**Group 7 Project Proposal**

* Plan: February 21, deadline.
* First Increment: March 7.
* Second Increment: March 21.
* Third Increment: April 11.
* Fourth Increment: April 25.
* Project Presentations: May 6 & May 8.
* Final Project: May 9.

1. **Introduction**

The idea of our project is to fetch a best suited recipe for the user’s requirement. The user requirements will be of the kind of meal the user is willing to have, cuisine, ingredients and time constraint. In our project we do machine learning on the data. Where the data consists of different recipes. When user gives his constraints, then the best suited recipe is fetched from the machine learned data.

1. **Project Goal and Objectives**

The main goal of the project is to provide a best matched recipe to the provided constraints by the user. Which satisfies the mentioned time, cuisine and ingredients. This project will be an efficient way of providing recipes to the user. Because in general recipes user has to search for the recipe which will match his constraints from a large number of recipes. We do machine learning to these recipes to reduce the work and effort of the user using this application. So that user will easily find one among the shortlisted recipes which suits him.

Our Project can add more features in the future. By collecting data from each user which tells us of what kind of recipes are being preferred by a particular user, we can suggest him that particular kind of recipes. Also we can suggest him of new kind of recipes which will suit his taste. So that he can try them sometime.

The objective of our project is doing machine learning on the data. Then storing the results of the confusion matrix on solar. The user will have a mobile application where he will input his constraints. These parameters are passed through web services which will query solar and obtain results. These resulting recipes are sent to the user’s application. Where he is given a few choices meeting his constraints.

1. **Project Background and Related Work**

**Recipe recommendation using ingredient networks**

In this system they have considered each ingredient and developed recipe based on the given ingredients. They also measured individual ingredients and decided whether to drop or add an ingredient. They also constructed two types of networks to capture the relationships between ingredients. The complement network try to gather the ingredients which occur frequently and divide them into two categories one sweet and other savory. The substitute network considers user recommendations from various communities and gives recommendations for healthy recipe.

4**. Proposed System**

1) Requirement Specification

Functional, Non-functional, Technical/business Requirements (prioritized)

­­Functional Requirements:

* Users will give all the ingredients whatever they have and will enter on the home screen shown in the Android app.
* There are some RDF datasets for recipe which we already got using web crawler from the internet.
* Using mahout and with these RDF datasets, all the information will be stored in Solr.
* User entered information will be matched with solr data and all the possible combinations will be exported as a datasets from Solr.
* The combinations basically the mahout’s output will be stored as the separate table.
* Web service will display all the information(combination) from the solr onto the screen.
* Solr will access and retrieve data very fast and it’s easier to use.
* User has the option to select from the combination which will result in steps to prepare that recipe.
* This steps will be displayed as the separate interface to users.

Non-Functional Requirements:

* + Should show all the combinations of recipe in an efficient manner in both time and space.
  + The system should not get disconnected at any point of time.
  + It should show all the possibilities without any interruption from the semantic of the ingredients.
  + Results shown should be displayed in such a way that should be easy to understand.

Technical/business Requirements:

* + We will use cloudera instance provided by the school.
  + Solr will be used to store all the datasets for the various combinations of ingredients.
  + For Agile process, we will use ScrumDo.
  + For version control we are using GitHub.
  + We are creating our own API’s if needed.

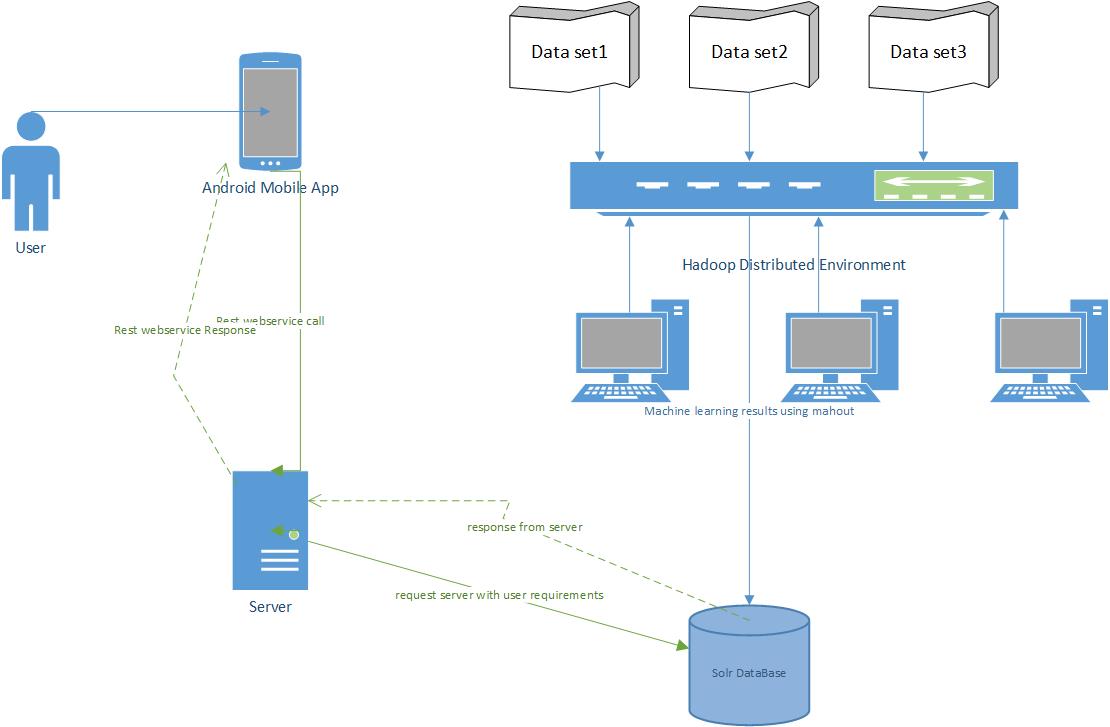
Technological and Architectural requirements

* Cloudera virtual machine with atleast one datanode.
* Hadoop
* HDFS
* Zookeeper
* Solr
* Mahout
* Rest WebServies
* Solr webservices
* Android ICS and above
* Python (web crawler)

1. **Framework Specification:**

* Front end will be and android application which will allow user to select his ingredients and other constrains like time, and cuisine.
* Using this information mobile app will make a web service call to server which will fetch result from Solr database.
* Methodologies and Algorithms
  + to do machine learning on dataset we have used inbuilt algorithms present in mahout
  + We are using mahout specified default algorithms
* Data Sources

System Architecture:



User enters the constraints like ingredients, time and cuisine in android mobile application. This data will be sent to the server . The datasets regarding the recipes are made to machine learn using mahout and stored in the solr database. Server sends the request data to solr database. Here solr selects the required data for the request and sends the reply to server. Server sends the response back to android app where the user can see the result.

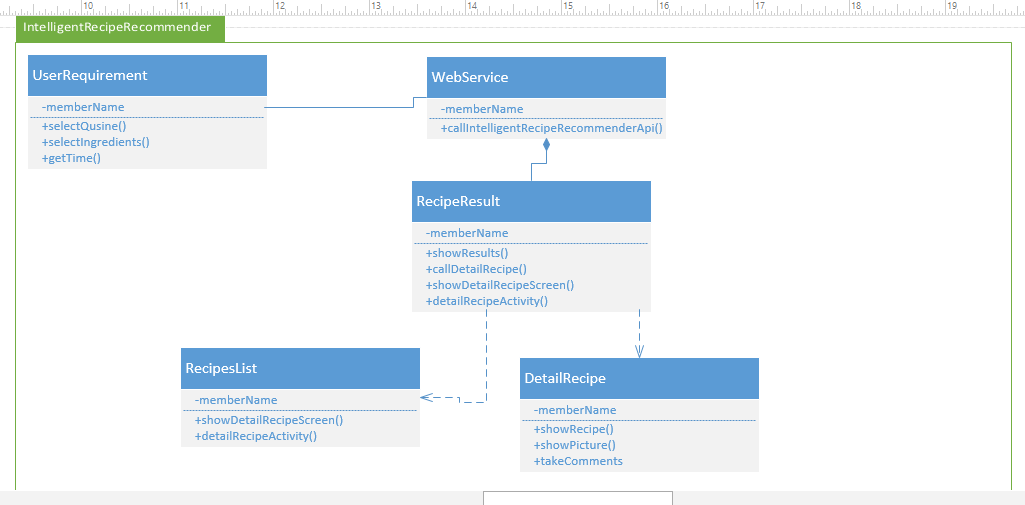
System Specification:

Datasets:

We have to write a web crawler to crawl all the web pages in [www.allrecipies.com](http://www.allrecipies.com/) and get the directions and ingredients list from recipes and save it to text documents.

We have used an existing project in python which has already done web-crawling on www.allrecipies.com but we need to modify that crawler to get some more extra details from that crawler.

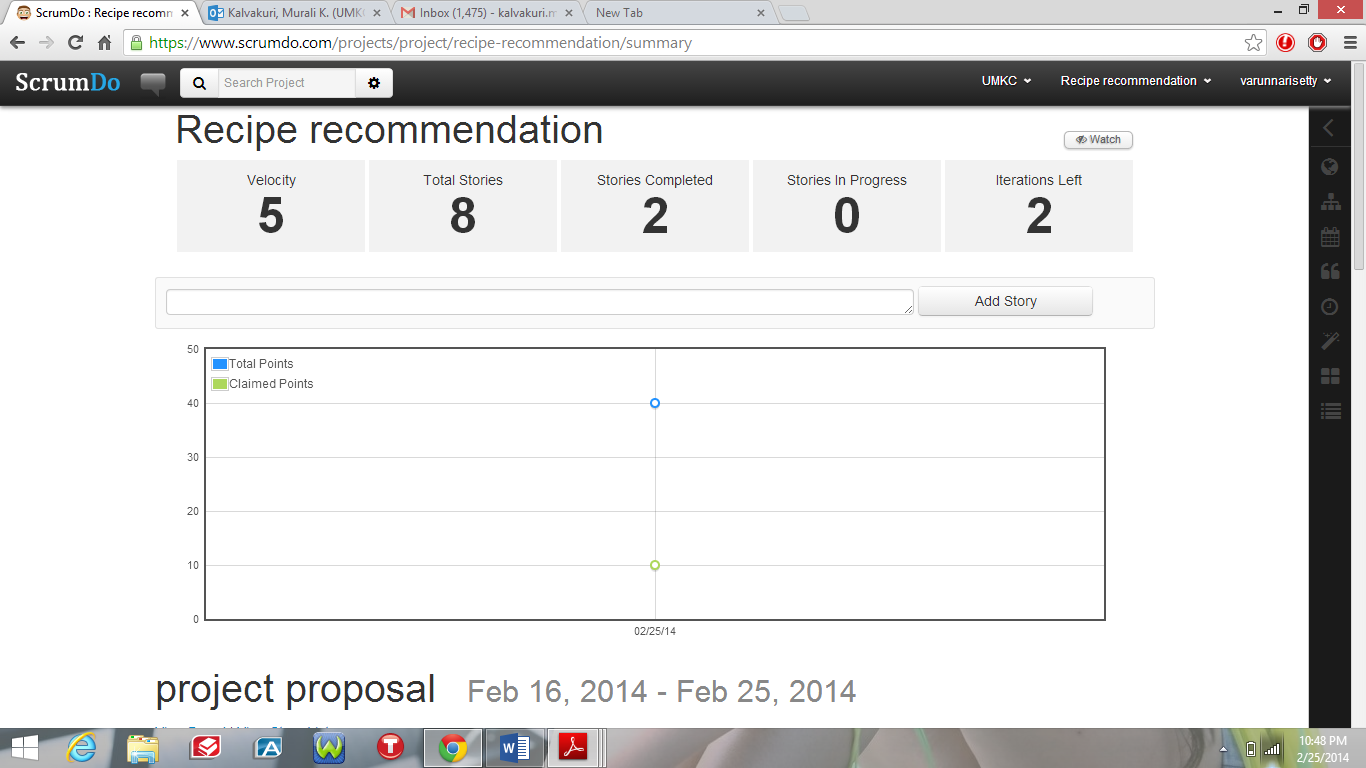
UML Class Diagram:



1. **ScrumDo URL:**

We have mentioned our increment stories and plan in ScrumDo

<https://www.scrumdo.com/projects/project/recipe-recommendation/summary>​



Post Project Plan document, data, and code to GitHub and post its URL to the google Spreadsheet (above)

7. **Bibliography**

* Recipe Recommendation system.

<http://arxiv.org/pdf/1111.3919.pdf>

* Mahout Machine learning

<https://mahout.apache.org/>

* Solr indexing data base

<https://lucene.apache.org/solr/>

Web crawler in python

<https://github.com/prateekvjoshi/Python-WebCrawler>