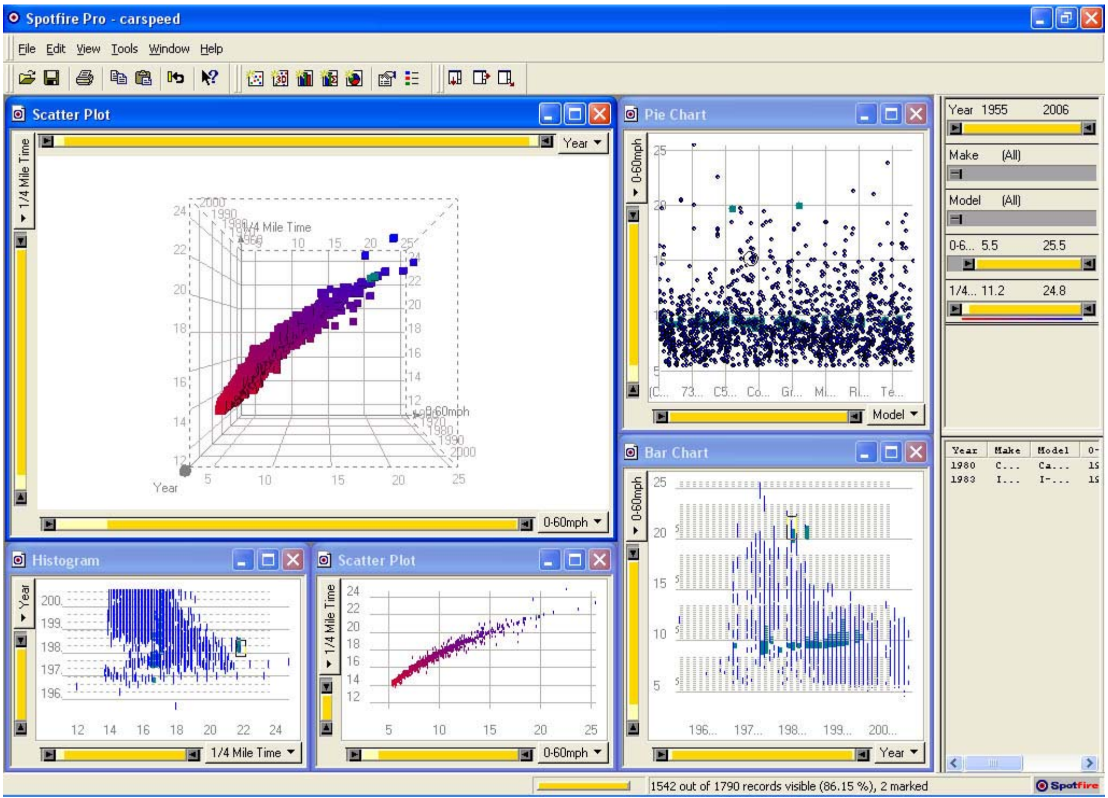


Figure 6.2 Spotfire

TimeSearcher 2

TimeSearcher is an information visualization program developed by the University of Maryland for genome research. It has many capabilities, especially its capability to analyze time series data effectively that can be adapted for intelligence analysts. Unlike the first version of TimeSearcher, the second version mainly uses a parallel-coordination to visualize the overview of the data. Time-box is the main navigation and filtering tool.

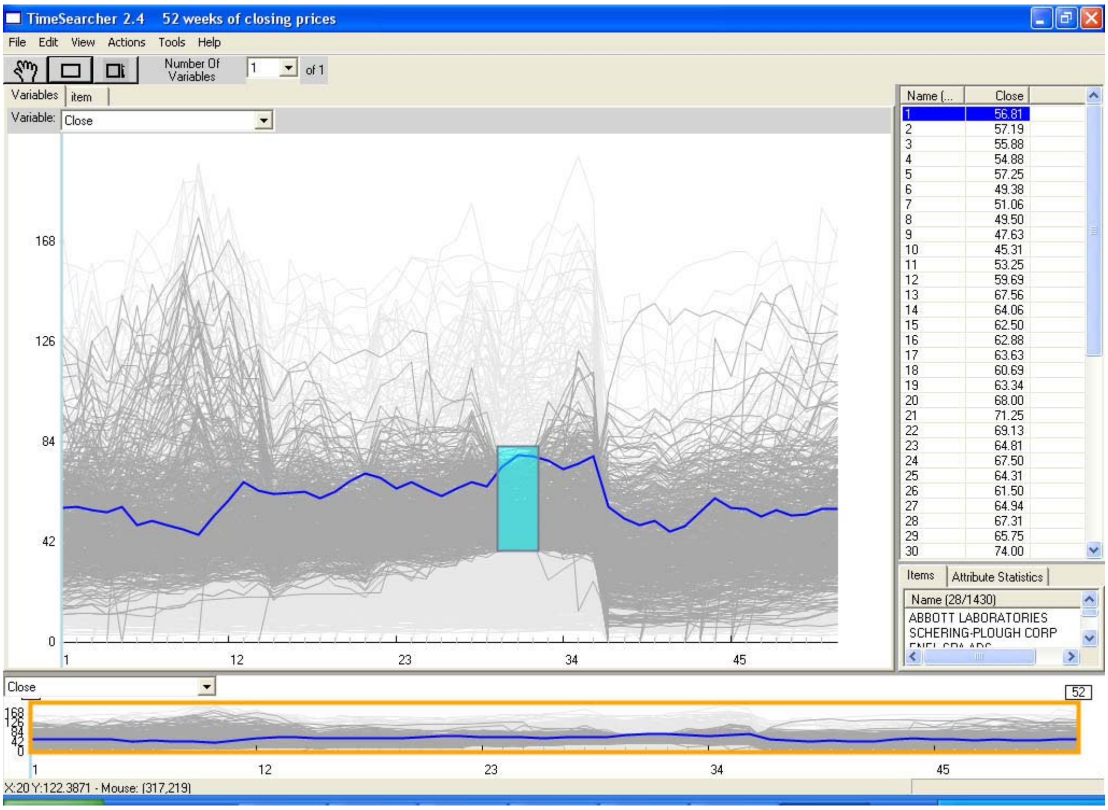


Figure 6.3 TimeSearcher 2

TableLens

TableLens (a.k.a. Eureka) is one of the most popular information visualization software packages on the market. It helps users to view the entire dataset and selected detailed information in a tabular format, which eases the work of obtaining insight. This tool may work well for intelligence analysts, because they often need to view the entire dataset and some detailed information at the same time.

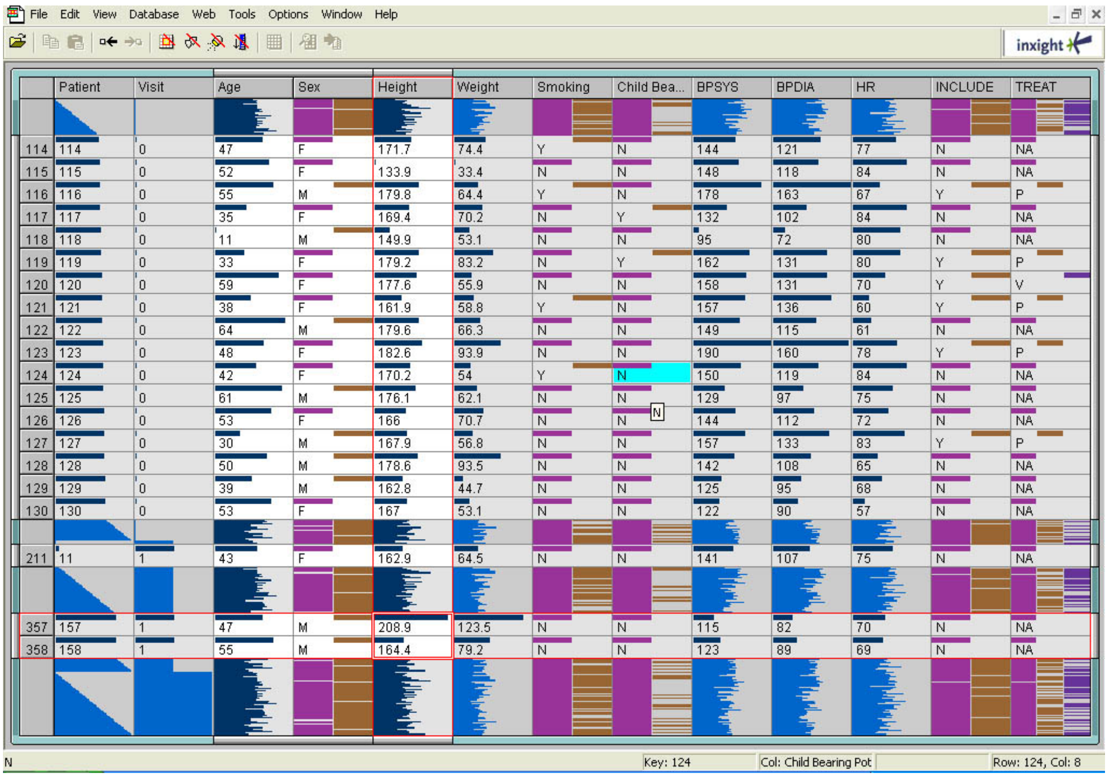


Figure 6.4 Tablelens

In-Spire

In-Spire takes in a maximum of 200,000 unstructured documents to show closeness of documents by clustering similar documents together and separating unrelated documents in the two main views, "Theme view" and "Galaxy view." This tool may be a good fit for intelligence analysis work, because intelligence analysts often need to deal with copious data/documents according to Robert E. Horn [41].

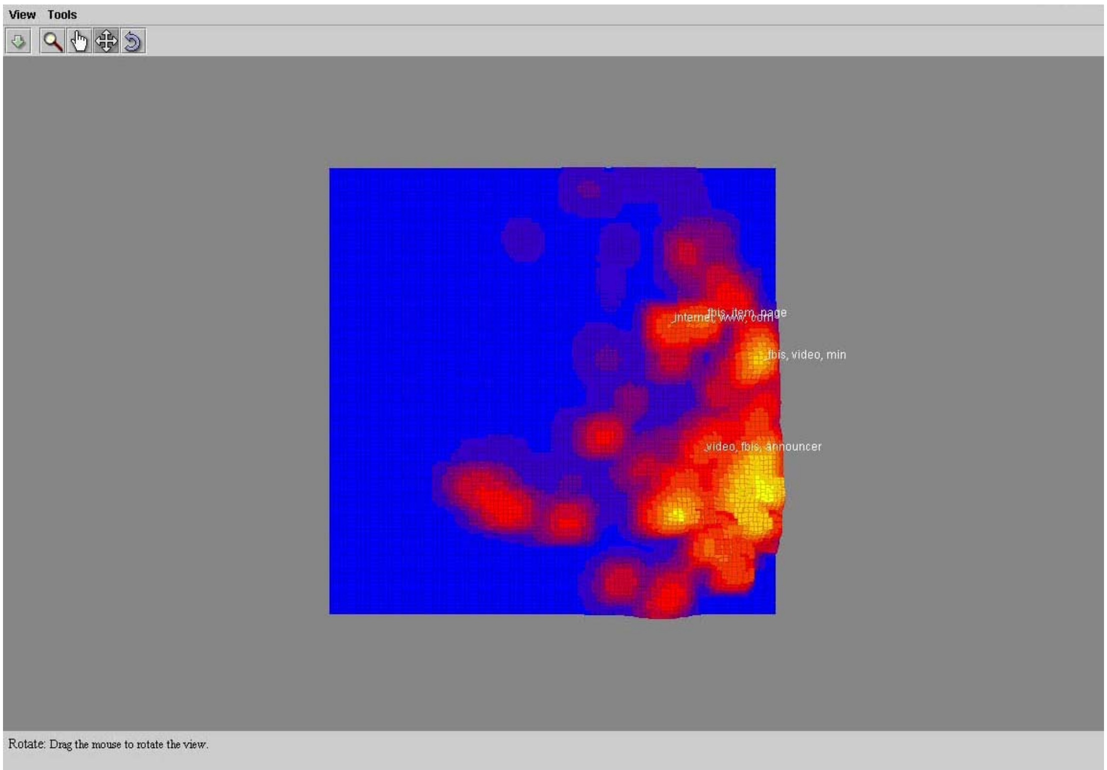


Figure 6.5 Theme view

Theme view – This uses a map of natural terrain to represent a set of documents. The view creates a 3D map in which the clusters are displayed as peaks. The height of the peaks indicates the relative strengths of the topics in the document set. Brightness has a linear relationship with the height. The higher the peak is, the brighter it is. The color is to show dominant theme combinations.

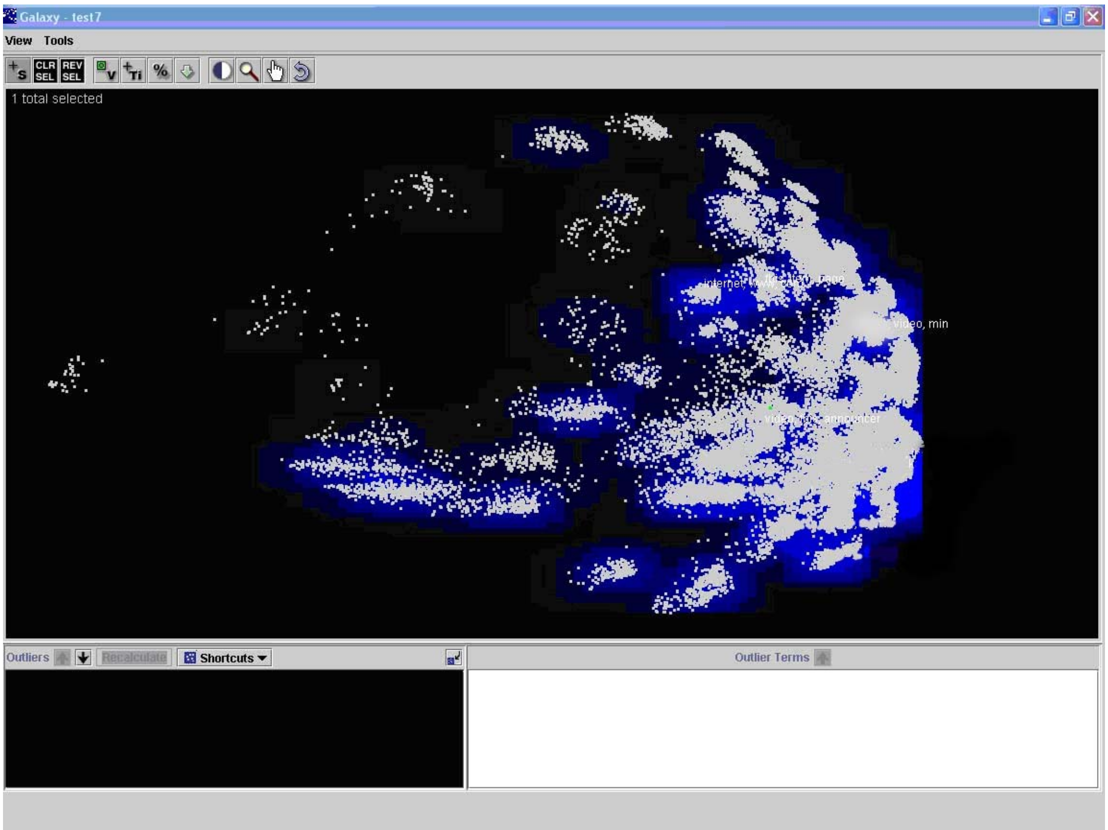


Figure 6.6 Galaxy view

Galaxy view - This uses the image of the galaxy to represent a set of documents. It's in 2D representation; clusters can be formed depending on the similarity. It has the same spatial arrangement as the Theme View. It groups similar or related documents near each other, which can be turned into a cluster of stars, and dissimilar documents far apart. The size of clusters and eventual brightness of the star cluster will represent the size of the document cluster. The closeness represents the closeness in topics.

HCE3 (Hierarchical Clustering Explorer)

HCE is visualization software developed by the University of Maryland for genome research, yet many features can be adopted into an intelligence analysis tool. It reads in multidimensional datasets and displays user interactive visualizations, such as color mapping, dendogram, color mosaic, table view, histogram ordering, profile search, evaluation tab, and K-means. This tool can be useful to intelligence analysts, because it graphically represents multidimensional dataset well in many different methods.

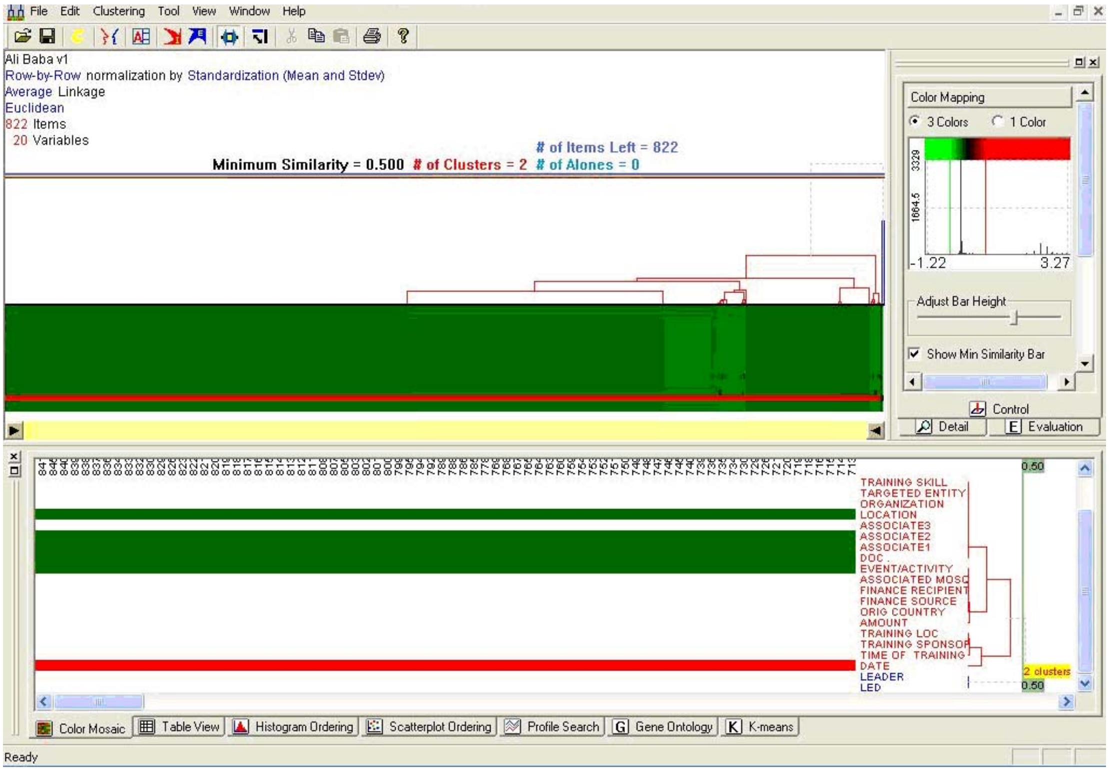


Figure 6.7 HCE3

Datasets

To examine information visualization for intelligence analysis purposes, it may be logical or appropriate to use a field dataset. Attempted to use a common data set for the tools, but for some of the tools, (Spotfire, TableLens, and TimeSearcher 2) there was not a way to ingest the information, mainly because the tools were not developed for intelligence analysis purpose. The fall back position was with demonstration datasets provided by the vendor of each tool.

In-Spire and HCE3, an intelligence analysis dataset which contains over 34,000 unstructured documents, has been applied for the experiment. The unstructured documents consist of trivial information including names, addresses, locations, times and messages. Some of the documents contain common subjects and relationships. Due to the sponsor's request, the name of the dataset cannot be revealed.

The following are some examples and descriptions of some of the data used for the examination and comparison of the tools. These datasets were selected based on each tools' capability to handle different format of datasets.

Tool	Dataset Description
Spotfire	Data for automobile from 1955 to 2006 including year, make, model, 0-60 mile time, and quarter mile time. (1790 rows and 5 columns)
TimeSearcher 2	A closing price of each week of the year for 1430 items
TableLens	Data for patient record. This dataset contains 600 patients with 13 attributes. (600 rows and 13 columns)
In-Spire and HCE3	34,000 unstructured documents consisting of names, addresses, locations, times and messages

Table 6.2 Visual Analytic Datasets in the Experiment

Evaluation Results

Note that many of the tools were not developed for intelligence analysis use. Some of the criteria could not be exercised, but they were used to find similar attributed or suggestion for the tool. Results are presented in terms of completeness of the tool and according to the criteria.

Spotfire

Finding general trend(s):

Visualizing data

- **Maximum input size –** one dataset at a time
- **Maximum output size –** less than 1,000,000 items & less than 10 attributes. The data scalability for this software is presumed to be less than 1,000,000 items and less than 10 attributes. This is measured by the effectiveness of showing the maximum number of data in the main visualization, scatter plot on a regular monitor, 1024 by 768 pixels.
- **Time to navigate (getting detailed information) –** It uses "Zoom" method with a scroll bar. Zooming in/out also effects the length of the bar inside the scroll bar. As the length of the bar gets longer, the more zoomed out it is and vise versa. Spotfire uses scatter plot as the main graph, which is familiar to most. Due to the familiarity and the nature of the scatter plot, identifying outliers, maximum, and minimum value can be done instantly. There were three main problems encountered. First, if the data contain two exact same values, one gets hidden. Second, navigating 3D scatter plot is not intuitive. Third, identifying mean value required some time, especially since there were many data. If the system would be able to upload intelligence analysis data (event), identifying summary of outliers, the maximum and minimum would be intuitive.

- **Geospatial and Temporal Representation –** The system does not support any feature to track movement according to time; however, it can present events according to time. Also, geospatial can be implemented using 2D scatter plot with US map on the background and location as one attribute. If time is the other attribute, they system can visually track a movement.
- **Social Representation –** The system does not have features for social analysis. Unless network graph becomes one of the representation forms, social analysis may not be possible with Spotfire.

Explore/Navigate

- **Efficiency –** Identifying the desired data on the visual representation may be a problem. However, accessing the detailed information from the visual representation can easily be done by clicking the data. Identifying outliers and the minimum and maximum value are uncomplicated in Spotfire; however, mean and median are not so obvious. Also, the system can benefit greatly by having a query feature for data tracking purpose.
- **Effort –** The available types of graphs are: 2D scatter plot, 3D scatter plot, histogram, bar chart, and pie chart.
    - 2D Scatter plot is the most used feature. It provides users with a quick glance of relationships between multiple attributes.
    - 3D Scatter plot also provides users with a quick glance of relationships between three attributes. Users can navigate the 3D space by moving around the graph, but it is not as easy as the 2D graphs.
    - A histogram is a 2D representation that allows users to measure aspects of one attribute versus another attribute. An example use of histogram is to visualize a number of units produced for each week of the year.
    - A bar chart is similar to a histogram. A bar chart enables users to compare several alternatives against a given criteria. An example use of

a bar chart is to visualize a number of units consumed compared with a list of production lines.

  - A pie chart is visualized in a circular view (pie). The percentage of each item/column is shown with respect to the total (pie). The area of the circle represents the whole, and the area of one sector of the circle represents the percentage of one item/column. Only one attribute can be shown for one pie chart. An example use of a pie chart is visualizing the percentage of profits for a corporation against all of the business units.
- **Extra Navigation Feature - Brushing/linking –** The brushing/linking feature is available among all the representations to help analysts understand correlations across multiple dimensions. This feature can aid analysts a great deal, because Spotfire can visualize only up to three attributes in one representation form. In many cases, multiple views are desirable. Therefore, brushing/linking helps analysts to think in multiple dimensions.

Re-visualizing

- **Multiple views –** Spotfire has a feature called Multi-view Window. It enables users to view multiple windows of analyses at a time. For example, a window with geospatial information may be coordinated and viewed at the same time as a social network analysis diagram to ascertain the connections between suspects from a geographic perspective. More than 25 windows can be opened and viewed at the same time, although more than 16 windows are ineffective. It adjusts the window size automatically for users. However, users can always pull one window out and resize.
- **Theorizing –** The slider bar feature allows analysts to highlight and hide data. Also, it can help analysts recognize patterns and changes. As a result, it reduces analysts' cognitive load during theorizing.

Specific threat or second step after finding general trend(s):

Threat Detection/Brainstorming

- **Manipulating/Alternating Data and View–** The system does not allow analysts to manipulate and alternate data or view. Once graphical representations have been visualized, data cannot be changed. In order to manipulate/alternate the visualization, data have to be changed.
- **Gathering supporting evidences –** The system does not include any feature which allows analysts to gather supporting evidences for their hypothesis. The system helps analysts to view and identify a general trend, outliers, and max and mean values, but it does not support brainstorming, hypothesizing or gathering evidence.

Threat/Hypothesis Testing

- **Animation or simulation –** The system does not support animation or simulation for testing the threat/hypothesis. Currently, the system does not support geospatial and temporal representation. However, the system can adopt these features relatively easily. Once the system supports these features, an animation or simulation that shows the geological movement according to time, can have a great effect on analyzing.
- **Re-visualizing the hypothesis –** The system does not support any feature that can re-visualize the hypothesis to help analysts test the threat/hypothesis. The only way to visualize hypothesis is to change the data according to hypothesis and visualize, which is inconvenient.
- **Comparing hypotheses (possibly with other analysts' hypotheses) -** The system does not support any feature that can help compare a number of hypotheses to help analysts test the threat/hypothesis. The only way to

visualize hypothesis is to change the data according to hypotheses and visualize them using multi-view.

- **View evidence -** The system does not support any feature that can help view evidence of a hypothesis/threat to assist analysts and to recognize the possibility and the level of relevance of the threat/hypothesis. A simple solution for the system is annotation.

Ranking the Threats/hypothesis (Threat Analysis)

- **Relevance of hypothesis/threat –** The system does not support any feature that can help recognize the relevance of hypothesis/threat to assist the analysts' ranking of the threats/hypothesis. Due to the system's capability, there is not any simple solution for this aspect.
- **Evidence credibility –** The system does not support any feature that can help show evidence credibility, which is the number of supporting evidences subtracted by the number of contradicting evidences, on a hypothesis/threat to assist the analysts' ranking of the threats/hypothesis. Annotation with analysts' effort may help with this aspect.

TimeSearcher 2

Finding general trend(s):

Visualizing data

- **Maximum input size –** one dataset at a time
- **Maximum output size –** less than 1,000 and less than 20 attributes.

  The main visualization for the system is parallel coordinates. Parallel coordinates can visualize more than 1,000 items and 20 attributes and the system is currently set to do so. However, once it visualizes over 1,000 items and 20 attributes, the effectiveness drops dramatically. Therefore, the data scalability for this system is presumed to be less than 1,000 items and less than 20 attributes. This is measured by the effectiveness of showing the maximum number of data in the main visualization, parallel coordinates on a regular monitor, 1024 by 768 pixels.

- **Time to navigate (getting detailed information) –** The main navigation feature for the system is a slider “Zoom” method using overview. Controlling the size of overview decides the level of detailed view.
- **Geospatial and Temporal Representation –** The system does not support any feature to support geospatial data; however, the system is well designed to analyze time series data or temporal data effectively. With little modification, the system can also track movement.
- **Social Representation –** The system does not have features for social analysis and the system is not fit to analyze social representation.

Explore/Navigate

- **Efficiency** – Parallel-coordination being a 2D visualization, TimeSearcher only lets two attributes (one attribute always being time) to be compared at a time. However, if there is a need to analyze multiple attributes, a multiple number of parallel-coordinations can be opened by setting the number of variables. Getting the detailed information of a desired data may not be easy once the number of data get big. If analysts can locate the desired data either from the graph or spreadsheet, analysts just need to select the data to access all the detailed information.
- **Effort –** In version 2.4, there is only one representation form. However, a spreadsheet is available next to the parallel coordinates.
- **Extra Navigation Feature - Brushing/linking –** The brushing/linking feature is available between parallel coordinates and the table.
    - There is one more main feature in TimeSearcher 2, called search-box. The feature allows users to locate all the similar trends that are

highlighted within search-box.

Re-visualizing

- **Multiple views –** As mentioned above, the system support analysts to engage multiple views (multiple number of parallel-coordination) by setting the number of variables.
- **Theorizing –** The time-box can be used to filter, highlight, and hide in order to help analysts' thinking process. By changing the location and size of the time-box, users can filter, highlight, and hide desired/undesired data. Also, if there is more than one attribute, the attribute can be used as a filter as well.

Specific threat or second step after finding general trend(s):

Threat Detection/Brainstorming

- **Manipulating/Alternating Data and View–** The system does not allow analysts to manipulate and alternate data and/or view, which can help brainstorming and hypothesizing. The only way to manipulate/alternate data is to change the data and to visualize the changed data. Due to the nature of the system, temporal threat detection can be implemented.
- **Gathering supporting evidences –** The system does not include any feature which allows analysts to gather supporting evidences for their hypothesis. However, a simple implementation, such as marking entity on the spreadsheet can be used to support evidence gathering.

Threat/Hypothesis Testing

- **Animation or simulation –** The system does not support animation or simulation for testing the threat/hypothesis. However, since the system can be used to analyze temporal data, animation may benefit analysts greatly.

- **Re-visualizing the hypothesis –** The system does not support any feature that can re-visualize the hypothesis to help analysts test the threat/hypothesis. The only way to visualize hypotheses is by changing the data and visualizing the changed data.
- **Comparing hypotheses (possibly with other analysts' hypotheses) -** The system does not support any feature that can help compare a number of hypotheses to help analysts test the threat/hypothesis. However, once visualizing hypotheses is implemented, comparing hypotheses can be done by using the multi-view feature.
- **View evidence -** The system does not support any feature that can help view evidence of a hypothesis/threat to help analysts recognize the possibility and the level of relevance of the threat/hypothesis. However, a simple implementation, such as tracking entity on the spreadsheet can be used to track and view evidences.

Ranking the Threats/hypothesis (Threat Analysis)

- **Relevance of hypothesis/threat –** The system does not support any feature that can help recognize the relevance of hypothesis/threat to help analysts rank the threats/hypothesis.
- **Evidence credibility –** The system does not support any feature that can help show evidence credibility, which is the number of supporting evidences subtracted by the number of contradicting evidences, on a hypothesis/threat to help analysts rank the threats/hypothesis. As mentioned above, tracking evidence on the spreadsheet may be used to check evidence credibility in some degree.

Tablelens

Finding general trend(s):

Visualizing data

- **Maximum input size –** One dataset at a time
- **Maximum output size –** less than 1,000 items and less than 50 attributes. The data scalability for the system is presumed to be less than 1,000 items and less than 50 attributes. This number is calculated considering about twenty pixels for the length of bar graph, one pixel for the width of the bar graph, and on a 19-inch screen (1024 by 768 pixels).
- **Time to navigate (getting detailed information) –** Tablelens uses "focus + detail" as known as "overview + detail" method. The system starts out with just an overview of the dataset. The application uses the "Aggregate" (combining) method to visualize data. It combines the multiple data to fit into the screen. It combines them by stacking one on top of another. Depending on the size of the dataset, it might be aggregated or not. However, when the data is being combined, it uses a color scheme to show the differentiation. Also analysts can obtain the detailed information by clicking the data to view the detail information. When an item is clicked or focused, the detailed information will be shown along with the overview of the dataset. Multiple numbers of detailed information can be shown at a time. Highlighted information can be opened up at the same time. The detailed information can be closed by clicking back the detailed information. The identifying of maximum and minimum is intuitive. There are features such as sort, rearrange, promote, and filter to help analysts navigate. (For detailed information on these features, note theorizing.)
- **Geospatial and Temporal Representation –** The system does not support any feature to track movement according to time. Although it may not be clear, Tablelens can present and sort events according to time.
- **Social Representation –** The system does not have features for social analysis. There is not an easy way to do a social analysis using Tablelens.

Explore/Navigate

- **Efficiency –** One of the biggest advantages of this tool is that it lets analysts compare all the attributes in one visual representation. Analysts can position the attributes depending on the importance. If analysts can locate the desired data on the visualization, analysts just need to select the data to access all the detailed information.
- **Effort –** There is only one representation form. However, Tablelens offers features like sort, rearrange, promote, focus, filter, and spotlight for easier explore/navigation.
- **Extra Navigation Feature - Brushing/linking –** The brushing/linking feature is not available since there is only one representation form.

Re-visualizing

- **Multiple views –** The system does not support analysts to engage multiple views for comparison purpose. However, the analysts can compare within the visual representation, since all the attributes and entities are visualized in one representation.
- **Theorizing –** The system helps analysts to sort by column, rearrange columns, create subgroups (called promote), filter, and track particular information during sorting (called spotlight). These features can help analysts analyze data and create a hypothesis in some degree.

Specific threat or second step after finding general trend(s):

Threat Detection/Brainstorming

- **Manipulating/Alternating Data and View–** The system does not allow analysts to manipulate and alternate data and/or view. It lets analysts view the data in a different perspective by sorting differently, but it does not allow analysts to change the data.

- **Gathering supporting evidences –** The system does not include any features which allow analysts to gather supporting evidences for their hypothesis. However, with little modification of the sub-grouping/promote feature, the system can gather evidences.

Threat/Hypothesis Testing

- **Animation or simulation –** The system does not support animation or simulation for testing the threat/hypothesis. Since tracking temporal and geospatial data are complicated, using Tablelens animation may not be beneficial.
- **Re-visualizing the hypothesis –** The system does not support any feature that can re-visualize the hypothesis to help analysts test the threat/hypothesis. However, with little modification of sub-grouping/promote feature can be turned into hypothesis feature.
- **Comparing hypotheses (possibly with other analysts' hypotheses) -** The system does not support any feature that can help compare a number of hypotheses to help analysts test the threat/hypothesis. Once again, the sub-grouping/promote feature can turn into a hypothesis feature, so comparing hypotheses may be possible with Tablelens.
- **View evidence -** The system does not support any feature that can help view evidence of a hypothesis/threat to help analysts recognize the possibility and the level of relevance of the threat/hypothesis.

Ranking the Threats/hypothesis (Threat Analysis)

- **Relevance of hypothesis/threat –** The system does not support any feature that can help recognize the relevance of hypothesis/threat to help analysts rank the threats/hypothesis.

- **Evidence credibility –** The system does not support any feature that can help show evidence credibility, which is the number of supporting evidences subtracted by the number of contradicting evidences, on a hypothesis/threat to help analysts rank the threats/hypothesis.

In-Spire

Finding general trend(s):

Visualizing data

- **Maximum input size –** In-Spire lets users upload as many datasets as the operating platform allows. In-Spire's first upload is much longer than other evaluated tools. However, after the first uploads, it saves the output, which accelerates the uploading time.
- **Maximum output size –** According to the In-Spire's homepage [50], In-spire uploads dataset that is more than 15 documents and less than 200,000 documents.
- **Time to navigate (getting detailed information) –** Most navigation is done with zooming and panning. The graphic is kept to be simple, so no lagging occurs while navigating. There are other features such as reset and rotation to help navigation. A basic zooming and panning technology is applied for most of the navigation.
- **Geospatial and Temporal Representation –** The system does not support any feature to track movement according to time. However, if an important location or time has been mentioned repeatedly, it is possible to identify the location and time.
- **Social Representation –** The system does not have features for social

analysis. However, with galaxy view, it is possible to see closely related people and unrelated people. It may be possible to identify groups of people.

Explore/Navigate

- **Efficiency –** The tool can read all the documents and can present in the form of a map of collections. The tool uses the "grocery store" concept, which is laying out similar documents near each other and dissimilar documents far apart. The negative aspect of this visualization is identifying the documents with multiple themes and determining the exact location on the map. For this type of document, it is not clear how to look for the document; the documents can be located either in both clusters or in between. However, it can definitely help to give users a quick overview of a dataset. It can effectively find commonly used topics, names, addresses, and other subjects in a number of documents. It can further contribute to reveal grouping, similarities, differences, and unanticipated relationships within the document collection. Assessing information on one particular document/data is not easy, just because there are too many documents on the representation. However, In-Spire supports a query feature to help analysts locate and access detailed information of their description.
- **Effort –** The tool produces two different types of graphical outputs, theme view and galaxy view. A Gist view is available as one of the features of galaxy view. It allows analysts to select one or a number of documents to display the words and the number of documents containing the word in selected documents. A similar feature is embedded in theme view, called Probe. Instead of displaying the actual number, it presents as bar graphs.
- **Extra Navigation Feature - Brushing/linking –** The brushing/linking feature is available between Gist view and Galaxy view and between Probe and Theme view. These features help, especially when a large number of documents have been visualized.

Re-visualizing

- **Multiple views –** The system allows analysts to engage multiple views of different presentation forms (theme view and galaxy view), for comparison purpose.
- **Theorizing** – Galaxy view allows analysts to highlight and hide by brightening and dimming the light of stars. In result it helps analysts' thinking process of analyzing.

Specific threat or second step after finding general trend(s):

Threat Detection/Brainstorming

- **Manipulating/Alternating Data and View–** The system does not allow analysts to manipulate and alternate data and/or view, which can help brainstorming and hypothesizing. Once graphical representations have been visualized, data cannot be changed. In order to manipulate/alternate the visualization, data have to be changed.
- **Gathering supporting evidences –** The system does not include any feature which allows analysts to gather supporting evidences for their hypothesis. A simple solution for this problem is to have an annotation feature.

Threat/Hypothesis Testing

- **Animation or simulation –** The system does not support animation or simulation for testing the threat/hypothesis. However, for this particular system, animation is not much help.
- **Re-visualizing the hypothesis –** The system does not support any feature that can re-visualize the hypothesis to help analysts test the threat/hypothesis. The system is developed to help analysts to understand the general trend of a large amount of documents. It does not help analysts to hypothesize.
- **Comparing hypotheses (possibly with other analysts' hypotheses) -** The

system does not support any feature that can help compare a number of hypotheses to help analysts test the threat/hypothesis.

- **View evidence -** The system does not support any feature that can help view evidence of a hypothesis/threat to help analysts recognize the possibility and the level of relevance of the threat/hypothesis.

Ranking the Threats/hypothesis (Threat Analysis)

- **Relevance of hypothesis/threat –** The system does not support any feature that can help recognize the relevance of hypothesis/threat to help analysts rank the threats/hypothesis.
- **Evidence credibility –** The system does not support any feature that can help show evidence credibility, which is the number of supporting evidences subtracted by the number of contradicting evidences, on a hypothesis/threat to help analysts rank the threats/hypothesis. However, the system may be able to help analysts to collect evidence using brightness of stars in the galaxy view.

HCE3

Finding general trend(s):

Visualizing data

- **Maximum input size –** One dataset at a time
- **Maximum output size –** About 10,000 items in color mosaic

  The overview, which is done in color mosaic and dendogram can display up to about 10,000 items in color mosaic and much more with dendogram. The number was calculated considering the size of a regular monitor, 1024 by 768 pixels.
- **Time to navigate (getting detailed information) –** It uses "Zoom" method

with a scroll bar. Zooming in/out also effects the length of the bar inside the scroll bar. As the length of the bar gets longer, the more zoomed out it is and vise versa.

- **Geospatial and Temporal Representation –** The system does not support any feature to track movement according to time.
- **Social Representation –** The system does not have features for social analysis. There is a possibility that the dendogram can turn into social representation.

Explore/Navigate

- **Efficiency –** It uses "hammer" methods. It reduces visual representation to fit in all the data for two overview representations, dendogram and color mosaic. With HCE, a large dataset, such as IA dataset, the overview gets too abstract for users to read and understand.
- **Effort –** The system offers a filtering feature. Two filtering options are available, present call filtering (only for Affymetrix GeneChip Experiment data) and standard deviation filtering. The standard deviation filtering is a basic filter function. The user can filer out rows based the standard deviation. The idea is to filter out items that do not change over time. For instance, genes that do not quite change over the samples or time points will get filtered out.
- **Extra Navigation Feature –** The system offers many different views, such as table view, histogram ordering, profile search, evaluation tab, and K-means to help user explorer the dataset. It reads in multidimensional datasets and display user interactive visualizations, such as color mapping, dendogram , and color mosaic. Hierarchical Clustering results are presented as dendogram. Dendogram is a binary tree, which relationship and the closeness of items in the graph. Color Mosaic is for multidimensional Dataset. It color-codes each value of the original table. It is also known as

heat map or patch grid. Scatter plot ordering provides the user with a quick glance of the relationship between two attributes. There are other views, such as table view, histogram ordering, profile search, evaluation tab, and K-means to help users explore the dataset.

Re-visualizing

- **Multiple views –** The system allows analysts to engage multiple views of different presentation forms (table view, histogram ordering, profile search, evaluation tab, and K-means), to let analysts to observe in a different perspective.
- **Theorizing –** HCE3 offers a dynamic query using a slider bar to hide or filter out the undesirable data.

Specific threat or second step after finding general trend(s):

Threat Detection/Brainstorming

- **Manipulating/Alternating Data and View–** The system does not allow analysts to manipulate and alternate data and/or view, which can help brainstorming and hypothesizing. Table view provides a nice platform to change data.
- **Gathering supporting evidences –** The system does not include any feature which allows analysts to gather supporting evidences for their hypothesis. However, gene ontology provides an annotation feature, which can be used to gather supporting evidence and to theorize.

Threat/Hypothesis Testing

- **Animation or simulation –** The system does not support animation or simulation for testing the threat/hypothesis. However, because this system has the potential for theorizing/hypothesizing, an animation or simulation will be a great aid for analysts to go over their threat/hypothesis.

- **Re-visualizing the hypothesis –** The system does not support any feature that can re-visualize the hypothesis to help analysts test the threat/hypothesis. However, if the table view has the capability of entering new data/hypotheses, it will be able to visualize the hypothesis.
- **Comparing hypotheses (possibly with other analysts' hypotheses) -** The system does not support any feature that can help compare a number of hypotheses to help analysts test the threat/hypothesis. Currently, the system only allows one dataset to be visualized. In order to compare hypotheses, multiple numbers of dataset must be visualized.
- **View evidence -** The system does not support any feature that can help view evidence of a hypothesis/threat to help analysts recognize the possibility and the level of relevance of the threat/hypothesis. However, with gene ontology view and table view, examining evidence may be possible.

Ranking the Threats/hypothesis (Threat Analysis)

- **Relevance of hypothesis/threat –** The system does not support any feature that can help recognize the relevance of hypothesis/threat to help analysts rank the threats/hypothesis. However, scatter plot ordering has an excellent ordering feature. If this can be integrated with gene ontology, the order of relevance of hypothesis/threat can be properly visualized.
- **Evidence credibility –** The system does not support any feature that can help show evidence credibility, which is the number of supporting evidences subtracted by the number of contradicting evidences, on a hypothesis/threat to help analysts rank the threats/hypothesis. If gene ontology's annotation feature and table view can be integrated, it may be able to visualize evidence

credibility.

To summarize the findings above, visualization tools commonly lack features that can help social network analysis, hypothesis creation, hypothesis comparison, and hypothesis testing - issues seen as important to the intelligence community. In order to have an effective intelligence analysis tool a number of visualization methods such as network, hierarchical, spatial, and temporal methods need to work together. Chapter 7 discusses on the findings of the five tools in detail.

Chapter 7. Discussion

This section discusses the revealed lessons and key findings from the experiments' results and contribution of this evaluation approach and thesis. This evaluation approach lets the evaluator assess information visualization tools according to the intended use instead of measuring all the aspects of the tools, including never and rarely used features. The approach focuses on key activities of the domain users and measures information visualization tools only according to the key activities, which will help guide and evaluate information visualization tools according to their intended use.

Key findings on the evaluation approach

This evaluation approach can be used for different purposes. First of all, it can be used to evaluate visual analytic tools for intelligence analysts. The approach developed and applied in this research effort can expedite the process of evaluating visualization tools for intelligence analysts and illuminate tools' potential strengths and limitations to help select the best-fitting tools. While this evaluation approach is considered a laboratory environment study, it employs information gained from the real work environment (result of analyzing key intelligence domain activities).

By applying domain knowledge as described, it is believed that this evaluation approach can deliver results similar to field/longitudinal studies. However, there are two key reasons that this approach is not sufficient to fully substitute for field studies/longitudinal studies (which is considered to be the most accurate but most expensive evaluation method). First, this masters level research effort over a relatively short span of time and there have not been a sufficient number of studies to make full claim to the impact of this approach. Second, there have not been field/longitudinal studies to compare the result on the same dataset. However, since the activities' analyzing process requires the work environment knowledge and the evaluation approach employs the knowledge, the evaluation approach appears to respond to the limitations of short-term evaluations while strengthening the short-term evaluations by reducing the evaluation time – a key objective of the research effort.

The second purpose has to do with shortening the evaluation process. Due to the smaller number of criteria, this approach's evaluation process itself has been shortened compared to existing evaluation methods. Since the approach can be done relatively quickly, it can be used as a filter-evaluation or pre-evaluation, setting up for longer and more in-depth evaluations for more selected tools or technology. While there is a cost for the pre-conditioning of criteria to align with the domain activities, this is quickly recouped when the number of relevant tools is sifted down to a manageable number.

Lastly, the evaluation approach can provide a general architecture or guideline for improvement for the domain area of study. In this case, it can be used as a guideline for an intelligence analysis purpose tool. It shows the essential features for intelligence analysts to work efficiently. While initially most analysis environments do not recognize areas where visualization tools would be most helpful, the activity analysis may shed light on how the visualization support would improve the performance of the analysts; hence providing guidelines for improvements.

All of the tools that were evaluated were not developed specifically for intelligence analysis purposes – few commercial or publicly available research tools are accessible. Besides meeting some of the criteria from the evaluation, they were largely chosen based on the general popularity, availability of the tools, and tools' information visualization method. Due to this fact, the results show lack of specific support for intelligence analysts' intended use. It is unfortunate that intelligence analysis tools are not widely available. In this study, as the approach was illuminating the tools' potential strengths and limitations, it suggested a guideline to strengthen the positives and to prevent/ameliorate the limitations.

Another lesson throughout the evaluation was that each tool's strength(s) lay in different aspects. Even though these tools were not developed specifically for intelligence analysis, combining the tools with little modification can provide many needs for the domain. Spotfire provides familiar types of graphs with a good navigation/query system. It shows potential for representing geospatial information and comparing hypotheses. Time Searcher 2 can be used to analyze temporal data and it may be potentially useful to have animation features to visually present the sequence. Tablelens is good for showing a general overview, because it can show all the attributes in one graph. Many of the initial intelligence analysis data may be in a large number of documents; In-Spire can assist analysts in figuring out general trends and outliers of a large number of documents. HCE3 is helpful in observing dataset in many different perspectives and it shows a potential for gathering evidence. The detailed information for this assertion is in key findings (strengths and limitations) within the tools section.

During this investigation, there were a number of strengths (of this approach) recognized. However, there were also a number of limitations. First, the most notable limitation is the analysis process. In the beginning of this approach, unlike other existing evaluation methods, evaluators are required to analyze the application domain and identify key visualization intensive activities for the domain, as well as a proper way to measure them. This process can be varied in its conclusions. Two methods have been considered for this process. The first method is accomplished by surveying domain experts through a questionnaire. The second method is through observation through literature and interaction with third parties with relevant exposure to the domain.

In this thesis, only the observation method has been applied. Unfortunately, actual intelligence analysis work could not be directly observed, so indirect observation through published papers and reports were conducted to verify key activities. Subsequently, criteria to measure these newly acquired key activities were developed and tested (separately from evaluation). This whole process required the most effort and time during the study (note Figure 6.1 in Chapter 6). However, once the criteria have been established, the rest of the process was easy and straightforward compared to existing evaluation methods. Of course, this would be even more expedient once a routine process was established to conduct such surveys and observations.

In Appendix A, most of the same tools have been evaluated either by one of the standard evaluation methods currently used in the industry (note Appendix B) or an evaluation method that is a composition of the standard industry evaluation method and an academic evaluation method (note Appendix C). These evaluations were conducted previous to this study; they are being mentioned for comparison purpose. As intended use evaluations were being conducted, the author tried to remain subjective, despite the fact that most of the tools have already been evaluated once before. Even after acknowledging the possible bias that may lie in the study, this intended use evaluation approach shows a reduction in evaluation time, once the analyzing process has been completed. As the number of evaluation increases, the property of intended use evaluation approach can be maximized as Figure 7.1 illustrates.

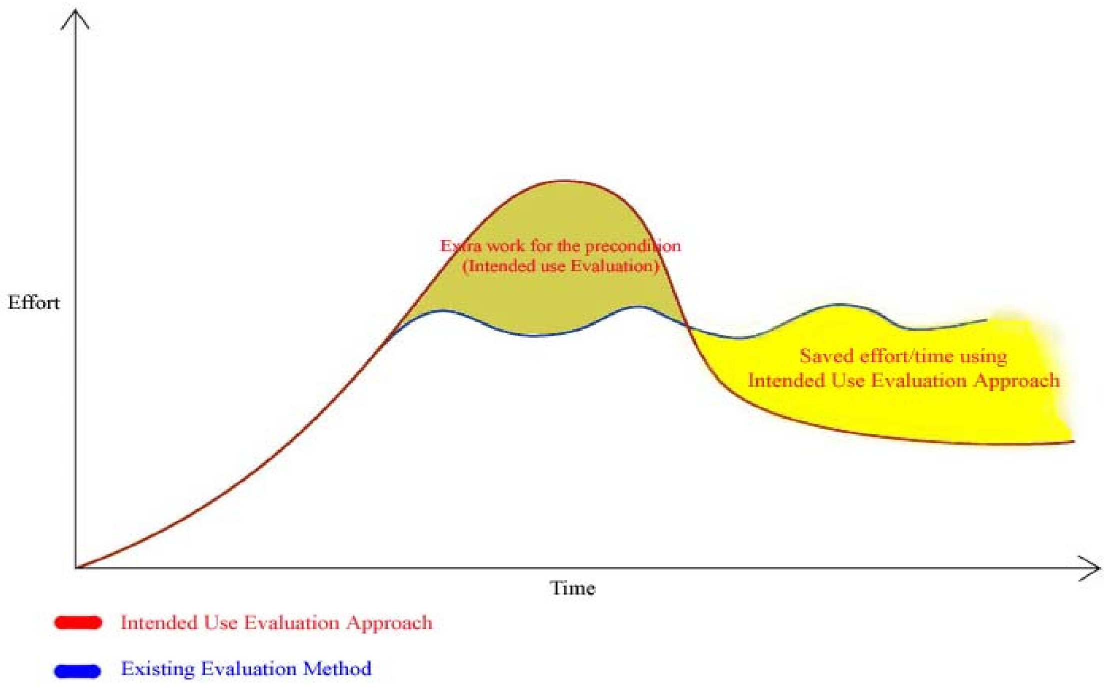


Figure 7.1 Conceptual Graph Illustrating Tradeoff between Up-front Preconditioning and Substantial Reduction in Time to Conduct the Evaluations

This graph shows the general trend of the approach in terms of time and effort (note effort can be thought of as temporal cost in this sense). As more evaluations are being conducted, more time and effort can be saved as the effort diminishes and the calendar time ends sooner. That is, due to the saved time in each evaluation, calendar time for a project will be saved as well. The total savings can be calculated by subtracting the precondition effort from the saved effort conducting the evaluations of the tools (see the graph above).

The approach is expected to reduce the risk in evaluation by preventing missteps. This is similar to how doing the requisite requirements analysis reduces the risk of rework in software development. The preconditioning narrows the criteria to those that are a priority for the domain upfront; thus reducing the need to evaluate aspects of the tools that are note relevant to the selection decision.

Analyzing the key activities early on will serve as a guide to create evaluation criteria. Due to the information gained early on, missteps will be prevented, which reduces the risk factor. This stabilizes the confidence in the outcome of the evaluation relatively early in the evaluation process.

The second limitation of this approach is that problems may co-occur with reduced evaluation time. By expediting the evaluation process, the learning time for novice users is reduced. For other experiment studies on top of the learning time, users learn more while testing tools. Shortening the evaluation process may create a possible neglect on any less superior, less used, or incorrectly labeled or placed feature for novice users. However, this should not be a problem for an expert evaluator.

The last limitation is that the evaluation approach suffers from the same problems as the previous evaluation method. Depending on the evaluator's opinion, the evaluation result may differ and could be subjective. Since this evaluation approach is based on metrics evaluation, it inherits the same limitation of that method.

Summary of the key finds of the evaluation approach

Key Strengths

1. A key short coming of many tool evaluation methods is that they rely heavily on a general set of criteria by which to determine the success or failure of a tool to meet the needs of a known group. It was chosen in this approach to conduct some early analysis to determine the activities that will use the visualization technology, and to use this to calibrate the criteria towards these first-order needs. With a little effort up front, there is a substantial reduction in time to conduct the evaluation.
2. The evaluation criteria may serve as the architecture of guideline for the domain area of study.

Key Limitations

1. The evaluation approach requires evaluators to analyze and find the most important activities for the domain, as well as a proper way to measure according to those key activities in the beginning of the process.
2. The evaluation approach may reduce the learning time for novice users.
3. The evaluation results may be subjective based on the evaluator's opinion.

Key findings (strengths and limitations) on the tools

It is important to note that tools, by the nature of the markets that they serve, are seldom able to fully serve a given customer. That is, they will be good at more general things, unless they are built specifically for the domain in which they are used. Visualization tools commonly lack features that can help social network analysis, hypothesis creation, hypothesis comparison, and hypothesis test - issues seen as important to the intelligence community. In order to have an effective intelligence analysis tool, a number of visualization methods, such as network, hierarchical, spatial, and temporal methods need to work together. Among the five tools that were evaluated, only Time Searcher 2 showed effective temporal data analyzing capability.

Next, the strengths for each tool were observed. The evaluation results show that each tool performs better for different areas. Spotfire provides familiar types of graphs with a good navigation/query and multi-view system. Time Searcher 2 shows effectiveness in analyzing temporal data and decent multi-view. Tablelens is good for showing a general overview, because it can show all the attributes in one graph. In-Spire can be effective in figuring out general trend, outliers, and main subject or time for intelligence analysts, because it reads a large number of documents like many of the initial intelligence analysis data may be. HCE3 is useful for observing dataset in many different perspectives.

After observing the strengths for each tool, the potential usages and future implementation guideline, if needed, were observed. Spotfire shows a potential use for representing geospatial information and comparing hypotheses. Both Time Searcher 2 and Spotfire show potential use for animation to illustrate, test, and rank threats/hypotheses. Also, Time Searcher 2 shows a potential capability for theorizing. HCE3 shows a good potential for gathering evidence and evidence credibility.

From the insight gained from the evaluations, each tool showed current and potential strengths for different intelligence activities. These strengths naturally led to thinking of integrating these tools into one intelligence analysis tool. In order to integrate tools, a common integrating platform is needed to pass common data format. Also, a social analysis feature is needed.

Contributions

The hypothesis of this study is that if visualization technologies can be evaluated according to their intended use, then a more accurate and timely selection of the appropriate tool and/or tool evaluation can be made. Intended use corresponds to a set of tasks that make up the core activities that are best supported by visualization in the operational environment. The set of tasks will be based on the key activities that the user finds most important and/or most frequently faced. This study has focused largely on intelligence analysis information visual analytic tools recognizing that the principles may generalize to the wider scope of visualization technology application domain. This is true especially to other analytical environments, such as business analytics and scientific visualization, because they share similar activities. To demonstrate the evaluation approach, an evaluation framework has been established based on existing work in the visualization community. The framework has been applied in calibrating the criteria toward these first-order needs later in the study. Succeeding the framework, identifying the intelligence analysis work process has followed. During the process, key activities in intelligence analysis community and evaluation criteria for intelligence analysis tools based on tools' intended use/key activities using the framework have been identified. Applying the criteria, five evaluations have been conducted. Subsequently, the evaluation approach has been compared with two existing evaluation methods in terms of evaluation time.

The following is the list of contributions of this study:

1. Established a visualization technology evaluation framework based on existing work in the visualization community, and utilized a calibration effort up-front to condition the criteria for better results in the longer term.
2. Identified intelligence analysis work process and key activities with respect to visualization supported activities.
3. Identified requirements needed for intelligence analysis visual analytic tools.
4. Established evaluation criteria for intelligence analysis tools based on tools' intended use/key activities.
5. Identified the strengths and weaknesses of this evaluation approach by comparing it with two existing evaluation methods.
6. Five tool evaluations were conducted, and the strengths and weaknesses of tools for intelligence analysis community were identified.

Chapter 8. Conclusions and Future work

This section discusses major conclusions for the visualization community and a number of further researches that can be applied for this study.

Conclusions

The study suggests the following major conclusions for the visualization community. In selecting and/or evaluating visual analytic tools, a little up front effort to analyze key activities regarding the domain field will be beneficial. Such analysis will substantially reduce evaluation time and effort throughout a longer period of time. In current practice, many evaluations rely on usability and control experiments [2]. In these methods, a general set of criteria determines the success or failure of a tool. However, since each tool was developed for a different intended use, it will serve differently. Therefore, each tool needs to be evaluated accordingly. In the described approach, the key activities that will use the visual technology, to calibrate the criteria toward these first-order needs were analyzed. As mentioned, this approach showed a reduction in time to conduct the evaluation. However, this requires more effort in the beginning of the process and it may reduce the learning time for novice users. This evaluation approach becomes more practical when there are more evaluations to be conducted.

Future Plan

This study has focused largely on intelligence analysis information visual analytic tools recognizing that the principles may generalize the wider scope of visualization technologies.

First, this evaluation approach needs to be applied to more intelligence analysis tools. So far, only a small number of visual analytic tools have been tested with this evaluation approach. Without actually evaluating intelligence analysis tools, it has been difficult to accurately assess the confidence level for this approach.

Second, to help build-up the confidence level for this approach, the result of this study needs to be compared to the result of field/longitudinal studies. There are a lot of both big and small variables that can affect the work in the real world environment. Many variables are considered to be accommodated with the activities' analyzing process. However, to prove the quality of this evaluation approach, the result of this study needs to be compared to the result of field studies/longitudinal studies.

Third, more evaluation needs to be conducted. In this study, only five evaluations have been conducted by this evaluation approach. With this approach, evaluation time and effort have been reduced. However, it was difficult to assess the exact reduced time because the results varied from tool to tool and from evaluation method to method. This study can be more effective if more studies can show the reduced time and effort that supports this approach.

Fourth, to reduce one limitation of this approach, an evaluation approach may provide an ordinal scores and strict guidelines for evaluators to follow. This process may reduce the subjective nature of the approach. The guideline should help the evaluator to score each evaluation criteria and overall score of visualization tools.

Last, the study started out recognizing that the principles may generalize to the wider scope of visualization technology application domains. However, the study focused on intelligence analysis information visual analytic tools to demonstrate the strength of the evaluation approach. To demonstrate the principle's capability in a wider scope of visualization technologies, studies demonstrating different domains of visualization need to be conducted.

Appendix A

Spotfire

1. Technology Information

Tech. Name	Spotfire Pro
Version	4.0
Description	Spotfire provides an easy to use environment to analyze a large amount of data.
POC	Customer Support: support@spotfire.com Phone: (617) 702-1600 Fax: (617) 702-1700 Toll-Free: 1800-245-4211
Website	http://www.spotfire.com
Pricing	The Spotfire 4.0, which is an old and tested version is free. The price range of the new version, Spotfire DecisioSite 7.2, is from $35,000 to $60,000 per year for typical configurations supporting five to twenty five users.
Requestor	RDEC team
RDEC	PP
Experiment Focus	Paper evaluation for the purpose of understanding
Status	Completed
RDEC POC	Albert Park
Installed on	Fall 2005
License Expiration	Evaluation copy (Spotfire 4.0) – no expiration.

2. Overview

   a. Screenshot – has been deleted due to sponsor's request.

   b. Observed Functionality

   Spotfire provides four different 2D visualizations and one 3D visualization with which most people are familiar.

   c. Features

   Spotfire offers the following built-in features:

      i. Scatter Plot

      It is probably the most used feature. It allows users to glance relationships between two attributes.

      ii. 3D Scatter Plot

      It is not as intuitive as 2D scatter plot and contains some navigation problems. It allows users to see a relationship between three attributes.

      iii. Histogram

      A 2D representation that allows users to measure aspects of one attribute versus one other attribute.

      iv. Bar Chart

      Similar to Histogram, enables users to compare several alternatives against a given criteria.

      v. Pie Chart

      The information is presented in a circular view (pie) with the percentage of each item/column with respect to the total. The main disadvantage is that only one attribute can be shown for one pie chart.

vi. Multi-view Window

It enables users to view multiple windows of analyses at a time. For example, a window with geospatial information many be coordinated and viewed at the same time as a social network analysis diagram to ascertain the connections between suspects from geographic perspective.

d. Overall Quality

There were no defects or errors detected during the evaluation. It would be helpful to obtain reliability information on this tool. The tool appeared to be robust with respect to the data input to the system for investigation. From a maintainability perspective, the tool appears to have sufficient technical support and update mechanisms in place (both tech support numbers and website for updates, etc.).

e. Suitability for further testing

No. Spotfire provides users with a quick insight about the dataset by comparing one, two, or three attributes at a time. Even though it shows its limitation in number of attributes, it makes using the conventional graphs very easy. The fact that everyone is very familiar with those 2D graphs adds a big advantage to the software. However, the 3D Scatter Plot is not very useful. It is awkward and hard to navigate. Reading the data from the 3D graph is challenging.

The software is more for simpler dataset. As mentioned above, it can only visualize one, two, or three attributes and ignores the rest. It does not fit well with Intelligent Analysis Dataset. It is not a stand-alone product for Intelligent

Analysis field, and it does not provide much depth for a complex multidimensional dataset where each attribute has a big impact on other attributes.

f. Potential Use

The software can be used to find out direct relationships among two or three attributes. It can be used as an additional feature for the system when the analysts need to have detailed information with few attributes.

3. Hardware/Software Required/Supported

a. SW Components Required

Taken from: Spotfire release notes

To run Spotfire Pro 4, OLE DB drivers from Microsoft need to be installed. These can be found on the Spotfire Pro installation CDROM, on the Spotfire web site at:

http://www.Spotfire.com/support or at the Microsoft web site at:

http://www.microsoft.com/data/

b. HW Platforms Supported

Windows 95 - OSR 2

Windows 98

Windows NT 3.51 - Service Pack 5

Windows NT 4.0 - Service Pack 2

4. PP Experimentation

a. Experimental Approach

This evaluation was a paper evaluation for the purpose of understanding the product. Although the evaluation involved using/testing built-in features with the sample data that comes with the software and protected data, there was no measuring or comparing with different products.

b. Data Set used

Used a common dataset throughout the experiments; however, description of dataset has been deleted due to sponsor’s request.

5. Quality of Software

a. Navigation Strategies

It uses “Zoom” method.

Zooming in/out also effects the length of the bar inside the scroll bar. As the length of the bar gets longer, the more zoomed out it is, and vice versa. The size of the data on the graph stays the same. Since there is not any need to view large dots, it works fairly well.

b. Interaction Strategies

Mouse + Keyboard

A conventional mouse and keyboard is used.

6. Usability

a. Ease of learning (Ease of obtaining proficiency)

Excellent. Even though there is a well-written manual, there is no need to open it. The Spotfire is easy to learn to use. The basic features are very intuitive, and more can be learned through the manual for experts. For the Intelligent Analyst adding images such as a map, background information may be very helpful.

b. GUI ease of use

Excellent. Particular focus was put into the ease of usability so even people with basic knowledge of graphs can easily operate. All one needs to use is the basic menu, which is located at the top the screen. There are also short keys and settings available for experts.

c. End-user documentation

Good. A very well-written 280 page manual comes with the software. There are a lot of pictures to guide users step by step. The only drawback is the fact that the manual is too detailed and too long.

d. Freedom from problems and issues

Good. No major problems were encountered. There was only one crash in the course of testing, but it did not hold up the system. The software got terminated immediately. The only problem was that it did not save or recover.

7. Performance

a. Data Scalability

< 1,000,000 items & < 10 attributes

b. Response Time

Excellent. Once the data were uploaded, the response time was even better. A new graph was generated at the release of the mouse button. The response time during the navigation, meaning zooming in/out, was excellent as well.

c. Number of concurrent users

Single user

8. Software Set-up / Maintenance

a. General set-up information

Good. It installs using a standard Windows installer.

b. Ease of software maintenance

Unexplored. Only one version of the software was found on the web.

9. Appendix

   Not included for this thesis

10. Attachments

   Not included for this thesis

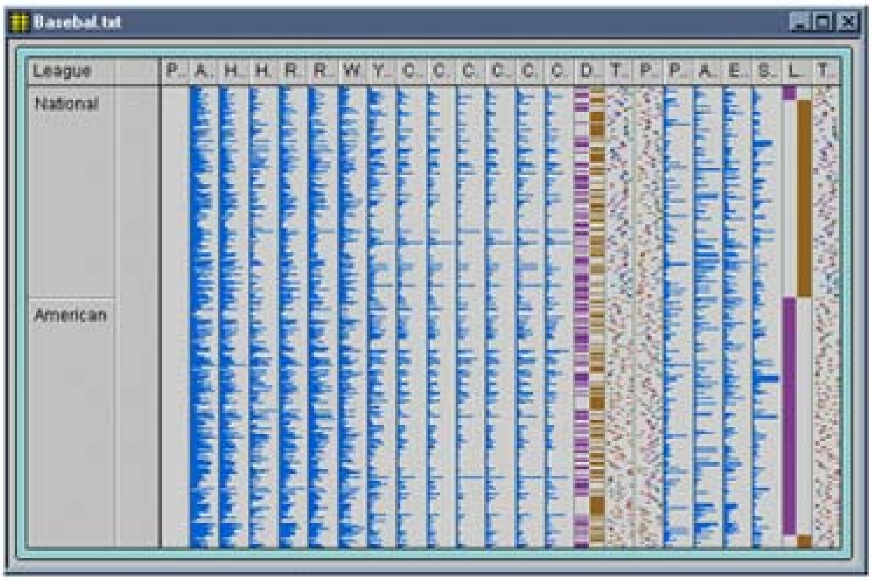
Tablelens

1. Technology Information

Tech. Name	Eureka (Table Lens)
Version	1.2
Description	Eureka (a.k.a. Table Lens) can visualize a large amount of multi-attribute dataset to spot a general trend or correlation.
POC	General Inquiries: Phone: (408) 738-6200 1888-414-4949 Email: info@inxight.com Sales and product information: Phone: (408) 738-6299 Email: sales@inxight.com
Website	http://www.inxight.com/products/sdks/tl/
Pricing	Inquire
Requestor	RDEC
RDEC	PP
Experiment Focus	Paper evaluation for the purpose of understanding
Status	Completed
RDEC POC	Albert
Installed on	Fall 2005
License Expiration	7 days trial version – Expired

a. Screenshot

Taken from: http://www.inxight.com/products/sdks/tl/


Basebal.txt
League
National
American
P.. A.. H.. H.. R.. R.. W.. Y.. C.. C.. C.. C.. C.. C.. D.. T.. P.. P.. A.. E.. S.. L.. T..

2. Overview

a. Observed Functionality

Tablelens allows users to visualize a large amount of multi-attribute dataset to spot a general trend or correlation. One of the biggest advantages is that Tablelens allows users to observe all the attributes in a tabular form.

b. Features

Taken from: http://www.inxight.com/products/sdks/tl/

i. **Sort** by clicking on columns.
ii. **Rearrange** columns by drag-and-drop.
iii. **Promote** columns to create subgroups.
iv. **Focus** by clicking on a cell or by clicking and dragging to "focus" a whole row or multiple rows.
v. **Filter** subsets to create smaller more specific datasets.
vi. **Spotlight** data (rows/columns) to track particular information as you sort.

c. Overall Quality

There were no defects or errors detected during the evaluation, except there was a problem opening up a dataset that was too large. It would be good to obtain the exact size of the dataset it can handle. The tool appeared to be robust and reliable. The only concern is the difference between the trial versions and the newest version. It seems there has been a great deal of upgrading. Due to the fact that the trial version was only available for a week, not enough testing was done to the software during this time. Deleting the expired version and installing new trial version did not work.

From a maintainability perspective, the tool appears to have sufficient technical support and update mechanisms in place (both tech support numbers and website for updates, etc.). Many newer versions have evolved from the first program, which indicates it has good maintainability.

d. Suitability for further testing

No. The software is for simpler datasets. Even though it visualizes multidimensional dataset, it does not fit well with Intelligent Analysis Datasets. Tablelens can provide users with a quick insight about the dataset itself and the general trend of it. The software is well-designed to display the entire dataset in a manner that is very intuitive to understand. Displaying the entire dataset, users can obtain the general or unexpected relationships. Even though it shows all the attributes at once, it is not so easy to control and compare more than three or four attributes at a time. The zoom in the selected data with overview feature made navigation easier but does not fit well with Intelligent Analysis. It is not a stand-alone product and it does not have the ability to provide analysts with IA insights

e. Potential Use

The software can be used as observing the overview and general trend of data. If there is a strong general relationship in a large amount of data, this software can be very useful.

3. Hardware/Software Required/Supported

Taken from the release note: The recommended system requirements for Eureka 1.2 are as follows:

Processor	Pentium 166
Memory	32MB
OS	Windows 95 (OSR-2 or higher, with IE 4.0 or higher)
	Windows NT 4.0 (Service Pack 3 or higher)
	Windows 98
	Windows 2000

4. PP Experimentation

   a. Experimental Approach

      This evaluation was a paper evaluation for the purpose of understanding the product. Although the evaluation involved using/testing built-in features with the sample data that comes with the software and protected data, there was no measuring or comparing with different products.

   b. Data Set used

      A common dataset was used throughout the experiments; however, a description of dataset has been deleted due to sponsor's request.

5. Quality of Software

   a. Navigation Strategies

      It uses "overview + detail" and "Zoom" method.

      The program starts out with just an overview of the dataset. Depending on the size of the dataset, it might be aggregated or not. When an item is clicked, the detailed information will be shown along with an overview of the dataset. Multiple numbers of detailed information can be shown at a time. Highlighted information can be opened up at the same time. The detailed information can be closed by clicking back the detailed information.

b. Interaction Strategies

   Mouse + Keyboard

   A conventional mouse and keyboard is used.

6. Usability

   a. Ease of learning

      Excellent. Even though there are two well-written manuals, there is no need to open them. The learnability of the Tablelens is easy. As long as the users have a little computer experience and remember the fact that the numbers in the table have changed into a bar graph, they will have any trouble. The basic features are very intuitive, and more can be learned through the manual for experts.

   b. GUI ease of use

      Excellent. It followed the standard format. The basic menu is located on the top of the screen and the mouse is all one needs to use. There are also short keys and settings available for experts.

   c. End-user documentation

      Good. There are two Help files included in this release. ("Tutorial.pdf" and "UserGuide.pdf"). The tutorial is for novice users who are just getting started to use the software. It is a 32- page document. The user guide has detailed information about the software, and it is a 118-page document. Both of them contain a lot of picture for the users. The only drawback is the fact that the user guide is too detailed and too long.

   d. Freedom from problems and issues

      Good. No major problems were encountered. The only concern is that the trial version is only available for only a week, which is not enough time to test any software properly. Deleting the expired version and installing a new trial version did not work.

7. Performance
   a. Data Scalability

      < 1,000 items & < 50 attributes

   b. Response Time

      Excellent. It took hardly any time to upload the test dataset. Once the data were uploaded, the response time was even better. As soon as the mouse button was released, the detailed information was opened/closed. The response time during the navigation and the scrolling up and down and changing of positions, was excellent as well.

   c. Number of concurrent users

      Single user

8. Software Set-up / Maintenance
   d. General set-up information

      Good. It installs using a standard Windows installer.

   e. Ease of software maintenance

      Unexplored. Only one version of the software was found on the web.

9. Appendix

   Not included for this thesis

10. Attachments

Not included for this thesis

In-Spire

Tech. Name	IN-SPIRE
Version	2.0
Description	In-Spire uses a clustering algorithm to generate a visualization of maximum of 200,000 through "Theme view" and "Galaxy view." It helps to give users a quick overview of a dataset. It can further contribute to reveal grouping, similarities, differences, and unanticipated relationships within the document collection.
POC	Lisa Glasford In-Spire administrator Phone: (509) 372-4717 Fax: (509) 375-2091
Website	http://in-spire.pnl.gov
Pricing	Commercial License: • In-Spire License: $15,000.00 per workstation. • One year maintenance: $3,750.00 • Support/Training: $1,500.00 per token. For U.S. Federal Government Use: • $500.00 Service fee per license that covers costs of distribution and administration.
Requestor	RDEC
RDEC	PP Plus
Experiment Focus	A more elaborated PP evaluation by measuring and testing more features using protected dataset.
Status	In process – testing how the software can be helpful to the analysts.
RDEC POC	Albert Park
Installed on	December 2005
License Expiration	90 Days.

a. Screenshot – has been deleted due to sponsor's request.

TECHNICAL: FUNCTIONAL

1. Platforms and Operating Environments

   System Requirements (taken from: http://in-spire.pnl.gov/getacopy.html)

   In-Spire will run on a PC with Windows 2000 (SP3 or later) or Windows XP. There are no minimum requirements for RAM, clock speed or disk space; however, the better the resources, the more quickly the analysis can be performed. Disk space requirements are driven by size and number of datasets needing to be analyzed. Roughly two or three times as much disk space as the total size of the datasets planned to be analyzed is needed. Currently, the maximum dataset size is about 200,000 documents at a time.

2. Input Methods

   It uploads dataset that is more than fifteen documents and less than 200,000 documents using mouse and keyboard.

   a. Input devices supported

      A traditional mouse and keyboard is used.

   b. Input formats supported

      It can upload both ASCII dataset and XML dataset. There were troubles opening XML type dataset, even with a detailed guide and following the guide step by step. The system should be able to handle ACSII Dataset, FBI Portal Harvest previous, FBI Portal Harvest new, Fall 04, Google Harvest, Web Harvest, and XML dataset are supported; however, only ASCII format was tested.

   c. Limitation of input data (size, volume, etc.)

      It uploads dataset that is more than fifteen documents and less than 200,000 documents using mouse and keyboard.

3. Major Output Produced

   The output presents in a form of a map of collection, which is shaped by major themes and sizes, relationships between themes, and the closeness between themes. It uses the "grocery store" concept; lay out similar documents near each other and dissimilar documents far apart. The negative aspect of this visualization is identifying the documents with multiple themes and determining the exact location on the map. For this type of document, it is not clear how to look for the document; a document can be located either in both clusters or in between.

   a. Support for metaphors (creation, application, transition, etc)

      Theme view – This uses a map of natural terrain to represent a set of documents. The height of the peaks indicates the relative strengths of the topics in the document set. Brightness has a linear relationship with the height. The higher the peak, the brighter it is. The color is to show dominant theme combinations.

      Screenshot – has been deleted due to sponsor's request.

      Galaxy view - This uses the image of constellations to represent a set of documents. It is in 2D representation; clusters can be formed depending on the similarity. It has same spatial arrangement as the Theme View. It groups similar or related documents near one other, which can be turned into a cluster of stars and dissimilar documents far apart.

      Screenshot – has been deleted due to sponsor's request.

b. Output formats supported

A Gist view (in galaxy view) - allows users to find desired information in the documents.

Probe (in theme view) – allows users to find the desired words and displays in bar graphs.

c. Handling Clusters

Theme view – it creates a 3D map in which the clusters are displayed as peaks. The height of the peaks indicates the relative strengths of the topics in the document set. Brightness has a linear relationship with the height. The higher the peak, the brighter it is. The color is to show dominant theme combinations.

Galaxy view – the clusters are displays as actual clusters of stars. It groups similar or related elements closely together to form a cluster. Unrelated ones have distance between them. The knowledge contributions decide how closely the documents are related; therefore, the size and eventual brightness of the star cluster will represent the size of the document cluster. The closeness represents the closeness in topics.

d. Multi-view

Multi-view is supported, but it does not have synchronized functionality. As one area of the graph representation gets highlighted, there is no change in the other graph.

4. Methods Supported

a. Virtual Reality (or Virtual Environment)

4.1.1 Immersive Environment

b. Animation

No, it does not have an animation feature. There is a time slider, which shows the status and number of the documents according to the time. Combining this feature with one of the visual representation can be a good animation. An animation can be helpful to spot a general trend over time.

c. Collaboration

This is single-user software, no collaboration is available.

d. Graphical modeling

The two graphs, Theme View and Galaxy View, are available. There are many features to aid the understandability of the graphs, but building mock views, metaphor editing, dynamic or stochastic modeling are not available.

5. Major Visualization Features

a. Navigation

Most navigation is done with zooming and panning.

b. Transition between graphical representations on Multi-View

Instead of closing one window and opening another window, it keeps the old window and opens up a new window. No transition is made.

c. Controlling interactions with the Visualization

5.3.1 Result set control

The number of clusters, outliers, patterns, and maximum number of results produced cannot be controlled. The whole visual can be moved, but no features are available to alter the data value from the graph. The recalculate button is available to rearrange the layouts; the system does not allow users to have full control over the output.

5.3.2 Display options

Result set control

Highlighting part of the layout can be recalculated and rearrange, but the system does not allow users to have control over the number of clusters, outliers, patterns, or maximum number of results produced.

Display options

Only two types of displays are available. Undesirable data/words can be filtered out.

5.3.3 Force change

No change can be made to the output result. The system allows users to recalculate the highlighted section of the output, but relationship, outlier, patterns cannot be changed.

5.3.3.1 Automat synch with “forced” changed by user

N/A

d. Tracking

The selected data can be tracked with a color scheme. The system allows one to track selected data or group of data by coloring them; the color stays even after each change/refinement, time step and processor action.

e. Searching

A key word query is available on any terms. A set of documents in the corpus are returned as a result. This helps to find specific document(s). It reduces the searching/browsing time and returns according to users' interests. However, finding the right query key word may be a problem.

f. Slicing

Highlighting and coloring allows one to select or to deselect data to simplify or to help understand the output.

g. Illustration by time and location (for geospatial information)

Time pattern can be observed through Time Slicer. It lets the user observe the relations between themes and time. Documents stay in the same spatial location but are shown or hidden according to the time.

h. Hierarchies/Grouping

It groups the clusters automatically. It does not give users the freedom of changing, creating or regrouping the hierarchies/clusters.

TECHNICAL: NON-FUNCTIONAL

6. Completeness of the Software

a. Accuracy

Unexplored.

b. Robustness

Excellent. No major problem was encountered during the evaluation. The software did not crash or hang. It is a very robust product.

c. Integrity

This is a stand-alone product. It does not seem like In-Spire was designed to be part of a larger system or integrate with other software.

7. Usability
   a. Understandability

      Good. The two visual representations are easy to understand. While the two visualizations help each other to present more concrete information to users, there are other features to clarify the visualizations. Gist view is available as one of the features of galaxy view. It allows users to select one or a number of documents to display the words and the number of documents containing the word in selected documents.

      A similar feature is embedded in theme view, called Probe. Instead of displaying the actual number, it presents them as bar graphs.

   b. Learnability

      Good. The overall understandability is good except for uploading the dataset. The basic features are very intuitive. In case they are not, users can mouse over a button to get more information on the button. High-level features are available for experts. The "In-Spire Tips and Techniques at Glance" is a guide which assists users in learning new features. It is a well-written document containing a lot of pictures and a step-by-step guide.

   c. Operability (ease of use mainly of GUI)

      Good. The simple designed GUI brings the ease of operability for users. They are intuitive. The magnify represents zooming, hand represents panning, and so on. Not a whole lot of options are available. (More)

   d. Suitability for the task(s)

Excellent. This is a stand-alone product to visualize a large number of documents (up to about 200,000 documents). The output generally presents a map of collection, major themes and sizes, relationships between themes, and the closeness between themes. It will give an intellectual glance of a large number of documents.

e. Alignment with User Skill-Level (novice, intermediate, and expert users)

Good. A simple task can be done by novice with only zooming and panning in overview. For the experts, Time Slicer, a query based on word presence, on thematic similarity, tracking, grouping, document finder, and hiding document feathers is available.

f. Attractiveness (Esthetic Alignment)

Good. Both visualizations, the theme view and galaxy, are attractive, but the performance of the program can backup the expectation, which is drawn out from the visualization. The only criticism is zooming. A jump from one view to another happens during zooming in the galaxy view. Users can lose focus.

g. Consistency and Uniformity of Interface

Yes. The program has consistent interface as other popular programs such as MS Office. The save button is located on the top and so on. However, the product evaluated had a different interface compared to the guide.

h. Help and Feedback

Fair. Some of the feedbacks were excellent. For example, when loading input failed, it stated a clear explanation for the action. The mistake was able to fixed and retried. However, feedback regarding slicing, time slicer, recalculation, and many other features were absent.

i. Security

Unexplored.

8. Operational Performance

a. Reliability

Fair. The only problem encountered during the evaluation was uploading XML input datasets. The XML dataset, which was uploaded by the ASCII input method successively, resulted in failure by the XML input method.

b. Maturity

Good. The system is very mature, especially considering the fact that it is only the second version of the product; there are not many flaws. Uploading input dataset is the only area that needs to be improved. Even though there are many input options, only ASCII works. Other methods may work, but they are unexplored. More useful feedback regarding expert mode features is needed.

c. Fault Tolerance

The system shows sensitivity in the area of uploading inputs. Other than uploading inputs, the system is robust and was able to maintain a specified level of performance throughout the whole evaluation.

d. Recoverability

Unexplored. The system never crashed during the course of the evaluation. However, once the input value is successfully uploaded into the system, the system saves the data. The visualized graphs are not stored, so users need to a select a stored input dataset to regenerate the graphs. Generating the graphs takes minimal time.

e. Response Time

Good. It takes a reasonable amount of time to upload the data, and once the data is uploaded, response time for navigating the visualization is great. Almost no lagging occurs while navigating visualized graphs.

f. Resource Efficiency

There are no minimum requirements for RAM, clock speed or disk space; however, the better the resources, the more quickly the analysis can be performed. Disk space requirements are driven by size and number of datasets wanting to be analyzed. Roughly two or three times as much disk space as the total size of the datasets planning to be analyzed is needed. Currently, the maximum dataset size is about 200,000 documents at a time. (taken from: http://in-spire.pnl.gov/getacopy.html)

g. Availability

The software does not take up all the resources; other trivial software, such as AIM and MSN worked perfectly. Running two In-Spire programs worked without any issue as well.

9. Portability

Due to the fact that the limited number of installation was granted to evaluate the product, it is difficult to evaluate operability in the various environments common to the application. The system requires to be run on a PC with Windows 2000 (SP3 or later) or Windows XP, which indicates poor portability. However, there are no minimum requirements regarding the resources except for disk space requirement, which is equivalent to the size and number of datasets wanting to be analyzed; roughly two or three times as much disk space as the total size of the datasets planning to be analyzed is needed. Without authorized access to the source code, it is impractical to measure expandability, evolvable, and interoperability. In-Spire can facilitate document understanding and information retrieval a great deal. It will help Replaceability.

10. Maintainability

Without authorized access to the source code, the only way to measure maintainability is through documentation. Although there are a number of documents to guide users, there is not a document which depicts overall software architecture or design. There are a total of four aspects which make up the maintainability criteria. The criteria are: analyzability, changeability, stability and testability of the product. Analyzability, which is closely related to the amount of resources used and the duration of such use in performing, seems good. There are no minimum requirements for RAM, clock speed or disk space; however, the better the resources, the more quickly the analysis can be performed. Disk space requirements are driven by size and number of datasets wanting to be analyzed. Roughly two or three times as much disk space as the total size of the datasets planning to be analyzed is needed. Currently, the maximum dataset size is about 200,000 documents at a time. (taken from: http://in-spire.pnl.gov/getacopy.html)

NON-TECHNICAL

11. MATURITY OF PRODUCT

This is the second version of In-Spire, yet it seems robust and mature. There were not any problems or issues, such as crash, hang, or other unusual behavior encountered during the evaluation.

12. COSTS

Commercial License:

- In-Spire License: $15,000.00 per workstation
- One year maintenance: $3,750.00
- Support/Training: $1,500.00 per token

For U.S. Federal Government Use:

- $500.00 Service fee per license that covers costs of distribution and administration

13. LEGAL ISSUES/COMPLIACNE

a. Access to source code

No. There is no access to the source code.

b. License Rights

This is a trial version of In-Spire. It is given to VT for federal government use.

14. PRODUCT SUPPORT

a. End-user Documentation

Good. There are help documents available online which include many picture and step-by-step guides to help users. http://in-spire.pnl.gov/IN-SPIRE_Help/index.html and http://in-spire.pnl.gov/IN-SPIRE2.1Tips.pdf.

Also an online discussion board is available to ask general questions about using and installing In-Spire (http://inspire.proboards4.com). The discussion board was used, but it seemed that there were not staff members to help users.

The link for "making a Subset Data Set" in the online help manual (http://www.pnl.gov/infoviz/in-spire/In-SPIRE_Help/subsets.html) leads to "page not found."

b. Training

Taken from: http://in-spire.pnl.gov/training.html

While it is not unusual for an analyst to start using IN-SPIRE™ without any technical training or support, most users will benefit from a short training session that covers the key aspects of using the tool. Training sessions usually consist of a four to six hour hands-on class that covers the general capabilities of the system along with tips and techniques for data import and analysis. Classes are usually held at the user's site and are not covered by the service fee or licensing fee paid to acquire the software.

In some cases, an organization may have greater support needs, such as datasets that require some level of pre-processing in order to extract the desired fielded information or convert from unsupported formats (e.g., HTML to ASCII). PNNL can assist in these cases as well, on a time and materials basis. Contact (http://in-spire.pnl.gov/contact.html) for more information.

Paying for Training and Support: **Basic Support Tokens (BST)** may be purchased for $1,500 each. One BST provides up to eight hours of support time — whether via telephone or at the site. This can be applied to training or general product support.

If on-site support or training is needed, please contact Dennis McQuerry (http://in-spire.pnl.gov/contact.html) to determine the number of **Custom Support Tokens (CST)** required to meet the requirements. CSTs can be applied to either labor or travel expenses, but the number of tokens necessary to meet the custom requirements must be determined after specific needs have been discussed.

c. Installation

Good. It installs using a standard Windows installer with installation unique update code required.

15. VENDOR CONFIDENCE

Excellent. Pacific Northwest National Laboratory (PNNL) has been researching and providing innovative technologies in visual analytics for more than ten years.

HCE

1. Technology Information

Tech. Name	HCE (Hierarchical clustering Explorer)
Version	3.0
Description	HCE is visualization software developed by the University of Maryland to analyze gene
POC	Jinwook Seo Phone: (301) 405 – 2725 Email: jinwook@cs.umd.edu
Website	http://www.cs.umd.edu/hcil/hce/
Pricing	Free
Requestor	RDEC
RDEC	PP
Experiment Focus	Paper evaluation for the purpose of understanding
Status	Completed
RDEC POC	Albert Park
Installed on	Fall 2005
License Expiration	No Expiration

Screenshot – has been deleted due to sponsor's request.

2. Overview

a. Observed Functionality

HCE allows users to observe data in many different perspectives.

b. Features

HCE offers the following built-in features:

I. Data Filtering

① Present call filtering (only for Affymetrix GeneChip Experiment data)

② Standard deviation filtering – This is a basic filter function.

II. Hierarchical Clustering and Dendogram Display

i. Hierarchical Clustering results are presented as a dendogram. A dendogram is a binary tree, in which each data item corresponds to a terminal node of the binary tree and the distance from the root to a sub tree indicates the similarity of the sub tree. (Taken from: http://www.cs.umd.edu/hcil/hce/hce3-manual/hce3_manual.html)

ii. Color Mosaic Display

This display is for multidimensional Dataset.

iii. Scatter plot ordering

It provides the user with a quick glance of a relationship between two attributes.

iv. There are other views, such as table view, histogram ordering, profile search, evaluation tab, and K-means.

c. Overall Quality

There were no defects or errors detected during the evaluation. The FBIs dataset failed to upload due to unacceptable input format. It would be good to obtain the exact input requirement on this tool. The tool appeared to be robust with respect to the data input to the system as long as the input file is in the right format for the software. From a maintainability perspective, with the tool being an academic tool, it lacks technical support and updates compared to a commercial tool. However, the fact that this is the third version that has evolved from the first program, suggests it has good maintainability.

It does not have eye-catching graphics, but it is well designed for multidimensional dataset. Although the overview can be displayed in both dendogram and color mosaic, the dendogram is more effective than color mosaic. Both of them became ineffective when the dataset got larger. To find out the direct relationship between two attributes or items, the scatter plot and histogram are helpful.

d. Suitability for further testing

Yes. This tool contains many aspects which can be adapted into IA system. It is definitely not a stand-alone product yet, and needs more work to be part of IA system. The biggest advantages are that it handles multidimensional dataset and useful features such as grouping and identifying clusters.

e. Potential Use

As mentioned above, the biggest strength of this software is grouping and identifying clusters. However, the problem with the current method is that it only accounts for one attribute for clustering. One other potential use may be filtering. The software's filtering method is designed for a genome researcher and gene analyst. While the modification of the current method can be useful, incorporating ontology and other information retrieval will be helpful.

3. Hardware/Software Required/Supported
   a. SW Components Required
      Microsoft® Windows 2000®, Windows XP

   b. HW Platforms Supported
      Intel® Pentium® processor

4. PP Experimentation
   a. Experimental Approach
      This evaluation was a paper evaluation for the purpose of understanding the product. Although the evaluation involved using/testing built-in features with the sample data that comes with the software and protected data, there was no measuring or comparing with different products.

   b. Data Set used
      Used a common dataset throughout the experiments; however, a description of dataset has been deleted due to sponsor's request.

5. Quality of Software
   a. Navigation Strategies
      It uses "Zoom" method with a scroll bar.
      By zooming in/out the length of bar the inside the scroll bar is also affected.

As the length of the bar gets longer, the more zoomed out it is and vice versa.

b. Interaction Strategies

Mouse + Keyboard

A conventional mouse and keyboard is used.

6. Usability

a. Ease of learn

Poor. It was easy to upload the dataset; afterwards it got difficult to operate. For example, there is a feature called "compare two results." It was not easy to be figured out immediately nor explained in the user guide.

However, navigating and other interactions are as intuitive as other tools, such as Spotfire or Table Lens.

b. GUI ease of use

Good. After the learning curve, it got easier due to the ease of GUI. However, sometimes the text got stacked on top of another, which can be irritating.

c. End-user documentation

Good. Well-written documentation is provided through web pages. (http://www.cs.umd.edu/hcil/hce/hce3-manual/hce3_manual.html)

It explains specialized features for the expert users. There are a few easier features, which did not get mentioned, such as "compare two result." There are more obvious features to use compared to the ones that were written in the manual.

d. Freedom from problems and issues

Good. No major problems were encountered, except that it crashed more than other software. Obtaining the reliability of the tool is needed; however, the tool handled the situation well by exiting the software immediately. The only problem was that it did not save or recover.

7. Performance
   a. Data Scalability

      About 10,000 items in color mosaic

   b. Response Time

      Excellent. It hardly took any time to upload the test dataset. Once the data was uploaded, the response time was even better. As soon as the mouse button is released, a new graph was generated. The response time during the navigation, meaning zooming in/out, was excellent as well.

   c. Number of concurrent users

      Single user

8. Software Set-up / Maintenance
   a. General set-up information

      Good. It installs using a standard Windows installer.

   b. Ease of software maintenance

      Unexplored. Only one version of the software was found on the web.

9. Appendix

   Not included for this thesis

10. Attachments

Not included for this thesis

Appendix B

PP Evaluation structure

1. Technology Information

   Screenshot

2. Overview
   a. Observed Functionality
   b. Features
   c. Overall Quality
   d. Suitability for further testing
   e. Potential Use
3. Hardware/Software Required/Supported
   a. SW Components Required
   b. HW Platforms Supported
4. PP Experimentation
   a. Experimental Approach
   b. Data Set used
5. Quality of Information Visualization
   a. Overview Strategies
   b. Navigation Strategies
   c. Interaction Strategies
6. Usability
   a. Ease of learning (Ease of obtaining proficiency)
   b. GUI ease of use
   c. End-user documentation
   d. Freedom from problems and issues
   e. Reliability

7. Performance
    a. Data Scalability
    b. Response Time
    c. Number of concurrent users
8. Software Set-up / Maintenance
    a. General set-up information
    b. Ease of software maintenance
9. Appendix
10. Attachments

Appendix C

PPP Evaluation structure

TECHNICAL: FUNCTIONAL

1. PLATFORMS AND OPERATING ENVIRONMENTS
2. INPUT METHODS
   a. Support for metaphors (creation, application, transition, etc)
   b. Input devices supported
   c. Input formats supported
   d. Limitation of input data (size, volume, etc.)
3. MAJOR OUTPUTS PRODUCED
   a. Output formats supported
   b. Handling Clusters
   c. Multi-view
4. METHODS SUPPORTED
   a. Virtual Reality (or Virtual Environment)
      i. Immersive Environment
   b. Animation
   c. Collaboration
   d. Graphical modeling
5. MAJOR VISUALIZATION FEATURES
   a. Navigation
   b. Transition between graphical representations on Multi-View
   c. Controlling interactions with the Visualization
      i. Result set control
      ii. Display options
      iii. Force change

iv. Automat synch with “forced” changed by user

d. Tracking

i. between refinement levels

ii. between time steps

iii. between processor

e. Searching

f. Slicing

g. Illustration by time and location (for geospatial information)

h. Hierarchies/Grouping

TECHNICAL: NON-FUNCTIONAL

6. COMPLETENESS OF THE SOFTWARE

a. Accuracy

b. Robustness

c. Integrity

7. USABILITY

a. Understandability

b. Learnability

c. Operability (ease of use mainly of GUI)

d. Suitability for the task(s)

e. Alignment with User Skill-Level (novice, intermediate, and expert users)

f. Attractiveness (Esthetic Alignment)

g. Consistency and Uniformity of Interface

h. Help and Feedback

i. Security

8. OPERATIONAL PERFORMANCE

a. Reliability

i. Maturity

ii. Fault Tolerance
iii. Recoverability
b. Throughput
c. Response Time
d. Resource Efficiency
e. Availability

9. PORTABILITY
   a. Adaptability
   b. Replaceability
   c. Interoperability
   d. Conformance
10. MAINTAINABILITY
   a. Analyzability
   b. Changeability
   c. Stability
   d. Testability

NON-TECHNICAL

11. MATURITY OF PRODUCT
12. COSTS
13. LEGAL ISSUES/COMPLIACNE
   a. Access to source code
   b. License Rights
14. PRODUCT SUPPORT
   a. End-user Documentation
   b. Training
   c. Installation
15. VENDOR CONFIDENCE

Appendix D

This is a survey that is developed for analyzing intelligence analysts' activities. Unfortunately, there was not an opportunity to conduct the survey for this thesis.

Please list five to seven tasks in the order that they are performed most frequently for each category; also state the frequency and the time it takes to perform these tasks. {e.g., 1. collect recent intelligence (once a week, 3 hours), 2. compare to previous activity logs (once a month, 5 hours), 3. hypothesis generation (everyday, 1 hour), 4. structured argumentation (once in 3 months, 3 hours), 5. reporting (once in 3 months, 1.5 hours)}

- Tasks related to analyzing
- Tasks where you may currently visualize the result (either through your mind's eye or through graphical means)
- Tasks using computer/software

Are there any tasks that you envision a computer/software supporting you more in your efforts?

What kinds of tasks require collaboration with other analysts or research staff? How do you get that assistance today?

Please describe your work environment, such as daily work-time situation/pressure, amount of data to analyze, targeted accuracy, targeted efficiency, etc. (Please note the areas where you believe the tasks are repetitive in nature or require information retrieval that could be conducted by a computer.)

References

[1] S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, 1999.

[2] C. Chen and Y. Yu, "Empirical studies of information visualization: a meta-analysis," *Int. J. Human-Computer Studies*, vol. Vol. 53, pp. 851 - 866, 2000.

[3] C. Plaisant, "The challenge of information visualization evaluation," in *Proceedings of the working conference on Advanced visual interfaces*. Gallipoli, Italy: ACM Press, 2004.

[4] O. Juarez, C. Hendrickson, and J. H. G. Jr., "Evaluating Visualizations Based on the Performed Task," presented at proceedings of Fourth International Conference on Information Visualisation (IV'00), 2000.

[5] C. Chen and M. P. Czerwinski, "Empirical evaluation of information visualization: an introduction," *Int. J. Human-Computer Studies*, vol. Vol. 53, pp. 631 - 635, 2000.

[6] C. Chen and M. P. Czerwinski, "Spatial Ability and Visaul Navigation: An Empirical Study," *The New Review of Hypermedia and Multimedia*, vol. 3, pp. 67 - 89, 1997.

[7] Y. Wang, S. T. Teoh, and K.-L. Ma, "Evaluating the Effectiveness of Tree Visualization Systems for Knowledge Discovery," presented at proceedings of Eurographics/IEEE-VGTC Symposium on Visualization, 2006.

[8] C. Plaisant, B. Shneiderman, K. Doan, and T. Bruns, "Interface and Data Architecture for Query Preview in Networked Information Systems," *ACM Transactions on Information Systems (TOIS)*, vol. 17, pp. 320 - 341, 1999.

[9] C. Plaisant, J. Grosjean, and B. B. Bederson, "SpaceTree: supporting exploration in large node link tree, design evolution and empirical evaluation," presented at proceedings of the IEEE Symposium on Information Visualization 2002 (InforVis '02), 2002.

[10] U. Wiss, D. Carr, and H. Jonsson, "Evaluating three-dimensional information visualization designs: acase study of three designs," presented at proceedings of the IEEE Conference on Information Visualization, London, England, 1998.

[11] A. Kobsa, "User Experiments with Tree Visualization Systems," presented at INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization 2004, Austin, Texas, USA, 2004.

[12] C. M. D. S. Freitas, P. R. G. Luzzardi, R. A. Cava, M. A. A. Winckler, M. S. Pimenta, and L. P. Nedel, "Evaluating Usability of Information Visualization

Techniques," presented at proceedings of Advanced Visual Interfaces - AVI '02, 2002.

[13] J. Somervell, D. S. McCrickard, C. North, and M. Shukla, "An evaluation of information visualization in attention-limited environments," presented at proceedings of the symposium on Data Visualisation 2002, 2002.

[14] M. M. Sebrechts, J. V. Cugini, J. Vasilakis, M. S. Miller, and S. J. Laskowski, " Visualization of search results: a comparative evaluation of text, 2d, and 3d interfaces," presented at proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999.

[15] M. Tory and C. Swindells, "Comparing ExoVis, Orientation Icon, and In-Place 3D Visualization Techniques," presented at proceedings of Graphics Interface, 2003.

[16] A. Kobsa, "An Empirical Comparison of Three Commercial Information Visualization Systems," presented at proceedings of InforVis 2001, 2001.

[17] J. Mackinlay, "Automating the Design of Graphical Presentations of Relational Information," *ACM Transactions on Graphics,* vol. Vol. 5, pp. 110 - 141, 1987.

[18] J. Yang-Pelaez and W. C. Flowers, "Information Content Measures of Visual Displays," presented at proceedings of the IEEE Symposium on Information Visualization (InforVis '00), 2000.

[19] E. Morse, M. Lewis, and K. A. Olsen, "Evaluating visualizations: using a taxonomic guide," *Int. J. Human-Computer Studies,* vol. Vol. 53, pp. 637 - 662, 2000.

[20] M. Tory and T. Moller, "Rethinking Visualization: A High-Level Taxonomy," presented at IEEE Symposium on Information Visualization, Austin, Texas, 2004.

[21] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," presented at proceedings of IEEE Symp. on Visual Languages '96, 1996.

[22] P. Cowley, L. Nowell, and J. Scholtz, "Glass Box: An Instrumented Infrastructure for Supporting Human Interaction with Information," presented at proceedings of the 38th Hawaii International Conference on System Sciences, 2005.

[23] R. Brath, "Concept Demonstration Metrics for Effective Information Visualization," presented at proceedings for IEEE Symposium on Information Visualization, 1997.

[24] R. M. Pillat, E. R. A. Valiati, and C. M. D. S. Freitas, "Experimental Study on

Evaluation of Multidimensional Information Visualization Techniques," presented at Proceedings of CLIHC '05, Cuernavaca, Mexico, 2005.

[25] S. Wehrend and C. Lewis, "A Problem-oriented Classification of Visualization Techniques," presented at proceedings of IEEE Visualization '90, 1990.

[26] M. A. Winckler, P. Palanque, and C. M. D. S. Freitas, "Tasks and Scenario-based Evaluation of Information Visualization Techniques," presented at TAMODIA, Pargue, Czech Republic, 2004.

[27] J. Nielsen, "Finding Usability Problems Through Heuristic Evaluation," presented at proceedings of ACM CHI '92, 1992.

[28] J. Nielsen and R. Molich, "Heuristic Evaluation of User Interfaces," presented at proceedings of ACM CHI '90, Seattle, VA, 1990.

[29] P. Saraiya, C. North, and K. Dunca, "An Evaluation of Microarray Visualization Tools for Biological Insight," presented at proceedings of IEEE Symposium on Information Visualization, Washington, D. C., 2004.

[30] P. Saraiya, C. North, and K. Duca, "An Insight-based Methodology for Evaluating Bioinformatics Visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. Vol. 11, pp. 443 - 456, 2005.

[31] V. Gonzalez and A. Kobsa, "A Workplace Study of the Adoption of Information Visualization Systems," presented at I-KNOW'03: 3rd International Conference on Knowledge Management, Graz, Austria, 2003.

[32] J. Rieman, "A Field study of Exploratory Learning Strategies " *ACM Trans. Comput-Human. Interact*, vol. Vol. 3, pp. 189 - 218, 1996.

[33] O. Juarez-Espinosa, "CAEVA: Cognitive Architecture to Evaluate Visualization Applications," presented at proceedings of International Conference on Information Visualization (IV'03), 2003.

[34] M. X. Zhou and S. K. Feiner, "Visual Task Characterization for Automated Visual Discourse Synthesis," presented at Proceedings of the CHI '98, Los Angeles, CA, U.S.A., 1998.

[35] E. R. A. Valiati, M. S. Pimenta, and C. M. D. S. Freitas, "A taxonomy of tasks for guiding the evaluation of multidimensional visualizations," in *Proceedings of the 2006 conference on BEyond time and errors: novel evaluation methods for information visualization*. Venice, Italy: ACM Press, 2006.

[36] J. Gersh, B. Lewis, J. Montemayor, C. Piatko, and R. Turner, "Supporting insight-based information exploration in intelligence analysis," *Commun. ACM*, vol. 49, pp. 63-68, 2006.

[37] N. Rashevsky, "Life, informaiton theory, and topology," *Bulletin of*

Mathematical Biophysics, vol. 17, pp. 229 - 235, 1955.

[38] C. M. D. S. Freitas, P. R. G. Luzzardi, R. A. Cava, M. Winckler, M. S. Pimenta, and L. P. Nedel, "On Evaluating Information Visualization Techniques," 2002.

[39] M. Maybury, J. Griffith, R. Holland, L. Damianos, Q. Hu, and R. Fish, "Virtually Integrated Visionary Intelligence Demonstration (VIVID)," presented at Intl Conf on Intelligence Analysis, 2005.

[40] H. Chen, F.-Y. Wang, and D. Zeng, "Intelligence and Security Informatics for Homeland Security: Information, Communication, and Trasportation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, pp. 329 - 341, 2004.

[41] R. E. Horn, "Connecting the Smudges: How Analytic Info-Murals May Be of Help in Dealing With Social Messes," presented at 2005 International Conference on Intelligence Analysis, 2005.

[42] R. Becker and W. Cleveland, "Brushing Scatterplots," *Technometrics*, vol. 29, pp. 127 - 142, 1987.

[43] D. L. Gresh, B. E. Rogowitz, R. L. Winslow, D. F. Scollan, and C. K. Yung, "WEAVE: a system for visually linking 3-D and statistical visualizations, applied to cardiac simulation and measurement data," presented at IEEE Visualization '00, 2000.

[44] H. Hauser, F. Ledermann, and H. Doleisch, "Angular Brushing of Extended Parallel Coordinates," presented at IEEE Symposium on Information Visualization 2002 (InforVis '02), 2002.

[45] R. J. Bobrow and A. Helsinger, "Kinetic Visualizations: A New Class of Tools for Intelligence Analysis," Advanced Research and Development Activity (ARDA),National Geospatial-Intelligence Agency (NGA) 2005.

[46] Spotfire, "Spotfire, Inc. - Interactive, Visual Analysis."

[47] P. Buono, A. Aris, C. Plaisant, A. Khella, and B. Shneiderman, "Interactive Pattern Search in Time Series," presented at Conference on Visualization and Data Analysis, VDA 2005, Washington DC, 2005.

[48] H. Hochheiser, E. H. Baehrecke, S. M. Mount, and B. Shneiderman, "Dynamic querying for pattern identification in microarray and genomic data," 2003.

[49] Tablelens, "The fastest way to turn data into decisions," Inxight.

[50] IN-SPIRE™, "Visual Document Analysis."

[51] J. Seo and B. Shneiderman, "Interactively exploring hierarchical clustering results [gene identification]," *Computer*, vol. 35, pp. 80-86, 2002.