

eDrug

by Team 4:

Venkata Vamsi Krishna Bhuvanam

Priyadarsini Nidadavolu

Bhulakshmi Makkena

Tej Kumar Yentrapragada

TABLE OF CONTENTS

1	Motivation	3
2	Objectives	4
3	Expected Outcome	5
4	Domain Chosen.....	6
5	Data Collection	7
6	Tasks and Features	8
7	Implementation	9
7.1	Architecture & UML Diagrams:	9
7.1.1	Sequence Diagram:	10
7.1.2	Class Diagram:.....	11
7.2	Workflow:.....	12
7.3	Existing Services:	13
7.4	Feature Implemented:	16
8	Project Management	21
8.1	Contribution:	21
8.2	ZenHub and GitHub:.....	22
8.3	Concerns/Issues:	24
8.4	Future Work:	25

1 MOTIVATION

The motivation of building a recommendation system for personalized treatment is to diagnosis a disease based on symptoms provided by the users and then recommend a suitable and appropriate drug for the user based on few criteria's like past behavior and historical data of past similar symptoms of the patients along with the drug allergies that the patient has in the past.

Our system will recommend few drugs to select from a list which are approved by the pharmacist. This option will help a user to stream into the list of interest across more than one domain. The system would be able to return a suggestive drug in the list using various technologies that we plan on implementing.

2 OBJECTIVES

The main objective of the recommendation system for E-drug is to recommend a drug to the user based on the symptoms that he provided.

Our system takes user symptoms as input and it analyses the input with the types of diseases and produces a dataset with various possible diseases.

It analysis the type of diseases and ultimately recommends an appropriate drug based on the following criteria:

- Past Behavior of the patient
- Historical data of the past similar symptoms of the patients
- Drug allergies
- Drug similarities that the doctor recommended previously

Our recommendation system would return the dataset of list of recommended drugs with a hyperlink which leads to the composition, availability of the drug in various stores, the cost etc. This would help the user to view the list and can go and check for nearby pharma and buy it.

Machine learning and ALS algorithms run in the background and the result would be an recommended drug dataset.

3 EXPECTED OUTCOME

At the end, this machine should be able to take the dataset and recommend the patient/doctor with the suitable drug without any future side effects.

4 DOMAIN CHOSEN

For this project, Recommendation system will help a lot to analyze the data of symptoms for diseases and its drug compositions.

Also we have used a Spark Core Libraries and Spark Machine Learning Libraries.

5 DATA COLLECTION

Static Data:

<http://human-phenotype-ontology.github.io/downloads.html>

<https://www.nlm.nih.gov/medlineplus/encyclopedia.html>

Hospital Charge Data:

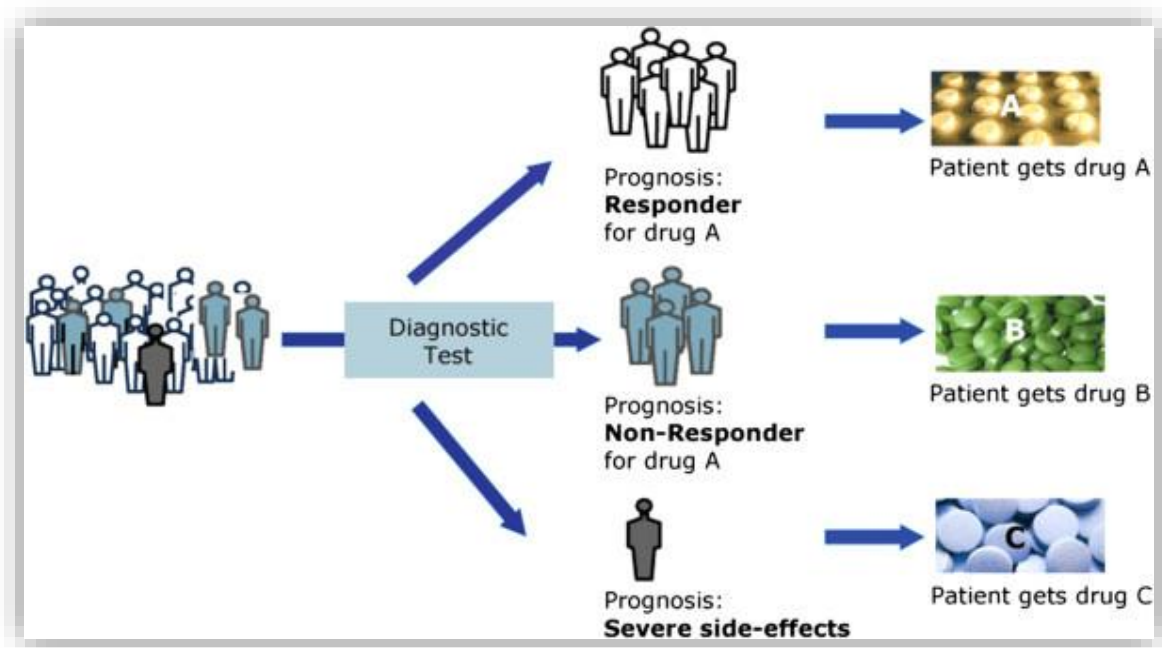
<https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3>

6 TASKS AND FEATURES

This project definitely need a lot of data and to analyze that and get to a strong knowledge base. Using that knowledge, the machine will recommend the drug. Here there will be a lot of features to be considered. For example, first the machine will analyze the symptoms and different types of diseases. This machine will take a lots of criteria into consideration like past behavior of the patient, historical data of other patients and their acceptance or rejection of the drug. Also this machine will analyze the drug allergies of a particular patient and based on the finalized knowledge created, the machine will recommend some drugs, so that the doctors can have a look and prescribe particular drugs to patients.

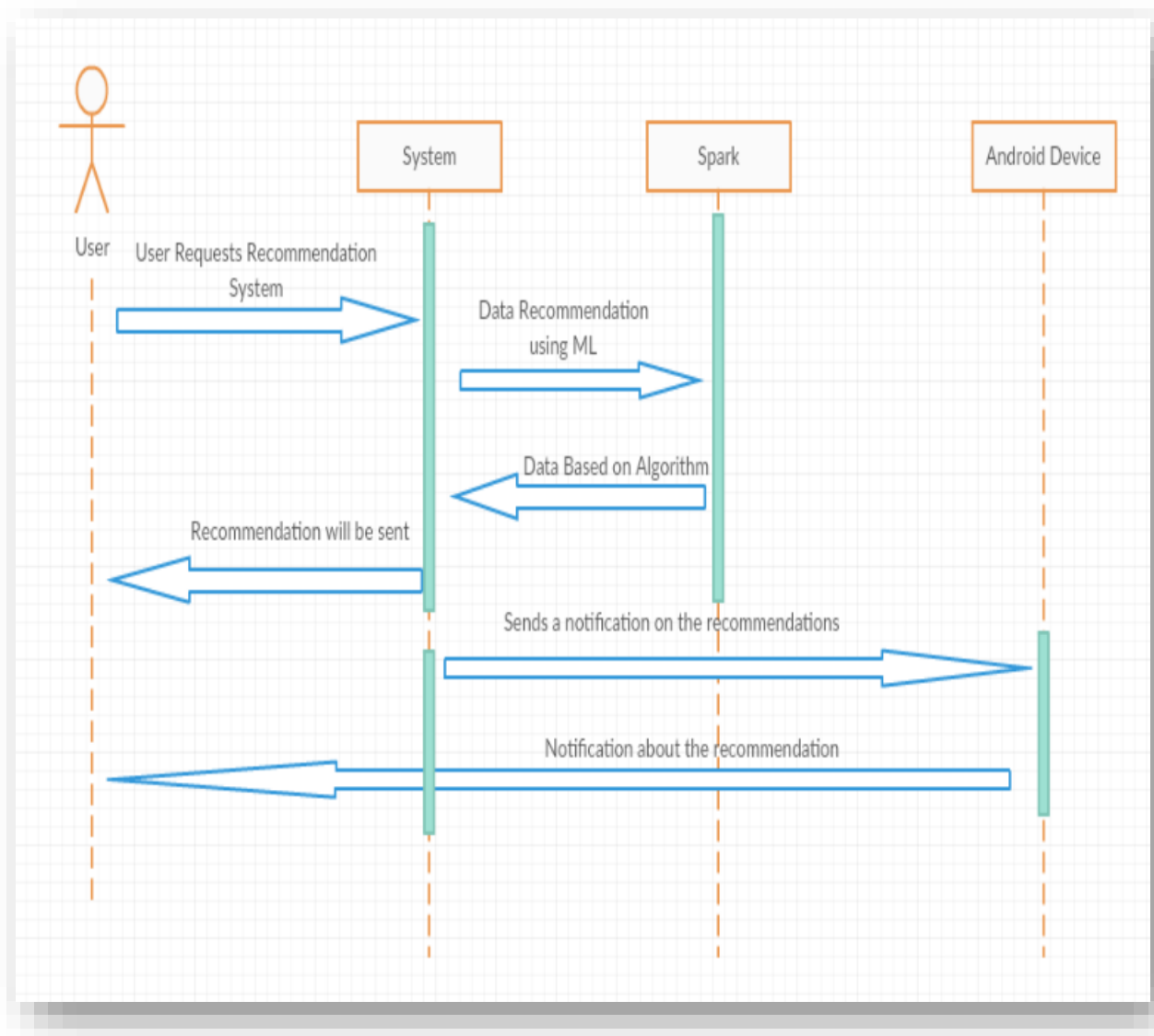
7 IMPLEMENTATION

7.1 ARCHITECTURE & UML DIAGRAMS:

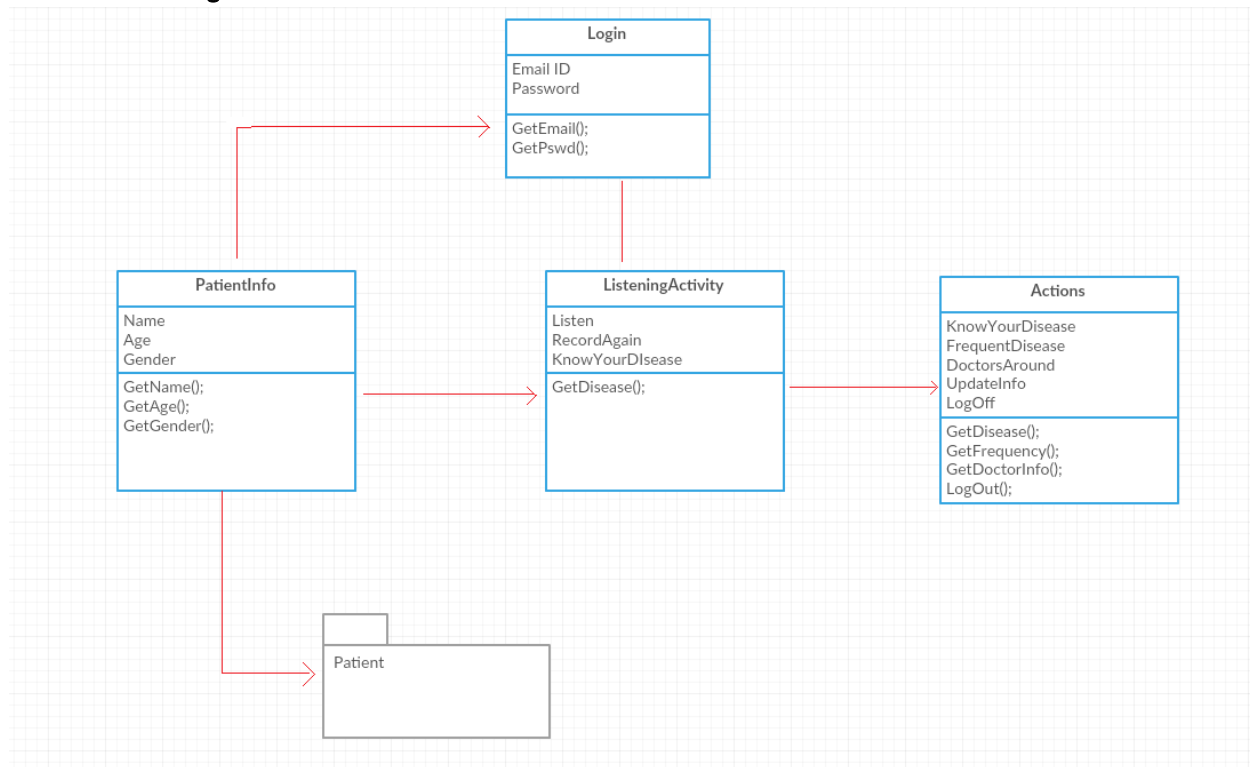


In traditional way, diagnostic test will be done first and then a drug will be used to a person where if he responds to it, then doctor will recommend to continue that drug A for better outcome, if by any reason the patient will not respond to that drug A, then doctor will recommend drug B. At-last if patient shows some severe side effects even after using drug B, then doctor will recommend drug C which is a long and time taking process. This can be changed by the above proposed approach.

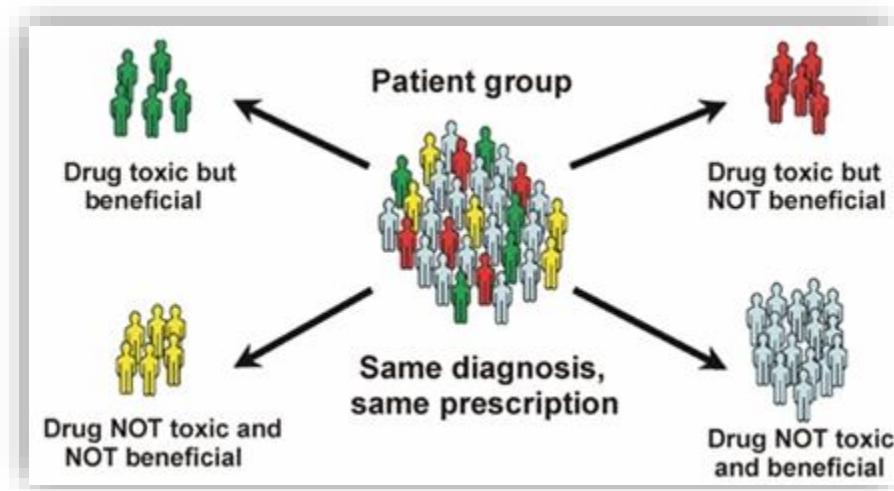
7.1.1 Sequence Diagram:



7.1.2 Class Diagram:



7.2 WORKFLOW:



Here we have a data set with group of people. Later this machine should segregate the data to groups like drug toxic but beneficial for some time, Drug toxic but NOT beneficial, Drug Not Toxic and Not Beneficial and Drug Not Toxic and Beneficial (this group will be ideal for data collection). Based on this we will give rating to the drugs for that particular group and later we will use that data for drug recommendation.

We have used the following Existing services:

Natural Language Processing (NLP) is the technique that we have used to process the input given by the User. Here the NLP takes the input as a symptom and it processes the data to enable the computer to derive meaning.

NLP Data provided:

The screenshot shows the IntelliJ IDEA 15.0.3 IDE interface. The top menu bar includes File, Edit, View, Navigate, Code, Analyze, Refactor, Build, Run, Tools, VCS, Window, and Help. The title bar indicates the project is 'Tutorial-3-CoreNLP' located at '[C:\Users\mpdar\Desktop\Acad\XDM\Tutorial\Code\Tutorial-3-CoreNLP] - [Tutorial-3-corenlp] - [src\main\java\Main.java - IntelliJ IDEA 15.0.3]'. The left sidebar shows the Project structure with 'Tutorial-3-CoreNLP [Tutorial-3-corenlp]' as the root, containing 'idea', 'out', 'project [Tutorial-3-corenlp-build] (sources root)', 'src', 'main', and 'java'. The 'Main' class is selected under 'src/main/java'. The main editor window shows the 'Main.java' file with the following code:

```

public class Main {
    // creates a StanfordCoreNLP object, with POS tagging, lemmatization, NER, parsing, and coreference resolution
    Properties props = new Properties();
    props.setProperty("annotators", "tokenize, ssplit, pos, lemma, ner, parse, dooref");
    StanfordCoreNLP pipeline = new StanfordCoreNLP(props);

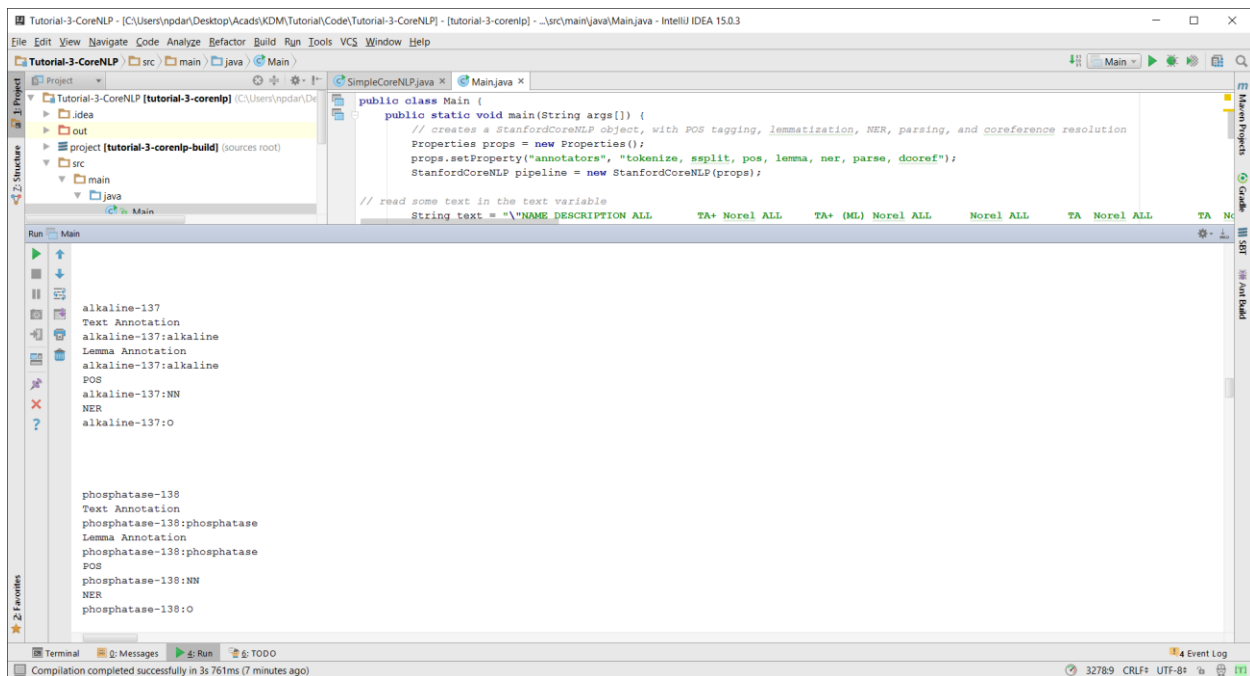
    // read some text in the text variable
    String text = "\\NAME DESCRIPTION ALL          TA+ Norel ALL          TA+ (MG) Norel ALL          Norel ALL          TA Norel ALL          TA No
        \"
        \" \\ALPPL ALPPL :alkaline phosphatase, placental like
        \" \\RPLP RPLP :ribosomal protein, large, P
        \" \\OPER OPER: growth factor, augmentor of liver regeneration (ERV homolog, 8 cervisiae)
        \" \\IGHM IGHM: immunoglobulin heavy constant mu
        \" \\SANT SANT :squeous cell carcinoma antigen recognised by T cells
        \" \\RPLP RPLP :ribosomal protein, large, P
        \" \\RPLP RPLP :ribosomal protein, large, P
        \" \\PDHA PDHA :pyruvate dehydrogenase (lipoamide) alpha
        \" \\OGT OGT :O linked N acetylglucosamine (GlcNAc) transferase (UDP N acetylglucosamine: polypeptide N ace
        \" \\PDHA PDHA :pyruvate dehydrogenase (lipoamide) alpha
        \"
        \" Type: In word PDHA more (Ctrl+F1)
        \"
        \" \\TMEFF TMEFF :transmembrane protein with EGF like and two follistatin like domains
        \" \\LRRC LRRC :leucine rich repeat containing \" // Add your text here!

    // create an empty Annotation just with the given text
    Annotation document = new Annotation(text);

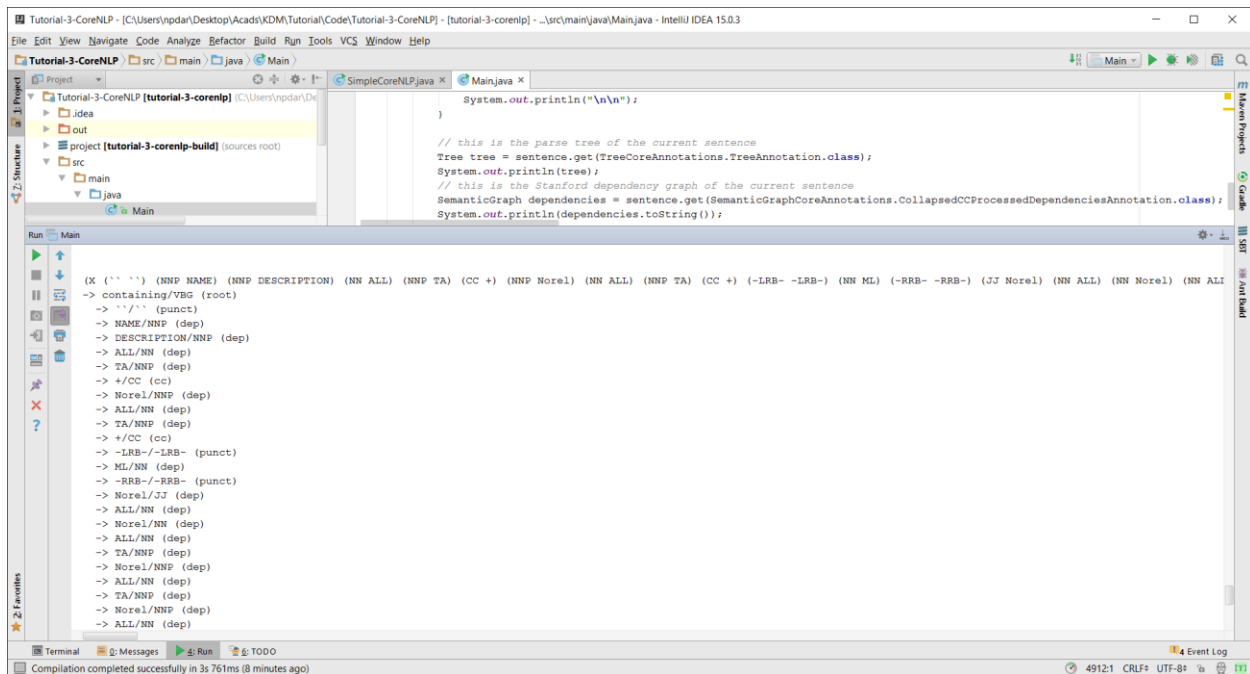
    // run all Annotators on this text
    pipeline.annotate(document);

```

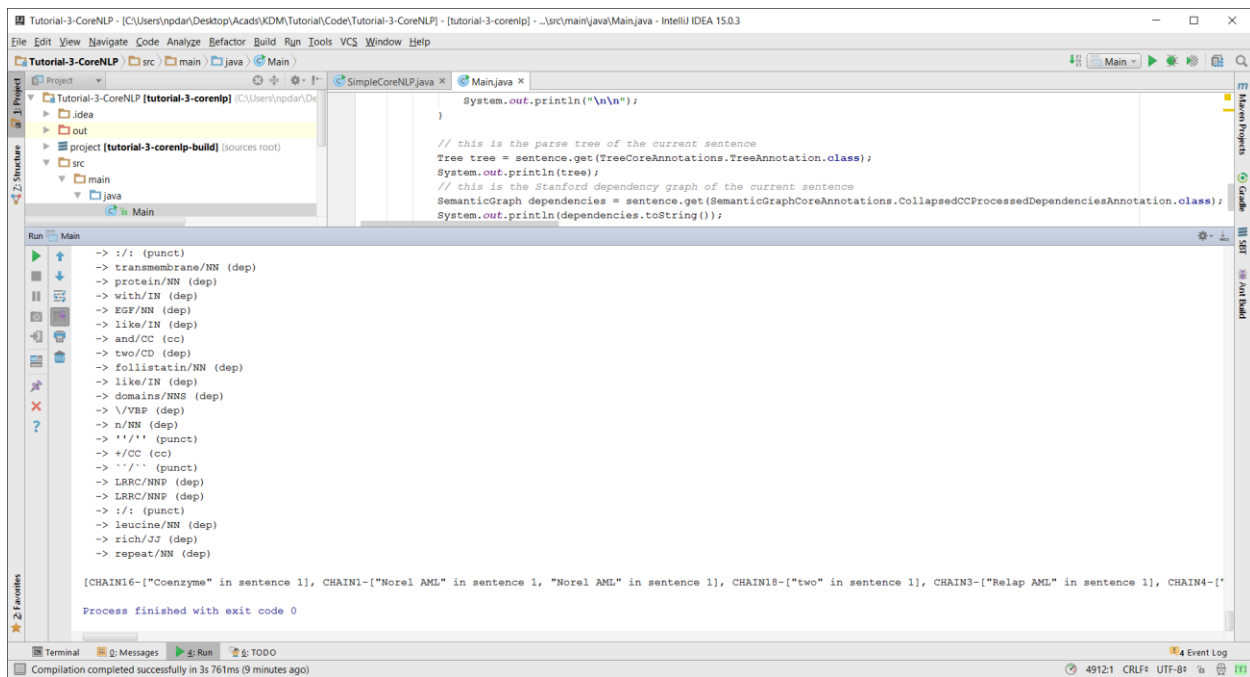
The bottom status bar shows 'Compilation completed successfully in 3s 761ms (6 minutes ago)'. The bottom right corner shows 'Event Log'.



Parse tree:



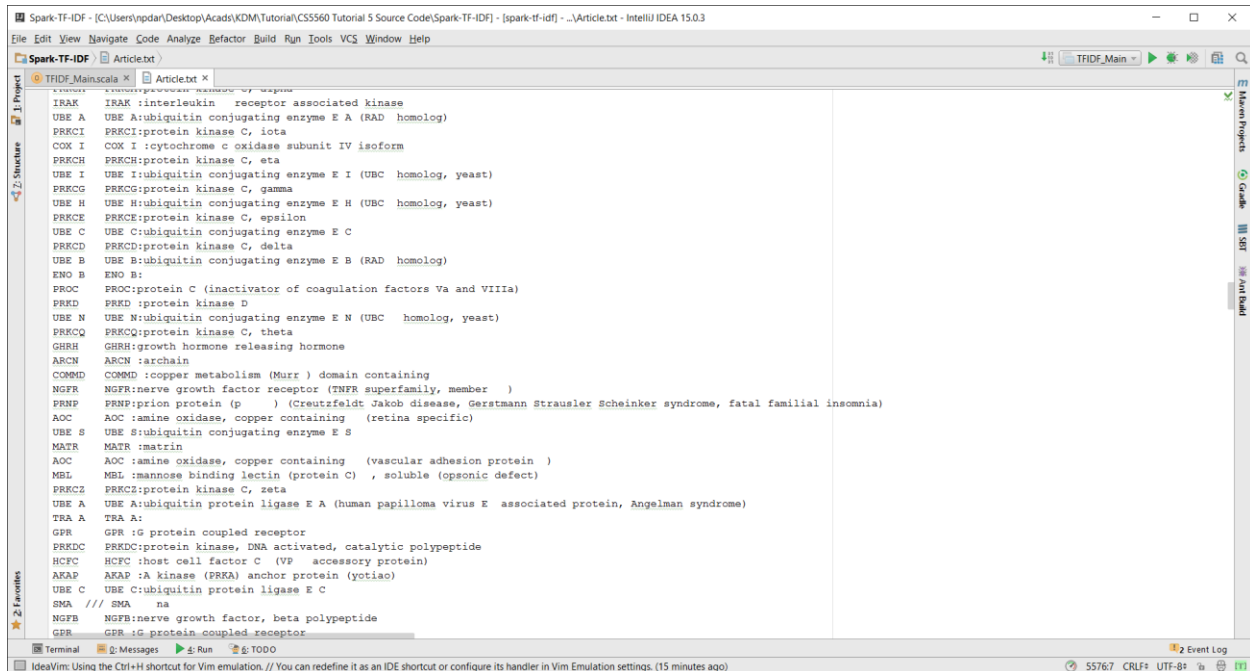
Process execution:



7.4 FEATURE IMPLEMENTED:

TF-IDF: We have used TF-IDF to extract the top drugs that can be recommended to the user based on the symptoms and the criteria of the overall system. It reflects how important the word is to a document or to a dataset. Based on the provided statistics, we consider the similarities with the training data and extract the relevant top 10 drugs that can be recommended.

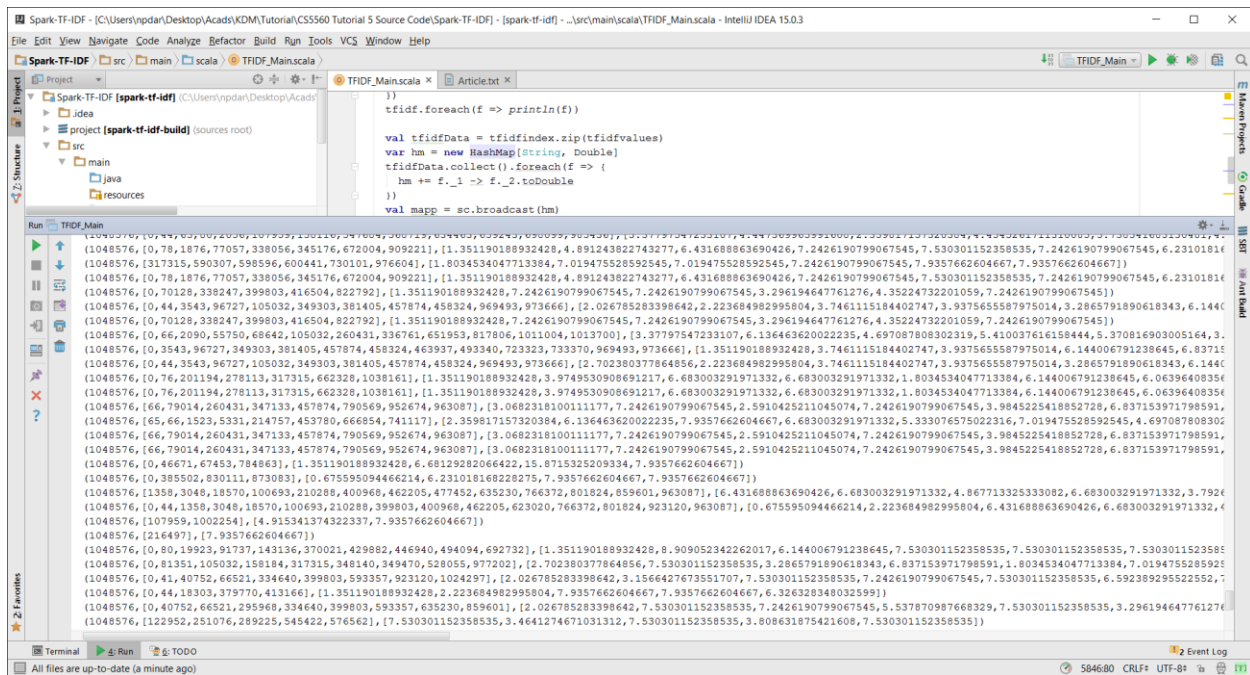
TF-IDF – Dataset:



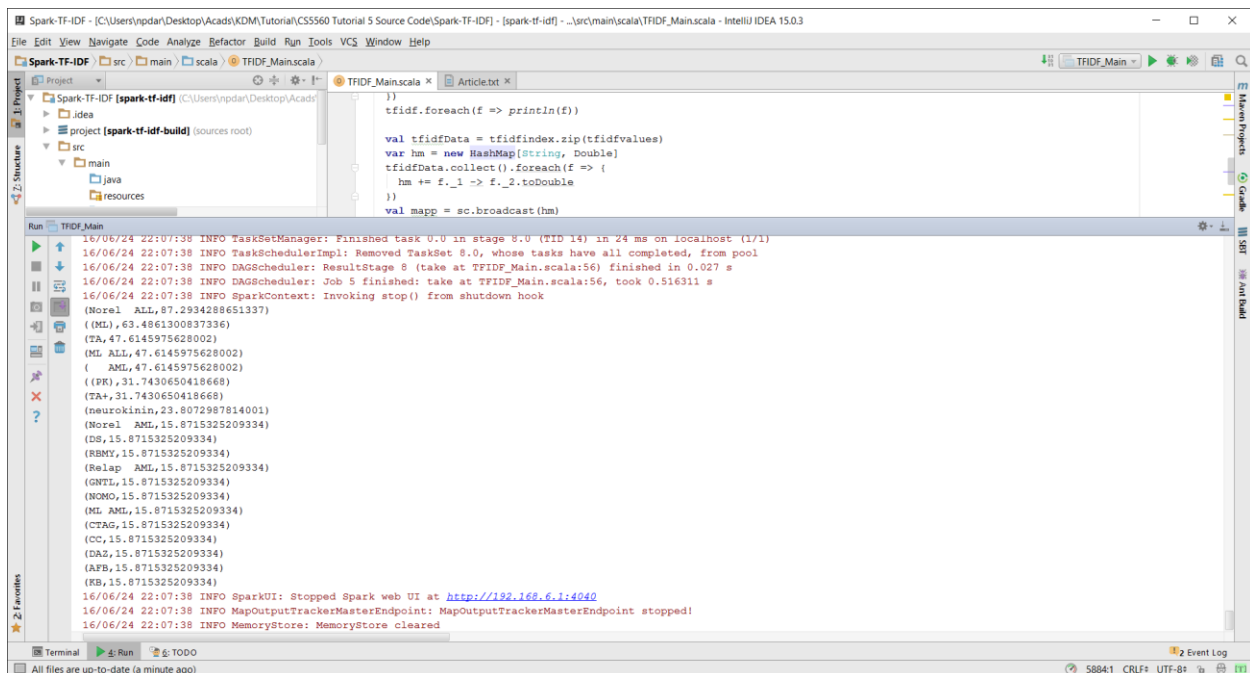
The screenshot shows an IDE window titled "Spark-TF-IDF - [C:\Users\ipdar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF] - [spark-tf-idf] - \Article.txt - IntelliJ IDEA 15.0.3". The main editor displays a list of proteins and their descriptions, organized in a table-like format. The proteins are listed in the first column, and their descriptions are in the second column. The descriptions are often truncated with ellipses (...).

Protein	Description
IRAK	IRAK :interleukin receptor associated kinase
UBE A	UBE A:ubiquitin conjugating enzyme E A (RAD homolog)
PRKCI	PRKCI:protein kinase C, iota
COX I	COX I :cytochrome c oxidase subunit IV isoform
PRKCH	PRKCH:protein kinase C, eta
UBE I	UBE I:ubiquitin conjugating enzyme E I (UBC homolog, yeast)
PRKCG	PRKCG:protein kinase C, gamma
UBE H	UBE H:ubiquitin conjugating enzyme E H (UBC homolog, yeast)
PRKCE	PRKCE:protein kinase C, epsilon
UBE C	UBE C:ubiquitin conjugating enzyme E C
PRKCD	PRKCD:protein kinase C, delta
UBE B	UBE B:ubiquitin conjugating enzyme E B (RAD homolog)
ENO B	ENO B:
PROC	PROC:protein C (inactivator of coagulation factors Va and VIIIA)
PRKD	PRKD :protein kinase D
UBE N	UBE N:ubiquitin conjugating enzyme E N (UBC homolog, yeast)
PRKCG	PRKCG:protein kinase C, theta
GHRH	GHRH:growth hormone releasing hormone
ARCN	ARCN :archain
COMMD	COMMD :copper metabolism (Murr) domain containing
NGFR	NGFR:nerve growth factor receptor (TNFR superfamily, member)
PRNP	PRNP:prion protein (p) (Creutzfeldt Jakob disease, Gerstmann Strausler Scheinker syndrome, fatal familial insomnia)
AOC	AOC :amine oxidase, copper containing (retina specific)
UBE S	UBE S:ubiquitin conjugating enzyme E S
MATR	MATR :matrin
AOC	AOC :amine oxidase, copper containing (vascular adhesion protein)
MBL	MBL :mannose binding lectin (protein C) , soluble (opsonic defect)
PRKCE	PRKCE:protein kinase C, zeta
UBE A	UBE A:ubiquitin protein ligase E A (human papilloma virus E associated protein, Angelman syndrome)
TRA A	TRA A:
GPR	GPR :G protein coupled receptor
PRKDC	PRKDC:protein kinase, DNA activated, catalytic polypeptide
HCFC	HCFC :host cell factor C (VP accessory protein)
AKAP	AKAP :A kinase (PRKA) anchor protein (yotiao)
UBE C	UBE C:ubiquitin protein ligase E C
SMA	SMA :SMA
NGFB	NGFB:nerve growth factor, beta polypeptide
GPR	GPR :G protein coupled receptor

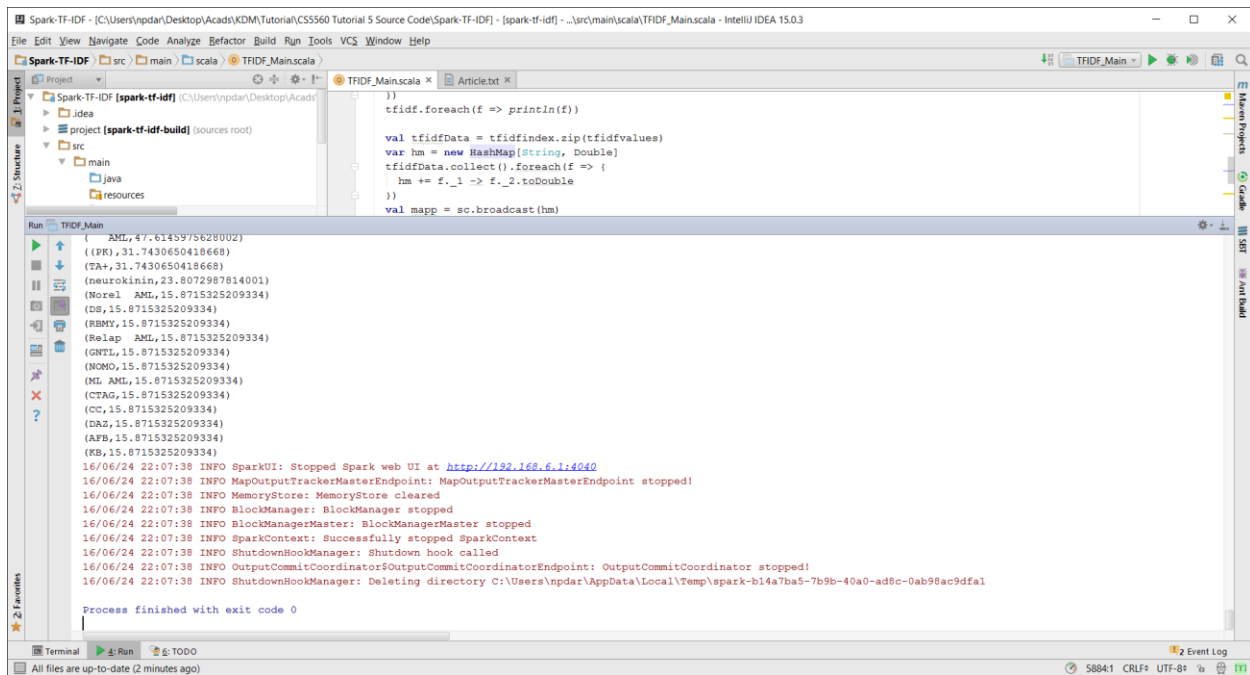
Output:



Top 10 terms:



Process Success:



```
Spark-TF-IDF - [C:\Users\npdar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF] - [spark-tf-idf] - ...src\main\scala\TFIDF_Mainscala - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

Spark-TF-IDF [src] main scala TFIDF_Mainscala

Project Structure
- Spark-TF-IDF [spark-tf-idf] (C:\Users\npdar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 5 Source Code\Spark-TF-IDF)
  - idea
  - project (spark-tf-idf-build) (sources root)
    - src
      - main
        - java
        - resources
      - test
      - target
      - build.sbt
      - input

Run TFIDF_Main

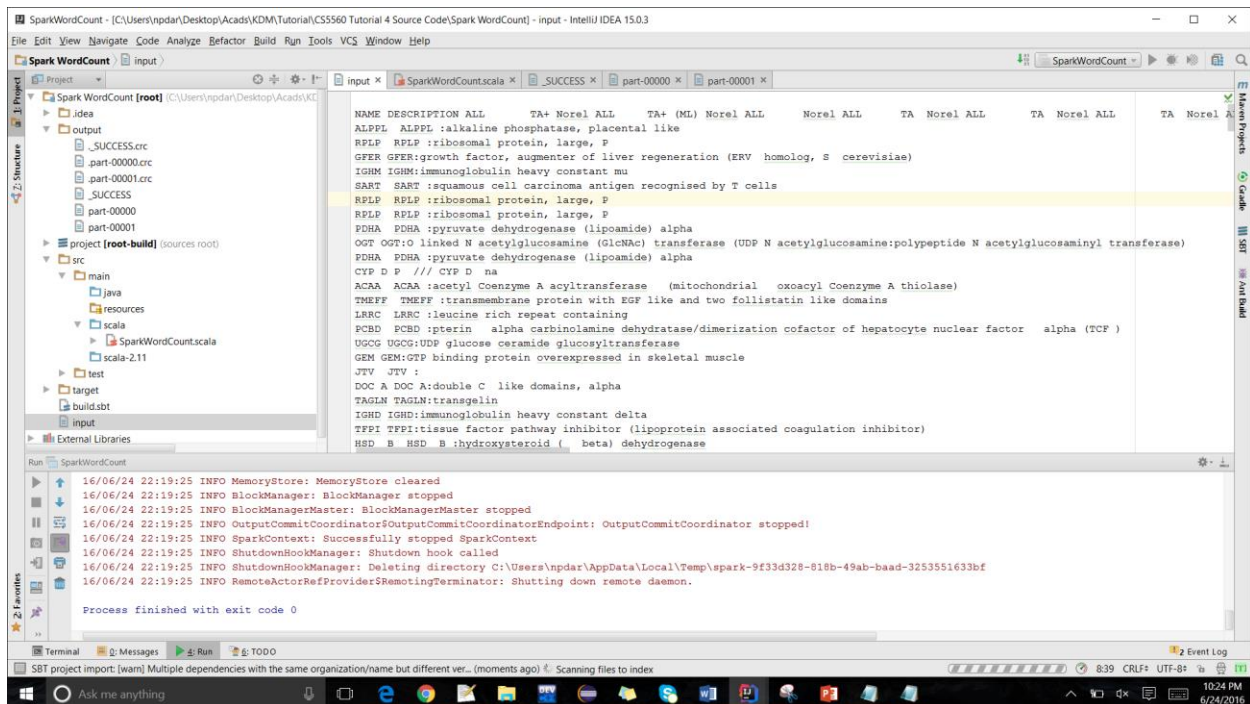
((AMEL,47.614597562002))
((PKI,31.7430650418668))
((TA+,31.7430650418668))
(neurokinin,23.8072987814001))
(Norel AME,15.8715325209334)
(DS,15.8715325209334)
(RBMY,15.8715325209334)
(Relap AME,15.8715325209334)
(GNTL,15.8715325209334)
(NOMO,15.8715325209334)
(ML AME,15.8715325209334)
(CTAG,15.8715325209334)
(CC,15.8715325209334)
(DAG,15.8715325209334)
(AFB,15.8715325209334)
(KB,15.8715325209334)

16/06/24 22:07:38 INFO SparkUI: Stopped Spark web UI at http://192.168.6.1:4040
16/06/24 22:07:38 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/06/24 22:07:38 INFO MemoryStore: MemoryStore cleared
16/06/24 22:07:38 INFO BlockManager: BlockManager stopped
16/06/24 22:07:38 INFO BlockManagerMaster: BlockManagerMaster stopped
16/06/24 22:07:38 INFO SparkContext: Successfully stopped SparkContext
16/06/24 22:07:38 INFO ShutdownHookManager: Shutdown hook called
16/06/24 22:07:38 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/06/24 22:07:38 INFO ShutdownHookManager: Deleting directory C:\Users\npdar\AppData\Local\Temp\spark-b14a7ba5-7b9b-40a0-ad8c-0ab9ac9dfa1

Process finished with exit code 0
```

Word Count:

Input file:



```
SparkWordCount - [C:\Users\npdar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 4 Source Code\Spark WordCount] - input - IntelliJ IDEA 15.0.3

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

SparkWordCount [input]

Project Structure
- Spark WordCount [root] (C:\Users\npdar\Desktop\Acads\KDM\Tutorial\CS5560 Tutorial 4 Source Code\Spark WordCount)
  - idea
  - output
    - _SUCCESS.crc
    - part-00000.crc
    - part-00001.crc
    - _SUCCESS
    - part-00000
    - part-00001
  - project (root-build) (sources root)
    - src
      - main
        - java
        - resources
      - scala
        - SparkWordCount.scala
        - scala-2.11
      - test
      - target
      - build.sbt
      - input
  - External Libraries

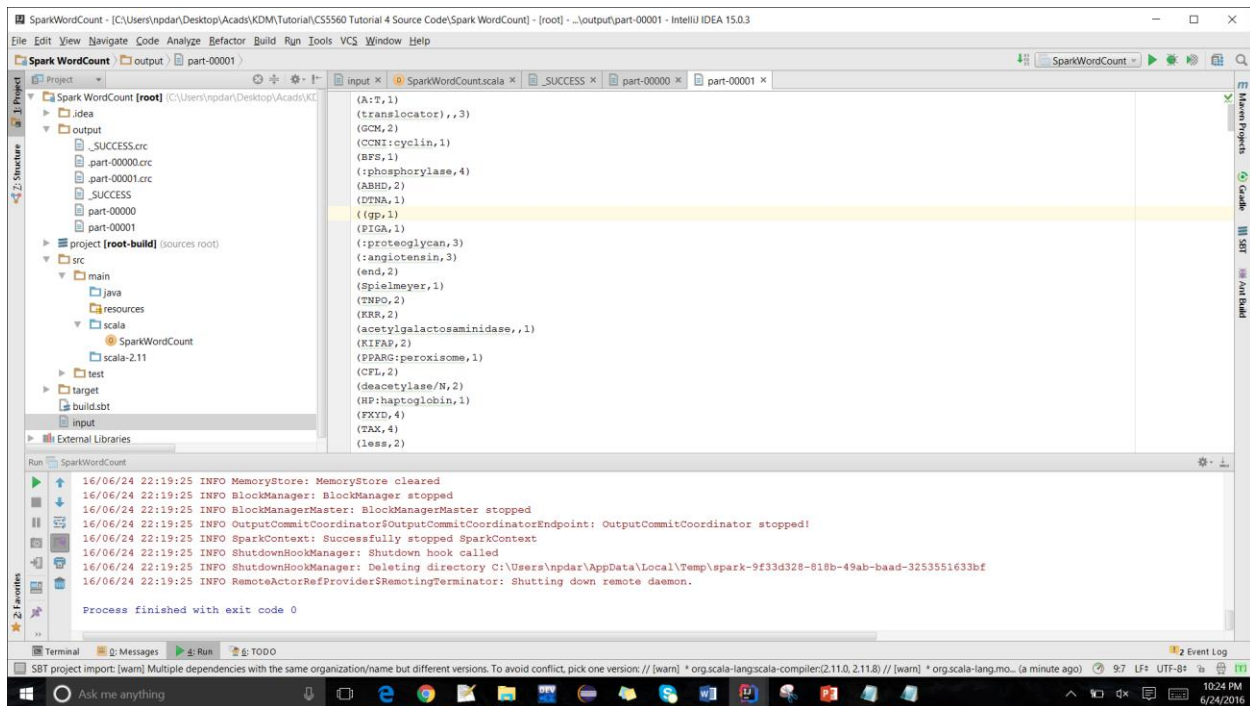
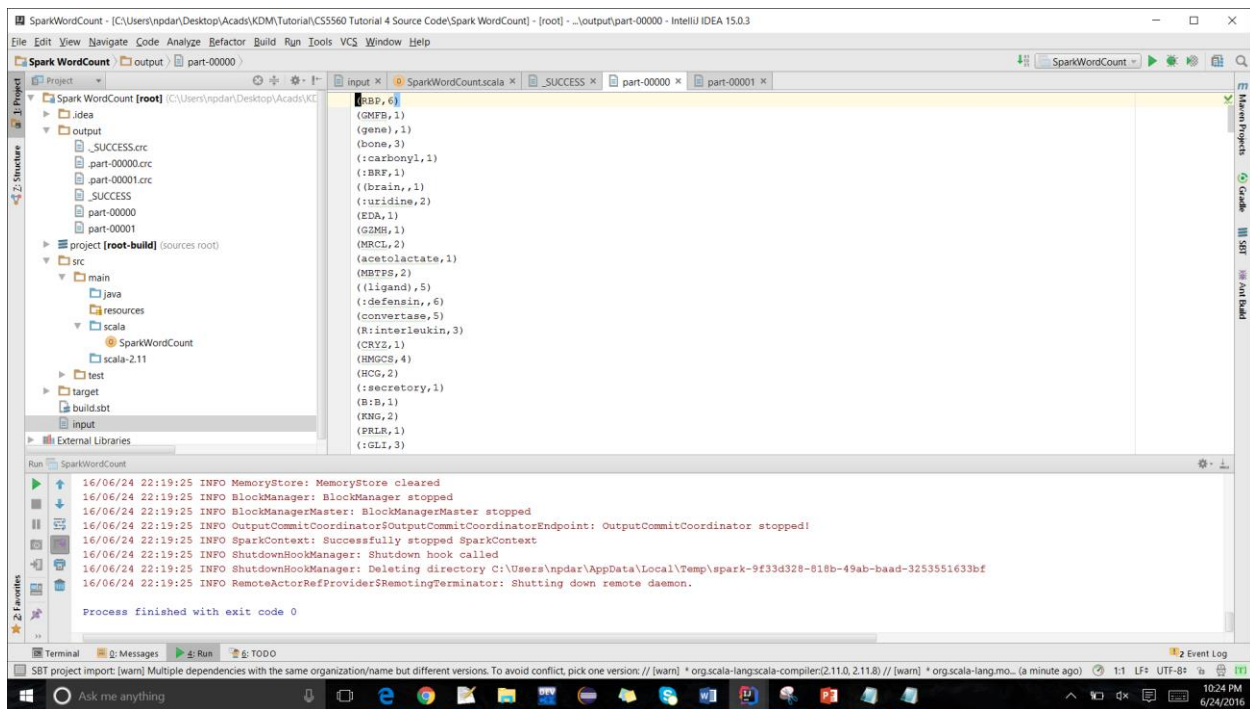
Run SparkWordCount

NAME DESCRIPTION ALL TA+ Norel ALL TA+ (ML) Norel ALL Norel ALL TA Norel ALL TA Norel ALL TA Norel ALL
ALPPL ALPPL :alkaline phosphatase, placental like
RPLP RPLP :ribosomal protein, large, P
GFER GFER: growth factor, augmenter of liver regeneration (ERV homolog, S cerevisiae)
IGHM IGHM: immunoglobulin heavy constant mu
SART SART :squamous cell carcinoma antigen recognised by T cells
RPLP RPLP :ribosomal protein, large, P
RPLP RPLP :ribosomal protein, large, P
PDHA PDHA :pyruvate dehydrogenase (lipoamide) alpha
OAT OAT: O linked N acetylglucosamine (GlcNAc) transferase (UDP N acetylglucosamine: polypeptide N acetylglucosaminyl transferase)
PDHA PDHA :pyruvate dehydrogenase (lipoamide) alpha
CYP D P /// CYP D na
ACAA ACAA :acetyl Coenzyme A acyltransferase (mitochondrial oxoacyl Coenzyme A thiolase)
TMEFF TMEFF :transmembrane protein with EGF like and two follistatin like domains
LRRC LRRC :leucine rich repeat containing
PCBD PCBD :pterin alpha carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor alpha (TCF)
UGCG UGCG: UDP glucose ceramide glucosyltransferase
GEM GEM: GTP binding protein overexpressed in skeletal muscle
JTV JTV :
DOC A DOC A: double C like domains, alpha
TAGLN TAGLN: transgelin
IGHD IGHM: immunoglobulin heavy constant delta
TFPI TFPI: tissue factor pathway inhibitor (lipoprotein associated coagulation inhibitor)
HSD B HSD B :hydroxysteroid ( beta) dehydrogenase

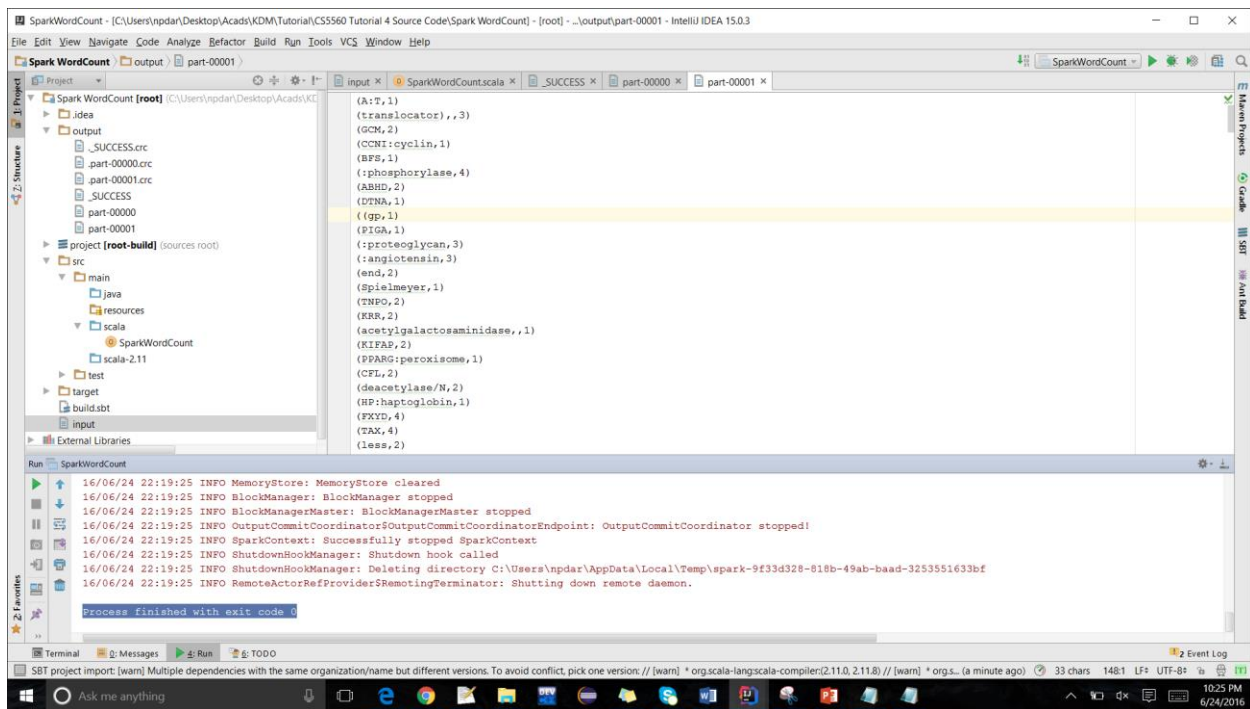
16/06/24 22:19:25 INFO MemoryStore: MemoryStore cleared
16/06/24 22:19:25 INFO BlockManager: BlockManager stopped
16/06/24 22:19:25 INFO BlockManagerMaster: BlockManagerMaster stopped
16/06/24 22:19:25 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/06/24 22:19:25 INFO SparkContext: Successfully stopped SparkContext
16/06/24 22:19:25 INFO ShutdownHookManager: Shutdown hook called
16/06/24 22:19:25 INFO RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.

Process finished with exit code 0
```

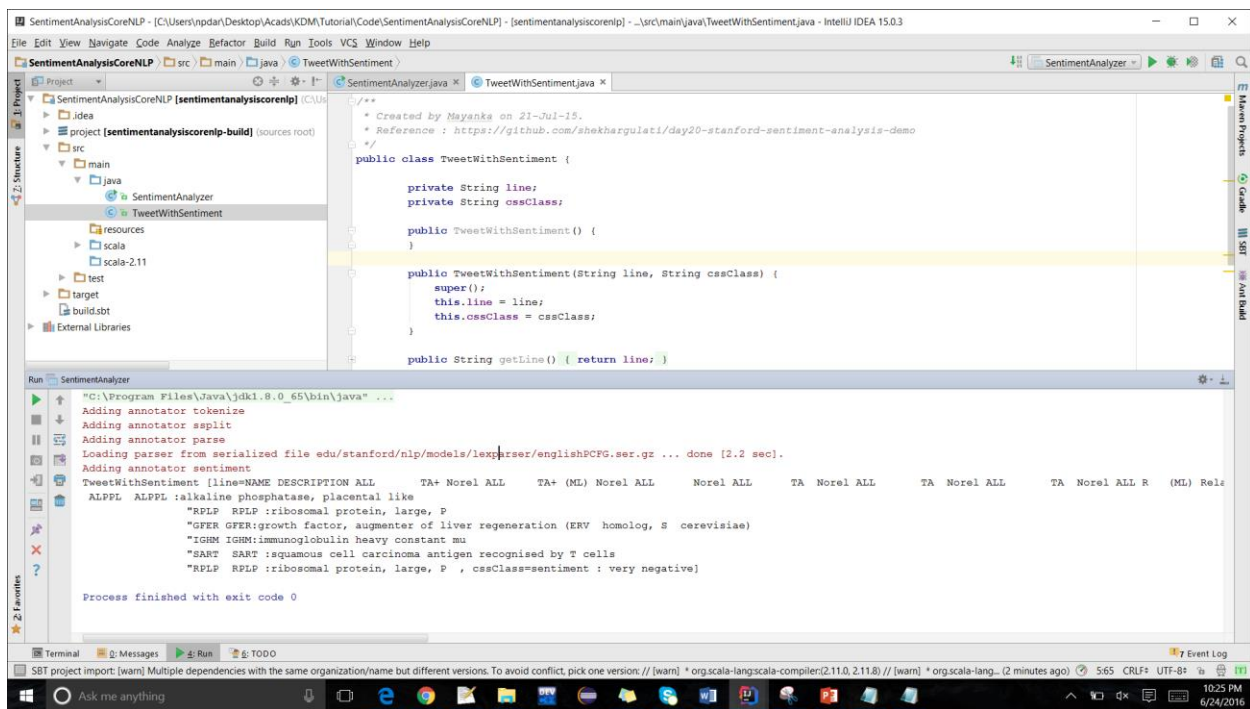
Output file:



Process execution:



Sentimental Analysis:



8 PROJECT MANAGEMENT

8.1 CONTRIBUTION:

Overall – 100%

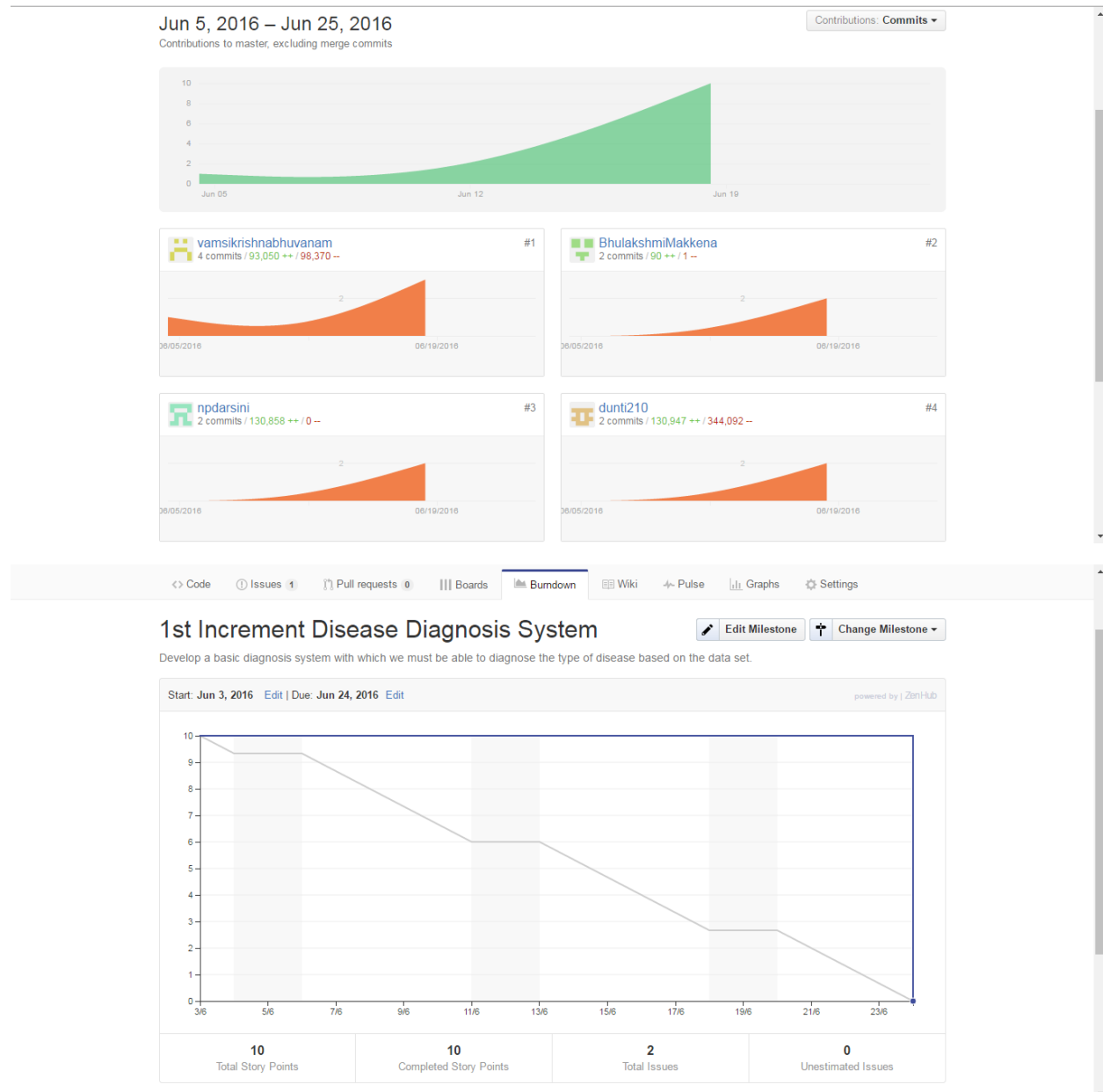
Venkata Vamsi Krishna Bhuvanam – 25%


Priyadarsini Nidadavolu – 25%

Bhulakshmi Makkena – 25%

Tej Kumar Yentrapragada – 25%




8.2 ZENHUB AND GITHUB:





This repository Search

[Pull requests](#) [Issues](#) [Gist](#) [+ ToDo](#)

Summer2016-KDM / KDM-Summer-2016-ProjectDraft

Unwatch

1

Star

0

Fork

0

Code

Issues1

Pull requests0

Boards

Burndown

Wiki

Pulse


Graphs

Settings

No description or website provided. — Edit

12 commits 1 branch 0 releases 5 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

 dunt210 Commit Latest commit 2c50b8d 10 minutes ago

1st Increment/Source/Spark-TF-IDFChanged	Commit	4 minutes ago
1stIncrement	Commit	4 minutes ago
README.md	Initial commit	17 days ago

README.md

KDM-Summer-2016-ProjectDraft

8.3 CONCERNS/ISSUES:

NA

8.4 FUTURE WORK:

Concentrating to continue to work on Drug Dataset collection and continue to build a recommendation system which will recommend drug to the user.