



AFRL-RH-WP-TR-2014-0098

BEHAVIORAL INFLUENCE MODELING AND SIMULATION

**Scott Klein
Ball Aerospace & Technologies Corporation
Systems Engineering Solutions
Fairborn, OH 45324**

JUNE 2014

FINAL REPORT

Distribution A: Approved for public release; distribution unlimited

**AIR FORCE RESEARCH LABORATORY
711th HUMAN PERFORMANCE WING
HUMAN EFFECTIVENESS DIRECTORATE
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
UNITED STATES AIR FORCE**

STINFO Copy

Notice and Signature Page

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88 th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>)

AFRL-RH-WP-TR-2014-0098 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//

MICHAEL YOUNG, Ph.D.
Work Unit Manager
Human Trust and Interaction Branch

//signature//

LOUISE CARTER, Ph.D.
Chief, Human-Centered ISR Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YY) 30 June 2014		2. REPORT TYPE Final		3. DATES COVERED (From – To) 22 Apr 2009 – 30 June 2014	
4. TITLE AND SUBTITLE Behavioral Influence Modeling and Simulation				5a. CONTRACT NUMBER FA8650-08-D-6801/0020	
				5b. GRANT NUMBER NA	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Scott Klein				5d. PROJECT NUMBER	
				5e. TASK NUMBER TO-20	
				5f. WORK UNIT NUMBER H06Z (7184X23W)	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Ball Aerospace & Technologies Corporation Systems Engineering Solutions 2875 Presidential Drive, Suite 180 Fairborn, OH 45324				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Materiel Command Air Force Research Laboratory 711th Human Performance Wing Human Effectiveness Directorate Human-Centered ISR Division Human Trust and Interaction Branch Wright-Patterson AFB OH 45433				10. SPONSORING/MONITORING AGENCY ACRONYM(S)	
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S) AFRL-RH-WP-TR-2014-0098	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution A. Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES 88ABW-2014-4352; Cleared 16 September 2014					
14. ABSTRACT (Maximum 200 words) This final report summarizes research and development support that Ball Aerospace provided to the 711 th Human Performance Wing, Human Effectiveness Directorate (RH), through the Warfighter Interface Research and Technology Operations (WIRTO) Contract (FA8650-08-D-6801), Task Order 0020, "Behavioral Influence Modeling and Simulation." This task order supported research within the Human-Centered ISR Division (RHX) that focused on cognitive models that incorporated various types of tacit knowledge, such as cultural ethics and behavior. Models of tacit knowledge need to be visualized in a format that provides analysis, comparison, manipulation, and recognition of that knowledge. Research into data collection tools and translation services was executed. Results are provided. All task objectives were met and all required deliverables were submitted.					
15. SUBJECT TERMS Operator Model Architecture (OMAR), human social cultural behavioral modeling and simulation, cognitive models, tacit knowledge					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor) Dr. Michael Young 19b. TELEPHONE NUMBER (Include Area Code)
a. REPORT	b. ABSTRACT	c. THIS PAGE			
U	U	U	U	64	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

THIS PAGE LEFT BLANK INTENTIONALLY.

TABLE OF CONTENTS

LIST OF FIGURES	v
1.0 SUMMARY.....	1
2.0 INTRODUCTION.....	2
3.0 METHODS, ASSUMPTIONS AND PROCEDURES	3
3.1 Scenario Generation and Display.....	3
3.1.1 OMAR.....	3
3.1.2 Neverwinter Nights	3
3.2 Analysis Application.....	3
3.2.1 PDAT.....	3
3.2.2 Scenario Tools.....	3
3.3 Researching Categorization and Stereotyping.....	3
4.0 RESULTS AND DISCUSSION.....	5
4.1 PDAT Data Generator.....	5
4.1.1 Main Screen.....	5
4.1.2 Backcloth File	5
4.2 PDAT.....	5
4.2.1 Main Screen Display	5
4.2.2 Menu.....	6
4.2.3 Tool Bar.....	7
4.2.4 Graph.....	7
4.2.5 Data	8
4.2.6 Status Bar	8
4.2.7 Scenario Tools.....	8
4.2.8 Graph Manipulation	13
4.3 Graph Structure.....	13
4.3.1 Conceptual Graph Structure.....	14
4.3.2 Scenario Based Graph Structure	15
4.4 Data Collection Selection	18
4.4.1 Data Collection.....	18
4.4.2 Criteria for English data collection tool.....	18
4.4.3 Phase One Collection Tool.....	19
4.4.4 Phase Two Data Collection Tool	21
4.5 Survey Results	22
4.5.1 Qualtrics Survey Results	23
4.5.2 Tablet Survey Results.....	23
4.6 Survey Interpreter	24
4.6.1 Menu.....	24

4.6.2	Main Screen.....	25
4.6.3	Interpret Panel	25
4.7	Survey Analysis	26
4.7.1	Menu.....	27
4.7.2	Loading and Viewing Survey Results	27
4.8	Report.....	47
4.8.1	Completed Survey Results	48
4.9	Panel Information.....	49
4.10	Translation	49
4.10.1	Criteria for translation tool	49
4.10.2	Translation Options	49
4.10.3	Human Translation Options.....	49
4.10.4	Machine Translation Options	49
5.0	CONCLUSIONS	51
5.1	Aggregation.....	51
5.2	Question Categories	53
6.0	SYMBOLS, ACRONYMS AND ABBREVIATIONS.....	55
7.0	GLOSSARY	56

LIST OF FIGURES

Figure 1. Neverwinter Nights	4
Figure 2. Neverwinter Nights' Office.....	4
Figure 3. Neverwinter Nights' Conversation.....	4
Figure 4. PDAT Application.....	6
Figure 5. Menu.....	7
Figure 6. Tool Bar.....	7
Figure 7. Polyhedral Graph vs. 2D Graph	7
Figure 8. Run Scenario	9
Figure 9. Manual Exercise	10
Figure 10. Session Comparison	11
Figure 11. Single Session Comparison	12
Figure 12. Dual Session Comparison	13
Figure 13. Dual Side-By-Side.....	14
Figure 14. Conceptual Graph.....	15
Figure 15. Scenario Based Graph	16
Figure 16. Interest Trigger	17
Figure 17. Decision.....	17
Figure 18. Category Question Display.....	20
Figure 19. Demographic Question Display	21
Figure 20. Tablet Category Question Display	22
Figure 21. Tablet Demographic Question Display	23
Figure 22. Survey Interpreter Main Screen.....	24
Figure 23. Survey Interpreter Menu.....	24
Figure 24. Loaded Results in Survey Interpreter	25
Figure 25. Interpret Screen in Survey Interpreter	26
Figure 26. Survey Analysis Main Screen	27
Figure 27. Survey Analysis Menu	27
Figure 28. Survey Analysis Load Screen.....	28
Figure 29. Survey Analysis with Results.....	29
Figure 30. Question Table.....	30
Figure 31. Question Information.....	31
Figure 32. Summary Report.....	32
Figure 33. Connectivity	34
Figure 34. Aggregates.....	35
Figure 35. Aggregate Report.....	36
Figure 36. Tentatives	36
Figure 37. Answer Table.....	37

Figure 38. Answers	38
Figure 39. Subjects.....	39
Figure 40. Subject Summary.....	40
Figure 41. Panel Report	41
Figure 42. Subject Aggregates.....	41
Figure 43. Demographic Information Groups	42
Figure 44. Data Annotation.....	42
Figure 45. Modify Question Types.....	43
Figure 46. Nationalities.....	44
Figure 47. Time in US	45
Figure 48. Languages.....	46
Figure 49. Survey Summary	46
Figure 50. Survey Summary Selection	47
Figure 51. Summary Survey (American).....	48
Figure 52. Translation Option Table.....	50
Figure 53. Newspaper Answer.....	51

1.0 SUMMARY

This final report summarizes research and development support that Ball Aerospace provided to the 711th Human Performance Wing, Human Effectiveness Directorate (RH), through the Warfighter Interface Research and Technology Operations (WIRTO) Contract (FA8650-08-D-6801), Task Order 0020 (TO 20), "Behavioral Influence Modeling and Simulation." Under TO 20, Ball Aerospace supported research within the Human-Centered Intelligence, Surveillance and Reconnaissance (ISR) Division (RHX) that was focused on cognitive models that incorporated various types of tacit knowledge, such as cultural ethics and behavior. In the context of the overall WIRTO research, this task focused on supporting research in cognitive modeling and dynamic visualization of understanding. Models of tacit knowledge need to be visualized in a format that provides analysis, comparison, manipulation, and recognition of that knowledge. Research into data collection tools and translation services was executed. Results are provided.

2.0 INTRODUCTION

Under WIRTO TO 20, Ball Aerospace supported research on cognitive models that incorporated various types of tacit knowledge such as cultural ethics and behavior. The cognitive models were created with the Operator Model Architecture (OMAR) software for the Air Force Research Laboratory (AFRL). The cognitive models animated avatars in the simulation environment Neverwinter Nights (NWN). A software visualization tool called the Polyhedral Dynamics Analysis Tool (PDAT), also developed by Ball for AFRL earlier under this task order, was modified to enable the tacit knowledge in the cognitive models to be displayed. PDAT enables models of tacit knowledge to be visualized in a format that provides analysis, comparison, manipulation, and recognition of the tacit knowledge. Under this task order, the PDAT was modified to incorporate scenario data and analysis. Scenario data is generated from scenario runs using OMAR and NWN. PDAT will provide tools to analyze knowledge models and recognize knowledge patterns. Research into data collection tools and translation services was executed.

TO 20 consisted of three subtasks: one technical subtask, a program management subtask, and final report delivery. This final report documents the activities and accomplishments of those subtasks and the subcontractor, and fulfills Contract Data Requirements List (CDRL) Data Item A001.

3.0 METHODS, ASSUMPTIONS AND PROCEDURES

3.1 Scenario Generation and Display

Scenarios are run using OMAR to send commands to NWN, which displays the scenario in a virtual environment.

3.1.1. OMAR

OMAR is a software suite that supports the development of simulation and agent-based systems. It is the software architecture in which the cognitive models and scenarios have been developed for this project. The two main functions are to send and receive data on the scenario to NWN, and to output the results of the scenario into a scenario (.scn) file for use in the PDAT application.

3.1.2. Neverwinter Nights

Neverwinter Nights is a software application that provides a virtual environment for our scenarios. Virtual environments have been generated in the application to simulate our generated scenarios. Agents are controlled via command from the OMAR application, and results of the actions are returned to OMAR for recording. Figures 1, 2, and 3 are examples of the office virtual environment that was generated.

3.2 Analysis Application

Analysis of the scenario is performed by the PDAT application.

3.2.1. PDAT

PDAT was updated to handle scenario data that came from the OMAR/Neverwinter Nights' scenarios. PDAT was originally developed to graph and analyze cognitive models using a number of analysis tools and reports that could determine strong links between elements and help discover patterns of stereotypes.

3.2.2. Scenario Tools

In order to fully use the data provided in the scenario files, a number of scenario tools were added for use in PDAT. These tools include Run Scenario, Manual Scenario, Single Session Comparison, Dual Session Comparison, and Dual Side By Side Comparison.

3.3 Researching Categorization and Stereotyping

Research was conducted to expand the scenarios with better knowledge models.

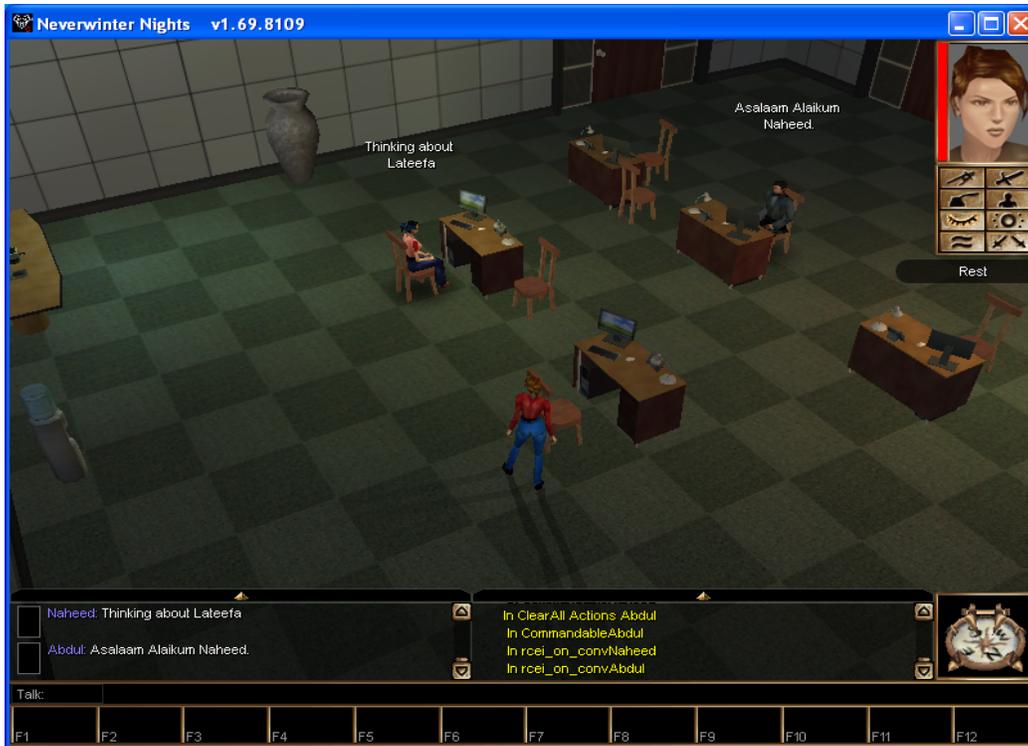


Figure 1: Neverwinter Nights



Figure 2. Neverwinter Nights' Office



Figure 3: Neverwinter Nights' Conversation

4.0 RESULTS AND DISCUSSION

4.1 PDAT Data Generator

The PDAT Data Generator takes the scenario files generated from OMAR and parses out the applicable data for use in generating a backcloth (.bclth) file for use in PDAT. The backcloth lists nodes that are elements of the scenario and places them on the graph according to the type and time of the event. The Generator can also create backcloth files without loading a scenario file, but the output is less sophisticated and more time intensive.

4.1.1. Main Screen

The main screen has three major sections: the scenario panel, the nodes panel, and the concepts panel. The scenario panel is used to load the scenario file. The concepts (agents) can be selected on load. Once loaded, the nodes and concepts are filled in. The lists can be modified in each respective panel. Once the data is correct, the user saves the data into a backcloth file.

4.1.2. Backcloth File

The backcloth file is generated by the PDAT Data Generator, and loaded by PDAT. The file is xml based, and it contains the nodes and concepts that will be graphed. Along with the nodes and concepts is the scenario data. The scenario data lists nodes and concepts, along with the time at which they are activated. The order of the nodes also determines the path that is graphed. The node and concept list contains all the possible activation nodes and concepts, but each scenario within the file may use a smaller percentage of them.

4.2 PDAT

4.2.1. Main Screen Display

The main screen is made up of five elements: Menu, Tool Bar, Graph, Data, and Status Bar (see Figure 4). This design has the main elements displayed (graph, concept, and node lists) with other features one to two clicks away.

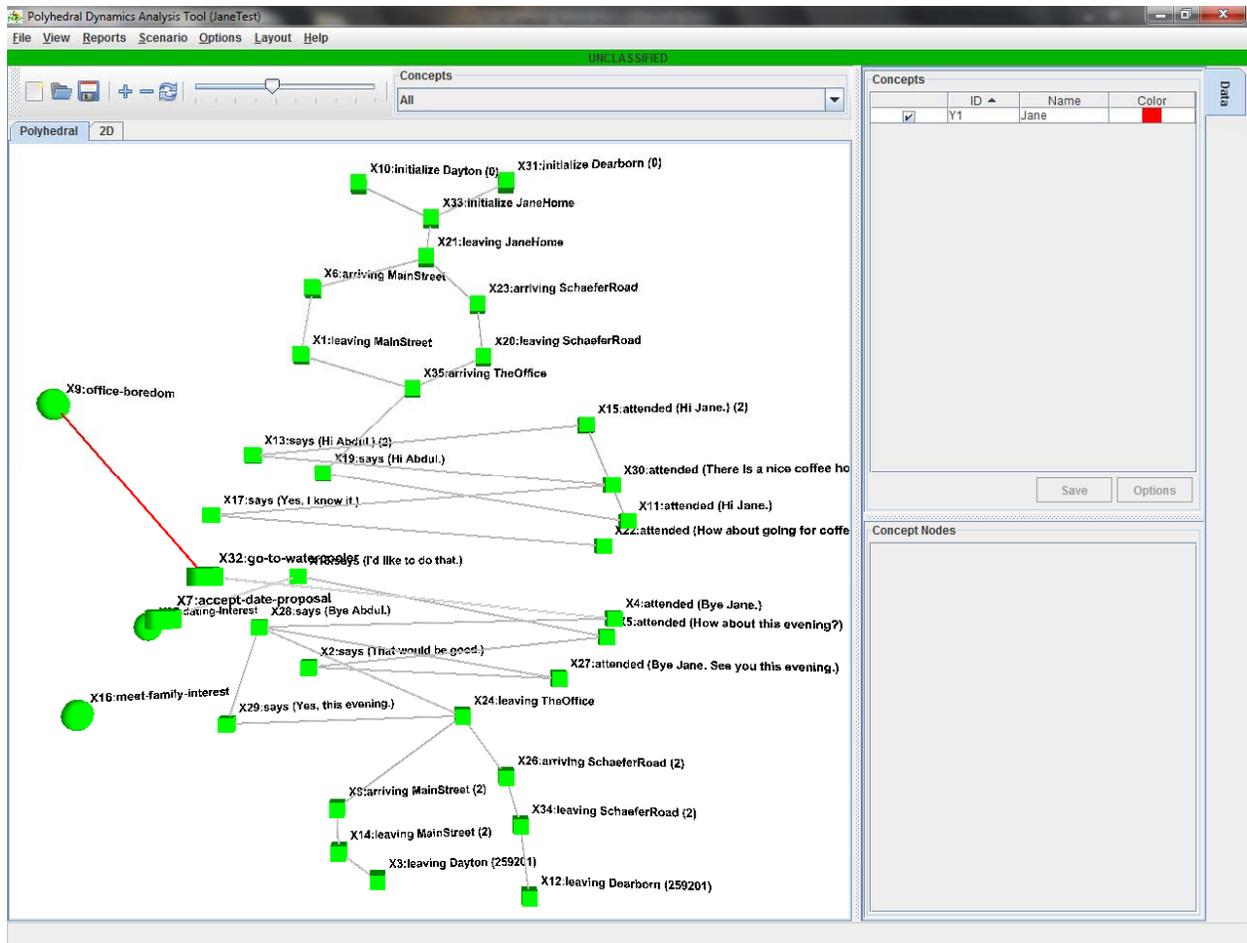


Figure 4: PDAT Application

4.2.2. Menu

The menu consists of seven menu items: File, View, Reports, Scenario, Options, Layout, and Help (see Figure 5). The File menu contains the file load/save actions for the application. The View menu has viewing selections for the application. The Reports menu lists reports for the applications, broken up into four sections: Graph Data, Node, Edge and Concepts. All scenario tools are found under the Scenario menu. Application options are located in the Options menu. The different layouts in which the graph can be generated are listed in the Layout menu. The Help menu contains the various help features.

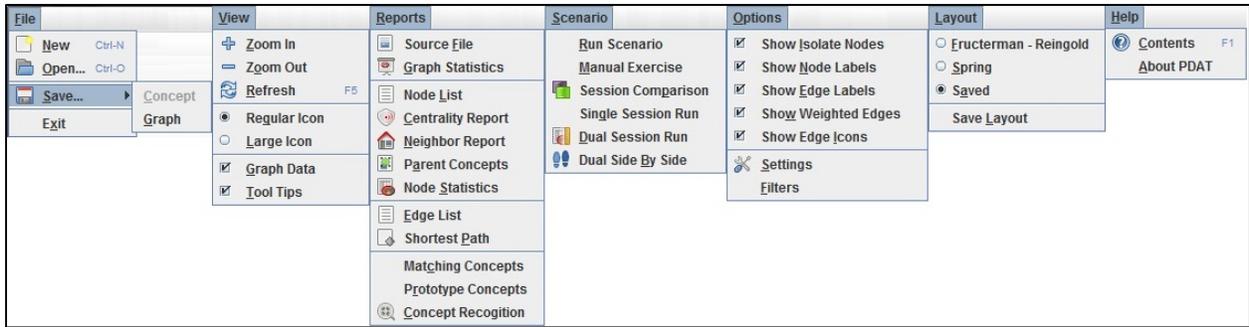


Figure 5: Menu

4.2.3. Tool Bar

The tool bar has quick launch icons for the application (see Figure 6). It also contains some graph manipulation tools, such as zoom and node sizing. The quick launch icons contain some of the menu items. The last element on the tool bar is the Concepts combo box.



Figure 6: Tool Bar

4.2.4. Graph

The majority of the application display is the graph itself. The graph can be built in two different ways, Polyhedral and two dimensional (2D) (see Figure 7). The polyhedral is a three dimensional (3D) display of the graph, whereas the 2D is a two dimensional display of the graph. 2D displays typically generate faster than polyhedral displays.

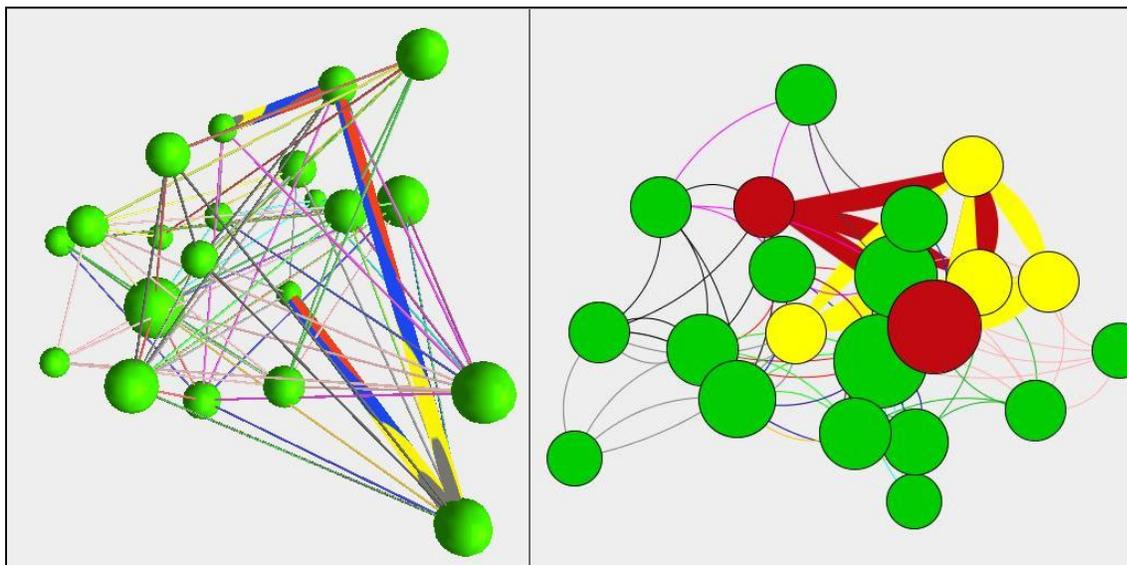


Figure 7: Polyhedral Graph vs. 2D Graph

4.2.5. Data

The right side of the application contains the graph data. It is split into two lists: concepts and concept nodes. The concept list contains all concepts that are used to build the graph. The concept node list populates with data when a concept is selected. The lists provide some functionality, mainly centering selected nodes and highlighting selected concepts. Concepts can also be modified through this display.

The right side of the application can display reports/tools, depending on whether reports are generated as pop-ups or tabs on the right side.

4.2.6. Status Bar

The status bar is located at the bottom of the application. It displays information and warnings about the application. Information is displayed for a given amount of time and the status bar displays information such as node data and load status. Errors are displayed in red, without a timer, and inform the user of a failure in the application, such as a file that did not load correctly.

The status bar also displays progress for those functions which are time consuming, such as drawing a large graph, or loading data.

4.2.7. Scenario Tools

The Scenario menu contains six different tools for the application. These tools are similar to those used for reports; they can be displayed as a tab or in their own window. The scenario tools provide two key functions: displaying graph activation from the provided scenario and analysis of the scenario.

Run Scenario: Run Scenario is an application that takes Scenario files (.scn) that were generated by OMAR and allows the user to run the scenario against the backcloth loaded into the PDAT graph. Run Scenario has a number of settings for manipulation of how the scenario is run and viewed. These settings include: following scenario time, or specifying the amount of time between each event (see Figure 8). Nodes are activated in the graph according to the specifications of the scenario. The amount of time the node is activated can be modified.

Run Scenario can also be modified to take images for future analysis. Images can be predefined in the settings, such as after an activation event or in a constant interval. Image events can be added on the main screen for a specific time or event number.

The application is broken up into six parts: Session, Scenario, Show Node, Event, Activation and Button Panel. Session contains the session information, name, time, nodes activated and concepts activated. Scenario loads the scenario file. Show Node sets the node for display in the activation panel. Event is broken into three tabs: All, Recent and Imaging. All lists all events in the scenario. Recent lists the last event, the current event, and the next two events, with the current event displayed larger for easy viewing. Imaging lists and modifies imaging events. The activation panel displays a line graph for interest nodes and attribute and disposition values for decision nodes. The button panel contains the settings and scenario run controls.

The application will run through the scenario, displaying each event as it is processed. It also updates the graph and activation panel. Images are taken and stored in the file system for later analysis.

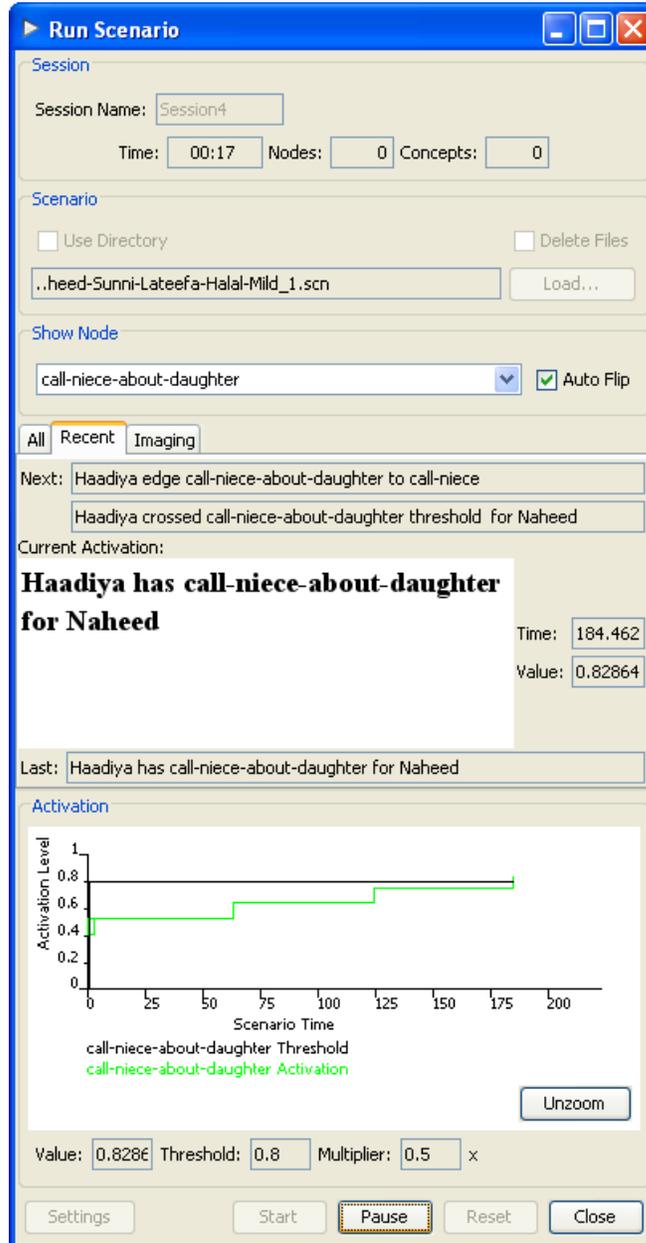


Figure 8: Run Scenario

Manual Exercise: Manual Exercise is a simple exercise tool that uses data that is completely user generated. The user creates a list of attributes to activate and sets the level of activation. Once the list of activations are created and applied, the matching concepts are listed in order of closest match. This tool can be used in defining stereotypes based on the loaded backcloth (see Figure 9).

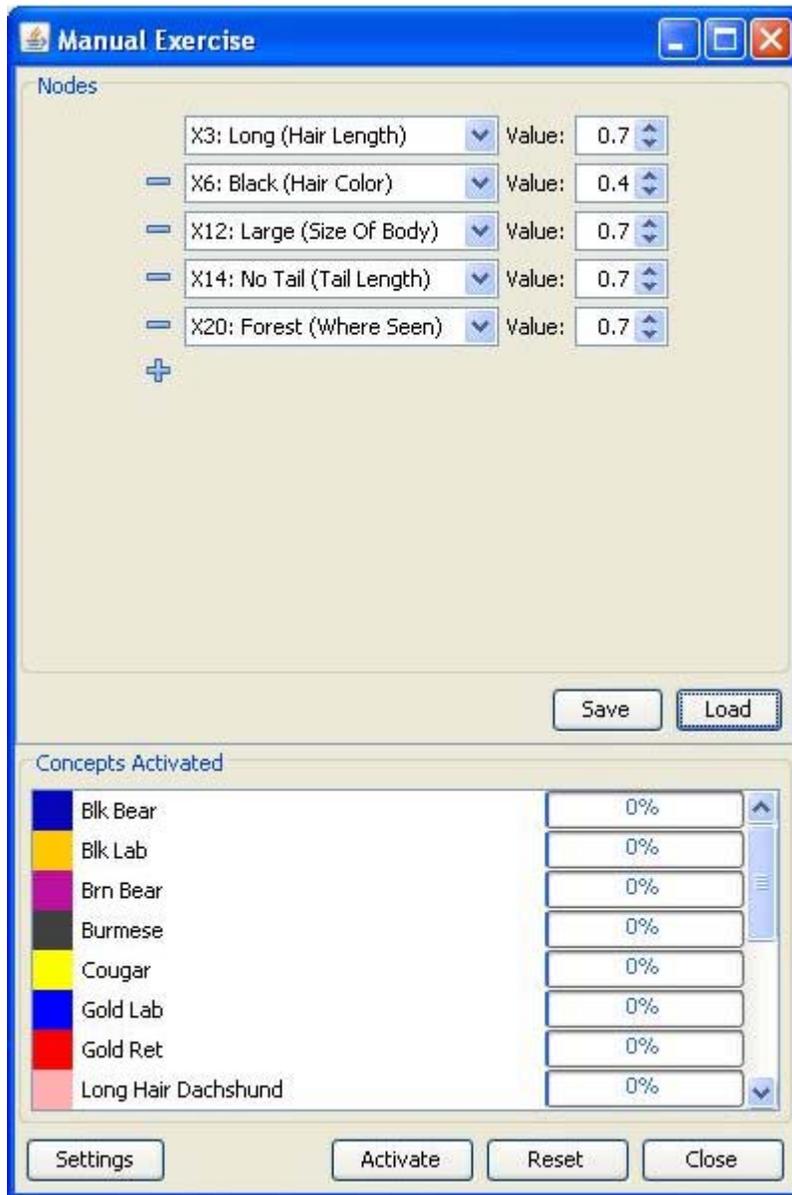


Figure 9: Manual Exercise

Session Comparison: Session Comparison takes the session data saved from Run Scenario and compares it to a second set of session data. The display is broken up into three sections: a session panel, a common panel and another session panel. Each session has its own display, with the panel containing common data in between the two. The session panel displays a graph with lists of unique activated nodes and concepts. The graph and lists are pulled from the time or event selected in the common section. If an activated node is activated in both sessions, the information is placed in the common area (see Figure 10).

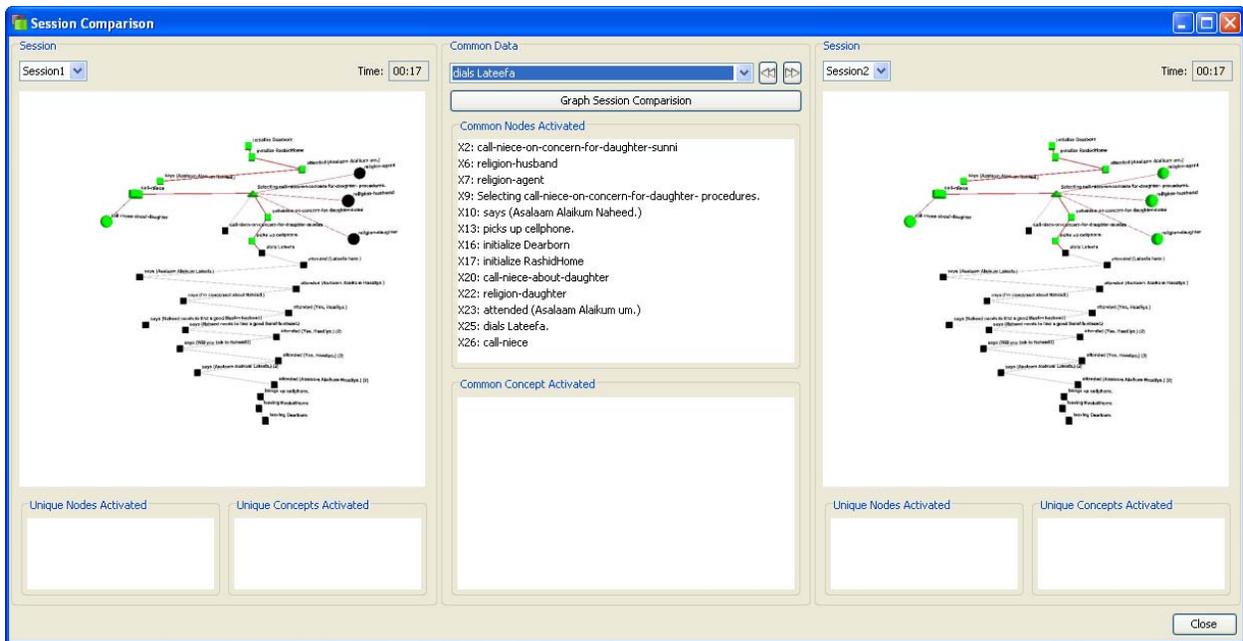


Figure 10: Session Comparison

Single Session Run: Single Session Run is very similar to Run Scenario, in that it takes a scenario file and activates the graph based on events from the scenario file. It differs in that it saves no session or image information and has no activation panel. The main difference is that it does not run through the scenario in a step by step, time based manner. The user selects a range of time or events (i.e., from 1 minute to 3 minutes, or event 6 through event 12), and all nodes/concepts that are active in this range will be activated immediately (see Figure 11).

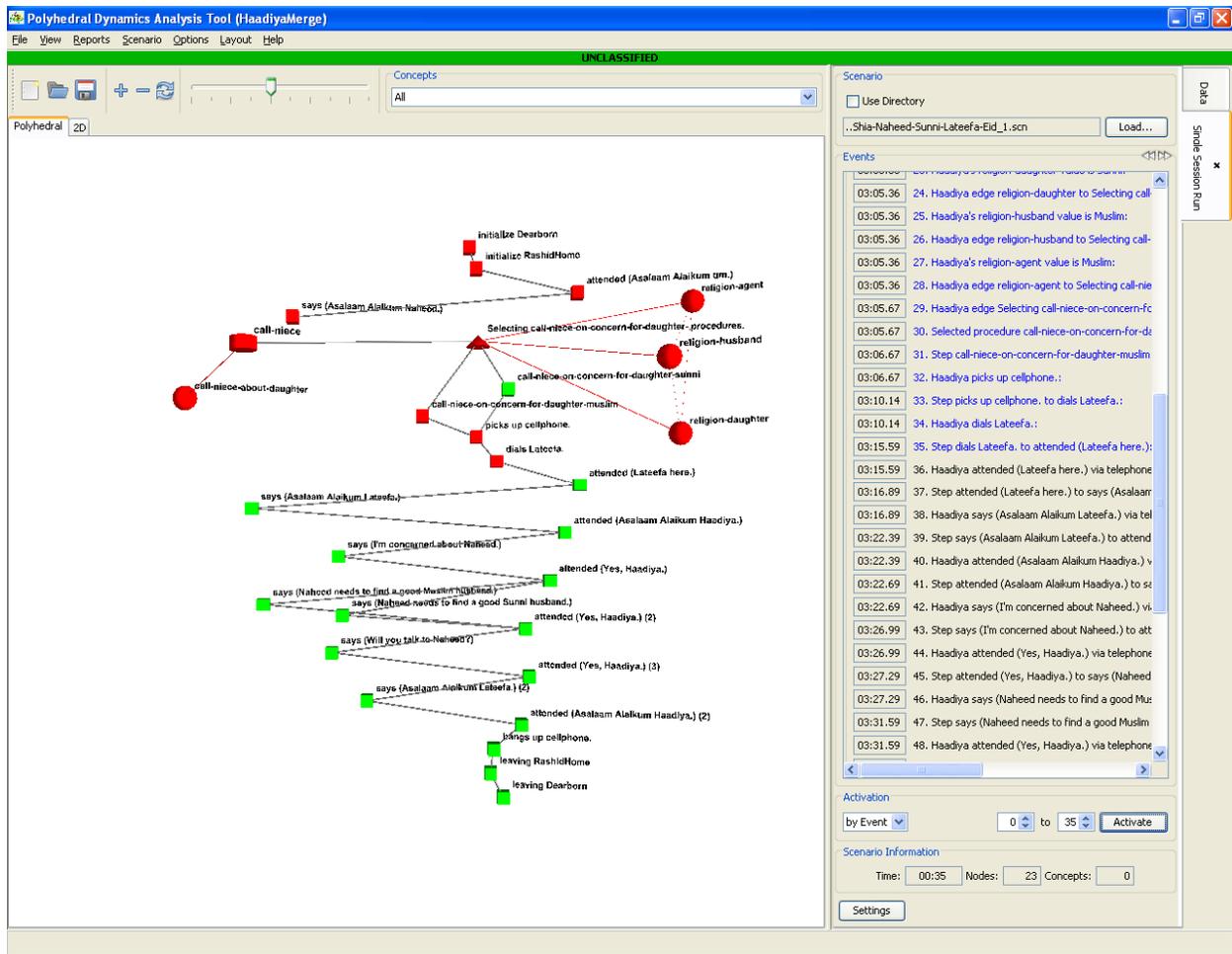


Figure 11: Single Session Comparison

Dual Session Comparison: Dual Session Comparison is similar to Single Session Run, but uses two scenarios instead of one. Like Single Session Comparison, the scenarios are loaded, the ranges are selected and the graph is updated to display the activations. Both scenarios are displayed in the same graph for comparison. The activations are in different colors, and if they both activate the same element, a third color is used to represent this. The ranges can be set to match, or each scenario can represent a different range (see Figure 12).

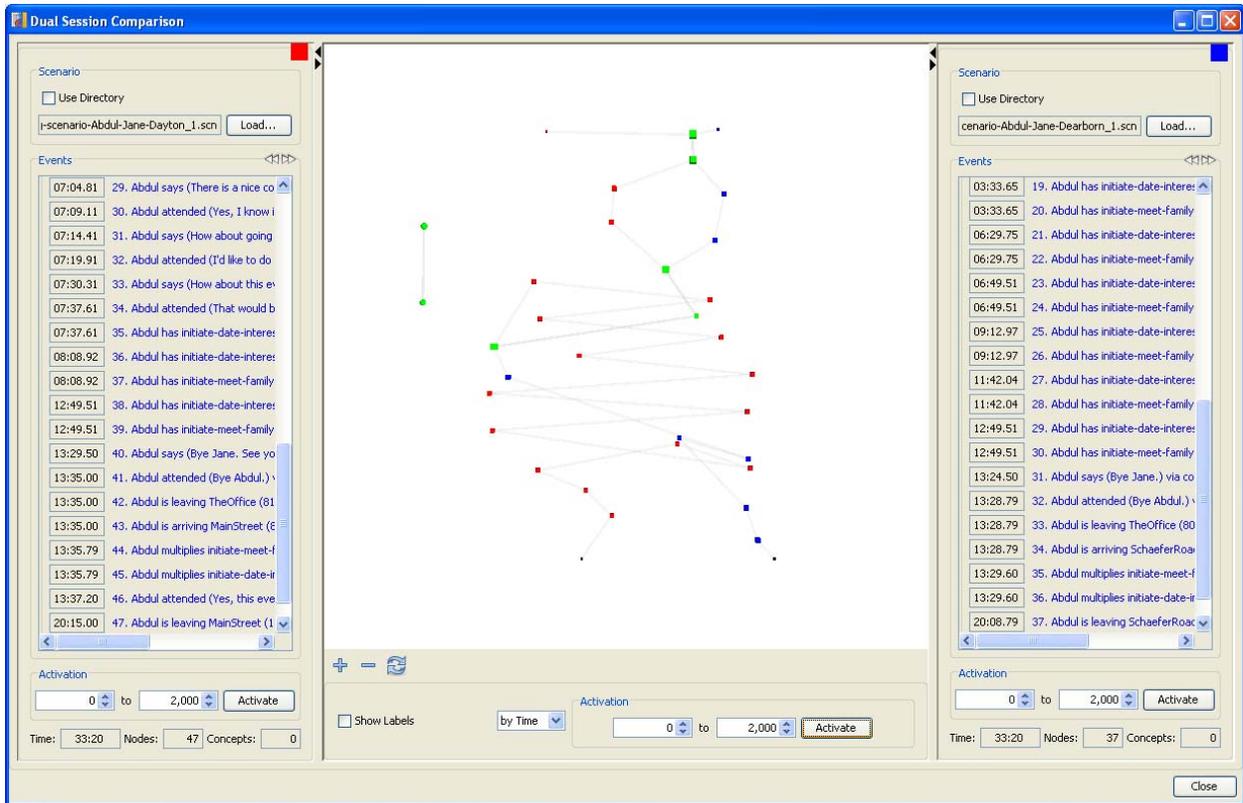


Figure 12: Dual Session Comparison

Dual Side-By-Side: Dual Side-By-Side is very much the same as Dual Session Comparison, except that each scenario displays its own graph (see Figure 13).

4.2.8. Graph Manipulation

The graph display can be manipulated for better use. Each element of the graph (concepts, edges, and nodes) can be filtered out. The graph nodes can be moved around, and the graph itself can be rotated to bring other nodes to the front. Nodes can be centered on the screen, and node size can be adjusted with a slider on the tool bar. The graph nodes can be the concept nodes, or be reduced to represent the concepts themselves, linking to other concepts that share a similar concept node.

4.3 Graph Structure

PDAT can display two structure types for analysis: Conceptual and Scenario Based.

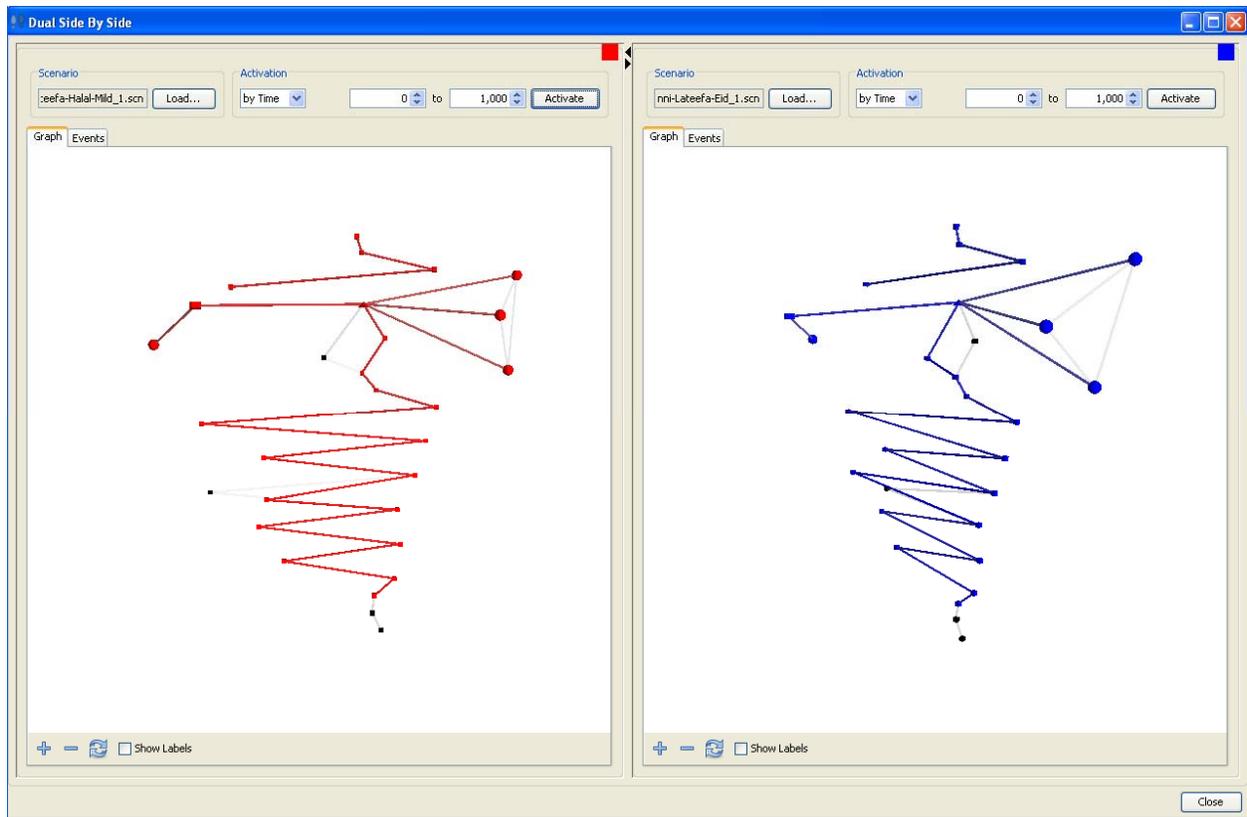


Figure 13: Dual Side-By-Side

4.3.1. Conceptual Graph Structure

The conceptual graph structure is generated from matrix (.matrix) files, or subset Node (.nodes) and Concept (.concepts) files. The graph is structured to have a number of nodes that represent features or attributes, as well as concepts that create the edges between the nodes. Nodes have no predefined locations on the graph, but depending on the strength of the edges between nodes, the nodes may be bunched together on the graph.

In the simple example displayed in Figure 14, there are two concepts: polar bear and brown bear. The polar bear has white hair color (X4), very large body size (X13), long hair (X3), no tail (X14), and found in the zoo (X19). The brown bear has brown hair (X9), very large body size (X13), long hair (X3) and no tail (X14). Attributes that are exclusively polar bear are grouped together, and shared attributes are also grouped together.

Some assumptions might be made looking at this simple graph. The edges between X3, X13 and X14 may be stronger based on the fact that two concepts group those three together. When presented with a concept that is very large and has long hair, the graph may automatically assume the concept has no tail.



Figure 14: Conceptual Graph

4.3.2. Scenario-Based Graph Structure

The scenario-based graph structure is built from a scenario (.scn) file. The scenario based graph is different from the conceptual graph in multiple ways. First, there are multiple types of nodes on the graph. Second, the location on the graph is somewhat defined by the type of node, and the time in the scenario when the node was generated. Third, all edges in the graph are created equally (see Figure 15).

Node Types: There are five types of nodes in the scenario based graph: Interest, Trigger, External Event, Decision, and Disposition.

Interest: Interest nodes are spherical nodes, located on the left side of the graph (see Figure 16). These nodes represent an interest of a subject that will rise and fall based on the scenario. Interest nodes have thresholds, that when met, will fire off the trigger.

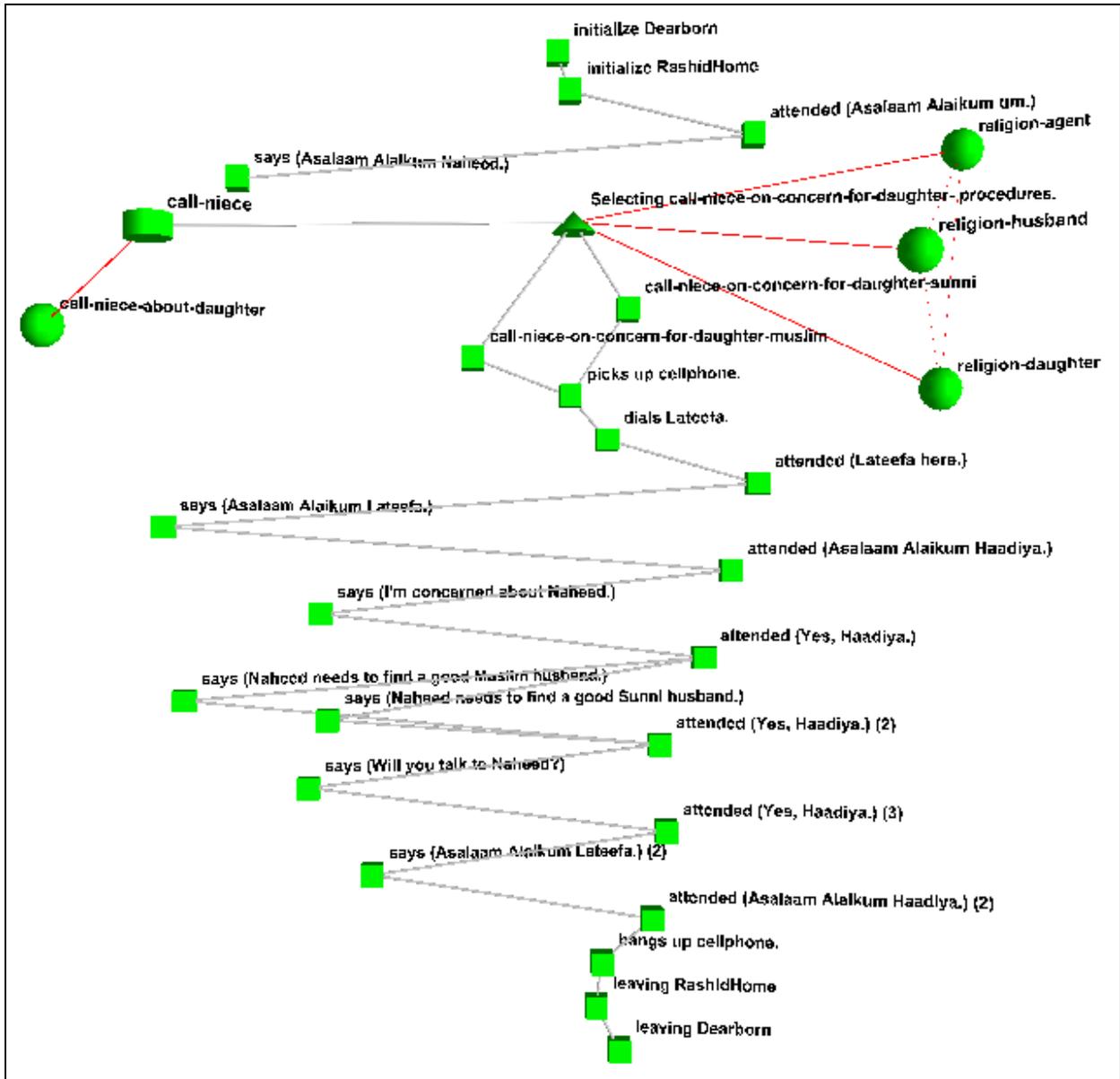


Figure 15: Scenario-Based Graph

Trigger: Trigger nodes are cylindrical and are located on the left side of the graph, just to the right of the interest nodes. When an interest crosses a threshold, the trigger node is activated. In Figure 16, the subject reaches the threshold for calling her niece. This is the point at which the interest is so high, it triggers an external action. The trigger links to the external event that occurs because of that interest.

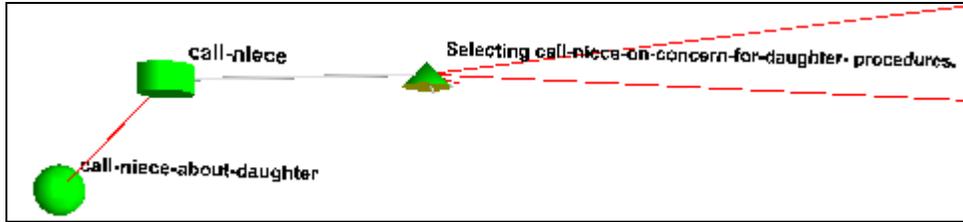


Figure 16: Interest Trigger

External Event: External events are cubes by default (see Figure 17), except in the special case of a decision event. External events provide flow for the scenario. They are lined up sequentially from top to bottom, with the earliest events at the top. External events take up the middle sections of the graph. Within the middle section events that may be triggered are on the left side, events that may affect disposition are on the right side, and everything else falls in between. Typical external events include leaving and arriving at a location, conversation (saying and hearing), and picking up the phone.

Decision: Decision nodes are special external nodes. Decision nodes are conical and are located in the middle of the graph. A decision node on the graph represents a decision point, in which there are multiple choices for action. The decision is influenced by the subject's disposition. The decision node is linked to multiple other external nodes, but only one external node will be selected. In Figure 17, the subject is going to call her niece about her concern for her daughter. There are two choices: calling about her daughter's Muslim acquaintance, or her Sunni acquaintance.

Disposition: Disposition nodes are spherical nodes on the right side of the graph. Disposition nodes are attributes used by decision nodes to evaluate a situation and make a choice. In Figure 17, there are three dispositions: the subject's religion, the subject's daughter's potential husband's religion, and the subject's daughter's religion. They all influence the decision. If every religion is the same, the conversation may be more pleasant. If the potential husband's religion does not match their religion, the conversation may be more problematic.

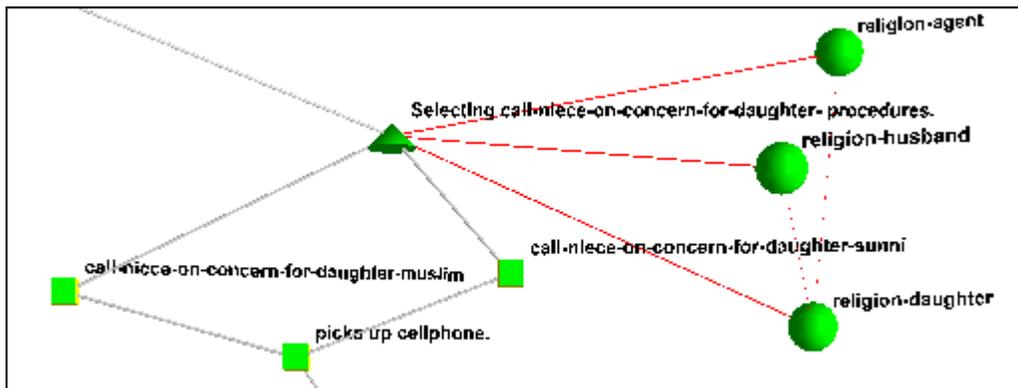


Figure 17: Decision

4.4 Data Collection Selection

Data collection is split into two phases. In Phase one, data was collected in English, using Qualtrics survey software that was adapted to our needs. In Phase two, data will be collected in the respondent's native language, and English. Data will be collected using an Android application that was developed and installed onto a Samsung tablet.

4.4.1. Data Collection

In addition to creating visualization methods, techniques were developed to collect data to populate future knowledge models. Data were gathered from subjects who self-identified as one of three nationalities; American, Chinese or Indian. The data were broken into two subsets. The first subset was data that were retrieved from these subjects when they read and answered experiment questions in English. The second subset of data came from additional subjects as they answered a subset of the original questions in their native language and English. It was discovered that a majority of the subjects (including the Chinese and Indian subjects) type in English, when using a computer. Therefore, in order to get results in which the subjects were not thinking in English, a second subset of data would be collected by having the subjects hand write their answers on a tablet computer. Hence, two different data collection tools would be necessary.

4.4.2. Criteria for English data collection tool

A variety of criteria was used to select the software used for data collection. They included:

- a. Response count
- b. Data format
- c. Languages supported
- d. Software documentation
- e. Extensible
- f. Clean display
- g. Software support
- h. Cost

The data collection software was evaluated against these criteria. Any tool that didn't meet the first three criteria was thrown out, as it would be unable to accomplish the tasks required. Once a tool met these main requirements, it was tested on how well it met the other requirements and how difficult it would be to overcome any deficiencies.

Data Collection Options: A total of 24 data collection tools were evaluated for potential use. Seven were selected as candidates for final analysis, and 17 options were discarded for failing to meet a required specification. The 17 discarded tools were Digivey, Fluid Surveys, Inquisite, Survey Systems, Survey Pro 5, Global Park, Cvent, Blue/Surveys, Checkbox Survey Online, Checkbox Survey Server, Cogix ViewsFlash, Feedback Server On-Demand, Feedback Server On-Site, KeySurvey, phpESP, StatPac, and Survey Said. The seven selected tools are detailed in the following sections.

Lime Surveys: Lime Surveys is a free on-line survey package with unlimited surveys and responses. It has adequate data output formats and can prevent someone from taking the survey

multiple times. The tool is basic, but it may not cover all the languages needed for this research. The question types are limited.

See URL: <http://www.limesurvey.org>

Qualtrics: Qualtrics is an on-line survey that provides a large number of features including an extensive question type selection, language support, and an on-line library of tips and support. Even with purchase, there is a limited number of responses allowed, and only one user may edit the surveys.

See URL: <http://www.qualtrics.com/survey-software/>

Snap Surveys: Snap Surveys provides two options, a survey hosted on-line or a downloadable survey to host on your own server. It covers the languages needed and has acceptable output formats. The cost is high in comparison to the quality of the surveys, and the software does not provide any time controls.

See URL: <http://www.snapsurveys.com/us/>

Survey Gizmo: Survey Gizmo provides usable data output formats along with unlimited surveys and language support. Timing controls are lacking, and it does not provide translations.

See URL: <http://www.surveygizmo.com/>

Survey Monkey Pro: Survey Monkey Pro is an application that was recommended by co-workers who currently use it. It provides 1000 responses at a low monthly cost. It is also easy to use. It provides great service for the basic features, but is not extensible enough to provide the controls needed for the research.

See URL: <http://www.surveymonkey.com>

Toluna: Toluna is a company which has a basic survey tool with few features. Toluna may be a great resource in panel procurement, if needed in the future. The lack of features makes this tool unacceptable for our use.

See URL: <http://www.toluna-group.com>

Zoomerang: Zoomerang provides unlimited surveys and responses. It can display images, but not video. It can be used in a kiosk version and provides adequate data output formats. It cannot control time. It also does not support all the languages we need, and only allows for one user.

See URL: <http://www.zoomerang.com/online-surveys/>

4.4.3. Phase One Collection Tool

Qualtrics was selected as the data collection tool for Phase I. The requirements for the experiment required a number of complex features that eliminated most of the other tools. These requirements included the need to record data entry and exit times, auto-flipping to the next question after a predefined amount of time, and disabling forward and backward buttons.

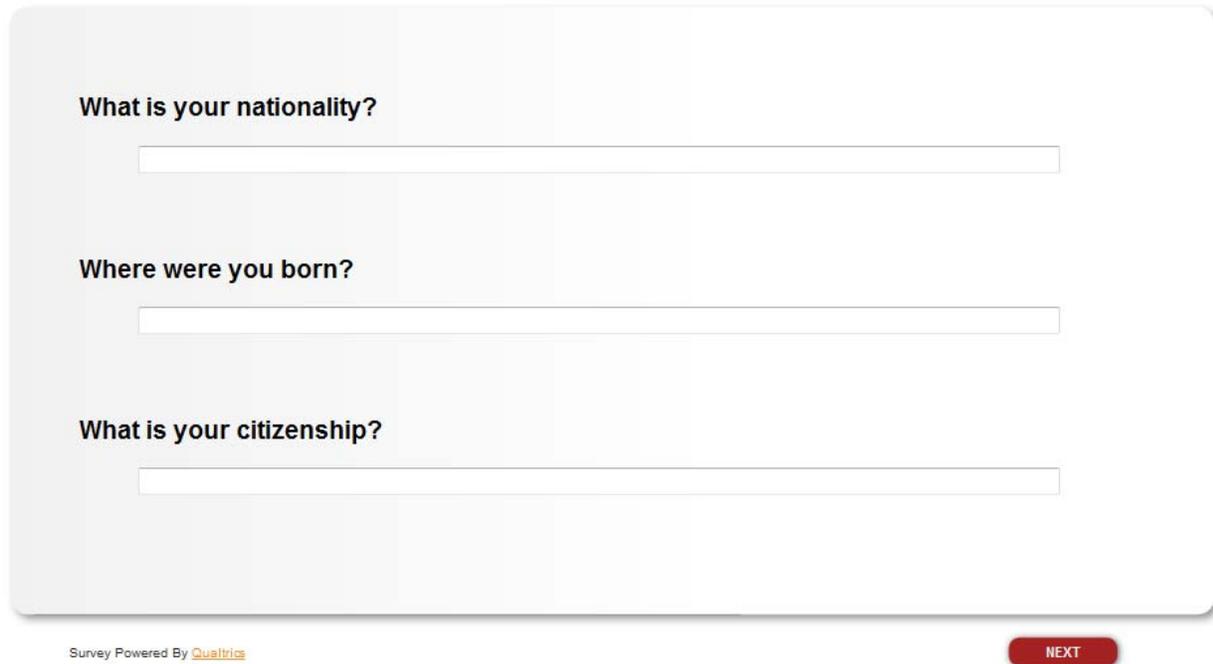
Category Survey Data Collection: The survey is composed of 15 demographic questions and 60 category questions. The survey will take around 75 minutes.

Category Questions: Category questions are timed, and entry times are recorded. Users have 60 seconds to answer each question, and there is enough space for up to 20 categories under each question (see Figure 18). Between each category question is a black screen that lasts for 5 seconds.

The image shows a survey question interface. At the top, the text reads: "Please list some examples of the category: musical instruments". Below this text is a vertical list of 20 numbered input fields, each containing a small white rectangular box for text entry. The numbers 1 through 20 are aligned to the left of each box. At the bottom left of the interface, there is a small text label: "Survey Powered By Qualtrics".

Figure 18: Category Question Display

Demographic Questions: Demographic questions are untimed, and the user can answer them at their own pace (see Figure 19).



What is your nationality?

Where were you born?

What is your citizenship?

Survey Powered By [Qualtrics](#) NEXT

Figure 19: Demographic Question Display

4.4.4. Phase Two Data Collection Tool

A Samsung Galaxy tablet was selected for Phase II data collection. A category survey application was developed and installed onto the tablet in three different languages. The languages were English, Chinese and Hindi.

Category Survey Data Collection: The Category Survey is composed of 15 demographic questions and 20 category questions. The survey takes approximately 40 minutes to complete.

Category Questions: Category questions are timed, and entry times are recorded. Users are given 60 seconds to complete each question, with space for up to 20 categories under each question (see Figure 20). Between each category question, a black screen is presented that lasts for 5 seconds.

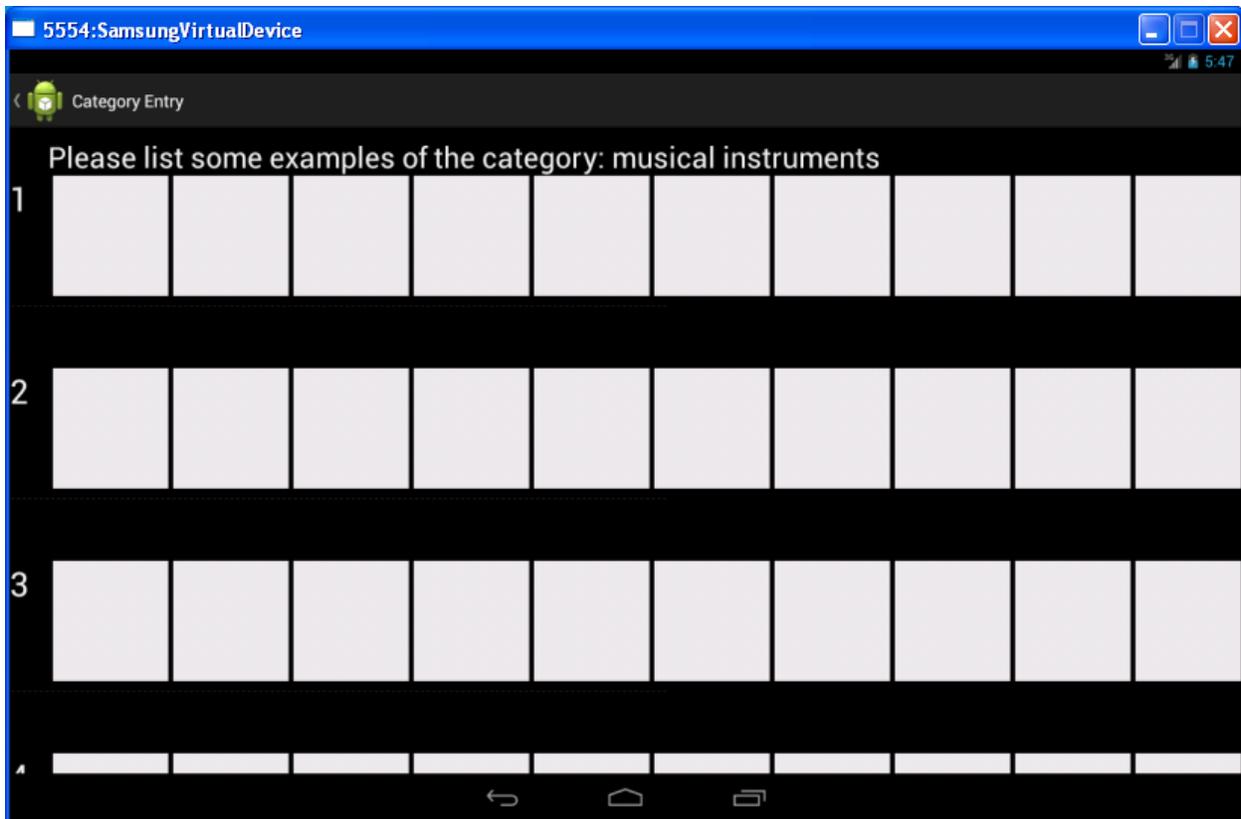


Figure 20: Tablet Category Question Display

Demographic Questions: Demographic questions are untimed, and the user can answer them at their own pace (see Figure 21).

4.5 Survey Results

Each survey generates results that are accumulated and distributed. The format of these results is different for the two survey collections tools.

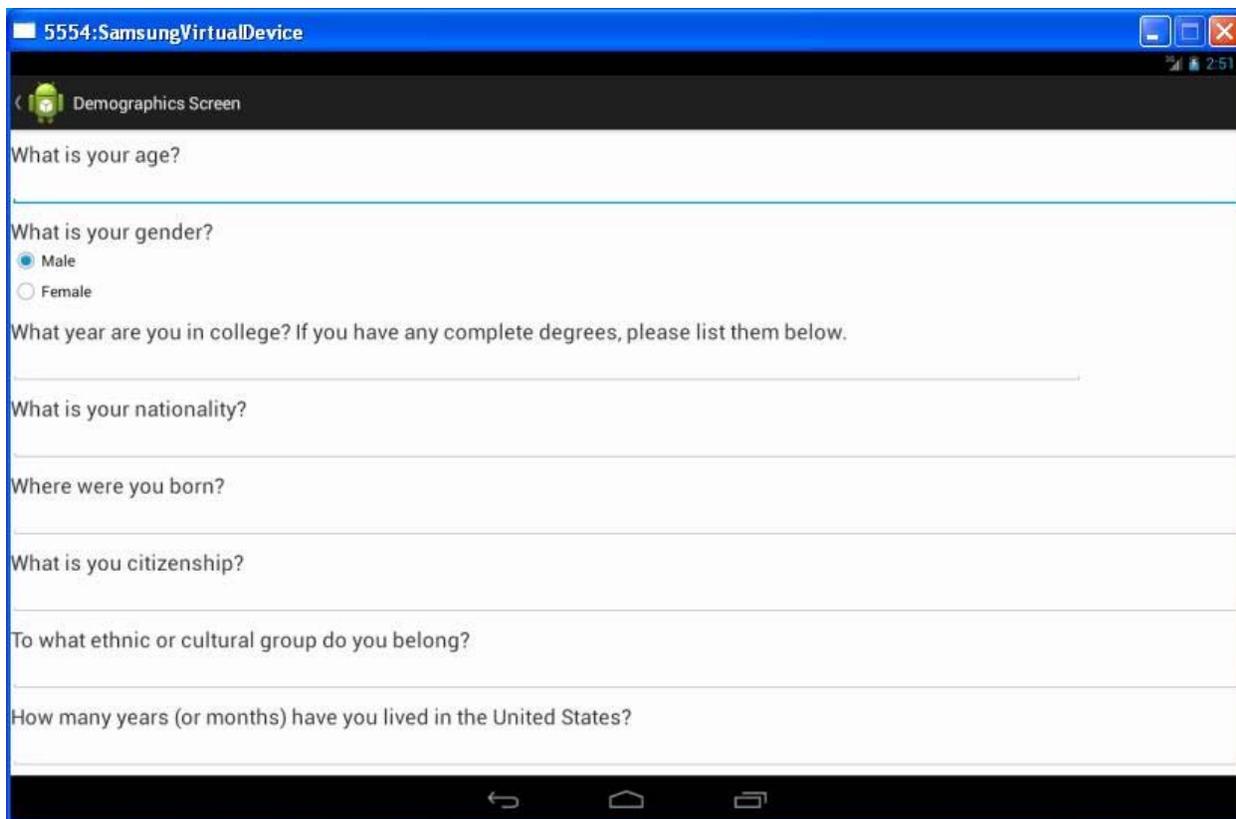


Figure 21: Tablet Demographic Question Display

4.5.1. Qualtrics Survey Results

Qualtrics allows for a number of different formats for downloading its results. The format that was selected for this research was comma separated values (CSV). Due to the CSV format, Microsoft Excel was used to browse through the results. In addition, this format enables data to be easily loaded into other applications for deeper analysis.

In addition to listing every question used in the survey, the results file lists every other field that is displayed or not used in the survey. The automatically generated data, such as the subject Identification (ID) and survey start and end times are located at the beginning of the results file.

The majority of the file consists of the answers given for each question (up to 20 per question) as well as the entry and exit times of those answers. From these three pieces of information, a larger data object can be stitched together. This data object, called the Survey Answer, contains the actual answer, the time the answer was started, the time the answer was ended, the amount of time it took to create the answer, and the order in which the answer was given. Since these pieces are scattered in the results file or are calculated, a Survey Analysis tool was developed for consolidating the data.

4.5.2. Tablet Survey Results

The tablet survey obtains inputs from the subjects through an interface that lets the users write each letter instead of typing it (except for the demographic information which is still typed using the tablet's digital keyboard). These written results can be in multiple languages.

Each subject has a folder that stores their survey results. Within that folder are two types of data files: a properties file and image files. The properties file contains the entry/exit times, demographic and other hard data. The image files are snapshots of every letter entered by the subjects. The filename indicates the question # / answer # / letter # it represents.

Before the results can be sent to the Survey Analysis tool, the letter files need to be interpreted by a language analyst. The results from the tablet surveys are sent through the Survey Interpreter tool.

4.6 Survey Interpreter

The Survey Interpreter tool is used for converting subjects' written results into a format that can be used in the Survey Analysis tool (see Figure 22).

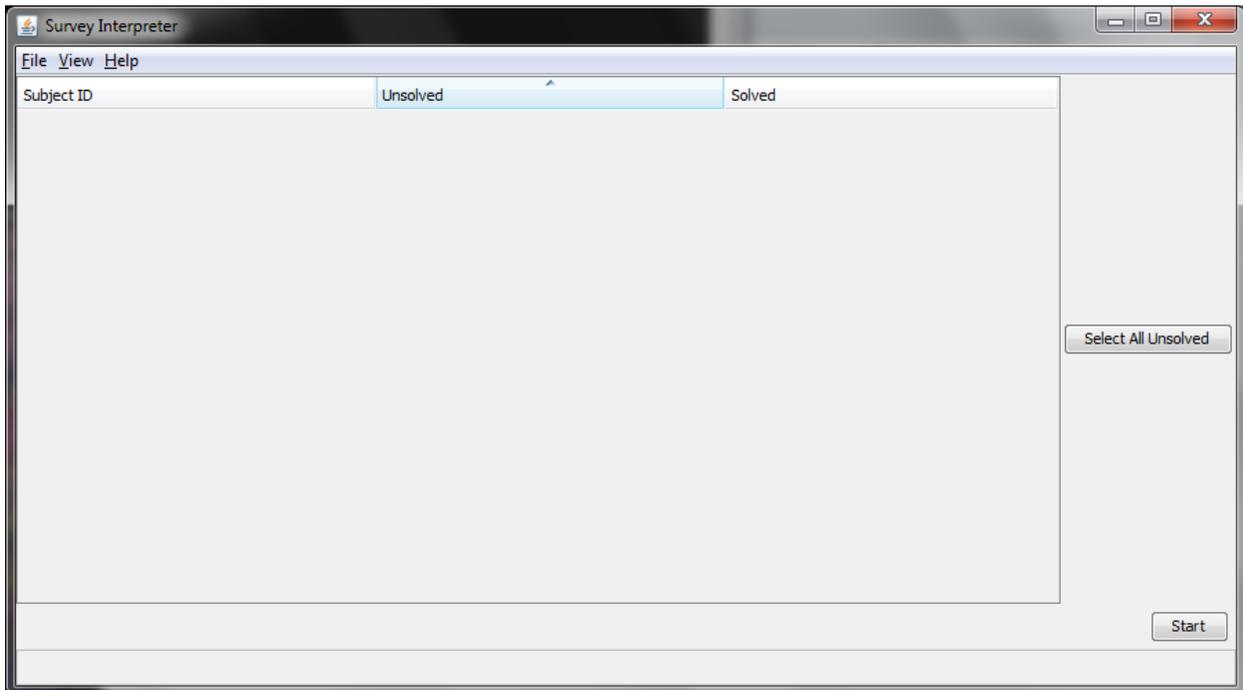


Figure 22: Survey Interpreter Main Screen

4.6.1. Menu

The menu (Figure 23) has three menu items: File, View, and Help. The File menu contains the file load/save actions for the application. The View menu has viewing selections for the application. The Help menu contains the various help features.

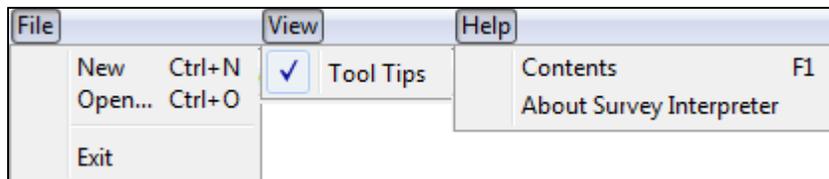


Figure 23: Survey Interpreter Menu

4.6.2. Main Screen

The main screen of Survey Interpreter contains a table of all the subjects and number of solved and unsolved answers (see Figure 24). The analyst can select a specific subject by clicking their row in the table. Otherwise pressing the “Select All Unsolved” button will select all subjects. Clicking the “Start” button will bring up the interpret panels.

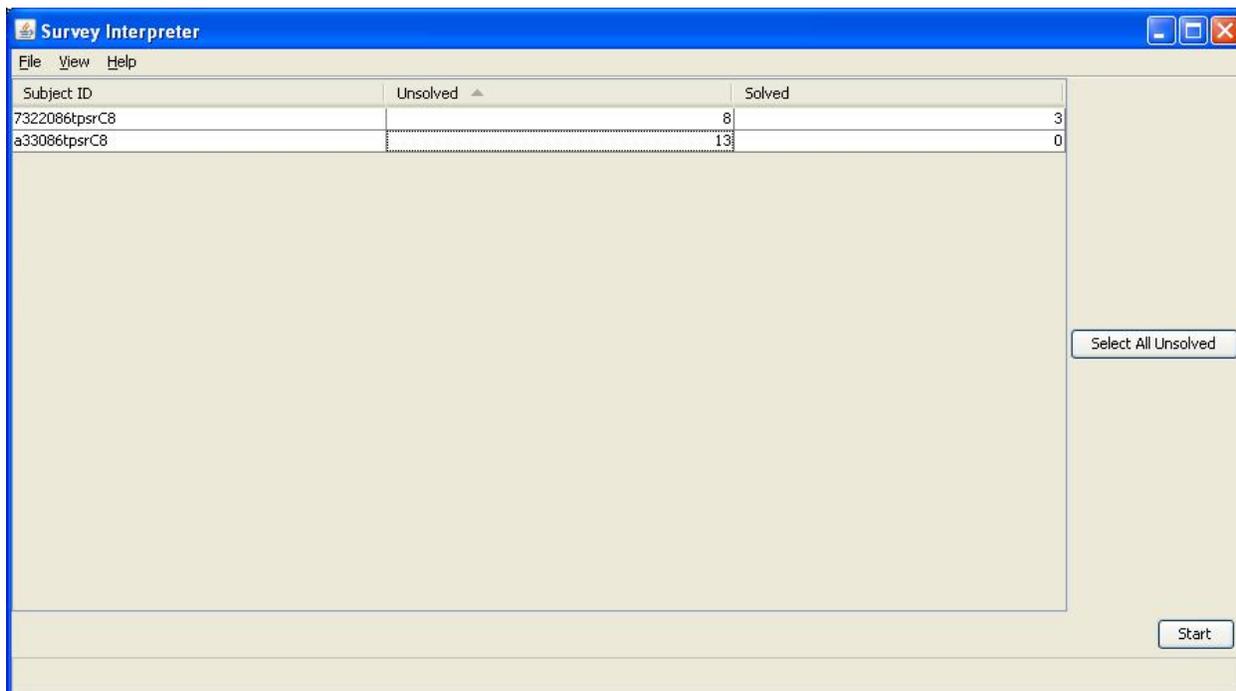


Figure 24: Loaded Results in Survey Interpreter

4.6.3. Interpret Panel

The Interpret Panel displays answers for survey questions (see Figure 25). The answers are the images taken from the tablet that are placed in order for the analyst. The survey question can be selected at the top of the screen, and the unsolved answers are listed. If the analyst can interpret the answers, they type them into the text field. If there is no answer, they can check the box “No Answer”, which indicates that the analyst has reviewed that answer and found that no answer was provided. When satisfied with their interpretations, the analyst may click “Save and Return” to return to the main screen, or “Save and Next” to move to the next question with unsolved answers.

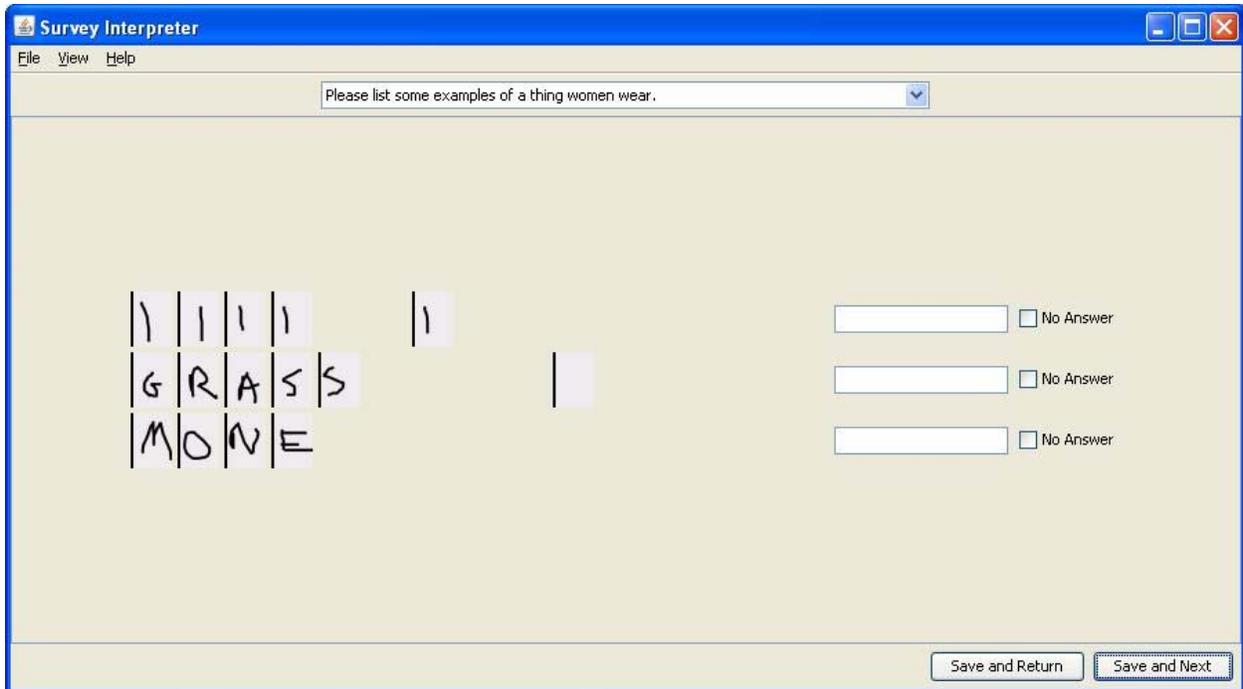


Figure 25: Interpret Screen in Survey Interpreter

4.7 Survey Analysis

Once all the data is collected, it is still in a raw format and not very useful. In order to make the data more useful, it must be condensed into groups of data that can be used for display in sortable tables and reports. Survey Analysis allows the analyst to take the raw data, combine similar elements of data and generate reports (see Figure 26).

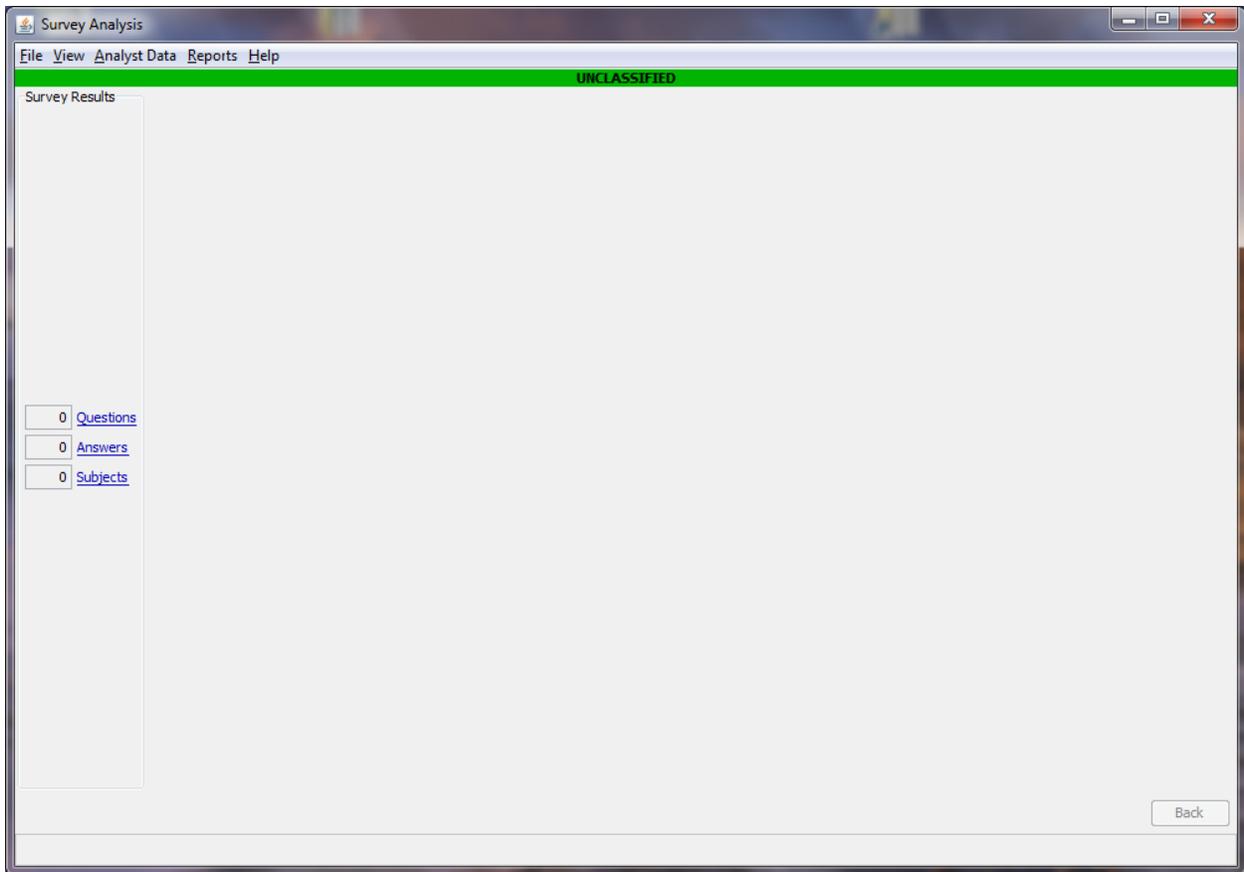


Figure 26: Survey Analysis Main Screen

4.7.1. Menu

There are five menu items in Survey Analysis (see Figure 27). File contains the file load and save features, View has the tool tips option, Analyst Data lists all the tools for data arranging and manipulation, Reports lists general reports that can be generated, and Helps lists the help features for the application.

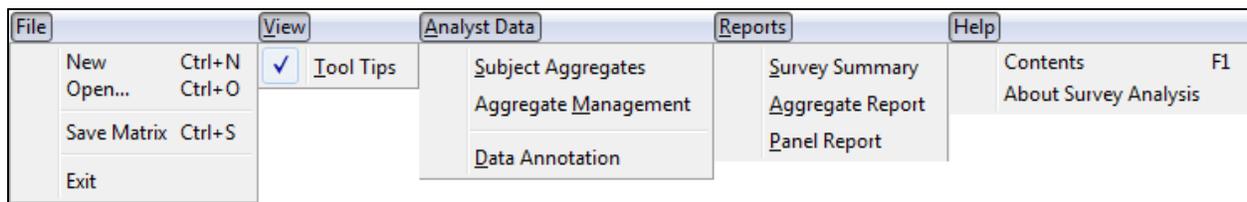


Figure 27: Survey Analysis Menu

4.7.2. Loading and Viewing Survey Results

In order to use the application, the results from the survey must be loaded.

Loading Survey Results: The results from Qualtrics that were saved can be loaded into the application. To load the data, click on the Open menu item in the File menu (see Figure 28).

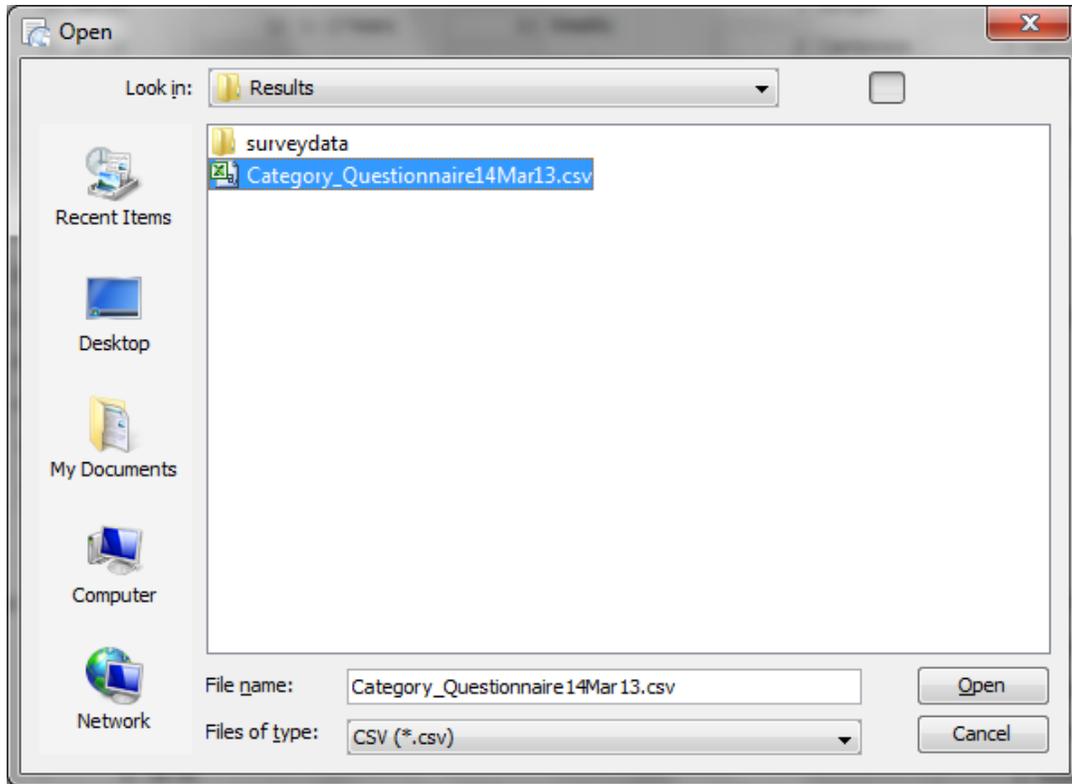


Figure 28: Survey Analysis Load Screen

Loaded Survey Results: The survey results are loaded into the application and the screen displays a summary of the results on the left side of the screen. The summary includes the number of questions, answers and subjects (see Figure 29).

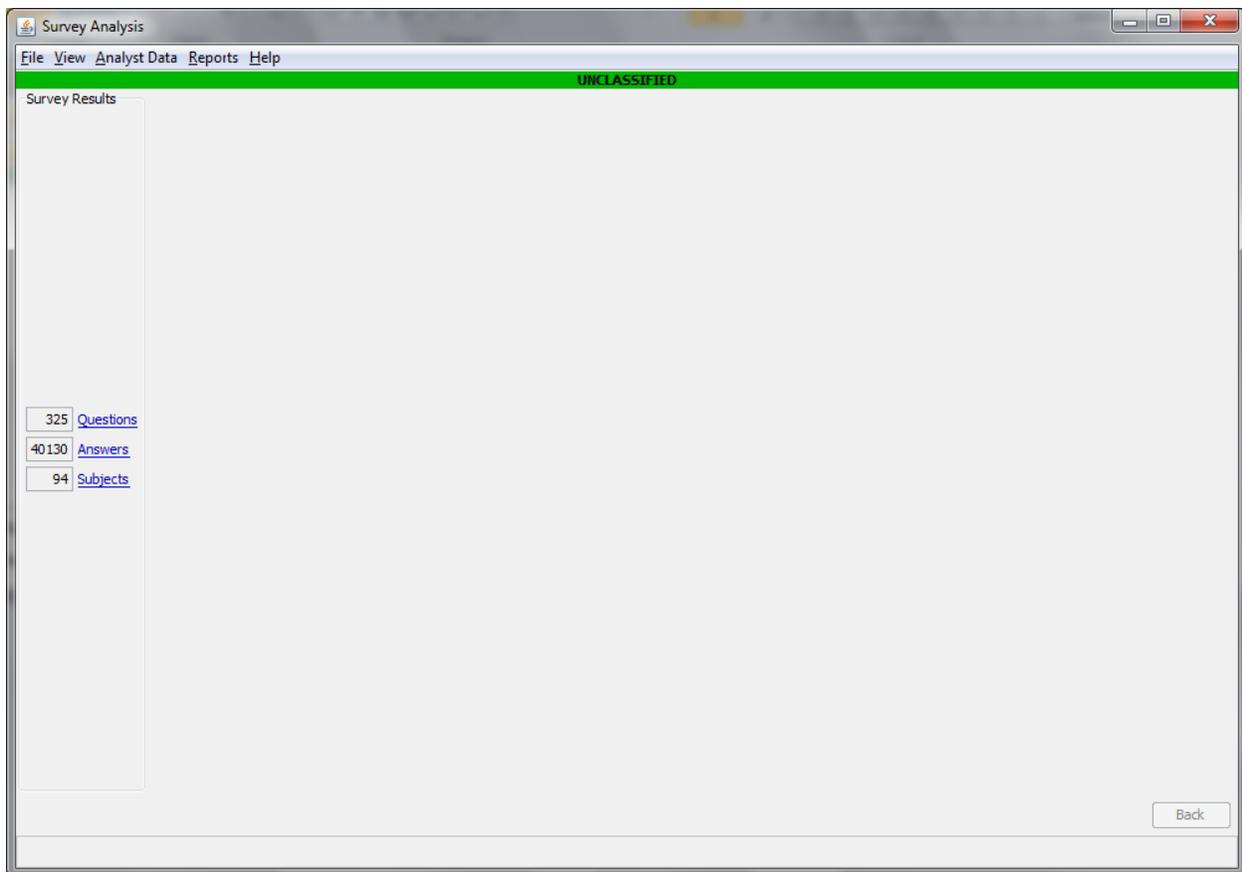


Figure 29: Survey Analysis with Results

Questions: All of the questions in the survey are loaded into the application. The data and reporting are broken up by all questions and by single questions. The Question Table lists every question in the survey, and the Question Information page displays data and reports for a single question.

Question Table: The Question Table is accessible by clicking the “Questions” link on the left side of the screen (see Figure 30). The table lists the survey questions by Question Number, text, category and program generated report type. Type is used to order the questions in the table in a way that shows the most important question at the top of the table.

Question #	Text	Category	Type
Q1	Please list some examples of the characteristics of a good spouse?	good spouse	AQ0001
Q2	Please list some examples of the characteristics of a good life?	good life	AQ0002
Q3	Please list some examples of the characteristics of a good career?	good career	AQ0003
Q4	Please list examples of an item associated with dining out at a restaurant?	dining out at a restaurant	AQ0004
Q5	Please list examples of an item associated with going for a walk in the park?	going for a walk in the park	AQ0005
Q6	Please list examples of an item associated with preparing for sleep at night?	preparing for sleep at night	AQ0006
Q7	Please list some examples of a liquid.	liquid	AQ0007
Q8	Please list some examples of a thing women wear.	thing women wear	AQ0008
Q9	Please list some examples of a thing that flies.	thing that flies	AQ0009
Q10	Please list some examples of the characteristics of a good citizen?	good citizen	AQ0010
Q11	Please list some examples of the characteristics of a good leader?	good leader	AQ0011
Q12	Please list some examples of the characteristics of a good way to respond in a conflict?	good way to respond in a conflict	AQ0012
Q13	Please list some examples of games?	games	AQ0013
Q14	Please list some examples of medicine?	medicine	AQ0014
Q15	Please list examples of an item associated with working out at the gym?	working out at the gym	AQ0015
Q16	Please list examples of an item associated with going to a fair?	going to a fair	AQ0016
Q17	Please list examples of an item associated with a party?	a party	AQ0017
Q18	Please list examples of an item associated with taking an airplane trip?	taking an airplane trip	AQ0018
Q19	Please list some examples of a thing that makes noise.	thing that makes noise	AQ0019
Q20	Please list some examples of a thing that is green.	thing that is green	AQ0020
Q21	Please list some examples of a thing made of wood.	thing made of wood	AQ0021
Q23	Please list some examples of an item to take on a vacation.	an item to take on a vacation	AQ0023
Q24	Please list some examples of an item not to eat while dieting.	an item not to eat while dieting	AQ0024
Q25	Please list some examples of a picnic item.	picnic item	AQ0025
Q26	Please list some examples of a thing to do on a holiday.	thing to do on a holiday	AQ0026
Q28	Please list some examples of a food to eat when sick.	food to eat when sick	AQ0028
Q29	Please list some examples of a thing that make someplace desirable to live.	thing that make someplace desirable to live	AQ0029
Q30	Please list some examples of a thing to remove from a burning house.	thing to remove from a burning house	AQ0030
Q31	Please list some examples of a weapon.	weapon	AQ0031
Q33	Please list some examples of a guide.	guide	AQ0033

Figure 30: Question Table

Question Information: The Question Information screen can be found by selecting a question ID in the Question Table (see Figure 31). The screen is broken up into four sections: Question Information Survey Results Summary, Reports, Edit, and Answer table.

Question Information Summary: The summary panel at the top of the screen displays the question number, the sub question number, code (if applicable), category and text of the question.

Reports: The reports panel lists two reports for generation: Summary and Connectivity.

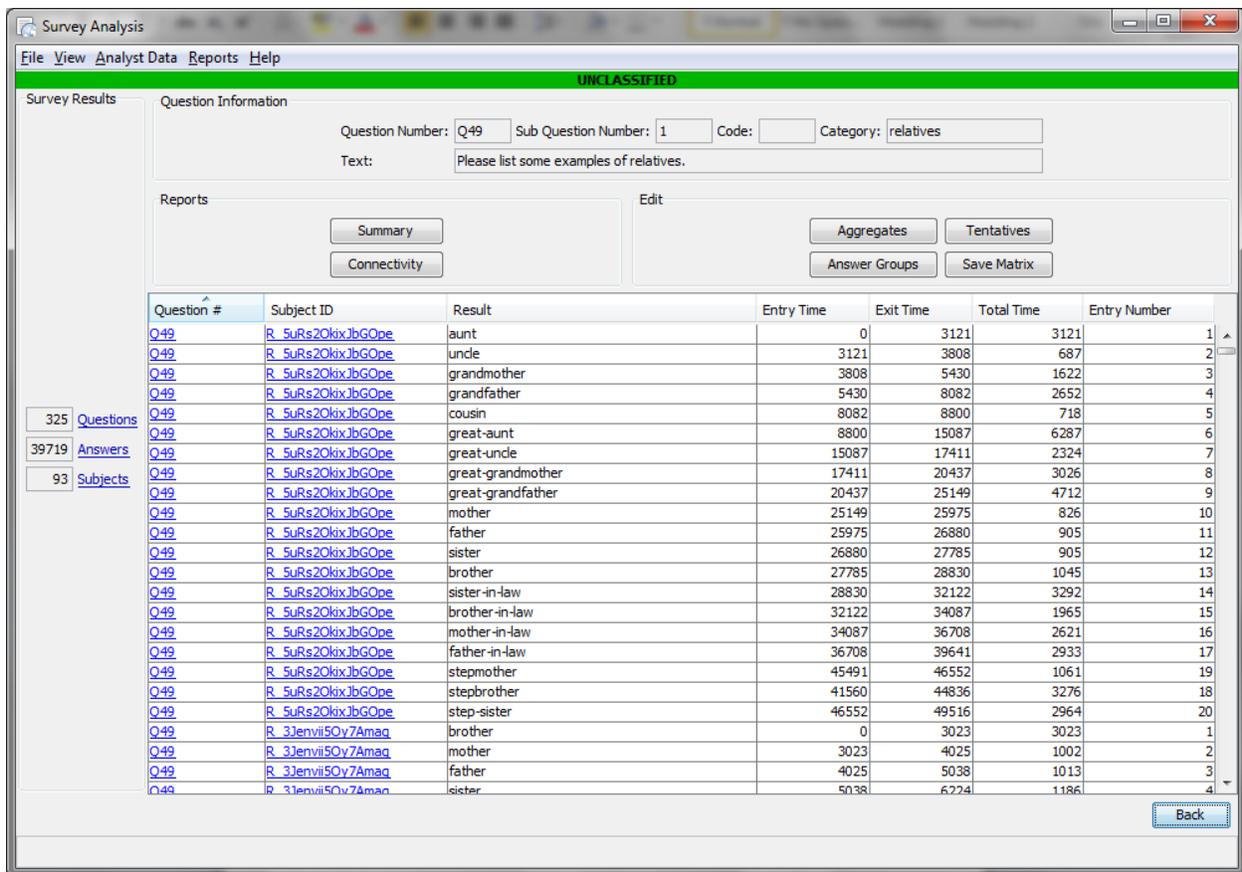


Figure 31: Question Information

Summary Report: The summary report takes all the answers to the question and calculates a number of pieces of information that summarizes the results (see Figure 32). The top left corner lists the overall statistics of the question. Answers count is the number of total answers for the question. Unique answers is the count of different answers provided (aggregated answers count as one). Average per subject shows how many answers were provided for the question per subject on average. The average answer time is the average time it took for each one of the answers in this question.

The bottom section of the report lists all the unique answers for the question, and the individual statistics for each answer. Along with the number and average time, as in the summary statistics, it also lists average location, first count, and the ranges for time and location.

Results can be filtered by answer location and by subject demographic, such as nationality and age. This provides the ability to view statistics for what was the first answer for Americans aged 20 – 22 years old.

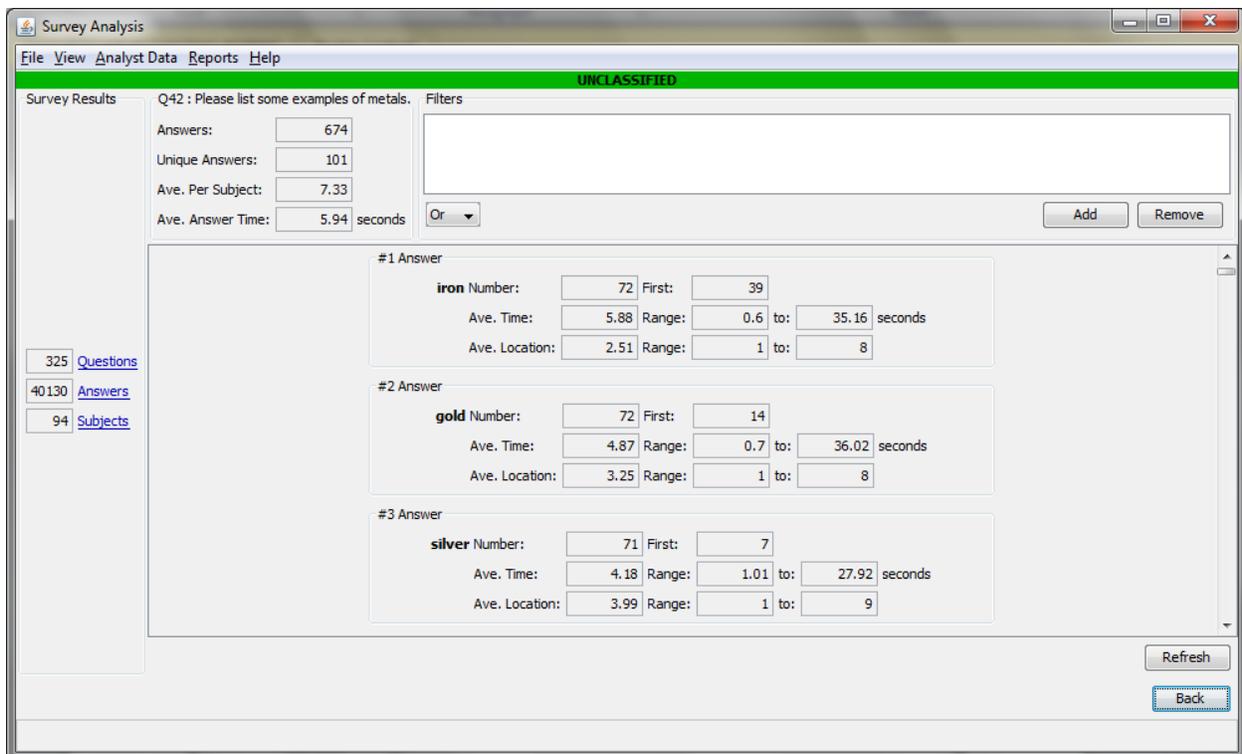


Figure 32: Summary Report

Connectivity Report: The Connectivity Report shows the relationship between answers for a question. Each unique answer for the question is listed in a descending order with the most frequent answer listed first and the least answered listed last. A target answer is an answer for which we are analyzing, and the connected answers are all the other answers for the question for the subjects who had the target answer. Each target answer has a table which lists every connected answer. Using the example in Figure 33, the first target answer is “Aunt”, and the table lists every answer provided by the subjects that responded “Aunt”.

The table has four columns: Answer, Match Count, Average Distance and Significance. The border shows the target answer and the number of subjects that provided the answer. Match count is the number of times that a subject answered the target answer and the answer in the first column of the table. In this example, of the 78 subjects that answered “Aunt”, 74 also answered “Uncle”. Average distance is the difference in answer location on average. For example, a subject could have these answers:

- Mother
- Father
- Aunt
- Uncle

Distance is calculated by the number of lines between answers. The distance between “Aunt” and both “Father” and “Uncle” is 1, and the distance between “Aunt” and “Mother” is two. The average distance of 1.08 indicates that most of the time “aunt” and “uncle” were answered next to each other.

Significance is a combination of Match Count and Average Distance. ($\text{Average Distance} / (\text{Match Count} / \text{Target Answer Count})$). This provides a better indicator of connectivity because it eliminates the extreme outliers of the other two indicators. If “maternal grandmother” was answered right after “aunt”, but they were only answered together once, their Average Distance would be a perfect 1, but wouldn’t be a strong indicator. Likewise in this example, grandfather was an average distance of 3.77 from “aunt”, but is listed below “sister” whose distance is 5.08 because “sister” was answered one more time than “grandfather” (sister matched 68%, grandfather matched 67%).

Edit: The edit panel has four options: Aggregates, Tentatives, Answer Groups and Save Matrix.

Question Answer Aggregates: The aggregate function allows the user to combine subject answers into an umbrella answer called the aggregate answer. The basic idea is that when generating reports about the answer “brother,” the report should include case sensitive answer “Brother,” the plural version “brothers,” the misspelled versions such as “Borther” and interrupted answers such as “brothe.”

There are other reasons to aggregate. For other categories, such as Metals, aggregate answers can include the element name “silver” and the chemical symbol “Ag.” In the category Things in a Park, aggregate answer “air” could include “fresh air,” “clean air,” “oxygen,” and “air to breathe.”

The top section of the screen displays the created aggregates and the bottom section lists the answers that are still singletons and not combined with other answers (see Figure 34). The table above lists the aggregate answer, the answers it covers, and the tentative answers that may be covered in the future. The tentative feature is for situations in which the analyst is unsure if the answer is part of the aggregate, but wants to mark it for further consideration. It is not included with the aggregate answer in the generated reports.

The application can generate aggregates automatically. Capitalization and plurality are mostly caught, but misspellings or familiar forms of words, such as “auntie” for “aunt” are not caught.

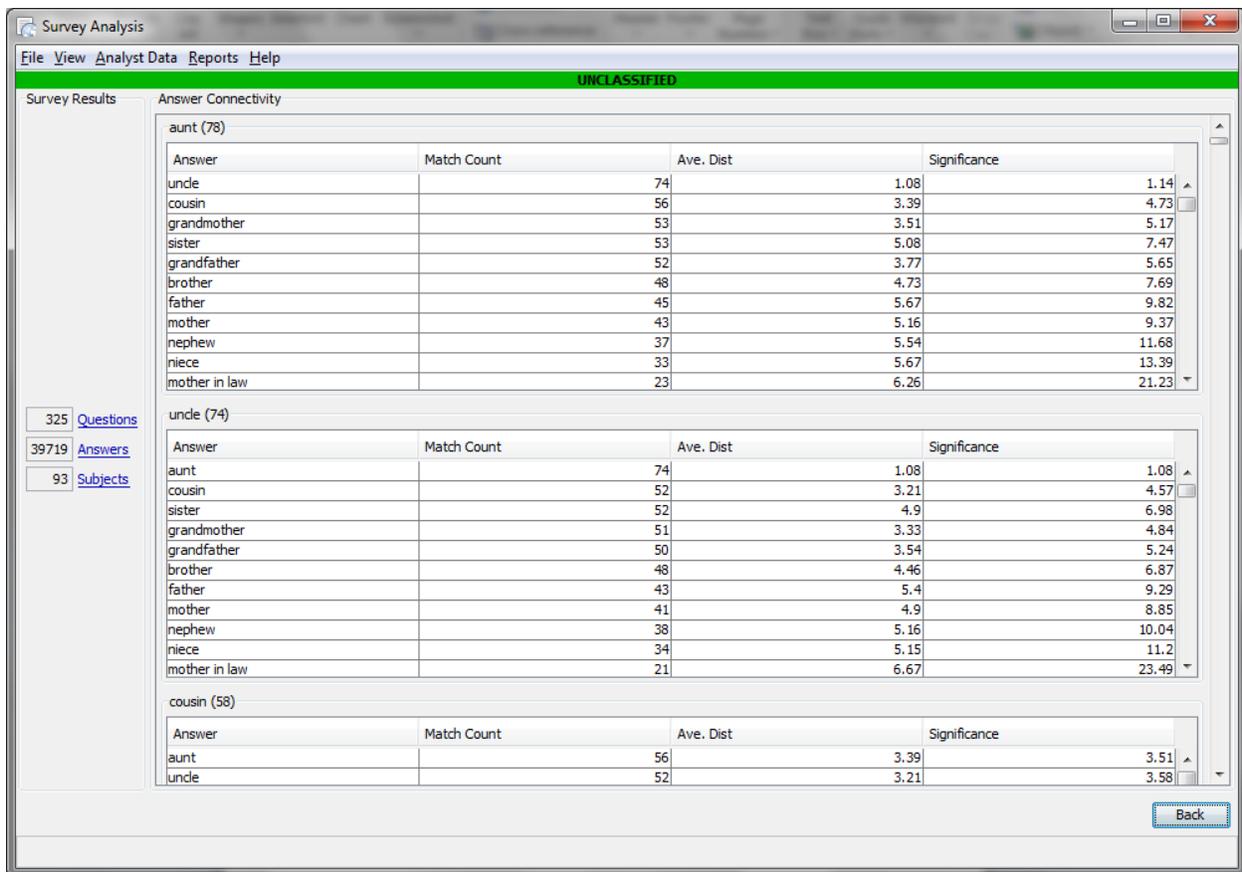


Figure 33: Connectivity

Aggregates: The data set comes from the surveys that were taken by a number of subjects from different cultures and countries. As a result of the answers being freely entered by the subjects of differing backgrounds, data that convey the same information will be written in multiple ways. For instance in the Relatives category, you may get these four answers:

- Grandpa
- Grandfather
- Grampa
- Grandfarther

The first two answers are the common English names for the father of one of your parents. The last two are misspelled versions of the first two. If you didn't aggregate these results together, the summary reports would be faulty. They may tell you that 95% of subjects said cousin, but only 73% said grandfather. However, if you combined the results of all four versions, it might indicate that 92% of the subjects said grandfather. This is why aggregation is valuable.

Aggregation does not modify the original result sets, so they remain untouched. Aggregated data can be changed back at any time.

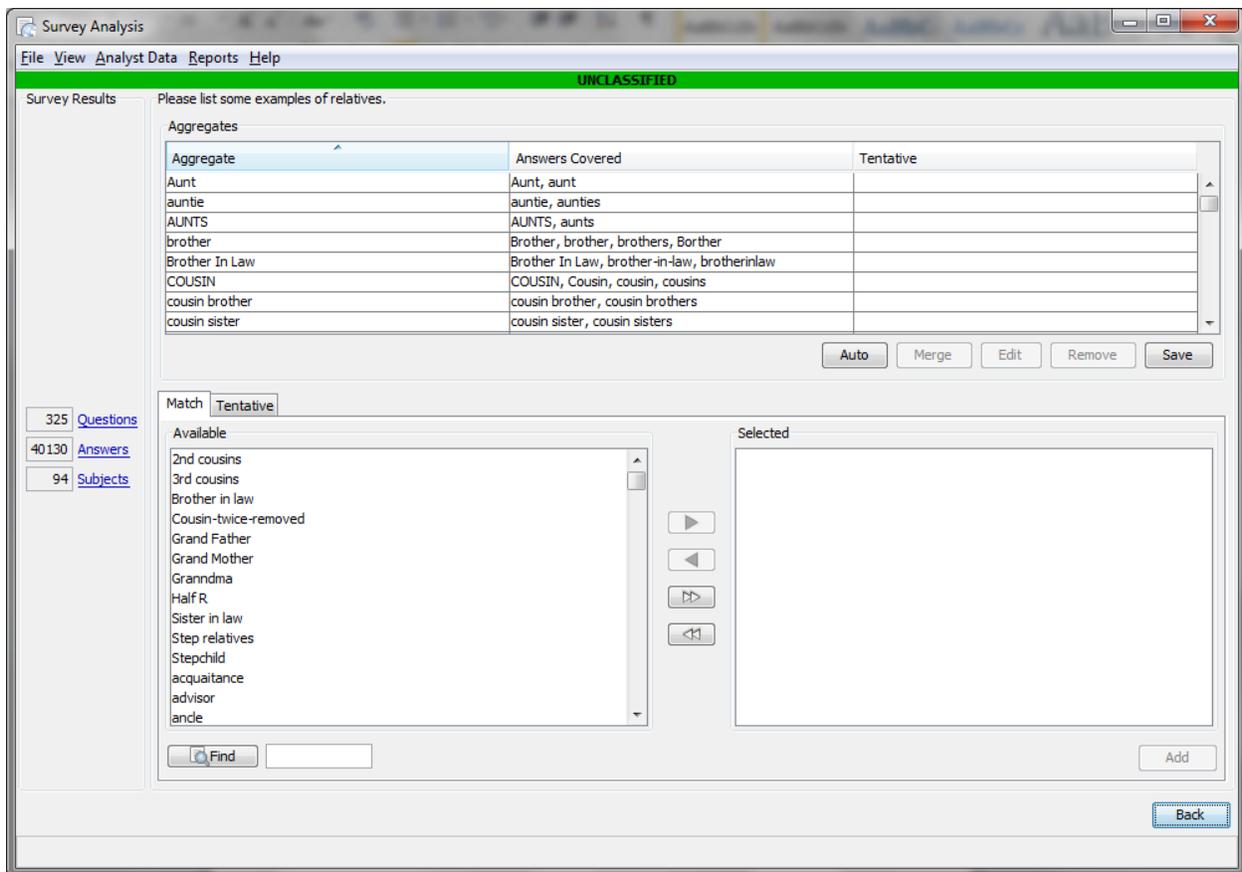


Figure 34: Aggregates

Aggregate Report: The Aggregate Report will list every answer aggregate created for this data set group by their question. The report is a text based report, and is generated as a temporary file and displayed in Notepad (see Figure 35).

The report is structured like this:

Question Text

Aggregate Name: answers covered [tentative answers]

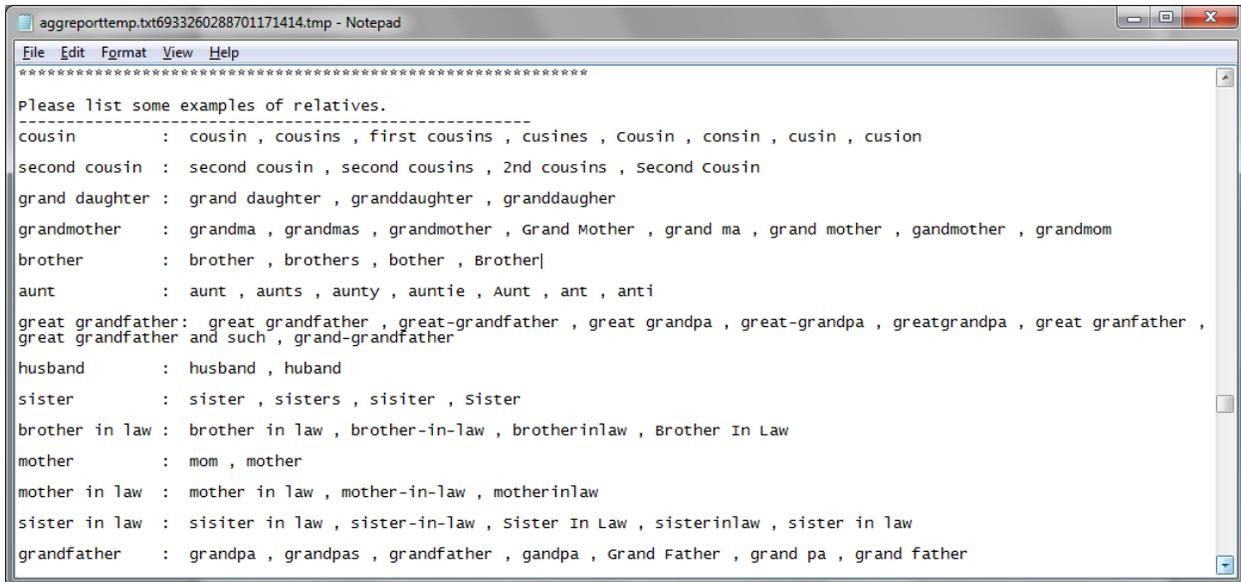


Figure 35: Aggregate Report

Tentatives: The Tentatives edit tool enables the analyst to go back to tentative answers and add them to the aggregate answer they were assigned to (see Figure 36).

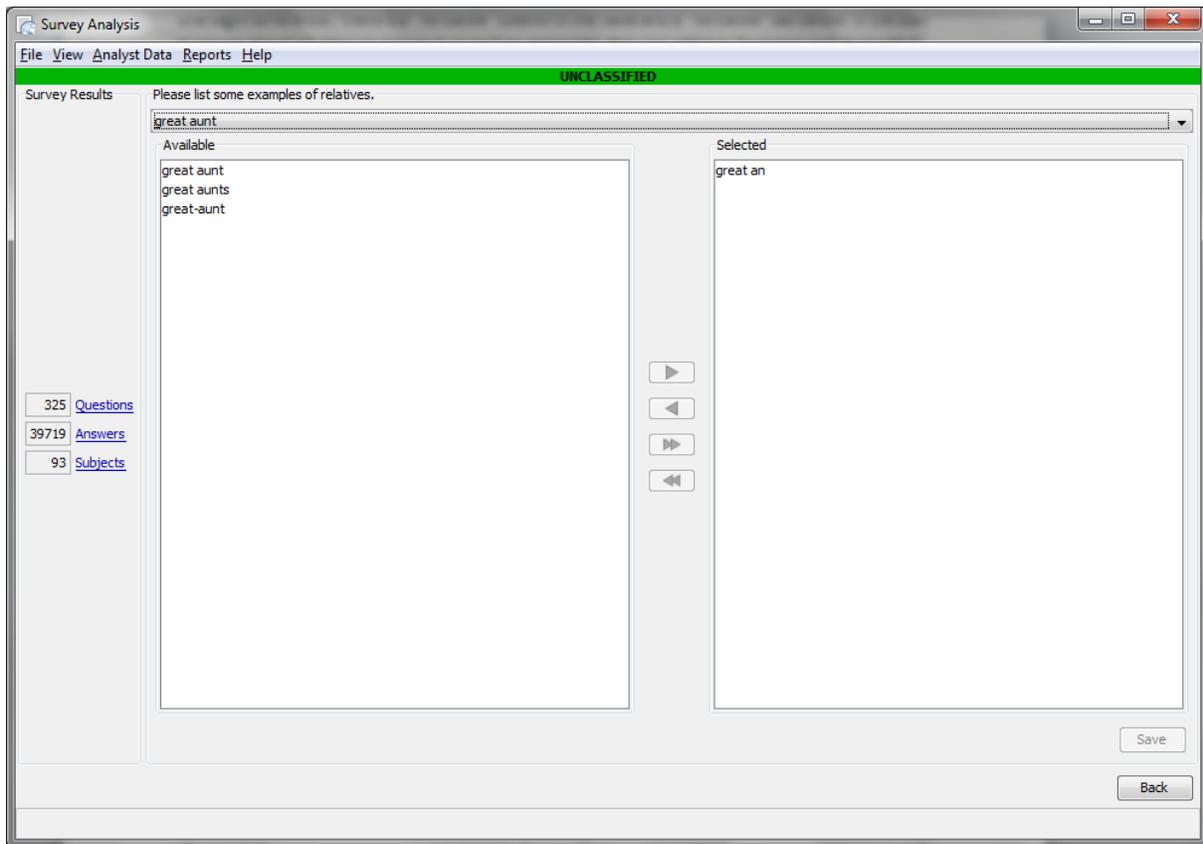


Figure 36: Tentatives

Answer Table: The answer table is shown at the bottom of the screen (see Figure 37). The table lists every answer for the question, with the fields Question #, Subject ID, Result, Entry Time, Exit Time, Total Time, and Entry Number. The table can be sorted by any column. The data in Question # and Subject ID link to the Question and Subject pages respectively.

Question #	Subject ID	Result	Entry Time	Exit Time	Total Time	Entry Number
Q49	R_5uRs2OkixJbGOpe	aunt		0	3121	3121
Q49	R_5uRs2OkixJbGOpe	uncle	3121	3808	687	2
Q49	R_5uRs2OkixJbGOpe	grandmother	3808	5430	1622	3
Q49	R_5uRs2OkixJbGOpe	grandfather	5430	8082	2652	4
Q49	R_5uRs2OkixJbGOpe	cousin	8082	8800	718	5
Q49	R_5uRs2OkixJbGOpe	great-aunt	8800	15087	6287	6
Q49	R_5uRs2OkixJbGOpe	great-uncle	15087	17411	2324	7
Q49	R_5uRs2OkixJbGOpe	great-grandmother	17411	20437	3026	8
Q49	R_5uRs2OkixJbGOpe	great-grandfather	20437	25149	4712	9
Q49	R_5uRs2OkixJbGOpe	mother	25149	25975	826	10
Q49	R_5uRs2OkixJbGOpe	father	25975	26880	905	11
Q49	R_5uRs2OkixJbGOpe	sister	26880	27785	905	12
Q49	R_5uRs2OkixJbGOpe	brother	27785	28830	1045	13
Q49	R_5uRs2OkixJbGOpe	sister-in-law	28830	32122	3292	14
Q49	R_5uRs2OkixJbGOpe	brother-in-law	32122	34087	1965	15
Q49	R_5uRs2OkixJbGOpe	mother-in-law	34087	36708	2621	16
Q49	R_5uRs2OkixJbGOpe	father-in-law	36708	39641	2933	17
Q49	R_5uRs2OkixJbGOpe	stepmother	45491	46552	1061	19
Q49	R_5uRs2OkixJbGOpe	stepbrother	41560	44836	3276	18
Q49	R_5uRs2OkixJbGOpe	step-sister	46552	49516	2964	20
Q49	R_3Jenvii5Oy7Amaq	brother		0	3023	3023
Q49	R_3Jenvii5Oy7Amaq	mother		3023	4025	1002
Q49	R_3Jenvii5Oy7Amaq	father		4025	5038	1013
Q49	R_3Jenvii5Oy7Amaq	sister		5038	6274	1186

Figure 37: Answer Table

Answer Groups: Answer Groups is an unfinished edit tool for possible future development. The object of this feature is similar to aggregation, only at the next level. For example, in the category games, there could be the answers “Madden,” “Super Mario Bros.,” “football,” “soccer,” “Monopoly,” and “checkers.” These six answers could be put into three answer groups, “video games,” “ball games,” and “board games.” Other answer groups could include “card games,” “imagination games,” and “playground games.” There are no reports currently developed that include this type of data.

Save Matrix: The save matrix button will save the question information to a matrix file for use in PDAT.

Answers: All of the answers in the survey are loaded into the application. The Answer Table has every answer in the survey. The fields in the table are Question Number, Subject ID, Result, Entry Time, Exit Time, Total Time, and Entry Number (see Figure 38). The values in the Question Number link to the question page and the values in the Subject ID field link to the subject information page.

UNCLASSIFIED							
Question #	Subject ID	Result	Entry Time	Exit Time	Total Time	Entry Number	
Q1	R_5uRs2OkixJbGOpe	honest		0	3199	3199	1
Q1	R_5uRs2OkixJbGOpe	faithful		3199	6069	2870	2
Q1	R_5uRs2OkixJbGOpe	good sense of humor		6069	12075	6006	3
Q1	R_5uRs2OkixJbGOpe	romantic		12075	13261	1186	4
Q1	R_5uRs2OkixJbGOpe	considerate		13261	15975	2714	5
Q1	R_5uRs2OkixJbGOpe	good with kids		15975	20889	4914	6
Q1	R_5uRs2OkixJbGOpe	good morals		20889	23525	2636	7
Q1	R_5uRs2OkixJbGOpe	goal-oriented		23525	27737	4212	8
Q1	R_5uRs2OkixJbGOpe	has a steady income		34211	37378	3167	9
Q1	R_5uRs2OkixJbGOpe	shares responsibilities		37378	42776	5398	10
Q1	R_5uRs2OkixJbGOpe	communicates often		42776	46598	3822	11
Q1	R_5uRs2OkixJbGOpe	argues effectively		46598	48891	2293	12
Q1	R_5uRs2OkixJbGOpe	trustworthy		48891	54460	5569	13
Q1	R_5uRs2OkixJbGOpe	have comm		54460	60000	5540	14
Q1	R_3Jenvi5Oy7Amaq	loving		0	2996	2996	1
Q1	R_3Jenvi5Oy7Amaq	supportive		2996	5609	2613	2
Q1	R_3Jenvi5Oy7Amaq	caring		5609	8027	2418	3
Q1	R_3Jenvi5Oy7Amaq	warm		8027	9562	1535	4
Q1	R_3Jenvi5Oy7Amaq	affectionate		9562	11557	1995	5
Q1	R_3Jenvi5Oy7Amaq	sensual		11557	21848	10291	6
Q1	R_3Jenvi5Oy7Amaq	good communicator		21848	33981	12133	7
Q1	R_3Jenvi5Oy7Amaq	reliable		33981	36605	2624	8
Q1	R_3Jenvi5Oy7Amaq	attractive		36605	40740	4135	9
Q1	R_eFAQEkutGJFoYAc	trustworthy		0	7582	7582	1
Q1	R_eFAQEkutGJFoYAc	trusting		7582	12606	5024	2
Q1	R_eFAQEkutGJFoYAc	respectful		12606	15601	2995	3
Q1	R_eFAQEkutGJFoYAc	driven		15601	22808	7207	4
Q1	R_eFAQEkutGJFoYAc	sense of humor		22808	26053	3245	5
Q1	R_eFAQEkutGJFoYAc	good looking		26053	28986	2933	6
Q1	R_eFAQEkutGJFoYAc	good family person		28986	55818	26832	7
Q1	R_4G6xFbyDK9beePO	patient		0	6631	6631	1
Q1	R_4G6xFbyDK9beePO	calm		6631	8472	1841	2
Q1	R_4G6xFbyDK9beePO	caring		8472	11264	2792	3
Q1	R_4G6xFbyDK9beePO	selfless		11264	13526	2262	4

Figure 38: Answers

Subjects: All of the subjects' information in the survey is loaded into the application. The Subject Table lists every subject in the survey. The Subject Summary report can be generated by selecting the Generate Report button in the upper right corner of the screen (see Figure 39).

Survey Analysis

File View Analyst Data Reports Help

UNCLASSIFIED

Survey Results Subject Information [Generate Report](#)

ID	Nationality	Start Date	End Date	Age	Gender
R_034L2YJfMAHb81	Indian	Oct 9, 2012	Oct 9, 2012		23 Female
R_065XqoVa36pk27q	India	Jun 13, 2012	Jun 13, 2012		25 Female
R_0B04GqCUTfSOFKv	Indian	Feb 5, 2013	Feb 5, 2013		25 Male
R_0Gm3kEHUX75hJb	American	Feb 1, 2013	Feb 1, 2013		26 Male
R_0HbNALTJJOheJ8w	American	Dec 26, 2011	Dec 26, 2011		19 Male
R_0IE7dJtdbCFahz	indian	Feb 19, 2013	Feb 19, 2013		22 Male
R_0JkJHTCdxpDpfPm	Indian	Apr 12, 2012	Apr 12, 2012		25 Female
R_0kNG8h8x7Eae9	Chinese	Sep 27, 2012	Sep 27, 2012		23 Female
R_0MtfEi3ACnGFXBq		Jan 25, 2012	Jan 25, 2012		19 Female
R_0NcoZE1xYzqmKF	Mexican	Oct 9, 2012	Oct 9, 2012		0 Male
R_0SwervY7RaVHdoU	india	Apr 12, 2012	Apr 12, 2012		27 Female
R_0Uu0AunszZqIOQ	india	Apr 23, 2012	Apr 23, 2012		21 Female
R_0wDBjh07yXL0qBn	China	Sep 30, 2012	Oct 1, 2012		23 Male
R_0xE8hYT41PpJ5Aw	India	Apr 23, 2012	Apr 23, 2012		22 Female
R_1CfvpKD3veAZNPF	Indian	Jan 13, 2013	Jan 14, 2013		27 Male
R_1Tt5GJPrLkJHzn	Indian	Feb 19, 2013	Feb 19, 2013		24 Male
R_24sAJwTlBbzmeF	United States	Feb 1, 2013	Feb 1, 2013		0 Male
R_29pEaDQl8lbnex	indian	Feb 19, 2013	Feb 19, 2013		0 Male
R_2qfm83eYtCNl0xv	China	Oct 2, 2012	Oct 2, 2012		26 Male
R_2npYhrhOfn10Wwf	People Republic China	Oct 4, 2012	Oct 4, 2012		23 Male
R_2ocnupsClni7lGf	Caucasian (European-American)	Feb 22, 2012	Feb 22, 2012		21 Female
R_3ehq4U4Z35dm2Ks	American	Jan 25, 2012	Jan 25, 2012		38 Male
R_3EhF75XivZD0bhH	Indian	Oct 9, 2012	Oct 9, 2012		26 Female
R_3fxRqOXOX18EcY	American	Feb 20, 2012	Feb 20, 2012		32 Female
R_3Jenvii5Oy7Amaq	American	Dec 21, 2011	Dec 21, 2011		33 Male
R_3xcOm2IG1OmzypT	Chinese	Feb 7, 2013	Feb 8, 2013		31 Male
R_3ZTNIOnfottUuvH	American	Feb 1, 2013	Feb 1, 2013		22 Female
R_4G6xEbyDK9beePO	Cacasuan	Dec 25, 2011	Dec 25, 2011		24 Female
R_4DXFLaSU51Jhha	Indian	Feb 16, 2012	Feb 16, 2012		19 Female
R_4SkRSO6v10VWcl	Serbian	Feb 24, 2012	Feb 24, 2012		36 Female
R_50wXlIncnMN1tbe	African American	May 29, 2012	May 29, 2012		19 Female
R_550HZ95oAihZ9l2	white american	Feb 16, 2012	Feb 16, 2012		21 Male

325 Questions
39719 Answers
93 Subjects

[Back](#)

Figure 39: Subjects

Subject Table: The fields in the Subject Table are ID, Nationality, Start Date, End Date, Age and Gender. The values in the ID column link to the subject information page.

Subject Summary: The Subject Summary report gives summary information about the survey results broken down by subject groups (see Figure 40). The report is currently broken down by nationality, but could be modified in the future to be broken down by gender/age/etc.

The first panel is the “All” panel, which shows the results for all subjects. The rest of the panels are broken down by subject nationality. Each panel has two sections: Summary and Question Type. The Summary section lists subject count, unique words, total words, and unique per subject. The Question Type selection lists all the question types as set in the Data Annotation dialog. Each type lists the number of questions in that type, the average number of answers for each question of that type, and the average amount of time for each of those answers.

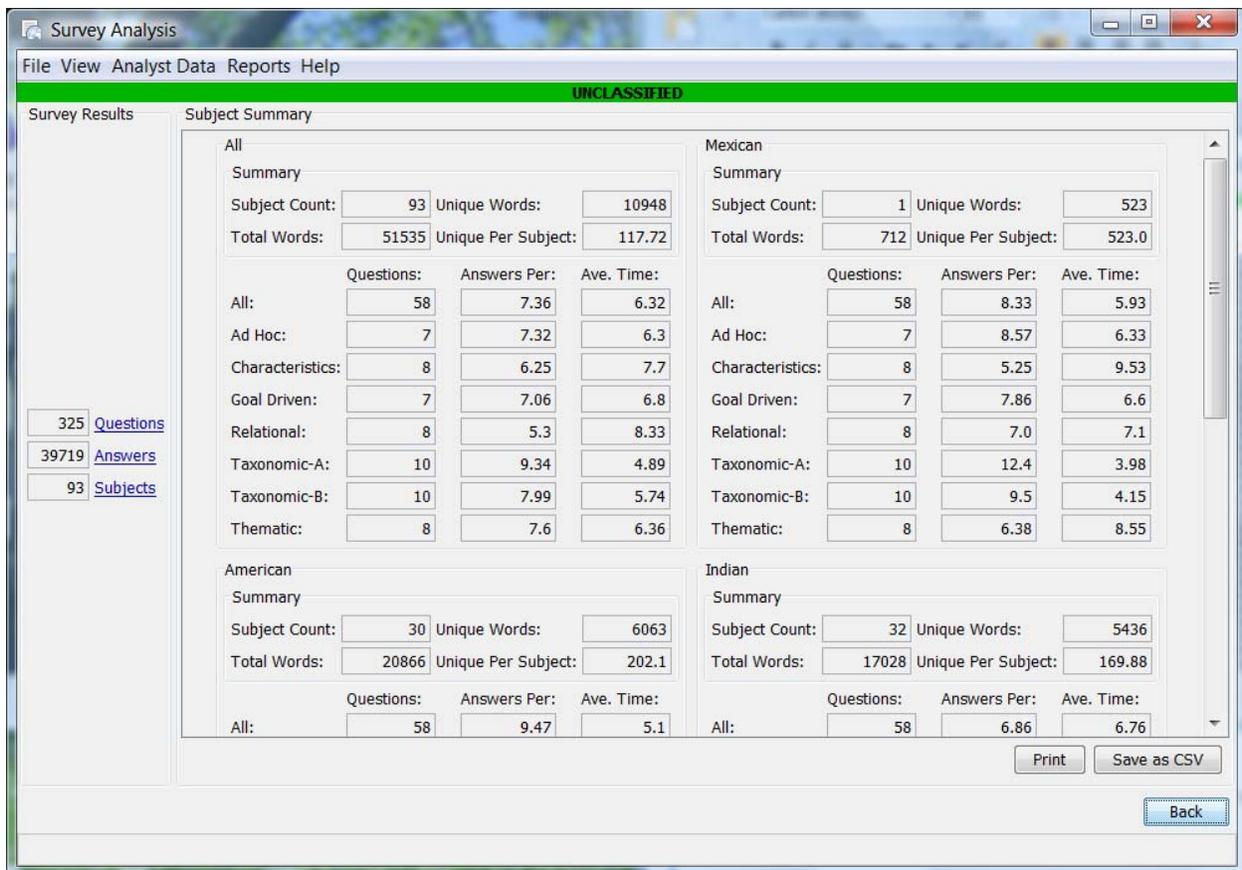


Figure 40: Subject Summary

Panel Report: The Panel Report breaks down the subjects by their demographic information. Subjects are grouped overall and by their nationality (see Figure 41). Each demographic question lists the multiple choice options and number of responses (fixed), or they list the provided answers (aggregated in the Aggregate tool) and number of responses (dynamic). The dynamic results are Languages Spoken/Written, Nationality, and Language Preference.

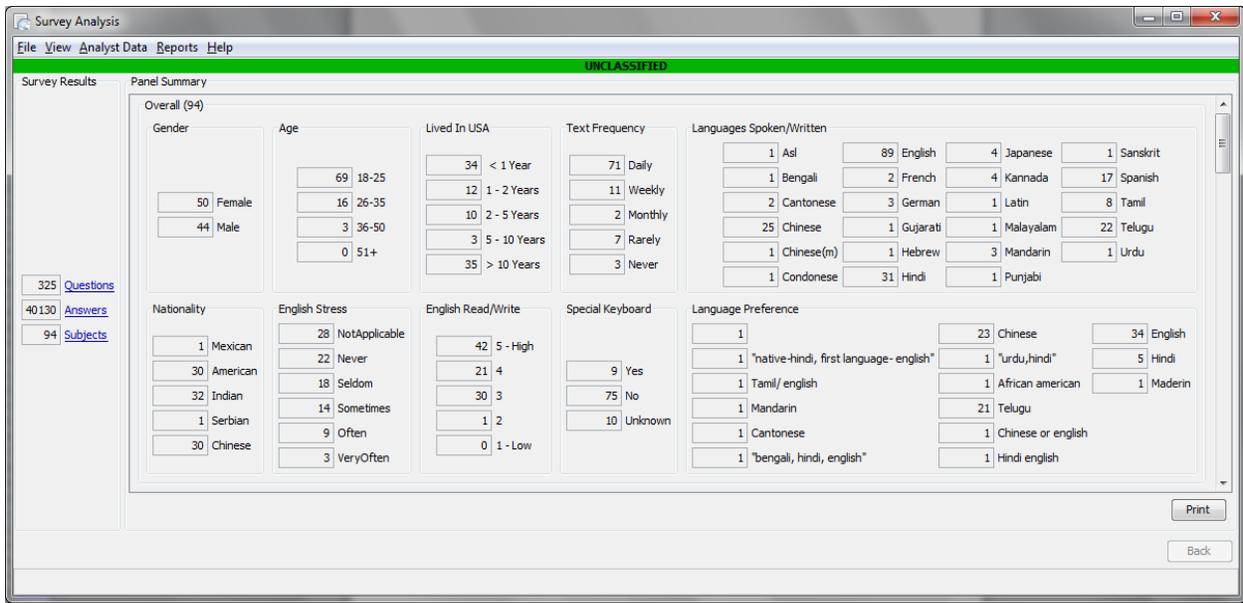


Figure 41: Panel Report

Subject Aggregates: The Subject Aggregates tool creates aggregate answers for subject demographic data (see Figure 42).

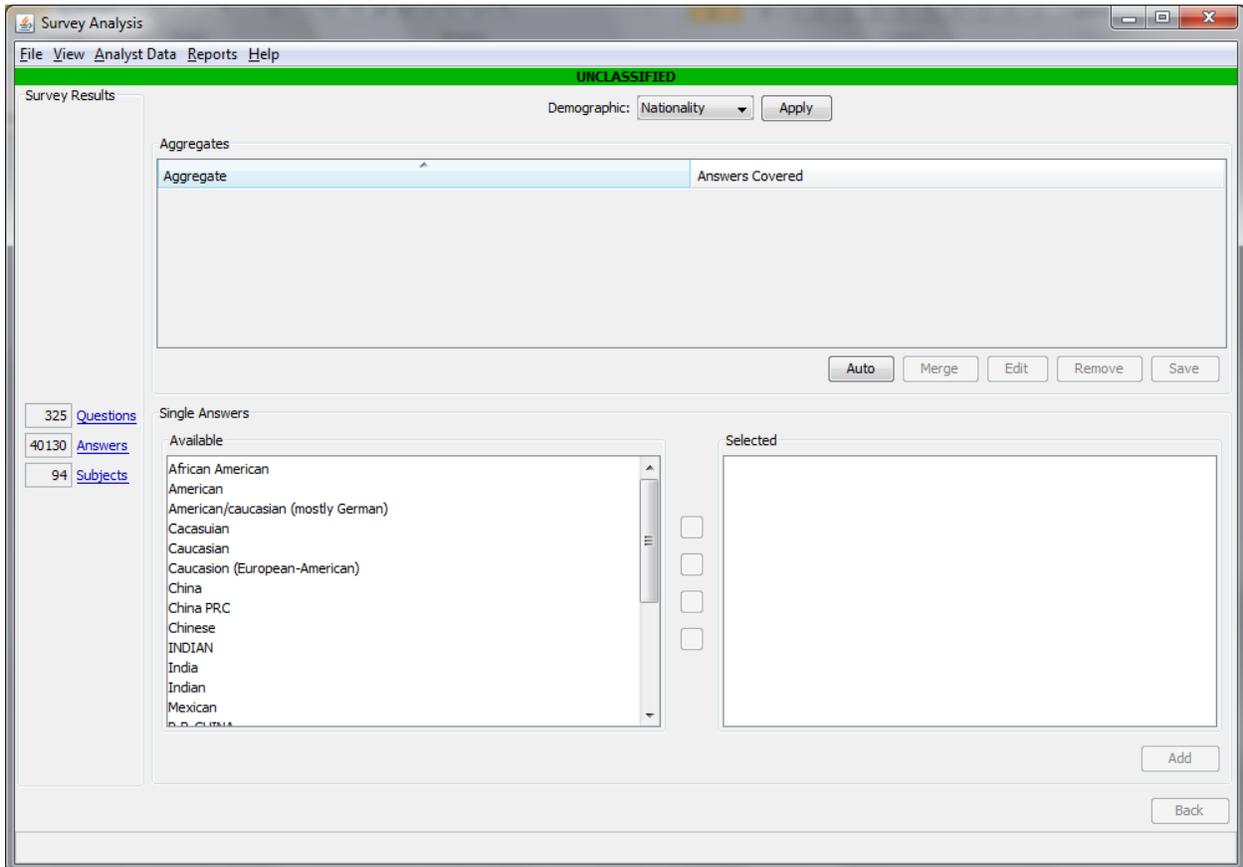


Figure 42: Subject Aggregates

The combo box at the top of the panel lists the demographic information groups that can be aggregated (see Figure 43).

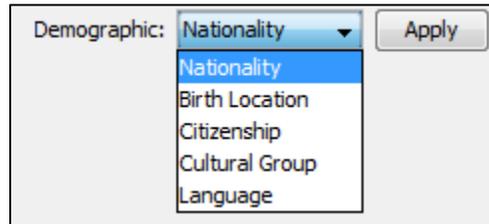


Figure 43: Demographic Information Groups

The Subject Aggregates tool works the same way that the Question Answer Aggregate tool does.

Annotation: Some data is automatically parsed, but sometimes the parsing fails. Data annotation is used to clear those data pieces up, and to set up categories for questions used in the survey. There are four data elements that can be annotated: Question Type, Nationalities, Time in US and Languages (see Figure 44).

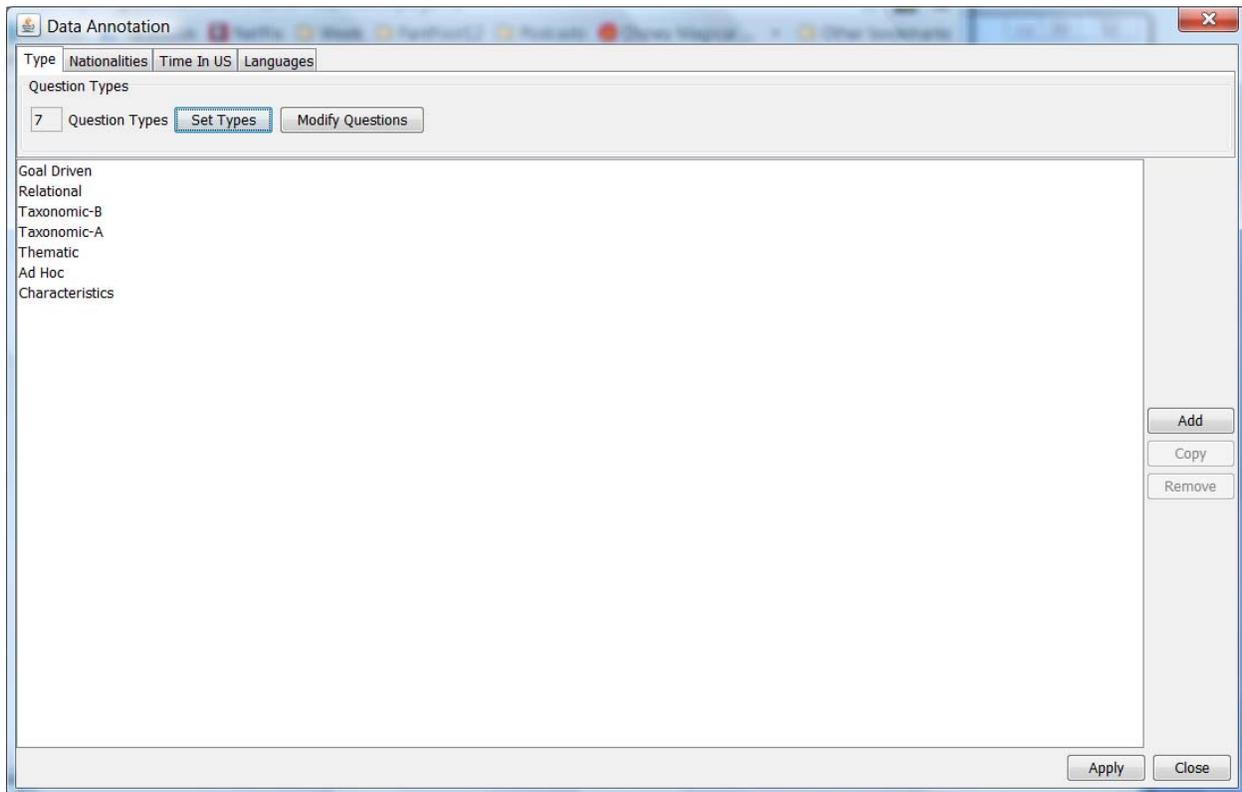


Figure 44: Data Annotation

Question Type: Question type data is not found anywhere in the data. Question types are crafted by the survey generator and can be entered using the Type tab of the Data Annotation screen. The Set Types button enables the user to input the question types, and the Modify Questions button is used to assign the types to the questions in the survey (see Figure 45).

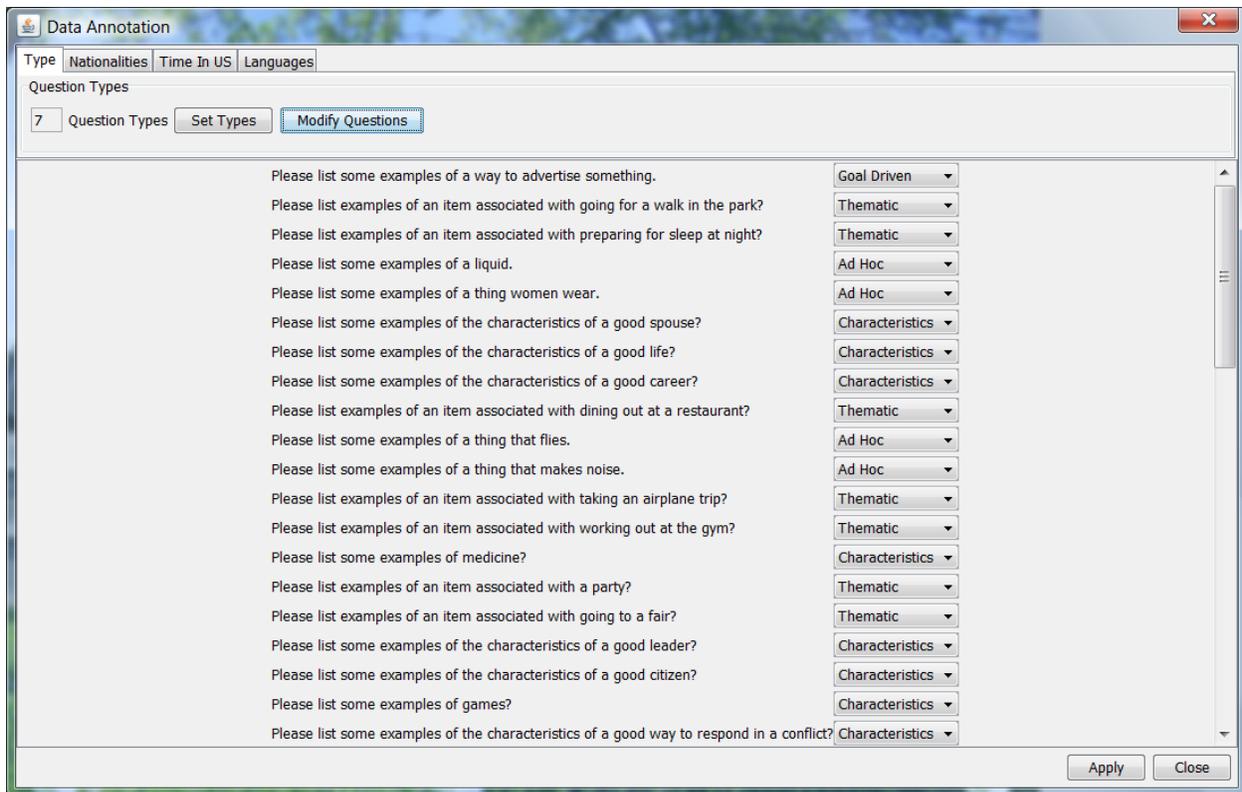


Figure 45: Modify Question Types

Nationalities: Nationality is an important piece of information for grouping subjects in Survey Analysis. Some subjects neglected to enter a value for nationality, perhaps because they were confused by the question or just missed it. The Nationalities tab allows the user to view the subject’s demographic information and determine nationality if possible. Once the user updates this information, the data to the right of the field is marked with a red asterisk to indicate that this piece of information was modified by the user (see Figure 46).

The screenshot shows a 'Data Annotation' window with three tabs: 'Nationalities', 'Time In US', and 'Languages'. The 'Nationalities' tab is active. At the top, there is a dropdown menu for 'Nationalities' with the value 'R_0MtEi3ACnGFXBq' and another dropdown menu for 'Nationality' with the value 'American'. An 'Update' button is next to the 'Nationality' dropdown. Below this is a large form with the following fields and values:

ID:	R_0MtEi3ACnGFXBq
Name:	Anonymous
Start Date:	Jan 24, 2012 2:46:50 PM
End Date:	Jan 24, 2012 4:01:19 PM
Age:	19
Gender:	Female
Nationality:	American *
Birth Location:	london OH
Citizenship:	American citizen
Cultural Group:	N/n
Preferred Language:	English /Spanish
Lived in US:	19
Languages you can speak:	English /Spanish
Ability to read/write English:	5
Stress in English:	NotApplicable
Computer Expertise:	High
Year in College:	sophomore
Text Messaging:	Daily
Special Keyboard:	no

At the bottom right of the window are 'Apply' and 'Close' buttons.

Figure 46: Nationalities

Time in US: The third tab on the Data Annotation screen is the “Time in US” tab. Survey Analysis parses this data set as best as it can, and converts the information into months. Some values entered cannot be parsed, and have to be entered manually on this screen. The combo box lists all the subjects with problem data. The User Value is the value the subject entered for the Time in the US question. The Annotate Value is what is currently set for that subject’s time in the US. The default is -1 month. In this example, the subject has been in the US for 15 days see (Figure 47). The user can decide to round down to zero months or up to one month.

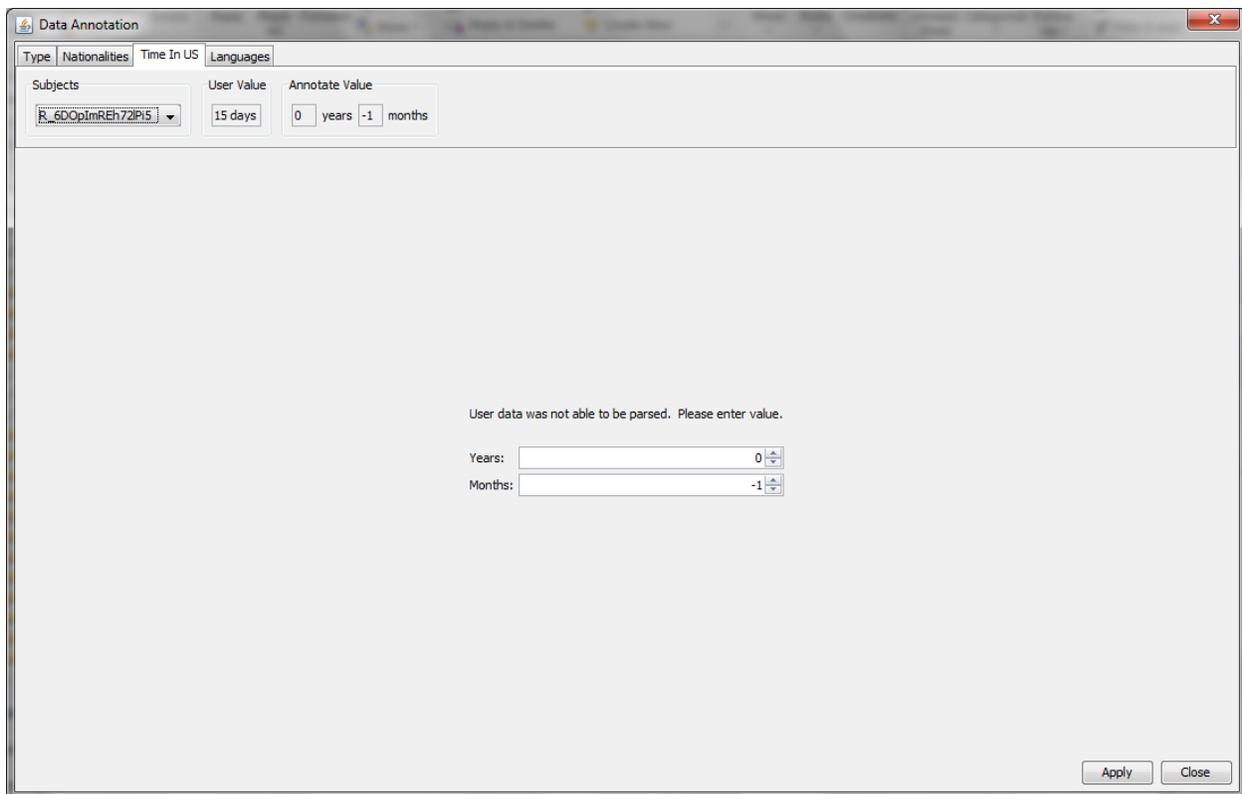


Figure 47: Time in US

Languages: The fourth tab on the Data Annotation screen is the “Languages” tab (see Figure 48). This tab contains the tools to annotate the information provided by the subjects, concerning the languages they speak. The available languages are listed at the top of the panel. Each response is listed in the first column of the data, and the appropriate languages are checked in each row. This is automatically parsed from the data provided. If any of the answers appear to be misspelled, the appropriate check boxes can be marked here.

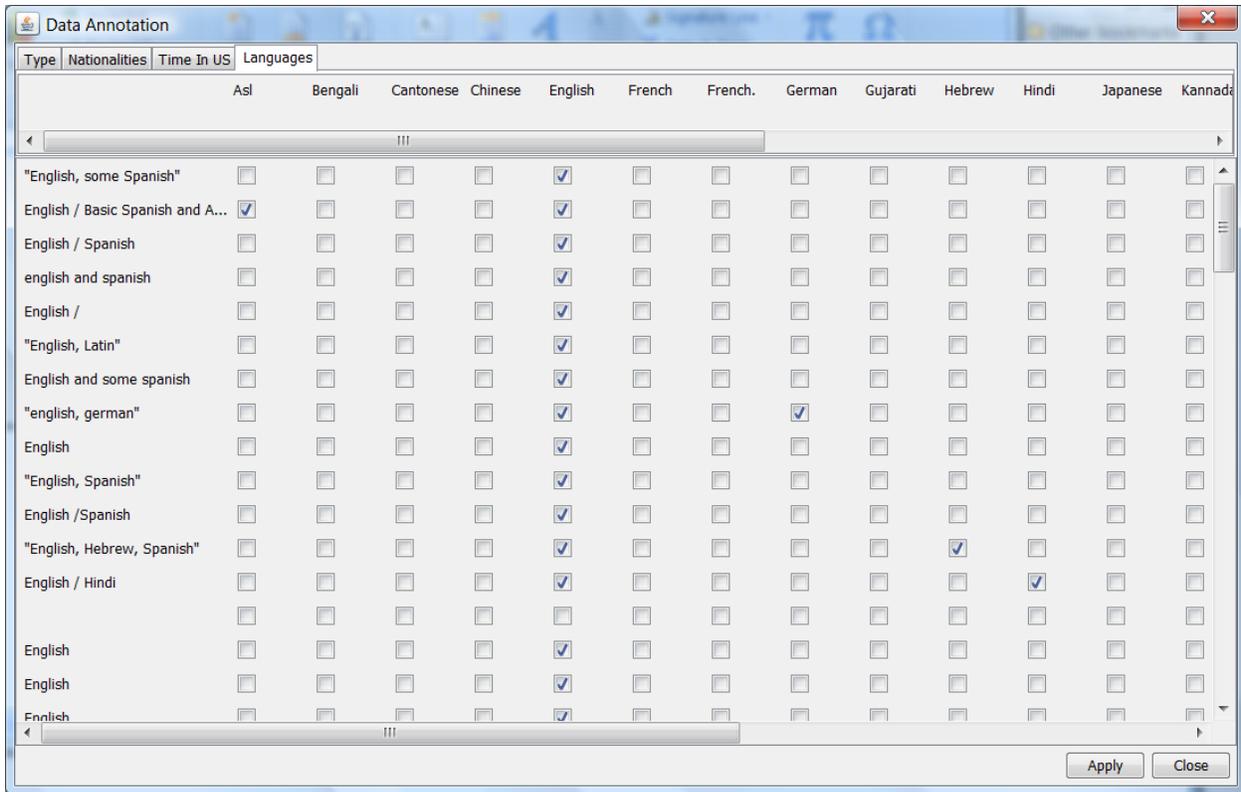


Figure 48: Languages

Survey Summary: The Survey Summary gives a summary of the answers given for each question (see Figure 49). Questions are grouped together by their type. Filtering options for nationality and aggregates are available.

The 'Survey Summary' window shows a table of results for a 'Goal Driven' question. The question is: '1. Please list some examples of a way to advertise something.' The table lists various responses and their associated statistics across different nationalities.

Response	Total (93)	First	RT overall	RT First	Rank	Mexican (1)	American (30)	Indian (32)	Serbian (1)	chinese (29)
TV	0.39	0.15	6.6	10.8	2.4	1.00	0.40	0.25	1.00	0.48
newspaper	0.38	0.11	7.4	13.3	3.6	1.00	0.33	0.34	0.00	0.45
internet	0.37	0.03	5.1	11.5	3.8	0.00	0.50	0.31	0.00	0.31
radio	0.37		4.8		4.5	0.00	0.60	0.28	1.00	0.21
billboard	0.32	0.16	6.0	7.3	2.8	0.00	0.80	0.09	0.00	0.10
flyer	0.30	0.01	5.0	3.1	5.4	0.00	0.50	0.22	1.00	0.17
magazine	0.22		4.8		3.9	0.00	0.40	0.03	0.00	0.24
word of mouth	0.22	0.03	6.6	8.4	5.3	0.00	0.33	0.22	1.00	0.07
e-mail	0.18	0.02	4.5	6.9	4.9	0.00	0.27	0.16	1.00	0.10

Figure 49: Survey Summary

Options: There are two groups of options under the Survey Summary Selection: Type and Settings (see Figure 50). The left panel is labeled Type. The “Summary” option generates the report with all nationalities included. Listed below “Summary” are all of the available nationalities. Selecting one of these choices generates the report using only results from subjects of that nationality.

Settings are displayed on the right of the panel. There are three settings that can be chosen:

Exclude – If checked, the report will only show the answers that were answered by a certain percentage of the selected subjects. Default selection is 5%.

Aggregates - If selected, the report will break down each aggregate value. For example:

Apple was answered by 80% of subjects

Apple is an aggregate answer, and covers four answers

Apple – 70%

Red Apple – 5%

Green Apple- 4%

Yummy Apple – 2%

Rounding will cause some minor differences ($70+5+4+2 = 81\%$)

CSV – Instead of displaying the report, if this setting is checked a CSV file is generated and saved to the file system. The CSV file can then be loaded into a program such as Excel.

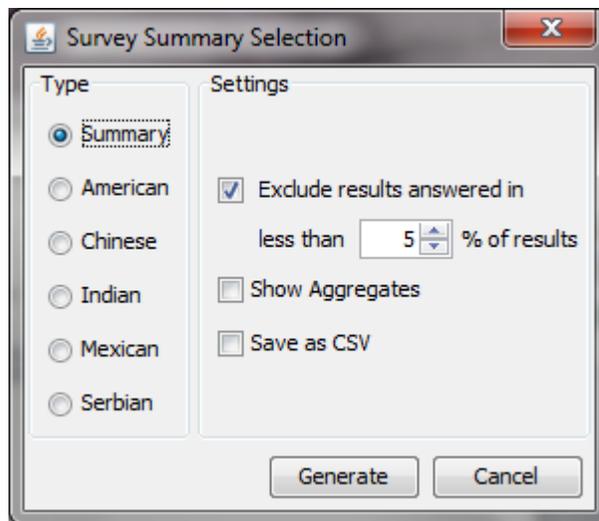


Figure 50: Survey Summary Selection

4.8 Report

The report groups questions by their question type (see Figure 51). Under the text of the question are the answers for that question. Each answer is one row of the table.

- a. Response – The answer provided by the subjects.

- b. Total (#) – The # is the number of subjects who answered this question. The value in the column is the percentage of subjects who answered this question with this response.
- c. First – The percentage of subjects who answered this question who chose this response first.
- d. RT Overall – The average response time.
- e. RT First – The average response time for this response if it was chosen first.
- f. Rank – The average location of the response.
- g. Nationality (##) – In the overall summary, the last columns will display all the nationalities and the percentage of subjects of this nationality who responded with this answer. The ## is the number of subjects of this nationality.

16. Please list some examples of relatives.					
Response	American (30)	First	RT Overall	RT First	Rank
cousin	1.00	0.10	2.8	5.6	5.9
--> cousin	(0.77)	(0.07)	(2.2)	(3.9)	(6.3)
--> cousins	(0.17)	(0.03)	(6.0)	(8.9)	(3.8)
--> Cousin	(0.07)		(2.0)		(7.0)
aunt	1.00	0.37	2.5	3.4	4.3
--> Aunt	(0.10)	(0.07)	(3.2)	(4.0)	(2.7)
--> aunts	(0.17)	(0.10)	(3.0)	(3.4)	(1.8)
--> aunt	(0.73)	(0.20)	(2.3)	(3.3)	(5.1)
uncle	0.97		1.6		4.8
--> uncle	(0.07)		(1.6)		(4.5)
--> uncles	(0.17)		(1.7)		(2.4)
--> uncle	(0.73)		(1.6)		(5.3)
grandfather	0.87		3.3		8.0
--> Grandpa	(0.07)		(3.4)		(10.0)
--> grandpa	(0.47)		(3.0)		(7.7)
--> grandpas	(0.03)		(1.8)		(4.0)
--> grandfather	(0.27)		(4.0)		(9.0)
--> gandpa	(0.03)		(2.8)		(5.0)
sister	0.87		2.0		6.3
--> sister	(0.77)		(2.0)		(6.6)
--> sisters	(0.03)		(2.1)		(5.0)
--> sister	(0.07)		(1.8)		(4.0)
grandmother	0.87	0.07	3.2	3.9	7.2
--> grandmas	(0.03)		(1.3)		(3.0)
--> Granddma	(0.03)		(4.8)		(8.0)
--> Grandma	(0.03)		(3.8)		(10.0)
--> grandma	(0.50)	(0.07)	(3.5)	(3.9)	(6.6)
--> grandmother	(0.27)		(2.6)		(8.3)
mother	0.83	0.20	2.1	4.3	3.8

Figure 51: Summary Survey (American)

4.8.1. Completed Survey Results

Data from the first set of surveys has been gathered through the Qualtrics survey. Subjects took the survey in English, regardless of their first or preferred language. The Qualtrics data collection is complete and the results were downloaded on 14 March 2013.

4.9 Panel Information

There were 94 subjects who participated in this survey.

- Gender – 53% male / 47% female
- Age – 73% were aged 18-25, 17% were 26-35, 3% were 36-50, and 6% declined to answer.
- Languages Spoken/Written – 95% listed English, 33% Hindi, 28% Chinese, 23% Telugu, 18% Spanish, 9% Tamil, and no more than 4% for any other language.
- Nationality – 34% Indian, 32% Chinese, 32% American, and 1% for both Mexican and Serbian.
- Language Preference – 36% preferred English, 24% Chinese, 22% Telugu, 5% Hindi and 1% for a number of other languages.

4.10 Translation

The surveys will be conducted in a number of different languages, so a translation tool will be needed.

4.10.1. Criteria for translation tool

A variety of criteria were used to select the software for data collection. The criteria included:

- a. Languages supported
- b. Human translated vs. machine translated
- c. Amount of time required for translation
- d. Cost

The data collection software was evaluated against these criteria. The four language translations that the software must provide are Chinese, Urdu, Sanskrit and Hindi. Human translations are better for sentences, but machine translation may be adequate for some of the one word responses.

To receive an adequate quote on cost and time, the number of words requested was 1000 for each of the required languages.

4.10.2. Translation Options

There were six evaluated services that provide human translation of the four languages, and three machine translation tools that provide translations for three of the languages (sans Sanskrit). A breakdown is shown in Figure 52.

4.10.3. Human Translation Options

The human translation options are Advanced Language, AltaLang, Betmar, Lionbridge, Transperfect, and Verbatim Solutions.

4.10.4. Machine Translation Options

The machine translation options are Ace Translator, Babylon and Google.

TranslationServices						
ID	Service Name	Website	Cost	Notes	Software	
9	Verbatim Solutions	http://www.verbatimsolutions.com/	\$160, \$220, \$210, \$220 Total: \$810.00	2-3 Day completion	<input type="checkbox"/>	
16	Transperfect	http://www.transperfect.com/	\$200, \$290, \$290, \$290 Total: \$1,070.00	5 day complete	<input type="checkbox"/>	
12	Advanced Language	http://advancedlanguage.com	\$1075 Total: \$1,075.00	3-5 day complete	<input type="checkbox"/>	
11	AltaLang	http://www.altalang.com	\$200(s)/\$240(t), \$260,\$480,\$260 Total: \$1,240.00	3 day complete	<input type="checkbox"/>	
3	Betmar	http://www.betmar.com	\$595, \$625, \$625, \$625 Total: \$2,470.00		<input type="checkbox"/>	
7	Lionbridge	http://www.lionbridge.com/lionbridge/en-US/services/loc ?		Requested info, have not heard i	<input type="checkbox"/>	
4	Google	http://translate.google.com/	Free Total: \$0.00		<input checked="" type="checkbox"/>	
13	Ace Translator	http://www.acetools.biz/	\$59.95 Total: \$59.95		<input checked="" type="checkbox"/>	
8	Babylon	http://www.babylon.com/	\$9.70 a month, \$118 lifetime Total: \$118.00		<input checked="" type="checkbox"/>	

Figure 52: Translation Option Table

5.0 CONCLUSIONS

The use of OMAR to run scenarios in Neverwinter Nights and input the results to PDAT was successful. The resulting displays in PDAT are very informative, and will provide easier analysis of the data, as the scenarios become more complex and lengthy. The variety of reports provides a deeper analysis that will be valuable in revealing trends in the data.

The Neverwinter Nights' displays have issues that need to be addressed, in order to streamline scenario display. The environments can be manipulated, but when running the scenarios, in order to view the events in the scenario, an active response is required. There have been discussions regarding changing the tool used for scenario display, with OpenSim being a top candidate to replace Neverwinter Nights. OpenSim will be evaluated in the future to determine if it has the functionality to replace the current simulation tool.

Data collection will start in the near future. No problems are anticipated with the retrieval of the data. The tool has been tested, and all issues were resolved with simple fixes. Qualtrics has all the necessary tools for collection according to our specific needs, including timing and automatic page change features. The results returned from Qualtrics will be analyzed and detailed in a future report.

5.1 Aggregation

Aggregation is a necessity for evaluation of the data provided by the subjects. Due to the nature of the surveys, answers for all the category questions and some of the demographic questions are text based. This means that subjects are not forced to follow any particular standard of entering data. Questions with radio button selections have a finite number of answers that can be provided, while text based answers are nearly infinite.

In general, answers need to be the same across users. There are a finite number of different types of fruit for example, but that finite group grows with every misspelled word, misplaced punctuation, and other anomalies.

For questions that are more abstract, such as "Please list some examples of a way to advertise something," similar answers vary much more (see Figure 53).

newspaper	0.44
--> new papers	(0.01)
--> news papper	(0.01)
--> news papper adds	(0.01)
--> in the paper	(0.01)
--> local news paper	(0.01)
--> news papers	(0.01)
--> Classified in newspaper	(0.01)
--> news paper	(0.01)
--> newspaper ad	(0.02)
--> by newspaper	(0.01)
--> News Paper	(0.01)
--> newspapers	(0.03)
--> newspaper advertisement	(0.01)
--> newspaper	(0.27)

Figure 53: Newspaper Answer

In the above example, 44% of the subjects answered a form of newspaper as an example of a way to advertise something. The most common answer was “newspaper”, which was the answer in 27% of survey results.

There are 14 answers that have been combined for this aggregate answer. Besides “newspaper,” the other 13 answers had one or more reasons why they needed to be aggregated.

- 7 added a space between “news” and “paper”
- 1 used capitalization (News Paper)
- 2 misspelled paper (“papper”)
- 1 misspelled news (“new”)
- 2 pluralized paper to papers
- 4 referenced a specific element of a newspaper (“ad” or “Classified”)
- 1 defined what kind of newspaper (“local”)
- 2 added unnecessary words (“by newspaper” and “in the newspaper”)

Without aggregation, 16 answers would have been lost by default, as none of the aggregated answers other than “newspaper” would have reached the default 5% threshold. “Newspaper” would have appeared to be answered only by a quarter of subjects, as opposed to nearly half.

While the benefits of aggregation are made clear by this example, there are downsides and risks to this approach. First, aggregation is a slow, tedious and time consuming process. Each question has its own group of aggregates, so the process must be run for all 58 questions in the survey. Automation handles some capitalization and plurals, but many of the aggregate answers are not covered by automation.

Aggregates were not automatically generated, rather the analyst could run automation, if they so desired. The reasoning was a fear that some answers may actually be different if capitalized. In the statements “A Christian loves God” and “Zeus is a god,” the word “god” is different if capitalized. There may be other examples in other languages as well.

If this application is expanded, it would be recommended that automation be run every time, as the number of mistakes are few and can easily be fixed. Unfortunately, it is still recommended to provide aggregates within each question, to avoid cross contamination between questions. For example, if the category was “things in an office,” “apple,” “computer,” and “laptop” may be grouped together under computer. However, “computer” would be listed in the “fruit” category, instead of “apple.”

The more complicated issue is determining which answers should be aggregated. Is “Classified in newspaper” a part of “newspaper” or “classified?” It contributes to the fidelity of the answers. “Newspaper” could cover advertisements printed in the paper, classified ads, an article written about a product, a paper insert, or a free sample provided with the Sunday paper. Then the question of how to handle advertisements on a newspaper’s webpage only complicates the situation.

Due to the nature of the survey, the subject provided answers within a time limit (60 or 90 seconds). As such, participants may not spend much time differentiating the specific elements of a newspaper. The aggregates that were created for this analysis showed preference to grouping at

a higher level. Using the example above, the answers were grouped under “newspaper” instead of breaking them up into smaller groups like “newspaper,” “newspaper ad,” and “newspaper classified.”

5.2 Question Categories

Each question in the survey was assigned a question type. Results from the survey can be broken down to show how these question types compare to one another.

The eight question types are as follows:

- Ad Hoc: Non-generalized questions that provides answers specific to the question (e.g., items that are green)
- Characteristics: Characteristics of a specific object (e.g., characteristics of a good life)
- Goal Driven: Question that list items to achieve a goal (e.g., how to advertise)
- Relational: Relational questions ask the subject to list items that are related to a category that is a bit more vague than Taxonomic (e.g., a tool)
- Taxonomic A : List items of a specific category (more universal, e.g., color)
- Taxonomic B : List items of a specific category (less universal, e.g., mythical being)
- Thematic : Questions that ask about items associated with a situation (e.g., a party)

Sorting these question types by the average number of answers a subject provides for each question shows that 5 of the 8 categories have averages between 7.00-7.99 answers. The outliers are the Taxonomic-A category at 9.34 answers on the high end, and Characteristics at 6.24 and Relational at 5.34 on the low end.

In general, American answers had the highest averages, ranging from 6.57 to 11.63. Indian answers were next, ranging from 4.95-8.72, while Chinese answers averaged from 4.41 to 7.53.

The three outlier types were consistent across the nationalities, as Taxonomic-A was the highest average for all three nationalities, and Characteristics and Relational were 7th and 8th as well.

Within the breakdown for each nationality, the overall results for both American and Indian participants were very close to the same order. This was not unexpected, as together they comprised around 2/3rds of the overall survey population. Results obtained from Chinese participants were mildly interesting, as Goal Driven was their 2nd highest average (6th for All) and Taxonomic-B was 6th (2nd for All). This is only mildly interesting as the difference between 6th (5.66 answer average) and 2nd (5.85) is negligible.

Looking at the speed in which the answers were provided, the results mirror the average answer count results. Overall Taxonomic-A was the fastest at 4.89 seconds per answer and Relational was the slowest at 8.29 seconds. Reaction times for Americans were the fastest, ranging from 3.96 to 7.03 seconds, Indian reaction times ranged from 5.28 to 8.4 seconds, and the reaction times of Chinese subjects averaged 5.93 to 10.33 seconds.

The only differences in ranking the question types occurred between Thematic and Ad Hoc. Thematic question types had the third most answers, but were 5th in answer speed, while Ad Hoc

question types were the reverse, having the fifth most answers but the third fastest answer speeds.

6.0 SYMBOLS, ACRONYMS AND ABBREVIATIONS

ACRONYM	MEANING
2D	Two Dimensional
3D	Three Dimensional
711HPW/RHC	711th Human Performance Wing Warfighter Interface Division
711HPW/RHXM	711th Human Performance Wing Human Analyst Augmentation Branch
AFRL	Air Force Research Laboratory
CDRL	Contract Data Requirements List
CSV	Comma Separated Values
ISR	Intelligence, Surveillance, and Reconnaissance
JUNG	Java Universal Network/Graph Framework
NWN	Neverwinter Nights
OMAR	Operator Model Architecture
PDAT	Polyhedral Dynamics Analysis Tool
TO	Task Order
WIRTO	Warfighter Interface Research and Technology Operations

7.0 GLOSSARY

Concept	An object made up of nodes and edges can be identified with a unique ID. Typically a concept will have a name (i.e., Beagle), and will be visible on the graph in its own color.
Edge	An object that connects two nodes. In a concept, all nodes (attributes) of the concept will connect to each other through an edge. An edge can have a weight assigned to it, signifying the strength of the bond between the nodes that it connects.
Layout	A mechanism for returning (x, y) coordinates for nodes.
Matrix	An X by X grid of data that establishes the connection of a concept and its nodes. The matrix can also define concepts and nodes.
Node	<p>An attribute that makes up a portion of a concept. It has an ID, and typically will have a name (i.e., Brown). It is displayed on the graph as a filled circle.</p> <ul style="list-style-type: none">a. Cut. A node that when removed, produces a graph with more connected components than the original graph.b. Leaf. A node that has one edge.c. Initial. A node which a directed edge starts at.d. Isolated. An unattached node which has no edges.e. Normal. A node that has both incoming and outgoing edges.f. Sink. A node which has no out edges. All edges connected to this node end at this node.g. Source. A node which has no in edges. All edges connected to this node start at this node.h. Terminal. A node which a directed edge ends at.