

RECALL

Final Project Report

Group 6

James Clark
Tirumala Reddy Konireddy
Jagadish Tirumalasetty
Wang Zhang

University of Missouri - Kansas City

CS 560: Knowledge Discovery and Management
Dr. Yugyung Lee

May 14, 2014

Project Deployment

Github

<https://github.com/CS560/Project>

YouTube

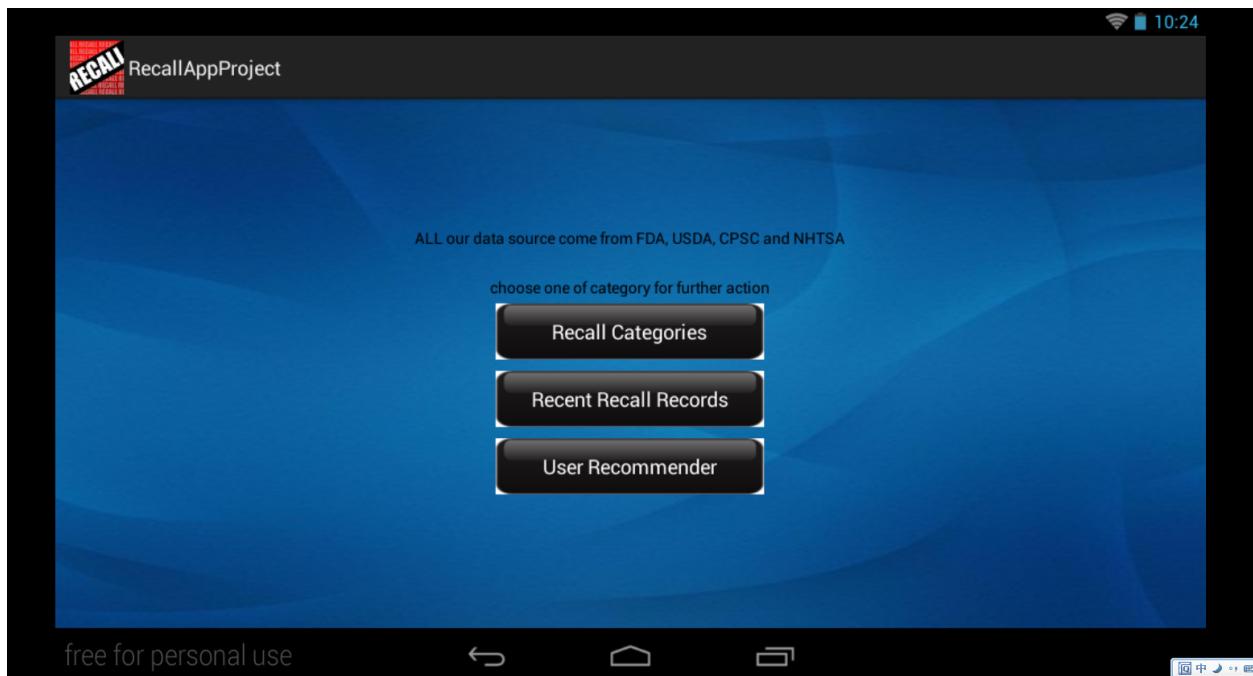
<https://www.youtube.com/watch?v=mlaME5YII4s>

User Manual

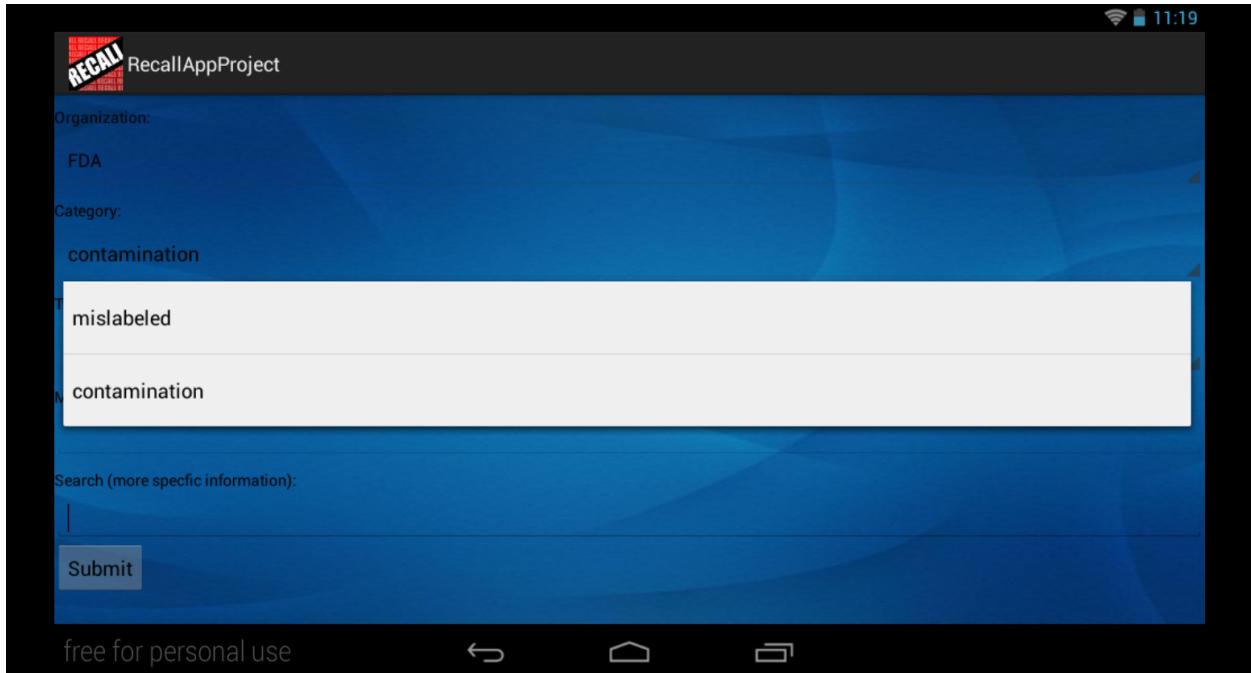
Introduction

The primary motivation of this app is the consolidation, organization, and classification of recall publications. Our recall application will enable user to look up any recall records which are provided by USDA, FDA, CPSC and NHTSA. For every organization, we have different categories and subcategories base on different data sets. Also user will be able to check most recent recall records. The last feature for user is recommendation. Base on what they are subscribe for recall categories, the app will recommend them 3 other categories base on our recommendation algorithm.

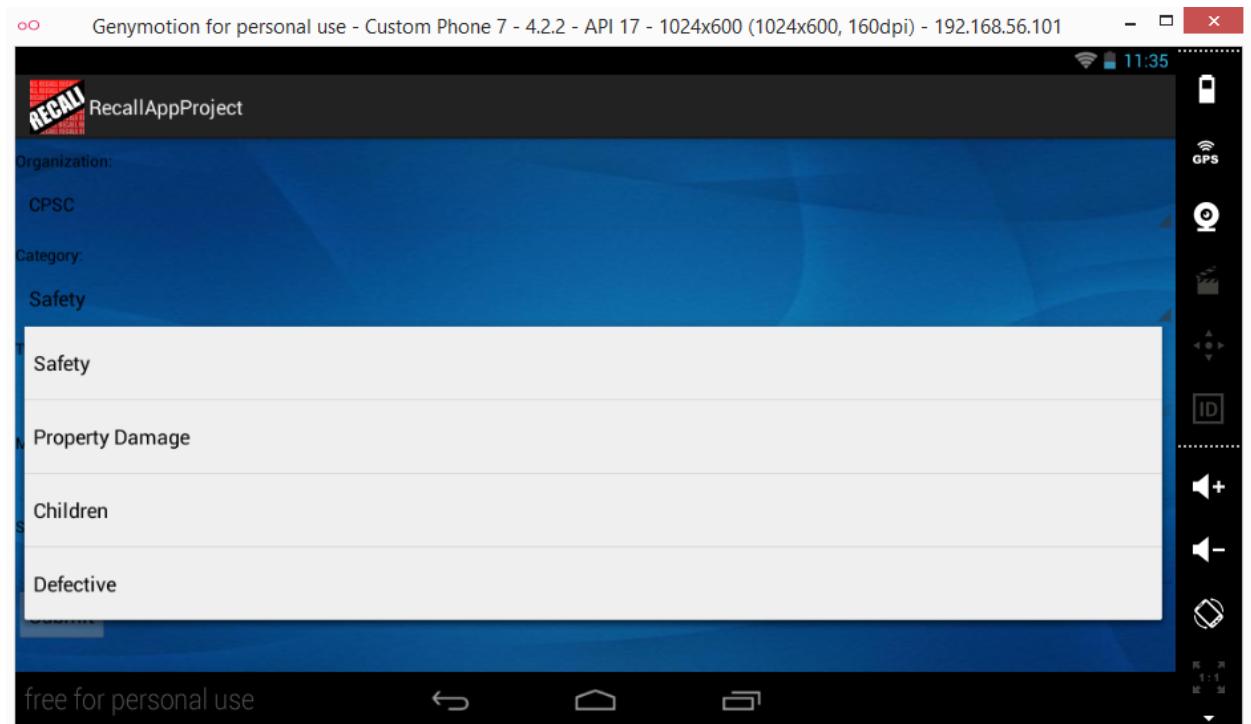
How to Use the System



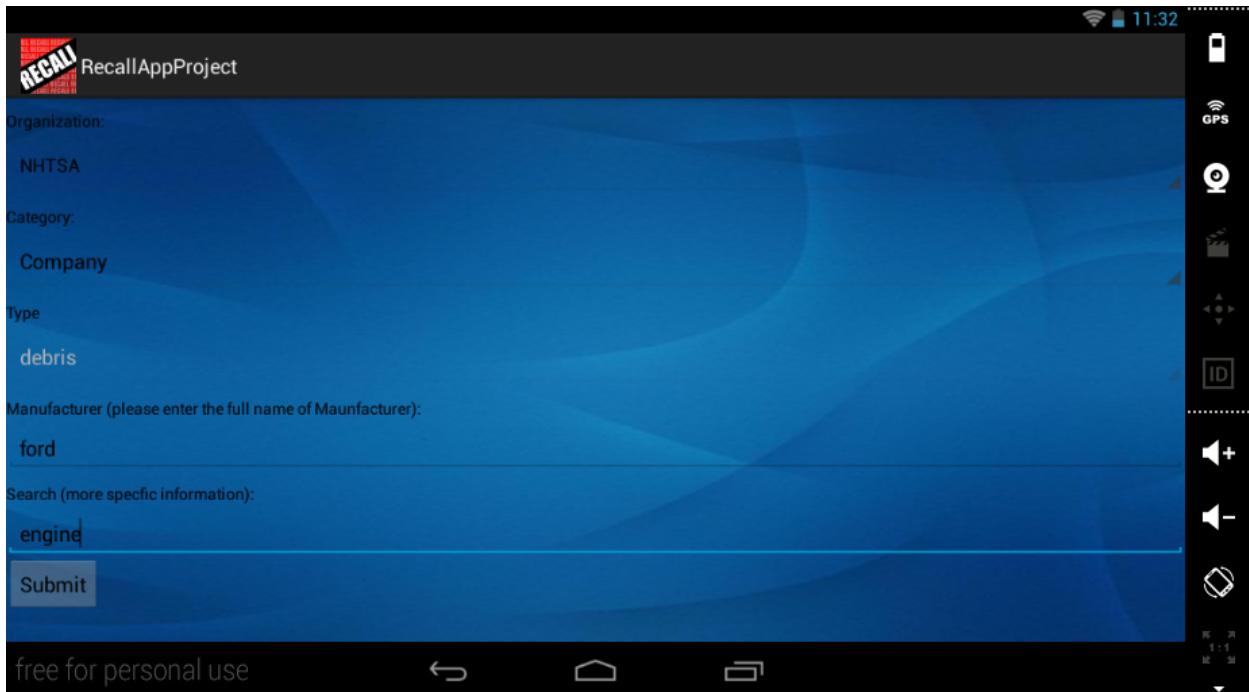
Main Screen - there are three feature for user select.



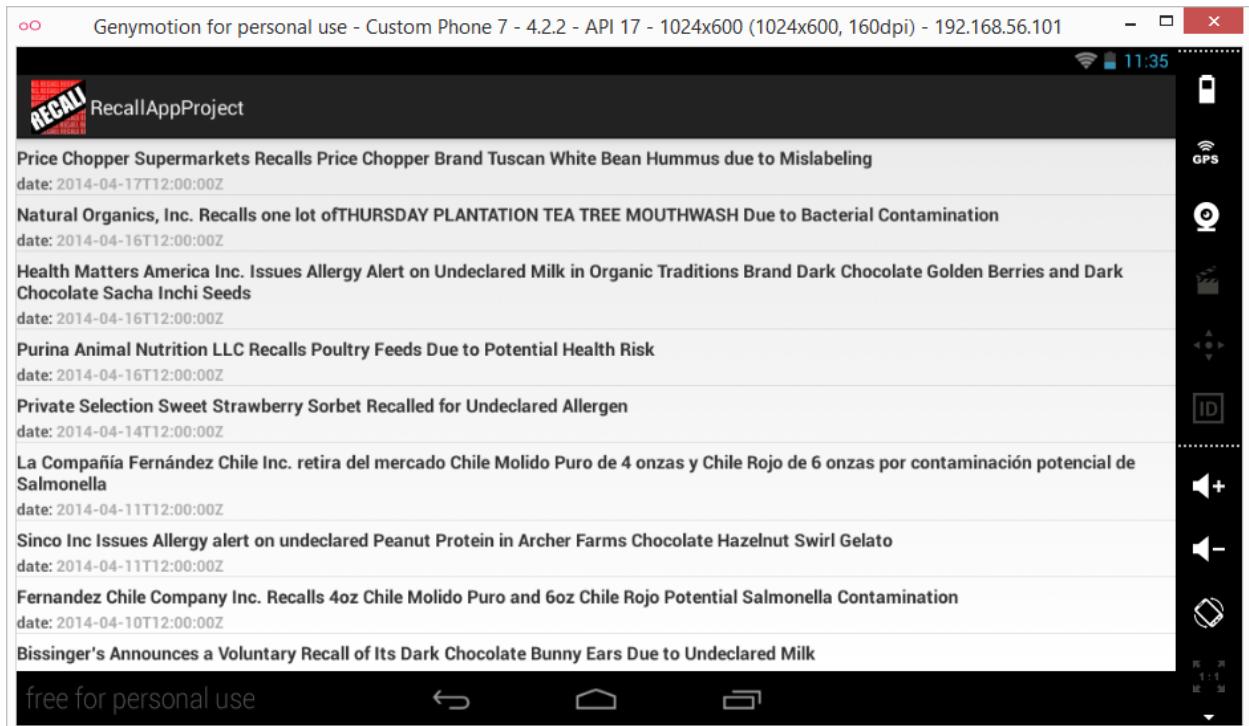
Recall Category - In the view, user will be able select different organization, categories and subcategories. For FDA and USDA, it has two categories mislabeled and contamination, for each categories, it has six subcategories. also user search any specific information by type keyword in search box.



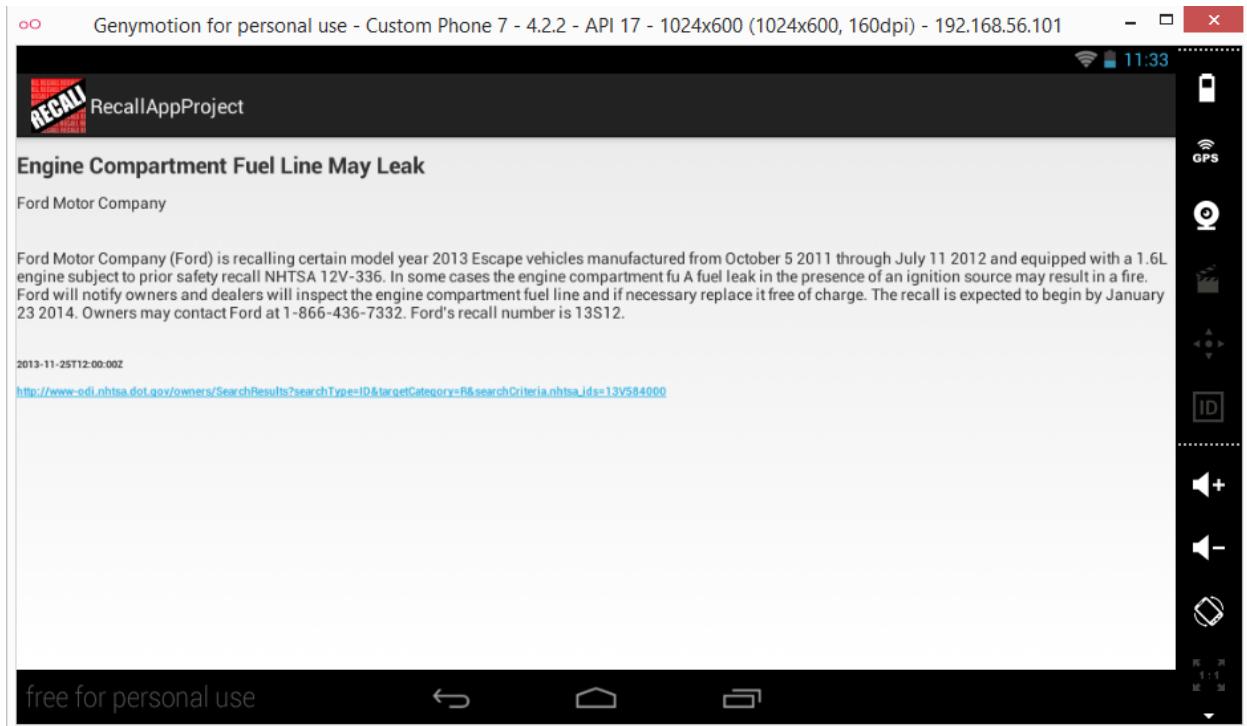
Recall Category - For the CPSC, it has four categories, but there is no subcategories, the search function still available.



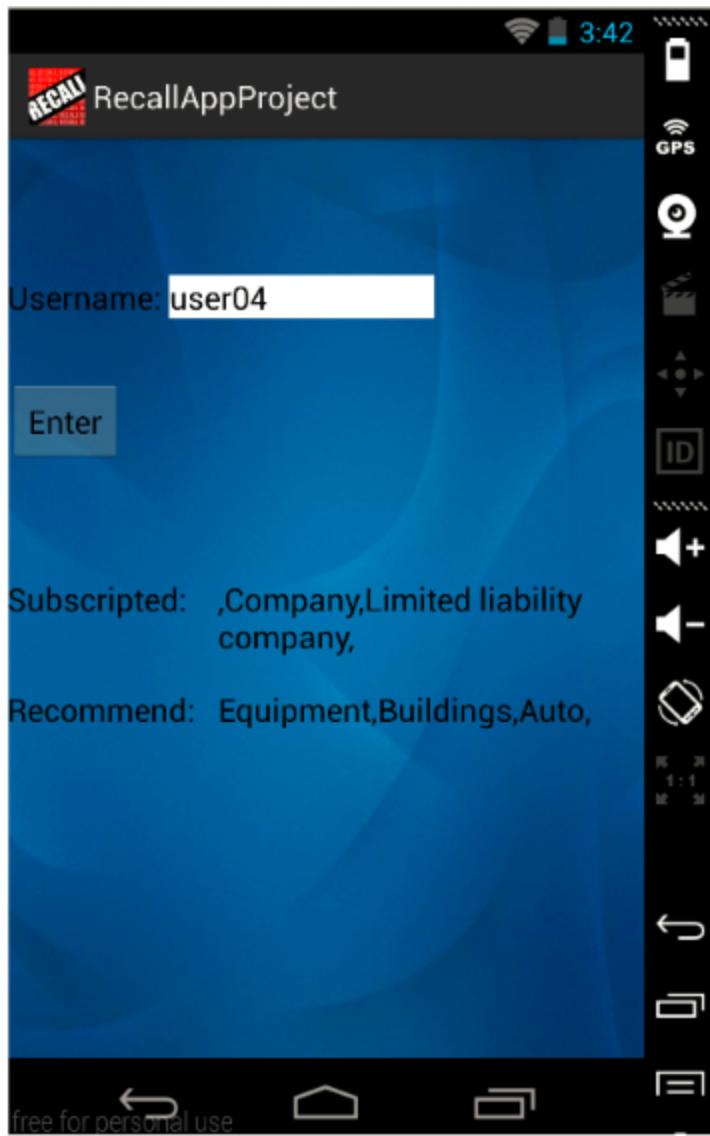
Recall category - For NHTSA, user will be able to select different car company. In manufacturer part, user would type any manufacturer of the car they want to look into.



Recent Recall Records - This is a view that display most recent recall information. it display as a list. For each recall, it display title and recall date. if user want to see more information about one particular recall, user would just click that one.



Recent Recall Records - This is a view that after user click one specific recall. In this view, it displays the title, description, recall date and link.



User Recommender - In this view, it displays the users' subscription, also base on their subscription, it displays recommendation for the users.

Error Recognition and Handling

The application was tested for consistency under the assumption users are inside the University of Missouri - Kansas City network. This is accessible for students and employees from any location connected through a virtual private network client. In the event of errors, it is useful to notify any of the developers if you know them personally. Otherwise, download the source code and feel free to have a look around.

Known Bugs and Deficiencies

At the time of our final publication, the hosted Glassfish Server 3.0 resource provided by UMKC was compromised and unrecoverable by the available technical support. As a result, our user recommendation system is inoperable.

Project Management Report

As part of our Project Management, We had used ScrumDo Website for planning our tasks among our team mates.

Increment 01: (02/25/2014 - 03/11/2014)

We had completed the following Tasks for our Increment 01 in ScrumDo

1. James Clark - Data Search and selection of Data, Work on Data Processing Methods, Report
2. Wang Zhang - Basic interface, Finalize the features going to provide in our Application, Design of UML Class and ER Diagrams, Report
3. Tirumala Reddy - Data Search, Study and Finalise the Back end Algorithms going to use, Report
4. Jagadish - Managed Scrumdo

The screenshot shows the ScrumDo dashboard for the organization "UMKC_KDM". At the top, there are navigation links for "Your Stats", "Organization Stats", "Point Breakdown", and "Organization Velocity". Below these are summary statistics: Active Projects (1), Total Stories (3), Total Tasks (10), Users (3), and Overall Velocity (3). A section titled "RECALL APP" displays "Iteration 1" from "Feb 25, 2014 - Mar 11, 2014". It lists one story: "#3 Data Search and Design of Back end Algorithm". A tip at the bottom says: "ScrumDo Tip: Drag a story up or down in an iteration to set its rank." A "Tutorial Videos" link is visible at the bottom left.

Iteration 1

The image shows a ScrumDo Kanban board for Iteration 1, which spans from February 25, 2014, to March 11, 2014. The board has four columns: Todo (empty), Doing (empty), Reviewing (empty), and Done. The Done column contains three tasks:

- #3 Data Search and Design of Back end Algorithm by tiru417, 0 Comments - Tasks, 2 points
- #1 Android Application by 2207722zw, 0 Comments - Tasks, 3 points
- #2 Data Search and finalize the project idea by jwc62f, 0 Comments - Tasks, 2 points

Increment 01

Increment 02:(03/11/2014 - 04/01/2014)

We had completed the following Tasks for our Increment 02 in ScrumDo

1. James Clark - Collection of Data from CPSC, USDA API's and Classification of Data using Naive Bayes Classifier, Check servers other than Solr to work for our project.
2. Wang Zhang - Adding More Features to Android App, Create Interface with Solr Server, Debugging Android Application
3. Tirumala Reddy - Collection of Data from ODI-NHTSA web site and Classification of Data using Naive Bayes Classifier in Mahout,Check "R" Machine learning Tool for Classification and Managed ScrumDo

The screenshot shows the ScrumDo interface for Increment 2, spanning from March 11, 2014, to April 01, 2014. The top navigation bar includes links for Fwd: ppt - kon, Project - Goog, Final Report - , Increment 02 - , ScrumDo : REC, ScrumDo : REC, ScrumDo : REC, Blackboard Le, Facebook, and a search bar. The main header displays "Increment 2 Mar 11, 2014 - Apr 01, 2014". Below the header, there are four summary boxes: Stories (3), Total Points (9), Points In Progress (0), and Points Completed (9). A "Not Enough Data" message box is present, stating "We don't have enough data to draw this burn up chart right now." with tips: 1. Size your stories, 2. Set the iteration dates to include today, 3. Burn-Up charts are generated nightly. To the right is a "Scrum Board" section with a "Stories" board containing three items: #6 Android Application, #4 Classification of data collected from cpsc.usda sites, and #5 Classification of data collected from odl-nhtsa website. Each item has a "Done" status, 0 comments, and a task count of 7, 3, and 3 respectively. A "Quick Links" sidebar on the right lists various project management tools and links.

Increment 02

The screenshot shows the ScrumDo interface for Increment 2, spanning from March 11, 2014, to April 01, 2014. The top navigation bar includes links for Fwd: ppt - kon, Project - Goog, Final Report - , Increment 02 - , ScrumDo : REC, ScrumDo : REC, ScrumDo : REC, Blackboard Le, Facebook, and a search bar. The main header displays "Increment 2 - March 11, 2014 - April 1, 2014". Below the header, there is a Kanban board with four columns: Todo, Doing, Reviewing, and Done. The Done column contains three items: #6 Android Application, #4 Classification of data collected from cpsc.usda sites, and #5 Classification of data collected from odl-nhtsa website. Each item has a "Done" status, 0 comments, and a task count of 7, 3, and 3 respectively. A "Quick Links" sidebar on the right lists various project management tools and links.

Increment 02

Increment 03:(04/02/2014 - 04/18/2014)

We had completed the following Tasks for our Increment 03 in ScrumDo

1. James Clark - Improving Accuracy of Classifier using Word Count Example and developed Web services to Interface Android Application with MongoDB.
2. Wang Zhang - Update Android Application according to the Classified results. Introduced User Recommender feature to Android Application
3. Tirumala Reddy - (As the data from ODI_NHTSA web site is not updated frequently) Classification of data of nhtsa.gov web site based on naive bayes and Developed a Recommender Algorithm for Recommending some categories to a particular user and managed Scrumdo.

ScrumDo Search Project UMKC_KDM RECALL APP tir417

Scrum Board

Increment 03 Apr 01, 2014 - Apr 18, 2014

Stories	Total Points	Points In Progress	Points Completed
3	15	0	15

Not Enough Data

We don't have enough data to draw this burn up chart right now.

Some Tips:

1. Size your stories
2. Set the iteration dates to include today
3. Burn-up charts are generated nightly

Burnup | Burndown | Stacked | Time

Stories

Add Story

#9 Categories selection and User Recommender Feature
Update Android Application according to the Classified results. Introduced User Recommender feature to Android Application
Done Tasks | 0 Comments 2207722zw 5

#7 Classifier and MongoDB
Improving Accuracy of Classifier using Word Count Example and developed Web services to Interface Android Application with MongoDB.
Done Tasks | 0 Comments jvc62f 5

#8 Classification and Recommender Algorithm
Classification of data of nhtsa.gov web site based on naive bayes Developed a Recommender Algorithm for Recommending some categories to a particular user.

Quick Links

- Project Summary
- Edit
- Iteration Planning
- Chat
- History
- Predictions
- Planning Poker
- Iterations
- Backlog
- Iteration Planning
- Iteration 03
- Iteration 2
- Iteration 1

Increment 03

The image shows a ScrumDo Kanban board for Increment 03, spanning from April 1, 2014, to April 18, 2014. The board has four columns: Todo, Doing, Reviewing, and Done. The Done column contains three tasks:

- #9 Categories selection and User Recommender Feature (Owner: 2207722zw)
- #7 Classifier and MongoDB (Owner: jwc62f)
- #8 Classification and Recommender Algorithm (Owner: tiru417)

The right sidebar includes a 'Quick Links' menu with options like Project Summary, Epics, Iteration Planning, Chat, History, Predictions, Planning Poker, Iterations, Backlog, and Backlog Stories. It also shows the status of Iterations 03, 02, and 01.

Increment 03

Increment 04:(04/19/2014 - 05/04/2014)

We had completed the following Tasks for our Increment 03 in ScrumDo

1. James Clark - Produced Final Classified Results for CPSC, FDA and USDA and Upload it to Solr Server, Stored Recommendation Index Table in MongoDB and developed a Webservice using Glassfish server to produce Recommendations and Project Video
2. Wang Zhang - Update Android Application according to the Final Classified results, Created Interface with Mongo DB for Providing User Recommendations, Final Design of Application and Debugging and Project Video.
3. Tirumala Reddy - Produced Final Classified Results for NHTSA data, Created User's Database for training Recommender, Final Recommendation Index Table and Recommender Results and Project Video and managed ScrumDo.

ScrumDo Search Project

Increment 04 Apr 19, 2014 - May 04, 2014

Scrum Board

Stories Total Points Points In Progress Points Completed

3 24 0 24

Not Enough Data Some tips
We don't have enough data to draw this burn up chart right now.
1. Add more stories
2. Set the iteration dates to include today
3. Burn-Up charts are generated nightly.

Burnup | Burndown | Stackup | Time

Add Story

Stories

Filter Board

#12 Final Classified Results, User Data Base, Final Recommendation Index Table
Produced Final Classified Results for NHTSA data, Created User's Database for training Recommender, Final Recommendation Index Table
Recommender Results and Project Video.
Done Tasks | 0 Comments tiru417 8

#10 Webservices, Solr Uploading and Final Classified Results
Produced Final Classified Results for CPSC, FDA and USDA and Upload it to Solr Server, Stored Recommendation Index Table in MongoDB and developed a Webservice using Glassfish server to produce Recommendations and Project Video
Done Tasks | 0 Comments jwvc82f 8

#11 Update App to work with Classified results and User Recommendations
#11 Update App to work with Classified results and User Recommendations

Increment 04

ScrumDo Search Project

Increment 04 - April 19, 2014 - May 4, 2014

Todo Doing Reviewing Done

Filter Board

#12 Final Classified Results, User Data Base, Final Recommendation Index Table
tiru417 0 Comments - Tasks 8

#10 Webservices, Solr Uploading and Final Classified Results
jwvc82f 0 Comments - Tasks 8

#11 Update App to work with Classified results and User Recommendations
2207722zw 0 Comments - Tasks 8

Increment 04

Project Summary:

ScrumDo Search Project

RECALL APP

Velocity Total Stories Stories Completed Stories In Progress Iterations Left

13 12 12 0 4

Quick Links >

Project Summary

Epic

Iteration Planning

Chat

History

Predictions

Planning Poker

Iterations

Backlog

Increment 04

Increment 03

Increment 02

Increment 1

Website URL (Scrumdo):

URL: <https://www.scrumdo.com/projects/project/recall-app/summary>

Final Project Evaluation

Discuss how well your project satisfies your original requirement specification.

We successfully gathered data from various recall publications and classified it with machine learning techniques. In addition, we emulated a recommendation system in place of a missing user registration component of our application. Since we originally planned to create a user registration platform and did not, we did not satisfy this part of our original specification. Furthermore, our intent was to base a machine-learned user recommendation system on preferences registered by newly created users, and provide a subscription service by which users could receive email notifications about recalls relevant to their subscription. Lacking a proper user registration workflow, we emulated recommendations by producing static user content and coded it into a class of our application.

Were you satisfied with your design process?

We covered all aspects of this project from two different directions. Tirumala began understanding machine learning algorithms as they would apply to our data. Wang created the Android interface, and James executed data mining processes that would integrate directly to Solr. Like a puzzle, all of the pieces came together as a finished product.

How helpful was the agile process?

The agile process was helpful in planning an overall project's requirements, but timelines were ultimately manipulated to be whatever we needed them to be. There was no missing a deadline because we just moved the deadline whenever we wanted to. In that way, agile process was not effective as an enforcer of progress since there was no project manager to oversee progress and performance requirements of project contributors.

How would you do the agile process next time?

Until an employer leverages performance requirements against production work, we would continue to perform as needed to accomplish necessary academic gains. That is to say, timelines and workloads are whatever and wherever we are comfortable having them so long as we are productive in meeting our end goal using reasonable time and resources to get there. Nevertheless, agile process is a good exercise to build awareness.

Did you stick to your project plan schedule?

Communication among active participants was regular. At the beginning of every new increment, we had clearly defined goals and expectations. Who was working on what, and when it could be expected was always clear. Any delay was properly communicated. There were regular setbacks or delays in the project ultimately resulting in unfinished components, but these were always made clear to the invested contributors. At the end of every increment, contributors always met to update documents and realign our goals and activities.

What was the real management structure within your group?

There was never an officially elected group leader. We used online collaborative tools courtesy of the Google Drive and Google Plus platforms to communicate remotely and share documents

and data. We incrementally scheduled meetings and we were fortunate that our schedules aligned in such a way as to meet on critical deadlines whether in person or online. Anyone that expressed a project need or concern was accommodated by others. Dr. Lee was clear in her expectations of project deliverables and our group was responsible in producing deliverable works. The workload was evenly distributed among those who chose to participate.

Does it bear any resemblance to structure that you had planned?

We informally planned everything after increment one. In spite of lacking formality, our planning did evolve a precise structure. For every increment, we would identify work previously completed based on our original plan. During the increment reporting process, we would identify the next segment of work to be completed by each person in preparation for the following increment. In between increments, we met after class to share progress and demonstrate working contributions to our project.

Did you have any problems getting each member to do his/her share of the work?

Three of four group members chose to contribute and were self motivated and fully committed to the final outcome of our project. One group member actively chose not to participate on this project, ignoring incoming communications and contributing nothing beyond the pre-planning phase.

Do you have any suggestion on how this could have been handled better?

No. Since the group was ready to proceed without all members present, any lack of participation by a group member was compensated for. Non-contributing participants will answer for their actions at their final evaluation.

Discuss what you might have done differently if this were a real world project?

If this were a real world project, we would have to work harder and produce more work hours to accomplish our original objectives. Our machine learning classifications and recommendations would have to be more accurate. A fully realized work flow would have to be produced to handle the constant stream of data. From back to front, it is possible to build a system that does all we wanted to do. Perhaps four instead of three active contributors would have built such a system. Certainly, a non-participatory staff member in a real project at a company would have been reprimanded.

Any recommendations for next year?

The tutorials that guided the beginning phase of this course were strong, but they can be stronger. It was not perfectly clear how a fully operating map reduce and machine learning operation can be realized to produce a working back end for a mobile client to consume. More clear expectations of the project might include an architectural requirement such as server-side process in map reduce, restful service for manipulation and control of data, and mobile client to control or consume either. We think it will be good if more detailed material is provided particularly in case of machine learning like clustering (hierarchical and obstacle avoiding cluster), evaluation of classification, recommending and clustering methods.