



proshanta sarkar &lt;proshant.cse@gmail.com&gt;

---

## Code

2 messages

**Subhajit Datta** <subhajit.datta@gmail.com>  
To: proshanta sarkar <proshant.cse@gmail.com>

Fri, Dec 27, 2013 at 10:56 AM

Dear Proshanta,

Thank you for coming over yesterday. As discussed, please find attached the code with the explanation. Let me know if you are able to run the code.

Regards,

Subhajit

-----

Contents of Eclipse Project:

TicketTopicExtractor.scala : A scala function that does the topic extraction from a csv file where each row of the csv file is a document and it expects that the csv file has 3 columns, id, bin to which the ticket belongs and the description. It creates a, output file "document-topic-distributions.csv" with the topic distributions for each document with its id.

MainClass.java : An example showing how to call this scala function from any java class. You have send the parameter values for each set of documents. I think overall you can use 30 topics, alpha= 0.1, beta= 0.1, iterations= 1000 for all type of tickets.

Files Attached:

TopicVectorExtractionProject.zip : Eclipse Project. Just extract the contents and change the folder path in the ".classpath" file to your folder

tmt-0.3.3.jar : StanfordNLP package that I used to extract topics (I am unable to attach this due to its size, you can download it from <http://nlp.stanford.edu/software/tmt/tmt-0.3/> )

How to run the code:

Please download and use Eclipse Indigo for this

In Eclipse download and install the Scala IDE plugin from:

<http://download.scala-ide.org/releases-28/stable/site>

Import the Eclipse project from your folder to Eclipse

Please set the JRE in the Build path as JRE 6 (unfortunately this setup does not work with JRE 7) and also change the Java Compiler to JRE 6.0 compiler

Add the tmt-0.3.3.jar to the project buildpath.



**TopicVectorExtractionProject.zip**

9K

---

**proshanta sarkar** <proshant.cse@gmail.com>

Fri, Dec 27, 2013 at 12:06 PM

To: Subhajit Datta <subhajit.datta@gmail.com>

Sir,

I am briefing the procedure the way I understood if there is any mistake please let me know. below is the procedure

1. Install Eclipse and setting it up with JRE 6 and java compiler to JRE 6.0 compiler.
2. Unzipping TopicVectorExtractionProject.zip to some folder for an example  
/proshant/TopicVectorExtractionProject/
3. setting the CLASSPATH upto /proshant/TopicVectorExtractionProject/bin/
4. adding the tmt-0.3.3.jar file to the project.
5. lastly editing the MainClass.java by providing our actual file names and attribute values.

Now, I have three questions one is don't we need to import anything in the java file to use scala function TicketTopicExtractor.extractTopics?

second one is in our input CSV file the first id value will be the bug number, The description is the concatenated value of title and description of every bug, what is the bin attribute?

another question is those bug will be all bugs or those bugs which has owner.

[Quoted text hidden]