

Using the analysis tool Tatiana to visualize and calculate interaction indicators

Gregory Dyke, Ecole Nationale Supérieure des Mines de Saint-Etienne, 42023, St. Etienne cedex 2, France, Dyke@emse.fr

Kristine Lund, ICAR, CNRS, University of Lyon, ENS-LSH, 15 Parvis René Descartes, BP 7000, 69342 Lyon cedex, France, Kristine.Lund@univ-lyon2.fr

Jean-Jacques Girardot, Ecole Nationale Supérieure des Mines de Saint-Etienne, 42023, St. Etienne cedex 2, France, Girardot@emse.fr

Abstract: Tatiana is a tool for assisting the human analysis of interaction data. It was designed to be generic in terms of visualization, source data and automated transformations. In this paper, we argue that the generic nature of Tatiana makes it relevant to problems associated with interaction analysis and visualization both as an engineering case study and as a platform in which to implement and share indicators and visualizations.

Introduction

After extensive experience in analyzing CSCL data on the part of two of the authors, a PhD thesis was started by the main author in 2006 in order to model CSCL researcher's analysis process and allow capitalization on analysis results (in other words make it possible to use these results as a basis for further analysis work). One of the results of this thesis, which is drawing to its conclusion, is Tatiana, a Trace Analysis Tool for Interaction Analysts. This tool is mainly aimed at assisting researchers in analyzing corpora consisting of a variety of synchronized media sources (interaction log files, video, transcribed dialogue etc.) (Dyke, Girardot, Lund, & Corbel, 2007). However its functionalities are also well suited to solving a number of problems related to the calculation and visualization of interaction analysis indicators in regards to asynchronous communication (Lund *et al.* 2008).

In this paper we will first present Tatiana's main functionalities and the kinds of corpora which are currently being analyzed with Tatiana. We will argue that Tatiana's generic design can be enlightening to the themes of this workshop on several counts: rapid implementation of many kinds of indicators for researchers, application of a given indicator to data from different tools, implementation of alternative visualization interfaces, prototyping of indicators and their visualization before implementing them in a real-time situations and finally evaluation of the use of IA tools by students, teachers and moderators.

Tatiana: a generic analysis environment

Tatiana (Trace Analysis Tool for Interaction ANALysts) (Dyke, Lund, & Girardot, 2009b) is an environment designed for manipulating various kinds of analysis artifacts, in particular those which present a view on event-based data. We call these views *replayables*, because they can be replayed in a similar fashion to a video. They are one of the major kinds of artifacts which researchers construct to analyze computer-mediated interaction (Dyke, Lund, & Girardot, 2009a)

Tatiana is built on a number of core concepts and components (cf. Figure 1). Tatiana replayables can be created either automatically (through import) or by hand. Once created, all replayables in Tatiana benefit from Tatiana's four core functionalities: transformation, analysis, visualization and synchronization.

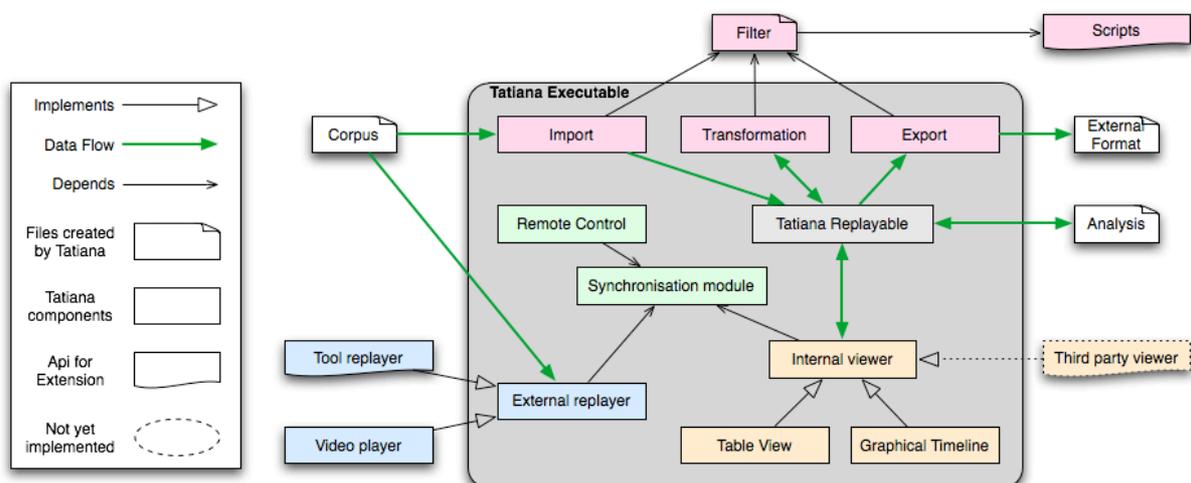


Figure 1. Tatiana architecture showing 1) dependencies between components, 2) components designed with extensibility in mind and 3) future developments.

Transformations

Replayables can be transformed (again, automatically or manually) and exported. Automated import, transformation and export works through the application of what we call *filters*. These are objects which combine *scripts* into a workflow. Scripts are small programs written in XQuery to perform a specific operation, such as transform a file in the corpus into data Tatiana can understand, exclude certain kinds of events from a replayable, find certain kinds of events in a replayable, combine multiple replayables, etc. As we do not expect researchers with no programming knowledge to write these scripts, we are currently developing a graphical editor for filters which will allow researchers to customize the execution of these scripts. Such a filter might combine a new script for data import from the interaction log data produced by a new kind of tool with an existing script which only shows the actions of a particular subset of students. Manual transformations include the ability to delete, reorder, re-group and split events.

Analysis

All replayables within Tatiana can be augmented by analysis data generated by the researcher. There are currently three kinds of analyses supported by Tatiana: free-form annotation, categorization, and graphs. Categorization is simply a way of adding annotations from a restricted list of words and can be used for coding, labeling and adding keywords. The list of categories available can be edited at any time thus allowing for an evolving analysis scheme. Graphs allow researchers to explicitly mark relationships between events (cf. Figure 2).

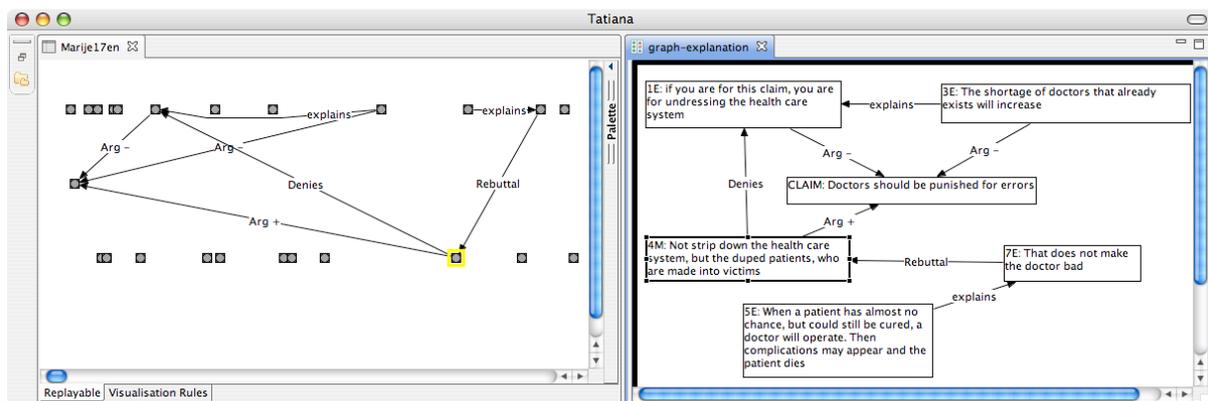


Figure 2. Synchronized view of a chat and the graph structure of this chat as analyzed by a researcher. The visualization to the left is completely automatic, taking the graph structure on the right and placing the nodes on the X-axis by date and on the Y-axis by contributor. The visualization on the right is a graph created by hand from a chat which was unstructured. Had structure been present (e.g. in a forum), this graph could also have been created automatically.

Visualization

All replayables within Tatiana can be visualized in different viewers. There currently exist two kinds of viewers: a table view in which data is presented as it might be presented in Excel, with one row per event and columns for each of the event's properties and a graphical timeline. The graphical timeline is a first attempt at assisting the automated creation of visualizations. It presents each event as a graphical object whose graphical properties (color, shape, size, position, etc.) can be set according to the properties of the event (user, tool, timestamp, analysis category, etc.). Tatiana is extensible, allowing new kinds of views to be created, affording new ways of visualizing data.

Synchronization

Finally, all replayables in Tatiana can be synchronized with each other (cf. Figure 3) and also with data viewed in external replayers such as media players and tool replayers (a special mode of certain CSCL tools which are capable of reading the interaction log data they generated and reproducing on screen what the user saw). Tatiana provides a mechanism to pilot external tool replayers. Synchronized replay means that when a timestamp is selected in the "remote control", the video player (and other external replayers) are instantly navigated to that timestamp, and the events matching that timestamp in the currently visualized replayables are highlighted. Furthermore, selecting an event in a visualized replayable will again navigate all the other views to that moment in time. For example, during analysis of a discussion of genetically modified organisms using Tatiana, if a

researcher clicks on the timestamped event in the table view “argument by Alice: they go against ethics” this action causes the replayer to show the state of the diagram immediately after that argument was constructed. Information on the dynamics of the interaction is thus provided, which is oftentimes difficult to discern in static log traces. Zooming in on particular episodes becomes possible. In general, such linking between replays is very useful for limiting the amount of information displayed in a single visualization, with the knowledge that further information is available in other visualizations on demand.

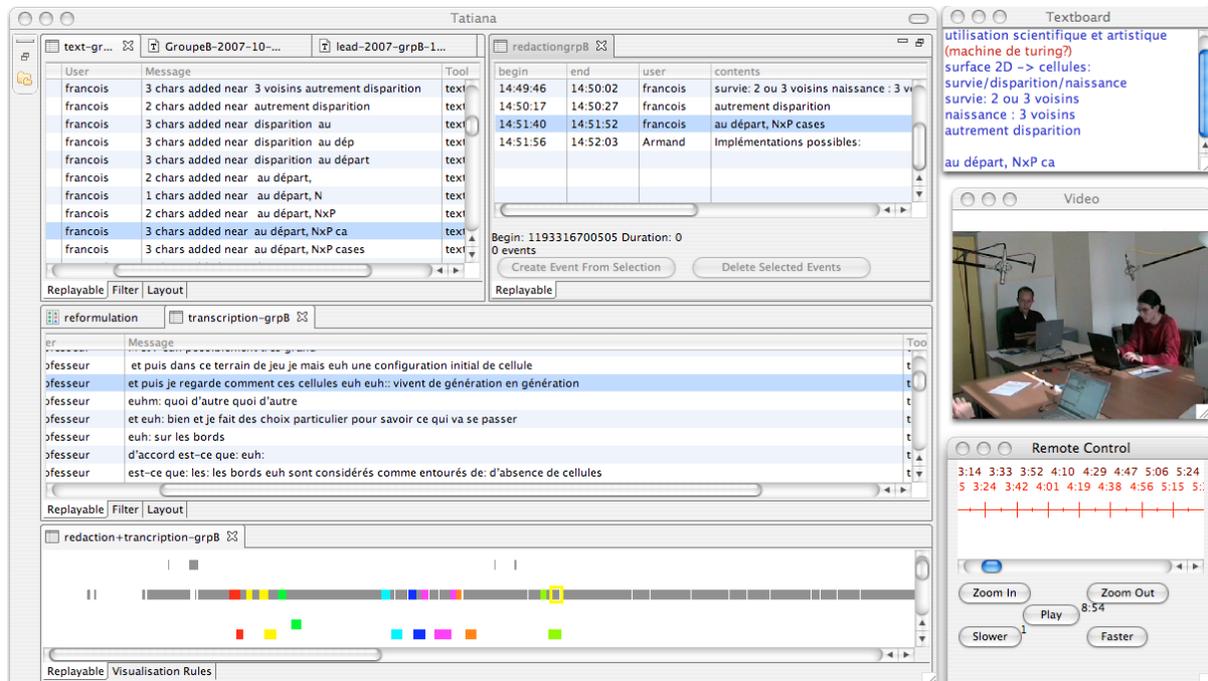


Figure 3. Various replays in Tatiana: traces of a shared text editor (top left), transcription (middle left), writing units (top center), visualization of reformulation (bottom left), synchronized with external tools, DREW replayer (top right), video player (middle right), remote control (bottom right).

Current applications of Tatiana

Tatiana is designed to be used with any kind of data (provided it can be presented in an XML format). Tatiana is currently being used to analyze data produced by a variety of tools such as DREW (Corbel, Girardot, & Jaillon, 2002; Corbel et al., 2003), CoFFEE (De Chiara, Di Matteo, Manno, & Scarano, 2007), Knowledge Forum (Scardamalia, 2004), blogs (via Atom feeds), Excel and Elan (<http://www.lat-mpi.eu/tools/elan/>). Some of this data is synchronous, other parts are asynchronous and some were produced through a series of synchronous sessions over several months.

The analysis methods which are used cover a variety of socio-cognitive interaction analysis methodologies from ethnographical, through case studies and sequential data exploration to coding and counting and statistical analyses.

Enabling new kinds of data to be analyzed within Tatiana is as simple as writing an import filter (such filters typically take thirty minutes to write for someone with extensive knowledge of the data format and with minimal knowledge of Tatiana and XQuery). Tatiana’s pivot format (a sequence of items with each item having a set of properties we call *facets*) imposes no restrictions on the data that is represented in it beyond each item having two compulsory facets: a *timestamp* in order to enable synchronization and an *anchor* providing a unique id to which analysis information can be attached. Generic transformations exist on this pivot format such as filtering, search, contingency tables, etc. New transformations can be developed which are either generic or specific to replays containing items which have a specific set of facets.

Calculation and visualization of interaction indicators with Tatiana

Tatiana’s ability to handle many different kinds of data, its mechanisms for rapidly integrating new kinds of transformations and the extension possibilities of visualizations make it both a useful tool and an engineering case study for the implementation of interaction indicators.

When these indicators are intended for the researcher, the transformation mechanism allows any kind of calculation to be made and the result immediately visualized in tabular form (cf. Figure 4). Provided the existence of filters which import different kinds of data into the same abstract data format (meaning the items have the same set of facets), a transformation which calculates a given indicator can easily be applied to different kinds of data. As presented above, two alternative visualization interfaces already exist in Tatiana. New ones can be easily implemented along the same model.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	total	
Teacher	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	2	7
SarahH	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	7	13
Sarah H	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	4
Ryan S	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	6
Michael C	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	3	3	9
Emma F	0	1	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	1	0	1	8	15
Amy G	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
katie R	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	4	6
Lauren D	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	2
leanne r	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	4
Sharnie C	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Sarah J	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	4
Lucas B	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	4	8
chantelle g	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2
Anna H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2
SERVER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Total	1	5	1	1	7	1	1	1	2	1	5	1	1	1	1	1	5	1	11	37	87	

Figure 4. Contingency table for a threaded chat. The horizontal axis shows the the thread ID, the vertical axis shows the user name. Figures are the number of interventions per user and per thread. Data is from CoFFEE – as yet unpublished study which is a follow up on Gelmini Hornsby, Ainsworth, Buda, Crook, & O’Malley (2008).

One of the important uses of interaction indicators is to provide mirroring, regulating and guiding mechanisms to learners and teachers. Combining a prototypical implementation of an indicator and its visualization in Tatiana with Tatiana’s synchronized replay functionality allows researchers to test such visualizations on previously collected data and replaying it in order to examine what would be shown to users in a variety of real situations.

Finally, as an analysis tool, Tatiana can help evaluate the use of IA tools in real-life settings by presenting analysts with a synchronized view of user’s use of communication tools and their use of the provided IA tools.

Conclusion

While Tatiana’s functionality is mainly aimed at assisting the human analysis of interaction log data, its generic design makes it a tool which is also ideally suited to the rapid implementation and prototyping of interaction indicators. While it is certainly possible to calculate and visualize indicators without the use of tools such as Tatiana, they provide an environment which makes it easier to implement and evaluate interaction indicators since the overhead of importing data, normalizing it and then exporting the result in a readable format is much reduced. Tatiana can also be used to share data sets and interaction analysis methods at a low technological cost making CSCL researchers less dependant on close collaboration with computer engineers.

References

- Corbel, A., Girardot, J.J., & Jaillon, P. (2002). DREW: A Dialogical Reasoning Web Tool, *ICTE2002, International Conference on ICT's in Education*. Badajoz, Spain.
- Corbel, A., Jaillon, P., Serpaggi, X., Baker, M., Quignard, M., Lund, K., & Séjourné, A., (2003). DREW : Un outil Internet pour créer des situations d’apprentissage coopérants [DREW : An internet tool for creating cooperative learning situations]. In Desmoulin, Marquet & Bouhineau (Eds.) *EIAH2003 Environnements Informatiques pour l’Apprentissage Humain, Actes de la conférence EIAH 2003*, Strasbourg, 15-17 avril 2003, Paris : INRP, p. 109-113.

- De Chiara, R., Di Matteo, A., Manno, I., Scarano, V. (2007). CoFFEE: Cooperative Face2Face Educational Environment. *Proceedings of the 3rd International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2007)*, New York, USA.
- Dyke, G., Girardot, J.-J., Lund, K., Corbel, A (2007). Analysing Face to Face Computer-mediated Interactions, *EARLI '07*, Budapest, Hungary.
- Dyke, G., Lund, K., Girardot, J.-J. (2009a). Tatiana: an environment to support the CSCL analysis process. *Proceedings of CSCL 2009*, Rhodes, Greece.
- Dyke, G., Lund, K., Girardot, J.-J. (2009b). Un outil flexible pour l'analyse de traces - Gestion, synchronisation, visualisation, analyse et partage de corpus d'interactions médiatisées par ordinateur avec Tatiana. *Atelier ICT 2009*, Caen, France.
- Gelmini Hornsby, G., Ainsworth, S., Buda, M., Crook, C., & O'Malley, C. (2008). Making your views known: The importance of anonymity before and after classroom debates. *ICLS 2008*, Utrecht, The Netherlands.
- Lund, K., van Diggelen, W., Dyke, G., Overdijk, M., Girardot, J.J., & Corbel, A (2008). A researcher perspective on the analysis and presentation of interaction log files from CSCL situations within the LEAD project. *ICLS 2008*, Utrecht, The Netherlands.
- Scardamalia, M (2004). CSILE/Knowledge Forum. *In education and Technology: An encyclopedia* (pp. 183-192). Santa Barbara: ABC-CLIO