

Project Proposal – CS 6315.501

Team Name: Logic, Proof, and Trust

Team Members:

1. Shreevatsa Hosakere Vittal Rao (sxh163630)
2. Avinash Chandrasekharan (axc166930)
3. Piyush Supe (pgs160230)
4. Aayush Bhandari(axb167830)

Other ideas:

- 1) Using the data from 1433, 1434 which can be used to obtain how many people will be willing to opt for international program exchange, which countries are chosen by popularity, which program is opted for and by what percent of the willing students, how many US and foreign participants are willing to join each program.
- 2) For 2007-08, Migration dataset could be used to compare the migration of people in/out of state or county. The State or County popular in migration could be found out. The data sets are 1146 – 1149. This will be utilized for Department of Treasury.

The following was chosen as the data set intricacies were more challenging to extract necessary information. There was also room for expanding the project further and dealt with data sets that could help obtain a perspective regarding health services offered to veterans.

- 3) Comparing Accidents and its rate of General Aviation and Air Carrier Occurrences Involving Illegal Acts

Dataset: 1575, 1577. Others Air Carriers Dataset can also be included which are under CFR.

Project Information

Type of Project: Custom Project

Project Title: State-wise distribution of Veteran expenses and benefits - FY07 and FY08

Data sets:

- 1) Veterans Affairs Geographic Distribution of Expenses FY07 by State and County
- 2) Veterans Affairs Geographic Distribution of Expenses FY08 by State and County
- 3) Veterans Health Administration 2008 Hospital Report Card - Waiting Times
- 4) Veterans Health Administration 2008 Hospital Report Card - Medical Center Staffing
- 5) Veterans Health Administration 2008 Hospital Report Card - Quality of Care – Populations
- 6) FY08_Veterans_Compensation_and_Pension_by_County

Data Source:

Dataset 1152 – <http://data-gov.tw.rpi.edu/raw/1152/data-1152.rdf> (No. of triples: 44,007)
Dataset 1154 – <http://data-gov.tw.rpi.edu/raw/1154/data-1154.rdf> (No. of triples: 40,654)
Dataset 1204 - <http://data-gov.tw.rpi.edu/raw/1204/data-1204.rdf> (No. of triples: 1118)
Dataset 1205 - <http://data-gov.tw.rpi.edu/raw/1205/data-1205.rdf> (No. of triple: 1257)
Dataset 1208 – <http://data-gov.tw.rpi.edu/raw/1208/data-1208.rdf> (No. of triples: 2,369)
Dataset 1538 – <http://data-gov.tw.rpi.edu/raw/1538/data-1538.rdf> (No. of triples: 53,488)

Tools needed:

Apache Jena Fuseki Server, Google Visualization API's, Gruff.

Expected Results -Visualization (Tentative- This includes all the ideas presented by team and may change):

- 1) To summarize the datasets and present the following results:

- a) Present the average population of each state during the given years (2007,2008)

- b) Present the average expenditure of each state during the given years (2007,2008)
- c) Present average pension and compensation details during the given years.
- d) Present average Medical expenditure, Number of people across different genders getting medical benefits during the time frame.
- e) Present number of people across different age gaps getting medical benefits.
- f) Present number of hospital staff (doctors and nurses) per veteran in every state.
- g) Present average wait time for veterans in every state.

Target Audience:

Department of Veteran Affairs.

Milestones:

- 1) Understand Semantics of Data sets latest by: March 12th
- 2) Write SPARQL Queries for above discussed ideas latest by: March 22th
- 3) Prepare HTML page for visualization of results achieved in step 2 by: April 5th
- 4) Final review and report: April 7th
- 5) Final Submission: April 9th