Group Name: The mighty Boosh

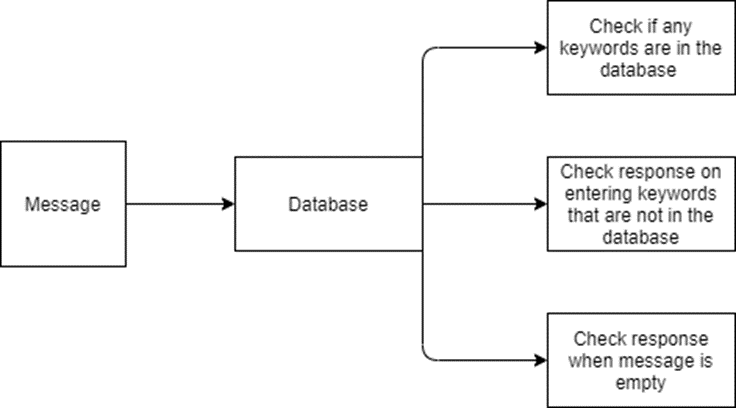
Individual Name: Marios Pelekanos

Solution testing:

The problem that is going to be tested is weather our program will be able to correctly identify the spam messages and produce accurate results. In order to achieve all thins the system has to be tested and trained to identify any keywords that will have the most impact on the result decision in each individual case. This mission is going to determine if the system is well trained and fully ready to categorise all the data into the right sections. Moreover, it is going to show how easily the system can be fooled and produce a false output.

Test cases:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Test Case ID | Test Scenario | Test Steps | Test Data | Expected Results | Actual Results | Pass/Fail |
| TC01 | Check if users’ message is spam | 1.Get message  2.Check database and match keywords | Message contains=Mobile Movie Club | Spam | As Expected | Pass |
| TC02 | Check if users’ message is ham | 1.Get message  2.Check database and match keywords | Message contains=No match | Ham | As Expected | Pass |
| TC03 | Check if users’ message is spam | 1.Get message  2.Check database and match keywords | Message contains=No match | Spam | Not Expected | Fail |
| TC04 | Check if users’ message is ham | 1.Get message  2.Check database and match keywords | Message contains=Just forced myself to eat ham | Spam | Not Expected | Fail |



No1Test Case Report

|  |  |  |  |
| --- | --- | --- | --- |
| **GENERAL INFORMATION** | | | |
| **Test Stage:** | CheckmarkUnit Functionality Integration System Interface  Performance Regression Acceptance Pilot | | |
| **Test Date:** | 04/20/2020 | **System Date, if applicable:** | 04/20/2020 |
| **Tester:** | Marios Pelekanos | **Test Case Number:** | Test NO1 |
| **Test Case Description:** | CheckmarkThis test case will indicate if the program will be able to determine spam messages | | |
| **Results:** | Pass | **Incident Number, if applicable:** | ID01 |
| **TEST** | | | |
| **Procedural Steps:** | In order to execute this test a scenario of a message must be given to the program. In this instance the program will try to match the following keyword “Mobile Movie Club”, If that keyword exists in its database then it will mark it as spam. If it doesn’t exist, it will continue to search for other keywords. | | |
| **Expected Results of Case:** | In this test the message that was given to the program had the keyword “Mobile Movie Club”. The program will search if that keyword exists in its database and proceed with the result. In this instance the given keyword exists in the database and the program should return with a spam result as it is expected. | | |
| **ACTUAL RESULTS** | | | |
| **Output Specifications:** | This test was successfully completed leading to an overall Pass. | | |

No2Test Case Report

|  |  |  |  |
| --- | --- | --- | --- |
| **GENERAL INFORMATION** | | | |
| **Test Stage:** | CheckmarkUnit Functionality Integration System Interface  Performance Regression Acceptance Pilot | | |
| **Test Date:** | 04/21/2020 | **System Date, if applicable:** | 04/21/2020 |
| **Tester:** | Marios Pelekanos | **Test Case Number:** | Test NO2 |
| **Test Case Description:** | This test case will indicate if the program will be able to determine spam messages | | |
| **Results:** | CheckmarkPass | **Incident Number, if applicable:** | ID02 |
| **TEST** | | | |
| **Procedural Steps:** | In order to execute this test a scenario of a message has to be given to the program. In this instance the program will try to match the given keyword. If that keyword exists in its database, then it will mark it as spam. If it doesn’t exist it will continue to search for the rest keywords. | | |
| **Expected Results of Case:** | In this test case the message that was given to the program will be cycled through the database. In this instance the given keyword doesn’t exist in the database and the program should return with a non-spam result as it is expected. | | |
| **ACTUAL RESULTS** | | | |
| **Output Specifications:** | This test was successfully completed leading to an overall Pass. | | |

No3Test Case Report

|  |  |  |  |
| --- | --- | --- | --- |
| **GENERAL INFORMATION** | | | |
| **Test Stage:** | CheckmarkUnit Functionality Integration System Interface  Performance Regression Acceptance Pilot | | |
| **Test Date:** | 04/22/2020 | **System Date, if applicable:** | 04/22/2020 |
| **Tester:** | Marios Pelekanos | **Test Case Number:** | Test NO3 |
| **Test Case Description:** | CheckmarkThis test case will indicate if the program will be able to determine spam messages | | |
| **Results:** | Fail | **Incident Number, if applicable:** | ID03 |
| **TEST** | | | |
| **Procedural Steps:** | In order to execute this test case a scenario of a message has to be given to the program. In this instance the program will try to match the given keyword. If that keyword exists in its database then it will mark it as spam. If it doesn’t exist it will continue to search for the rest keywords. | | |
| **Expected Results of Case:** | In this test case the message that was given to the program will be cycled through the database. In this instance the given keyword doesn’t exist in the database and the program should return with a non-spam result as it is expected. But in this occasion this test tricked the program by providing a keyword that did exist in its database. This output is incorrect as the message that was provided was not a spam message. | | |
| **ACTUAL RESULTS** | | | |
| **Output Specifications:** | This test was not successfully completed leading to an overall Fail. | | |

No4Test Case Report

|  |  |  |  |
| --- | --- | --- | --- |
| **GENERAL INFORMATION** | | | |
| **Test Stage:** | CheckmarkUnit Functionality Integration System Interface  Performance Regression Acceptance Pilot | | |
| **Test Date:** | 04/23/2020 | **System Date, if applicable:** | 04/23/2020 |
| **Tester:** | Marios Pelekanos | **Test Case Number:** | Test NO4 |
| **Test Case Description:** | CheckmarkThis test case will indicate if the program will be able to determine spam messages | | |
| **Results:** | Fail | **Incident Number, if applicable:** | ID04 |
| **TEST** | | | |
| **Procedural Steps:** | In order to execute this test a scenario of a message must be given to the program. In this instance the program will try to match the following keyword “Just forced my self to eat ham”, If that keyword exists in its database then it will mark it as spam. If it doesn’t exist, it will continue to search for other keywords. | | |
| **Expected Results of Case:** | In this test the message that was given to the program had the keyword “Just forced my self to eat ham”. The program will search if that keyword exists in its database and proceed with the result. In this instance the given keyword doesn’t exist in the database but the program categorised it as spam. | | |
| **ACTUAL RESULTS** | | | |
| **Output Specifications:** | This test was successfully completed leading to an overall Fail. | | |

Solution Review:

Comparison:

There are many software-based systems that their basic function is to determine whether the massage that the user has received is spam or not. In our testing our implemented software seem to be well build and robust and at the same time being very accurate in most occasions with very good results. Some other big companies like Yahoo, outlook and Gmail have their own software to determine if the messages of their users are spam or not. Yahoo has its own spam algorithms that it uses to detect spam messages.

The basic methods that Yahoo uses is URL filtering, email content and spam complaints from Users. Unlike other systems Yahoo filters its email messages by domains not IP addresses. Moreover, it checks whither an anonymous sender exist in a database. If there is no history of spamming, then the message is sent to the recipient. On the other hand, Gmail uses another approach in message spamming, it filters emails by IP addresses and not domains. In its classification of emails, Google is said to use state-of-the-art spam detection machine learning algorithms such as logistic regression and neural networks. In addition, Gmail uses optical character recognition (OCR) to protect users from image spam.

Discussion:

Our system results ensure that our software implementation is very accurate. But in this instance as we can see from the competition their system implementation is far more complicated thus delivery more promising results. In fairness of that it is said that Gmail’s algorithm filters the spam messages from 99.5% to 99.99% accuracy.

According to the information that is mentioned above you can’t really compare our implementation with other big companies that exist out there. It is very hard to match their abilities of spamming software because of their years of experience, their funding and their number of employees.

Recommendations:

One thing that would greatly benefit our program is to use more algorithms to determine whether the messages that the user receives is spam or not. In addition, systems and methods may be provided in accordance with other embodiments to determine a degree of similarity between a user query or search string and content items that are stored in a database. If the degree of similarity exceeds a predetermined threshold, the Database content item may be identified as a content item that matches the user's query or search string.

References:

Dada, Emmanuel Gbenga, et al. “Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems.” Heliyon, vol. 5, no. 6, June 2019, p. e01802, 10.1016/j.heliyon.2019.e01802. Accessed 15 Sept. 2019.

“Systems and Methods for Providing a Spam Database and Identifying Spam Communications.” Google Patents, patents.google.com/patent/US9407463B2/en. Accessed 17 May 2020.

‌