Team number: 23

Project Title: News tone

1. **Introduction** 
   1. **Project Overview**

News tone aims to create an implementation of sentiment analysis for a certain news content crawled on a given website. As digital media has become a growing trend of the news publishing industry, uncountable articles are generated on the Internet every day. Along with the time elapsing, our program and algorithm are focusing on investigating the polarity of the vocabularies used in the article and how the polarity of an article changed over time.

* 1. **Scope of the test plan document**

This document illustrates the test strategies which have been practised and will have been practised in the project to demonstrate the actual performance of the sentiment analysis program. The risks of the testing have been identified. The testing schedule is attached as Ghatt chart as an appendix at the end of the document.

1. **Test strategy** 
   1. **Test scope: what will be tested and why**

The accuracy of the analysing algorithm as the core of the program had been tested before interaction 1. As the vital component of the sentiment analysis, the accuracy of the algorithm determines the functionality of News tone to a large extent.

* 1. **Testing report:**

All the testing report has recorded. Part of the testing rocord has been addressed as an appendix at the end of the document.

* 1. **Test assumptions**
* All the testing result will be recorded.
* Most of the issues which identified will be resolved, at least will not bring negative impact.
* Errors and bugs will be fixed after the testing.

1. **Test execution**
   1. **Defect testing**

The web crawler cannot avoid crawling advertising information from the webpage as part of the sample which may influence the result of the analysis to some extent. A study of the impact is needed to demonstrate such a situation in the final report. Moreover, some words may have multiple meanings under diverse circumstances. For instance, "Trump" could be a family name and a word at the same time. As a family name, the polarity is neutral but as a word, the polarity is most likely positive. These are some potential defect our algorithm may have and will hardly have solutions to deal with. The test will go through such defect and determine the extent of the impact on the result.

* 1. **Unit testing**

The unit test will be divided into 3 parts which will test webpage content crawler, sentiment analysis algorithm, and JSP web server explicitly.

1. The data crawled by the crawler will be compared with the original webpages to test accuracy and integrality of the crawler.
2. The outcome of sentiment analysis for a certain text will be contrasted with human recognition to test how the algorithm works.
3. All the requests and responses handled by the JSP web server will be monitored to test the reliability and stability.
   1. **Integration and release testing**

As advertisment and polysemy are extremely hard to avoid, the negative impact of advertising and polysemy will be investigated in the coming testing. A reliable report will be included in the final report based on the testing.

1. **Testing plan schedule**

Refering appendix B

1. **Risks**

Risks management identifies potential risks that may bring negative impact to the testing procedure.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Risk | Impact | Cause | Mitigation strategy |
| 1 | High | Web scrawler cannot download information from the designated website in a certain time | Unreliable Internet connection or poor Internet connection | Find a backup Internet connection socket. Improve the Internet conection quality. |
| 2 | High | Web scrawler cannot download information from the designated website | Destination server maintainance | Change another time to have the testing executed. |
| 3 | Medium | Akgorithm fault | Algorithm has potential defect which have not identified | Terminate the testing, improve the algorithm porformance, and replace the testing after the adjustment of the algorithm. |

**Appendix A - Unit testing**

|  |  |  |  |
| --- | --- | --- | --- |
| Test description | Input | Expected output | Current output |
| Web scrawler testing | URL address of certain webpage which contains news titles | All the news titles on the webpages for a certain time span. | Same as the original website illustrated. |
| Sentiment analysis algorithm testing | Text content scrawled from the designated website | An accurate polarity analysis based on the input context | Some of the analysis outcome is incorrest. For instance, a same word with different meanings. |
| JSP web server testing |  |  |  |
| Test type: unit test  Date: 11/05/2020 | | Responsible: Ying Wang, Yiru Li  Reviewer: Jin Zhou | |

**Appendix B – Testing Plan Schedule**

