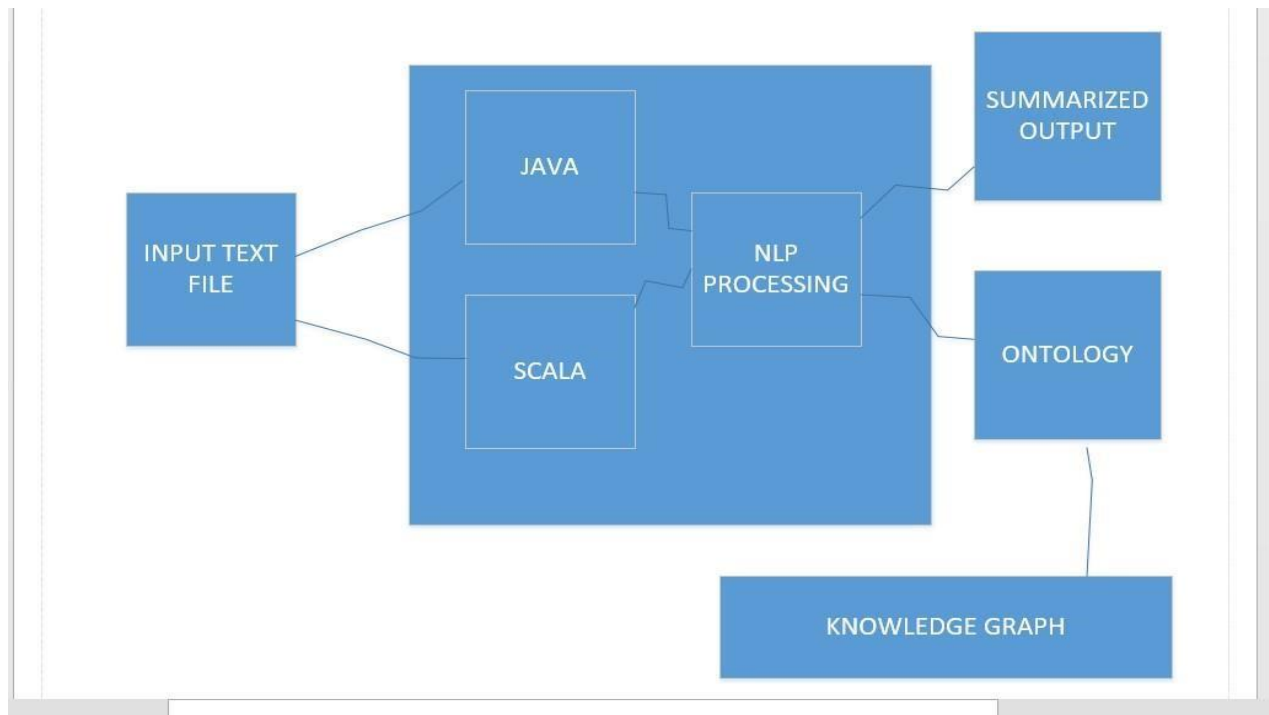
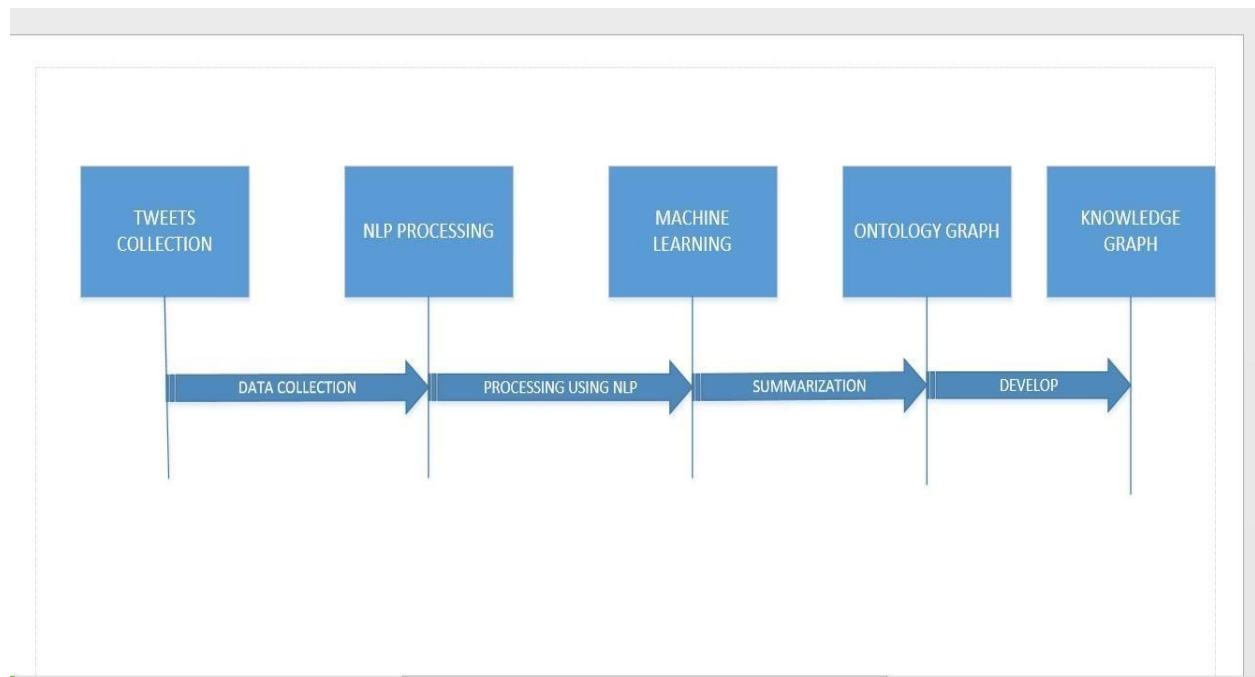


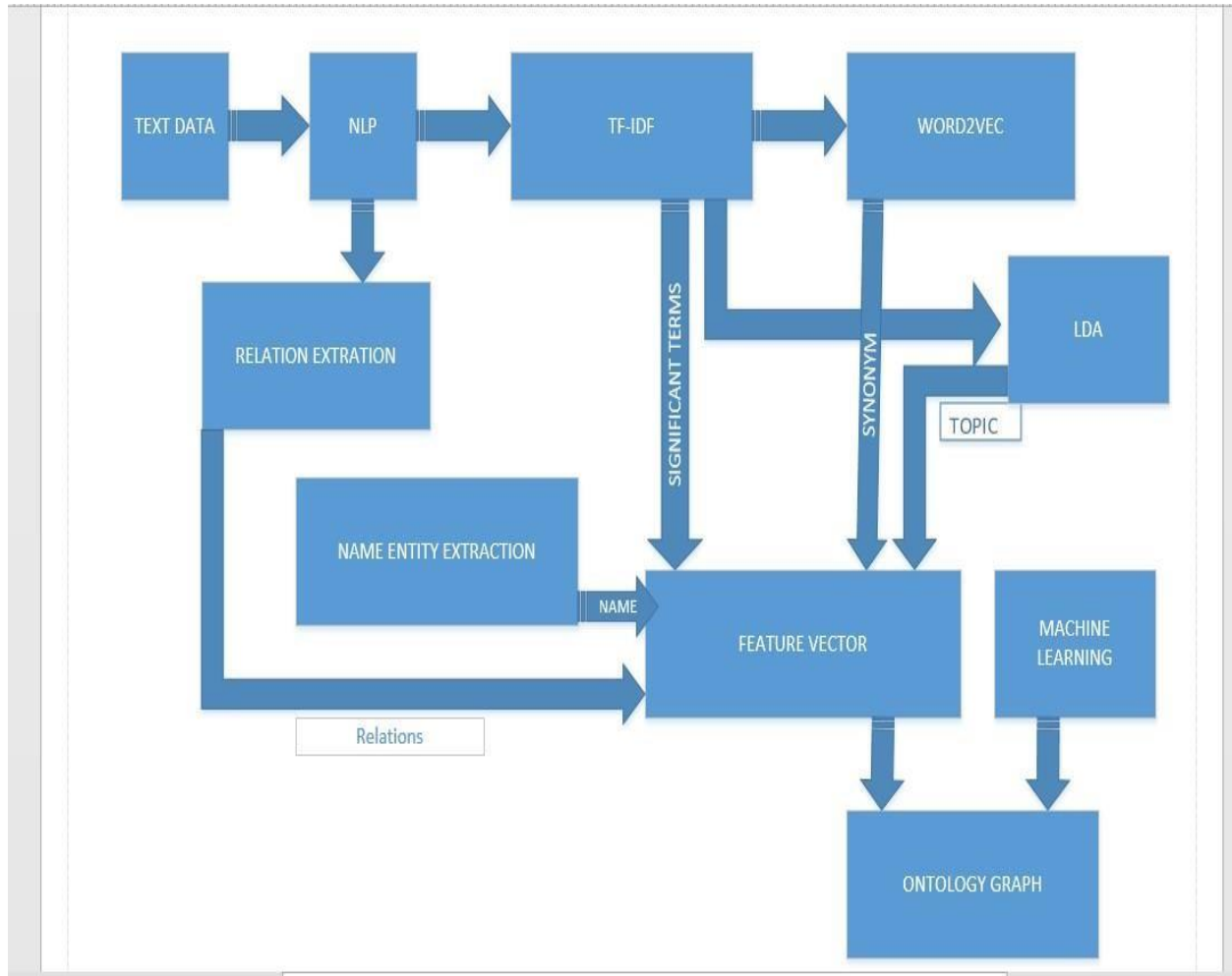
## SOFTWARE ARCHITECTURE DIAGRAM ::



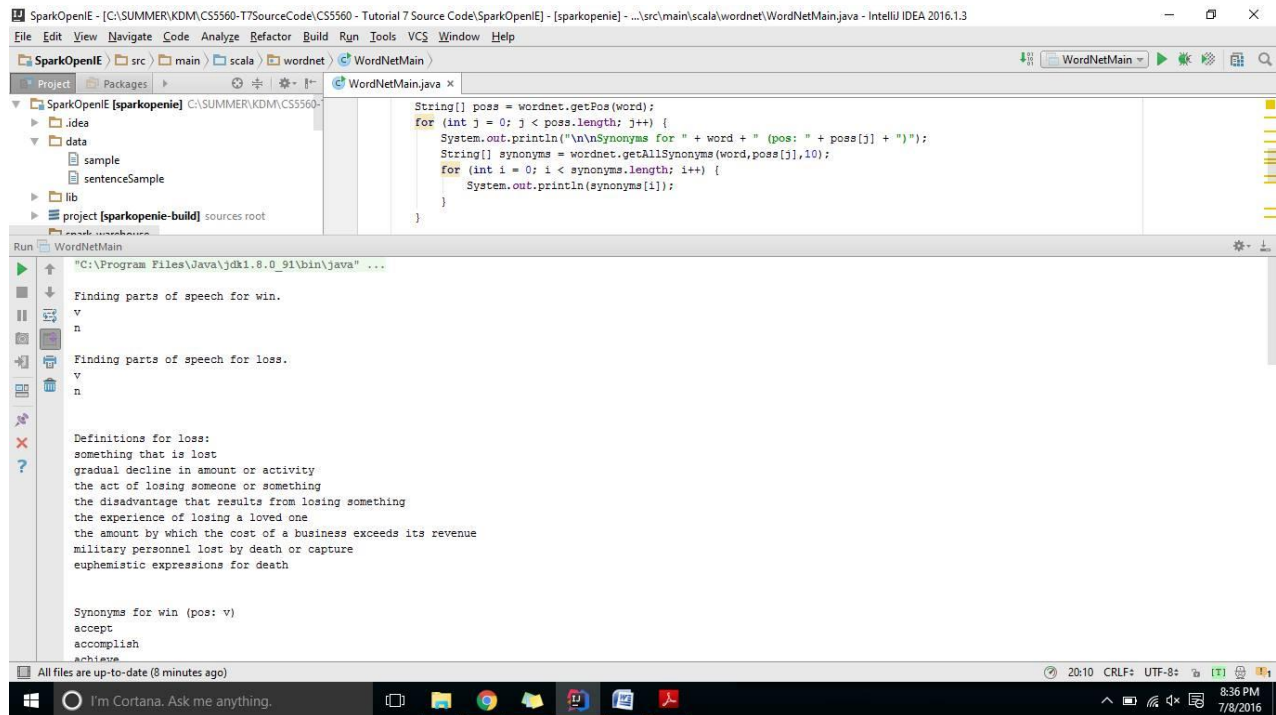
## SEQUENCE DIAGRAM ::



WORKFLOW DIAGRAM ::



## Wordnet output :



The screenshot shows the IntelliJ IDEA interface with the file `WordNetMain.java` open. The code iterates through wordnet data and prints parts of speech and synonyms for the word "loss". The Run window displays the following output:

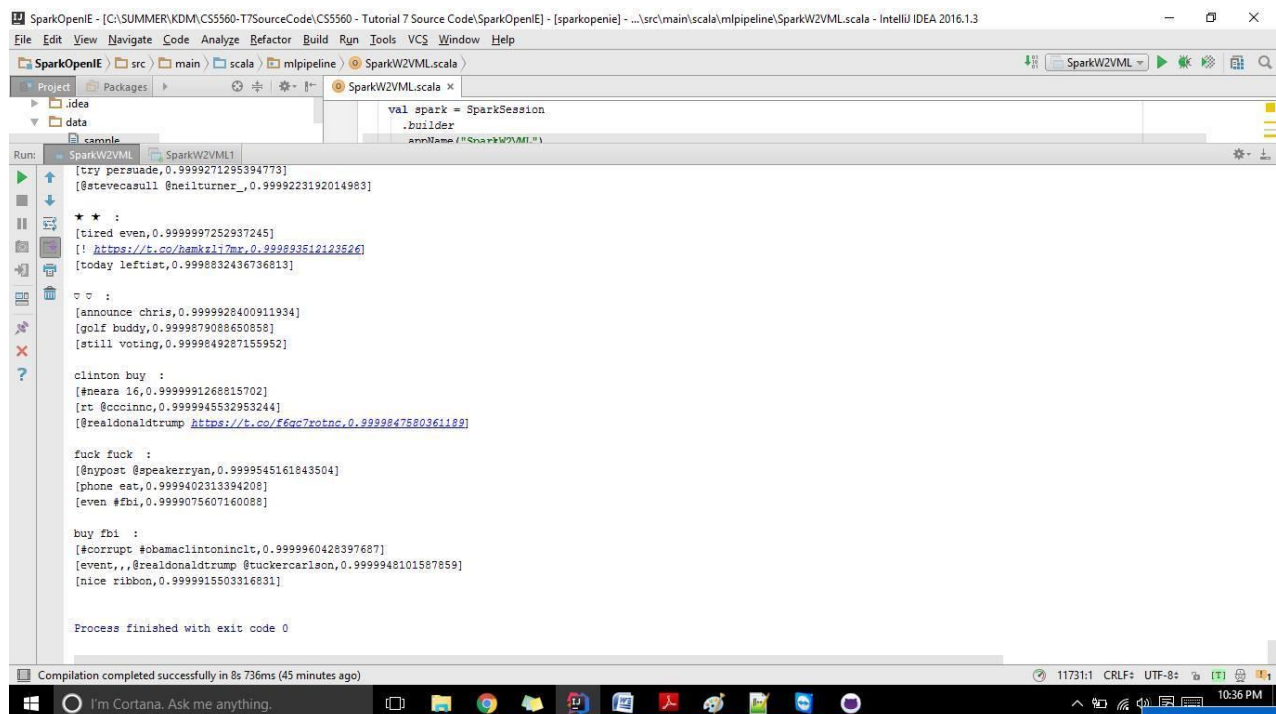
```
Finding parts of speech for win.
v
n

Finding parts of speech for loss.
v
n

Definitions for loss:
something that is lost
gradual decline in amount or activity
the act of losing someone or something
the disadvantage that results from losing something
the experience of losing a loved one
the amount by which the cost of a business exceeds its revenue
military personnel lost by death or capture
euphemistic expressions for death

Synonyms for win (pos: v)
accept
accomplish
achieve
```

## N-Gram and word2vec output :



The screenshot shows the IntelliJ IDEA interface with the file `SparkW2VML.scala` open. The Run window displays the following output:

```
[try persuade,0.9999271295394773]
[stevecasull1 @neilturner_,0.999223192014983]

★ ★ :
[tired even,0.999997252937245]
[! https://t.co/hams1i7mx,0.998893512123526]
[today leftist,0.9998832436736813]

▽ ▽ :
[announce' chris,0.9999928400911934]
[golf buddy,0.999987908650858]
[still voting,0.9999849287155952]

clinton buy :
[#nears 16,0.9999991268815702]
[rt @ccinnnc,0.9999945532953244]
[@realDonaldTrump https://t.co/f6gc7kotnc,0.9998847580361182]

fuck fuck :
[@nypost @speakerryan,0.9999545161843504]
[phone eat,0.9999402313384208]
[even #fbi,0.9999075607160088]

buy fbi :
[#corrupt @obamacintoninict,0.9999960428397687]
[event,,@realDonaldTrump @tucker Carlson,0.9999948101587859]
[nice ribbon,0.999915503316831]

Process finished with exit code 0
```

## Name Entity Extraction output :

```
SparkNER
↑ (spend ,O )
↓ (the ,O )
  (last ,O )
  (16 ,NUMBER )
  (month ,DURATION )
  (deliberately ,O )
  (lie ,O )
  (to ,O )
  (the ,O )
  (American ,MISC )
  (people . ,O O )
  (watch : ,O O )
  (https://t.co/18kiu6b69a ,O )
  (, POLL)
16/07/08 21:58:36 INFO SparkContext: Invoking stop() from shutdown hook
16/07/08 21:58:36 INFO SparkUI: Stopped Spark web UI at http://192.168.0.20:4040
16/07/08 21:58:36 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/07/08 21:58:36 INFO MemoryStore: MemoryStore cleared
16/07/08 21:58:36 INFO BlockManager: BlockManager stopped
16/07/08 21:58:36 INFO BlockManagerMaster: BlockManagerMaster stopped
16/07/08 21:58:36 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/07/08 21:58:36 INFO SparkContext: Successfully stopped SparkContext
16/07/08 21:58:36 INFO ShutdownHookManager: Shutdown hook called
16/07/08 21:58:36 INFO ShutdownHookManager: Deleting directory C:\Users\vilas\AppData\Local\Temp\spark-d3b84157-3504-4f3e-ac9c-54f198249045

Process finished with exit code 0
```

```
Run SparkNER
↑ (El_Universal_Mx : ,O O )
  ({entérate ,O })
↓ (Lynch ,PERSON )
  (y ,O )
  (Comey ,PERSON )
  (comparecerán ,O )
  (ante ,O )
  (Congreso ,PERSON )
  (por ,O )
  (caso ,O )
  (HillaryClinton ,PERSON )
  (https://t.co/wilhlpatd6 ,O )
  (, POLL)
  (rt ,O )
  (@Reince : ,O O )
  (. @hillaryclinton ,O O )
  (spend ,O )
  (the ,O )
  (last ,O )
  (16 ,NUMBER )
  (month ,DURATION )
  (deliberately ,O )
  (lie ,O )
  (to ,O )
  (the ,O )
  (American ,MISC )
  (people . ,O O )
```

## Topic discovery output :

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

Spark\_KMeans\_FV | data | LDA\_Results.txt

Project Packages Project Files SparkLDAMain.scala LDA\_Results.txt build.sbt

1-Project

Structure

Spark\_KMeans\_FV [spark\_kmeans\_fv] (E:\study\Study\KDM)

idea

data

20\_twitter\_wiki

hillary

politics

politics2

trump

DT\_Results.txt

LDA\_Results.txt

NB\_Results.txt

Results.txt

RF\_Results.txt

Run SparkLDAMain

Corpus summary:

Training set size: 10704 documents

Vocabulary size: 11485 terms

Training set size: 68775 tokens

Preprocessing time: 36.617139417 sec

Finished training LDA model. Summary:

Training time: 231.321711026 sec

Training data average log likelihood: -51.39299144407459

4 topics:

TOPIC\_0;:0.06898491214163481

TOPIC\_0;rt;0.05230523781160823

TOPIC\_0;.:0.0443182452233433

accessory 4.330187058220508E-7

engine 4.3300706399910715E-7

founder 4.3299449089844737E-7

steakhouse 4.329928297333121E-7

brokerage 4.329800007177234E-7

coffee 4.3296174743087827E-7

menswear 4.3295446272151215E-7

drinks 4.329446783631709E-7

leasing 4.329246106191E-7

external 4.329217171459645E-7

144 4.32877349955435E-7

annex 4.328732236055471E-7

books 4.3286129582288156E-7

chocolate 4.3285034999841574E-7

catering 4.3283666341396286E-7

145 4.3282839027119965E-7

parlor 4.327825690362709E-7

presentation 4.3278136670538397E-7

450,000 4.3275898474339983E-7

Process finished with exit code 0

Feature vector output :

Spark\_KMeans\_FV | src | main | scala | featureVector | SparkTFIDF\_FVMain.scala

Project Packages Project Files SparkTFIDF\_FVMain.scala

Spark\_KMeans\_FV [spark\_kmeans\_fv] (E:\study\Study\KDM)

idea

data

20\_twitter\_wiki

hillary

politics

politics2

trump

Run SparkTFIDF\_FVMain

```
// Read the file into RDD[String]
val input = sc.wholeTextFiles("data/20_twitter_wiki/*", 500).map(line => {
  val location_array = line._1.split("/")
  val class_name = location_array(location_array.length - 2)
  var ff = line._2.replaceAll("[*a-zA-Z\\s:]", " ")
  ff = ff.replaceAll(":", " ")
  //Getting Lemmatized Form of the word using CoreNLP
  val lemma = CoreNLP.returnLemma(ff)
  (mapping.value.get(class_name).get.toDouble, lemma)
})
```

16/07/08 12:42:25 INFO BlockManagerMaster: Trying to register BlockManager

16/07/08 12:42:25 INFO BlockManagerMasterEndpoint: Registering block manager localhost:51092 with 2.4 GB RAM, BlockManagerId(driver, localhost, 51092)

16/07/08 12:42:25 INFO BlockManagerMaster: Registered BlockManager

16/07/08 12:42:27 WARN : Your hostname, DESKTOP-OCQAA2F resolves to a loopback/non-reachable address: fe80:0:0:0:9d8e:f363:315a:a773%wlan1, but we couldn't find any external IP

Reading POS tagger model from edu/stanford/nlp/models/pos-tagger/english-left3words/english-left3words-distsim.tagger ... done [0.6 sec].

[0.0,obama schedule he campaign appearance with hillaryclinton before comey speak action speak louder than word he kne BloodyDifficultWoman HillaryClinton http t co hih srn zz

[0.0,e don t know but Russia probably do Clinton rt gamma ray if this doesn t demand realdonaldtrump vote NOTHING will traitorousfbi realdonaldtrump do you answer the tweet or

[0.0,we don t know but Russia probably do Clinton rt creativewe jjohn pastormarkburn HillaryClinton FBI she walk free for ignorance to where we would be in prison for yr c rt

root

```
-- labels: double (nullable = false)
-- sentence: string (nullable = true)
-- words: array (nullable = true)
  |-- element: string (containsNull = true)
-- filteredWords: array (nullable = true)
  |-- element: string (containsNull = true)
-- ngrams: array (nullable = true)
  |-- element: string (containsNull = false)
```

()

Confusion matrix:

```
2.0 0.0 1.0 0.0 |
4.0 0.0 0.0 1.0 |
0.0 1.0 1.0 1.0 |
1.0 1.0 0.0 1.0 |
```

Accuracy: 0.2857142857142857

Process finished with exit code 0