|  |  |
| --- | --- |
| PROJECT ANALYSIS  Week 3 Assignment | ABSTRACT  This document was created for UMUC Course, CMSC 495, and analyzes aspects of the (TNC)  Group 3 Members  Name: Christiano, Andrew  Name: Fernandez, Yrume  Name: Orwick, Brian  Name: Sell, Julia  Class: CMSC 495 - Current Trends and Projects in Computer Science Professor: Dr. Hung Dao  Due: 10 September 2018 |

**Version Control**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Revision # | Date | Name | Descriptions | Contact Info |
| TNC\_0001 | 9/4/2018 | Brian Orwick | Created | Orwick12@outlook.com |
| TNC\_0002 | 9/5/2018 | Yrume Fernandez | Revisions | Yrume.fernandez@gmail.com |

Table of Contents

[1. Outside Systems 4](#_Toc524032537)

[2. Input Data 4](#_Toc524032538)

[3. Output Data 4](#_Toc524032539)

[4. Data Processing 5](#_Toc524032540)

[5. Subsystem Requirements 5](#_Toc524032541)

[6. Data Interface 6](#_Toc524032542)

[7. Potential Risk and Mitigation 6](#_Toc524032543)

[8. Future Enhancements 6](#_Toc524032544)

# Outside Systems

The News Code (TNC) software connects to and sends a get-request using Hyper-Text Transfer Protocol (HTTP) to the trusted news sources hard coded into the source code. It requires internet connectivity using either internet protocol (IP) version four or six (IPV4 / IPV6) and leverages the newspaper python library to checks for, connect, and download news articles and required information from websites. The following is a list of trusted sources that are hardcoded for TNC to check:

* <https://www.foxnews.com>
* <https://www.usatoday.com>
* <https://www.cnn.com>
* <https://www.bbc.com>



Figure - Architectural Context Diagram (ACD) External Interfaces

# Input Data

TNC does not require any input data from a user directly. A user will navigate to the TNC Server Web Page (which initiates code execution) and receive statistical information on articles posted within the hardcoded list of trusted news sources. TNC takes input data from web pages that are part of the hardcoded list of trusted news websites. This data is then parsed and structured for storage and retrieval using the SQLite Data Base that is part of overall TNC system.

# Output Data

TNC will parse through data collected from trusted news sources, and will export the data in accordance with data structure to a local SQLite Database (DB). This database will house all the required information to statistically analyze the trustworthiness of posted articles. The following provides information on the structure of the database table used during execution of the TNC:

|  |  |  |  |
| --- | --- | --- | --- |
| NEWS ARTICLE (TABLE NAME) | | | |
| ID (TEXT) | NUMBER (INT) | NAME (TEXT) | COUNT (INT) |
| Web-Page URL | Article Number | Article Name | Article Count |

Figure - Database Table (News Article)

# Data Processing

Information captured and stored within the TNC is used to identify the trustworthiness of an article. Initially the TNC software, using the newspaper library, connects and downloads a copy of newspaper articles housed within the source websites. This information is captured and stored for subsequent processing within the SQLite DB. After each site, within the list of trusted sites, is captured and stored the TNC software parses and iterates through the contents of each articles sequentially comparing each word within articles. As the system compares each article a counter is incremented when common words are found and a percentage of common words is calculated. If two articles have over 70%-word commonality (> 70), then both articles are assumed related. The system adds a count to each article identified as related, and articles acquiring a higher count of related articles are assumed more trustworthy than lesser related articles. All data is stored in database for storage and retrieval.

# Subsystem Requirements

TNC Software requires connections to local database (SQLite DB) which provides a structure to maintain information for analysis. This DB also provides the capability of running analysis on articles while the system is offline. The TNC requires access to local systems networking services to retrieve the required information from external websites while online.

# Data Interface



Figure 3 - Data Flow Diagram (DFD) Interfaces

# Potential Risk and Mitigation

Potential risk when operating the system are:

1. The system may fail to retrieve information from trusted websites (due to https certificate error, no internet connection, or other unknown circumstances)
   1. Mitigation: System will check for error code 200 to ensure website is available prior to submitting get-request for html.
2. System may fail to provide adequate information if fake-news goes viral and a number of articles are found on the internet based on false information.
   1. Mitigation: System will have a fake-news option that will lower the trustworthiness of the article.

# Future Enhancements

1. Provide option for users to input website uniform resource locator (URL) as an additional trusted news source.