Asynchronous Collaborative Data Visualization*

A Review on the State-of-the-Art

Daniel Moser
Institute of Computer Science and Applied Mathematics
Neubrückstrasse 10
3011 Bern, Switzerland
daniel.moser@students.unibe.ch

ABSTRACT

This paper will give an review on the state-of-the-art in asynchronous collaborative data visualization techniques. The author will review multiple papers spanning the last seven years showing the progress that has been made in collaborative data visualization. This time span reflects the rise of research with the focus on asynchronous collaboration which was only little researched before.

The author will look at the different approaches in terms of human-computer interaction, annotation types and shortcomings. Where possible the article will look into the technical details and compare ease of installation and extensibility.

Categories and Subject Descriptors

H.5.3 [**INFORMATION INTERFACES AND PRESENTATION**]: Group and Organization Interfaces

General Terms

Review, Classification

Keywords

Asynchronous, Visualization, CSCW

1. INTRODUCTION

The evolution of the Web 2.0 has influenced the way people collaborate and communicate. While collaboration happened mostly in a synchronous co-located manner over local area networks in the earlier days, the key paradigms shifted towards globally connected asynchronous working collaborators. Having them in multiple time zones and loosely organized leads to many different challenges. The challenges connected to this shift were first broadly studied around 2007 [5] and the collaboration systems have incrementally been updated since then(e.g. [2], [6]). Another very important reason for the need of asynchronous collaborative data visualization is the explosion of data production and storage capabilities in recent years. The cBrain platform [3] takes that into account with providing distributed data storage and high-performance computing plugins.

The subject of asynchronous collaborative data visualization has only come up in the recent years. Viégas et al. [5] and Heer et al. [2] were among the first to take on the challenges of Web 2.0 technologies to research possible enablers of asynchronous collaboration in the data visualization field.

In this paper, the author will first give an insight into the referenced papers and the therein described visualization systems (see Section 2). In Section 3, the author will then classify the afore described systems according to different metrics the author finds important specifically for the asynchronous element of visualization collaboration. The criteria the papers were chosen by were, in part, the amount of citations and the distribution in time in which the systems were developed. This way, one might observe an evolution of the paradigms by which these asynchronous collaboration platforms were built upon.

2. SYSTEMS

In this section the author will compare the different systems and the evolution of these systems according to many different aspects.

A comparison of central properties of the presented systems can be found in Table 1 for an easy overview.

In asynchronous data visualization, the names Heer, Viégas and Wattenberg show up, all together or as subsets, in many papers and conferences in various constellations. They began publishing their investigations on the asynchronous part of collaboration in 2007 [5]. Since then, they published many more articles and developed multiple systems to evaluate their hypotheses to improve asynchronous collaboration (e.g. [2], [6]).

One of the first published works describes Many Eyes [5]. This system was developed by researchers at IBM. They launched this service as a public website to investigate the outcome of scaling the size of the audience. Other researchers

*This paper is a product of the Seminar Collaborative Data Visualization held during the autumn semester of 2014 for the Master of Computer Science program of the Universities of Bern, Neuchâtel and Fribourg supervised by Dr. D. Lalanne.

Table 1: Feature Comparison of the Different Platforms

System	Graphical Annotations	Doubly Linked Discussion	Tags / Links	User Data	Public Web	Installable
Many Eyes	no	no	no	yes	yes IBM intranet	no
sense.us	yes	yes	no	no	partly	no
Giovanni	no	no	no	no	yes	no
CoScribe	yes	no	yes	lecturers slides	no	client software
CommentSpace	no	yes	yes	not directly	partly	Plugin (Flash and Firefox)
CBRAIN	no	no	no	yes	no	yes

(e.g. [2], [6]) use closed systems with a mostly fairly small user base. A central point of this system is the fact that it allows user-supplied data for visualization. The import format is a tabulator delimited table, like they come from Excel or Open Office copy-pasting. Each data set and visualization has its own discussion forum attached for users to share a variety of information like textual comments or links to other websites. Every visualization and data set in Many Eyes resides at a specific and unique URL which means users can exchange links to multiple different visualizations of the same data set in the forum or in a blog outside of the Many Eyes ecosystem. The system allows for many different visualization styles like tag clouds and scatter plots. But it does not allow users to draw onto the visualizations like other systems [2].

The results for the sense.us platform have last been published in 2009 [2]. This system is web based and comes preloaded with demographic data of the United States from the last 150 years. It does not only come with one new property which the researchers are investigating, but a variety of multiple possible enablers of asynchronous collaborative data visualization. Each view is composed of a visualization part that shows a certain graph and on the right-hand side there is a discussion area showing comments. Sense.us can be accessed through any browser supporting Java applets. Like [5] each individual visualization (including filters, zoom-levels etc) has its own individual and unique URL. This property enables bookmarking views to return to a certain visualization. A novelty addition to this kind of system are the so-called bookmark trails where multiple bookmarks are lined up in a list to show a progression of visualization manipulations. Unlike other platforms, sense.us allows the users to make graphical annotations on top of the visualizations. Those graphical annotations include lines, arrows, text labels, geometrical shapes and free-form drawings. A key element of this system and a novelty in asynchronous collaborative data visualization are doubly linked discussions. This element does not only link from comments to a visualization state, but also any applied filter links the visualization to all comments made at this visualization state.

In 2011 a last paper has been released by Willet, Heer et al [6]. The newest system they developed, can be embedded to various websites as a flash plugin or it can be installed as an add-on for Firefox. This system introduces a new concept to asynchronous collaborative data visualization called tags and links. This concept was designed to bring a structure to otherwise structure-less discussions. To not get the list of different tags and links not out of hands, the developers of CommentSpace introduce a fixed set of tags (question, hypothesis, to-do) and links (evidence-for, evidence-against) to be used by the users. This allows late-comers to catch up with the progress of discussion groups and scanning through the questions and hypotheses posed by other users. CommentSpace also allows the use of doubly linked discussions like they are used in the sense.us platform. The researchers have found that like [2], it can be beneficial to give users the possibility to share links to comments or visualization outside the community.

Giovanni [1] originates at NASA. Its basic purpose is the visualization of environmental and atmospheric data. The referenced paper describes the third iteration of the NASA's Goddard Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni) system at the Goddard Earth Sciences Data and Information Services Center. This tool lets users explore and visualize Earth science data from sensors spread around the world and user supplied models. A central service apart from the visualization is that, after the initial parameter description on the data set, every processing step like filtering and zooming is recorded in a so-called data lineage. This allows for better understanding how a user got to a certain visualization and retrace conclusions made by other data analysts. Further the system enables asynchronous collaboration through RSS feeds for status updates and alerts on data sets.

For easier annotation sharing with lecture slides, a team at the Technische Universität Darmstadt developed CoScribe [4]. During initial field studies the researchers found that students using paper as annotation medium were producing notably more annotations than their colleagues using an electronic medium (e.g. laptops, tablet computers). They also found that despite a majority of students owning a laptop, those using electronic mediums for annotations were in a clear minority group. When sitting in class, students use the lecturer's slides printed out through the CoScribe system and a special digitizer pen. This system adds a special dot pattern that is only minimally visible to each slide in addition to a special toolbar in the center of each page. The dot pattern is used by the pen's camera to detect which slide any given annotation is added to. The toolbar adds buttons known from digital interfaces to add visibility properties to annotations. Additionally, any annotation can be tagged with this toolbar to four different categories similar to [6] (Important, Question, To Do and Correction). The annotations of all different students will then be aggregated in a database which can be accessed through a client program. Using this program, users can make the annotations of others permanently visible and then print them out.

cBrain [3] is another exotic system. It was developed as a joint-venture by neuroscientists at the McGill University of Montreal and the University of Lyon. Their goal was to develop a platform for neuroimaging visualizations to be able to collaboratively detect indications of brain diseases like Alzheimer's. This paper is a rather technical one, mostly describing the development, the technologies involved and how to extend the system. The researchers who developed it wanted to provide a system that can be installed and used as-is, without much technical know-how required. cBrain offers middleware for researchers to upload data onto high-performance computing clusters and submit jobs to process. It also allows distribution of the data and fine grained rights management. The author chose this paper not directly because of visualization paradigms or special HCI features but rather because it might serve as a starting point for new projects. As cBrain was written using open-source frameworks like Ruby on Rails and current web technology like HTML 5, it can be extended by any computer scientist to include new interaction paradigms or visualization types.

3. CLASSIFICATION

In this section, the author tries to extract the central challenges for asynchronous data visualization and find a classification schema to categorize the referenced papers and extract their strengths and weaknesses. The herein shown classifications were chosen by common features and goals of the referenced papers. Some of them can be found in Table 1 and others were brought up in the discussions of the results of multiple papers (e.g. user motivation).

3.1 Common Ground Finding

Finding a common ground in asynchronous collaborative visualization can be a big challenge as the individual users do not communicate directly. The users mostly use a low feedback medium like textual comments and with the world wide web, they often do not share a common cultural background. Thus it is very important to provide a system to bring all participants of a discussion to a common ground from which they can draw conclusions and ask new questions.

Systems like sense.us [2] and CommentSpace [6] use the aforementioned concept of doubly linked discussion to try to enable common ground finding. It allows users to not only link from text to visualizations but also the other way around. This way, a user filtering the data to pursue a question might find that other users had the same questions and can directly start collaborating with them. In a system without the doubly linked discussion concept, he might never have found out that other users already started investigating on the same questions.

Another possibility of finding a common ground is through a progression of visualizations. E.g. showing multiple corresponding filtering steps to show how one arrived to a particular visualization state. Here we have two system supporting this kind of behaviour, one is Giovanni [1] and the other is sense.us [2]. Giovanni records every step after providing the initial parameters, which also allows users to retrace their own steps.

3.2 Group Sense Making

When a common ground for a discussion was elaborated, it is important to be able to show your ideas and thought process behind your arguments to the other collaborators. Otherwise the users will start talking at crossed purposes and any progress on the discussion will stall.

To enable good progress in the discussions and in letting other collaborators understand one's though process, CommentSpace [6] and CoScribe [4] implemented a system using labels and tags. This allows for a more structured discussion thread where it is easily possible to find the problem statement (tag: question), hypothesis (tag: hypothesis) and any arguments supporting or contradicting a hypothesis (links: evidence-for, evidence-against).

Also, in the special case of CoScribe [4], a user might tag important notes that a lecturer has emphasized to later enable himself and other students to find this part again when studying for the exam.

3.3 Graphical Annotations

Only having textual comments as the primary communication medium gives very little feedback on how the other users might understand one's comments and which conclusions they might draw from it. Being able to make graphical annotations to the visualization through various geometric shapes and free-form drawing gives other users additional information to grasp the ideas behind a user's reasoning. Even with only the possibility of drawing arrows, a user can more easily point to certain peaks or areas of a visualization without having to extensively paraphrase which peak at which value was meant for the current comment.

Only one of the papers mentioned in Section 2 allows users to use graphical annotations to support their textual comments. Sense.us [2] gives a user many possibilities to graphically annotate a certain visualization. Heer et al. found that one third of all comments included graphical annotations of which the vast majority of nearly 90% were used to pointing. A small amount of graphical annotations were used for more playful expression of thoughts. One user, for example, annotated the decrease of stock brokers during the great depression with a matchstick man jumping down a building and commenting "Great depression 'killed' a lot of brokers".

A challenge connected to graphically annotate data, is that the annotations are always linked to visualizations or their state but never really to the data itself. When using another visualization style, the annotations can not easily be transformed or assigned to the new layout.

3.4 User Motivation

An important factor for the success of a visualization platform is the user's motivation to use it. When deployed in a working environment, the motivation factor might not carry that much weight, but for a public platform a user needs to have the motivation of coming back after the initial exploration phase. This problem can be seen in the data of Many Eyes [5] where they had over 100'000 user sessions and nearly 1500 users that registered for the service. But only around a third of those registered users have even created visualizations and only one in every 10 registered users have left comments.

The here discussed papers do not really look into this, but it might be useful to dedicate research onto this field. Having a motivating platform to use will give more data points to analyze and might bring the data pool to a size where statistically relevant conclusions might be drawn from the data. Overcoming the problem of user motivation could be achieved by usability researchers analyzing existing platforms and issue guidelines for future visualization platforms. Another possibility would involve the principle of gamification to create more playful handling of the interaction with the system. But this approach is very dependent on the intended audience of the system. A more scientific audience might prefer more clean interfaces while a public web platform could very much profit from a more playful approach.

3.5 Outside Sharing

When working in a semi-closed group of collaborators it might provide additional insight to bring the attention of a person outside the group to the attention of the discussion. For certain problems, this person might not even be a user of the platform where the discussion is taking place.

To address the need of outside users commenting on visualizations, several systems added social sharing mechanisms. Viegas et al. [5] provide RSS feeds and a "blog this" button to easily extract data from Many Eyes and make it possible to embed visualizations into blogs and to notify outside users of changes in discussions. Users of CommentSpace [6] can log in with their Facebook account and share individual comments or visualizations with their friends or generate unique URLs to share views through other means of communication. While going through the papers it became apparent that researchers have already found applicable solutions to this challenge.

4. CONCLUSIONS

In this paper the author evaluated the state-of-the-art in asynchronous data visualization spanning the last 7 years. The article shows the evolution of web platform's used for asynchronous collaborative data visualization. Different systems were introduced and described to later enable the classification of important categories that enable asynchronous collaboration. It was shown that for some problems (e.g. outside sharing) there are already valid solutions working with today's social media. Other factors, like user motivation have not been studied to a great extent and might need further research.

As an addition to the classifications made in Section 3, the author proposes a system to find similar graphical annotations. Such a system could be seen as an extension of the doubly linked discussion model where one can find comments for visualizations where the same data filter was applied. Extending doubly linked discussion in such a way would further enable finding other users that are discussing on similar topics. This way, if two filters do not match enough to show another user's comments and annotations (e.g. different zoom level), the user could still be notified that another user pointed to a same area in the visualization. As already mentioned before, linking the annotations also to the data and not to the visualization itself could extend the possibility of finding and joining discussions around the selected data. Analyzing very discrete geometrical shapes like arrows and rectangles, a system can detect what area is highlighted and show information related to the other annotations.

5. REFERENCES

[1] S. Berrick, G. Leptoukh, J. Farley, H. Rui, and H. Rui. Giovanni: A web service workflow-based data visualization and analysis system. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(1):106–113, Jan 2009.

[2] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM*, 52(1):87–97, Jan. 2009.

[3] T. Sherif, P. Rioux, M.-E. Rousseau, N. Kassis, N. Beck, R. Adalat, S. Das, T. Glatard, and A. C. Evans. Cbrain: a web-based, distributed computing platform for collaborative neuroimaging research. *Frontiers in Neuroinformatics*, 8, 2014.

[4] J. Steimle, O. Brdiczka, and M. Muhlhauser. Coscribe: Integrating paper and digital documents for collaborative knowledge work. *Learning Technologies, IEEE Transactions on*, 2(3):174–188, July 2009.

[5] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1121–1128, 2007.

[6] W. Willett, J. Heer, J. Hellerstein, and M. Agrawala. Commentspace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 3131–3140. ACM, 2011.