# ***CSCI 572 – Spring 2014 – Assignment 3 – Solr, SIS and Google Maps***

Team Members:
- Abhishek Agrawal agra457@usc.edu
- Vaibhav Mathur vaibhavm@usc.edu
- Hari Haran Venugopal hvenugop@usc.edu

Link to the demo: https://www.youtube.com/watch?v=qk_6JdJeWnE&noredirect=1

Technical Description:
The assignment is done using Solr 4.8 running on Apache Tomcat7. The project is developed as an extension to the first assignment (Tika) and used Tika framework to parse the content from the vault containing PDF files. Solr does indexing of the documents and queries are created on a web page . Results are shown on the map using Google Maps API for virtualization.

General approach used in assignment:
- Geaonames.org dataset is parsed and all the place names are stored with the corresponding latitude and longitudes using Java code on JDK 7.
- Used Tika to parse the vault and create term frequency. 1-grams and 2-grams are created and searched across the list of keywords and list of names in descending order of term frequency.
- Based on any match in descending term frequency of the terms, geo-location is calculated and mapped with the PDF document.
- Solr is installed on Apache Tomcat and uses Servlet container within the server.
- One document per PDF file is added into the Solr for indexing with following fields:
  - ID: Used internally for Solr (unique)
  - File_name: Name of the PDF file.
  - File_Content: Content includes the keywords that are present in the document.
  - File_Lat_Long: Includes the mapped geo location of the file
- The result after indexing is verified using Solr user interface.
- A web page, using HTML, CSS, Bootstrap, is created which makes AJAX calls to the Solr servlet running on Apache Tomcat.
- The web page uses GoogleMaps API by invoking script at https://maps.googleapis.com/maps/api/js?key=AIzaSyBheAQq1Z6dqTD1s5MnMj5uzSx-AwXBWe&sensor=true and uses Marker to point the locations retrieved after making query calls to Solr.
- Last, the results are shown on the GoogleMaps. On click of any marker on the map, it opens up the corresponding PDF file from the server.