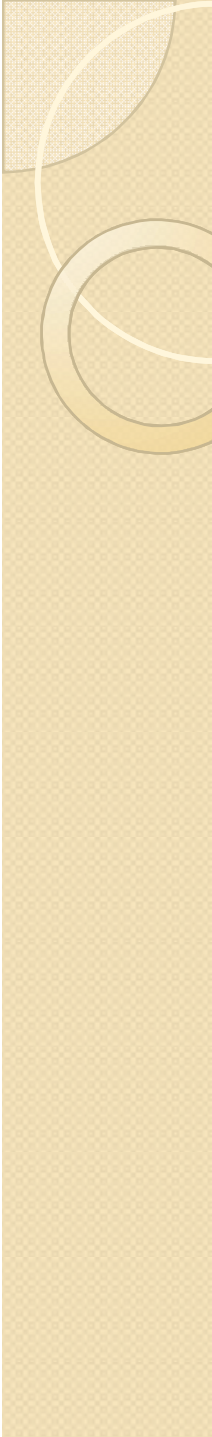




# ACT Data Analysis and Progress on "Three Study Tracks"

by Q. Liu, A. Puthenputhussery, and C. Liu



# Data Processing and NAICS code extraction

- Data Processing of the 647,916 contracts (this process takes about 32 hours to finish)
- First, 6 out of the 12 CSV files need to be manually split into smaller pieces due to computational constraints.  
In total 70 smaller files are created manually, which takes about an hour.
- Second, the XAMPP software package is utilized, where the software PHPMyAdmin converts the smaller files into MySQL format automatically.  
This process takes about 31 hours to finish.



# Data Processing and NAICS code extraction

- Finally, queries are made using the SQL language on the converted files.
- NAICS code extraction that produced .csv file and a pdf file.
- Note that the SQL file provided by Radu/Dan is not complete and it contains only 2 years of the data.



## Study Case 2: Competency Classification

- Out of the 222 attributes, 3 are selected as the important attributes for defining the NOC (narrow objective competence). Specifically, the 3 attributes selected are as follows:  
vendorname  
principalnaicscode  
productorservicecode
- The product, service information and NAICS code are extracted from the attributes "productorservicecode" and "principalnaicscode", respectively, to define the NOC.



## Study Case 2: Competency Classification

- We will further consider the following attributes to be included in the computation of the NOC:

maj\_agency\_cat

fundedbyforeignentity

contractactiontype



# Study Case 3: Risk Visualization

- 99% of the effort is required for Risk Modelling and the Composite Risk Factor Computation. 1% of the effort is needed for visualization in color coding.
- We have created a Risk Analysis Module -- please see our demo # 4.
- Another risk model we are working on is a searchable table that based on user's query of companies displays a relation between the number of government contracts and the annual revenue of the companies in query for each year.



# Study Case 3: Risk Visualization

- The 3rd risk model we are working on is an advanced user-friendly visualization graph that reveals more details to more specific user queries.
- The risk score of a company is related to both the government contracts and the contracts from the private section. In particular, for a specific company, if the number of the government contracts decreases and the revenue decreases accordingly, this means that the company depends on the government funding heavily.





## Study Case 3: Risk Visualization

- Our risk score computation will also incorporate the following attributes:

vendorname

numberofemployees

annualrevenue

verysmallbusinessflag

contractingofficebusinesssizedetermination

dollarsobligated

baseandexercisedoptionsvalue

baseandalloptionsvalue

numberoffersrecieved





# Study Case 1: Commercial Transactions Identification

- "We need CS faculty help to fill this gap in information using various technologies at their disposal."
- Search for the private contracts for the companies in the ACT data.
- The ACT data contains only government contracts, but the contracts with the private sector are not available.  
Maybe a Web Crawler should be deployed to search for such information.



**THANK YOU**