CSCI 572 ASSIGNMENT III REPORT
Visual Search and Interaction with Weapons Datasets

Team number: 33
Team members: Gaurav Shenoy, Mahesh Kumar Lunawat, Pramod Nagarajarao, Presha Thakkar, Karthik Manipur Kini

Task 1:

A web front end was developed using HTML, CSS and BOOTSTRAP framework. UI/UX design rules to build a user interactive experience with backend built using python. The user interface was developed keeping in mind with the aim to enable smooth interactions by providing powerful tools to the user. Also the interface is leveraged by D3.js, Banana and FACTVIEW to a dashboard which is developed and a total of six D3 visualizations have been incorporated.

The backend python calls SOLR via a REST call and parses the JSON returned to local files which are grabbed by D3 for visualization.

Task 2:

D3 has a strong data format rules which gives the user to give a feel of great visualization. We ran the queries through SOLR and the JSON data which has been returned is processed to suit the individual needs of the D3 widgets. REST API is done through a GET call to JSON handler and once the JSON is successfully parsed, we run several python scripts to obtain the required. .JSON, .CSV, .TSV and other data formats for D3. The D3 were successfully integrated to UI/UX which represents the SOLR indexed data. Guide on Mike Bostock's gallery was very useful which heled us to understand on loading the data into D3.

Task 3:

We designed our analytics dashboard to accept the query entered by the user along with providing suitable alert box message. The query posed a challenging informational complexity which were based on location and time data that is indexed in SOLR and queries expressed posed a geographical and temporal aspect to them. A project video demo has been presented to explain the queries and their results.

The project was designed keeping in mind with overcoming any cross site scripting that we faced. Questions answered below:

1. How easy to use was D3?

D3 which produces dynamic, interactive data visualization helped us to express data from document present which drives the document from data. Cross origin resource sharing issue was faced in the early stage of the assignment but was The computational overhead of

converting data to a specific format for each D3 was a huge hurdle but the overall design in which D3 is structured is user friendly. The designs were modelled on HTML canvas was awesome regardless of data formatting. D3 was flexible of multiple data file formats could be leveraged ranging from XML to CSV to JSON and the indexed-parsed-crawled data was on similar lane to these formats. The features that D3 provided far outweighed its complexity. D3 provided the great development with detail oriented UI/UX and massive scope for limitless customization which helped us to overcome data explanation challenges. Finally, however the performance of D3.js framework decreases as the data it consumes increases.

Fig.1. Displays the different D3 visualizations used and Fig.2 displays the correlation between SOLR, D3, Banana and FACETVIEW.

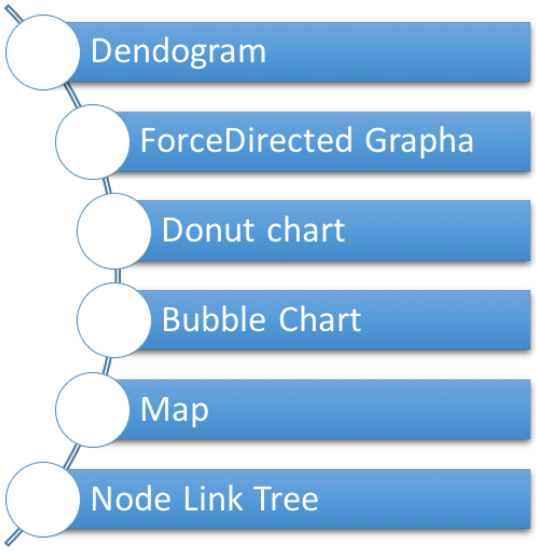

Fig 1 : D3 Visualizations

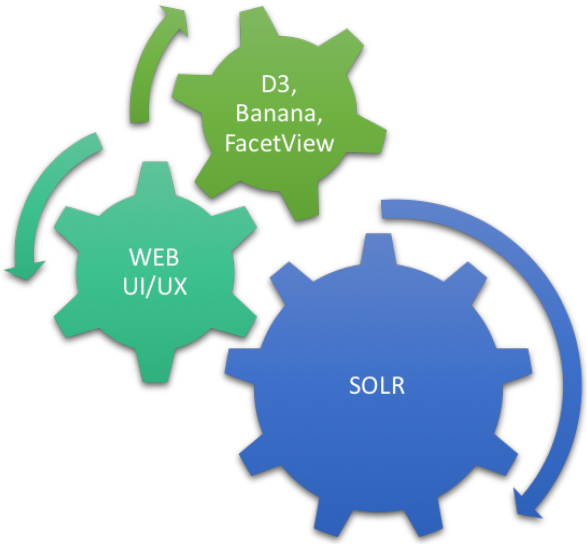

Fig 2 : Correlation between SOLR, D3, Banana and FacetView

2. How easy was it to use Banana?

Since the Banana provided plug and play nature there was no problem with the integration and the overall system was perfect to incorporate this addition. Banana is an open source hence the dashboards were developed with less effort and was simple. By specifying the fields that we want to query we ran the reports on the contents of our index and were able to display the required data. For example, using search bar, filter facets, hit counter, tag cloud, bar chart, time picker generated data for every unique value of the field or if there weren't many unique values as well. We could also plot data vs time and event data such as 'date_created' and 'date_lastupdated'. We could have a number of these panels set up in our dashboard and it could be run on almost any kind of data which was indexed by SOLR. The widgets used are the following:

Fig.3. Displays all the Banana widgets used and Fig.4 represents the overall structure of the Banana system

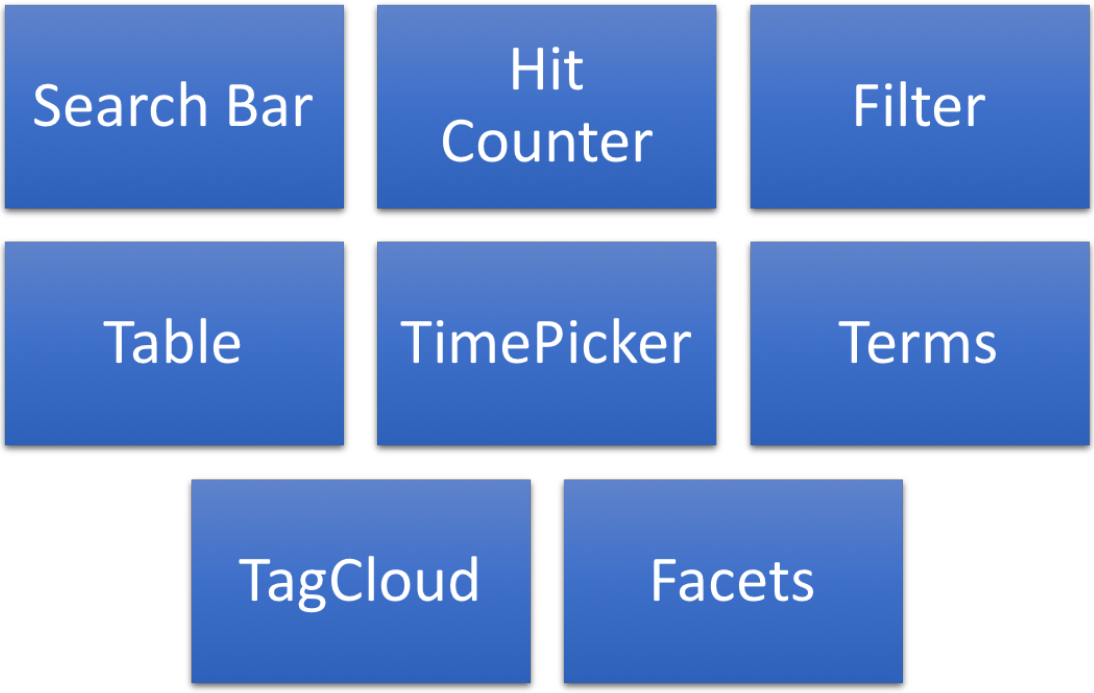

Fig 3 : Various Widgets associated with Banana Analytics Dashboard

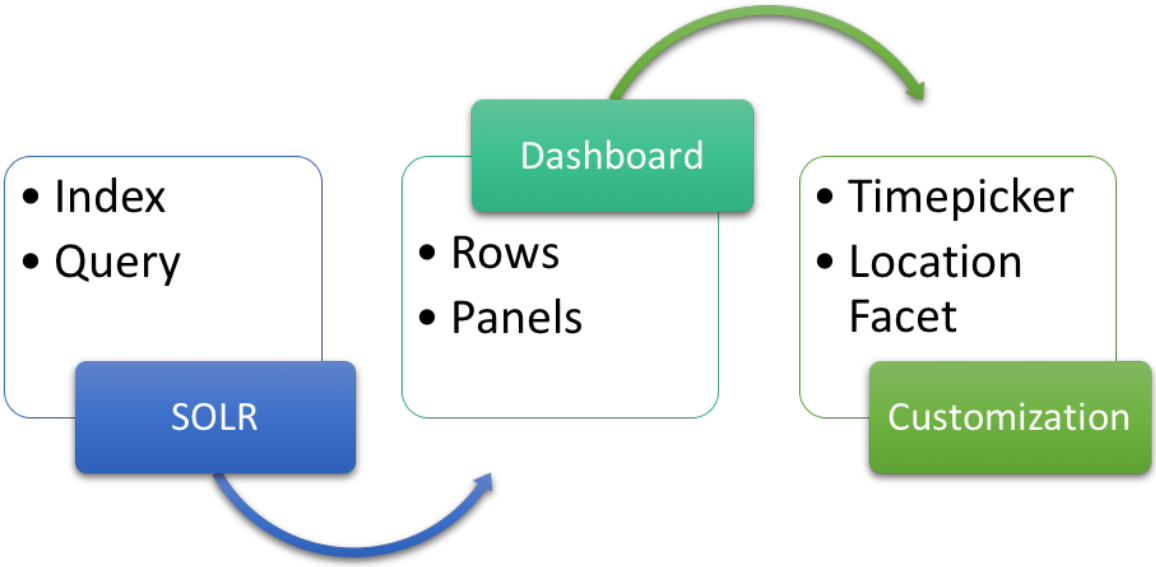

Fig 4 : Architecture of Banana dashboard

3. How easy was it to use FACETVIEW?

Facet view was easy to use because the FACETVIEW implementation makes up the integrating SOLR directly with mere mention of a URL to be used in the 'config' object. As the 'config' object is embedded within the facetview.jquery.js the configuration for displaying the facets is simple. Since adding facets is nothing more than adding facet objects in the 'config' file, it is scalable. It acts as a filter on the search results which helps users in guiding the search. While setting up the FACETVIEW we need to solve the CORS issue which helps to query the SOLR directly from any domain of the client.But even this is made simple since the JARS that have to be added to resolve the CORS issue is already provided in the Github repository. There are multiple options that can be configured with FACETVIEW and also adding multiple facets is a basic feature that is provided by FACETVIEW along with a feature 'subfilters' like sorting. Using the config objects we can change the look and feel of search interface along with functionality. Overall the FACETVIEW is easy to use due to its flexibility of adding facets and linking those facets directly with SOLR.

The Figure 5 displays the overall process of FACETVIEW toolkit working.

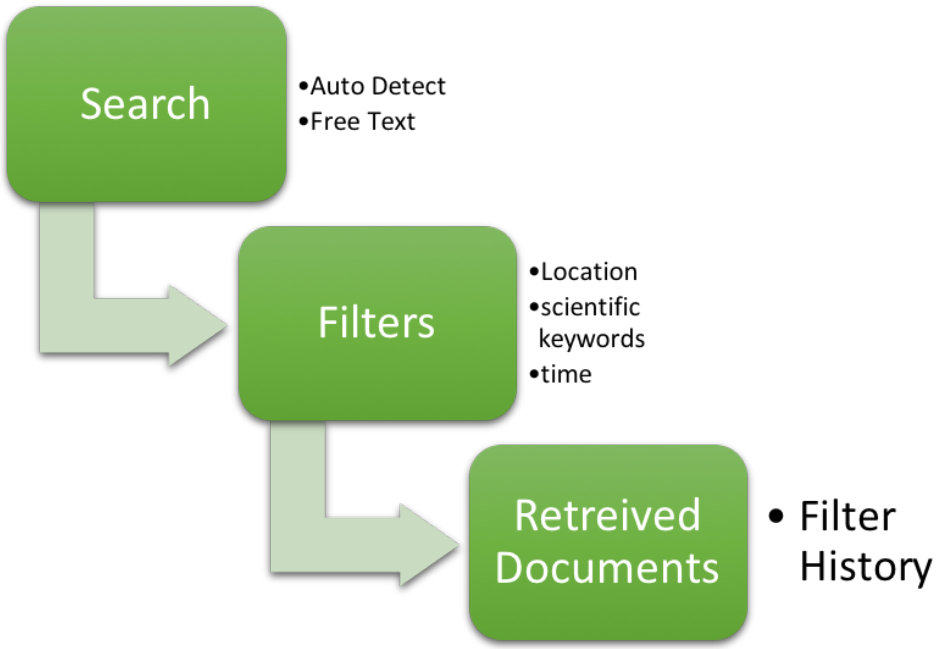

Fig 5 : Working of FacetView

4. What was the hardest part, loading data, or visualizing it?

Various factors considered for the complexity of loading the data and in developing visualizations. Data collection and parsing it was the major step as we had to analyze the data and prune it to fit the requirements of the input structure of visualizations and nature of the source of input was mind boggling in deciding whether to load the JSON directly or should be

consumed from a file. Since all the visualizations did not follow the same pattern, it was a complicated one. Regarding the relevance of data, we had to figure out which visualizations can be most suitable for the data and not all of inputs matched to one single pattern of input format. Loading the data and visualization tasks posed a different problem and both were equally hard and equally easy. One of the great learning experience was to learn about the nature of the data before we visualize it using different techniques and other technologies.

5. Develop capabilities in your Analytics Dashboard to satisfy the following four challenge questions.

Q1. What time-based trends exist in Gun ads? Can you correlate temporal and spatial properties with buyers? For example, can you identify based on ad time-window and/or based on geospatial area places where people try and purchase guns on behalf of someone unauthorized to purchase them?

By running the query, the following are the set of inferences we found:
● On year-month 2015-05-24, in the state of Connecticut, Delaware, Hawaii, Idaho and Utah there were same number of ad sales for Rifles and shotguns. The number of undefined weapons ad sales were significantly large compared to others.

● There were no rifles and shotguns sold in the state of Maine on 2015-05-10, 2015-05-12, 2015-05-13,2015-05-14, 2015-05-17, 2015-05-19, 2015-05-22, 2015-05-24, 2015-05-25 and 2015-05-27. We had boosted “Location/Geo location PageRank field for this query. Minnesota, Missouri and Michigan has more handgun ads than in Montana.

The D3 Visualization used to analyze this query Is a “Node-Link Tree”. Please refer the code to view the same.

Q2. Can you identify similar firearms image types (e.g., shotguns) that are sold in the same region and time? Does this indicate influx related to stolen goods?

The following are the set of inferences we found by running the query:
● On 2015-05-30 same number of shotguns were sold in the state of Kansas and Kentucky.
● The Seller’s name and description are obtained in the execution of the query. The “Description” field is the manufacturer’s name. Though we have the data, we can’t entirely be sure that if an ad query result which doesn’t have the manufacturer’s name can be classified as stolen goods. We yet again boost the geo location PageRank field for obtaining good refined results for the above query.

The D3 Visualizations used to analyze this query Is a “Location” and “Location Based Force Directed Graph”.

Q3. When a shipment of bulk firearms is stolen, the rate of ads and images may indicate an increase in sales of that particular make/model – can you identify these?

The following are the set of inferences we found by running the query:
● The seller by Name “Darin”, who works for “The Armory Company” bearing the phone number “+1-9122851288” sells many different type guns through ads having almost the same “Maintained Conditions”.
● Some of the make/model of the guns sold by “Darin” are “'M&P15-22'”, “G30gen4” etc.
● Some Geo-locations like “Texas” has a surge in the number of the arms being sold/bought during certain period of time (In span of days-months).
● The seller by Name “Joanna”, who works for manufacturer “Plano Pawn Shop” being the phone number “+1-9124246911” sells different models with guns conditions.

The boosted fields for this query are the “Manufacturer” and “Gun type”. More particular results related to the query are obtained.

We used a D3 Dendogram to visualize the data. It has to be noted that we also used Banana Analytics dashboard to view and analyze the query results.

Q4. Can you identify ads and/or weapons images that are posted by persons, whom are underage or in which the weapons are de-identified (by type and/or serial number, etc.)

The following are the set of inferences we found by running the query:
● The Seller with Name “Steve” works for two companies or associated with two manufacturers - “Texas Jewelry Pawn” and “B & S Guns”. In the State of Texas, there are 47 manufacturers that sell guns or subscribed to gun related ads.
● But, underage and unauthorized or weapon being de-identified might not be deducted with the information we have. The Index constructed by our team doesn’t posses the necessary fields to come up

The Seller named “steve” works at more than two-gun manufacturing companies. Steve’s personal details, phone number could be retrieved due to the added boosting. We used Facetview to view the results for the above mentioned query.

Q5. Can you identify ads and/or images that relate to the unlawful transfer, sale, and possession of explosives, WMD devices, and precursors?

The following are the set of inferences we found by running the query:
● This query is one of the most interesting query we came across. When we queried “Chemicals, Nuclear, WMD, Missile etc.”, we obtained results from the index.
● We went one step ahead of re-checking the URL’s obtained. The following URL’s are some of them.

The last query we boost only “keywords” metadata field to 200 times as it is very much relevant.

The boosting of “keywords” field helped us in getting appropriate results. Though the results for “nuclear, WMD etc” were very sporadic. Genuine WMD ads were not detected. We used Facetview to view the results for the above mentioned query.

Conclusion:

The technologies D3.js, Banana and FacetView can be used together to gather interesting insights on the weapons data. The visualizations provide valuable trends which can be used to understand and predict the weapon ads. Efficient use of these resources can go a long way in understanding and interpreting data beyond the realms of traditional search methods.

Video Presentation Link:

https://www.youtube.com/watch?v=Y_JVHGCIzco