Application for Automated Collection of Test Files  
 for CSS Class via HTTP and for Local Plagiarism Check:   
Testrek

**by**  
**Akshat Tandon**

Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Bachelor of Mechanical Engineering (Information and Automaton Technology)

in the  
Czech Technical University  
Faculty of Mechanical Engineering



© [Akshat Tandon]  
Czech Technical University  
Summer 2017

Copyright in this thesis rests with the author. Please ensure that any reproduction   
or re-use is done in accordance with the relevant national copyright legislation.

Approval

|  |  |  |
| --- | --- | --- |
| **Name:** | **[Firstname Lastname]** | |
| **Degree:** | **Bachelor of Mechanical Engineering (Information and Automaton Technology)** | |
| **Title:** | **Application for Automated Collection of Test Files for CSS Class via HTTP and for Local Plagiarism Check: Testrek** | |
| **Examining Committee:** | **Chair:** | **[Firstname Lastname]** [Academic Role] |
|  | **[Firstname Lastname]** Supervisor [Academic Role] | |
|  | **[Firstname Lastname]** Internal Examiner [Academic Role] [Department – if different from the home department] | |
|  | **[Firstname Lastname]** External Examiner [Academic Role or Professional Role] [Department] [Institution or Employer] | |
|  |  | |
| **Date Defended/Approved:** | [Month Day, YEAR] | |

Ethics Statement

The author, whose name appears on the title page of this work, has conducted the research as a co-investigator, collaborator, or research assistant in a project approved in advance.

Abstract

Abstract paragraphs should be unindented: use the style **1\_Para\_FlushLeft.** Master’s abstracts are limited to 150 words; the limit is 350 words for doctoral abstracts. Abstract text must fit on a single page. Keywords may appear alone on a second page.

You may include up to six keywords or phrases. Keywords should be separated with semicolons.

**Keywords**: thesis template; Microsoft Word; keywords; 1” hanging indent; instructions

Dedication

This is an optional page. Use your choice of paragraph style for text on this page (**1\_Para\_FlushLeft** shown here).

To hide the heading at the top of this page, select the text and change the text colour to white.

Acknowledgements

This is an optional page. Use your choice of paragraph style for text on this page (**1\_Para** shown here).

Table of Contents

[Approval ii](#_Toc483818976)

[Ethics Statement iii](#_Toc483818977)

[Abstract iv](#_Toc483818978)

[Dedication v](#_Toc483818979)

[Acknowledgements vi](#_Toc483818980)

[Table of Contents vii](#_Toc483818981)

[List of Tables viii](#_Toc483818982)

[List of Figures ix](#_Toc483818983)

[List of Acronyms x](#_Toc483818984)

[Glossary xi](#_Toc483818985)

[Chapter 1. Introduction 1](#_Toc483818986)

[1.1. Motivation 1](#_Toc483818987)

[1.2. Process Summary 1](#_Toc483818988)

[1.3. Aim 2](#_Toc483818989)

[Chapter 2. Process Analysis 3](#_Toc483818990)

[2.1. Process Description 3](#_Toc483818991)

[Chapter 3. Proposed Solutions 5](#_Toc483818992)

[3.1. Standardization of the process 5](#_Toc483818993)

[3.2. Possible Solutions 5](#_Toc483818994)

[Chapter 4. Technologies and Methodologies 8](#_Toc483818995)

[4.1. Development Environment 8](#_Toc483818996)

[4.2. Programming Technologies 9](#_Toc483818997)

[4.2.2. Why Python? 13](#_Toc483818998)

[Chapter 5. Technical walkthrough 14](#_Toc483818999)

[Chapter 6. Plagiarism Check 15](#_Toc483819000)

[Chapter 7. Testing Testrek 16](#_Toc483819001)

[Appendix A. An Example of an Appendix 17](#_Toc483819002)

[References 18](#_Toc483819003)

List of Tables

**No table of figures entries found.**  
This is an automatic table of contents. To use it, apply heading styles (on the Home tab) to the text that goes in your table of contents, and then update this table. If you want to type your own entries, use a manual table of contents (in the same menu as the automatic one).

Use References>Insert Caption to create caption labels and numbers. Right-click on the text above and select Update Field to update this list. See the Thesis Template Instructions for information on creating tables, figures, and captions.

Remember to delete this note before submission.

If there are no tables in the document, remove this page.

List of Figures

**No table of figures entries found.**

Use References>Insert Caption to create caption labels and numbers. Right-click on the text above and select Update Field to update this list. See the Thesis Template Instructions for information on creating tables, figures, and captions.

Remember to delete this note before submission.

If there are no figures in the document, remove this page.

List of Acronyms

|  |  |
| --- | --- |
| CTU | Czech Technical University |
| CSS | Computer Support for Study |
| GUI | Graphical User Interface |
| IDE | Integrated Development Environment |
| PyPI | Python Package Index |

Glossary

|  |  |
| --- | --- |
| Thesis |  |
| Python |  |
| Swim Lane Chart |  |
| IDE |  |
| Script |  |
| GUI |  |
| Open Source |  |
| Database |  |
| Hosting |  |
| PEP8 | Python code style conventions[[1]](#endnote-1) |
| Object Oriented Programming |  |
| Python Interactive Shell |  |
| modules |  |
| snippet |  |
|  |  |
|  |  |

# Introduction

The major motivation for this thesis comes from the fact that most of the work surrounding the examining of students is still done manually. This thesis introduces the need, scope and application that is built for the automation of the examination process for reducing the manual work. The software solution described and developed as a part of this thesis is tend to be designed in a way that it utilizes and requires the least resources and infrastructure for its operation. The application is made as dynamic as possible hence it can be used outside the set scope (i.e. for Computer Support for Study course), if the process of examination can be standardized for that course or subject. Different methods of approaching the automation are also described along with their benefits and shortcomings.

For the sake of simplicity, CSS course is used as an example all through this thesis, to analyze and demonstrate the challenges and solution proposals to different aspects of such an automation process.

When conducted manually, the whole examination process procedure is taken care of manually with the help of up to three or more people. From the preparation of question paper to reviewing the answers from students (most of the times in an unorganized way) and assigning grades. If a course or subject is registered by many students (which in fact is a case for the CSS course), for e.g. in order of hundreds then the whole process becomes even more cumbersome and hard to handle, leading to the slowdown of the review process and requirement of even more manual efforts.

The aim of this thesis is to provides a good understanding of the examination process where most of the work is done manually, along with ways to approach the automations of such nature. The problem that has been addressed in this thesis is mostly concerned with the organization and automation of the answer reviewing system and aid the plagiarism check. Also, a summary of how to achieve operational excellence by the deployment of small and cost effective automations is made part of this thesis.

The software solution ***Testrek***, developed mainly in Python as part of this thesis can help in organization of reviewing and running a preliminary plagiarism check on answers from the students taking part in the examination. The designing of the solution is dynamically approached so that the testing of any kind can use this solution if it can be standardized as mentioned in the 2nd Chapter. This thesis can also be used to conduct further research and development on automating the sub processes which are out of scope of this bachelor thesis.

There are information systems that already exists and can provide a potential solution for achieving the automation, if modifiable and extensible. For example, moodle[[2]](#endnote-2) is a shared learning platform which is hosted by CTU to share study material with students. Although by default, there is no possibility for students to upload their answers during the examination or for teachers to run a plagiarism check, it can be extended in order to do similar tasks that Testrek would do but the efforts required for achieving the same results as Testrek would require more time, resources, expertise in other front and back end programming languages.

The information in this thesis is ordered in way so that one starts by understanding the aim and challenges faced due to the current process, followed by a description of the solution proposals and technologies used to develop the solution itself.

# Process Analysis

Process Analysis constitute an important part of this thesis as it helped in understanding the different aspects of the problem which are tried to be solved while designing the software solution described in *Chapter 3*.

## /Users/akshat/Documents/Tand_Labs/Testrek/_01_Designs/images/Thesis_Test_current.jpgProcess Description

When conducted manually, the process map of whole examination process is explained using a swim lane chart on the right. Next chapter contains further information about the scope of this thesis and a rough estimation of efforts required to automate the sub processes which are not included in the scope.

The chart shows the current process procedure, with the estimations of work and in some cases time required for a sub process to complete.

# Proposed Solutions

## Standardization of the process

Before designing any solution for automating the examination process, it was necessary to standardize it. This in fact proved to be crucial while searching for different ways to approach the solution. The standardization helped in deciding the best possible way to approach the designing of the software solution, in the limited time frame. To make the solution fail safe it was necessary to set many rules for different parts of the process, but the challenges was to still maintain its dynamical nature.

After a thorough analysis, the two most qualified solutions which can be build are listed below with the requirement for standardisation, complexity, maintainability, infrastructure and support.

## Possible Solutions

##### 1st Solution (Testrek) – Preferred

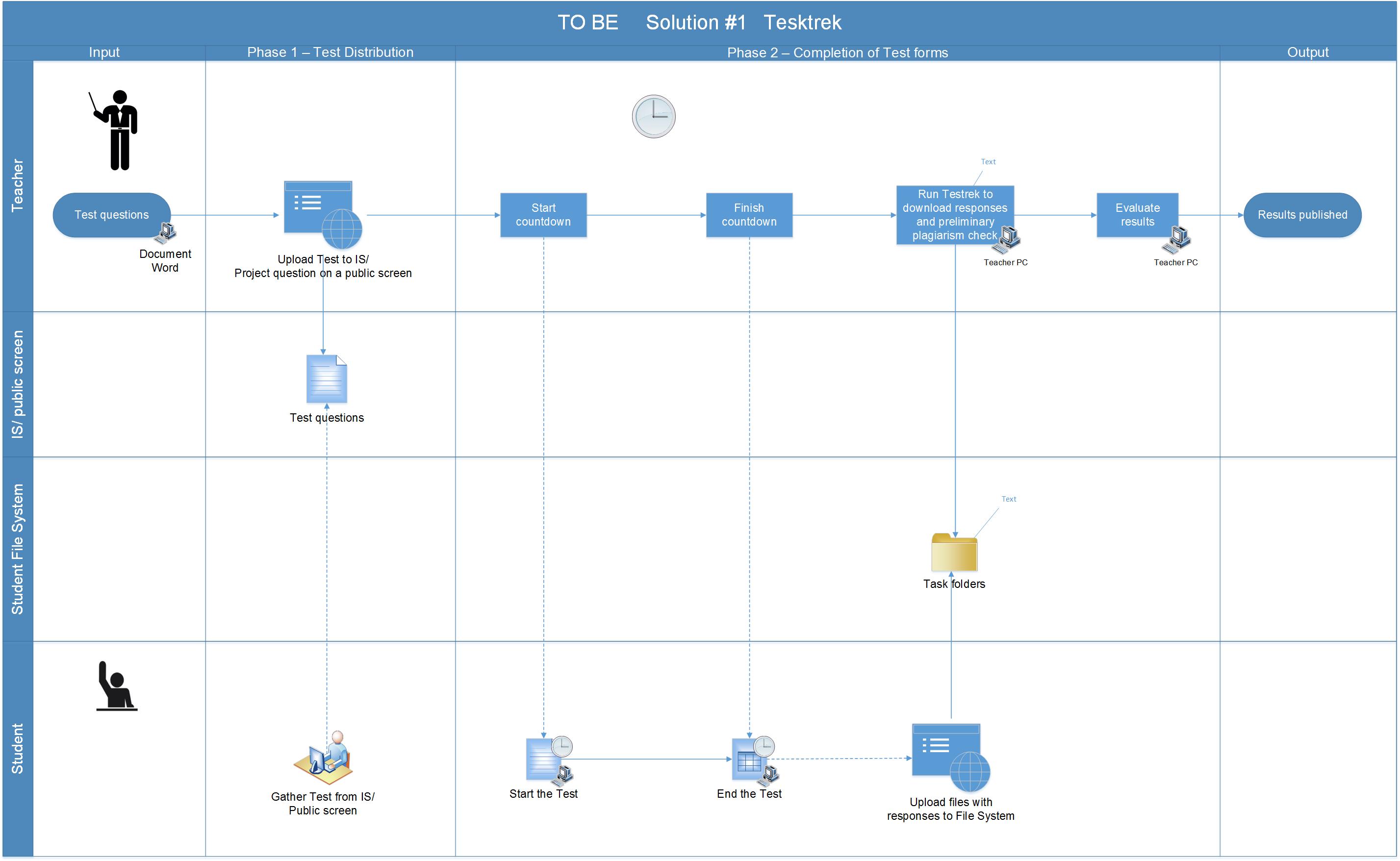
Basic Requirements for this solution to run:

1. Access to the internet
2. Disk space of about 30 Mega Bytes

This solution which is essentially a combination of scripts written in Python, is built as a part of this thesis. This solution was preferred because the requirements for running this application were very low and affordable. As the scripts are simple to understand, it would not require a lot of time to modify if the need be.

This solution also does not demand a lot of infrastructure hence it can run on most of the computers of today’s standard. Although it does not have a very user interaction to the of GUI, it is still easy to use and maintain. The challenging part for this solution is strictly standardizing many parts of the examination process, for example, where data is expected from the user.

The swim lane chart shows the process workflow if application like *Testrek* will be implemented for automation. A more thorough description of this solution can be found in the upcoming chapters.



##### 2nd Solution – Not preferred

As this solution will have to utilize a server, it will have a client side and the server side. Basic Requirements for this solution to run:

On client side:

1. Internet Connection and a modern web browser

On server side:

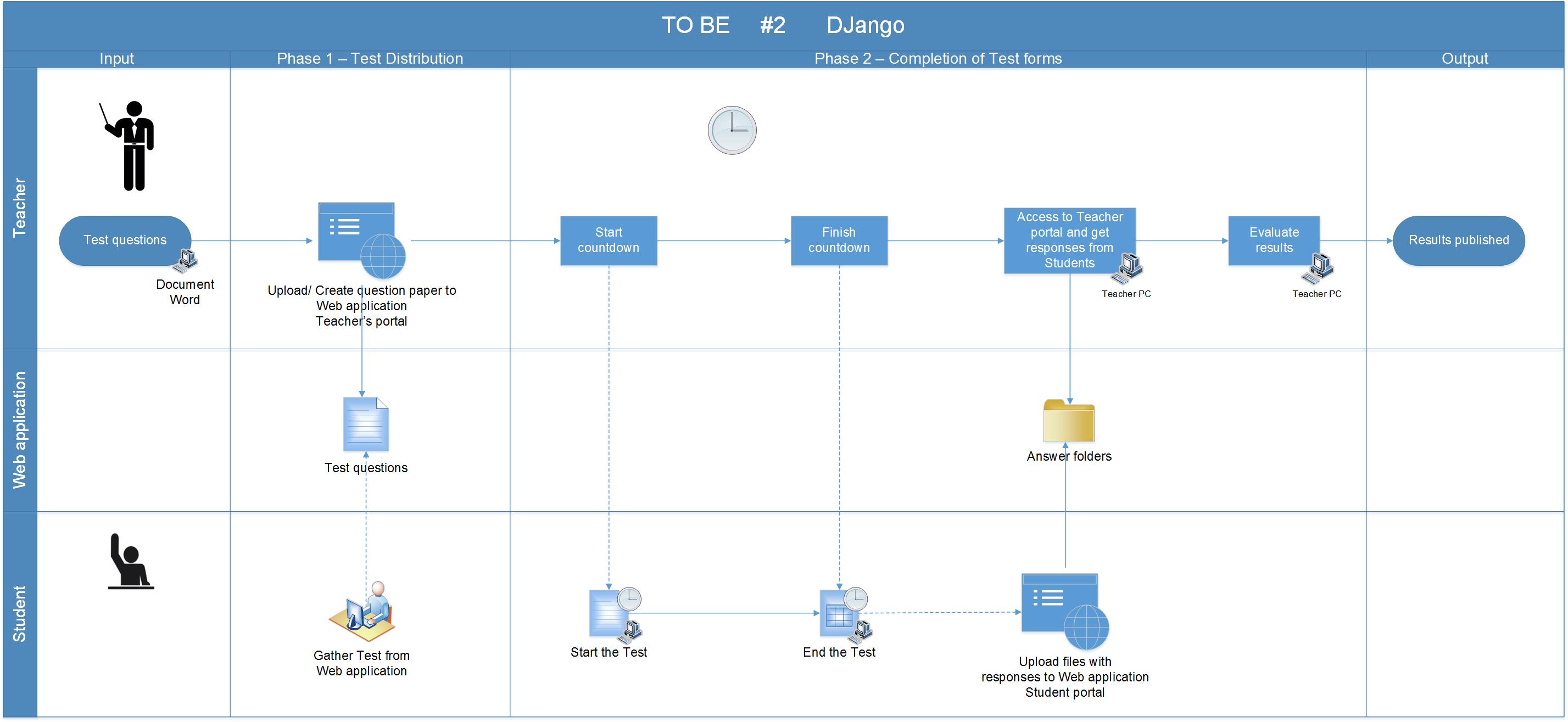
a. a database

1. a server supporting Python
2. an app hosting platform
3. maintenance and support

Building this application solution would require a robust web framework for Python, e.g. Flask, Django etc. It would utilize a database to store the information and a server to operate itself. Furthermore, the development process would require a lot more time and resources in comparison to the 1st solution.

Though the requirements are on the higher end when compared to the 1st solution, this solution which is essentially a web application will be more robust in its operation and delivery. This application would run on a web browser with an intensive GUI, hence will be more interactive for the users. One other essential characteristic of this solution would be the fact that it will not require as much standardization of the process as does the 1st Solution requires. Several validations can be run on the data as and when it is received, hence diminishing the need for much standardization.

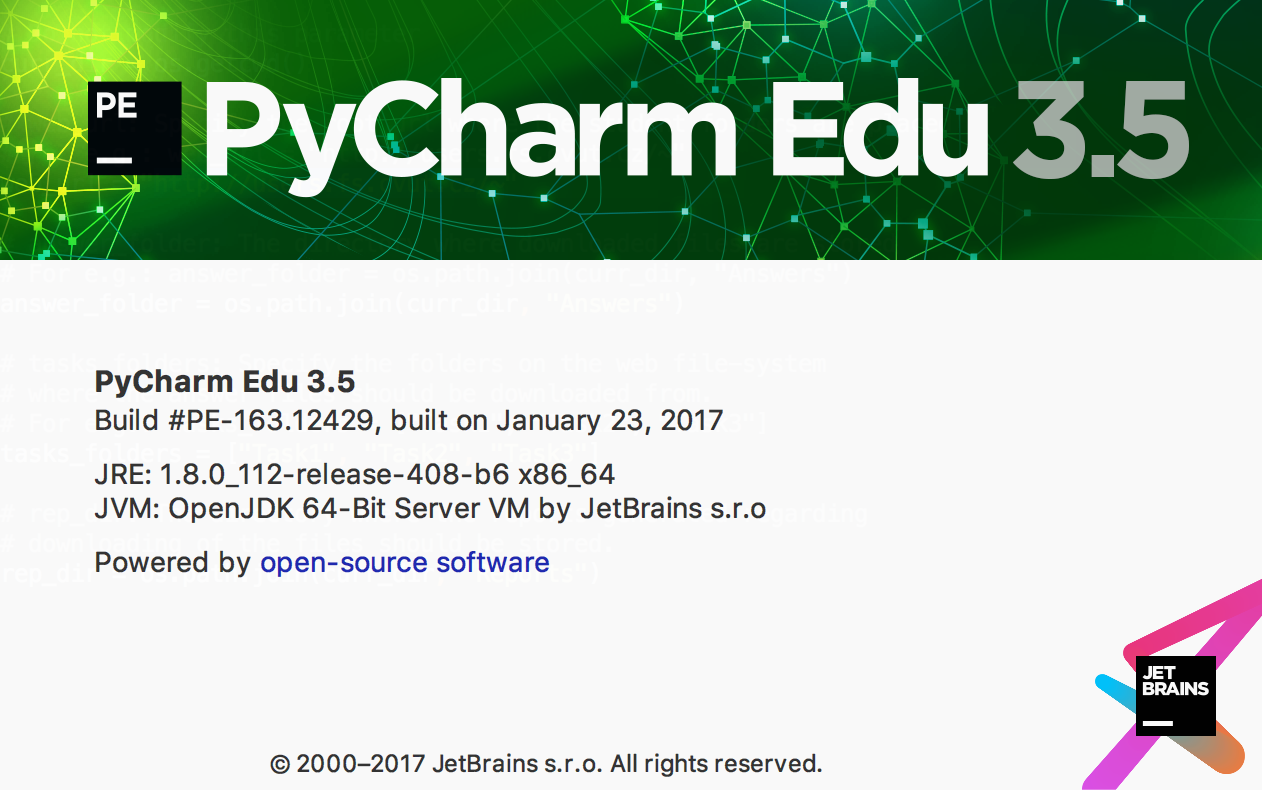
The swim lane chart below shows how the examination process would look like if a web application, written for example in python’s web framework Django, can be implemented to automate the examination process.



# Technologies and Methodologies

This chapter guides through different technologies and methodologies used, from development to the deployment of the application solution. The sub sections of this chapter consist of information about the platforms used for development, environment settings, programming technology and the libraries used.

## Development Environment

The solution had been built on a machine running MAC OS and the IDE or integrated Development Environment that was used to build the software solution is called PyCharm Edu 3.5[[3]](#endnote-3) provided by JetBrains s.r.o..

This IDE was selected as it has a great support for Python development and is an open-source software. It is best known for intelligent code completion, on-the-fly error checking and quick-fixes and easy project navigation. It helps keep quality under control with PEP8 checks, testing assistance, smart refactoring, and a host of inspections.

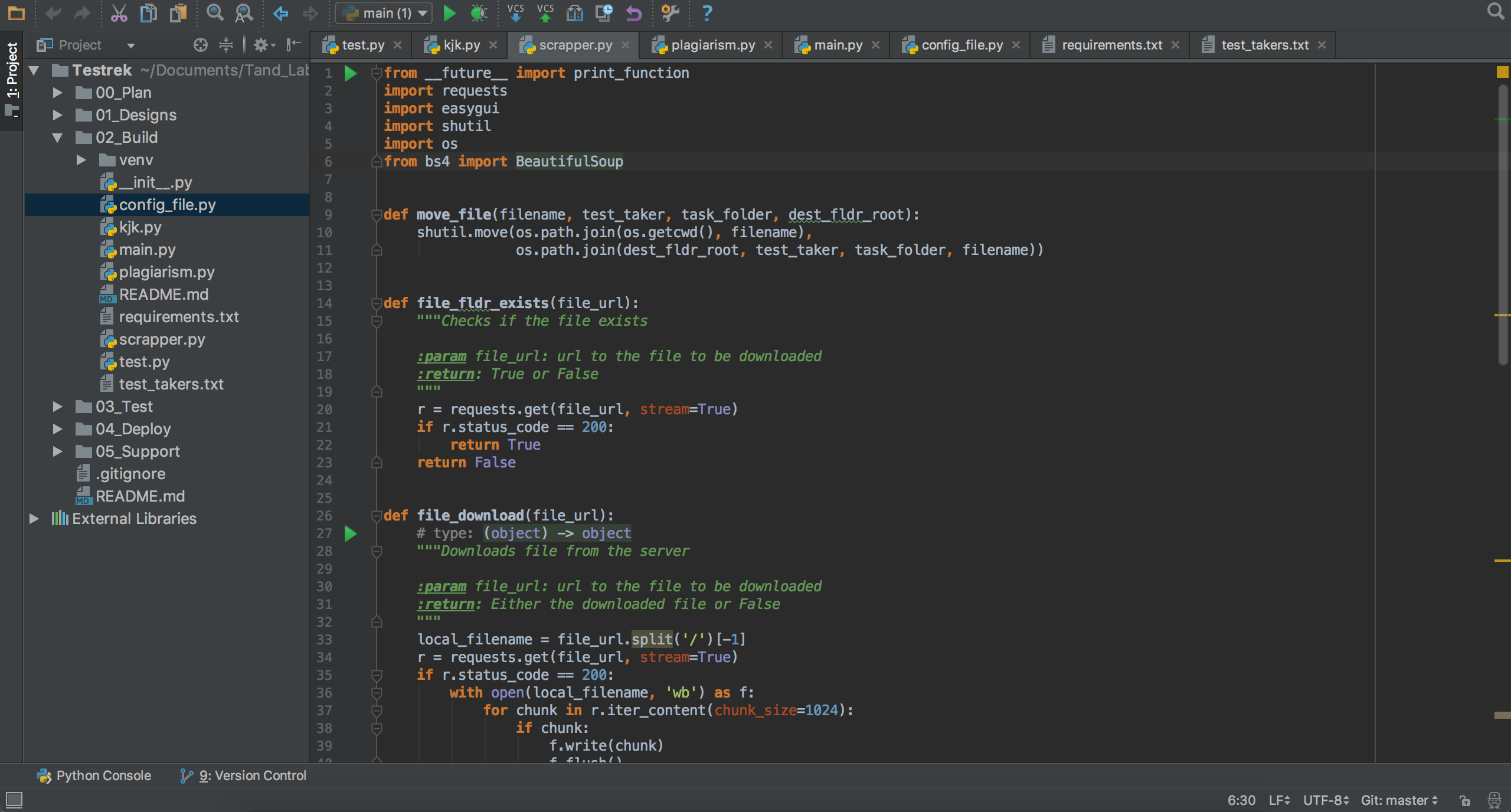


Figure PyCharm Workspace

## Programming Technologies

The software solution *Testrek* is purely written in the programming language called Python. Python is an interpreted, multi-purpose programming language that can be used to write web applications, GUIs, scripts and much more. It is strongly and dynamically typed with focus given to its readability and productivity. With an immense support from the community around it which builds a great range of libraries, it has proved to be a powerful language for scientific use and mathematical modelling. It is a self-contained object oriented programming language that has an interactive shell, strong introspection, cross platform capabilities and a variant for specific use like CPython, JPython, IronPythnon etc.[[4]](#endnote-4)

The versions of Python used for building and testing the scripts for this thesis are Python 2.7.10 and Python 3.5.2.

##### Hello World in Python

Writing “Python” in the command line starts the Python interactive shell which can be used to write python commands.

|  |
| --- |
| #!/usr/bin/env python  print "Hello World!" |

##### Indentation is necessary

Unlike most other programming languages, Python cares about the indentation and structure of the code. A sample is shown below:

|  |
| --- |
| #!/usr/bin/env python  for i in range(1, 10):  print ("I am number " + str(i))  if i == 9:  print ("9, here again!") |

##### Comments

Comments in Python are written as shown below. Any string that is not assigned to a variable or function is regarded as a comment.

|  |
| --- |
| # One line comment  """  This is a multi-line  comment.  """  "Any string not assigned to a variable is a comment" |

##### Data Types

Python has built in support for primitive data type like strings, Numbers, Null, Booleans, Lists, Tuples and Dictionaries[[5]](#endnote-5). Python dynamically assign the data type after a variable is initialized, hence declaration of variable is not included in Python.

|  |
| --- |
| # Strings  address = "This is a string."  address\_long = """This is a  long string."""  # Numbers  # Integers  age = 9  year = int("2010")  #Float  pi = 3.14159  # Null  data = None  # Booleans  is\_Python = True  # Lists  # initialisation  names = ["Charlie", "Brown", "Chris", 59, True]  # Appending  names.append("Martin")  # Tuples  # Can't be changed after initialisation  names = ("This", "is", "final", true)  # Dictionaries  # initialisation  dict = {'Name': 'Zara', 'Age': 7, 'Class': 'First'}  # Update  dict.update({  'Gender': 'Female',  'Hobby': 'Reading',  }) |

##### Basic Operators[[6]](#endnote-6)

|  |  |  |
| --- | --- | --- |
| **Operator** | **Description** | **Example** |
| + | Addition | a **+** b = 55 |
| - | Subtraction | a – b = 12 |
| \* | Multiplication | a \* b = 556 |
| / | Division | a / b = 45 |
| % | Modulus | b % a = 0 |
| \*\* | Exponent | a\*\*b =10 to the power 20 |
| // | Floor Division | 9//2 = 4 and 9.0//2.0 = 4.0, -11//3 = -4 |
| == | equivalence | (a == b) is not true. |
| != | Not equal | (a != b) is true. |
| > | Greater than | (a > b) is not true. |
| < | Less than | (a < b) is true. |
| >= | Greater than equal to | (a >= b) is not true. |
| <= | Less than equal to | (a <= b) is true. |

##### Control Flow

Python has support for conditionals, for and while loops, and list comprehension. An example of list comprehension is mentioned below:

|  |
| --- |
| # List comprehension  div\_by\_two = {x for x in range(10) if (x%2 == 0)} |

##### Classes and Functions[[7]](#endnote-7)

Classes and Functions makes up for an important part of any object-oriented programming language. Classes can inherit from other classes and ultimately from “object” class which is the top-level class in Python from which all classes inherit. Functions can accept argument or not. In Python, a function can be defined as below:

|  |
| --- |
| # Funtion that excepts an argument  class Any\_name(object):  def foo(i):  """Function documentation"""  remainder = i % 5  if (remainder!= 0):  return remainder |

### Why Python?

Python is undoubtedly an easy to use programming language with a great community that generously promotes and supports it. Because of the reliability that frameworks like Django, Flasks and Pylons etc. provide, they are being used as a primary platform for development for many software products. Python has no interfaces or real scoping of functions and methods, which lets developer concentrate more of the logic of the application than the syntax of the code itself. Due to all the above reasons, Python proved to be the right choice for developing *Testrek*.

## Python libraries

Python allows adding external modules (libraries) to a project. Packages are essentially a collection of dynamically written classes with variables and function, which can be re-used in another project. The keyword “import” is used to create a reference to these modules in a project and then the functions from these classes can be used.

|  |
| --- |
| # importing the whole datetime module and creating an alias dt  import datetime as dt  # importing only the classes timedelta and date from the datetime  from datetime import timedelta, date |

An exhaustive list of all the packages publicly available for Python is available on PyPI[[8]](#endnote-8).

Apart from the several packages that come by default with Python, a few other were also used while developing *Testrek*. I brief introduction to each one of them is mentioned in this section.

##### Requests (v. 2.14.2)

Requests is a HTTP library for Python, which send HTTP/1.1 requests without the need for much of work that is required to be taken care of while sending a HTTP request. For example, there is no need to manually add query strings the URLs or to form-encode the POST data.[[9]](#endnote-9) Connection pooling is also taken care of in the library itself which reduces the need to custom write the related code again and again.

***Testrek*** utilizes requests library to download files related to each task from the user’s filesystem (found under public\_html folder) on the University server.

##### Easygui (v. 0.98.1)

EasyGUI is a simple yet robust GUI written in Python. It is not event driven, instead all the GUI interactions are invoked by simple function calls.[[10]](#endnote-10) This GUI library is used to present dialogs with information during the runtime of ***Testrek***. It saves the user from knowing anything about tkinter, frames, widgets, callbacks or lambda, which are core to it. It runs smoothly on Python 2 and 3 and does not have any dependencies.

##### Fuzzywuzzy (v. 0.15.0)[[11]](#endnote-11)

Fuzzywuzzy is a package used for the string comparison. It uses Levenshtein Distance[[12]](#endnote-12) to calculate the differences between the sequence of strings. It is compatible with python 2.4 or higher. It utilizes difflib library that comes bundled with Python and uses the package python-Levenshtein to deliver results even faster. In ***Testtructor*** python-Levenshtein pachange is used along with fuzzywuzzy in order to get results as quickly as possible.

##### Tqdm (v. 4.14.0)[[13]](#endnote-13)

Tqdm package is available for Python 2.6 and higher is used to create a progress bar from the number of iterations. For implementation, it must simply wrap with the iterable. The snippet below one can see the implementation in a real-time scenario.

|  |
| --- |
| from tqdm import tqdm  for i in tqdm(range(500)):  ... |

##### Beautifulsoup4 (v. 4.6.0)[[14]](#endnote-14)

Beautifulsoup is a an extensively used python package for parsing through a web page. Since its emergence in 2004[[15]](#endnote-15), it has been under constant development and the latest version provides some great features and robust runtime performance. It is built upon am HTML or XML parser, providing extensive features to iterate, search and modify the parsing tree.

In ***Testrek***, it is mainly used to check for the existence of the files on the web url before *requests* library can be used to download that file. This was necessary to be done while providing the right results in the download success report.

# Technical walkthrough

Below is a list of all the standardization and suppositions that were required to be made before building the software solution:

The file\_download function defined below is used to download the files from the server. The requests.get method returns the response from the URL that is provided as a parameter. The function proceeds if the URL can be reached by checking the status code of the response method. Below are some status codes that can be returned by a server:

|  |  |
| --- | --- |
| Code | Description |
| 100 | Continue |
| 200 | OK |
| 201 | Created |
| 202 | Accepted |
| 400 | Bad Request |
| 404 | Not Found |
| 500 | Internal Server Error |

|  |
| --- |
| def file\_download(file\_url):  # type: (object) -> object  """Downloads file from the server  :param file\_url: url to the file to be downloaded  :return: Either the downloaded file or False  """  local\_filename = file\_url.split('/')[-1]  r = requests.get(file\_url, stream=True)  if r.status\_code == 200:  with open(local\_filename, 'wb') as f:  for chunk in r.iter\_content(chunk\_size=1024):  if chunk:  f.write(chunk)  f.flush()  return local\_filename  pass  return 0 |

# Plagiarism Check

# Testing Testrek

Appendix A.   
  
An Example of an Appendix

Appendices should be used for supplemental information that does not form part of the main research. Remember that figures and tables in appendices should not be listed in the List of Figures or List of Tables.

References

1. “[PEP 8 -- Style Guide for Python Code.](https://www.python.org/dev/peps/pep-0008/)” [↑](#endnote-ref-1)
2. “[ČVUT v Praze, Fakulta Strojní: Log in to the Site.](https://moodle.fs.cvut.cz/login/index.php)” [↑](#endnote-ref-2)
3. “[PyCharm](https://www.jetbrains.com/pycharm/).” [↑](#endnote-ref-3)
4. Nowell Strite, “[Introduction to Python.](https://www.slideshare.net/nowells/introduction-to-python-5182313)” [↑](#endnote-ref-4)
5. t[utorialspoint.com, “Python Dictionary.](https://www.tutorialspoint.com/python/python_dictionary.htm)” [↑](#endnote-ref-5)
6. [tutorialspoint.com, “Python Basic Operators.](https://www.tutorialspoint.com/python/python_basic_operators.htm)” [↑](#endnote-ref-6)
7. “[Learn Python the Hard Way.](https://learnpythonthehardway.org/book/ex40.html)” [↑](#endnote-ref-7)
8. “PyPI - the Python Package Index: [Python Package Index.](https://pypi.python.org/pypi)” [↑](#endnote-ref-8)
9. “Requests: HTTP for Humans — [Requests 2.17.3 Documentation.](http://docs.python-requests.org/en/master/)” [↑](#endnote-ref-9)
10. “[Easygui](https://github.com/robertlugg/easygui).” [↑](#endnote-ref-10)
11. “[Seatgeek/Fuzzywuzzy.](https://github.com/seatgeek/fuzzywuzzy)” [↑](#endnote-ref-11)
12. “[Levenshtein Distance.](https://people.cs.pitt.edu/~kirk/cs1501/Pruhs/Spring2006/assignments/editdistance/Levenshtein%20Distance.htm)” [↑](#endnote-ref-12)
13. [developers, *Tqdm*](https://pypi.python.org/pypi/tqdm). [↑](#endnote-ref-13)
14. [Richardson, *beautifulsoup4*.](https://pypi.python.org/pypi/beautifulsoup4) [↑](#endnote-ref-14)
15. “[Beautiful Soup](https://www.crummy.com/software/BeautifulSoup/): We Called Him Tortoise Because He Taught Us.” [↑](#endnote-ref-15)