Pamuru, Sricharan (UMKC-Student)

UMKC  [Company address]

ARTICLE RECOMMENDATION SYSTEM

Contents

[1. Project Proposal: 2](#_Toc454471820)

[a. Motivation: 2](#_Toc454471821)

[a. Objectives: 2](#_Toc454471822)

[b. Expected Outcomes: 2](#_Toc454471823)

[2. Domain: 2](#_Toc454471824)

[3. Data Collection: 3](#_Toc454471825)

[4. Report: 3](#_Toc454471826)

[5. Design of features 3](#_Toc454471827)

[a. Software Architecture: 3](#_Toc454471828)

[i. UML Modeling: 3](#_Toc454471829)

[ii. Class Diagram: 3](#_Toc454471830)

[iii. Sequence Diagram: 3](#_Toc454471831)

[iv. Activity Diagram: 3](#_Toc454471832)

[v. Use Case Diagram: 3](#_Toc454471833)

[b. Workflow: 3](#_Toc454471834)

[c. Existing Services used: 3](#_Toc454471835)

[d. New feature implemented: 3](#_Toc454471836)

[6. Project Management: 3](#_Toc454471837)

[a. Individual Contribution: 3](#_Toc454471838)

[b. Zenhub/Github Screenshots: 3](#_Toc454471839)

[c. Concerns/Issue: 3](#_Toc454471840)

[d. Future Work: 3](#_Toc454471841)

[7. Bibliography: 3](#_Toc454471842)

# Project Proposal:

## Motivation:

The motivation of building a recommendation system for articles is an aspect of providing some kind of intuitive outcome for people who read a specific post or blog. In our case, the thought of binding different sources altogether in order to obtain some sort of output which will yield to dynamic data fed in the form of RSS, has been an ideological aspect that we have considered significant to understanding how the data trends and topics of approach have to be brought together. This combination will enable a user to stream into topics of interest across more than one domain. The machine would be able to return a suggestive article in the list through various technologies that we plan on implementing in this project

## Objectives:

The goal of our project is to come up with a recommendation system for articles in different domain areas to bring the user a specific search blog or engine. It will help the user to understand how the scalability and necessity of the article is considered. The page rank would be used as a measure to understand the sequence of recommendations put forward. Based upon this, the articles returned would focus on terms that have a similarity index, matching the score of the previously iterated indices. There would be a relative rank associated that would be used to figure out the magnanimity of the article which would have to be returned.

Our recommendation engine would return the articles to be recommended in a list along with the associated hyperlink. This would help the user view topics of similar interest. Machine learning algorithms run in the background to bring a result yielding a knowledge graph where there is a proper connection between the elements, which in our case is articles, dynamically.

## Expected Outcomes:

The outcome of the project would be to bring out articles to be recommended in a list which will help the user view similarity index based articles. Our project is very useful for people who read blogs and articles. It would be an interactive web page where similar index based documents have a value returned which match the preexisting ones.

# Domain:

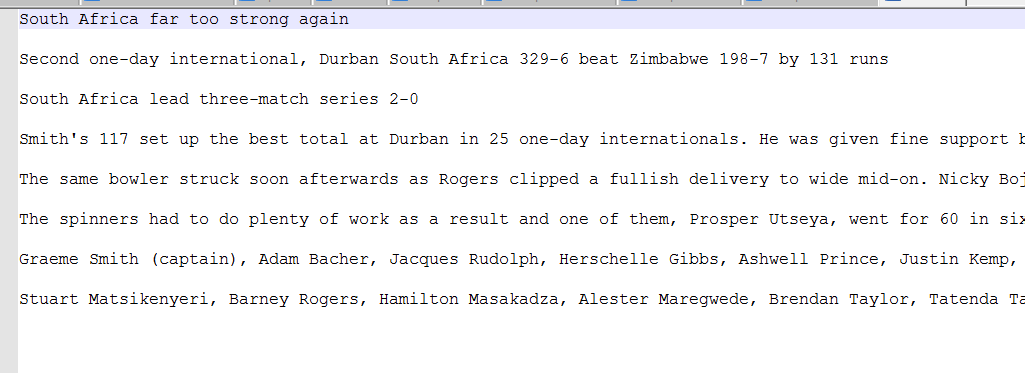
Our project aims at fetching real time data in the form of RSS(Rich Site Summary) which will be dynamic. The recommendation engine will use collaborative filtering and it serves as a branch of deep learning. It will enable the machine to undergo a deep analysis and filter the data collection that is appropriate for mining. Technology has developed immensely and it is required for the machine to be able to understand each fragment in order to return a better perspective of how the inner mechanism works. Neural networks would basically iterate through every fragment and in our case the articles have to be searched and read through to understand how the term frequency and inverse document frequency matches. (TF-IDF) is a key constraint that will enable information retrieval to be scaled. We would obtain the respective important term that will differentiate a specific document when compared with another.

Recommendation engine is a platform for the machine to be trained and tested. Simulation of such technology actually occurs by rapidly searching the inner context. Big data mining tools focus on scalability and larger data sets are compared in order to retrieve information which is adaptive to the respective target, which is the articles. The NLP tools will help us extract a relation and naming the entities that is crucial to construction of feature vectors. They are combined with machine learning algorithms and tools to form an ontology. The ontologies learned will altogether contribute in constructing the recommendation system, which is the ultimate motivation of our project.

# Data Collection:

The data collection for this specific increment has been generated from the BBC news sites where the various articles sections have been formed. It has helped us understand how the articles are generated and their category mapping.

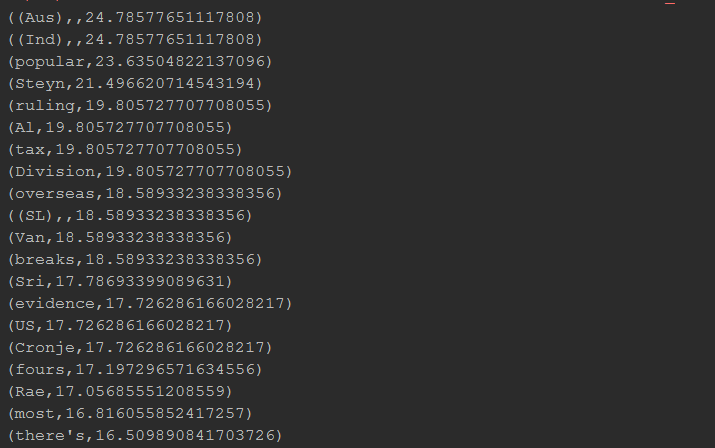
Sample article under Sports Category:



Similarly, the data collection is static right now and we have used the BBC datasets as a training set. The overall outcome would be to dynamically obtain data in the form of RSS feed. The RSS feed would use the standard web feed formats in order to publish updates of information that can be categorized as blogs, headlines and also includes audio, video files. The metadata would explain about the specific data content and focus on publish dates, time and author information.

We intend to pull the required information and accordingly sort it as per the user’s search. Dynamically, it would be feasible to categorize it based on the XML file format generated. The RSS reader would check how the user checks feed for information and the user interface will serve as a front-end for interpreting such information. The back end would include the process of fetching data and perform the NLP operations to understand the grammar and contextual meaning.

# Report:



# Design of features

## Software Architecture:

### Architecture of Model

### Knowledge Graph

### UML Modeling:

### Class Diagram:

### Sequence Diagram:

### Activity Diagram:

### Use Case Diagram

## Workflow:

## Existing Services used:

## New feature implemented:

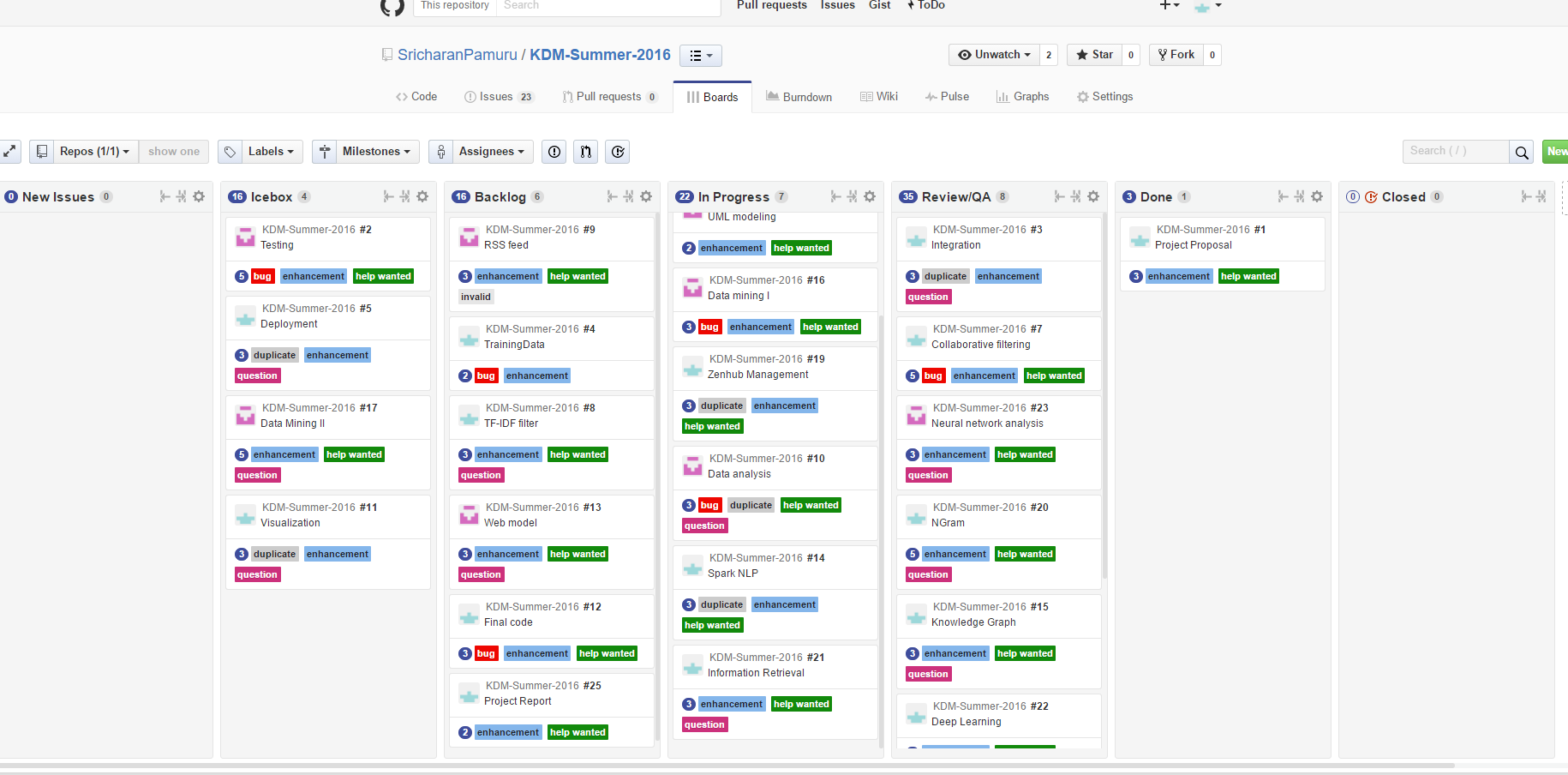
# Project Management:

## Individual Contribution:

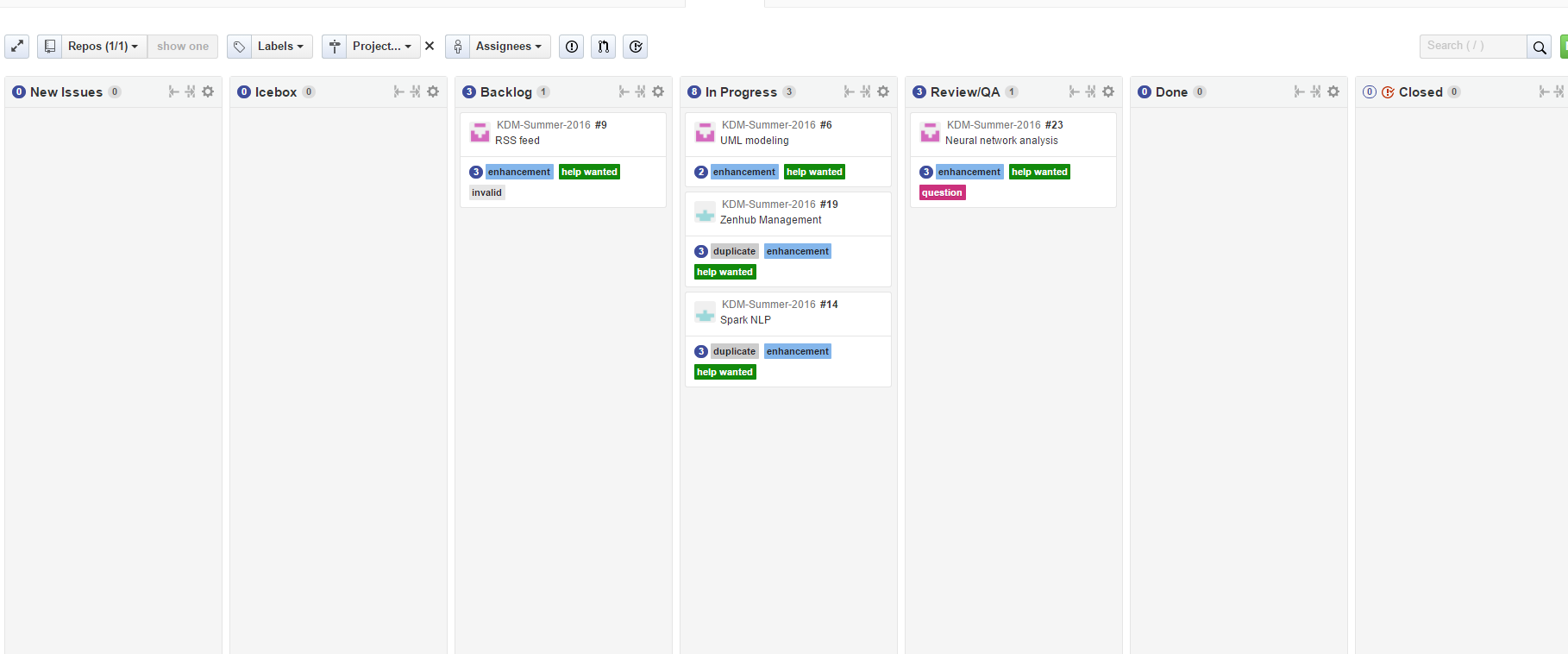
|  |  |  |
| --- | --- | --- |
| Name  (CLASS ID) | Contribution | Number of hours |
| SRICHARAN PAMURU  (20) | CODE EXECUTION FOR TF-IDF AND DATA STREAMING | 20 |
| PRUDHVI RAJ MUDUNURI  (22) | WORKFLOW DESIGN AND NLP PROCESSING ON DOCUMENTS | 20 |
| SNEHAL VANTASHALA  (41) | UML MODELING AND ARCHITECTURE DESIGN | 20 |
| BHARGAV KRISHNA VELAGAPUDI  (42) | DATA COLLECTION AND ARTICLES RESEARCH | 20 |

## Zenhub/Github Screenshots:

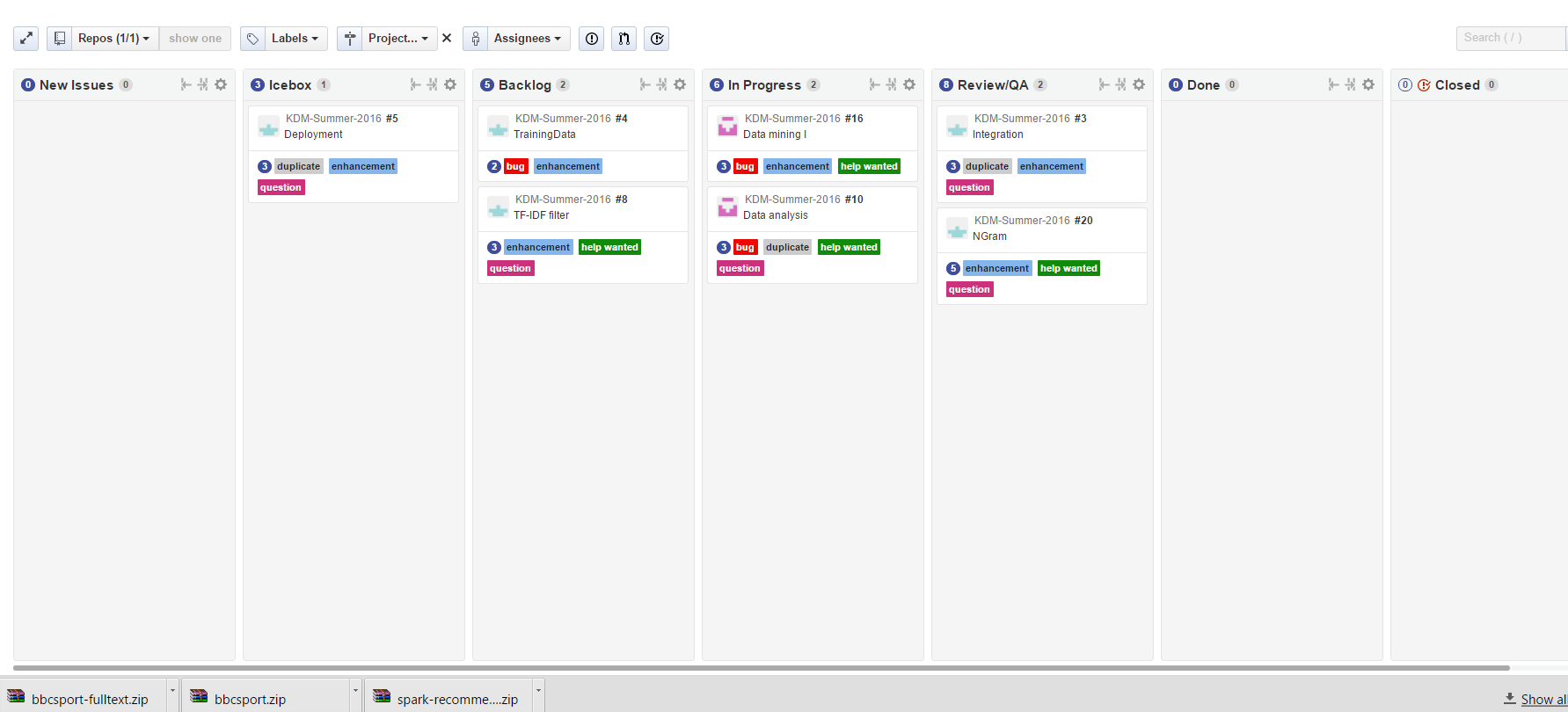
### Full timeline:



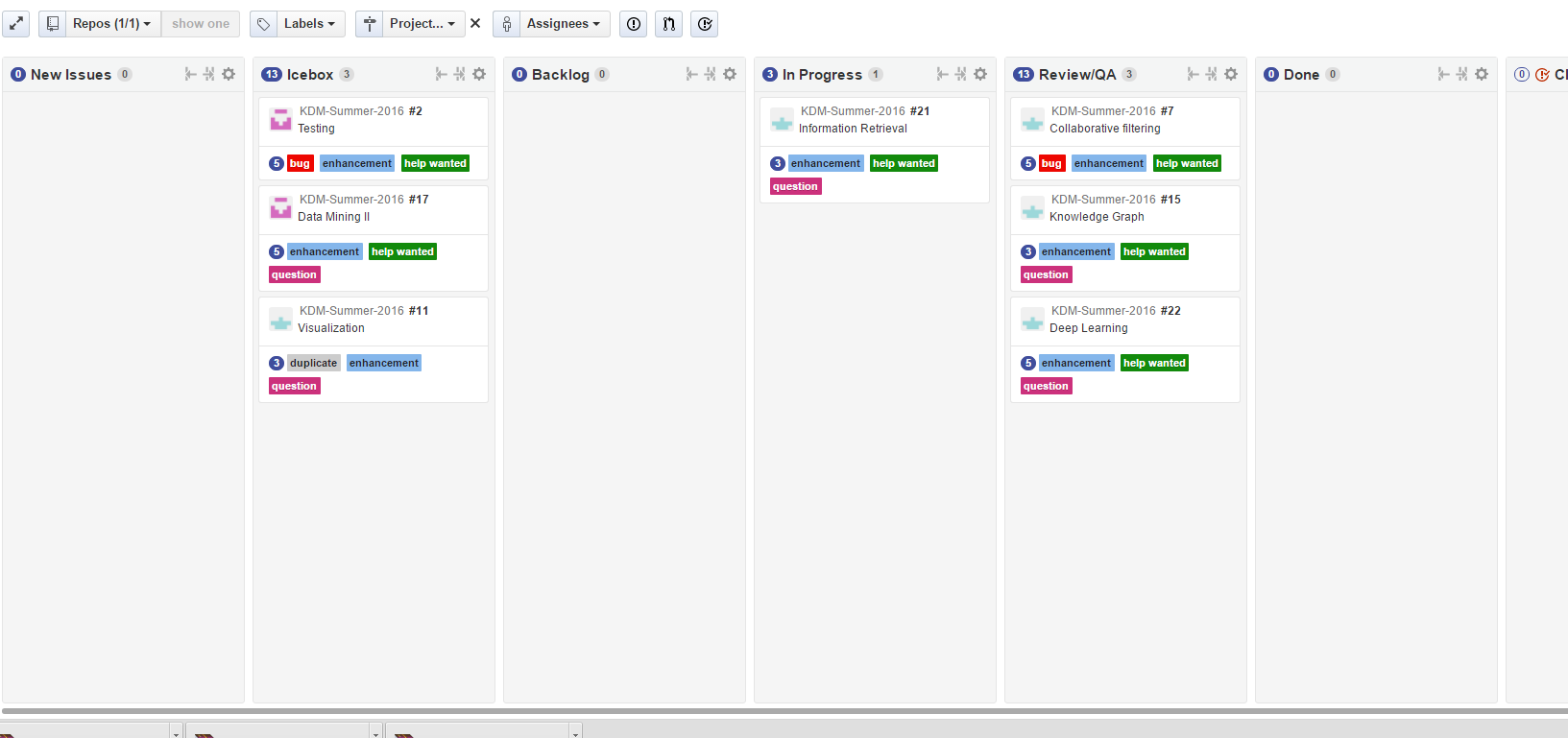
### Report 1:



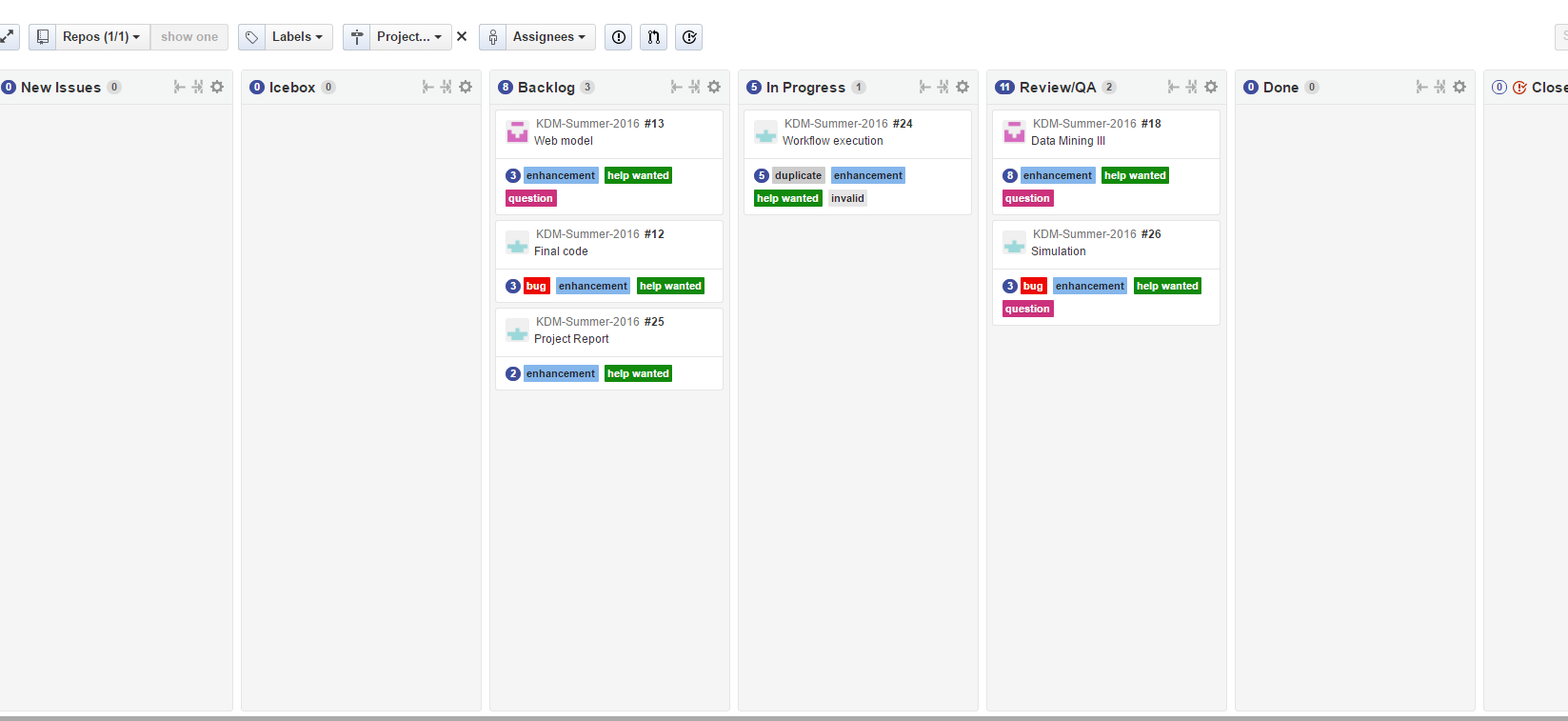
### Report 2:



### Report 3:



### Report 4:



## Concerns/Issue:

## Future Work:

# Bibliography: