Information technology for detecting the hidden content of text messages

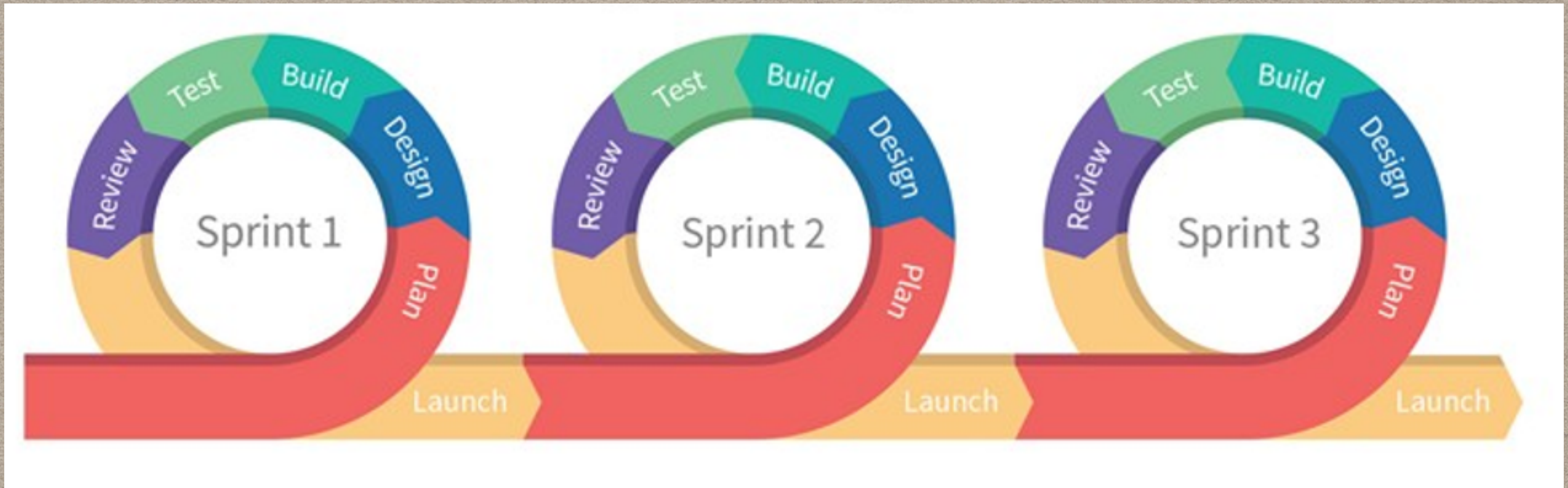
Fulfilled by students of the group 3ACIT-17M:
Oleksii O. Maksymov
Anastasiia T. Maksymova
Roman V. Slobodian

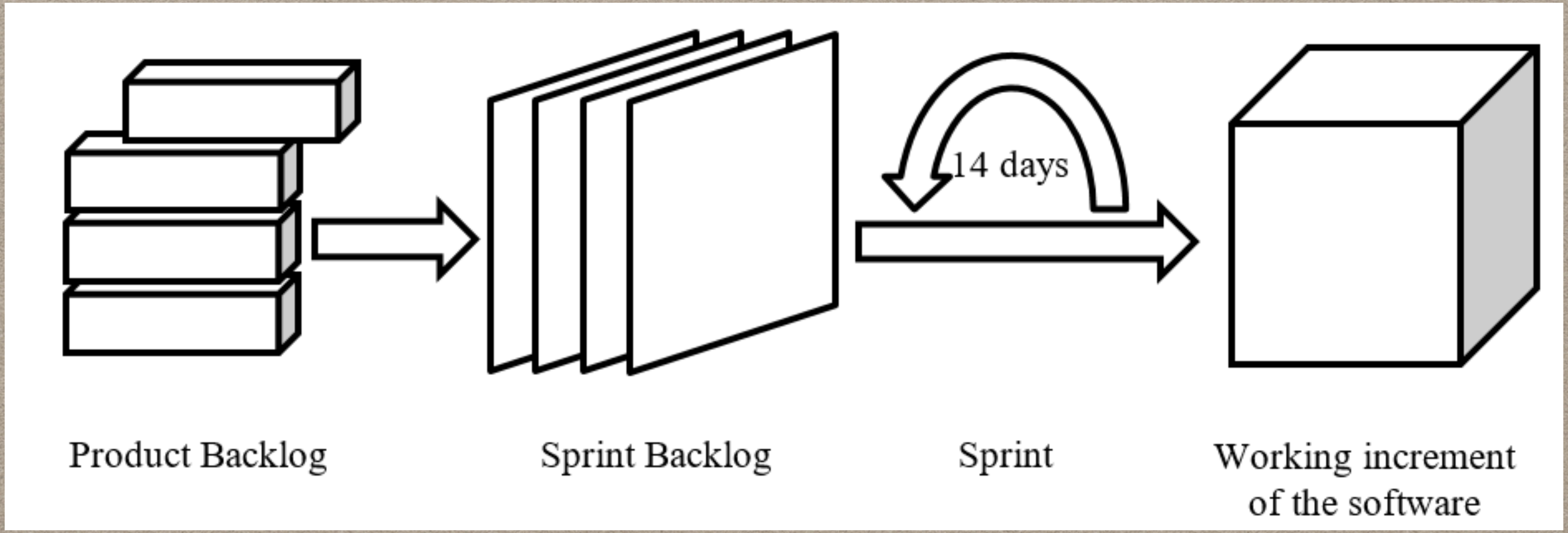
Supervisor:
Dr. of Sci., Professor Oleg V. Bisikalo

PART 1. DEVELOPMENT OF A PROJECT TO CREATE INFORMATION TECHNOLOGY FOR DETECTING HIDDEN CONTENT IN TEXT MESSAGES

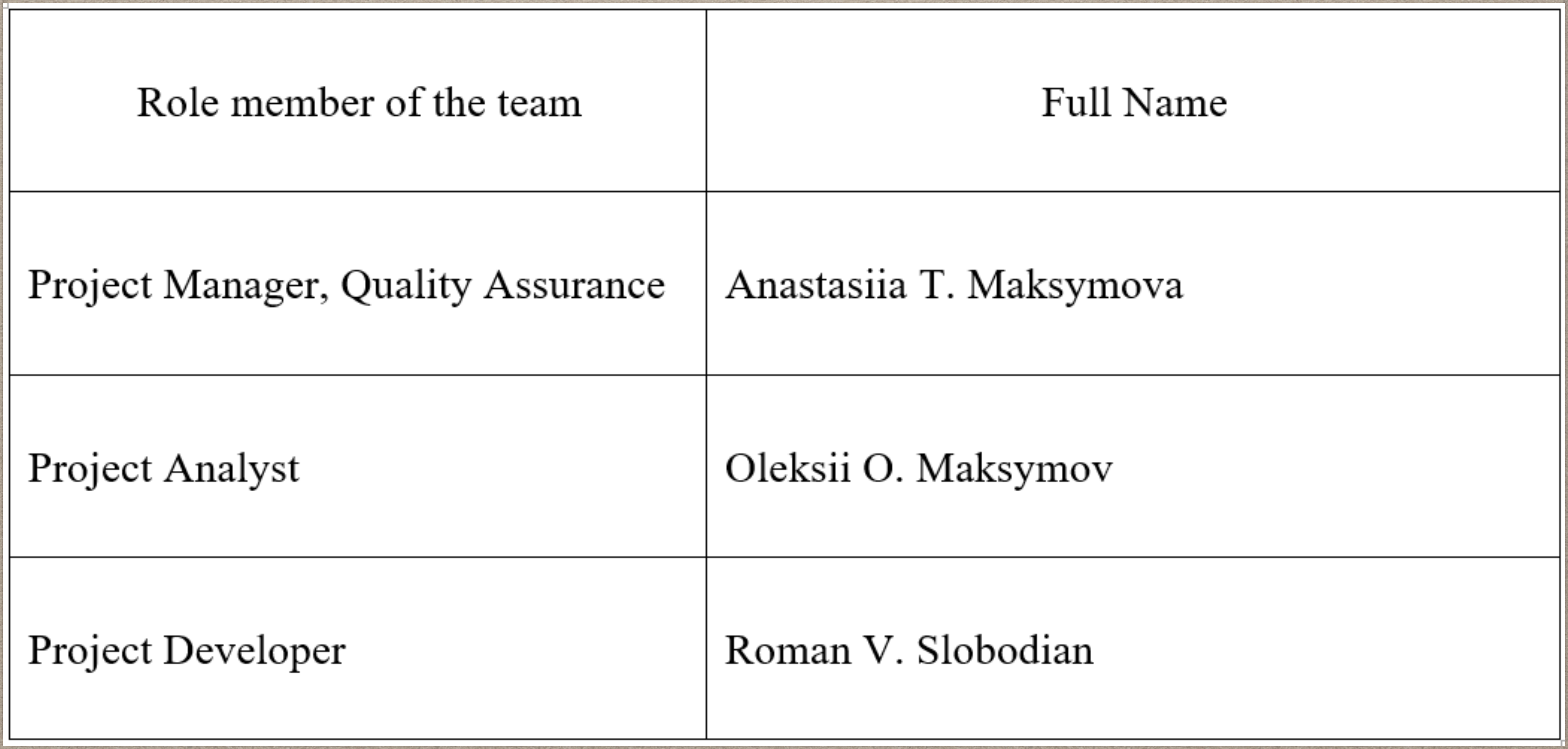
- The goal of the project management is to solve a specific task in a short time with minimum transaction costs (related to poor communication, changes, outflow of resources, etc.).

SCHEMATIC OF THE AGILE METHODOLOGY



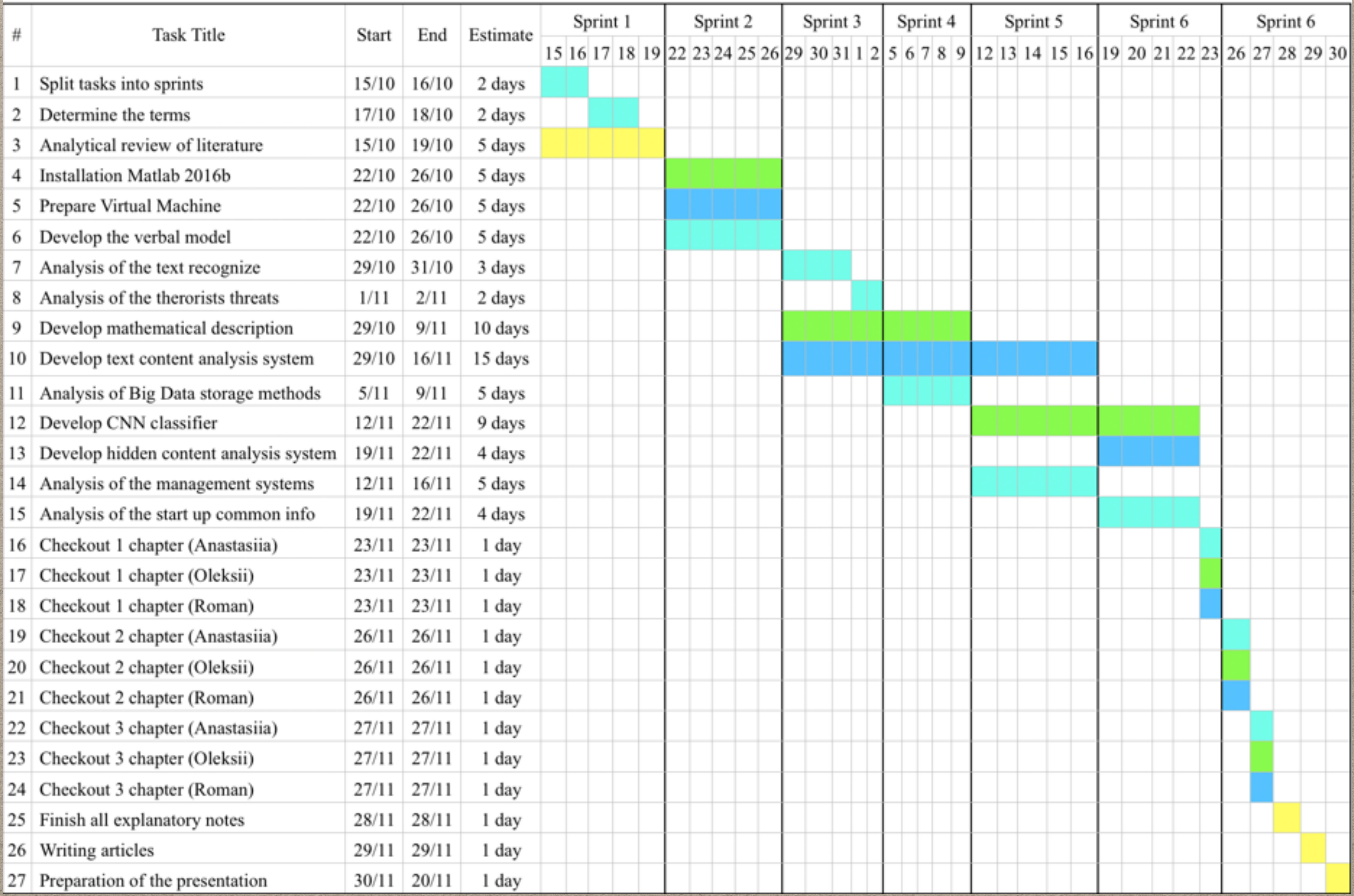


ROLES FOR EACH MEMBER OF THE TEAM



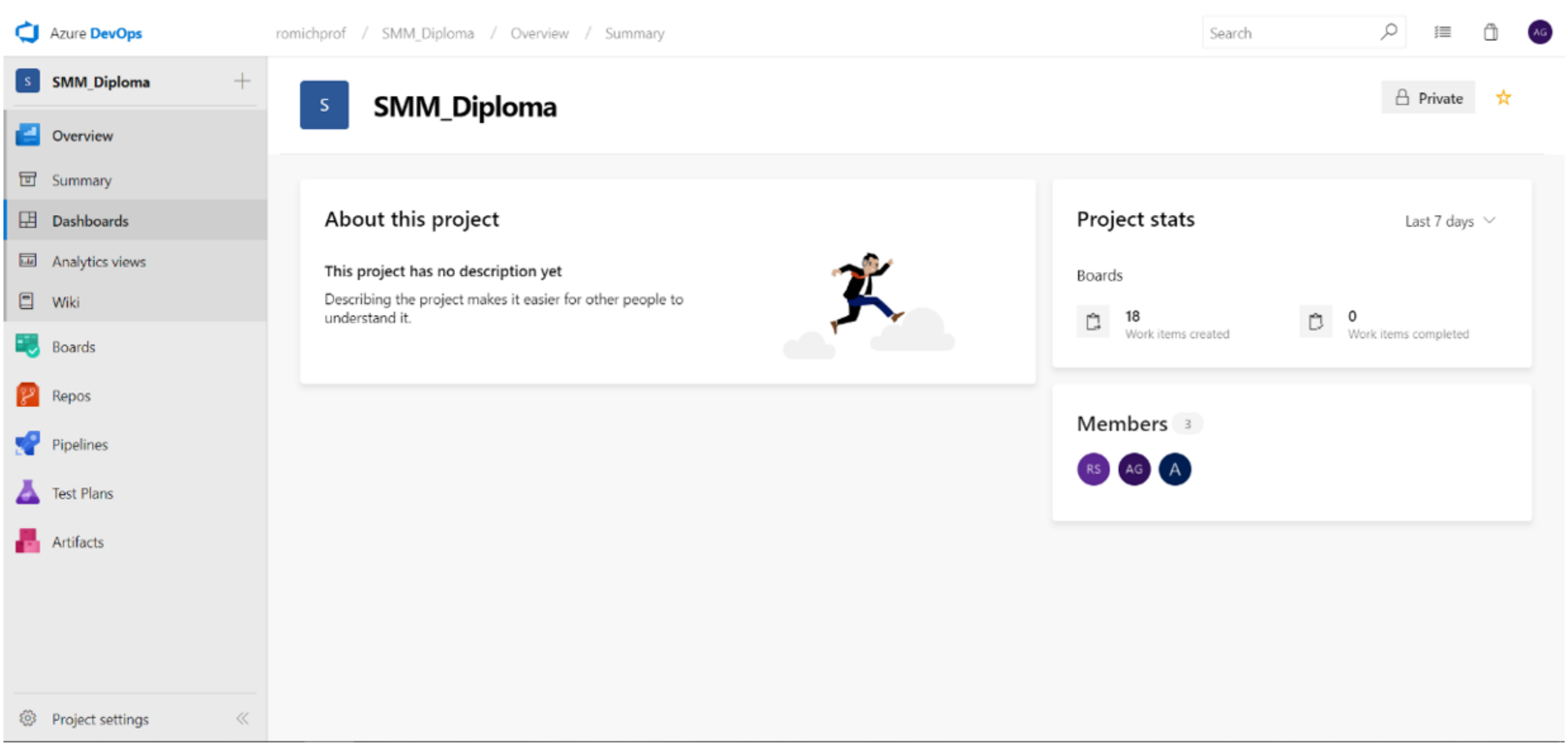
Role member of the team	Full Name
Project Manager, Quality Assurance	Anastasiia T. Maksymova
Project Analyst	Oleksii O. Maksymov
Project Developer	Roman V. Slobodian

PROJECT IMPLEMENTATION REPORT



#	Task Title	Start	End	Estimate	Sprint 1					Sprint 2					Sprint 3					Sprint 4					Sprint 5					Sprint 6					Sprint 6				
					15	16	17	18	19	22	23	24	25	26	29	30	31	1	2	5	6	7	8	9	12	13	14	15	16	19	20	21	22	23	26	27	28	29	30
1	Split tasks into sprints	15/10	16/10	2 days	▨	▨																																	
2	Determine the terms	17/10	18/10	2 days			▨	▨																															
3	Analytical review of literature	15/10	19/10	5 days	🟨	🟨	🟨	🟨	🟨																														
4	Installation Matlab 2016b	22/10	26/10	5 days						🟩	🟩	🟩	🟩	🟩																									
5	Prepare Virtual Machine	22/10	26/10	5 days						🟦	🟦	🟦	🟦	🟦																									
6	Develop the verbal model	22/10	26/10	5 days						▨	▨	▨	▨	▨																									
7	Analysis of the text recognize	29/10	31/10	3 days											▨	▨	▨																						
8	Analysis of the therorists threats	1/11	2/11	2 days														▨	▨																				
9	Develop mathematical description	29/10	9/11	10 days											🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩															
10	Develop text content analysis system	29/10	16/11	15 days											🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦										
11	Analysis of Big Data storage methods	5/11	9/11	5 days																▨	▨	▨	▨	▨															
12	Develop CNN classifier	12/11	22/11	9 days																					🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩						
13	Develop hidden content analysis system	19/11	22/11	4 days																										🟦	🟦	🟦	🟦						
14	Analysis of the management systems	12/11	16/11	5 days																					▨	▨	▨	▨	▨										
15	Analysis of the start up common info	19/11	22/11	4 days																										▨	▨	▨	▨						
16	Checkout 1 chapter (Anastasiia)	23/11	23/11	1 day																														▨					
17	Checkout 1 chapter (Oleksii)	23/11	23/11	1 day																														🟩					
18	Checkout 1 chapter (Roman)	23/11	23/11	1 day																														🟦					
19	Checkout 2 chapter (Anastasiia)	26/11	26/11	1 day																															▨				
20	Checkout 2 chapter (Oleksii)	26/11	26/11	1 day																															🟩				
21	Checkout 2 chapter (Roman)	26/11	26/11	1 day																															🟦				
22	Checkout 3 chapter (Anastasiia)	27/11	27/11	1 day																																▨			
23	Checkout 3 chapter (Oleksii)	27/11	27/11	1 day																																🟩			
24	Checkout 3 chapter (Roman)	27/11	27/11	1 day																																🟦			
25	Finish all explanatory notes	28/11	28/11	1 day																																	🟨		
26	Writing articles	29/11	29/11	1 day																																		🟨	
27	Preparation of the presentation	30/11	20/11	1 day																																			🟨

OVERVIEW AZURE USER INTERFACE



SMM_Diploma

About this project

This project has no description yet

Describing the project makes it easier for other people to understand it.

Project stats

Boards

18 Work items created

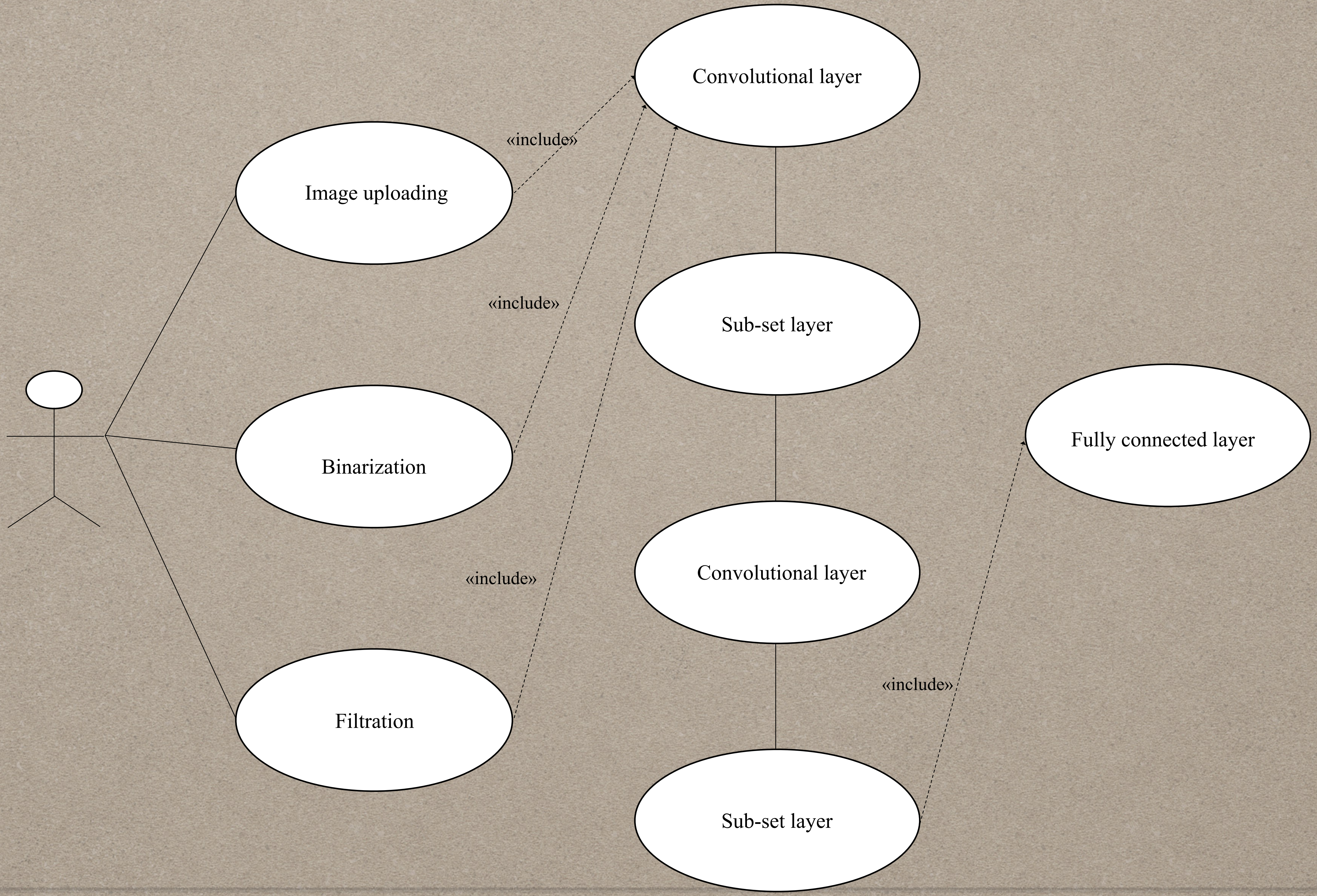
0 Work items completed

Members 3

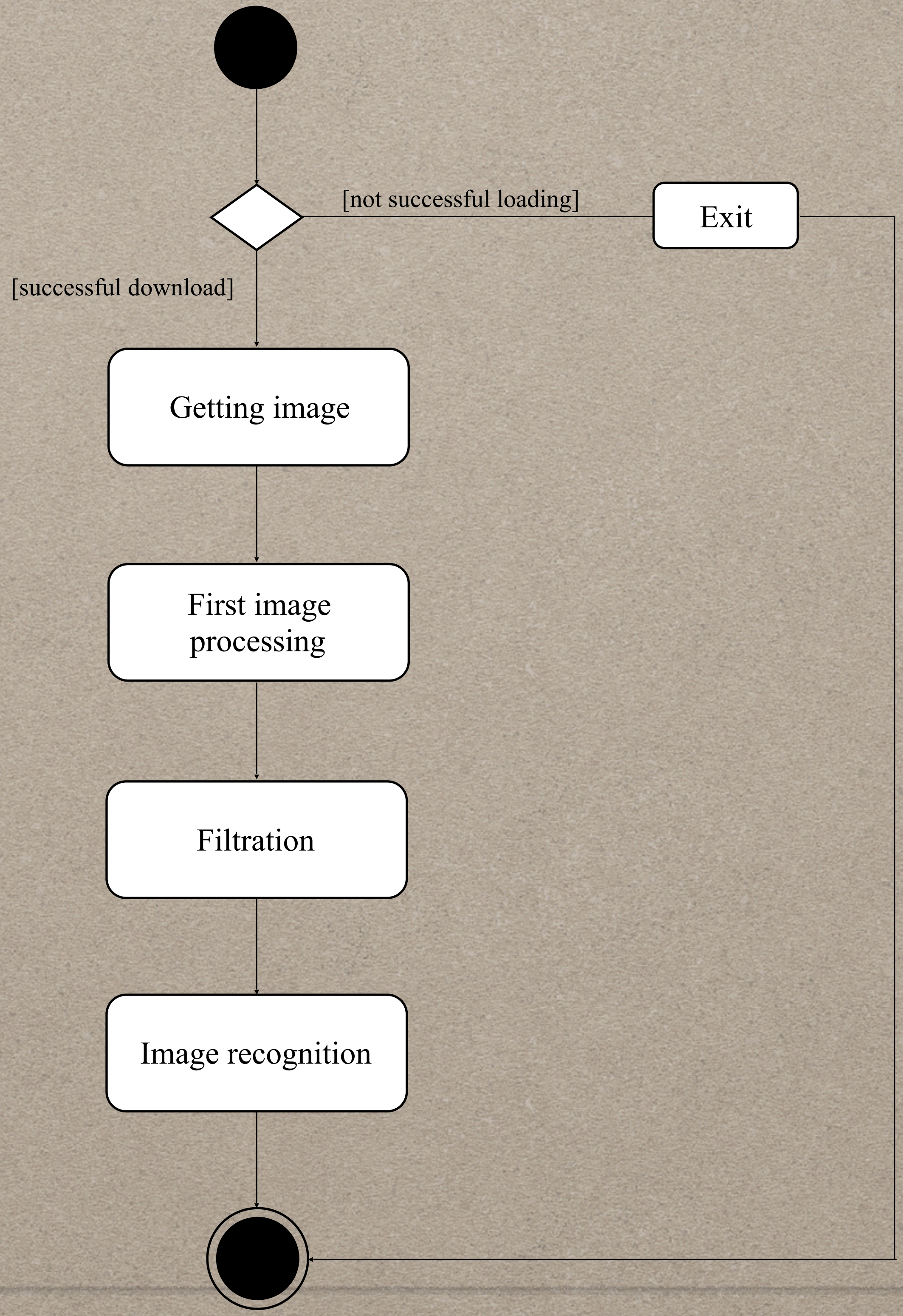
PART 2. DEVELOPMENT OF THE ANALYTICAL AND MATHEMATICAL SUPPORT FOR INFORMATION TECHNOLOGY FOR DETECTING HIDDEN CONTENT IN THE TEXT MESSAGES

- The purpose of master's qualification work is :
- 1. Create analytical and mathematical support for information technology.
- 2. to improve the quality of recognition of hidden content in text messages through the use of methods of working with Big Data, as well as neural networks and machine learning.

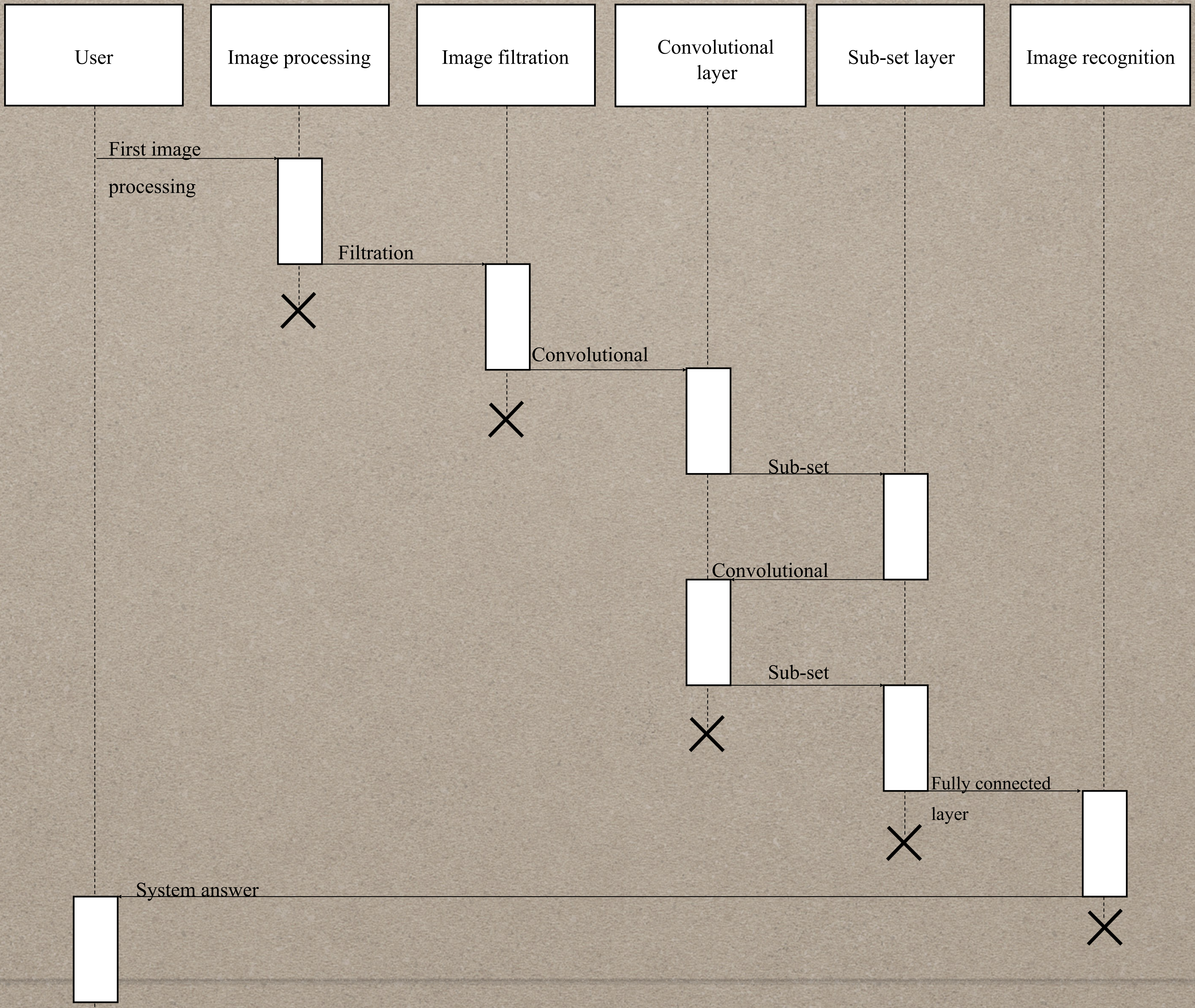
USE CASE DIAGRAM



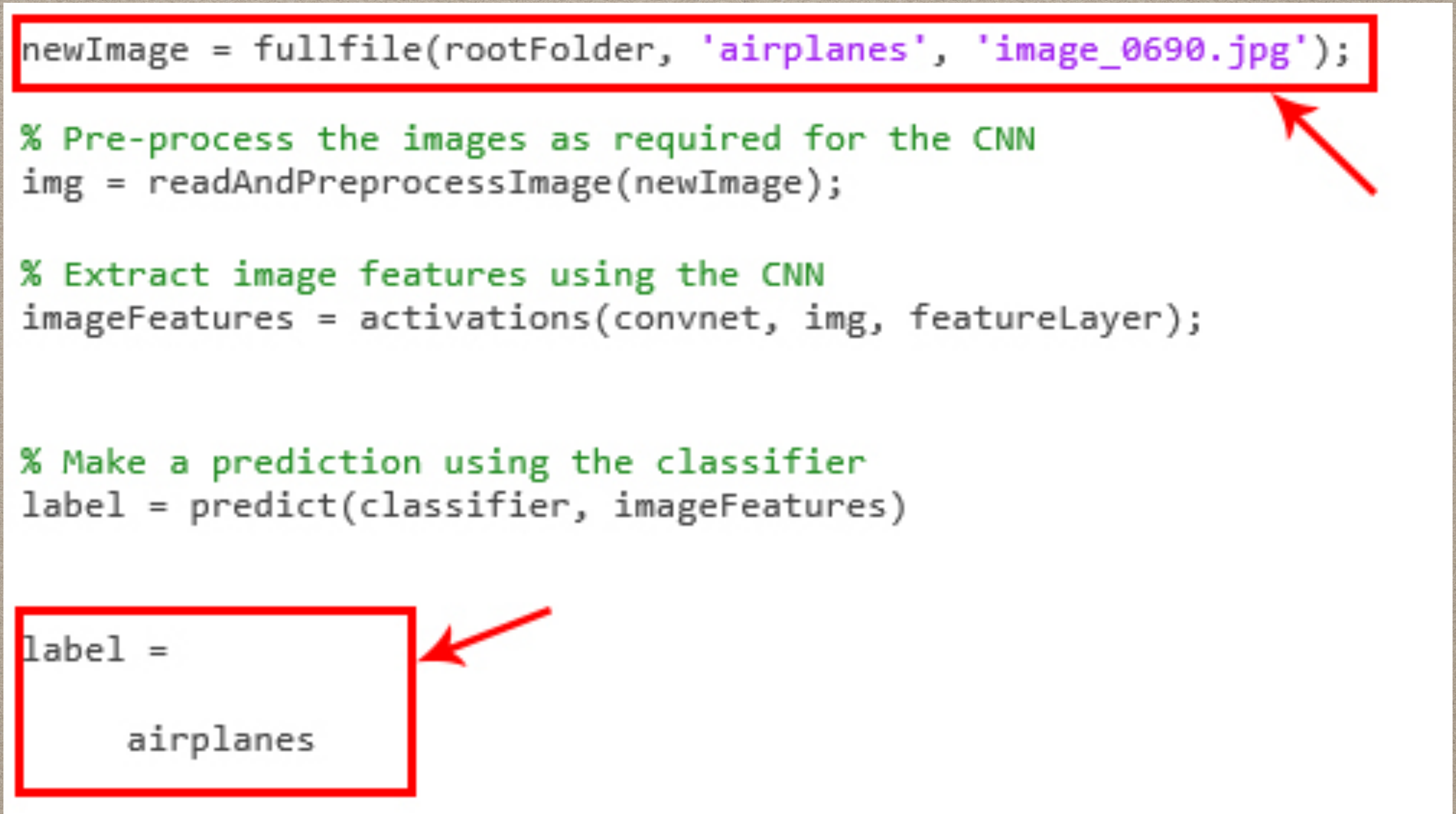
PROCESS ALGORITHM



SEQUENCE DIAGRAM

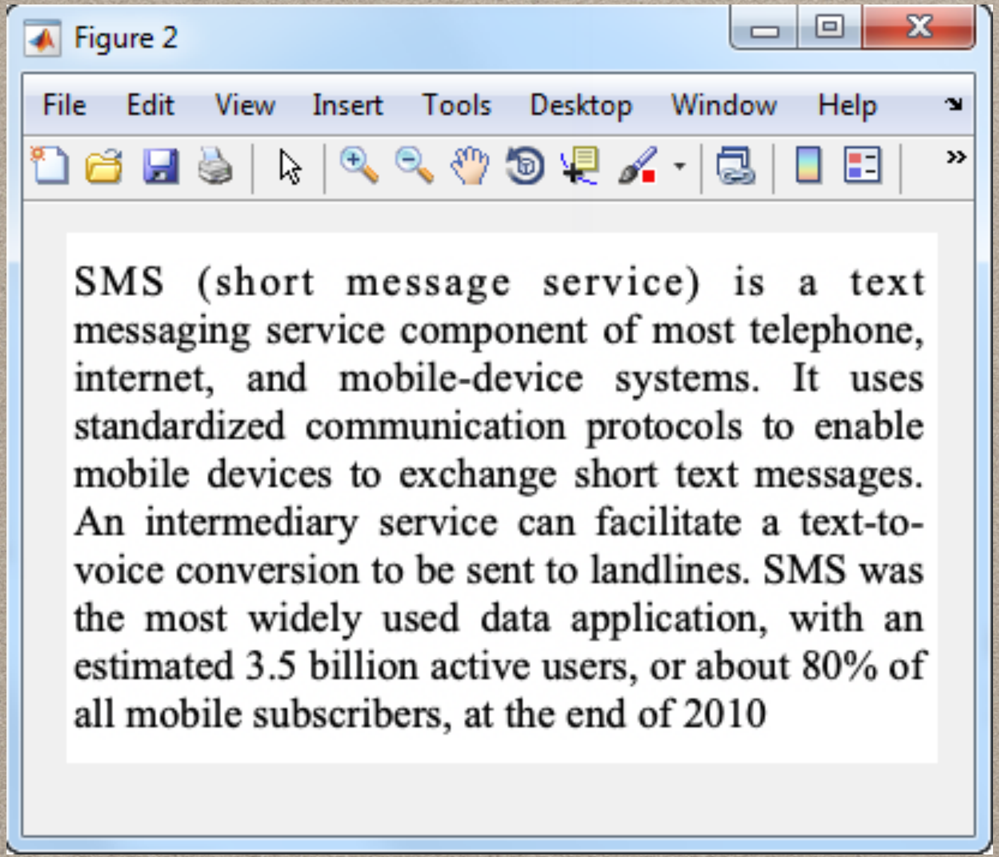


WORK OF THE MAIN FUNCTION

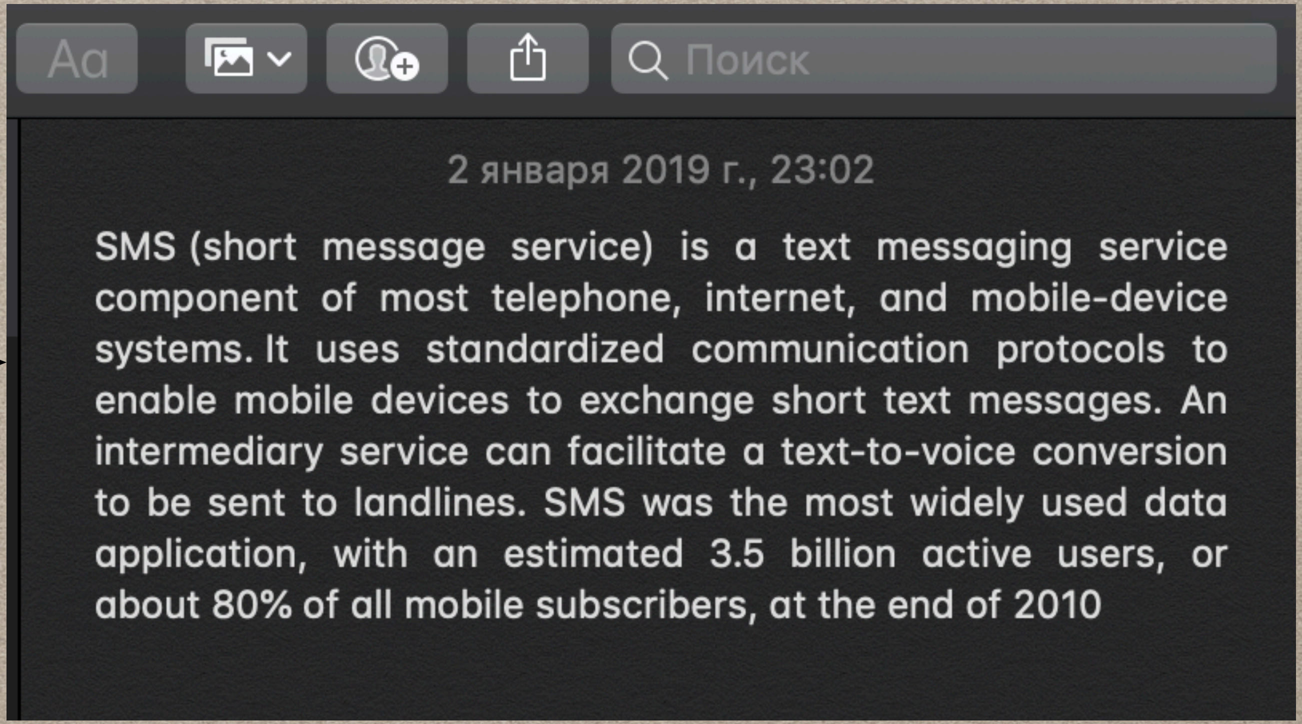


```
newImage = fullfile(rootFolder, 'airplanes', 'image_0690.jpg');

% Pre-process the images as required for the CNN
img = readAndPreprocessImage(newImage);

% Extract image features using the CNN
imageFeatures = activations(convnet, img, featureLayer);



% Make a prediction using the classifier
label = predict(classifier, imageFeatures)


label =

    airplanes
```

RECOGNITION RESULTS



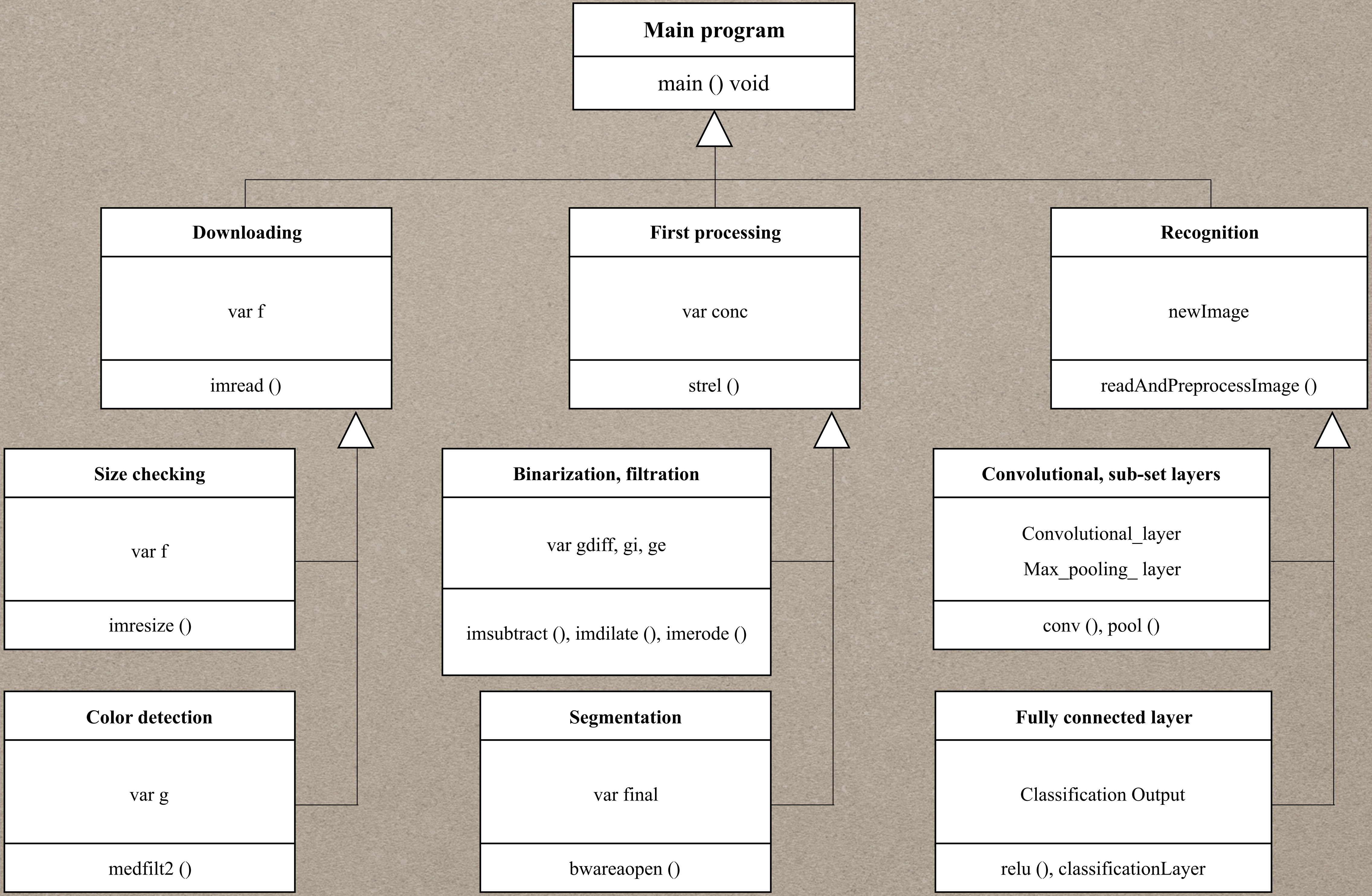
SMS (short message service) is a text messaging service component of most telephone, internet, and mobile-device systems. It uses standardized communication protocols to enable mobile devices to exchange short text messages. An intermediary service can facilitate a text-to-voice conversion to be sent to landlines. SMS was the most widely used data application, with an estimated 3.5 billion active users, or about 80% of all mobile subscribers, at the end of 2010



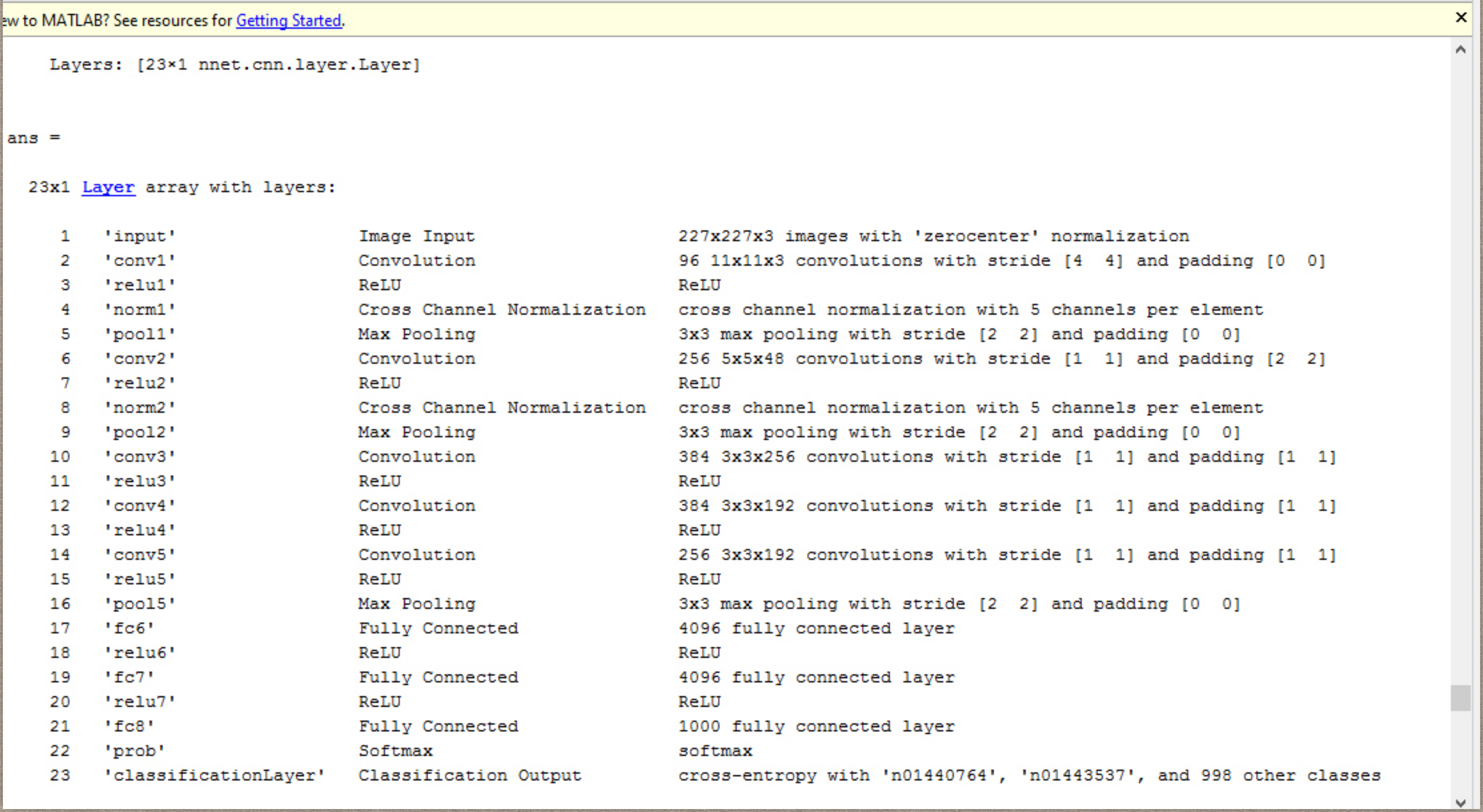
2 января 2019 г., 23:02

SMS (short message service) is a text messaging service component of most telephone, internet, and mobile-device systems. It uses standardized communication protocols to enable mobile devices to exchange short text messages. An intermediary service can facilitate a text-to-voice conversion to be sent to landlines. SMS was the most widely used data application, with an estimated 3.5 billion active users, or about 80% of all mobile subscribers, at the end of 2010

CLASS DIAGRAM

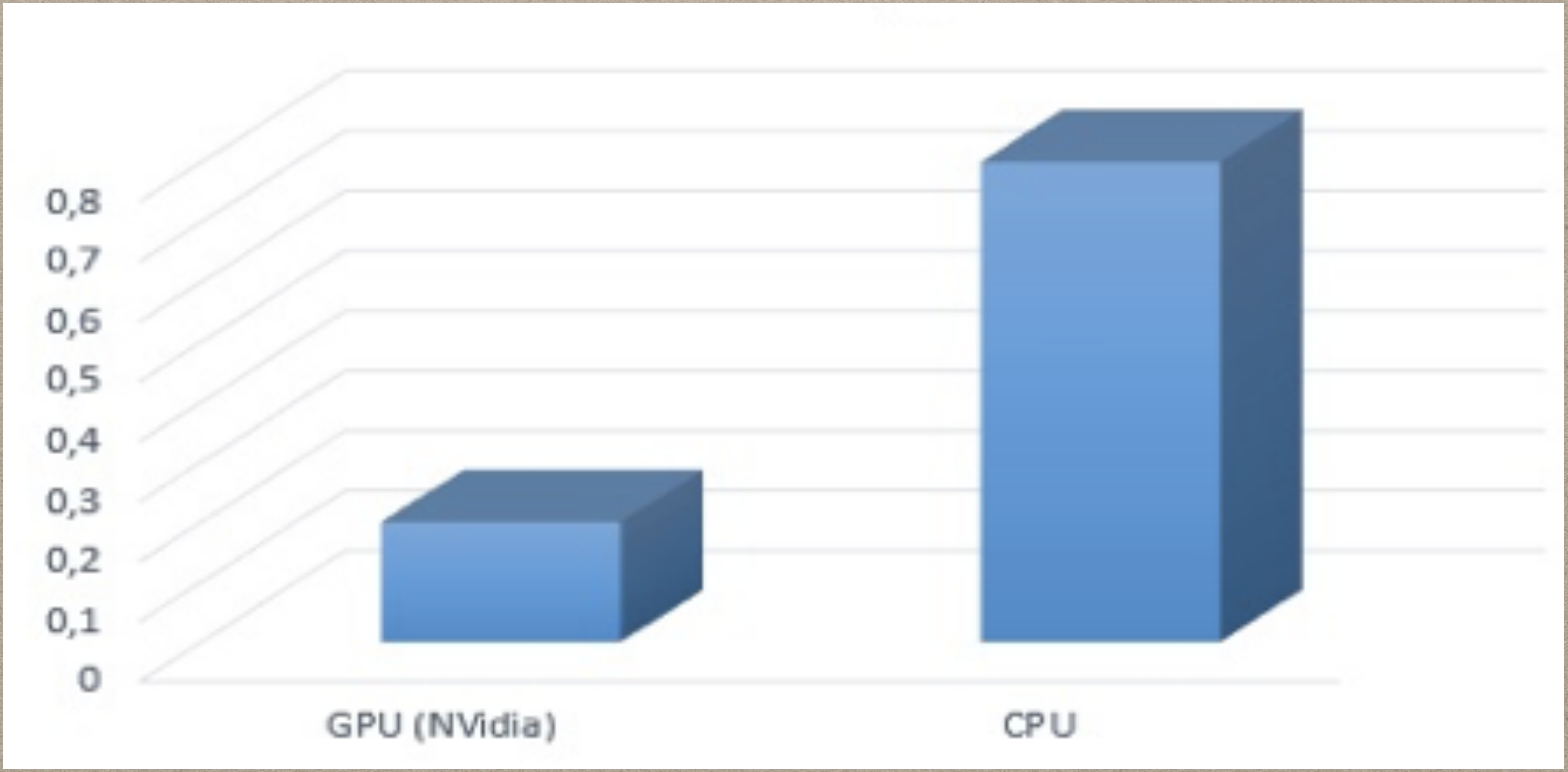


DEVELOPED CLASSIFIER



```
    Layers: [23×1 nnet.cnn.layer.Layer]


ans =

  23x1 Layer array with layers:

     1   'input'                 Image Input                   227x227x3 images with 'zerocenter' normalization
     2   'conv1'                 Convolution                   96 11x11x3 convolutions with stride [4  4] and padding [0  0]
     3   'relu1'                 ReLU                          ReLU
     4   'norm1'                 Cross Channel Normalization   cross channel normalization with 5 channels per element
     5   'pool1'                 Max Pooling                   3x3 max pooling with stride [2  2] and padding [0  0]
     6   'conv2'                 Convolution                   256 5x5x48 convolutions with stride [1  1] and padding [2  2]
     7   'relu2'                 ReLU                          ReLU
     8   'norm2'                 Cross Channel Normalization   cross channel normalization with 5 channels per element
     9   'pool2'                 Max Pooling                   3x3 max pooling with stride [2  2] and padding [0  0]
    10   'conv3'                 Convolution                   384 3x3x256 convolutions with stride [1  1] and padding [1  1]
    11   'relu3'                 ReLU                          ReLU
    12   'conv4'                 Convolution                   384 3x3x192 convolutions with stride [1  1] and padding [1  1]
    13   'relu4'                 ReLU                          ReLU
    14   'conv5'                 Convolution                   256 3x3x192 convolutions with stride [1  1] and padding [1  1]
    15   'relu5'                 ReLU                          ReLU
    16   'pool5'                 Max Pooling                   3x3 max pooling with stride [2  2] and padding [0  0]
    17   'fc6'                   Fully Connected               4096 fully connected layer
    18   'relu6'                 ReLU                          ReLU
    19   'fc7'                   Fully Connected               4096 fully connected layer
    20   'relu7'                 ReLU                          ReLU
    21   'fc8'                   Fully Connected               1000 fully connected layer
    22   'prob'                  Softmax                       softmax
    23   'classificationLayer'   Classification Output         cross-entropy with 'n01440764', 'n01443537', and 998 other classes
```

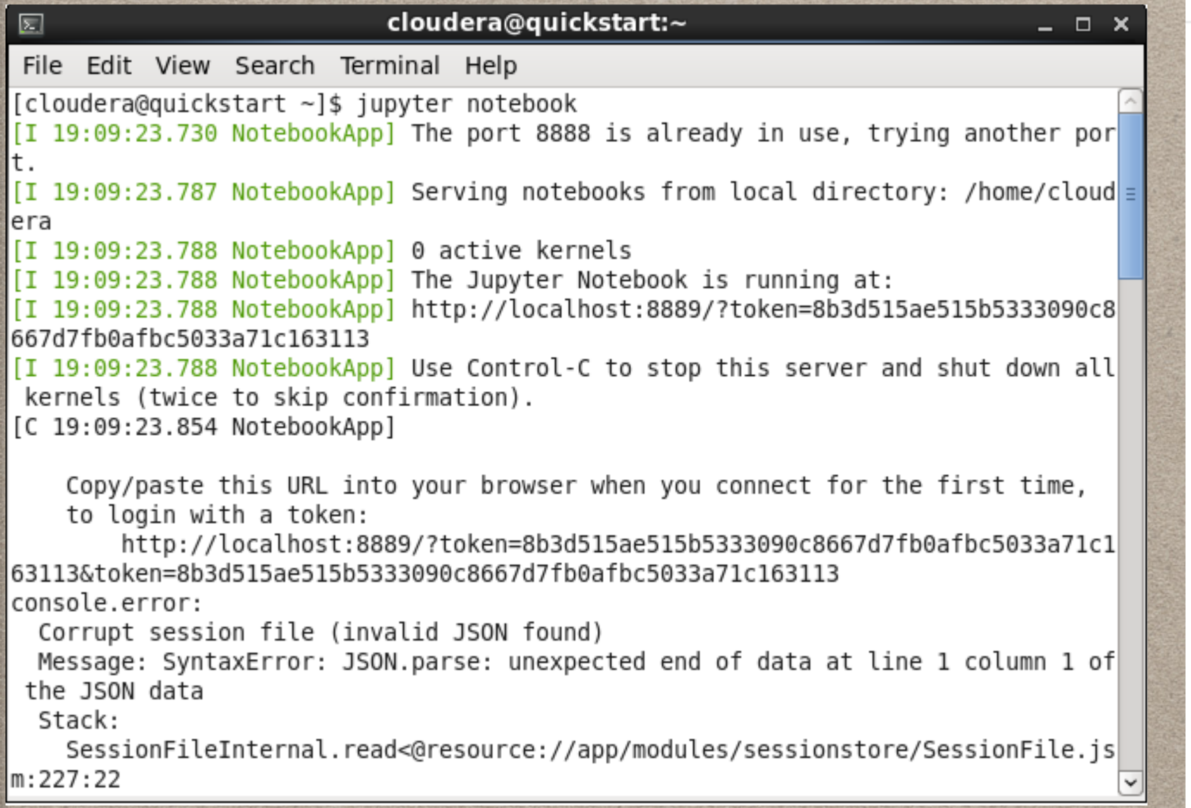
DEPENDENCIES OF THE CHOSEN PLATFORM



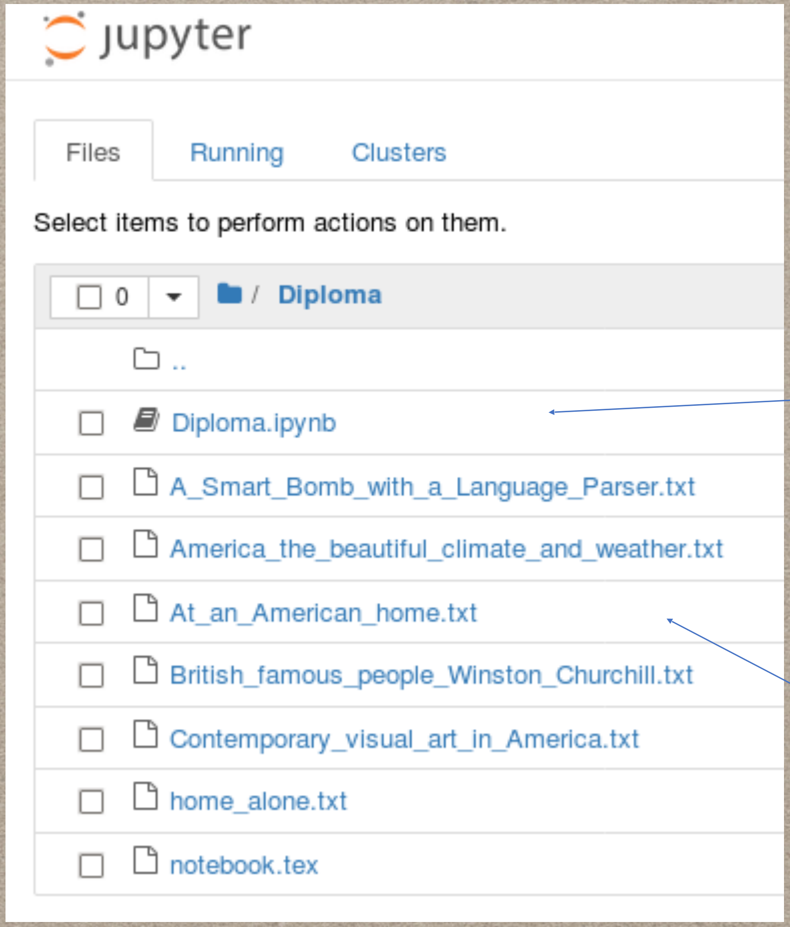
PART 3. DEVELOPMENT OF SOFTWARE AND TECHNOLOGY PROVISION OF INFORMATION TECHNOLOGY FOR DETECTING HIDDEN CONTENT IN TEXT MESSAGES

- The goal of this work is to investigate and apply on practice methods and tools of Big Data’s processing and analyzing on the information that generates in real time during the everyday communication between people to detect hidden content in it.

ENVIRONMENT



```
[cloudera@quickstart ~]$ jupyter notebook
[I 19:09:23.730 NotebookApp] The port 8888 is already in use, trying another por
t.
[I 19:09:23.787 NotebookApp] Serving notebooks from local directory: /home/cloud
era
[I 19:09:23.788 NotebookApp] 0 active kernels
[I 19:09:23.788 NotebookApp] The Jupyter Notebook is running at:
[I 19:09:23.788 NotebookApp] http://localhost:8889/?token=8b3d515ae515b5333090c8
667d7fb0afbc5033a71c163113
[I 19:09:23.788 NotebookApp] Use Control-C to stop this server and shut down all
 kernels (twice to skip confirmation).
[C 19:09:23.854 NotebookApp]

    Copy/paste this URL into your browser when you connect for the first time,
    to login with a token:
        http://localhost:8889/?token=8b3d515ae515b5333090c8667d7fb0afbc5033a71c1
63113&token=8b3d515ae515b5333090c8667d7fb0afbc5033a71c163113
console.error:
  Corrupt session file (invalid JSON found)
  Message: SyntaxError: JSON.parse: unexpected end of data at line 1 column 1 of
 the JSON data
  Stack:
    SessionFileInternal.read<@resource://app/modules/sessionstore/SessionFile.js
m:227:22
```



Select items to perform actions on them.

□ 0 ▾ / Diploma

..

□ Diploma.ipynb

□ A_Smart_Bomb_with_a_Language_Parser.txt

□ America_the_beautiful_climate_and_weather.txt

□ At_an_American_home.txt

□ British_famous_people_Winston_Churchill.txt

□ Contemporary_visual_art_in_America.txt

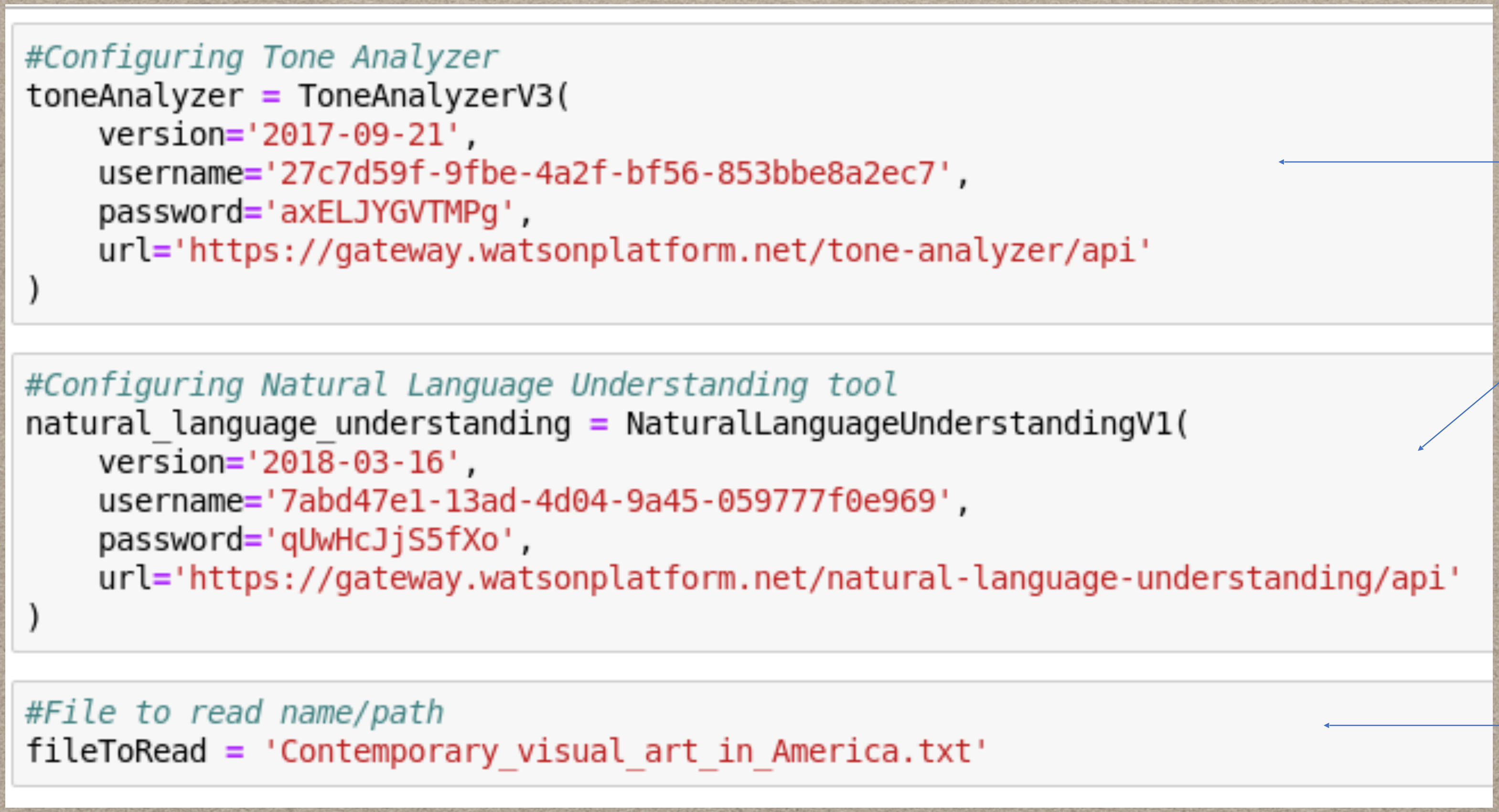
□ home_alone.txt

□ notebook.tex

Developed script

Test data

SCRIPT OVERVIEW



```
#Configuring Tone Analyzer
toneAnalyzer = ToneAnalyzerV3(
    version='2017-09-21',
    username='27c7d59f-9fbe-4a2f-bf56-853bbe8a2ec7',
    password='axELJYGVTMPg',
    url='https://gateway.watsonplatform.net/tone-analyzer/api'
)
```

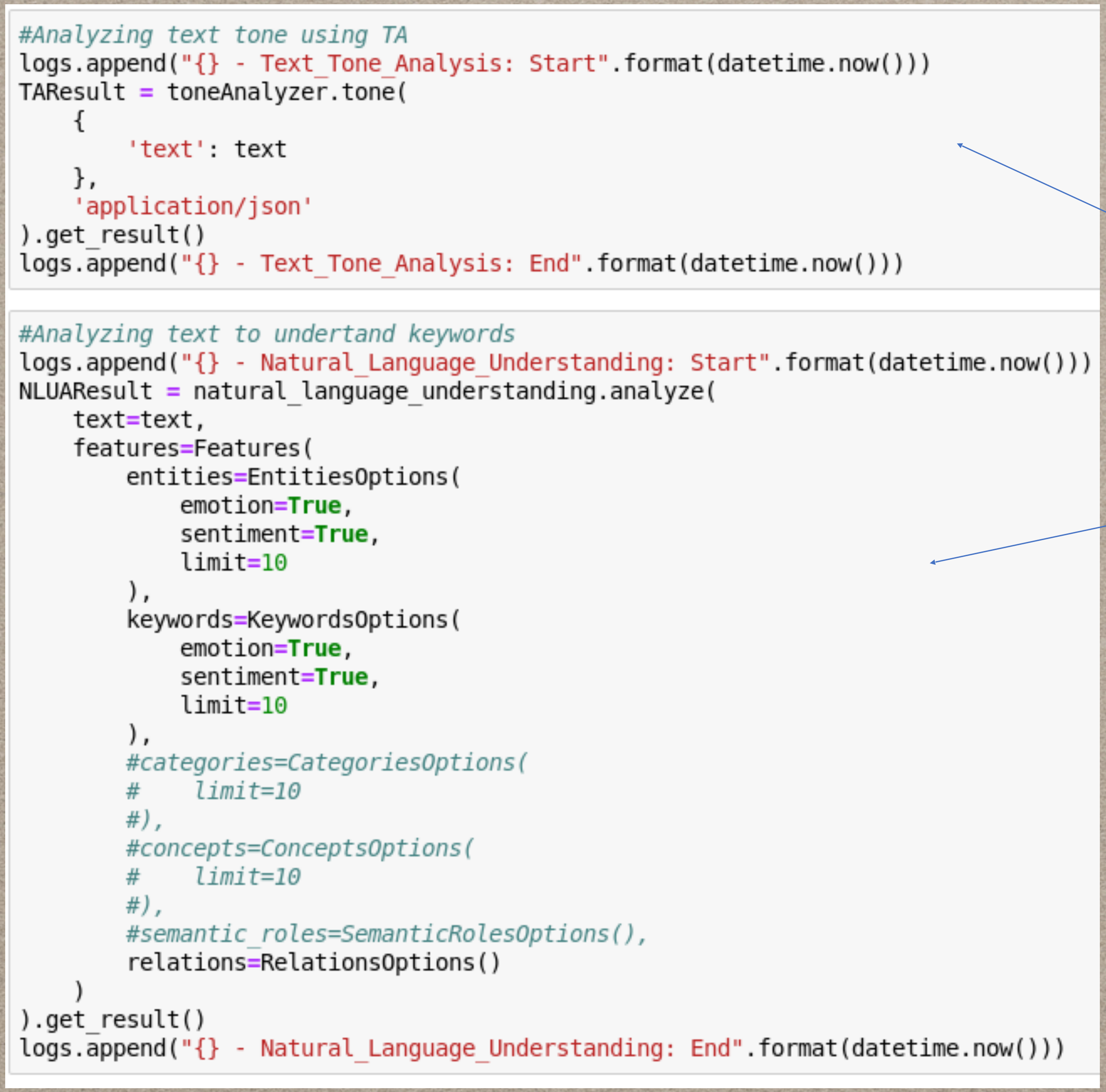
```
#Configuring Natural Language Understanding tool
natural_language_understanding = NaturalLanguageUnderstandingV1(
    version='2018-03-16',
    username='7abd47e1-13ad-4d04-9a45-059777f0e969',
    password='qUwHcJjS5fXo',
    url='https://gateway.watsonplatform.net/natural-language-understanding/api'
)
```

```
#File to read name/path
fileToRead = 'Contemporary_visual_art_in_America.txt'
```

IBM Watson API
connections configuration

File to read path

SCRIPT OVERVIEW

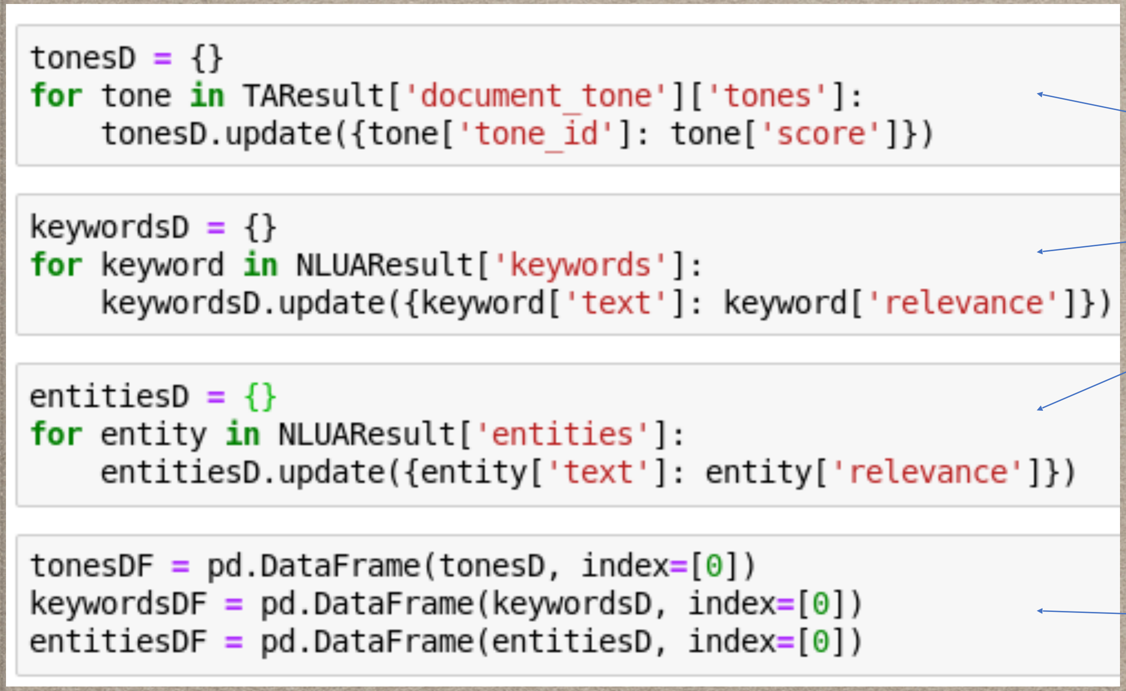


```
#Analyzing text tone using TA
logs.append("{} - Text_Tone_Analysis: Start".format(datetime.now()))
TAResult = toneAnalyzer.tone(
    {
        'text': text
    },
    'application/json'
).get_result()
logs.append("{} - Text_Tone_Analysis: End".format(datetime.now()))
```

```
#Analyzing text to undertand keywords
logs.append("{} - Natural_Language_Understanding: Start".format(datetime.now()))
NLUAResult = natural_language_understanding.analyze(
    text=text,
    features=Features(
        entities=EntitiesOptions(
            emotion=True,
            sentiment=True,
            limit=10
        ),
        keywords=KeywordsOptions(
            emotion=True,
            sentiment=True,
            limit=10
        ),
        #categories=CategoriesOptions(
        #    limit=10
        #),
        #concepts=ConceptsOptions(
        #    limit=10
        #),
        #semantic_roles=SemanticRolesOptions(),
        relations=RelationsOptions()
    )
).get_result()
logs.append("{} - Natural_Language_Understanding: End".format(datetime.now()))
```

API Calls to IMB Watsons for Text tone analysis and Natural Language Processing (Understanding)

SCRIPT OVERVIEW



```
tonesD = {}
for tone in TAResult['document_tone']['tones']:
    tonesD.update({tone['tone_id']: tone['score']})
```

```
keywordsD = {}
for keyword in NLUAResult['keywords']:
    keywordsD.update({keyword['text']: keyword['relevance']})
```

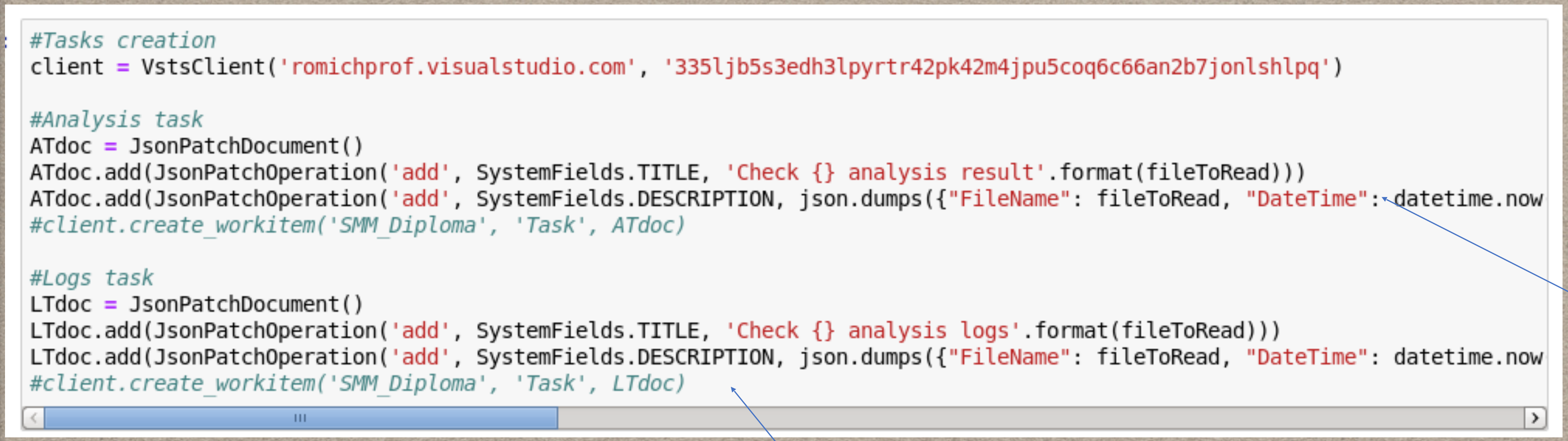
```
entitiesD = {}
for entity in NLUAResult['entities']:
    entitiesD.update({entity['text']: entity['relevance']})
```

```
tonesDF = pd.DataFrame(tonesD, index=[0])
keywordsDF = pd.DataFrame(keywordsD, index=[0])
entitiesDF = pd.DataFrame(entitiesD, index=[0])
```

Text analysis results parsing

Converting parsed results to Pandas DataFrames
(to simplify future manipulations)

SCRIPT OVERVIEW

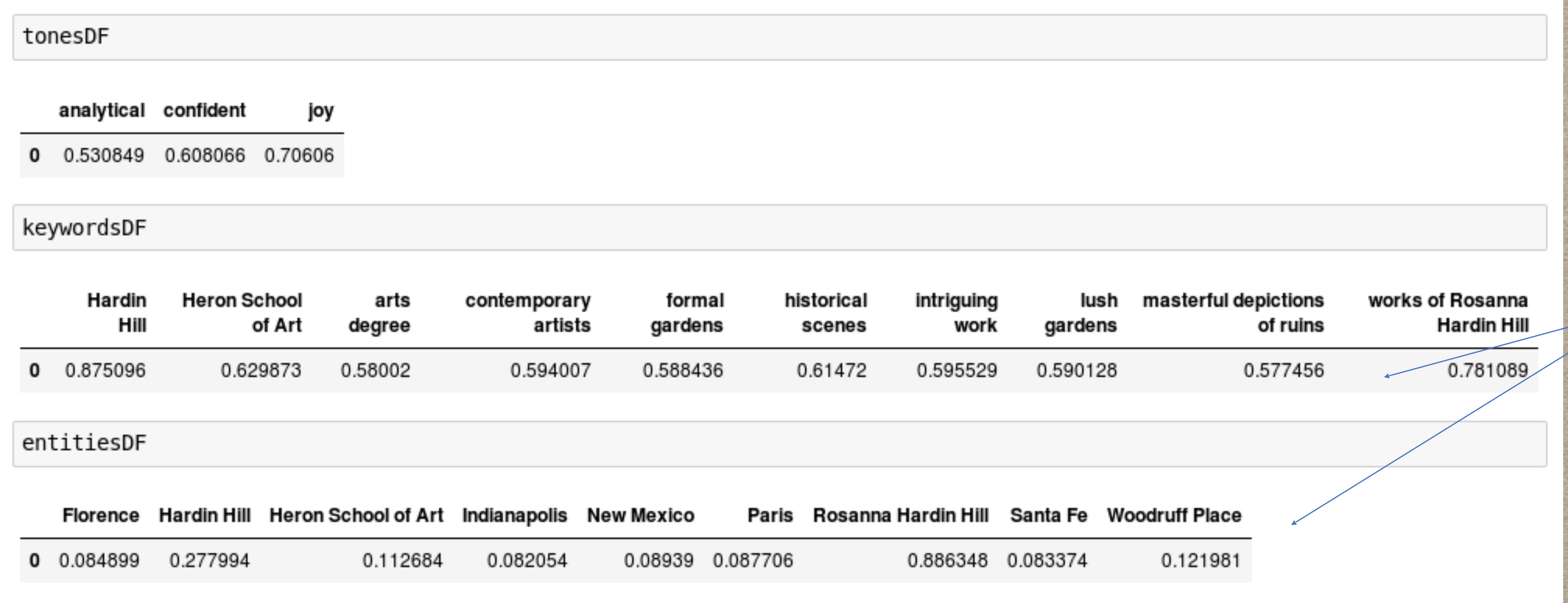


```
#Tasks creation
client = VstsClient('romichprof.visualstudio.com', '335ljb5s3edh3lpyrtr42pk42m4jpu5coq6c66an2b7jonlshlpq')

#Analysis task
ATdoc = JsonPatchDocument()
ATdoc.add(JsonPatchOperation('add', SystemFields.TITLE, 'Check {} analysis result'.format(fileToRead)))
ATdoc.add(JsonPatchOperation('add', SystemFields.DESCRIPTION, json.dumps({"FileName": fileToRead, "DateTime": datetime.now
#client.create_workitem('SMM_Diploma', 'Task', ATdoc)

#Logs task
LTdoc = JsonPatchDocument()
LTdoc.add(JsonPatchOperation('add', SystemFields.TITLE, 'Check {} analysis logs'.format(fileToRead)))
LTdoc.add(JsonPatchOperation('add', SystemFields.DESCRIPTION, json.dumps({"FileName": fileToRead, "DateTime": datetime.now
#client.create_workitem('SMM_Diploma', 'Task', LTdoc)
```

Task with text analysis results creation

Task with text analysis processing logs creation

RESULTS OVERVIEW



tonesDF

	analytical	confident	joy
0	0.530849	0.608066	0.70606

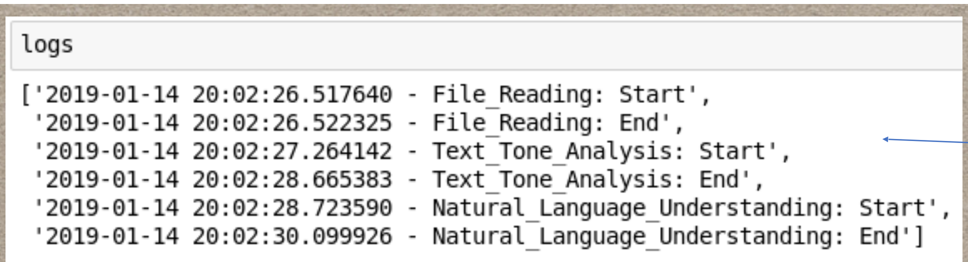
keywordsDF

	Hardin Hill	Heron School of Art	arts degree	contemporary artists	formal gardens	historical scenes	intriguing work	lush gardens	masterful depictions of ruins	works of Rosanna Hardin Hill
0	0.875096	0.629873	0.58002	0.594007	0.588436	0.61472	0.595529	0.590128	0.577456	0.781089

entitiesDF

	Florence	Hardin Hill	Heron School of Art	Indianapolis	New Mexico	Paris	Rosanna Hardin Hill	Santa Fe	Woodruff Place
0	0.084899	0.277994	0.112684	0.082054	0.08939	0.087706	0.886348	0.083374	0.121981

Parsed and converted Natural Language Processing (Understanding) results



logs

```
['2019-01-14 20:02:26.517640 - File_Reading: Start',
 '2019-01-14 20:02:26.522325 - File_Reading: End',
 '2019-01-14 20:02:27.264142 - Text_Tone_Analysis: Start',
 '2019-01-14 20:02:28.665383 - Text_Tone_Analysis: End',
 '2019-01-14 20:02:28.723590 - Natural_Language_Understanding: Start',
 '2019-01-14 20:02:30.099926 - Natural_Language_Understanding: End']
```

Execution logs

THAT'S ALL.
THANKS FOR ATTENTION!