INDEPENDENT STUDY FINAL REPORT

Assessing Validity and Usefulness of Automated Personality Generation Through LTA Analysis in State Sim

Vibhor Nigam

(M.S.E. Robotics, University of Pennsylvania)

Adviser: Prof. Barry. G. Silverman

(Dept. Electrical and Systems Engineering, University of Pennsylvania)

(Director, Ackoff Collaboratory for Advanced Systems Research, ACASA)

Prof. C. J. Taylor

(Dept. Computer and Information Science, University of Pennsylvania)

ABSTRACT

This report focuses on the scope of LTA analysis when carried out on leaders with limited quantity and quality of data available as compared to the leaders for whom a large amount of data is available in the public domain. It also focuses on the scope of using the personality traits generated by the LTA analysis directly into our tool State Sim.

During the analysis it was found that while LTA analysis was accurate in determining the traits for leaders of largely democratic and stable countries for whom data was readily available on internet, it wasn't so accurate in determining traits of leaders of a country where there is instability or leaders have dictatorial nature and interview responses are limited. In case of insurgents or militant leaders results obtained through LTA analysis of their speeches available online were highly inaccurate and could not be used.

It was analyzed if the results obtained via LTA analysis could be directly used in State Sim and due to significant differences between the values it was decided against it.

INTRODUCTION

The goal of this study is to determine the accuracy and validity of performing automated personality generation of world leaders by using LTA analysis through the tool Profiler Plus as well as automating the process of feeding the data and extracting the results from Profiles Plus to enhance the speed of decision making process in our current tool State Sim. This study will also focus on the usability of results obtained from Profiler Plus directly into State Sim.

Profiler Plus is a general purpose text analytics system which provides fine grained control and complete process transparency to allow for a customized and individualized analysis process. It is coded to perform analysis with three major personality analysis themes i.e. Leadership Trait Analysis, Verbs in Context System for Operational Code, Conceptual/Integrative Complexity and Verbal Behavior Analysis.

State Sim is a tool developed by ***Ackoff Center for Advancement of Systems Approach*** (ACASA), University of Pennsylvania, which

This study will focus upon the results of automated coding versus manually coding for Herman Trait Analysis and will check upon the scalability and scope of using this tool. It will also cover the results for automatically generated personalities versus the manually coded by Subject Matter Experts and benefits to State Sim by automating the process.

ALGORITHM (Herman Trait Analysis)

Herman Trait Analysis is one of the most widely used algorithms for leadership style assessment in the world of political science. The algorithm tries to answer three basic questions pertaining to leadership style assessment i.e. a) How do leaders react to political constraints in their environment – do they respect or challenge such constraints? b) How open are leaders to incoming information – do they selectively use information or are they open to information directing their response? c) What are the leaders reasons for seeking their positions – are they driven by an internal focus of attention within themselves or by relationships that can be formed with salient constituents?

The personality of a leader is decoded in the form of seven traits, Belief in ability to control events, Need for power, Conceptual complexity, Self-confidence, Task orientation, Distrust and In-group bias, each of which are explained below.

Belief in ability to Control events: It refers to the degree of control a person perceives over the situation he/she is in. Coding for belief in control over events focuses on verbs. It is based on an assumption that action proposed by a leader indicates belief in control over events. ***A score on this trait is determined by calculating the percentage of time the verbs in an interview response indicate that the speaker or a group with whom speaker identifies has taken responsibility for planning or initiating an ac*****tion***.

Conceptual Complexity: It is defined as degree of differentiation which the author shows in describing or discussing other people, places, policies, ideas or things. Coding for conceptual complexity focuses on particular words that suggest the author can see different dimensions in the environment. ***The score for conceptual complexity is the percentage of high and low complexity words in any interview response that suggest high complexity***. Overall score for a leader is his average score across interview responses.

Self-Confidence: It defines author's sense of self importance, or image of his or her ability to cope adequately with objects and persons in environment. Coding for self-confidence focuses on pronouns "my", "myself", "I", "me" and "mine". ***A score on this trait is determined by calculating the percentage of times these personal pronouns are used in an interview response that meet the three criteria's, How important do they see themselves as being to what is happening?, Does the use of pronoun reflect that leader is instigating an activity or is the recipient of a positive response from other person or group?***.

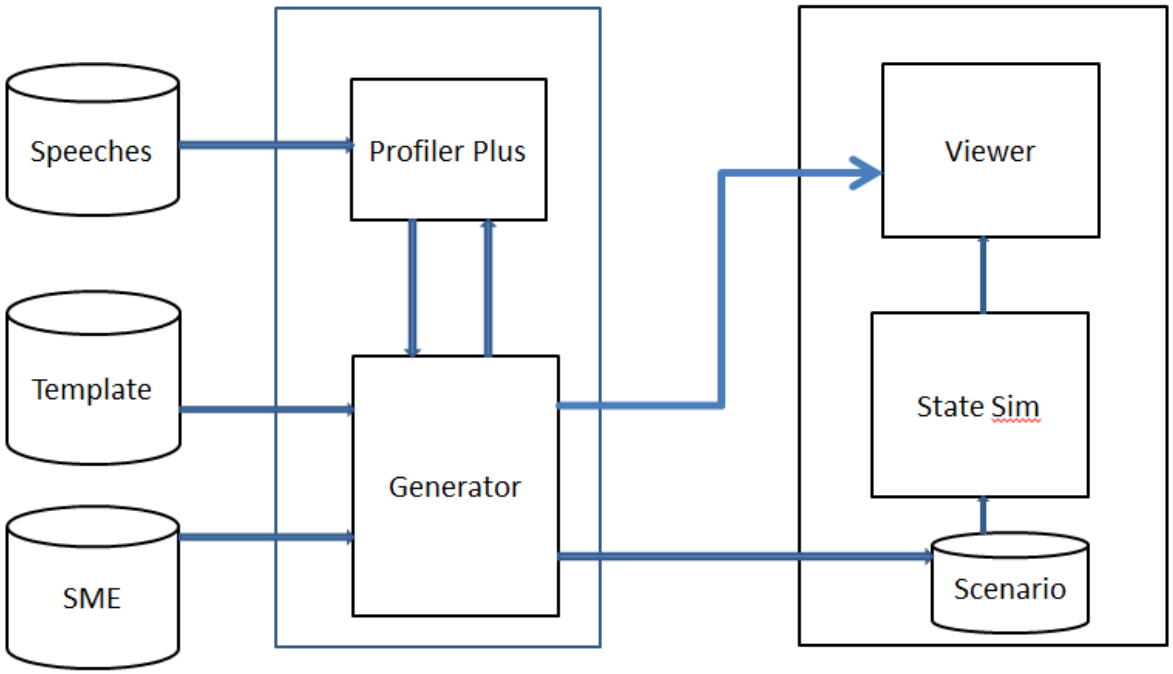
Task Orientation: It defines the author's relative emphasis on interaction with others when dealing problems as opposed to focusing on the feelings and needs of relevant and important constituents. Coding focuses on words that indicate work orientation as well as words that center on concern for another's feelings, desires and satisfaction***. Score for this trait is calculated as percentage of task-oriented words relative to the total number of task and group-maintenance words in a particular interview response***. Total score is the average of the scores across all interview responses.

Distrust: It defines the wariness about others or degree of authors inclination to suspect the motives and actions of others. Coding for distrust focuses on references persons other than the leader and to groups other than those with whom the leader identifies that convey distrust, doubt, misgivings or concern about what these groups are doing. ***Score on this trait is calculated as the percentage of times in an interview response that he or she exhibits distrust towards other groups or persons***. Overall score is average of responses across interviews.

In-Group Bias: It is defined as a view of the world where one's own group holds center stage is perceived as the best, and/or there are strong emotional attachments to this in-group. Coding focuses on words or phrases focusing on authors own group. ***Score for this group is calculated as percentage of times in an interview that a person refers to his or her own group***.

Need for Power: It defines the degree of authors concern for establishing, maintaining, or restoring one's power or in other words, the desire to control, influence or have an impact on other persons or groups. Coding focuses on words where author, proposes or engages in a forceful or strong action, gives not required advice or assistance, attempts to regulate behavior of others, endeavors to impress or gain fame with an action or is considered with his or her reputation or position. ***Score for this is calculated as the percentage of times the verbs in an interview response indicate a speaker has engaged in one of the described behaviors*.**

ARCHITECTURE OF STATE SIM AND GENERATOR



The above diagram explains the architecture of Generator and how various components i.e. Profiler Plus, Generator and State Sim interact with each other.

INTERACTION BETWEEN PROFILER PLUS AND GENERATOR

As can be seen through the diagram there exist a 2 way relation between Profiler Plus and Generator. Since Profiler Plus is being used to modify the existing GSP trees thus Speeches and Generator both act as an input source for Profiler Plus. Output of Profiler Plus is used to modify the GSP trees extracted via generator which are then in turn fed back to the Generator thereby satisfying the 2-way relationship

GENERATOR

Generator is the interface designed to create new user scenarios and to explore and modify the existing scenarios present in State Sim. It is a GUI Interface which provides an SME with options of selecting pre- existing groups and leaders and reuses them to create new scenarios.

PROFILER PLUS

Profiler Plus is a tool developed by Social Science and automation which has been integrated into the Stae Sim and Generator to create accurate GSP trees of leaders. Default GSP trees already exists for leaders but they need to be modified according to the leader being profiled. Initially, these changes were made with consultations from Subject Matter Experts.

With Profiler Plus it is an attempt to automate the process with same level of accuracy by obtaining few traits of leaders through Herman Trait Analysis. We only need to provide text speeches into the profiler plus and result can be inherently used in the State Sim generator while creating scenarios.

GSP Trees

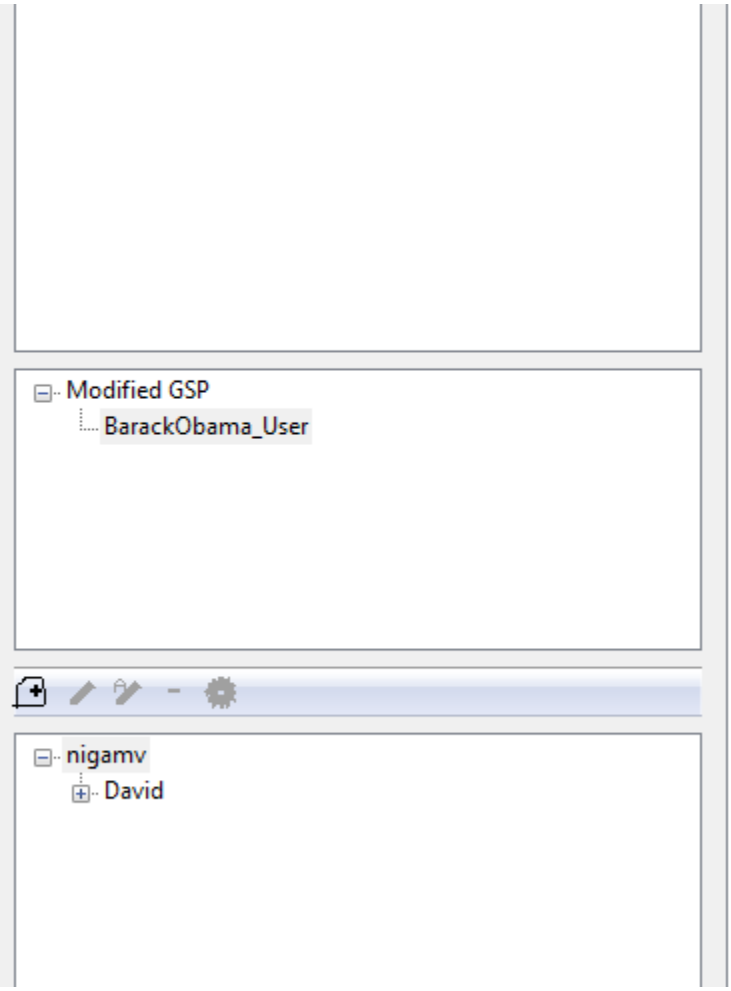
GSP trees are the most integral part of State Sim. They are way of creating a person virtually by creating its personality and thereby helping us determines his/her way of thinking. It divides a person according to his or her Goals, Standards and Preferences. State Sim while simulating a scenario takes into account the personality of leaders and followers of a particular groups and regions along with their economy and resources. These are important while running simulations and then predicting results of a conflict in State Sim.

USER SCENARIOS

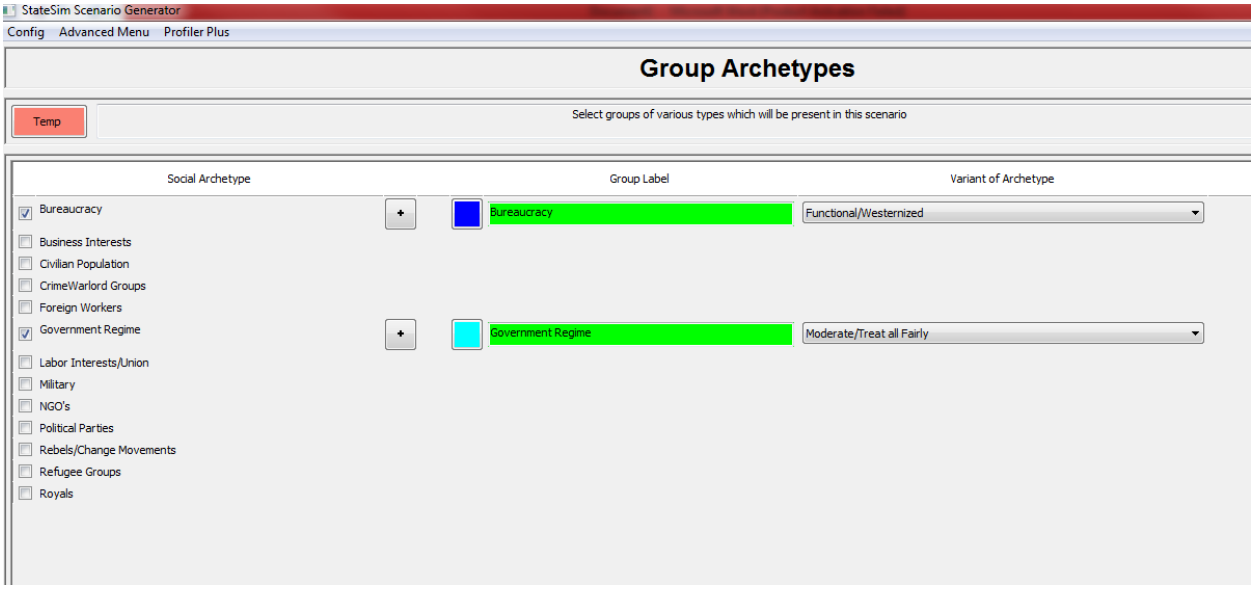
User scenarios refer to the scenarios that a user can create by combining pre- existing groups, leaders and the modified leaders which can be created through Profiler Plus. It works on the assumption that basic characters of a leader or a dictator remain same throughout with some changes which the culture and society bring upon the person. Similarly all new conflicts can be created by re-using the groups and leaders in the existing scenarios.

This approach helps in saving a lot of time and effort and in quick creation and working on the new scenarios which otherwise may require a lot of time and effort of the SME's. To make the process further easier a simple page by page flow in generator was provided to create new user scenarios.

To create a new user scenario we start with the small plus button in the user scenario panel.

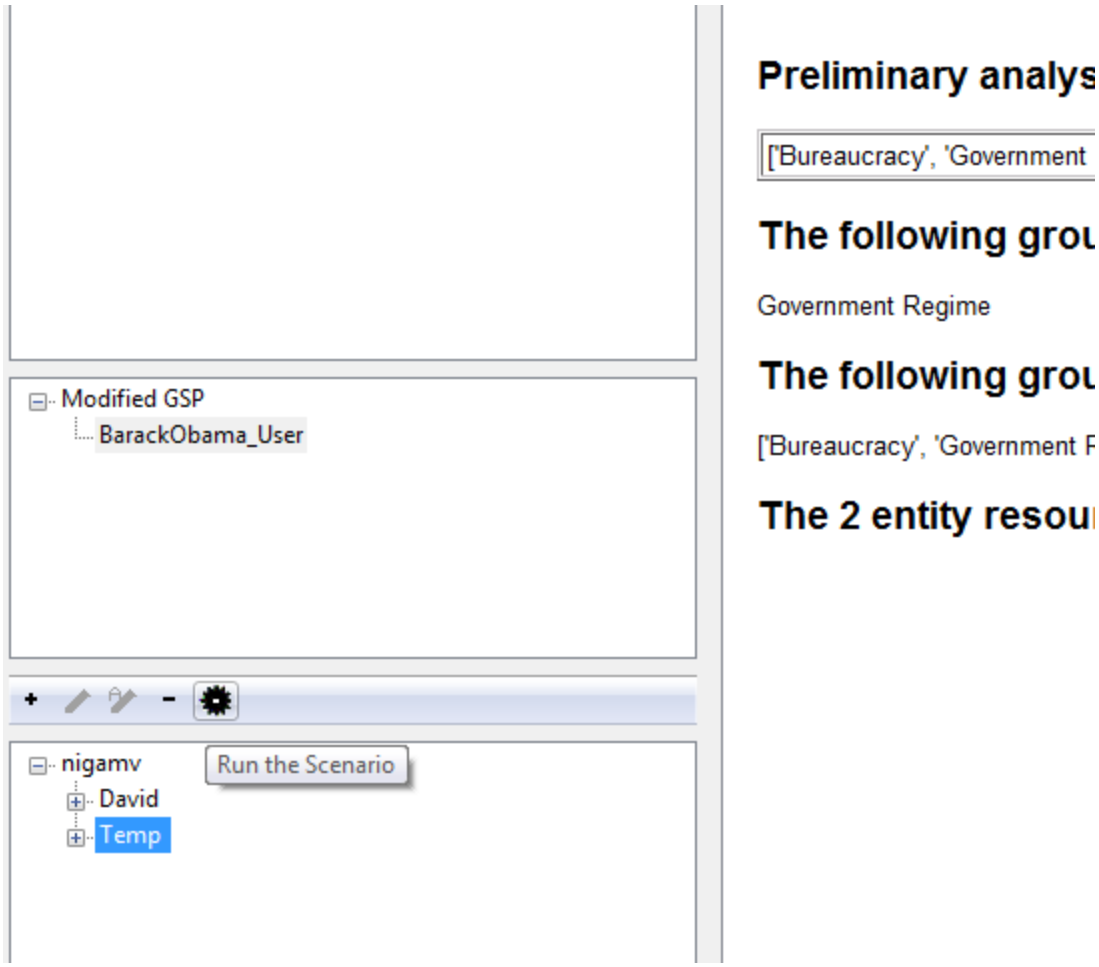


It opens up a dialog which takes a user through various steps of user creation one by one. Some of the Snapshots of the process are described below.

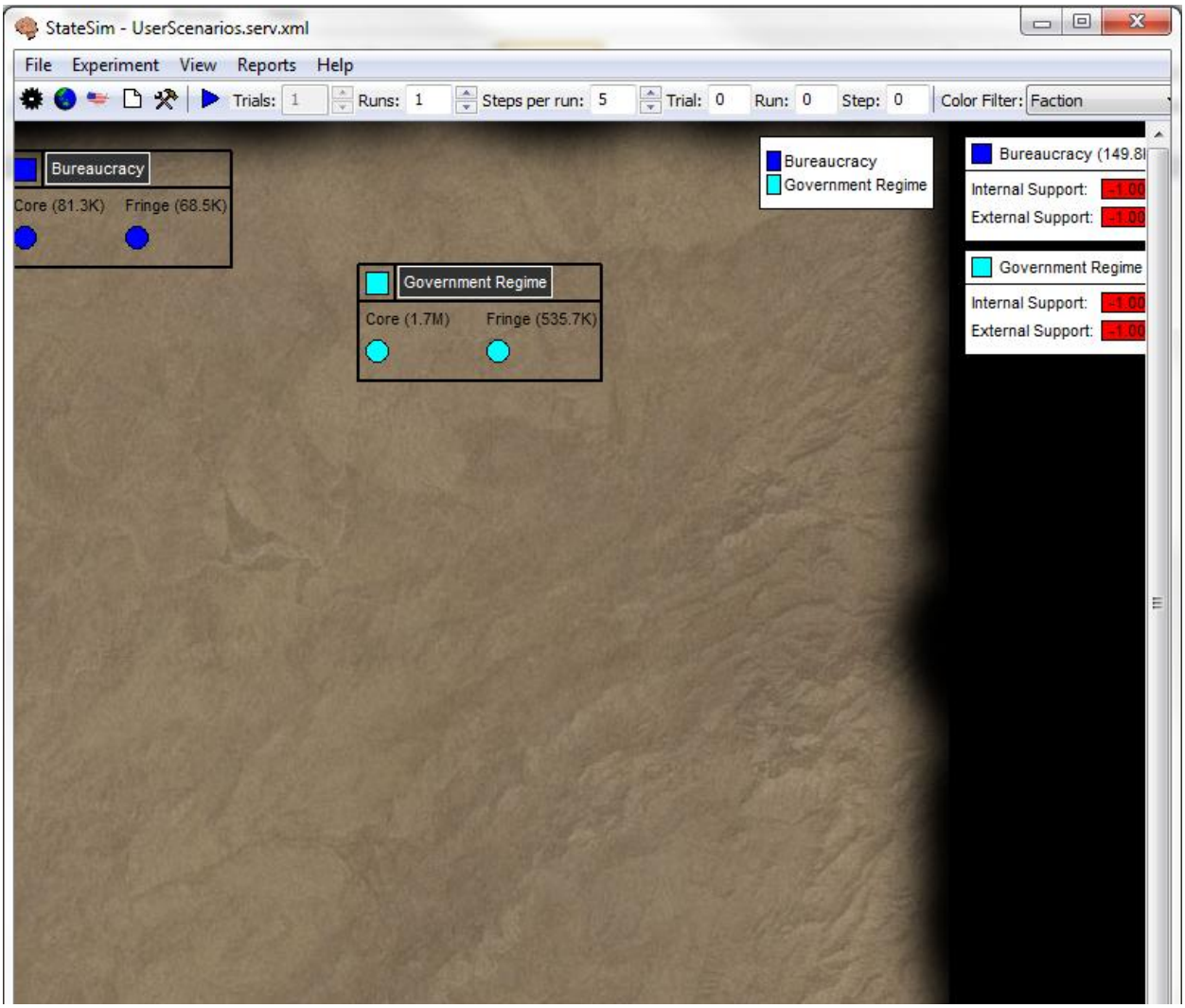


STATE SIM VIEWER

Once the scenario is built it can be opened in a State SIm Viewer via Run the Scenario Button.

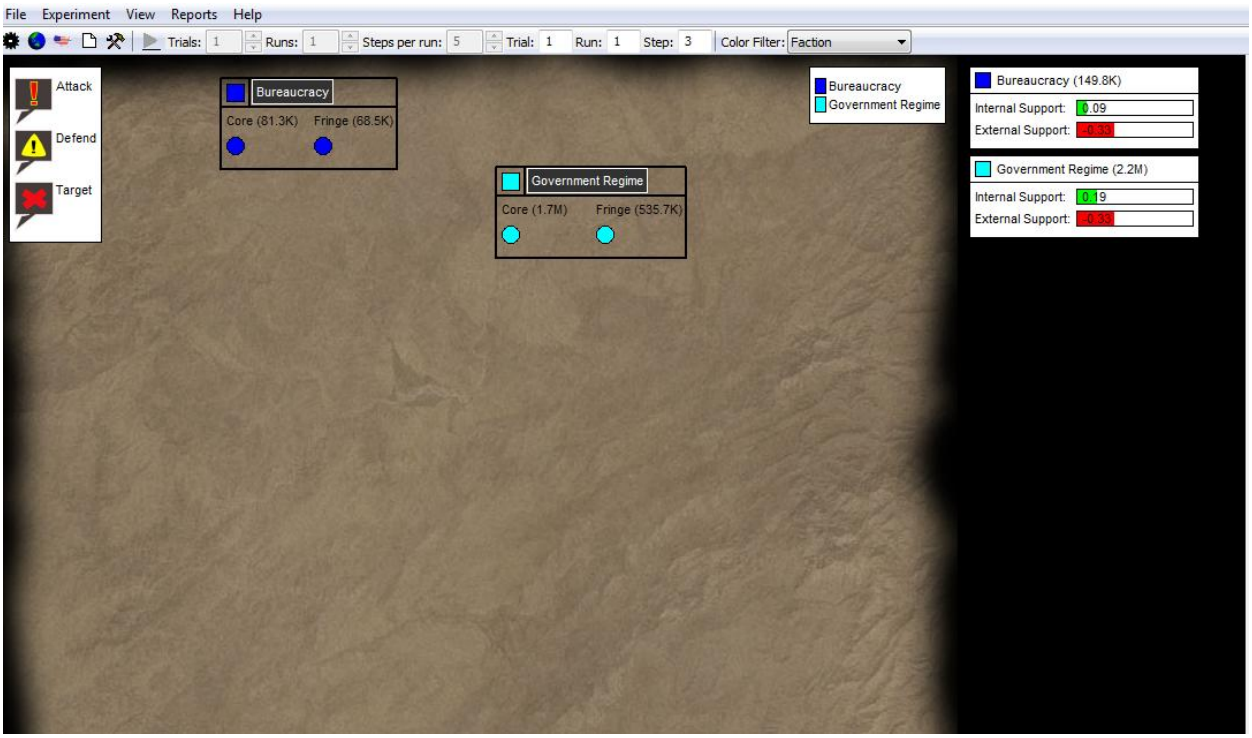


It opens up the scenario in State SIm Viewer through which various simulations can be run and results can be seen on the scenario.


StateSim - UserScenarios.serv.xml
File
Experiment
View
Reports
Help
Trials: 1
Runs: 1
Steps per run: 5
Trial: 0
Run: 0
Step: 0
Color Filter: Faction
Bureaucracy
Core (81.3K)
Fringe (68.5K)
Government Regime
Core (1.7M)
Fringe (535.7K)
Bureaucracy
Government Regime
Bureaucracy (149.8
Internal Support:
External Support:
Government Regime
Internal Support:
External Support:

SIMULATIONS & RESULTS

Once the simulations are run on the scenario the changes and results can be seen.



For instance here the internal support for Bureaucracy increases to 0.09 and to Government Regime increases to 0.19.

OVERVIEW OF PROFILER PLUS

Profiler Plus is a general purpose text analytics system which provides fine grained control and complete process transparency to allow for a customized and individualized analysis process. It is coded to perform analysis with three major personality analysis themes i.e. Leadership Trait Analysis, Verbs in Context System for Operational Code, Conceptual/Integrative Complexity and Verbal Behavior Analysis.

Profiler Plus is implemented in LISP and is entirely rule based application in contrast to most statistically-based NLP systems. Words or strings of words present in a document are matched against a pre coded dictionary containing words/ string of words which indicate towards a

particular trait. A count of such words or strings of words are kept which is then used further to calculate scores for a particular trait based on the rules described in Herman Trait Analysis.

It performs text analytics in a highly flexible, multiple pass process and other than government is currently being used to support broader applications in the areas of media analysis, campaign and election media evaluation, athlete achievement, profiling and forensic psycholinguistics.

For the purpose of our study we used the desktop version of Profiler Plus 7.3 which was easily integrated with the State Sim through batch files and could be accessed through State Sim. Outputs were available in easy CSV formats which could be easily read to carry out further analysis.

Since the personality is extracted from the responses of a leader a special attention needs to be given in keeping the responses as natural as possible. For this purpose a higher weightage is given on the interview responses rather than on speeches as there is a higher chance of speeches being tailor made to suit the occasion or being written by professionals thus inadequate in articulately reflecting the personality of the leader.

To get an accurate estimate responses stretching over a period of 5-6 years are collected. For accurate results minimum 50 responses of each 100 words or minimum 5000 words of responses need to be collected. A larger data set of responses might be beneficial but provides similar results.

USER INTERFACE

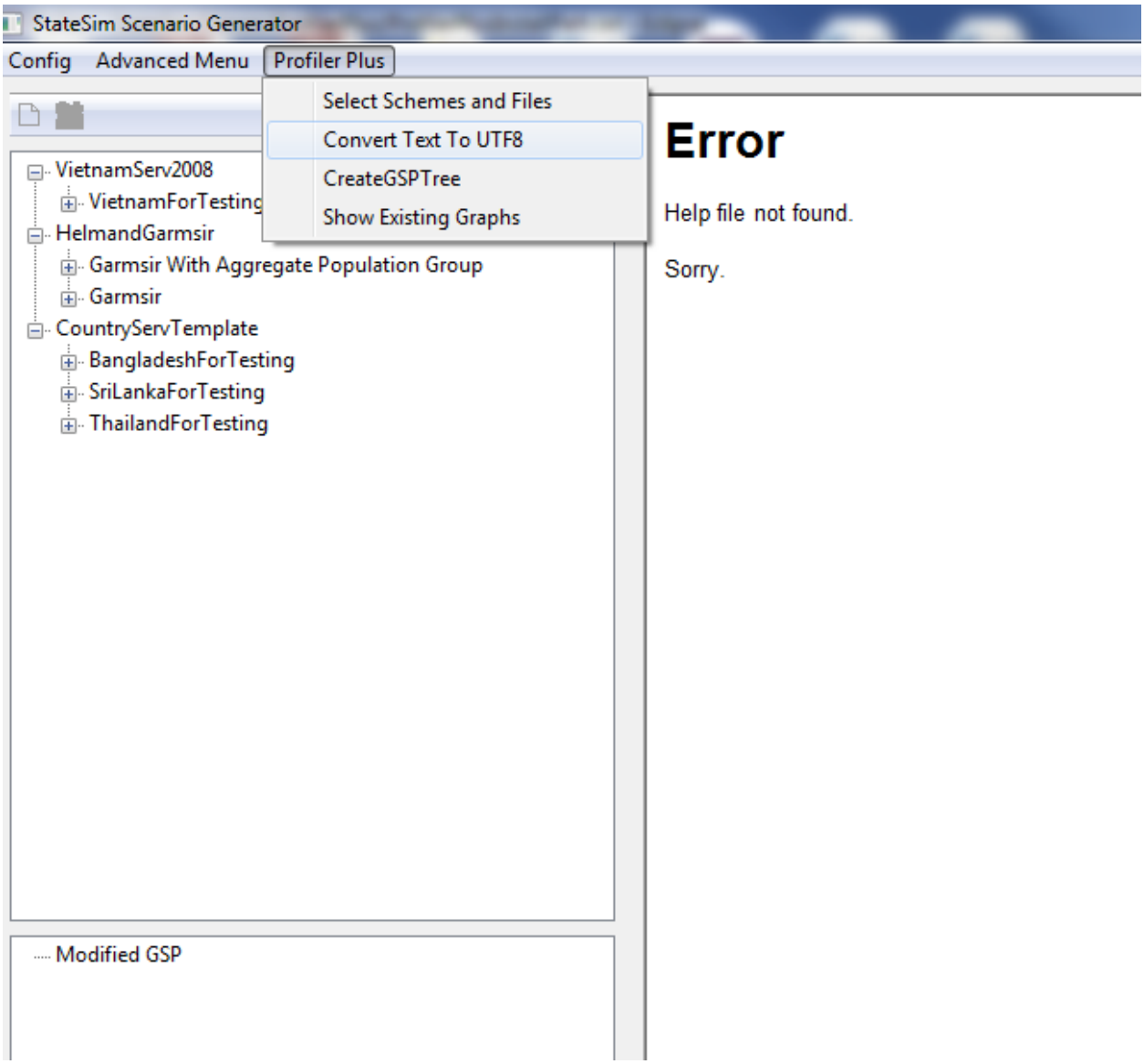
To integrate profiler plus efficiently with our tool and to make it easily accessible through State Sim a simple user interface was designed. This user interface is described below in great detail.

The collected speeches are to be put at Viewer/ProfilerPlus/"Name of Leader", for instance if we are collecting speeches of Barack Obama the path for his speeches should be Viewer/ProfilerPlus/BarackObama. Speeches should be saved in the form of a text file and should be kept to ASCII character set.

For the interface to work properly both installation path and result path of ProfilerPlus needs to be specified at Viewer/ProfilerPlus/ProfilerPlusInstallPath and Viewer/ProfilerPlus/ProfilerPlusResultPath respectively.

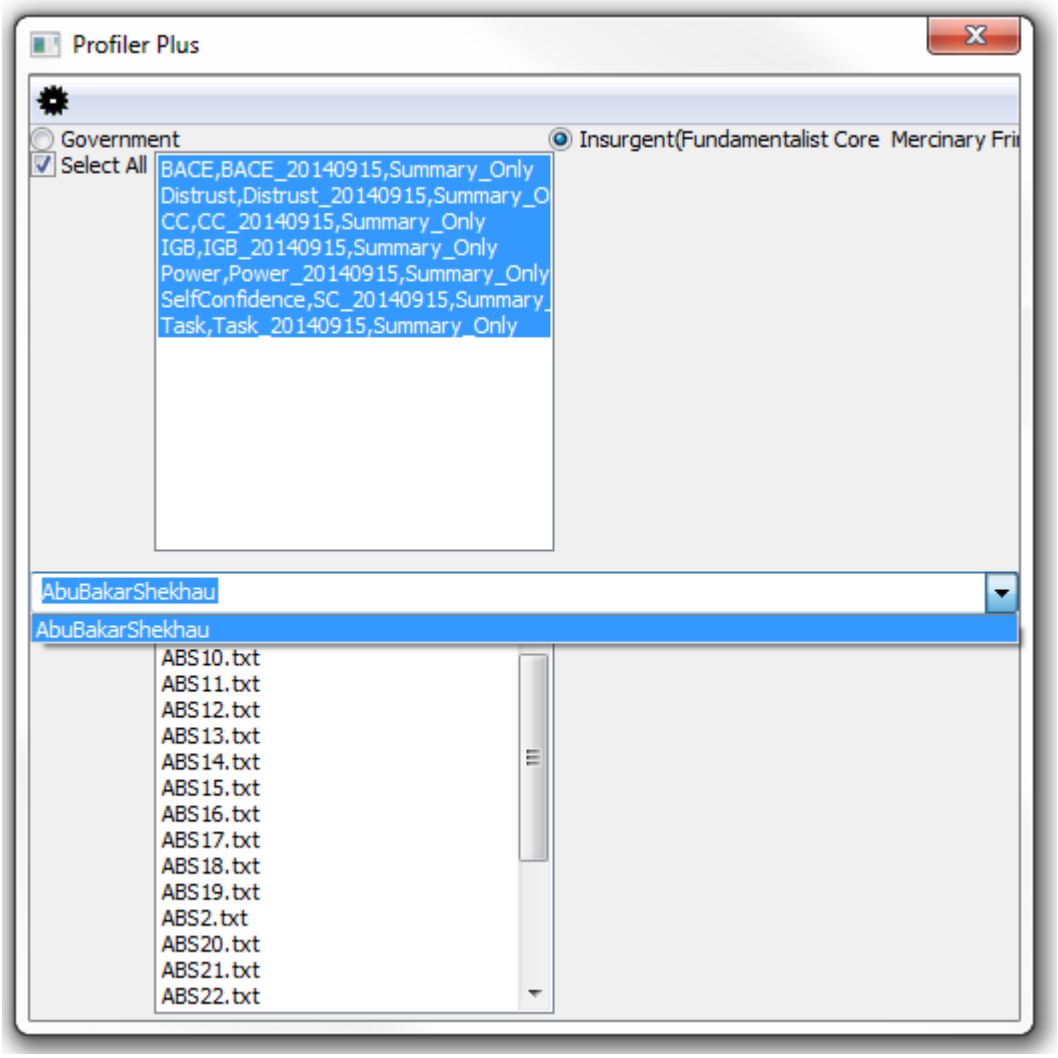
A drop down menu is provided in the State Sim generator specifically to handle actions related to Profiler Plus. The menu provides with actions to convert the collected text files into suitable encoding for Profiler Plus, for running Profiler Plus schemes on the selected files, for modifying the GSP trees according to values generated according to the traits and storing them under the Modified GSP trees section, and for re-visiting the graphs generated for various leaders on whom the analysis was done during Profiler Plus.

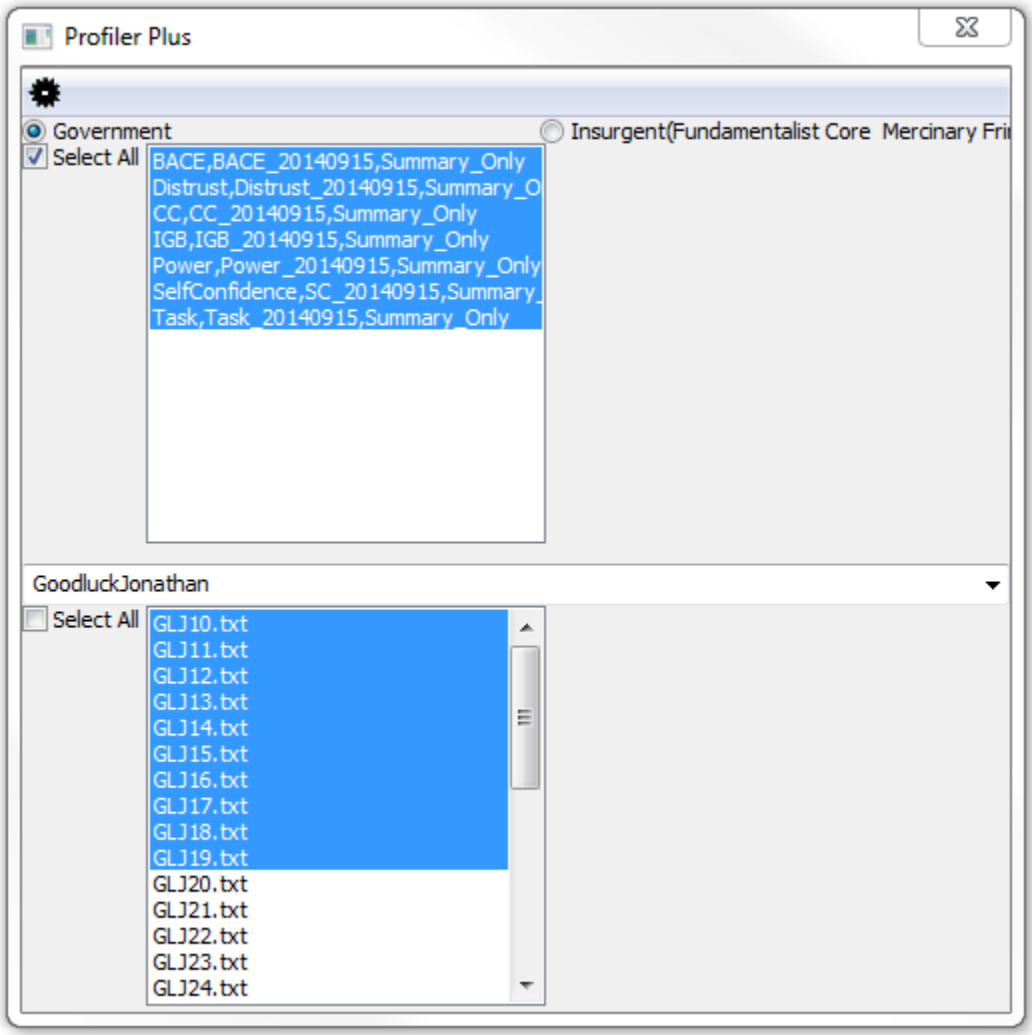
Once State Sim generator is open for converting the collected speeches from text to UTF-8 encoding “Convert text to UTF-8” is selected from the drop down menu.



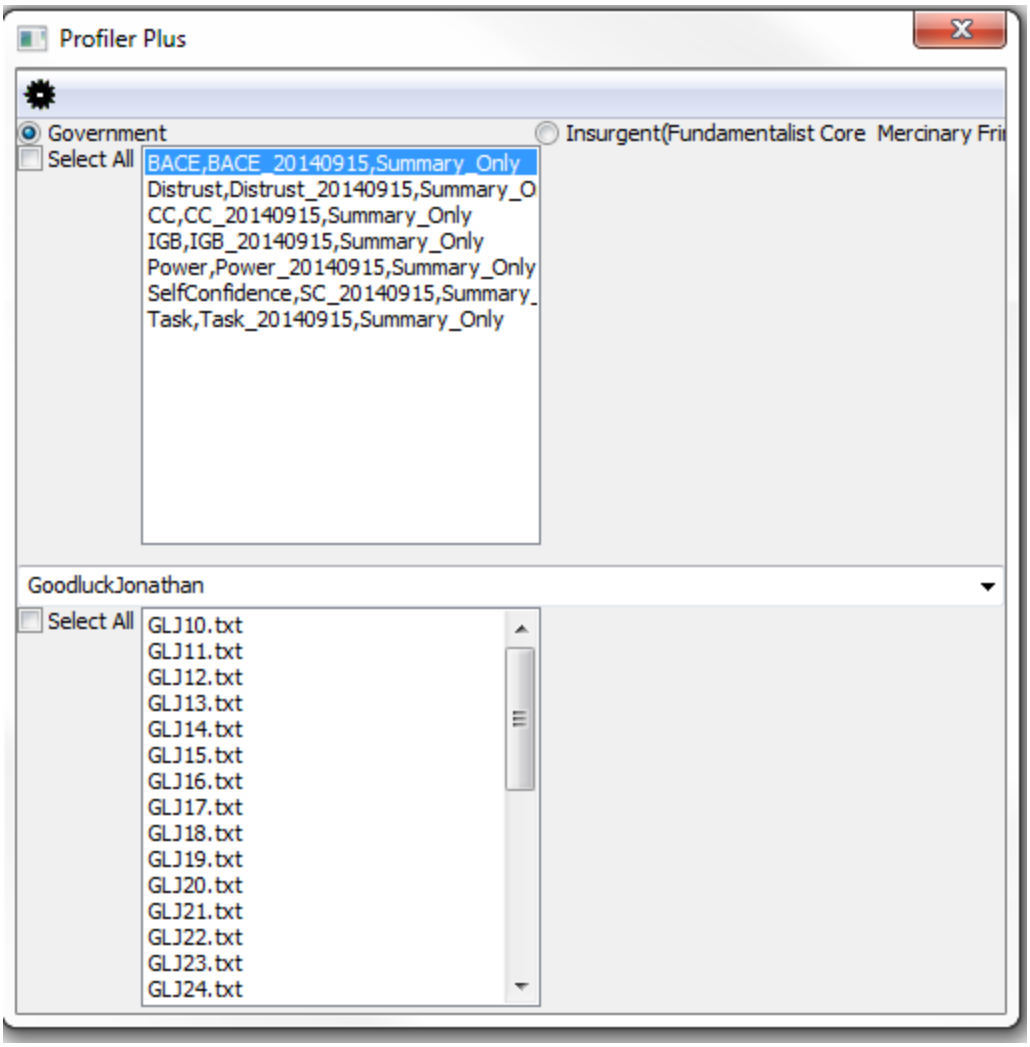
The action ends with a dialogue box displaying the message “Files Converted” which indicates that all files have been successfully converted to UTF-8 encoding.

The second action is “Select Schemes and Files”. This action opens up a dialogue box which enables the users to select the leaders broadly divided into two categories of Government and Insurgents and then select their speeches and the schemes that one wishes to apply on these speeches to perform an analysis.

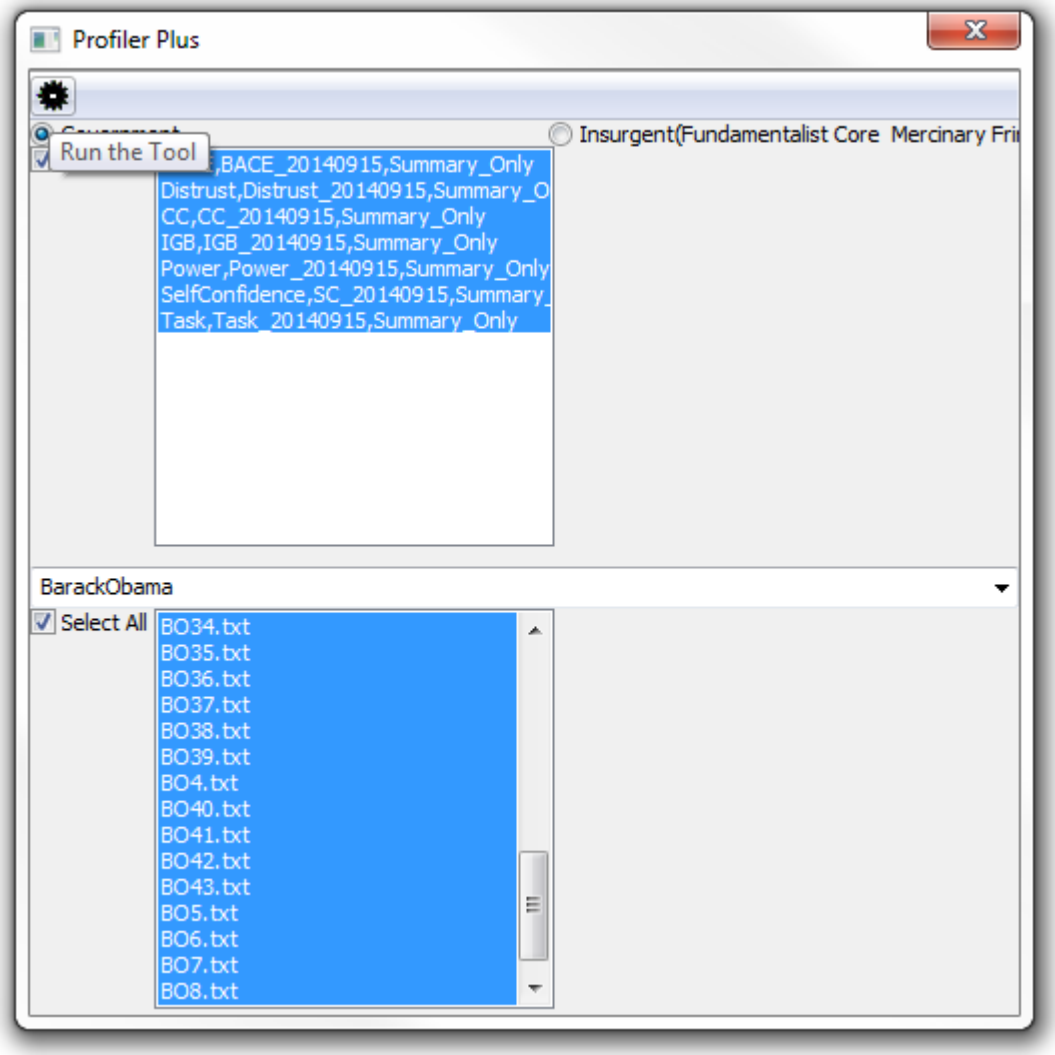

Profiler Plus
Government
Insurgent(Fundamentalist Core Mercinary Fri
Select All
BACE,BACE_20140915,Summary_Only
Distrust,Distrust_20140915,Summary_O
CC,CC_20140915,Summary_Only
IGB,IGB_20140915,Summary_Only
Power,Power_20140915,Summary_Only
SelfConfidence,SC_20140915,Summary_
Task,Task_20140915,Summary_Only
AbuBakarShekhau
AbuBakarShekhau
ABS10.txt
ABS11.txt
ABS12.txt
ABS13.txt
ABS14.txt
ABS15.txt
ABS16.txt
ABS17.txt
ABS18.txt
ABS19.txt
ABS2.txt
ABS20.txt
ABS21.txt
ABS22.txt



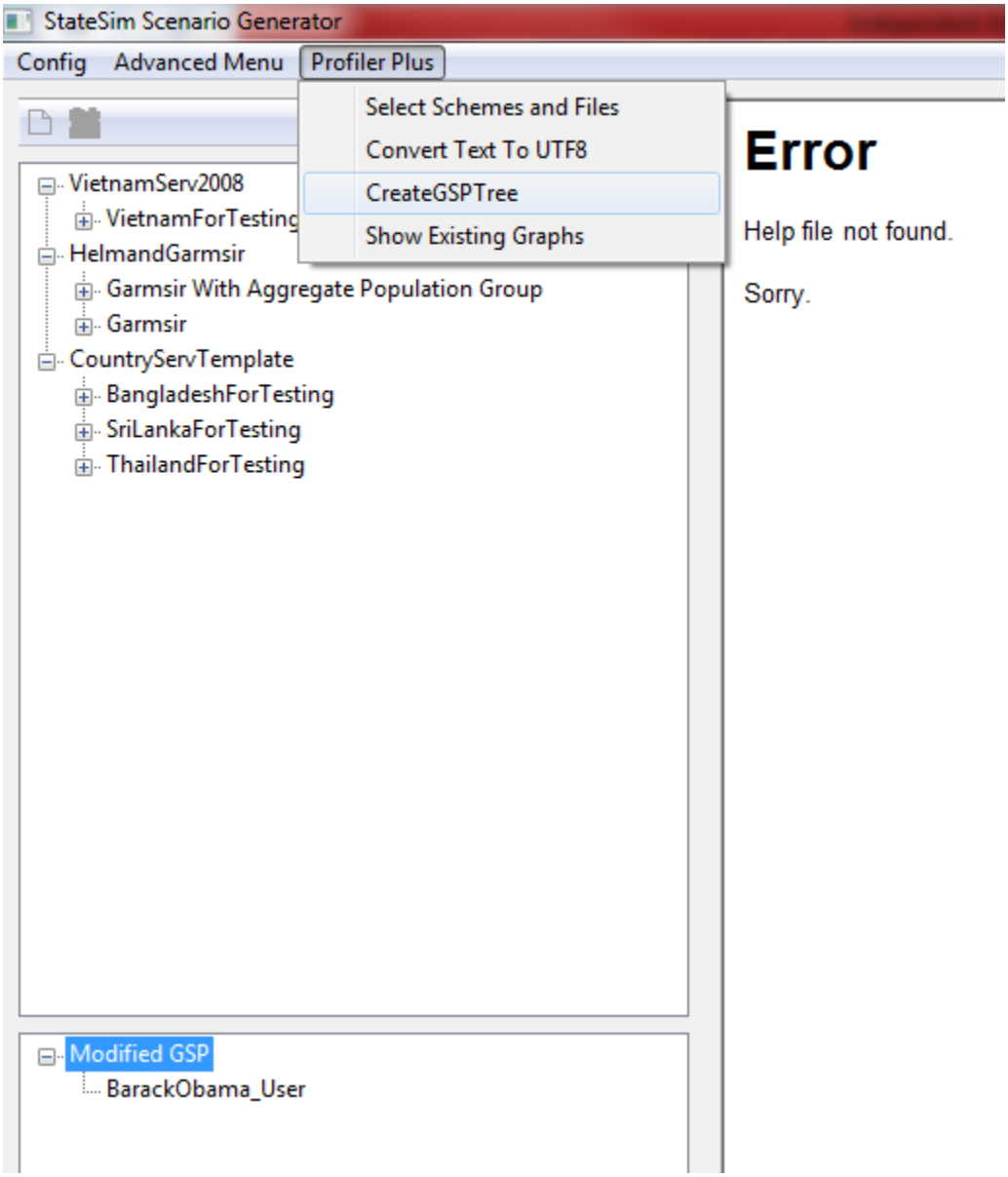
Select all button selects all the schemes/files available for the leader. A single or multi selection can also be made available by individually selecting the files or schemes.



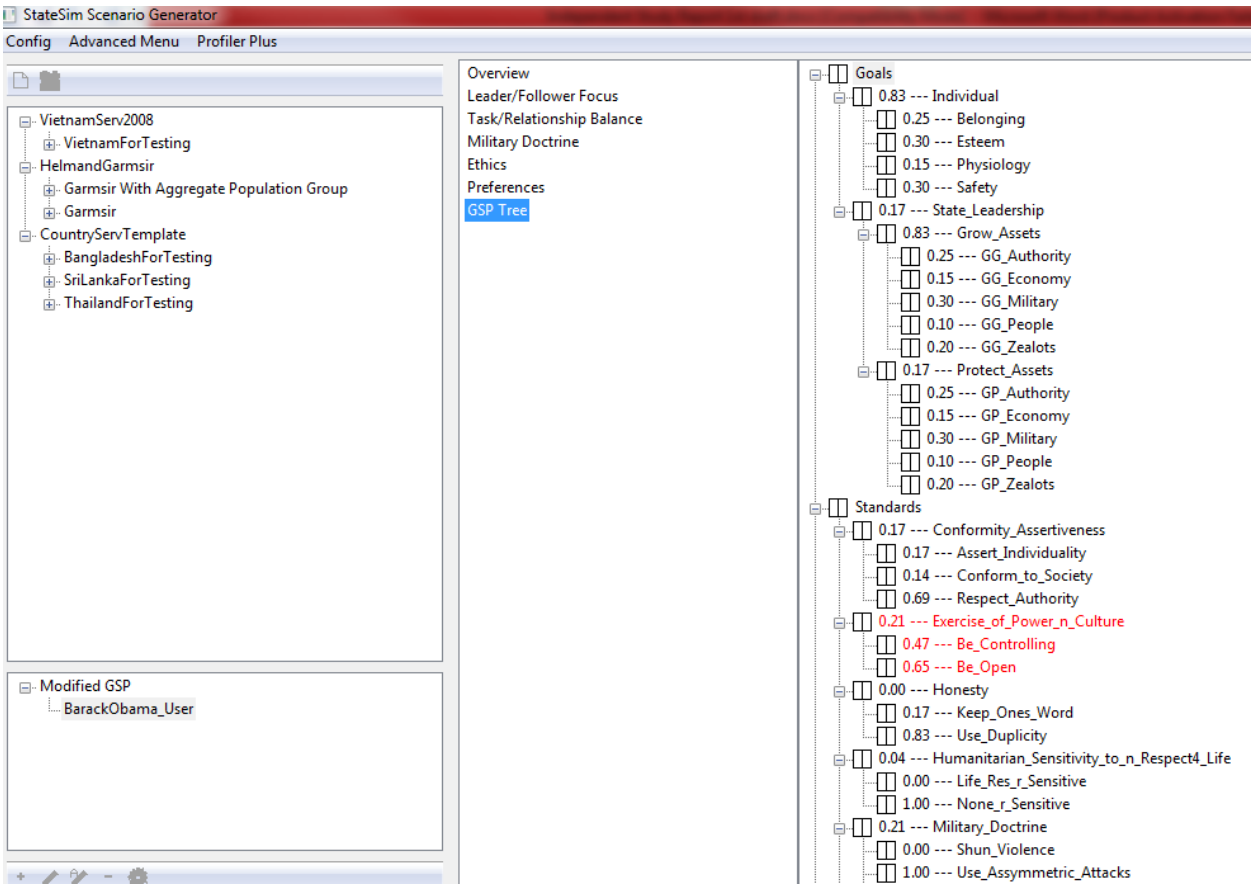
This interface provides a simple and flexible choice to the user to select between the schemes and files on which the user wants to run Profiler Plus. For running the Profiler Plus through this interface “Run Tool” button is used.



To create modified GSP tree of the leader the command “Create GSP Tree” needs to be fired. It modifies the default GSP tree of leader and creates a GSP tree of the current leader according to analysis.



The modified GSP tree can be accessed by double clicking on the GSP Tree label. For instance GSP tree of Barack Obama can be opened by clicking on “BarackObama_User”.



APPROACH

The whole study was divided into two main phases, **data collection and data cleaning**, and running the algorithm to get the results. Data collection is the most important step. Since the personality is extracted from the responses of a leader we want the responses to be natural. Generally, speeches of a leader are written by others and tailored and modified according to the occasions. Interview responses on the other hand, though prepared, are more natural and reflect the personality more accurately. A careful attempt was made to create a mix so as to avoid a bias, and also the responses were collected from over a period of 5-6 years so as to measure the effect of circumstances on a leader.

Once, the data was collected it was then cleaned and converted to 'UTF-8' encoding. Once the data was cleaned it was fed into the profiler plus through the designed interface described above and results were obtained in the form of an excel file. Scores for the seven traits were then calculated from these files as described in the algorithm.

To get the best responses minimum of 50 responses are required with each response being of 100 words i.e. minimum of 5000 words of responses were taken to create a profile of the leader.

LIMITATIONS

One of the major limitations realized during the study was the amount and quality of the data available for a person. While the amount was adequately enough for the leaders of democratic countries like U.S., U.K., Germany, India, etc., availability was a big issue for the countries where freedom of press is curbed or countries are very small like those of Nigeria, Thailand etc. Getting interviews of leaders of these countries was extremely difficult, as they seldom gave interviews. Even, when they gave interviews there was a high level of preparedness in it.

Another, major problem was translation from a foreign language to English. Since, the data collection also required a pre-screening, thus all sources could not be tapped for the process of data collection. Only those interviews could be tapped which were given to leading English news channels and whose transcript were available.

RESULT

The study was carried out on two sets of leaders. One set consisted of leaders like Barack Obama, Bill Clinton and Mahatma Gandhi, those who's speeches and interviews are easily available. The other set consisted of leaders like Thaksin Shinawatra, Khaleda Zia, Goodluck Jonathan and AbuBakarShkekhau whose speeches and interview responses are limited. Scores obtained for these leaders through profiler plus were then matched with scores given to these leaders by Subject Matter Experts.

Table1: Results of LTA Analysis on Various Leaders

	World Leaders	**Barack Obama**	**Bill Clinton**	**Khaleda Zia**	**Goodluck Jonathan**	**Mahatma Gandhi**	**Abu Bakar Shekhau**	**Thaksin Shinawatra**	**GSP Tree Gov Leader**	**GSP Tree Insurgent Leader**
Trait	**Mean**	**Scores**								
Belief in ability to control events	**0.35**	0.48	0.5	0.31	0.38	0.29	0.42	0.3	0.5	1
Conceptu	**0.59**	0.65	0.55	0.62	0.6	0.59	0.52	0.7		

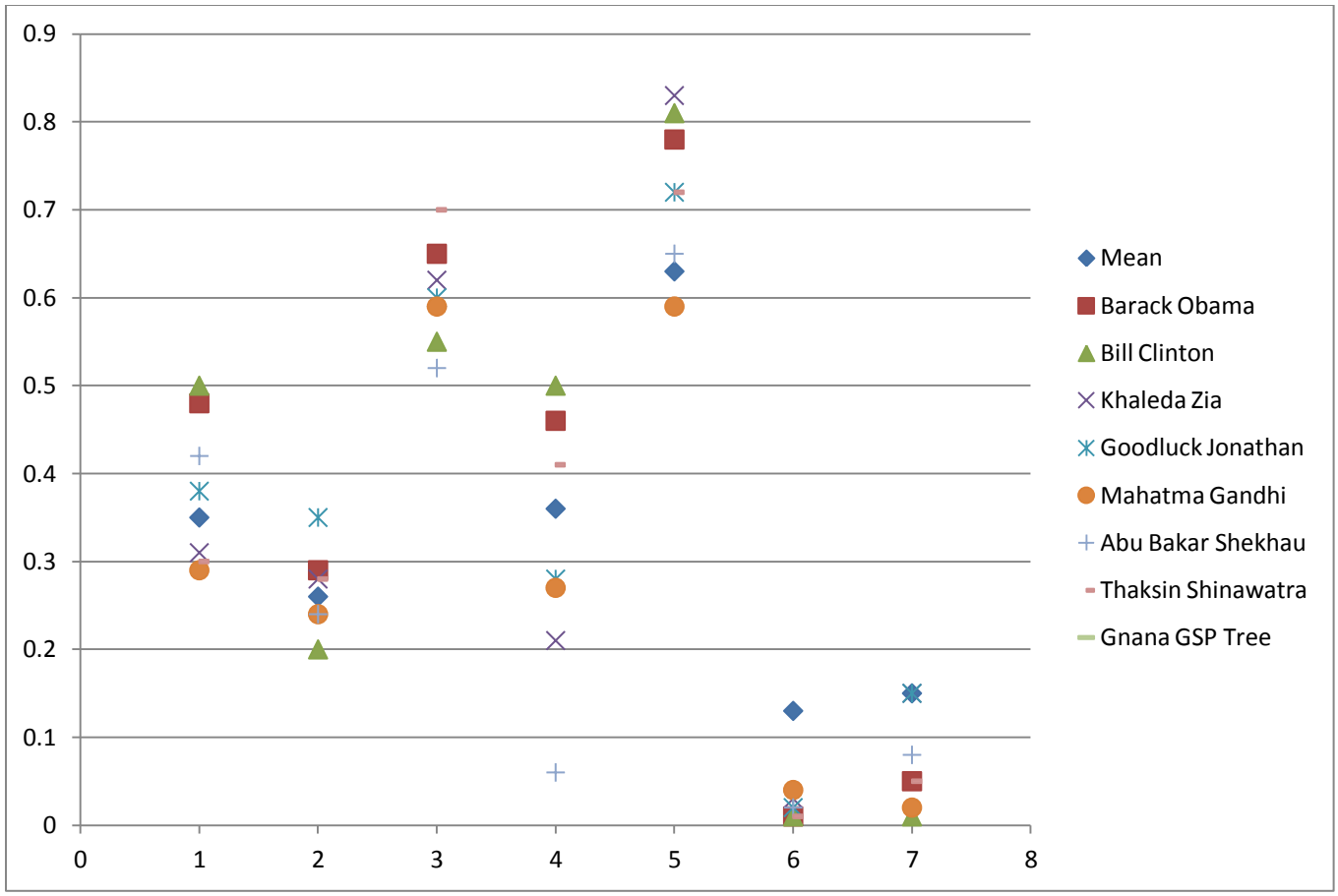
al Complexity										
Need for Power	**0.26**	0.29	0.2	0.28	0.35	0.24	0.24	0.28		
Self-Confidence	**0.36**	0.46	0.5	0.21	0.28	0.27	0.06	0.41		
Task Orientation	**0.63**	0.78	0.81	0.83	0.72	0.59	0.65	0.72	0.67	0.33
Distrust of Others	**0.13**	0.01	0.01	0.03	0.02	0.04	0.02	0.01		
In-group Bias	**0.15**	0.05	0.01	0.15	0.15	0.02	0.08	0.05	0.35	0.67

Explaining results for Barack Obama: Analyzing the results for Barack Obama we found that both his scores for "***Need for Power***" and " ***Belief in ability to control events***" is higher than mean. These scores indicate that such a person challenges constraints, is skillful in both direct and indirect influence, knows what he wants and takes charge to see it happen.

Scores for both "***Conceptual Complexity***" and "***Self-confidence***" are higher than the mean with conceptual complexity score being greater than self-confidence score. This indicates an open personality, where a person has his/her own set of beliefs to which he/she sticks but also is open to the new ideas and concepts. However, each new idea is regressed through his own set of beliefs and concepts and is accepted only when the leader is convinced of it being better than his /her individual perception.

Barack Obama scores 0.78 in "***Task-orientation***" as compared to mean of 0.63 which indicates that he is more of a problem solver than a relationship oriented person. His primary focus is on solving a problem rather than developing a consensus with everyone involved.

His scores for both "**Distrust of others**" and "***In-group Bias***" are way below mean. These indicates that a person focuses on taking advantages of opportunities and building relationships. World is not considered as a threatening place by such a leader and conflicts are percieved as context specific and are dealt with on a case to case basis.



Graph Plotting Trait Values of Leaders and the Mean Value for LTA

(On X axis 1- Belief in ability to control events, 2- Need for Power, 3- Conceptual Complexity, 4-Self-confidence, 5- Task orientation, 6-Distrust of Others, 7- In Group Bias)

Results obtained in this independent study throw up an interesting observation. While the results for the leaders whose data were readily available in form of interviews and speeches matched to the general perception of the leaders and were accurate, results for leaders of lesser known countries or for whom data were not easily available threw up some completely unexpected result.

Explaining results for Abu Bakar Shekhau : One specific case to discuss is of the Militant Leader "Abu Bakar Shekhau" of the militant group Boko Haram in Nigeria. While his scores for "***Belief in Ability to Control Events"***(0.42)and "***Task-Orientaion"***(0.65) were somewhat accurate his other scores were completely out of range.

He scored a mere 0.06 in "***Self-Confidence"*** which is a very low score for a person who claims to be running the largest militant group in the country and claims the responsibility for some of

the most dangerous attacks in the country. Similrly, he scored a mere 0.02 on "***Distrust of Others***" and 0.08 on ***" In- group Bias"*** which are even below the mean. Normally for a militant leader these scores are expected to be appreciably above the mean. Similarly, he scored a 0.24 in "***Need for Power"*** which is also below the mean. For a militant leader this score is also expected to be appreciably above the mean.

CONCLUSION

The above results show that while Profiler Plus and LTA analysis were accurate in prediciting results of leaders of democratic countries who indulge in a daily basis dialogue with the basis but is not so accurate in predicting results for dictators and militant leaders with the small amount of data available for them.

In case of leaders of a democratic country when compared with the 4 trait values in our existing models obtained through SME (Subject Matter Experts), 3 were found to be in the range but value for In-group bias obtained through LTA analysis differed significantly from those obtained through SME in case of government leaders.

However, in case of insurgent and militant leaders values differed significantly and the values obtained through LTA analysis could not be directly used in the model. These limitations posed a serious question mark on the usability of Profiler Plus into more complex scenarios and world problems and being a highly benefecial tool in global strategic decision making.

It was observed that in case of scenarios where data was not readily available Profiler Plus was not as effective in providing the values and thus could not be relied upon heavily in strategic decision making and profiling when it comes to militant leaders and dictators. Thus, it was concluded that for the purpose of State Sim obtaining values from LTA analysis is not a viable option.

REFERENCES

1. Margaret G.Herman, Assessing Leadership Style: A Trait Analysis, 1999
2. Stephen Benedict Dyson, Alexandra L. Rayleigh, Public and private Beliefs of political leaders: Saddam Hussein in front of a crowd and behind closed doors, 2014
3. Aaron Menenberg, A psychological profile of Ben Bernanke, 2012
4. Barry G Silverman, Gnana K. Baharthy, G. Jiyun Kim, The New Frontier of Agent-Based Modelling and Simulation of Social Systems with Country Databases,Newsfeeds, and Expert Surveys, 2001
5. Barry G Silverman, Michael Johns, Jason Cornwell, Kevin O'Brien, Human behavior models for agents in Simulators and Games: Part 1: Enabling Science with PMFServ, 2006
6. Barry G Silverman, Gnana Bharthy, Benjamin Nye, Human Terrain Data – What should we do with it?, 2006
7. Barry G Silverman, Gnana K Bharthy, Benjamin Nye, Roy.J.Eidelson, Modelling factions for 'Effects Based Operations': Part 1 Leader and Follower Behaviors