COLLEGE OF INFORMATION SCIENCES AND TECHNOLOGY
THE PENNSYLVANIA STATE UNIVERSITY

A Comparative study of Exploratory Sequential Data Analysis tools

Sowmyalatha Srinivasmurthy

sowmya@cse.psu.edu

4 May 2010

College of IST, University Park, PA 16802

Class Project for IST 521

A Comparative study of Sequential Data Analysis tools

Sowmyalatha Srinivasmurthy

sowmya@cse.psu.edu

Department of Computer Science and Engineering

The Pennsylvania State University

University Park, PA 16802

4 May 2010

Abstract

With Exploratory Sequential Data Analysis (ESDA) growing into a major area of research in the field of cognitive sciences, there have been numerous methods and tools that have been developed as a result of the research in the recent past. Some of the tools have also been designed as commercial products requiring appropriate licensing and involving costs. An attempt to give an overview and bring out the positive and negative aspects of some current tools is the present study.

The following tools were used for the study:

1. MS Excel from Microsoft Corporation

2. ABSTRACT (ANALYSIS OF BEHAVIOR AND SITUATION FOR MENTAL REPRESENTATION ASSESSMENT AND COGNITIVE ACTIVITY MODELLING) developed by LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information) Lab

3. InfoScope developed by Macro Focus

4. Theme developed by Noldus

5. INTERACT developed by Mangold International

The comparison of these tools was conducted based on the following facets: System Requirements, Supported Data Formats, Integrated Record function, and Integrated Search function, Level of Visualization (including statistical evaluation and interactive data interpretation) and Documentation and Support. We found that none of the tools studied here was comprehensive enough to handle all kinds of data. Each tool was found to have its own positive features. While Theme was found to be really powerful tool supporting a variety of input file formats and ability to present statistical inferences from the data, it lacks the record function. The tool ABSTRACT was found to be the only tool that was capable of hierarchical event display, but lacks statistical support and record function too. INTERACT was found to support features such

as recording and tagging of video data. It also has an integrated statistical package making the tool comprehensive. But the tool supports only video format input files thus restricting its use by the supported data format. Excel was found to be very primitive as an ESDA tool as it cannot support actual different formats of data like audio or video and well as fails to display the data on a time line easily. It also lacks the record function. But the tool has a very comprehensive search and statistical package. Thus we found that none of the tools was able to satisfy all the requirements of a comprehensive ESDA tool. In order to overcome this lack of a single comprehensive too, we recommend using a combination of several tools to exploit their complementarities to achieve all the required features. For example one could use ABSTRACT for its advanced timeline visualization features in parallel with Excel for its easy editing features, search and advanced statistical processing functions.

Table of Contents

1. Introduction: Exploratory Sequential Data Analysis

Researchers in the field of Human Computer Interaction (HCI) have always been faced with the challenge of handling the data they accumulate during their research. The data may be collected as video or audio recordings capturing the verbalizations, actions, as well as the general aspects of the working environment during an interaction of human subjects with the system under study. The enormous data recorded often needs to be analyzed to provide reasons for an observed behavior or to obtain inputs for a system design. Some of the methodologies used to explore and analyze this data are protocol analysis, video analysis, interaction analysis, conversation analysis, discourse analysis, sequential data and task analysis. While handling the large data, researchers also need to pay attention to the sequential nature of the interaction that has been recorded in the data. Attempting to explore and analyze the sequential nature of events, which unfold over time, has been a greater challenge with the lack of powerful tools for analyzing this nature of the extensive data (Hilbert & Redmiles, 2000).

This sub-field of research in HCI was been identified and worked on by many researchers in the past. The pioneers Sanderson and Fisher (1994) coined the term "ESDA" (Exploratory Sequential Data Analysis) to represent this approach. They define ESDA as "Any empirical undertaking seeking to analyze systems, environment and/or behavior data (usually recorded) in which the sequential integrity of events has been preserved." ESDA has been accepted as the "application of the philosophy of Exploratory Data Analysis (EDA) to sequential data" (Sanderson and Fisher, 1992). Tukey (1977) introduced EDA and is defined as a simple, visual, but still quantitative approach to data that allows an investigator to achieve a richer quantitative understanding by 'looking at data to see what it seems to say'. It has often been used as a type of observational data analysis. There have been various researchers who have adapted this methodology (Mavkay, 1989; Roschelle, Pea & Trigg, 1990; Sanderson, James & Seidler, 1989). EDA has been based on three principles (Hartwig & Dearing, 1979; Tukey, 1977): Continual Openness and Re-expression, Initial Skepticism and Exploration vs. Confirmation. ESDA has been extended from these principles and the principles have been laid out as 8 C's by Sanderson and Fisher (1992). The 8 C's defined are chunking, commenting, coding, connecting, comparing, constraining, converting and computing.

With ESDA growing into a major area of research in the field of cognitive sciences in recent times, there have been numerous methods and tools that have been developed as a result of the research. Some of the tools have also been designed as commercial products requiring appropriate licensing and involving costs. This study attempts to give an overview of current tools and bring out their positive and negative aspects. Due to time constraints and the ease of availability of the tools only five tools were studied.

2. Motivation to the study

There has been research in the past that dealt with the capture of user actions for HCI and usability studies. There are a variety of tools that are currently available to support the capture and coding of user interactions with software. Hilbert and Redmiles (2000) review a few of these methods. Although they present a comprehensive evaluation of tools, that can capture as well as code the data, they do not present any comparison on tools developed for ESDA. They compare tools that can capture events, search and compare sequences, but no comparison is presented for comprehensive tools that allow recording, visualize and code sequential data along with the ease to compare and manage the data.

Sanderson and Fisher (1994) gave researchers a direction for research with the 8 C's for ESDA approach. But they too do not present a comparison of tools that satisfy the principles of ESDA proposed in their paper. The paper also does not give direct indicators or hints for the researchers to choose any tool for this purpose.

The current work is intended to focus on the features and usability of currently available ESDA tools. This study is intended to provide guidelines, to first time users to choose the right tool, for the data collection and analysis schemes intended. But due to time and resources availability constraints, the study will be centered on the study of a few tools. The study was conducted with already available sample data sets that consist of observation of the activity either as output of sensors, video recording of behavioral patterns (like in driving scenario) or data patterns collected in a sociological context like the terrorism activities in a particular region.

3. Methodology

The study involved 5 tools that were available for analysis. When the study began the number of tool selected was twelve. But obtaining all of the tools was not possible as most of them are commercial tools and no demo versions were available even on request. With support from some tool creators, who volunteered to loan the tool for the study, the final set of tools used for the study were:

1. MS Excel from Microsoft Corporation (Office 2003 : 11.8320.8221, SP3)

2.ABSTRACT (ANALYSIS OF BEHAVIOR AND SITUATION FOR MENTAL REPRESENTATION ASSESSMENT AND COGNITIVE ACTIVITY MODELLING) developed by LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information) Lab (version: checked out from container Cp1252 versioned on November 25, 2009 10:12:39 PM)

3. InfoScope developed by Macro Focus (version: 3.3.4)

4. Theme developed by Noldus (version: 5.0)

5. INTERACT developed by Mangold International (version: 9.0.6)

All the tools were installed on a Dell Inspiron 1525 with an Intel Pentium Dual Core T2370, 1.75Ghz processor, with a 15 inch screen, a 32 bit Windows Vista Home operating system and 3GB RAM..

The comparisons of these tools were done based on the following facets:

System Requirements, Supported data formats, Integrated Record function, Integrated Search function, Level of Visualization (including statistical evaluation and interactive data interpretation) and Documentation and Support.

All the tools were installed and used by a single user. The installation of the tools was recorded in audio format and analyzed using Verbal Protocol analysis techniques. This verbal protocol of the installation of Infoscope and ABSTRACT are available in the Appendix .The recordings were analyzed for the issues faced by the user while installing the tools as well as the incompleteness of the installation guide to provide sufficient directions for this purpose. A task analysis on the

recording for the installation of ABSTRACT was used to come up with a more detailed installation guide for the tool. This guide is available in the Appendix .

4. Observations and Discussion

4.1 Availability and System Requirements

The tools used in the study had varying system requirements. This is an important aspect to be always noted by the users while they choose the tool for their use.

MS Excel was one of the tools that are available as a stand alone tool as well as a part of the MS-Office suite. This tool as well as the package is license controlled and can be purchased online or at a store at a price ranging from $395 to $499 depending on the tools included in the suite. The system requirements of this tool are minimal as this package is well supported by most of the current operating systems. The package is available for Windows Operating systems as well as Macintoshes. The two versions of Excel vary significantly in their macro-support capabilities. Excel for Windows has a built-in Visual-Basic (VBA) that allows creating and using macros which can prove to be an efficient way for processing data. In Excel for Macintoshes, VBA is not available and the macros system is much less efficient.

The tool, ABSTRACT (Georgeon, O, 2008) does not specify a stringent system requirement and was found to be the easiest to be fulfilled. Most of the requirements for this tool are other support tools (mostly freeware available online) that can be easily obtained and installed. The tool is an open sourced, free-licensed tool that is available to users and developers free of cost. This tool has the added advantage of supporting an online installation for collaborative data analysis. An online installation is currently hosted by The Pennsylvania State University for public usage thus avoiding any installations that may be required by other tools. The tool has the flexibility of being used on Windows as well as Macintosh systems.

The tool, Infoscope, used in this study is only a demo version available online as the creators of the tool were unable to provide comprehensive demo version for the study. The system requirements for the tool are not clearly specified by the tool makers but the tools also has the flexibility of being used on Windows, Macintosh or Linux systems. The tool is license controlled and can be purchased online at a price of $365/user.

The tool, Theme, developed by Noldus International also does not specify the system requirements clearly. But the tool can be used only on Windows based systems and is not supported for others operating systems.

The tool, INTERACT, is available only for Windows based systems but has low requirements for its successful running. They claim that INTERACT can run on any up-to-date Windows multimedia PC. The tool is license controlled and can be purchased online.

4.2 Supported data formats

With various data acquisition techniques available to the researchers, the selection of the tool to analyze this acquired data requires the tool to support the data format of the data on hand. It is an important issue that needs to be considered if the tool fails to support the raw data format. The raw data would need to be formatted and converted to the format supported by the tool chosen Data in their initial form may exist as audio or video recordings. They can also exist as data logs

or even verbal protocols that record the required interactions of the user with the systems under study. Most tools considered for this study do not support raw formats of data. They require that the data be formatted and saved into other formats using tools like MATLAB (2010a, The MathWorks, Boston, MA).

In this study, four out of the five tools used support standard comma separated values (csv) as an input format while a few also supported a variety of other formats of data input. The tool INTERACT was found to support only video files in a variety of formats. This allows researchers to work with video recordings.

ABSTRACT currently supports only csv format.

The tool, Theme, supports a variety of formats including,

- Theme un-compiled data files - *.erg
- Theme compiled data files - *.tsp
- The Observer Observational Data files - *.odf
- Fasta file format
- Text files in certain formats: *.sdf - Character delimited, with no column headings

     *.csv - Comma delimited, with column headings

     *.txt - Tab delimited, with column headings

4.3 Integrated Record function

Another aspect that the user needs to consider while choosing the tool is the record feature in the tool. This requirement is important for researchers who record their data in audio or video format so that the tool is comprehensive enough to record and help analyze this data. But not all areas of HCI research require this feature. For those who need the tool only to explore and analyze the already collected and formatted data, this feature adds no extra value. Among the tools used in the study, INTERACT was the only tool that supported recording of data as video before being analyzed. It also supported importing of existing videos into the tool. The recording feature of the tool gives rise to other system requirements like an audio and video input.

The tool Excel allows the data to be directly input into its spread sheets for visualization. Its does not support video input but can be used to log data.

Tools ABSTRACT, Theme and Infoscope do not support any form of data recording as a built in feature.

4.4 Level of Visualization

This is probably the most important aspect of the tool that would be the deciding factor for its selection. As we already spoke about the data collected in HCI has an inherent sequential nature that unfolds in time. The ability of the tool to be able to present this nature to the researcher makes the tool most useful.

The tool ABSTRACT was found to be able to display this to a very high extent. The tool not only

could display the data given on a time line thus successfully representing the chronological nature of the data but also displayed the Hierarchical sequential nature where a series of events led to a specific set of other events.

Figure 1: Visualization of Lane change in ABSTRACT shows a snapshot of the display of typical "highway lane change layout" on ABSTRACT (Henning,M.J., Georgeon,O., & Krems, J. ,2007). The data visualized here was collected in the multiple forms such as directly by the ergonomist, video cameras or microphone recording, sensor recording of numerical data, software "logs" among others. The data was preprocessed using MATLAB to obtain the data in a csv file that was imported into ABSTRACT. The symbols in the figure represent various actions executed by the driver while performing a lane change on a separated-lane highway. The labels are not part of the Abstract visualization; they were added in the figure for explanation. The x-axis of the display represents the time line.

ABSTRACT provides the researchers the freedom to define the shapes of symbols, colors and y-axis. This figure uses a display model where the lower level symbols are the dots at the bottom. The middle level symbols are placed around the middle horizontal axis. Everything to the left of the vehicle is placed above this axis and those events associated with the right are placed below it. Triangles oriented to the right indicate "frontward" (e.g. "Look ahead"), triangles oriented to the left indicate "backward" (e.g. "Left mirror glance"). The lines are the relations of inference between lower level symbols to higher level symbols. The upper part of the display is used for the highest level symbols which are not concerned by this left/right rule, e.g., the symbol "Decision" to change lane symbol (lavender triangle) which could be inferred from the occurrence of acceleration and left mirror gaze. This allows the tool to achieve Hierarchical event display too.

Figure 2 shows another example of the data that was visualized in ABSTRACT. The data used here was obtained from the Global Terrorist Database hosted by the University of Maryland (Georgeon, O., Morgan, J., Horgan, J., & Braddock, K., 2010). The sample used here shows the terrorist activity that caused fatalities greater than 100 between the year 1990 and 2007. We observe that each activity is represented as grey dots displayed on a time line. Each event displayed can be clicked on to get the complete information about the event in a pop-up box.

The tool, Theme, also supports similar features like ABSTRACT but for the Hierarchical display of data. Figure 3 shows the representation of the same data. The data was first modified to fit the formatting required by Theme, by creating separate Category tables and data files (*.dat). This conversion required effort as it had to be created for all the data in the current csv files, although, the tool claims to support csv format files. Importing this type of file always gave rise to errors that could not be debugged. The visualization presented here displays the events on a time line. But the information relevant to the displayed event is not available directly but available in a panel on the right side of the plot. The figure shows one such event that is mapped to the relevant event information. This makes the tool relatively less interactive than ABSTRACT in this regard.

Theme has the feature that generates statistical inferences from the data being analyzed. ABSTRACT fails in this regard as it can only visualize the data and cannot be used to obtain statistical inferences.

Figure 1: Visualization of Lane change in ABSTRACT

Day: 727
WEAPON_TYPE: Armed Assault
GTD ID:
PERPETRATOR: Unknown
TARGET:
ATTACK:
gtd_id: 1.98008E+11
COUNTRY: El Salvador
CITY: Suchitoto
FATALITIES: 300
INJURED: 0
TARGET_TYPE: Military
REGION: Central America & Caribbean
WEAPON_TYPE: Firearms
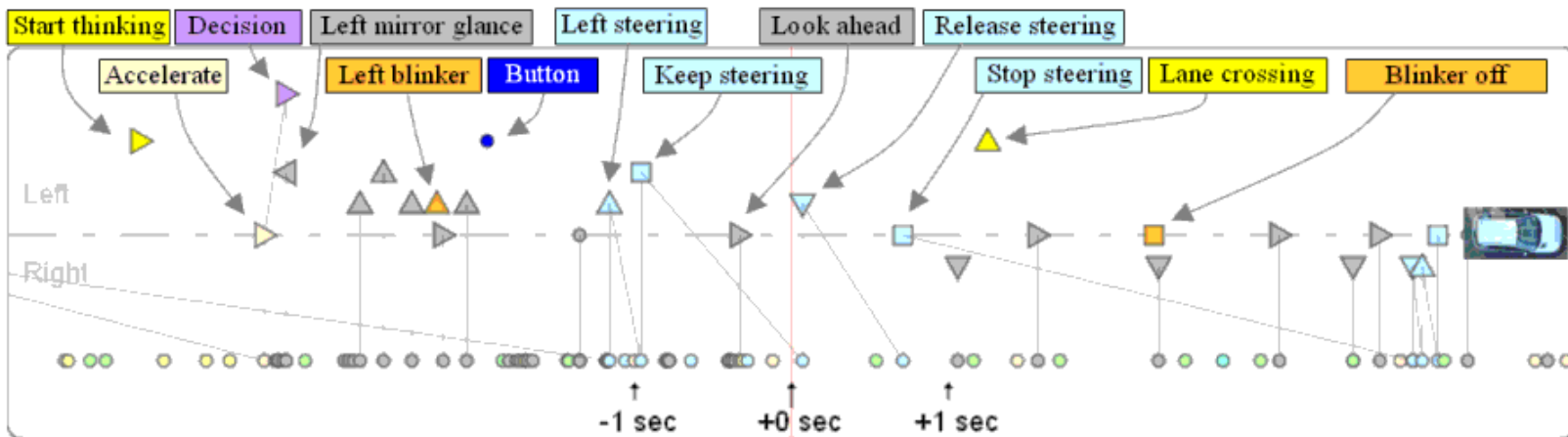
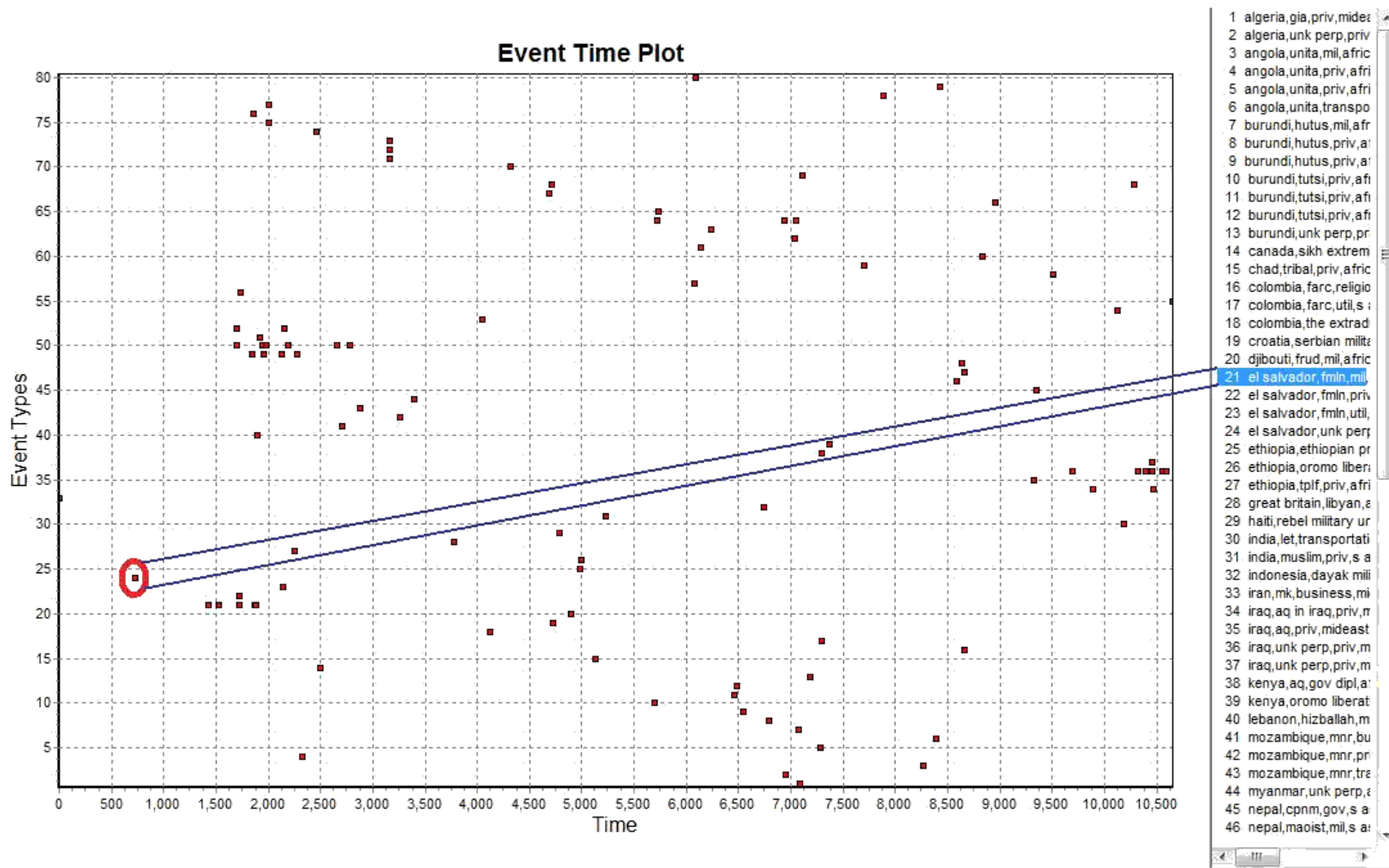**Figure 2: Terrorism Data visualization in ABSTRACT**

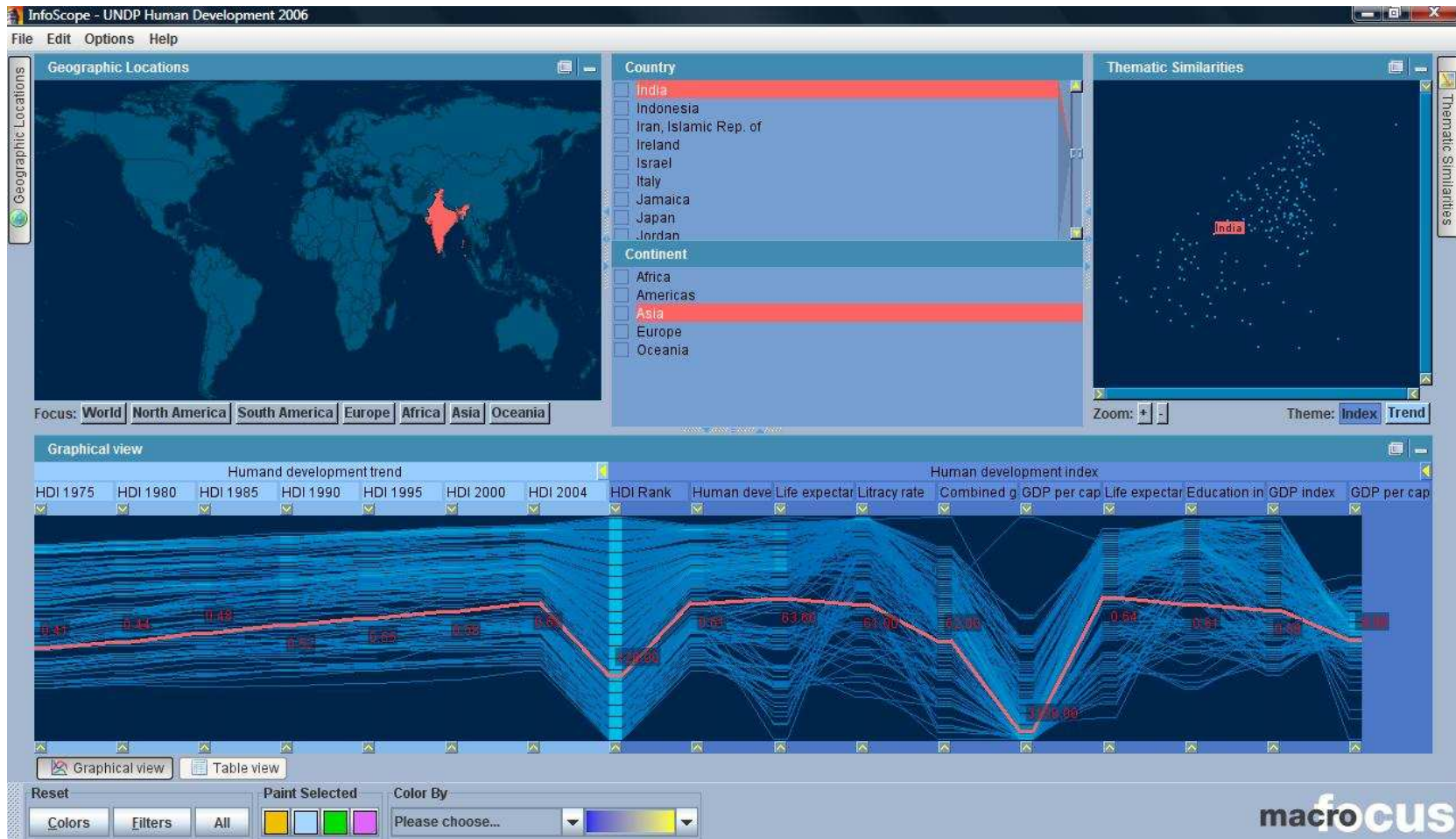**Figure 3: Terrorism Data visualization in Theme**

Figure 4: Human Development Data for the year 2007

The terrorist activity data was also used in Excel. But visualization of the time line was on the y-axis rather than the x-axis in this case. To obtain the time line representation of the data over x-axis required the data to be rearranged to get the transpose of the current format of the data in the file. The visualization was more textual than graphical in Excel. Symbols could be possibly used, which would require more effort to create simple graphical visualizations.

ABSTRACT does not support embedded video analysis while Theme supports videos created using EthoVision 2.3 or later only. This feature of Theme has not been explored in this study.

INTERACT; the other tool in the study supports only video format files as input. The tool allows analysis of multimedia files (digitized video on the computer). It can also be used to log codes during live observations. It is even possible to combine these two features. The tool supports live observation and coding the recorded digitized video. The tool successfully displayed the time line based visualization while also providing time line based statistics. The tool is very comprehensive as it allows tagging and editing video once recorded.

The tool, Infoscope, was not studied in depth as the demo version of the tool did not support any file formats that would be supported by the complete tool package. The only data sets studied on this tool was the ones provided as examples by the creators of the tool. The tool was found to be very interactive and allowed visualization of the data on a time line. The data used in this study is the human development data for a majority of countries around the world up to the year 2007. The tool allows the researcher to highlight relevant information and also displays required statistics. Figure 4 shows the visualization of the data. Clicking on the map highlights the relevant information as well as the displayed graphics. The tool also provided different views of the data as well as allows the researcher to be able to use the focus operation to zoom in on the relevant parts of the map based on the continent for the data used.

4.5 Integrated Search function

An integrated search feature is a very useful feature for quicker analysis of recorded data. Tools with this feature facilitate the researcher to spend more time on analysis of the trends and discoveries using the data rather than spending time to manually find the patterns needed.

Among the tools, Excel was found to have a comprehensive search option that allows search of patterns with various filters. ABSTRACT also claims to have a search function but is not completely functional in the currently available version of the tool. INTERACT also has an inbuilt search function that can search for codes, comments or even independent variables. Theme has no comprehensive search option. Infoscope on the other hand has a search option with a GUI, which allows the researcher to select and search from the visualized data visible. We were unable to explore the search function to a larger extent on Infoscope as only the demo version was available for the study.

4.6 Documentation and Support (Ease of use)

This is a relatively important aspect for any tool contending to be chosen as it forms the basis for the initial understanding and usability of the tool. Documentation of the tool also lends a helping hand to learn as well as use the tool better thus unblocking the power of the tool. Most of features go un-noticed by most users in the absence of good documentation. Most users never get to use them because they cannot figure out how to use them or often they do not know they even

exist. Thus good documentation adds value to the tool and therefore it becomes a very important facet that needs to be considered while choosing the tool. It can also be highlighted that availability of strong customer support for the tool is an added advantage to the user to tap the full potential of the tool.

During the study of the tools these aspects proved to be very vital. Since the study required demo versions of the tools, and assistance with the use of the tools themselves, customer support and documentation was the key to the successful study of each tool.

MS Excel was found to have a strong documentation support. Although we did not use the customer support services for the study of this tool, the documentation was found to be very interactive allowing us to make use of many features of the tool.

Theme, the tool created by Noldus International, and INTERACT, the tool created by Mangold International were found to have a very strong customer support although the tool documentation was found to be not comprehensive enough. Quick responses from the support group allowed us to understand the tools better in a short span of time.

ABTRACT tool support is also appreciable as it is an academic tool as of now and is supported by the researchers who created it. The documentation of the tool was also found to be simple and usable although not complete but evolving thanks to a wiki.

Ease of use of each tool is based on personal experiences and the conclusions presented here is restricted to the installation and use of the tools by one participant. The installation of all the tools except for ABSTRACT was found to be simple with an executable script performing the operations. On the other hand ABSTRACT was found to be complex and very dependent on the availability of other tools that may be already available on the system used for the installation. But this drawback was solved by the ready to use online installation provided by the creators of the tool.

Appendix A contains the improved installation documentation that is proposed as a result of a task protocol analysis done on the verbal protocol transcription of the installation process.

5. Conclusions

In the study we got an overview of the sample set of ESDA tools allowing us to present the most important facets that would need to be considered for the tool selection process depending on the application or data. We also found that none of the tools studied here were comprehensive enough to handle all kinds of data. Each tool was found to have its own positive aspects. While Theme was found to be a powerful tool supporting a variety of input files formats and the ability to present statistical data, it lacks the record function.

The tool ABSTRACT was found to be the only tool that was capable of hierarchical event display, but it lacks statistical support, search and record function too.

INTERACT was found to support features such as recording and tagging of video data. It also has an integrated statistical package making the tool comprehensive. But the tool supports only video format input files thus restricting its use by the supported data format.

Excel was found to be very primitive as an ESDA tool as it cannot support different formats of data like audio or video and well as fails to display the data on a time line easily. It also lacks the record function. But the tool has a very comprehensive search and statistical package.

Thus we found that none of the tools was able to satisfy all the requirements of a comprehensive ESDA tool. In order to overcome this lack of a single comprehensive too, we recommend using a combination of several tools to exploit their complementarities to achieve all the required features. For example one could use ABSTRACT for its advanced timeline visualization features with Excel for its easy editing features, search and advanced statistical processing functions.

We also found that strong documentation and good tool support plays a key role in understanding the true potential of the tool. The study presented here is not complete as we intend to target more tools in this area while extending the depth of study for each tool too. As for the future, we hope to come up with a more comprehensive study and observations that can be used as a starters guide to evaluate ESDA tools.

6. Acknowledgments

This study would not have been possible but for the co-operation from different tool creators who have been very helpful and supportive by providing the tools as well the customer support required to be able to use the tools themselves. I would like to thank Noldus International, Mangold International and ACS lab (in IST) for providing the required tool. My special thanks to Dr. Olivier Georgeon for helping me formalize and study the tools. I would like to also thank Jeremy Lothian a fellow student for his valuable comments on the report. The study has not been as complete as intended in the beginning as acquiring trial versions of tools for the study was the biggest challenge. The next challenge faced was the formatting the data for the specific tools. But the customer supports for the tools have been very helpful to sort out the issues and successfully port the data to the required format for the study.

7. Appendix

7.1: Verbal Protocol Analysis for installation of Infoscope

7.1.1. Introduction

Ericsson (2002) defines "Protocol analysis is a rigorous methodology for eliciting verbal reports of thought sequences as a valid source of data on thinking." With cognitive revolution and the advances in the acceptance of Verbal Protocol analysis, it has emerged as one of the primary methods for studying thinking in Cognitive Science (Simon and Kaplan 1989), Cognitive Psychology (Crutcher 1994), and Behavior Analysis (Austin and Delaney 1998). With more advances in the technique, it is widely used in the design of surveys and interviews (Sudman Bradburn and Schwarz, 1996), and user testing of computer software (Henderson Smith Podd and Varela-Alvarez, 1995).

In this section we study the initial comprehension of software by a general user using the Verbal Protocol Analysis. This is being done as a part of the project "A Comparative study of Exploratory Sequential Data Analysis tools". We study only the phase of software installation here. We present the transcript of the verbal protocol along with our observations.

7.1.2. Scenario

The software under study here is the freely available trial version of a sequential data analysis tool named Infoscope by Macroscope. It was chosen as an example of a commercial data analysis tool. Macroscope the creators of InfoScope claim it to be an interactive visualization tool to access, explore, and communicate large or complex datasets. Some of the features claimed by the creators of the tool are:

1. Provides an overview of the global relationships between objects.
2. Encourage and support exploration, discovery of unexpected structures and formulation of new hypotheses.
3. Embed fine grain tasks in a global frame of reference.
4. Take advantage of the human perceptual abilities.

7.1.3. System Configuration and User profile

The software was installed on a system with the following configuration: Intel Pentium Dual Core and 3GB RAM with a 32bit Windows Vista Home premium operating system.

The experiment was run on two subjects. The subjects were asked to think aloud as thoughts occurred while installing the tool. One of the subjects selected to do execute the task was selected from a group of moderately computer literate participants with prior experience of similar software installation and the other subject selected was a novice with no prior experience of tool installation but with a prior experience of being a subject in a verbal protocol experiment. The verbal protocol was recorded on a tape and then was transcribed. The subject was initially advised on the nature of the task and was given the website address on a sheet of paper to start the task.

7.1.4. Revealing observations and conclusions

One of the most expected observations was the time taken to complete the task. The installation task was completed in 7.3 minutes in total including the downloading of the trial version from the tool website by the expert user while it took about 16.3 minutes for the novice user to complete the same task. Clearly, experts have more knowledge and more highly organized knowledge structures within their domains than do novices. The expert user seemed to evaluate and anticipate any change in the installation pattern and plan ahead and reason accordingly. We made this observation when the expert user decided to save the executable on the desktop as a temporary location of the executable and installed the software in the Program Files folder while the novice user paused to ask the location for the installation.

But as we analyzed the transcribed protocol we observe that the expert user seem to show that the demands of the think-aloud task interfered with the subject's abilities to use knowledge and stammered while thinking aloud although the subject's speech was normal otherwise, which we conclude is a depiction of self-presentation issues suffered by this subject. While the novice subject took longer to execute the task, the subject seemed to be confident while thinking aloud allowing the subject to be able to vocalize better. When we began the experiment we assumed that the expert user would be more confident and comfortable while verbalizing as the subject had previous experience of such tasks but were proven wrong as we observed that the novice user was able to verbalize better. This observation made during the think-aloud procedure does not follow the conclusion of Ericsson et.al (1993) that thinking aloud does not usually affect normal cognitive processing. The verbalizing procedure seems to lead to overestimates and/or underestimates of the knowledge and processes used under normal task conditions. The need to verbalize for the think-aloud seem to interfere with subjects' abilities to use knowledge and/or processes they might use under normal conditions.

7.1.5 Verbal protocol (as transcribed from the recording of the expert user.)

- Open an instance of Internet Explorer.
- Type in www.macrofocus.com/public/products/infoscope
- Click on "Download free of charge"
- Click on Infoscope v.3.3.4 for windows
- Clicks "Save" to save the executable to the desktop
- It takes a long time. It is a 20.8 MB file
- Also available is the professional version but it needs to be purchased
- Click on the desktop icon on the task bar.
- Click on the Infoscope_windows_3_3_4.exe file.
- Click on the "Allow" button in the User Account Control pop up.
- The window says "Infoscope is now being prepared to be installed"
- Click "Next" on the Installation window.
- Type in C:\Program Files\Macroscope\Infoscope in the directory box allowed and press "Next"
- Check the "Create shortcut icon for all users option" and click "Next"
- Click "Next"
- Now Infoscope is being installed by extracting the relevant files.
- Click "Next"
- Check the Start Infoscope option
- Click finish to complete the installation.

7.2: Verbal protocol transcription and Installation guide of ABSTRACT using Task Analysis

7.2.1 Verbal Protocol Transcription

I am beginning the process of installation. I'm on the webpage. It says that I need firefox. But I have firefox. The next installation step is to install wamp or easy php and I am going ahead and trying to get wamp so there is a link here provided on where to get wamp I am opening the link… it says install php5 apache mysql4 in windows. Now… it has 3 steps mentioned.

The user reads the options available and says…

Now, I don't understand anything of what that meant. I am just trying to find something to click on and I am clicking on step 1. So, download wamp server 2.0. Save File. Now saving file on desktop. File saved. And returning to the webpage on what it says. So, Follow the instructions on the wamp website it says.. to go to your apache server running. Now, what I need to do is I guess I need actually to go in and install the wamp server. So there I am. I say run after double clicking on installation file. In the wamp setup, it says select additional tasks you would like to setup to perform while installing wamp server 2 then click next. So I think this doesn't look like something too confusing . In the installation process …33% ..50%.. 66% ..100%. So it asks me please specify the smtp server and the address to be used by php when using the function. If you are not sure leave default values, so I am leaving default values here. Click on finish. It seems like the job is done.

Getting back to the installation manual.

It says setup your php.ini . The file php.ini can be found in the apache directory. so It looks like as if this is for easy.php.

Clicking on the bottom right of the Windows task bar. Wamp server. Uncomment the lines. Remove the semicolon in front of the lines. It says by clicking on the wamp icon at the bottom right uncomment the lines remove the semi colon in front of the lines… where do I get all this ?

So the whole idea is to uncomment certain lines so I need to find the particular line that is being used. Doesn't look like a big deal. Semi colon removed. Going to the second line. So it looks like as if the second link cannot be found a problem with the find function. I have changed an off to an on. Its not an uncomment process. I have saved the .ini file. The max execution time needs to be changed to 300. This is max execution time. It is at 30.. we need to make it 300. So this job is done.

Setup your httpd.conf. It can be found in the apache configuration directory. With wamp you can access this file by clicking on the wamp icon bottom right of the windows task bar. Activate the url rewriting. So same process to just simply find a different file, Php extensions, Php settings. Httpd.conf. found a file, going back to uncomment the lines. Load module rewrite, is what we need to do. Removing the hash for the first case. There is one more.. okay.. lets look at the mod__rewrite.c file itself. Look at the set of add_modules that may be available. Its load_module not add_module. It only says uncomment the lines. Easyphp or wamp must be re started. So how do I restart wamp ? let me see the options.. restart all services. Doesn't really give me any feedback on whether it was stopped and restarted again. Next step is to , so it says in this section we describe how to retrieve the abstract project onto your computer once retrieved

abstract will be available for use as a web application from your apache server. It will use also be available for you to contribute to its development. So to do so we use eclipse with the subclipse and the Php plugins.

So the first step is to install eclipse and I have eclipse. So, next one install eclipse subclipse plugin. I am clicking on a link to find out what it is about. So the first attempt to find out what a subclipse plugin is about has been a failure because our link from that is not working…. So I can go to google and try to look at what is a subclipse plugin.

Going to homepage, going to history, display 500, someone has hacked into the page… going to next 500, next 500…

Install subclipse plug in, lets go to eclipse, we can pick the kind of eclipse, the subclipse settings may not match my version of eclipse…..we should try to get the newest version of eclipse here. A lot of time was spent because I did not know what to do.

Installing eclipse

After downloading eclipse, I am installing the new version of eclipse (Galileo), waiting for eclipse to start off. Now going to the installation webpage, before starting off at any workspace. Something is said about in the installation page. When eclipse starts this is what the installation page says, when eclipse starts - it opens the window workspace launcher that lets you choose a workspace, choose a root directory for the apache server.. this is PhP I don't have it.. I need to find the link to my wamp server. Is the path to index.php the right path? Go to properties. It says "c:\wamp\www\index.php".. I need to go in this. This is the workspace, and the workspace is setup.

It says define the repository location… open the SVN perspective. Let's go to workbench and open the SVN perspective. Looking at a number of perspectives here, trying to find the SVN perspective. The webpage doesn't say how do open the SVN perspective. I have to install SVN. I don't have subversion on my system.. I need to download subversion. I hope it doesn't take too much time. Is there another way to find SVN perspective ? Help -> Search -> SVN perspective…

Not successful

Ok right … download SVN. Lets go to tigris – subversion.tigris. Select the latest version. Choosing the one for Win32.

I have been able to get into this SVN repository view in eclipse. Need to get back to what the next instruction says in the installation. So open the SVN perspective. Right click on the SVN repository window and click in new repository location and there is a specific URL that needs to be entered there.

Entered the URL, operation in progress. Error validating server certificate, Unknown server certificate. Accept temporarily. It seems to have gotten through some success in this.

Download the abstract project in the SVN rep window… right click on the rep. In the contextual menu choose check out. It is giving me an error…. Finish. It was my system that was throwing the error.

This part is done in the subversion thing. So, what does all this mean ?... so this is the code of the actual tool. Each time there is an update on this code.. I can update it here after checking out. Create your data directory.. it must be a sub directory or the root web directory of your apache server. You create it from eclipse.. the PhP perspective is what we have right now… right click on your php explorer window … in the contextual choose new php project.. .. write a folder name - traces. Configure your copy of abstract…. Create subfolders of abstract … traces.. new folder.. or is it new subfolder ?

The other thing is new folder… temp.. this step lets you configure local paths related to the installation of abstract on your computer. In eclipse open the php perspective .. in the abstract project open include settings.inc.php.   Trying to find the path to the project folder. Source -> configure -> ….no… team …no…. build-path -> include-path..no…… go to…. No.

I am supposed to change the URL. Configure your .ht access. The file is hidden. It is.ht access. How do we see a file that is hidden? We need to check them where they are stored. Here we are. Need to change the abstract base URL. Remove the extra characters.

Ready to access ABSTRACT with Firefox. Starting firefox. Entering the URL.

There is an error in some file. Is the wamp running ? Yes.

Was the port changed in skype ? no. We need to change a port in skype. Let us change the port in Skype. Starting Skype. Tools->options->connections ->use port 80. Changing port, it shouldn't be a problem as Skype was not on. Trying to start ABSTRACT again.

7.2.2 Installation guide to ABSTRACT

The installation of the tool ABSTRACT has a few prerequisite software tools

that need to be available on the system before it can be installed. The

following instructions will guide through the pre-requisites to be installed

first.

Firefox:

This web-browser is required to run ABSTRACT as it is the only browser that

natively supports SVG (Scalable Vector Graphics).

Check if your system already has Firefox installed.

If you don't find it then you can install Firefox

* Download Firefox from http://www.getfirefox.com/

* Install the browser and restart the browser.

Install an Apache server:

In this section, we describe how to install and configure the Apache server to run ABSTRACT on a Windows machine.

There are 2 Apache server tools that can be installed: Easy PHP or Wamp.

- If you choose Easy PHP, then install it from http://easyphp.org/.

Follow the instructions on the Easy PHP website to get your Apache server running.

- If you choose Wamp, then install it from http://www.wampserver.com/en/.

Follow the instructions on the wamp website to get your Apache server running.

Configuring the Apache Server

To complete the installation of the Apache server we would need to configure a few files. Follow the below instructions.

Setup your php.ini:

The file "php.ini" can be found in the apache directory (i.e.: c:/Program Files/EasyPHP1-8/apache/).

With Wamp, you can access this file by clicking on the Wamp icon at the bottom-right of the windows task bar.

- Uncomment the lines (remove the ";" on front of the lines):

"extension=php_xslt.dll" or "extension=php_xsl.dll"

"allow_url_include = On"

- Change the value of the parameter "max_execution_time" :

"max_execution_time = 300;

This is the maximum execution time of each script, in seconds"

Setup your httpd.conf:

The file "httpd.conf" can be found in the apache configuration directory (i.e.: c:/Program Files/EasyPHP1-8/apache/conf/).

With Wamp, you can access this file by clicking on the wamp icon at the bottom-right of the windows task bar.

- Activate the url rewriting: Uncomment the lines : (remove the "#" on front of the lines):

"LoadModule rewrite_module modules/mod_rewrite.so"

"AddModule mod_rewrite.c".

- Restart Apache

Easy PHP or Wamp must be restarted for these modifications to take effect.

Installation of Eclipse and Subclipse:

In this section, we describe how to retrieve the Abstract project onto your computer. Once retrieved, Abstract will be available for use as a web application from your Apache server. It will also be available for you to contribute to its development!

To do so, we use Eclipse with the Subclipse and the PHP (PDT) plugins.

- Install the latest version of eclipse from http://www.eclipse.org/downloads/

You may choose your eclipse package according to your other needs. If you only need Eclipse for Abstract, we recommend you chose the "Eclipse for PHP Developers".

- Install the Subclipse plug-in by following the instructions at: http://subclipse.tigris.org/.

- To check whether the PDT plugin is already installed in the Eclipse installed go to the "Window / open perspective / Other" menu.

- If the perspectives "PHP" and "PHP Debug" are listed, then PDT is already installed.

- If not already installed, install the PDT plugin by following the instructions at: http://wiki.eclipse.org/PDT/Installation.

Checkout the Abstract project:

This step lets you retrieve the Abstract project from our SVN repository on the Liris server.

- Run Eclipse

When Eclipse starts, it opens the window "Workspace Launcher" that lets you chose a Workspace.

- Choose the root web directory of your Apache server (i.e.: for Easy PHP c:/Program Files/EasyPHP1-8/www/).

- To define the repository location, open the SVN perspective

- Right-click in the SVN Repository window

- Create a new repository location.

- Provide the repository URL: https://svn.liris.cnrs.fr/abstract/abstract

- Download the Abstract project

- In the SVN Repository window, right-click on the repository https://svn.liris.cnrs.fr/abstract/abstract

- In the contextual menu, choose "checkout".

- Choose a name for your local Project, i.e.: "abstract" and click OK

The Abstract project will be downloaded to your computer.

By default, you have read only privileges.

- Create your data directory

The data directory is the directory that will contain your traces.

  It must be a subdirectory of the root web directory of your Apache server (i.e.: For example if you have Easy PHP then c:/Program Files/EasyPHP1-8/www/).

  It can be created from Eclipse (Being able to view your trace files within an Eclipse project may turn out useful)

- In the PHP perspective: Right-click in the PHP Explorer window

24

- In the contextual menu, choose "New / PHP Project".

- Provide a folder name, i.e.: abstractUserData.

- Then create the subfolders abstractUserData/traces and abstractUserData/temp.

Configure your copy of Abstract:

This step is to configure the local paths related to the installation of Abstract on your computer.

- In Eclipse, open the PHP perspective

- In the abstract project, open the file includes/settings.inc.php.

- This file contends the following lines:

<?php

// use ABSOLUTE urls

define('TRACE_USER_DIRECTORY_PATH','d:/Abstract/abstractUserData/traces/');

define('TEMP_DIRECTORY_PATH','d:/abstractUserData/temp/');

define('ABSTRACT_BASE_URL','/abstract/');

?>

- Define TRACE_USER_DIRECTORY_PATH to specify the directory that contains your traces, i.e.: c:/Program Files/EasyPHP1-8/www/abstractUserData/traces.

- Define TEMP_DIRECTORY_PATH to specify the directory that contains the cache files for the visualization of your traces, i.e.: c:/Program Files/EasyPHP1-8/www/abstractUserData/temp.

- Define ABSTRACT_BASE_URL to specify your base url for abstract in your Apache server, i.e.: /abstract/

- If abstract doesn't find the traces and temp folders, it will not open correctly but redirect you to the /intall/setup.php page.

Configure your .htaccess:

This file is a hidden file so it will not appear in your PHP Explorer window in Eclipse.

- The file .htaccess located at the root of your abstract project must contain the lines:

RewriteEngine on

RewriteBase /abstract/

- The RewriteBase value must be set the same as the ABSTRACT_BASE_URL in the above section.

Access Abstract:

You are now ready to access Abstract within Firefox from the root directory of your local apache server

- Type in http://localhost/abstract/ in the Firefox browser

You can import your own CSV traces by following the instructions in Import CSV Trace.

8. References

David M. Hilbert, David F. Redmiles, Extracting usability information from user interface events, ACM Computing Surveys (CSUR), v.32 n.4, p.384-421, Dec. 2000

Penelope M. Sanderson, Carolanne Fisher, Introduction to this special issue on exploratory sequential data analysis, Human-Computer Interaction, v.9 n.4, p.247-250, September 1994.

Penelope M. Sanderson, Carolanne Fisher, Exploratory sequential data analysis: foundations, Human-Computer Interaction, v.9 n.4, p.251-317, September 1994

Penelope M. Sanderson, Carolanne Fisher, Exploratory sequential data analysis: foundations, Human-Computer Interaction, v.9 n.4, p.251-317, September 1994

Tukey, J. W. (1977) Exploratory Data Analysis. Addison-Wesley, Reading, MA

Mackay, W.E (1989). EVA: An experimental video annotator for symbolic analysis of video data. SIGCHI Bulletin, 21, 68-71

Roschelle, J., Pea, R,. & Trigg, R. (1990) VideoNoter: A tool for exploratory video analysis (Technical Report IRL 90-0021) Palo Alto, CA: Institute for Research on Learning

Sanderson, P.M., James, J.M., & Seidler, K.S. (1989) SHAPA: An interactive software environment for protocol analysis. Ergonomics, 11, pp. 1271-1302.

Hartwig, F. and Dearing, B.E. (1979) Exploratory Data Analysis, Newberry Park, CA: Sage Publications, Inc

Georgeon,O.(2008) Analyzing traces of activity for modeling cognitive schemes of operators AISB Quarterly, 127, 1-3

Henning,M.J., Georgeon,O., & Krems, J. (2007).The quality of behavioral and environmental indicators used to infer the intention to change lanes Paper presented at the 4th International Driving Symposium on Human Factors in Driver Assessment, Stevenson, Washington USA, 231-237

Georgeon, O., Morgan, J., Horgan, J., & Braddock, K. (2010) Process Modeling for the Study of Non-State Political Violence Paper presented at the 19th Annual Conference on Behavior Representation in Modeling Simulation, Charleston, NC.

Ericsson, K. Anders, and Simon, Herbert A., 1993 Protocol Analysis: Verbal Reports as Data, revised edition. Cambridge, MA: MIT Press.

Payne, John W. 1994. "Thinking Aloud: Insights into Information Processing" Psychological Science5 (5):241 - 248.