INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Aviz

Analysis and Visualization

Saclay - Île-de-France

Theme : Interaction and Visualization

Activity Report

2009

Table of contents

1. Team

Research Scientist
Jean-Daniel Fekete [Team Leader, Research Director, HdR]
Pierre Dragicevic [Research Scientist]

External Collaborator
Anastasia Bezerianos [École Centrale Paris, Assistant Professor]

Technical Staff
Julie Stromer [Sep. 2008 to May 2009, Engineer on the Taranis project]
Johan Euphrosine [Feb. to May 2009, Engineer on the Taranis project]
Pierre-Luc Hemery

PhD Student
Fadia Azaiez [12/2008 to 10/2009, Co-Advised by Jean-Daniel Fekete, Ioana Manolescu (INRIA, Gemo) and Véronique Benzaken (Paris-Sud, BD)]
Wael Khemiri [09/2008, Co-Advised by Jean-Daniel Fekete, Ioana Manolescu (INRIA, Gemo) and Véronique Benzaken (Paris-Sud, BD)]

Post-Doctoral Fellow
Fanny Chevalier [Jan. 2008 to Aug. 2010]
Nadia Boukhelifa [Oct. 2009 to Oct. 2010]
Tomer Moscovich [Jul. 2008 to Aug. 2009]

Administrative Assistant
Marie-Jeanne Gaffard [until Sep. 2009]
Hélène Milome [since Sep. 2009]

Other
Jiacheng Li [BSc Zhejiang University, China]
Cyprien Pindat [M1 Polytechnique]

2. Overall Objectives

2.1. Objectives

Like many other fields, all the sciences are being transformed by our rapidly-increasing abilities to collect, manage and understand vast amounts of data. A 2003 study estimated that the amount of data produced in the world was increasing by 50% each year [29]. According to SearchEngineWatch[1], the amount of information made available through Internet search engines has grown exponentially for the last decade, and major Web search engines currently index more than 2 billion documents. However, since our brains and sensory capacities have not changed in the meantime, gaining competitive advantage from all this data depends increasingly on the effectiveness with which we support human abilities to perceive, understand, and act on it.

The focus of the AVIZ project is to design methods and tools that make analyzing large data sets easy and massive data sets possible. Our interests include:

- *Methods to visualize and smoothly navigate through large datasets;*
- *Efficient analysis methods to reduce huge datasets to visualizable size;*
- *Evaluation methods to assess their effectiveness and usability; and*
- *Engineering tools for building visual analytics systems that can access, search, visualize and analyze large datasets with smooth, interactive response.*

[1] http://www.searchenginewatch.com

2.2. Research Themes

AVIZ's research on Visual Analytics is organized around four main Research Themes:

Methods to visualize and smoothly navigate through large data sets Large data sets challenge current visualization and analysis methods. Understanding the structure of a graph with one million vertices is not just a matter of displaying the vertices on a screen and connecting them with lines. Current screens only have around two million pixels. Understanding a large graph requires both data reduction to visualize the whole and navigation techniques coupled with suitable representations to see the details. These representations, aggregation functions, navigation and interaction techniques must be chosen as a coordinated whole to be effective and fit the user's mental map.

AVIZ designs new visualization representations and interactions to efficiently navigate and manipulate large data sets.

Efficient analysis methods to reduce huge data sets to visualizable size Designing analysis components with interaction in mind has strong implications for both the algorithms and the processes they use. Some data reduction algorithms are suited to the principle of sampling, then extrapolating, assessing the quality and incrementally enhancing the computation: for example, all the linear reductions such as PCA, Factorial Analysis, and SVM, as well as general MDS and Self Organizing Maps. AVIZ investigates the possible analysis processes according to the analyzed data types.

Evaluation methods to assess their effectiveness and usability Evaluation of Visual Analytics tools is currently a challenge. Traditional HCI evaluation has focused on measuring performance (speed, error rate) for well-specified tasks. Visual Analytics is about developing insights from data. Measuring the number or quality of insights is difficult and not well understood. To address this problem, we have been actively working in three different directions: organizing workshops to gather experience and principles from researchers, co-organizing the Information Visualization Contest[5] to establish benchmarks for Information Visualization, and developing a framework to help evaluate Information Visualization applications built using the InfoVis Toolkit. To improve evaluation, we want to improve both theoretical and practical methods. We plan to add experiment modules into the InfoVis Toolkit to simplify the planning and realization of controlled experiments.

Engineering tools for building visual analytic systems that can access, search, visualize and analyze large data sets with smooth, interactive response.

Currently, databases, data analysis and visualization all use the concept of data tables made of tuples and linked by relations. However, databases are storage-oriented and do not describe the data types precisely. Analytical systems describe the data types precisely, but their data storage and computation model are not suited to interactive visualization. Visualization systems use in-memory data tables tailored for fast display and filtering, but their interactions with external analysis programs and databases are often slow.

AVIZ seeks to merge three fields: databases, data analysis and visualization. Part of this merging involves using common abstractions and interoperable components. This is a long-term challenge, but it is a necessity because generic, loosely-coupled combinations will not achieve interactive performance.

These themes are presented separately, but are closely linked: a good multi-scale visualization technique relies on an analysis method to generate the suitable data structure. The effectiveness of a Visual Analytics tool has to be evaluated at several levels (component, system, environment). Finally, to build Visual Analytics systems that manage large data sets, the software infrastructure has to provide the right abstractions and mechanisms. Therefore, each of the four research themes work together. One of the scientific challenges is to fit them all together into a coherent framework supporting the analyst's work process.

2.3. Highlights

Visualization and Interaction on Wall-Size Displays AVIZ has worked on new visualization and interaction methods for Wall-Size display such as the new WILD platform recently installed at INRIA/LRI.

We have investigated important issues to transform a visualization system designed for one user on a personal machine to explore a large social network into a collaborative colocated visualization for several users using several input devices on a wall-size display [8].

We designed a novel interaction technique called “Motion Pointing” to select targets hard or impossible to reach using motion instead of position [13].

Navigation in large node-link networks Exploring very large networks involves navigating in very large information spaces. We designed two navigation techniques to improve that navigation using a standard mouse [14].

Strong collaboration on Database/Workflow/Visualization AVIZ has continued its strong collaboration with the GEMO INRIA project-team and the BD LRI group to tackle the problem of “Interactive data-intensive workflows for scientific applications”.

We have designed an initial workflow model for managing scientific data and started to present it to workshops [18].

3. Scientific Foundations

3.1. Scientific Foundations

The scientific foundations of Visual Analytics lie primarily in the domains of Information Visualization and Data Mining. Indirectly, it inherits from other established domains such as graphic design, Exploratory Data Analysis (EDA), statistics, Artificial Intelligence (AI), Human-Computer Interaction (HCI), and Psychology.

The use of graphic representation to understand abstract data is a goal Visual Analytics shares with Tukey’s Exploratory Data Analysis (EDA) [33], graphic designers such as Bertin [21] and Tufte [32], and HCI researchers in the field of Information Visualization [20].

EDA is complementary to classical statistical analysis. Classical statistics starts from a *problem*, gathers *data*, design a *model* and performs an *analysis* to reach a *conclusion* about whether the data follows the model. While EDA also starts with a problem and data, it is most useful *before* we have a model; rather, we perform visual analysis to discover what kind of model might apply to it. However, statistical validation is not always required with EDA; since often the results of visual analysis are sufficiently clear-cut that statistics are unnecessary.

Visual Analytics relies on a process similar to EDA, but expands its scope to include more sophisticated graphics and areas where considerable automated analysis is required before the visual analysis takes place. This richer data analysis has its roots in the domain of Data Mining, while the advanced graphics and interactive exploration techniques come from the scientific fields of Data Visualization and HCI, as well as the expertise of professions such as cartography and graphic designers who have long worked to create effective methods for graphically conveying information.

The books of the cartographer Bertin and the graphic designer Tufte are full of rules drawn from their experience about how the meaning of data can be best conveyed visually. Their purpose is to find effective visual representation that describe a data set but also (mainly for Bertin) to discover structure in the data by using the right mappings from abstract dimensions in the data to visual ones.

For the last 25 years, the field of Human-Computer Interaction (HCI) has also shown that interacting with visual representations of data in a tight perception-action loop improves the time and level of understanding of data sets. Information Visualization is the branch of HCI that has studied visual representations suitable to understanding and interaction methods suitable to navigating and drilling down on data. The scientific foundations of Information Visualization come from theories about perception, action and interaction.

Several theories of perception are related to information visualization such as the "Gestalt" principles, Gibson's theory of visual perception [25] and Triesman's "preattentive processing" theory [31]. We use them extensively but they only have a limited accuracy for predicting the effectiveness of novel visual representations in interactive settings.

Information Visualization emerged from HCI when researchers realized that interaction greatly enhanced the perception of visual representations. To be effective, interaction should take place in an interactive loop faster than 100ms. For small data sets, it is not difficult to guarantee that analysis, visualization and interaction steps occur in this time, permitting smooth data analysis and navigation. For larger data sets, more computation should be performed to reduce the data size to a size that may be visualized effectively.

In 2002, we showed that the practical limit of InfoVis was on the order of 1 million items displayed on a screen [23]. Although screen technologies have improved rapidly since then, eventually we will be limited by the physiology of our vision system: about 20 millions receptor cells (rods and cones) on the retina. Another problem will be the limits of human visual attention, as suggested by our 2006 study on change blindness in large and multiple displays [1]. Therefore, visualization alone cannot let us understand very large data sets. Other techniques such as aggregation or sampling must be used to reduce the visual complexity of the data to the scale of human perception.

Abstracting data to reduce its size to what humans can understand is the goal of Data Mining research. It uses data analysis and machine learning techniques. The scientific foundations of these techniques revolve around the idea of finding a good model for the data. Unfortunately, the more sophisticated techniques for finding models are complex, and the algorithms can take a long time to run, making them unsuitable for an interactive environment. Furthermore, some models are too complex for humans to understand; so the results of data mining can be difficult or impossible to understand directly.

Unlike pure Data Mining systems, a Visual Analytics system provides analysis algorithms and processes compatible with human perception and understandable to human cognition. The analysis should provide understandable results quickly, even if they are not ideal. Instead of running to a predefined threshold, algorithms and programs should be designed to allow trading speed for quality and show the tradeoffs interactively. This is not a temporary requirement: it will be with us even when computers are much faster, because good quality algorithms are at least quadratic in time (e.g. hierarchical clustering methods). Visual Analytics systems need different algorithms for different phases of the work that can trade speed for quality in an understandable way.

Designing novel interaction and visualization techniques to explore huge data sets is an important goal and requires solving hard problems, but how can we assess whether or not our techniques and systems provide real improvements? Without this answer, we cannot know if we are heading in the right direction. This is why we have been actively involved in the design of evaluation methods for information visualization [5] [28], [26], [27], [24]. For more complex systems, other methods are required. For these we want to focus on longitudinal evaluation methods while still trying to improve controlled experiments.

4. Application Domains

4.1. Application Domains

AVIZ develops active collaboration with users from various application domains, making sure it can support their specific needs. By studying similar problems in different domains, we can begin to generalize our results and have confidence that our solutions will work for a variety of applications. Our current application domains include:

- Social Network Analysis, in cooperation with Microsoft Research, the University of Calgary and the French National Archives;
- Biological research, in cooperation with INRA, the IGM Biological Research Laboratory at Univ. Paris-Sud and Institut Pasteur;

- Digital Libraries, in cooperation with the French National Archives, the Bibilothèque Nationale and the Wikipedia community;
- Global Security, in cooperation with training centers for urban communities at risk;
- eScience, in collaboration with Microsoft Research (see 6.1) and with the INRIA Gemo and LRI BD groups (see 6.5).

4.1.1. Social Network Analysis

In the social networks domain, we are working on exploratory visualization. Current studies in social networks presuppose that users know the nature of the networks they want to explore and the kinds of of transformations and layouts that will best suit their needs. This is often not true, and tools are very weak at helping users understand the nature of their networks and the transformations they could perform to get meaningful insights. This work began in 2004 with the arrival of Nathalie Henry in the Project.

We have been focusing on the use of the matrix representation to explore large graphs, building on our previous work using matrices for constraint-based programming. Matrices present challenging problems both interactively and mathematically. We are designing an interactive system to help users navigate and interact with large matrices. We are also preparing a survey on methods to reorder matrices, whether from graphs from tabular data.

We are now expanding our research towards multivariate social networks and, more importantly, collaborative visualization and exploration of social networks using the wall-size display WILD recently installed at INRIA/LRI.

4.1.2. Biological Visualization

Bioinformatics uses many complex data structures such as phylogenetic trees and genomes made of multi-scale parts (sequences of base pairs, genes, interaction pathways etc.) Biologists navigate through multitudes of these varied and complex structures daily in complex, changeable, data- and insight-driven paths. They also often need to edit these structures to annotate genes and add information about their functions. Visual Analytics is a powerful tool to help them, as we are currently pursuing in the Microbiogenomics project (see section 6.4) and in an exploratory work with Institut Pasteur on "New Generation Sequencing" and the exploration of huge biological datasets.

4.1.3. Digital Libraries

In the digital Library domain, we collaborate with Wikipedia contributors to improve Wikipedia, as well as with historians such as the French National Archives on the National Center of Renaissance on exploratory projects to visualize and analyze historical documents (see 6.1.)

4.1.4. Global Security

In the global security domain, we collaborate with MASA Group (Mathématique Appliquée S.A.) and civil crisis managers from the CODAH (Communauté d'Agglomération Havraise) on interactive visualization tools for improving the quality of crisis management training exercices (see 6.2.)

4.1.5. eScience

Part of our research consists in supporting traditional sciences with high-level tools to help analyze and make sense of large datasets. We apply our tools and techniques to biology, social sciences and to Wikipedia which has become a major supporting tool for scientists (see 6.1). We also design software infrastructures to help scientist perform their analytical tasks with high-level tools instead of having to learn complex tools requiring computer science skills (see 6.5).

5. Software

5.1. ScatterDice

Participants: Jean-Daniel Fekete [correspondant], Pierre Dragicevic, Niklas Elmqvist.

ScatterDice [2] is a visualization system for exploring multidimensional datasets. It shows a scatter plot of 2 dimensions among the multiple ones available and provides a very simple paradigm of 3D rotation to change the visualized dimensions. The navigation is controlled by a scatter plot matrix that is used as a high-level overview of the dataset as well as a control panel to switch the dimensions. ScatterDice received the "best paper award" at the IEEE InfoVis 2008 conference.

ScatterDice works on any tabular dataset (e.g. CSV file) and is available as a Java Web Start application. See its web page for details: http://engineering.purdue.edu/~elm/projects/scatterdice.html.

We have extended ScatterDice to support the exploration of multivariate social networks. The GraphDice system show networks using a node-link diagram representation as a scatter plot with links drawn between connected nodes.

5.2. The InfoVis Toolkit

Participant: Jean-Daniel Fekete [correspondant].

The InfoVis Toolkit [3] is an Interactive Graphics Toolkit written in Java to facilitate the development of Information Visualization applications and components.

The main characteristics of the InfoVis Toolkit are:

Unified data structure The base data structure is a table of columns. Columns contain objects of homogeneous types, such as integers or strings. Trees and Graphs are derived from Tables.

Small memory footprint Using homogeneous columns instead of compound types dramatically reduces the memory required to store large tables, trees or graphs, and usually also the time required to manage them.

Unified set of interactive components Interactive filtering (a.k.a. dynamic queries) can be performed with the same control objects and components regardless of the data structure, simplifying the reuse of existing components and the design of generic ones.

Fast The InfoVis Toolkit can use accelerated graphics provided by Agile2D[2], an implementation of Java2D based on the OpenGL API for hardware accelerated graphics [23]. On machines with hardware acceleration, some visualizations redisplay 100 times faster than with the standard Java2D implementation.

Extensible The InfoVis Toolkit is meant to incorporate new information visualization techniques and is distributed with the full source and a very liberal license. It can be used for student projects, research projects or commercial products.

The InfoVis Toolkit, as of version 0.9, implements nine types of visualization (Fig. 1): Time Series, Scatter Plots, Parallel Coordinates and Matrices for tables, Node-Link diagrams, Icicle trees and Treemaps for trees, Adjacency Matrices and Node-Link diagrams (with several layouts) for graphs.

The InfoVis toolkit is used for teaching the Information Visualization course (Masters level, Univ. of Paris-Sud) and is the basis for all AVIZ contracts. It is our main development platform for information visualization; most of our Information Visualization prototypes rely on it. It is available at http://ivtk.sourceforge.net.

5.3. XML/TEI Eclipse Plugin

Participant: Jean-Daniel Fekete [correspondant].

The Millefeuille Platform is a Plugin for the Eclipse programming environment designed to assist historians in encoding their documents. It provides a set of mechanisms found in standard programming environments but not known by historians, including source version control (SVN), project management and asynchronous collaboration tools. The Plugin improves the standard XML editor of the Eclipse Platform with several encoding help for building multiple indexes, verifying the consistency of XML encoding based on high-level properties and dynamically apply a stylesheet to the XML encoded files. Indexes are one kind of cross-document structures that the Plugin can dynamically create from the XML structure.

[2] http://www.cs.umd.edu/hcil/agile2d

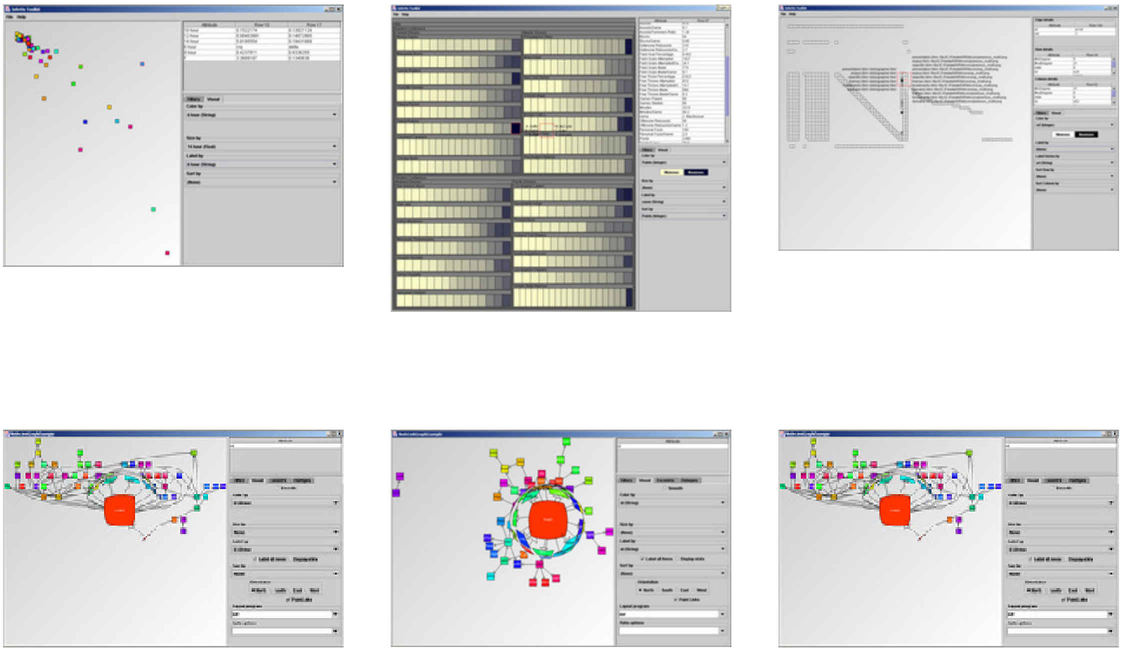

Figure 1. Several visualizations produced using the Infovis Toolkit

The Millefeuille Platform is currently used to encode in XML/TEI a sample of the administration of France during 100 years, from just before the revolution to 100 years later. The encoding is done using a generated XML Schema based on TEI P5 (see http://www.tei-c.org/Guidelines/P5/. It is generated with the Roma tool (http://tei.oucs.ox.ac.uk/Roma/).

Since September 2009, Millefeuille has been worked on actively by Cyril Masset from "Service éditorial et publications électroniques" (SEPE) of IRHT (http://www.irht.cnrs.fr/recherche/sepe.htm. It is used by several projects in the Humanities including Ethiopian studies, Renaissance studies and history of the French Administration.

The Platform is described in available at http://tei-eclipse.gforge.inria.fr.

6. Contracts and Grants with Industry

6.1. ReActivity

Participants: Nadia Boukhelifa, Fanny Chevalier, Pierre Dragicevic, Jean-Daniel Fekete [correspondant], Tomer Moscovich.

This project belongs to the joint INRIA-Microsoft Research Laboratory and is a collaboration of the VIBE Group at Microsoft Research in Redmond, the in|situ| and AVIZ INRIA groups. It is a three-year project started in 2007, focused on analyzing researchers' activities to help them reflect on these activities, analyze them or communicate them more effectively. The project has to deal with logging, storing, summarizing, visualizing and interacting with activity data to solve interesting problems in science.

Both VIBE and INRIA are faced with difficult problems in term of data capture, management, retrieval, effective visualization of stored data, effective aggregation, higher-level summarization (inferring the high-level user activity from the captured low-level user activity) and reflective presentation of that information. The teams are collaborating in designing Information Visualization infrastructures capable of managing large amounts of information and interacting with it. The ReActivity project involves logging, visualizing and interacting with logged data. It is split into three phases: collecting the logs in a consistent, extensible and robust way, mining the logs to extract higher-level information and visualizing the information for understanding, interaction and sharing. It addresses these issues for simple desktop-based information initially and then increase the scope of the project by aggregating information from outside sources.

We have also started to work on providing group awareness mechanisms to Wikipedia contributors. We organized two participatory design workshop with important contributors of the French Wikipedia and gathered a set of requirements and processes. From that, we have designed a set of interactive components and visualizations that seem important to improve the collective writing of Wikipedians. Some of the information required to these components and visualizations is not provided by the standard Wikipedia tables accessible on the web; it has to be computed. We are working on mechanisms to compute this information effectively to be able to test the components with real users doing real tasks. Important information include the amount of changes made by each users on each page, the ratio between the number of characters entered by a contributor and finaly remaining on wikipedia pages etc. This information is important to quickly assess the profile of contributors to quickly monitor changes and raise the overall quality of Wikipedia [11], [17].

We have worked on the software infrastructure needed to support high-level awareness for Wikipedia and this had been much harder than expected due to the amount of data to manage and summarize. We are expecting that a summarization server will be available for the partners in the begining of 2010.

6.2. TARANIS: Technologies for the Appraisal of Risks through Animation and Simulation

Participants: Pierre Dragicevic [correspondant], Julie Stromer, Johan Euphrosine.

The TARANIS 2-year ANR project (program “Concepts, systèmes et outils pour la sécurité globale”) started in 2006, supported by MASA, ESRI-France and INRIA. It aims at creating a new training system for crisis managers, based on innovative simulation tools, allowing trainers to easily recreate complex crisis situations. Simulation tools give the trainer unprecedented control over the training session while making the virtual crisis reactive to the trainees actions and providing an unlimited variety of extreme crisis situations, a challenge that even very expensive ground exercises cannot meet. TARANIS is a “Global Security” project conducted by the company “Mathématique Appliquée S.A.” to design crisis simulation environment.

The goal of the Aviz team in this project is to design graphical user interfaces and visualization tools for improving the quality of crisis training exercices. To this end, we have been working closely with MASA Group and members of the CODAH (Communauté de l’agglomération havraise), who are regularly organizing crisis management exercices to train people who could be involved in a civil crisis in the city of Le Havre. We conducted observational studies and participatory design sessions in Le Havre, which together suggested that the quality of the training exercices can be significantly improved by 1) increasing group activity awareness during the exercise and 2) adding logging mechanisms to enhance feedback to the trainees during the final debriefing session.

With these two goals in mind, we designed and implemented an augmented crisis animation setup in order to assist the animators in their tasks without significantly changing their working habits. This setup includes:

- A collaborative interactive timeline visualization, which allows animators to see the evolution of the scenario in real-time, assign tasks and change their status, and add annotations they can reuse during their debriefing session. This timeline relies on a client-server architecture that has been developed together with MASA, and supports multiple displays and multiple sources of input. The current setup uses a shared (projected) display as output, a tablet PC, a wireless keyboard and a jog dial as input, and an additional client laptop for the observer.

- A phone conversation logging mechanism, which records all communications occurring between the animation room and the crisis room, and allows the animators to replay and comment on conversations during the debriefing session.
- A simulation and GIS tools, developed by MASA and ESRI-France.

The augmented crisis animation room has been tested during several crisis training exercices and received very positive feedback; it has now become accepted as an improvement of the training process.

For more information, see http://www.aviz.fr/taranis and http://www.masagroup.net/solutions/customer-cases/taranis.html.

6.3. VisMaster: Visual Analytics - Mastering the Information Age

Participants: Jean-Daniel Fekete [correspondant], Fanny Chevalier, Pierre-Luc Hemery.

VisMaster is a European Coordination Action Project focused on the research discipline of Visual Analytics: One of the most important challenges of the emerging Information Age is to effectively utilise the immense wealth of information and data acquired, computed and stored by modern information systems. On the one hand, the appropriate use of available information volumes offers large potential to realize technological progress and business success. On the other hand, there exists the severe danger that users and analysts easily get lost in irrelevant, or inappropriately processed or presented information, a problem which is generally called the information overload problem. Visual Analytics is an emerging research discipline developing technology to make the best possible use of huge information loads in a wide variety of applications. The basic idea is to appropriately combine the strengths of intelligent automatic data analysis with the visual perception and analysis capabilities of the human user.

With VisMaster, we want to push the limits of today's Visual Analytics. To achieve this goal, we formed a Coordination Action to join European academic and industrial R&D excellence from several individual disciplines, forming a strong Visual Analytics research community. The project is divided into an array of thematic working groups that focus on advancing the state of the art in Visual Analytics. Specifically, the working groups will join excellence in the fields of data management, data analysis, spatial-temporal data, and human visual perceptionresearch with the wider visualisation research community.

The VisMaster Project main goals are to:

- form and shape a strong European Visual Analytics community
- define the European Visual Analytics Research Roadmap
- expose public and private stakeholders to Visual Analytics technology
- set the stage for larger follow-up Visual Analytics research initiatives in Europe.

In the VisMaster project, AVIZ is in charge of the Work Package 4: Infrastructure for Visual Analytics. This work-package is responsible for providing and maintaining the communication infrastructure for the collection of resources in other work-packages and the dissemination of the project results to the public. The scientific management board, consisting of all work-package leaders and chaired by the scientific manager also works in the domain of this work-package and is responsible for the coordination of workshops and the invitation of new community partners.

We have conducted one workshop on software infrastructure for visual analytics and started a google code project called “Obvious” to validate a unified infrastructure for Visual Analytics. Since October, Pierre-Luc Hemery has joined AVIZ as an engineer for 2 years to implement the specifications and organize a software community around it, see http://code.google.com/p/obvious.

The project VisMaster CA acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framwork Programme (FP7) for research of the European Commission (EC) under FET-Open grant number 225429. For more information, see http://www.vismaster.eu.

6.4. Integrated Resources for Microbial Genomics

Participant: Jean-Daniel Fekete [correspondant].

Microbiogenomics is a 3-year ANR project (program “Masses de données”) stated in 2006. The project is designed to address the challenges raised by the ongoing deluge of genomic data. It plans at designing an integrating resources for microbial genomics. The objective is to gather the largest amount of relevant data and to make it available for a number of data mining approaches, despite its heterogeneity. A graphic interface will be designed for efficient and simple but still expressive queries, letting users extract relevant pieces of knowledge through a visual interactive system. This will make cross-fertilization between domains possible, and allow detailed analysis of a wide range of available genomic data.

With Stéphane Descorps-Declère, we have started to design a annotation editor based on a multi-scale interaction paradigm. We believe that all the multi-scale navigation and visualization techniques AVIZ have designed recently can be effectively applied to the problem of comparative annotation for microbian genomes, showing relations between genes or proteins, as well as effective navigation methods similar to Mélange [7] and topology-based navigation in networks [14].

Partnership: IGM (Univ. Paris-Sud), LRI (Univ. Paris-Sud), MIG (INRA).

6.5. Interactive data-intensive workflows for scientific applications

Participants: Jean-Daniel Fekete [correspondant], Ioana Manolescu [INRIA GEMO Project Team], Véronique Benzaken.

Today’s scientific data management applications involve huge and increasing data volumes. Data can be numeric, e.g. output of measure instruments, textual, e.g. corpora studied by social scientists which may consist of news archives over several years, structured as is the case of astronomy or physics data, or highly unstructured as is the case of medical patient files. Data, in all forms, is increasingly large in volume, as a result of computers capturing more and more of the work scientists used to do based on paper, and also as a result of better and more powerful automatic data gathering tools, e.g. space telescopes, focused crawlers, archived experimental data (mandatory in some types of government-funded research programs) and so on.

The availability of such large data volumes is a gold mine for scientists which may carry research based on this data. Today’s scientists, however, more often than not rely on proprietary, ad-hoc information systems, consisting perhaps of a directory structure organized by hand by the scientist, a few specialized data processing applications, perhaps a few scripts etc.

For example, social scientists are interested in analyzing online social networks such as Wikipedia where new forms of group organization emerge. Visualizing the hypertext network that connects articles together requires accessing the hypertext data, computing some “shape” to visualize the network and using visualization tools to navigate the representation effectively. We have designed the Zoomable Adjacency Matrix Explorer (ZAME [22]) that allows the exploration by computing a linear ordering of the articles contained in Wikipedia using a fast and complex dimension reduction algorithm (see figure 1). However, all the required steps to access the data, compute the ordering, store it for reuse, visualize it and navigate on the representation is done using ad-hoc methods, very tedious to implement and out of reach of the sociologists who are interested by the study.

Off-the-shelf databases are not well adapted for scientific data management for several reasons.

First, database systems are not very flexible: changing the schema in a relational database management system (RDBMS) is very difficult, whereas exploratory usage of data routinely requires adding it new dimensions e.g., building summary categories to help the user tame the data complexity and volume. More flexible formats, such as XML or RDF, bring their own problems, which for the time being are mostly performance ones!

Second, database systems are tuned towards specific declarative search operations, typically expressed using a query language. In contrast, exploring scientific data involves operations such as clustering and finding interesting data orders, which cannot be specified based on stored attributes, but have to be discovered by complex, possibly iterative computations.

Finally, databases support query-based interactions, but lack more friendly interfaces, allowing the user to inspect a large data set, with varying level of detail for different, dynamically specified subsets [30].

The purpose of the project is *to investigate models, algorithms, and propose an architecture* of a system helping scientists to organize and make the most out of their data. The research work spans over three related, yet distinct areas, among which we expect it to build bridges: workflow modeling; database execution and optimization; and information visualization.

7. Other Grants and Activities

7.1. International actions

- Jean-Daniel Fekete was invited on Aug. 17-19, 2009 at the Summer Course on Information Visualization at the Pekin University to present the work of AVIZ and a survey of research work in Europe;
- Jean-Daniel Fekete was invited on Aug. 20, 2009 as a keynote speaker at IEEE CAD/Graphics 2009, Yellow Mountains, China;
- Pierre Dragicevic was invited on Nov. 12 to Dec. 20 at Hasso Plattner Institute in Berlin/Potsdam to collaborate with its prestigious HCI group.

7.2. National actions

- Fanny Chevalier and Jean-Daniel Fekete continued the seminars on HCI and Visualization in the Parisian Area with prestigious presenters (see http://www.aviz.fr/~chevalie/seminarhcivisu/home.php).

8. Dissemination

8.1. Journal editorial board

- Co-Editor in Chief of the new “Journal d’Interaction Personne Système” supported by the French-speaking Association of HCI: Jean-Daniel Fekete
- Associate Editor of the International Journal of Human-Computer Studies (IJHCS): Jean-Daniel Fekete

8.2. Journal reviewing

- Information Visualization Journal, Palgrave Macmillan: Jean-Daniel Fekete
- Document Numérique, Hermès, France: Jean-Daniel Fekete
- IEEE Transactions on Visualization and Computer Graphics: Jean-Daniel Fekete, Fanny Chevalier, Pierre Dragicevic
- International Journal of Human-Computer Studies, Elsevier: Pierre Dragicevic, Jean-Daniel Fekete, Fanny Chevalier
- Interacting with Computers, Elsevier: Pierre Dragicevic
- Technique et Science Informatiques: Pierre Dragicevic

8.3. Conference organization

- IEEE Symposium on Information Visualization 2009: Jean-Daniel Fekete (Paper co-chair), Pierre Dragicevic (Program Committee member)
- IEEE VAST 2009: Jean-Daniel Fekete (Program Committee member)
- IEEE EuroVis 2010: Jean-Daniel Fekete (Program Committee member)
- IEEE Pacific Visualization Symposium 2010: Jean-Daniel Fekete (Program Committee member)
- ACM UIST 2009: Jean-Daniel Fekete (Program Committee member)
- ACM UIST 2009: Pierre Dragicevic (Demo Chair)
- ACM CHI 2009: Pierre Dragicevic (Program Committee member)
- IEEE 3DUI 2009: Pierre Dragicevic (Program Committee member)
- CISIS 2009: Pierre Dragicevic (Program Committee member)
- IHM 2009: Pierre Dragicevic (Demo Chair, Program Committee member)

8.4. Workshop organization

- For the VisMaster project (see 6.3), we have organized a 3-days workshop in Paris gathering 16 of the world leaders in the domain of software infrastructure for visual analytics. We have started to design a set of interfaces for well-known visual analytics components as a Google-code project called "obvious" (see http://code.google.com/p/obvious).

8.5. Conference reviewing

- ACM CHI 2009: Jean-Daniel Fekete, Pierre Dragicevic
- ACM UIST 2009: Pierre Dragicevic, Jean-Daniel Fekete
- ACM SIGGRAPH 2009: Pierre Dragicevic, Jean-Daniel Fekete
- ACM EICS 2009: Pierre Dragicevic
- Conférence Francophone d'Interaction Homme-Machine (IHM) 2009: Pierre Dragicevic, Jean-Daniel Fekete
- IEEE Conference on Information Visualization 2009: Pierre Dragicevic
- IEEE Symposium on Visual Analytics Science and Technology 2009: Jean-Daniel Fekete
- Eurographics/IEEE Symposium on Visualization 2009: Pierre Dragicevic, Fanny Chevalier

8.6. Scientific associations

- AFIHM (French speaking HCI association): Jean-Daniel Fekete is President and Pierre Dragicevic is an active member.

8.7. Evaluation committees and invited expertise

- MDD program (ANR, National Research Agency): Jean-Daniel Fekete, member of the evaluation committee since 2005
- MdC750, associate professor position with INRIA chair at Bordeaux (2009): Pierre Dragicevic, member of the evaluation committee.

8.8. PhD defenses

- Rodrigo Andrade de Bortelho de Almeida (CNAM, Nov. 2009), Ph.D. Thesis "Contributions aux techniques pour enrichir l'espace moteur et l'espace visuel des dispositifs d'interaction bureautique": Jean-Daniel Fekete, reviewer.

- Gilles Bailly (Université de Grenoble, May 2009), Ph.D. Thesis "Techniques de menus : Caracterisation, Conception et Evaluation.": Pierre Dragicevic, jury member.

9. Bibliography

Major publications by the team in recent years

[1] A. Bezerianos, P. Dragicevic, R. Balakrishnan. *Mnemonic rendering: an image-based approach for exposing hidden changes in dynamic displays*, in "UIST '06: Proceedings of the 19th annual ACM symposium on User interface software and technology, New York, NY, USA", ACM, 2006, p. 159–168, http://doi.acm.org/10.1145/1166253.1166279.

[2] N. Elmqvist, P. Dragicevic, J.-D. Fekete. *Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation*, in "IEEE Transactions on Visualization and Computer Graphics", vol. 14, n^{o} 6, 2008, p. 1141-1148, http://doi.ieeecomputersociety.org/10.1109/TVCG.2008.153, (Best Paper Award)FR.

[3] J.-D. Fekete. *The InfoVis Toolkit*, in "Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis 04), Austin, TX", IEEE Press, October 2004, p. 167-174, http://www.lri.fr/~fekete/ps/ivtk-04.pdf.

[4] N. Henry, J.-D. Fekete, M. J. McGuffin. *NodeTrix: a Hybrid Visualization of Social Networks*, in "IEEE Transactions on Visualization and Computer Graphics", vol. 13, n^{o} 6, 2007, p. 1302-1309, http://doi.ieeecomputersociety.org/10.1109/TVCG.2007.70582.

[5] C. Plaisant, J.-D. Fekete, G. Grinstein. *Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository*, in "IEEE Transactions on Visualization and Computer Graphics", vol. 14, n^{o} 1, 2008, p. 120–134, http://doi.ieeecomputersociety.org/10.1109/TVCG.2007.70412.

Year Publications

Articles in International Peer-Reviewed Journal

[6] N. Elmqvist, J.-D. Fekete. *Hierarchical Aggregation for Information Visualization: Overview, Techniques and Design Guidelines*, in "IEEE Transactions on Visualization and Computer Graphics", 2009, to appear FR US .

[7] N. Elmqvist, N. Henry, Y. Riche, J.-D. Fekete. *Mélange: Space Folding for Visual Exploration*, in "IEEE Transactions on Visualization and Computer Graphics", 2009, to appear FR US .

[8] P. Isenberg, A. Bezerianos, N. Henry, S. Carpendale, J.-D. Fekete. *CoCoNutTrix: Collaborative Retrofitting for Information Visualization*, in "Computer Graphics and Applications: Special Issue on Collaborative Visualizationp", vol. 29, n^{o} 5, Sept-Oct 2009, p. 44–57 FR US CA .

Articles in National Peer-Reviewed Journal

[9] J.-D. Fekete, D. Ogilvie. *Le projet et la plate-forme "Millefeuille" : recherches et outils informatiques pour de nouveaux usages des almanachs"*, in "Bibliothèque de l'École des Chartes", vol. 166, 2009, p. 89-98 FR .

[10] N. Henry, J.-D. Fekete. *Représentations visuelles alternatives pour les réseaux sociaux*, in "RÉSEAUX: communication - technologie - société", vol. 26, n^{o} 152, 2009, p. 59–92, http://reseaux.revuesonline.com/article.jsp?articleId=12719FR.

International Peer-Reviewed Conference/Proceedings

[11] F. CHEVALIER, S. HUOT, J.-D. FEKETE. *WikipediaViz: Conveying Article Quality for Casual Wikipedia Readers*, in "Proceedings of the IEEE Pacific Visualization Symposium 2010", IEEE Press, March 2010, to appear.

[12] P. DRAGICEVIC, Y. SHI. *Visualizing and Manipulating Automatic Document Orientation Methods Using Vector Fields*, in "Proceedings of the International Conference on Interactive Tabletops and Surfaces (Tabletop '09), New York, NY, USA", ACM, 2009, to appear FR CN .

[13] J.-D. FEKETE, N. ELMQVIST, Y. GUIARD. *Motion-Pointing: Target Selection using Elliptical Motions*, in "CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, New York, NY, USA", ACM, April 2009, p. 289–298, http://doi.acm.org/10.1145/1518701.1518748FRUS.

[14] T. MOSCOVICH, F. CHEVALIER, N. HENRY, E. PIETRIGA, J.-D. FEKETE. *Topology-Aware Navigation in Large Networks*, in "CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, New York, NY, USA", ACM, April 2009, p. 2319–2328, http://doi.acm.org/10.1145/1518701.1519056FRUS.

[15] T. MOSCOVICH. *Contact Area Interaction with Sliding Widgets*, in "Proc. ACM symposium on User Interface Software and Technology (UIST 2009)", ACM Press, October 2009, to appear FR .

National Peer-Reviewed Conference/Proceedings

[16] C. APPERT, S. HUOT, P. DRAGICEVIC, M. BEAUDOUIN-LAFON. *Flowstates: Prototypage d'applications interactives avec des flots de données et des machines à états*, in "Proceedings of the 21th French speaking conference on Human-Computer Interaction (IHM '09), New York, NY, USA", ACM, 2009, to appear FR .

[17] F. CHEVALIER, S. HUOT, J.-D. FEKETE. *Visualisation de mesures agrégées pour l'estimation de la qualité des articles Wikipedia*, in "Extraction et gestion des connaissances (EGC'2010), Actes, Hammamet, Tunisie, 26 au 29 janvier 2010", Revue des Nouvelles Technologies de l'Information, Cépaduès-Éditions, 2010.

Workshops without Proceedings

[18] I. MANOLESCU, W. KHEMIRI, V. BENZAKEN, J.-D. FEKETE. *Reactive workflows for visual analytics*, in "Journées Bases de Données Avancées, Naumur Belgique", B. AMANN (editor), 10 2009, http://hal.inria.fr/inria-00425666/en/, H.5.2 H.: Information Systems/H.2: DATABASE MANAGEMENT/H.2.8: Database Applications/H.2.8.2: Scientific databasesFR.

Scientific Books (or Scientific Book chapters)

[19] J.-D. FEKETE. *Treemap*, in "Encyclopedia of Database Systems", Springer, July 2009, http://www.springer.com/computer/database+management+&+information+retrieval/book/978-0-387-35544-3 FR .

References in notes

[20] S. K. CARD, J. D. MACKINLAY, B. SHNEIDERMAN (editors). *Readings in information visualization: using vision to think*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[21] J. BERTIN. *Sémiologie graphique : Les diagrammes - Les réseaux - Les cartes*, Les réimpressions, Editions de l'Ecole des Hautes Etudes en Sciences, Paris, France, 1967.

[22] N. ELMQVIST, T.-N. DO, H. GOODELL, N. HENRY, J.-D. FEKETE. *ZAME: Interactive Large-Scale Graph Visualization*, in "Proceedings of the IEEE Pacific Visualization Symposium 2008", IEEE Press, March 2008, p. 215–222, http://www.aviz.fr/~fekete/ps/zame.pdf FR .

[23] J.-D. FEKETE, C. PLAISANT. *Interactive Information Visualization of a Million Items*, in "Proc. IEEE Symposium on Information Visualization 2002 (InfoVis 2002), Boston, USA", IEEE Press, October 2002, p. 117-124.

[24] J.-D. FEKETE, C. PLAISANT. *Les leçons tirées des deux compétitions de visualisation d'information*, in "Proceedings of IHM 2004, Namur, Belgium", International Conference Proceedings Series, ACM Press, September 2004, p. 7-12.

[25] J. J. GIBSON. *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates, New Jersey, USA, 1979.

[26] Y. GUIARD, Y. DU, J.-D. FEKETE, M. BEAUDOUIN-LAFON, C. APPERT, O. CHAPUIS. *Shakespeare's Complete Works as a Benchmark for Evaluating Multiscale Document-Navigation Techniques*, in "Proceedings of BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV'06), Venice, Italy", ACM Press, May 2006, p. 65-70.

[27] N. HENRY, J.-D. FEKETE. *Evaluating Visual Table Data Understanding*, in "Proceedings of BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV'06), Venice, Italy", ACM Press, May 2006, 6 pages, to be publised.

[28] B. LEE, C. PLAISANT, C. SIMS PARR, J.-D. FEKETE, N. HENRY. *Task taxonomy for graph visualization*, in "BELIV '06: Proceedings of the 2006 AVI workshop on BEyond time and errors, New York, NY, USA", ACM, 2006, p. 1–5, http://doi.acm.org/10.1145/1168149.1168168.

[29] P. LYMAN, H. R. VARIAN. *How Much Information*, 2003, http://www.sims.berkeley.edu/how-much-info-2003.

[30] T. MUNZNER, C. JOHNSON, R. MOORHEAD, H. PFISTER, P. RHEINGANS, T. S. YOO. *NIH-NSF Visualization Research Challenges Report Summary*, in "IEEE Computer Graphics and Applications", vol. 26, n^{o} 2, 2006, p. 20-24, http://doi.ieeecomputersociety.org/10.1109/MCG.2006.44.

[31] A. TRIESMAN. *Preattentive Processing in Vision*, in "Computer Vision, Graphics, and Image Processing", vol. 31, n^{o} 2, August 1985, p. 156-177.

[32] E. TUFTE. *The Visual Display of Quantitative Information*, Graphics Press, 1983.

[33] J. W. TUKEY. *Exploratory Data Analysis*, Addison-Wesley, 1977.