Växjö universitet
School of Mathematics and System Engineering
Course: DA4024 – Visual Analytics, autumn semester 2007
Examiner: Dr. Andreas Kerren, Associate Professor

March 9, 2008

Geo-temporal Visual Analysis
An Overview

Arianit Kurti
arianit.kurti@msi.vxu.se

Abstract. This paper presents an overview of current research efforts in the field of geo-temporal visual analysis. This overview is based on the analysis of four recently published research papers in the IEEE Symposium on Visual Analytics Science and Technology (2006 and 2007). The selected papers cover projects that are applicable in oceanography, epidemiology, service economy (hotel management), and disaster warning systems. In these projects, different visualization tools have been used, such as PaleoAnalogs, Improvise, GeoTime, and FemaRepViz. This paper starts with a brief introduction on databases and their importance, after continues with the identification of the problems related to overload of the data. Then, Information Visualization and Visual Analytics are introduced. After this introductory section, a main part will deal with an overview of the research papers. This overview will be focused on the description of the projects by primarily showing how these software tools works and how they enhance the human insight into abstract data and if these tools are applicable in different application domains. The last section of this paper will provide a summary based on the software used, application domain, geo-information used and usability of the tools.

Keywords: visual analytics, geo-temporal visualization, interaction techniques, geospatial analytics, historical geography

1. Introduction

The need for digital representation of the physical occurrence and activities existed for a long time. Examples of this need can be found in different domains such as meteorology where a need for storing the temperature change over the years exists. The need for the digital representation was primarily motivated with the need for analyzing these occurrences and activities. In the first stages of dealing with this problem, the solution came in the form of archives and libraries of different documents. These structured organization of the documents led to creation of so called first databases. Initially, these databases have been paper based and not electronic/digital ones. The problems with the paper based databases have grown proportionally with the growth of the data/documents stored. The main problem was the sequential access and especially the enormous time needed to find a desired document. As these archives become bigger the time needed for search was becoming longer. The problem of search and achieving fast results of the queries ignited the need for the design of electronic/digital databases. Hence, since the introduction of the first computers the main applications focus has been the domain data storing and manipulations (Beynon-Davies, 2004). The database term initially covered both no digital and digital data archives, but with the time digital databases gained more attractions and nowadays when the term database is mentioned it primarily means computer databases.

Computer based databases are closely related with the introduction of the Database Management System (DBMS) software. In the late 1960s, the dominant DBMS was based on the hierarchical data representation model. After them, the revolution within the DBMS came with the introduction of the relational data representation models in the 1970s (Rivero et al. 2006). The main characteristics of this data representation model have been entities and relationships. This model is most widely used one today. In general, computer based databases offered new ways of data storage and search/seek. The main advantages that DBMS offers for data management and storage are defined by Ramakrishnan and Gehrke (1999) as: data independence, efficient data access, data integrity and security, and data administration. The first two points primarily represent the advantages of the DBMS regarding the usability, while the third and forth points represent the advantages regarding security. The usability advantages of the DBMS have had tremendous impacts in the way how the business is done and practically enhanced almost all the aspect of human life. The ability for fast seek of the needed information was a clear advantage compared to non-digital databases. The data value toward the user is often defined based on the ease of obtaining that data. Especially in the relational databases, this aspect has been developed further in order to allow a rich class of questions to be asked. The questions asked to the DBMS terminology are defined as queries for example by SQL.

Today in the 21st century, the sources and amount of data produced has been increased significantly. The databases are becoming huge. Despite the best optimizations of the queries, the results gained from the databases are sometimes very large and thus not fitting the users needs. Therefore, the real problem that today's economy is facing is the vast amount of data. It has become extremely difficult that just using queries to extract the needed data and/or identify certain patterns and trends that exist among the data. Thus, it is clear that there is an immediate need for data selection, transformation and representation in order to ease this “overload of data” problem. One of the approaches to provide better analytical insight into abstract data and thus tackling the “overload of data problem” was based on the concept of contextualization. The contextualization analytical approaches that will be described in this paper are based on the idea of geographical maps usage as layer for localizing abstract data.

2. The Need for Data Visualization

Human vision is the sense with the highest bandwidth and thus visualization of abstract data is used as a technique for ease the “overload of data” problem. Use of visualization has led to introduction of a multidisciplinary scientific discipline that combine the features human cognitive apparatus with database queries, data exploration and browsing etc. This discipline is referred as Information Visualization. The aim of this discipline is primarily to facilitate the creation of the mental model of the data, thereby gaining insight into the data (Spence, 2001). The main focus in this aspect is to support the acquisition of the insight by identifying the models and trends through visualization. From the computer based perspective, information visualization is defined as: *“the use of computer-supported, interactive, visual representations of abstract data to amplify cognition”* (Kerren, 2007). In this sense, the main points of this definition are: interaction, abstract data and amplification of the cognition. Interaction means that this computer tool should offer some interactive features that enable users to directly interact with abstract data. This interaction of course should be visual and thus potentially amplifying cognition. Based on this in Figure 1 below the basic visualization model is illustrated with all the building blocks.

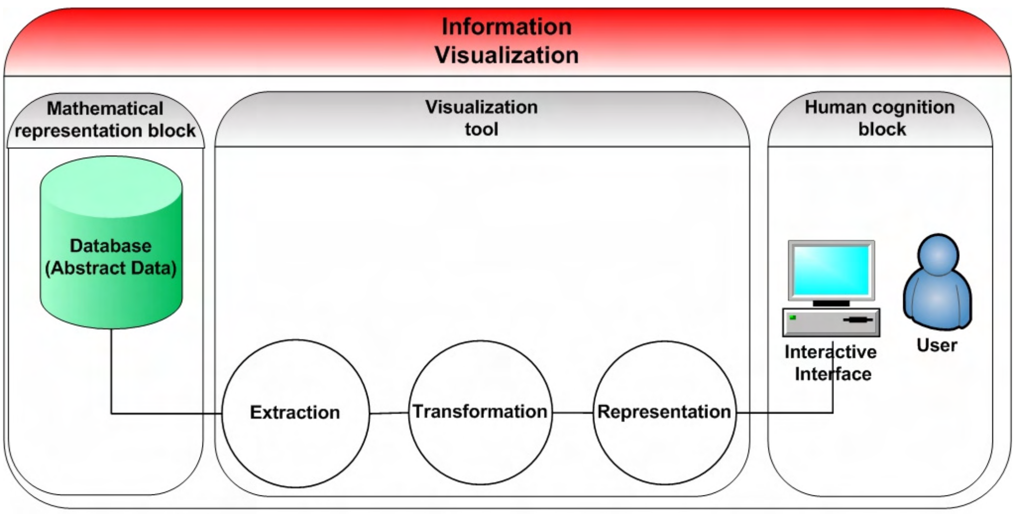

Fig 1. Information Visualization model

The three building blocks that build a information visualization tool could be defined as:

- The *Mathematical representation block* that includes the database with the abstract data. This block is referred like this primarily because the grounded foundations of the databases are mathematical ones.
- The *Visualization tool* that includes the data extraction, transformation, and representation. This block represents a three phase process: extraction, transformation and representation over which the data need to be processed.
- The *Human cognition* block that basically relies on human vision as highest bandwidth human sense. This block should include the interactive interface that would enhance the users' acquisition of insight into abstract data.

The aims of each of these blocks are different but together they tend to help transforming of data into information and thus potentially helping people to think and easily identify the patterns and trends from the abstract data. In this aspect, the aim of the mathematical representation block is digital abstraction of the physical occurrences and activities. The aim of the visualization tool is to process the data over three phases. While the human cognition block aim is to provide interactive means for manipulations and control over the processed abstract data. In one way the visualization tool as main component in this system represents a kind of middleware between the users and the database systems.

2.1 Information Visualization vs. Scientific Visualization

For the better understanding of the information visualization it is very important to make a clear distinction from scientific visualization. While the information visualization deals with abstract quantities as its data sources, scientific visualization deals with visual representation of something "physical" (Spence, 2001). In this sense for example, if we represent visually the flow of the fluids in the pipe we do a scientific visualization. But if we visualize the number of the pipes that fluids are flowing through then we have to do with information visualization. This is primarily because the "flow of the fluid" is something "physical", while the number of pipes having this flow is more an abstract quantity.

In this sense, it is interesting to mention that geographical maps basically represent scientific visualization, simply because it visualizes the physical state (oceans, seas, rivers, valleys, mountains, etc.). Geo-temporal visualization that is a part of information visualization uses maps as visualization and localization layer for abstract data. Therefore it could be regarded that geo-temporal visualization is on the borderline between scientific visualization and information visualization. Geo-temporal visualization will be discussed in the next section.

2.2 Visual Analytics and Information Visualization

According to Wong and Thomas (2004), visual analytics is defined as: "*a contemporary approach that combines the art of human intuition and the science of mathematical deduction to directly perceive patterns and derive knowledge and insight from them*". In this sense, visual analytics tries to create abstract visual metaphors that in association with human information interaction could enable the detection and discovery of patterns and unexpected insight within vast information spaces. Therefore, it can be said that visual analytics focus more toward the human cognition block and representation step within the visualization tool as illustrated in Figure 1. This basically means that the difference between Visual Analytics and Information Visualization is mainly on their focus. While in Information Visualization the focus is also on data extraction techniques and data transformation, this is not the case in Visual Analytics. Visual Analytics is also an interdisciplinary approach that includes methods and technologies of other fields, such as knowledge management, statistical analysis, data mining, cognitive science, decision science, and many more (Wong and Thomas, 2004).

3. Geo-temporal Visualization

Geo-temporal visualization could be regarded as visualization of abstract data in time and geography. This definition could be deducted from the name as well where *Geo* stands for geography usage for visualization, while *temporal* stands for time usage for visualization. In this aspect, geography factors used in visualization are represented with digital maps that serve as representation layer. It should be made clear distinction with scientific visualization here, because in geo-temporal visualization a geographical map is only used as a representation layer, and it does not define anything else beside the location. In geo-temporal visualization it is very common to use the so-called XYT coordinative system. This means that the XY plane represents the location or the terrain plane (i.e. map) while the T axis represent time. T=0 represents the current or instant focus of the situation. The positive values of the T represents the future situations while the negative ones the previous ones. The illustration of this 3D coordinative system is illustrated in the Figure 2.

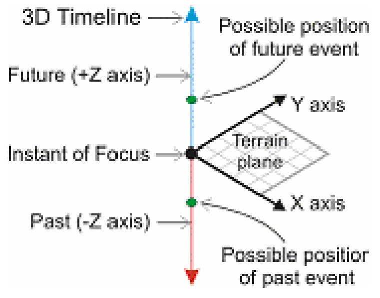

Fig 2. Geo-temporal 3D representation model (Shuping and Wright, 2005)

This model of representation presents the basic visualization technique in geo-temporal visualization. It can be combined also with other visualization techniques to create different visualization tools that could be used in distinct application domains. The importance of geo-temporal visualization has become increasingly interesting in the Web 2.0 era. A large number of Web 2.0 services has been developed based on the idea of adding “where?” information. These services have been primarily developed as a result of blooming of digital map services like Google Earth/Maps, Yahoo Maps, and Microsoft Live Maps etc. These services are based on the idea of mashups that combine data from multiple sources and represent these data over a digital map. In this way, these services try to increase the insight into abstract data by providing better analytical opportunities by relating the data with the location.

4. Current Research Efforts in Geo-temporal Visualization

With the reference to the key concepts presented in the previous sections, this section focuses on recent developments in the geo-temporal visualization field. For this reason, we will now provide an overview of the four research papers within the area of the geo-temporal visual analysis. The papers have been selected from the last two editions (2006 and 2007) of IEEE Symposium on Visual Analytics Science and Technology. The underlying projects cover issues relating to geo-temporal visual analysis applications in oceanography, epidemiology, service economy, and disaster warning systems. In these projects, different visualization tools have been used, such as: PaleoAnalogs, Improvise, GeoTime, and FemaRepViz. The presentation of the papers will be focused primarily showing how these tools works, how they enhance the human insight into abstract data and how these tools are applicable in different application domains. The overview of the papers will follow a template that will cover these issues:

1. *Motivation*; that will present the authors’ motives for doing such work and in that specific domain.
2. *Data Set*; which will deal with identification of the abstract data used in the project. This section will deal with the digital abstraction of the physical occurrences. .
3. *Data Transformation*; this section will include the discussion about the transformation techniques used.
4. *Software Tool*; will deal with detail presentation of the software tools used in the projects, interaction support and visual representation approach used.

4.1 Paper "Visual Analytics of Paleoceanographic Conditions" by Theron (2006)

This paper describes a tool that enables the understanding of events in the past relating to climatic changes in order to be able to forecast abrupt climatic changes. The study is done in the paleoceanography research based on the analysis of the paleo time-series offered by different research laboratories. This paper presents a highly interactive visual analysis environment that enable getting insight into the paleodata and moreover offers methods for the reconstruction of climatic conditions in the past.

4.1.1 Motivation

During the last decade, we have been witnesses of some very catastrophic natural disasters. Some of these disasters have been of geologic origin while most of them are weather related. Having the possibility to aggregate the data from different sources, it is possible to diagnose and to prevent these natural disasters in some extend. The possibility to prevent these natural disasters by using visual analytics tools served as motive for the author of this paper. This is a very noble motivation since it shows how important visualization is, and providing insight into the large set of abstract data can potentially lead to savings human lives.

4.1.2 Data Set

As suggested in the Figure 1 in Section 2, the data set from where the data are extracted is a very important for the visualization. In this paper, the data set is basically created by aggregation from different sources, especially because for the paleoceanography analysis there is need for comprehensive overview. These data has been collected during the course of thousand years (mainly in ice and sediments core) and potentially would lead to better understanding of earth and ocean dynamics. These data has been collected from different research labs by using different techniques such as ocean drilling, ocean tracers, astronomic curves etc. The collection of this data has been done with the help of tools like PaleoPlot and AnalySeries. This very complex and aggregated raw data represents the mathematical representation block as suggested in Figure1.

4.1.3 Data Transformation

It's well known fact that for the visualization we need some kind of data transformation. In the discussed project as a data transformation approach is used Modern Analogue Technique (MAT). This technique is well-known in paleoceanography especially for the reconstruction of the past environmental conditions. There have been multiple software tools that were built for implementing the MAT technique, like SIMMAX and RAM. The problem that arises with these tools was lack of insight over the process of MAT for the paleoceanographyst. In both of these implementations, MAT has been regarded as a black box that has an input and output. Main contribution of the author of this paper was the fact that the use of the PaleoAnalogos offered the scientist insight over MAT technique and thus potentially some more knowledge regarding the past and future environmental factors. MAT implemented in PaleoAnalogos uses the nearest neighbor prediction algorithms. The nearest neighbor prediction algorithms can be also found as K-nearest neighbor in different literature. It works on the principle that is illustrated in Figure 3.

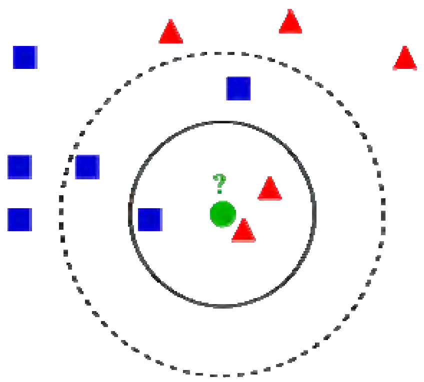

Fig 3. The nearest neighbor prediction algorithms

If for the case presented in Figure 3, the green dot represents the unknown object, then based on the value of K (that represents the number of the nearest neighbor) it can be predicted if the unknown object is a square or triangle. If, for example K=3 then based on the algorithm it can be deducted that the unknown object is a triangle because two of the three nearest neighbors (the circle with the full line) are triangles. But if K=5 then the unknown object would be a square since three out of five nearest neighbors (the circle with the dash line) are squares.
This algorithm is used in this project as a mean for predicting and reconstructing the paleoenvironmental features such as sea surface temperatures (SST). It was applied using the PaleoAnalogs tool.

4.1.4 The Software Tool

The author's tool PaleoAnalogs[1] offers visual insight into different data that are relevant for paleo analysis. The software tool is a Java-based program that enables among others the scientist with the visual representation of the nearest neighbor algorithm used in the MAT technique. In the Figure 4, the author shows how the PaleoAnalogs offers this insight and thus making the entire prediction process of environmental factors more transparent for the scientists.

[1] Free software tool based on Java, that can be downloaded from: http://carpe.usal.es/

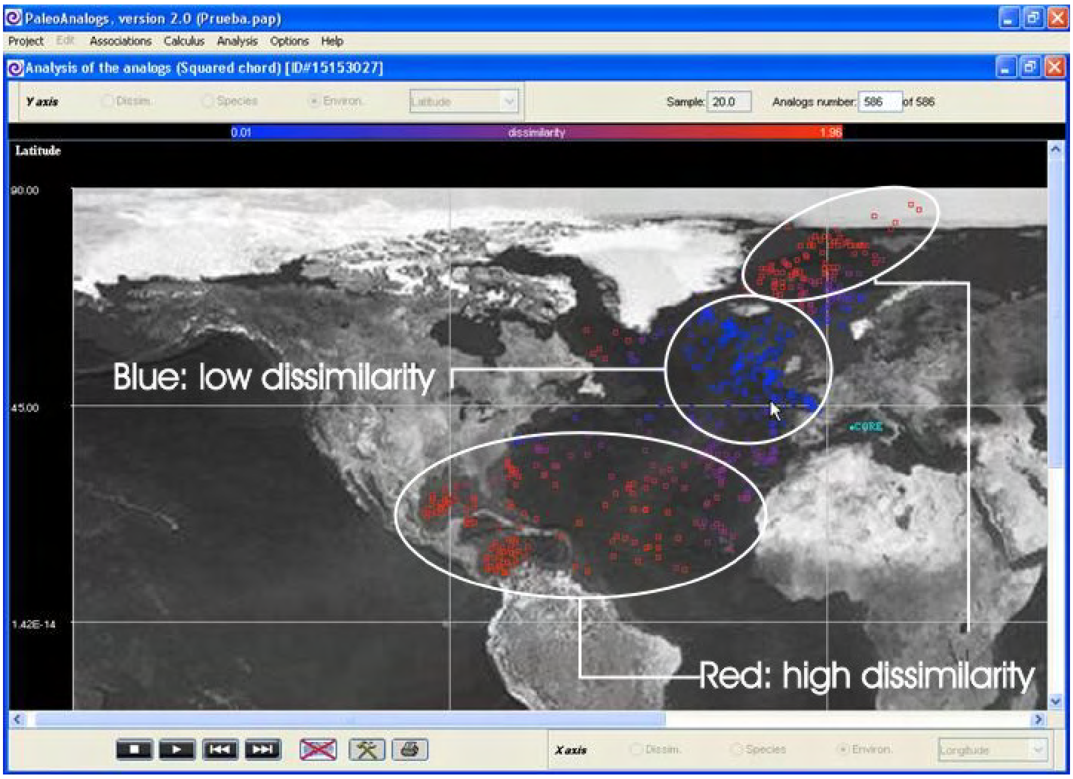

Fig 4. Use of k-nearest neighbor algorithm by PaleoAnalogs (Theron, 2006)

For example, the CORE point (cyan label) in the Mediterranean Sea represents the unknown sample. Using the PaleoAnalogs tool and visualizing other samples of analogs, the cold ones with blue while the hot ones with red, we can have the visual insight if the unknown sample would have been cold or hot using the nearest neighbor prediction algorithm. In this specific case it is obvious that the unknown sample has been a cold one. Thus, the PaleoAnalogs tool provides insight into the MAT technique and thus becoming more transparent for the scientists. The overall PaleoAnalogs visual analytics environment is presented in Figure 5. The PaleoAnalogs offers multiple interactions and visualization techniques for facilitating the process of getting the insight into the abstract data. The visualization and the interaction techniques used in this tool are: space and time reasoning, Focus+Context, filtering and axis interaction, Brushing, and Parallel Coordinates. The space and time reasoning is done using a map as a representation layer. It is interesting to mention that the time dimension is not implemented as the third dimension but more as a series of snapshots. In this sense, it can be regarded that space time reasoning is not implemented as 3D structure. The focus and context technique facilitates the selection of the data sources. It is implemented using a fish eye view. Filtering and axis interaction is used for dynamic filtering in PaleoAnalogs. The dynamic queries on the Parallel Coordinates Plots (PCP) that are used for representing multidimensional data are implemented using axis filtering using range sliders.

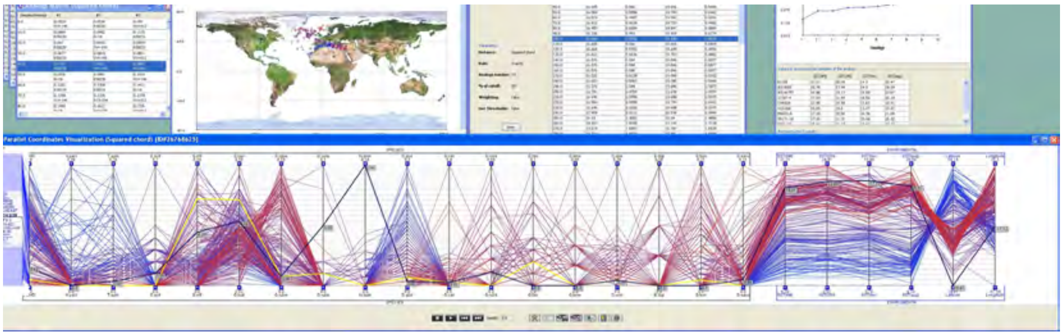

Fig 5. PaleoAnalogs visual analytics environment (Theron, 2006)

4.2 Visual Analysis of Historic Hotel Visitation Patterns (Weaver *et al.*, 2006)

This paper represents an effort of how visual analysis can be used to understand the space and time characteristics of human interactions in the social context. They focus on identification and comparison of patterns of the recurring events. In this aspect, the authors describe a tool used to understand hotel visiting patterns in the Rebersburg area in Pennsylvania (US) during the 1898-1900 time periods. For a better understanding and more insight into the abstract data they use a wrapping spreadsheet technique called *reruns*.

4.2.1 Motivation

Today there are multiple data available for describing human behavior primarily for commercial and marketing reasons. Using these data for extracting wealthy information about the cultural connectivity patterns is mentioned as a motivation by the authors of this paper. The goal according to Weaver *et al.* (2006) is: *"to allow users to explore information that contains geospatial, temporal and abstract components in a flexible, integrated, interactive graphical environment"*.

4.2.2 Data Set

This work is based on the analyzing hotel visiting patterns in the Rebersburg hotel in Rebersburg. The data set used for this purpose was the guest registry for the period between June 1898 and November 1900. In total this guest registry contained 2411 entries. The visitors have been from 219 different places in 16 US states and from Canada and England as well. Approximately there have been 100 guests every month. More then 1800 visits were made from people having the residence in the radius less then 50 miles (80 km) from Rebersburg. Due the impossibility to decipher the names, locations and other data from the guest registers, 6% of the data set was not legible.

4.2.3 Data Transformation

Here there was a need for transforming data from a non-digital database (the guest registry) into a digital database. The data from the guest registry have been transformed into a spreadsheet. The transformation was manual and took approximately 80 hours. It was done by enabling data to be sorted in the way that spatial and temporal patterns can be identified using computational methods. Due the manual process of transformation of data about guests and residences have been checked for consistency, in order to eliminate the eventual errors. After this process, the summary tables showing the total number of visits per person and per residence have been created.

4.2.4 The Software Tool

The software tool used for this purpose was Improvise[2] that is an open source Java based tool. This tool enables the analysts to have insight into the abstract data by browsing multiple coordinated views. The main advantage of using Improvise according to authors is that fact that it offers the precise control over how interaction affects the presentation of space, time, and abstract data. By being an open source tool, Improvise offers the possibility for modifications and extension. In this aspect, the main visualization component used in this project is a glyph-based technique called *reruns*. Reruns is a visualization environment that offers wrapping for different data spreadsheets in order to explore cyclic and temporal patterns among data. The hotel visualization interface is illustrated in Figure 6. The authors do not refer to this application with a special name beside “the hotel visualization” because, among other things, the Improvise “documents” are saved as regular, self-contained XML.

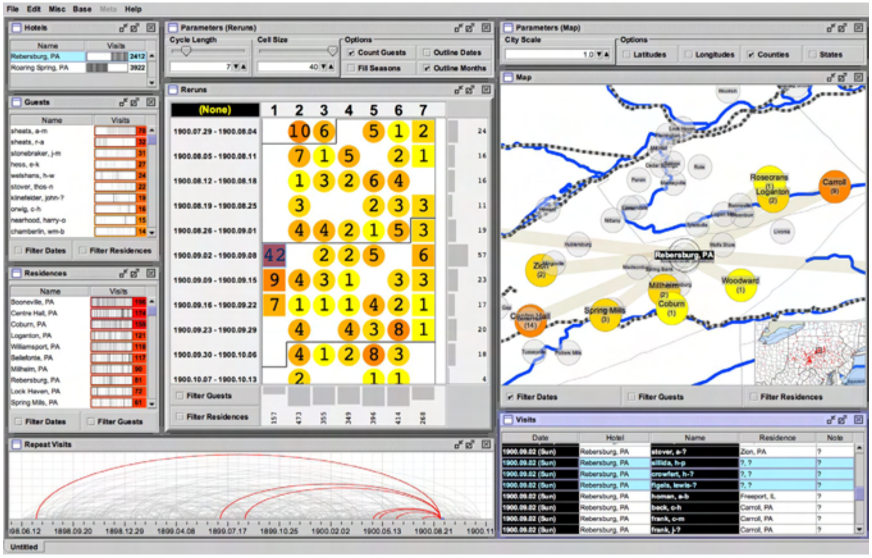

Fig 6. The hotel visualization interface based on Improvise (Weaver *et al.*, 2006)

---

[2] This tool can be downloaded for free from http://www.personal.psu.edu/faculty/c/e/cew15/improvise/

The hotel visualization interface is organized in cells that offer different views on the data and different interaction techniques. On the left hand side of the interface there is a cell with a table view that contains the names of the visitors and the total number of visits. This table offers the possibility for sorting according to name and the number of visits, as well as in combination with other tables in more complex queries. The table beneath this one, still on the left hand side of the interface, represents the residences with tabular view of specific locations and number of visits. Also this table offers different filtering capabilities, such as: numbers of visits from one residence. The reruns table represents the breakdown of the number of visitors' into the specific days and over a two week cycle. The days are shown on the horizontal axis (with Sunday being the first day of the week), while the time in two week cycles. In the vertical histogram on the right hand side of the reruns table, the total number of visits per time period cycle, are represented. While horizontal histograms in the bottom part of the reruns table are used to show a summarized view of total visits for each day during the time period. The multilayered map on the right is used for cross representation of the data, mainly paths from residences to the hotel, relative to railroads and rivers. Another interesting view in the Hotel Visualization interface is the arc diagrams used to represent the returned visits. The arc diagram is a visualization technique developed by Wattenberg (2002) and is used for representing complex patterns of repetition in string data. Arc diagrams can be applied to visualize diverse data strings as music, text, and compiled code. An example of arc diagram usage is illustrated in Figure 7.

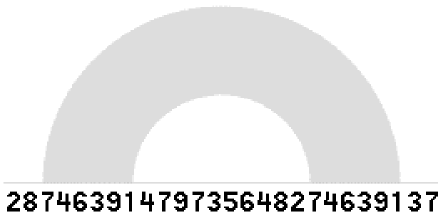

Fig 7. Arc diagram visualization technique (Wattenberg, 2002)

In Figure 7, a simple usage of the arc diagram visualization technique is visualized. This visualization technique implies that for matching string (in the case illustrated it is 746391) to be drawn an arc in order to visualize its repetition. Without the visualization it would be very difficult to track the repetition pattern. In the Hotel Visualization project, this technique is used to show the previous and next visit to the hotel from a specific guest. In this way, it enabled a visual insight for discovery of traveling patterns of the guests of the Rebersburg hotel.

4.3 Avian Flu Case Study with nSpace and GeoTime (Proulx *et al.*, 2006)

This paper offers an overview of GeoTime and nSpace as analysis tools with visual capabilities. To show the usefulness of these tools, the authors describe the epidemiology analysis scenario. They use GeoTime for exploring geo-temporal aspects of the data while with nSpace they describe the rapid information scan and extraction.

4.3.1 Motivation

Despite the fact that the scenario developed and described in the paper is fictional, the primary motivation of the authors is to offer an analytic tool that could fast scan the publicly available documents in order to control and eventually predict the flu propagation. In this aspect, the paper just presents the idea how the different tools for visualization can be used in order to get a insight into abstract data relating a flu epidemics.

4.3.2 Data Set

The data set for this project is related to epidemics of the Avian flu. This data set has been created by the authors in order to investigate the analytic process and evaluate the efficiency of the GeoTime and nSpace tools. The data set has been consisted form publicly available documents. It has been extracted using The Rapid Information Scanning Tool (TRIST). TRIST uses Human Information Interaction (HII) techniques to interact with massive data in order to quickly uncover the relevant, novel and unexpected (Jonker *et al.*, 2005). The research community in the field of information retrieval supported the design of the TRIST. According to Prolux and colleagues (2006), the TRIST tool offers possibilities for rapid and efficient scanning and prioritization of a large number of search results in one display. Moreover, it supports multiple analytical processes, such as: query planning, rapid scanning capabilities, and result display across multiple dimensions. The interface of the TRIST is presented in Figure 8.

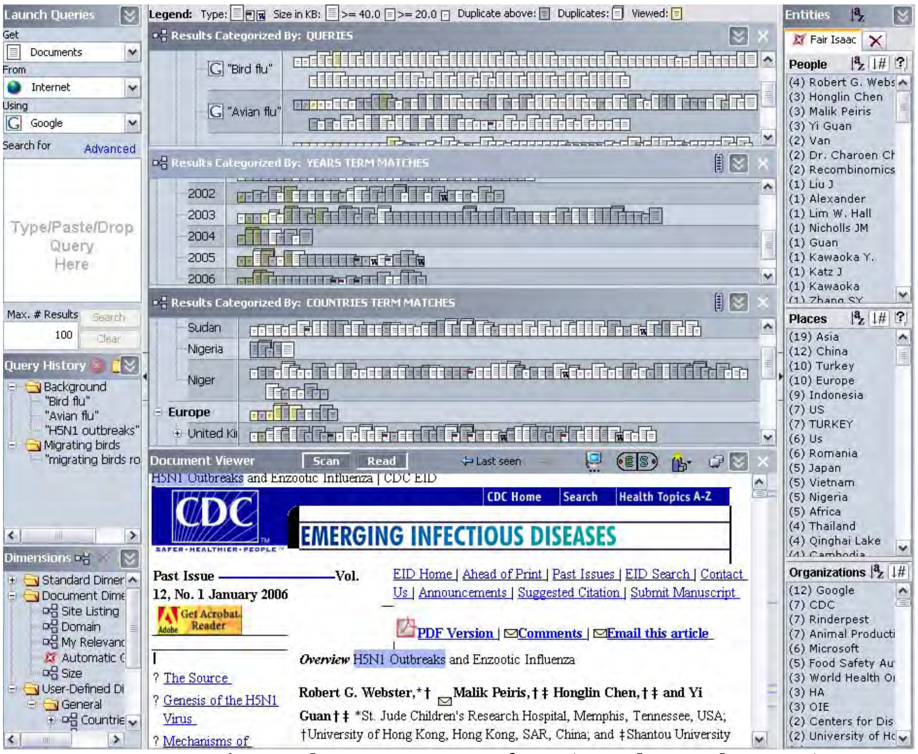

Fig 8. The TRIST interface (Proulx *et al.*, 2006)

The TRIST interface has three columns: left, middle and right column. In the left column the user can launch queries, review the query history and identify the dimensions panes. The middle column displays categorized results that in this case are categorized according to the keyword, year of the documents and countries. While the right column is used to represent the entities that are used to characterize the result sets (Jonker *et al.*, 2005). The data extracted with TRIST are based on the hypothesis developed by Sandbox (Prolux *et al.*, 2006).

4.3.3 Data transformation

The data transformation in this project has been done using the Sandbox tool. This tool is used to create the set of hypotheses that would lead to data aggregations. It is similar to concept maps but has a better computational model in order to avoid node-link representation limitations. Sandbox offers the possibilities to manage layers of multiple dimensions of data. For the Avian Flu scenario, the initial step is to create a brainstorming using Sandbox in order to create the hypothesis concerning the avian flu. The main idea is to visualize the hypothesis and their interconnections. The illustration of the brainstorming using Sandbox is presented in Figure 9.

Fig 9. Brainstorming in Sandbox regarding the Avian Flu (Prolux *et al.*, 2006)

The SandBox tool enables the user to make annotations based on the hypothesis generated and after based on these to initiate the search on TRIST. The idea is to be able to discover trends and themes of the avian flu outbreak. For doing this discovery, it is important to relate the documents retrieved through TRIST based on the SandBox, to the location of their occurrence. The authors suggest the use of nSpace and GeoTime for creating a comprehensive environment as visual analytical tools.

4.3.4 The Software Tool

The software tool used in this project has been quite complex and has been developed based on the idea of aggregation of different analytical modules. As a result of this approach the use of nSpace has been introduced. According to the authors, nSpace basically combines several interactive visualization techniques in order to create a unified workspace with primary aim to support the analytical process. Moreover, in this project a geo-temporal visualization tool called GeoTime[3] has been used. GeoTime supports the visual analysis of data and events over time and location by using the 3D representation as presented in the Section 3. The interface of the GeoTime software is presented in the Figure 10.

---

3 A commercial software tool developed by Oculus Info Inc, demo can be found at http://www.oculusinfo.com/Taxi_Nov28.htm

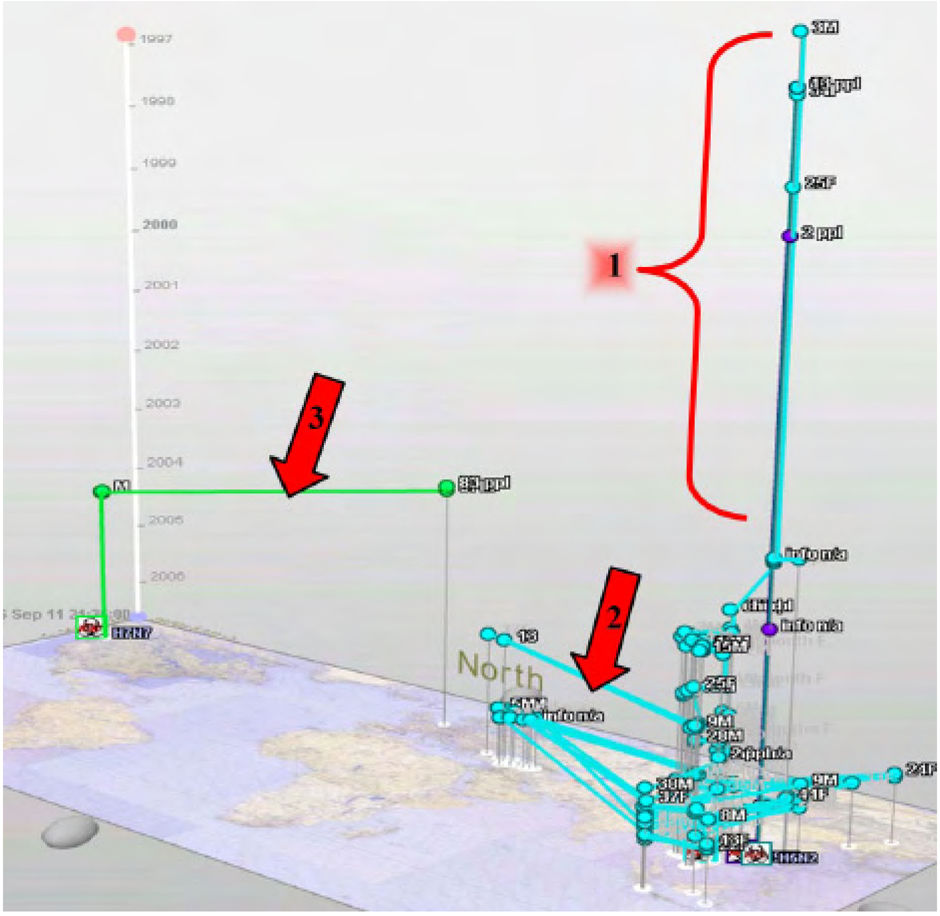

Fig 10. View of the data using GeoTime (Prolux *et al.*, 2006)

In the vertical axis, the time is shown while the XY plane represents the location (i.e., map). With number 1 in the Figure the frequency of avian flu cases from 1997 is shown. We can easily see that the increase of the frequency happens from 2003 and mainly in the Asia (represented by number 2). While with number 3 the disease spreading (still isolated cases) is shown, toward Europe and Canada. With this insight, the knowledge gained directs that potential hot spot of the avian flu is Asia with potentials for spreading toward Europe. Without localizing the data into the map, this kind of knowledge would not be possible to be gained easily. All this data is connectable with the other tools described in the paper like TRIST and SandBox, and thus, eventually creating a comprehensive visual analytics environment that is called nSpace.

4.4 FemaRepViz: Automatic Extraction and Geo-temporal Visualization of FEMA National Situations Update (Pan and Mitra, 2007)

This paper offers a description of a visualization toolkit, called FemaRepViz, that is used to extract and represent the summarized data from the FEMA (Federal Emergency Management Agency) web site. The information is extracted, categorized, localized and then presented in the map. The extraction of the news is based on the FactXtractor in order to extract the data entities form the text files. FactXtractor is a web service for Named Entity Recognition (NER) and Entity Relation Extraction (ERE). The information is extracted then processed using GeoTagger. For the visualization they use Concept Vista, Google Maps and Google Earth.

4.4.1 Motivation

As a main motivation for this project, the authors mention the necessity for rapid respond in crisis management. For achieving this there is need to process a vast amount of data in a very short time period. The need for such systems has been increased especially in the last couple of years and because of the disasters that have happened. Providing information in real time regarding the eminent natural disaster leads toward saving of humans' life. In this sense, the development of tools for getting insights into the aggregated abstract data available on different web sites is becoming a challenge for the scientific community. The ultimate idea of this project is to offer a decision support system in crisis cases.

4.4.2 Data Set

The data set used for this project is based on the FEMA web site and its National Situation Reports. These reports are aggregated from federal agencies, state and local governments as well as news media as well. The main purpose of these reports is to provide useful information for emergency management. These include information about the weather, eventual earthquake activities and other eventual incidents. In these reports, the name of location where the specific occurred is always included. Geographical information available in the FEMA National Situation Reports makes them a suitable data set for geo-temporal visualization.

4.4.3 Data Transformation

Data available on the FEMA National Situation reports needs to be extracted and processed in the computational model. The extraction of the information from the reports in this project is done by FactXtractor. FactXtrator processes text documents using the open source text processing platform, called General Architecture for Text Engineering (GATE)[4], while the relations among entities are identified using the Striped Dependency Tree (SDT) method. According to Pan and Mitra (2007), the SDT method is defined as a subtree of a dependency tree where each node has at least one descendant node. Main advantages of the SDT method are computational efficiency and better results in similarity measurement. In this paper, authors use SDT method with aim to create the

[4] Free software tool that can be downloaded from http://gate.ac.uk/

concept maps that can be visualized with ConceptVista[5] and formatted in the Web Ontology Language (OWL). The entire process of data extraction and transformation is illustrated in the Figure 11.

Fig 11. Data processing flow

Initial step is the identification of the parts of the speech, nouns etc. The second step is to create a Named Entity Extraction based on these PoS using GATE. The next step is to deep parse in order to create syntactic relationships in the form of Subject-Verb-Object that can be processed using the dependency trees.

4.4.4 The Software Tool

The FEMARepViz tool is built as a conjunction of different modules. Each of these modules has a specific task that enables the visualization of the information available in FEMA National Situations Reports. The first module of this tool is a statistical text extraction/processing module that primarily deals with text extraction and generation of concept maps as described in the previous section. The next module is a so-called disambiguation module that primarily aims to avoid eventual mistakes based on the extracted names of the locations and entities. This is especially important for geo-visualization since, as we know, there can be more then one location in the world with the same name. This process is initially done with use of GeoTagger. GeoTagger is a geocoding Web service that maps the location name that appears in the text in the FEMA reports into its geographical coordinates. The output of this process is a KML file that can be loaded using Google Maps. The disambiguation module starts a custom made algorithm for identifying the exact locations. It runs in four geographically defined levels: world, continent, country, and province. This basically means that if the country, state and province are extracted from the text segment, the location (its coordinates) is calculated at the Country level. For example, if in one FEMA report there are mentioned US and Sweden, the location is calculated only within

5 Can downloaded from http://www.geovista.psu.edu/ConceptVISTA/

those two countries. In Figure 12, the FEMARepViz interface is presented. The FEMARepViz visualizes the reports from FEMA using Google Earth. Each event is represented with a specific icon showing the type of incident or event that occurred. The name of location mentioned in these report are geo-coded into a KML file that are uploaded to Google earth using Network Link. The interaction with this application can be made using the right-mouse click over the incident icon. The translucent pane that appears shows the summary of the report for that incident. Each of the report has a time stamp so user can navigate using time-slide.

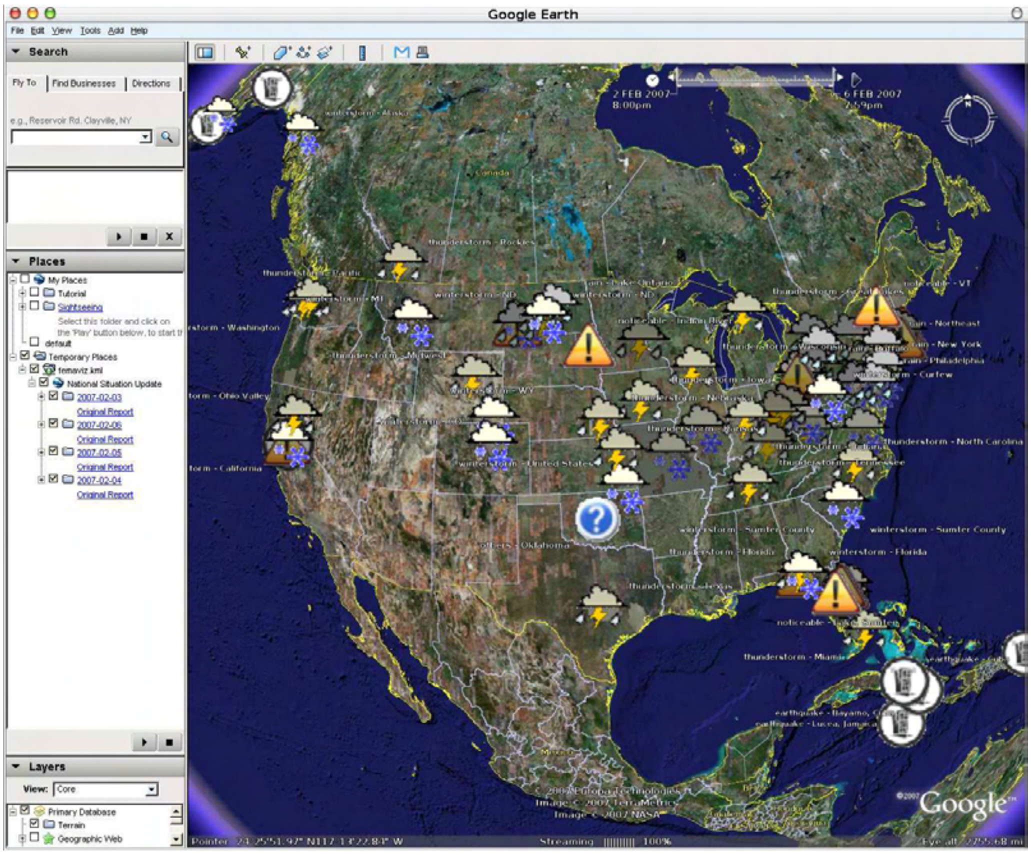

Fig 12. FEMRepViz interface showing the processed reports between Feb 3 – 6, 2007 (Pan and Mitra, 2007)

5. Discussion

In the previous section, we described four different projects that use different visualization tools for getting insight into the data. In these four recent research projects, there have been introduced four different software tools for visualization. For a better overview of the results of these research projects I will provide a simple evaluation in this section based on the specific set of criteria that are important in the visual analytics. These criteria are: the type of software, the data set, interaction techniques, domain independence, geo-temporal model, digital maps representation, evaluation and extensibility. With regard to the information visualization model presented in Figure 1, factors mentioned above could be interesting for analysis and evaluation of visualization tools. I will discuss each of these factors individually and I will present a tabular breakdown at the end of this section.

Type of Software – In the four projects, there have been introduced several software tools: PaleoAnalogs, Improvise, GeoTime, and FEMARepViz. The PaleoAnalogs tool looks to be free but not as an open source. Imporvise is both free and open source, while GeoTime is a commercial tool. FEMARepViz looks to be a closed software tool that at the moment is not available either as a download or a commercial product.

Data Set – The data sets used in these systems has been very different. Some of the software tools has had self-contained data set (like Improvise) while the other tools offered possibility for aggregating data from multiple data sources. GeoTime and FEMARepViz moreover offer the possibility for extraction of data from the web available text documents.

Interaction Techniques – All of the software tools presented in these projects offers different levels of interaction and customization. While the PaleoAnalogs, Improvise and GeoTime offer wide variety of the interaction techniques typical for visual analytics, such as space and time reasoning, Focus+Context, filtering and axis interaction, FEMARepViz seems to have a disadvantage regarding the interaction techniques used. I regard this do be a drawback since without offering the users with full control over the data, the analytical capabilities of the software tool might be limited.

Domain Independence – All projects described in these papers are tailored to a specific domain of usage like paleoceanography, service economy (hotel management), epidemiology, and crisis management. Based on the data sets used and the data transformation techniques it can be said that PaleoAnalogs is very much related to the domain of applications (MAT transformation). The other three software tools (Improvise, GeoTime and FEMARepViz) offer more flexibility in since the visualization is not related so much with the type of data available

Geo-temporal Model – All software tools presented in these research projects have the GeoTime representation model. Only GeoTime uses the full 3D model of location-time representation. The other three software tools have the so-called 2D model of location-time representation. This means that the time variable is implemented in the series of the snapshots. This might be a usability issue if there is heavy correlation of variables over time.

Digital Maps Representation – Since for the geo-temporal visualization analysis the usage of digital maps is inevitable, I used this aspect as evaluation criteria. PaleoAnalogs and Improvise seems to have internal digital maps that they are using for visualization. The GeoTime tool has also internal but with the possibility of change and access other maps via ESRI[6] ArcGIS 9 architecture gateway. While the FEMARepViz uses the Google Maps and Earth services by implementing KML location files.

Evaluation – In the three projects described, there have not been an evaluation use case of the software tools used. Only the Hotel Visualization project, where Imporvise tool was used, there have been an evaluation based on e-Delphi, a web-based toolkit designed to support Delphi exercises. Also in GeoTime, the authors refer to a evaluation done by The National Institute of Standards and Technology (NIST). Since GeoTime is a commercial software tool this reference seems to be more for marketing purposes. In general, I would prefer to have an evaluation that would measure the effectiveness and the usability of the visualization tools.

Extensibility – The last criteria that I used for the categorization of the software tools is the possibility for extending with new modules. Overall they offer good prospect for extensibility and adaptations but this is primarily related to the type of the software. In this aspect, the Improvise tool has a clear advantage since it is an open source software and thus, potentially, easily extensible with new modules and visual representation techniques. Also GeoTime seems to have this possibility referring to the work done by Eccles *et al.* (2007) where they show the extensibility of the software with the story window. But being a commercial product the extensibility of the GeoTime software is potentially a factor of financial constrains.

In Table 1, the breakdown of the software tools is represented, based on the set of criteria described in this section.

---

[6] Large GIS Company. More info at http://www.esri.com/

Table 1 – Software tools breakdown based on different criteria

Software tool	Type	Data Set	Interaction	Domain Independence	Geo-temporal Model	Digital Maps	Evaluation	Extensibility
PaleoAnalogs	Free	Multiple/ external	High	Low	2D	internal	No	NA
Improvise	Open Source	Single/ Self-cont.	High	High	2D	internal	Yes	Yes
GeoTime	Commercial	Multiple/ external	High	High	3D	internal/ external	Yes	Yes
FEMARepViz	Unknown	Multiple/ external	Limited	Medium	2D	external	No	NA

6. Conclusions

The possibility for data generation has increased to the level that the "overload of data" problem has become evident in the majority of the application domains. Making sense of these data with help of only basic database tools (such as queries) is practically impossible. There are two basic challenges faced in the process of solving this problem: variety of the data and their complexity. The solution suggested is information visualization as a multidisciplinary approach that combines the human cognition capabilities with database queries, data exploration and browsing etc. in order to offer better insight into the abstract data. Geo-temporal visual analysis as a subfield of visual analytics is a great tool for analyzing large set of data in multidimensional spaces and relative to the location of occurrence. This paper offers a description of four different tools for different applications domains. Even without being an expert domain for these projects, it is easy to identify the usefulness of these tools. In general, visualization tends to help users gaining knowledge from abstract data and thus eventually leading to better decision. In this sense, the visualization tools could be regards as a sort of decision support systems.

7. References

Beynon-Davies, P. (2004) "Database Systems – 3rd edition" Palgrave Macmillan, New York, US

Rivero, L., Doorn, J. and Ferraggine, V. (2006) "Encyclopedia of Database Technologies and Applications" Idea Group, London, UK

Ramakrishnan, R. and Gehrke, J. (1999) "Database Management Systems – 2nd Edition" McGraw-Hill, New York, US

Spence, R. (2001) "Information Visualization" Pearson Education Limited, Essex, UK

Kerren, A. (2007) "Information Visualization" Växjö University – MSI, lecture slides

Shuping, D. and Wright, W. (2005) "Geo-Temporal Visualization of RFID" Direction Magazines Online, retrieved on March 5, 2008 from: http://www.directionsmag.com/article.php?article_id=951&trv=1

Theron R. (2006) "Visual Analytics of Paleoceanographic Conditions", IEEE Symposium on Visual Analytics Science and Technology - VAST 2006, vol. 1, pp. 19-26

Weaver, C., Fyfe, D., Robinson, A., Holdsworth, D., Peuquet, D. and MacEachren, A. (2006) "Visual Analysis of Historic Hotel Visitation Patterns", IEEE Symposium on Visual Analytics Science and Technology - VAST 2006, vol. 1 , pp. 35-42

M.Wattenberg (2002) "Arc diagrams: Visualizing structure in strings", IEEE Symposium on Information Visualization, pp.110–116

Prolux, P., Tandon, S., Bodnar, A., Schorh, D., Harper, R. and Wright., W. (2006) "Avian Flu Case Study with nSpace and GeoTime", IEEE Symposium on Visual Analytics Science and Technology - VAST 2006, vol. 1 , pp. 27-34

Jonker, D., Wright, W., Schroh, D., Proulx, P. and Cort, B. (2005)"Information Triage with TRIST", In Proc. International Conference on Intelligence Analysis

Pan, C. and Mitra, P. (2007) "FemaRepViz: Automatic Extraction and Geo-Temporal Visualization of FEMA National Situation Updates", IEEE Symposium on Visual Analytics Science and Technology - VAST 2007, pp. 11-18

Eccles, R., Kapler, T., Harper, R., Wright, W. (2007) "Stpries in GeoTime", IEEE Symposium on Visual Analytics Science and Technology - VAST 2007, pp. 19-26

Wong, P. and Thomas, J. (2004) "Visual Analytics" Computer Graphics and Applications, IEEE, Vol.24, Issue 5, Pp. 20-21