# Terms of Service Analyzer

Keerthana Madhavan

**School of Computer Science –University of Windsor**

**Keywords*:*** Terms of Service, Outline, Communication, Information, important sentences

## Abstract

Nowadays, the internet provides many application and software platforms that are available to the public. Although the internet is a boon for communication and information, it is also the root cause for many challenges. One of the main issues is to accept the terms and conditions of these applications without proper reading. This is a problem as Terms of Service is one of the fundamentals that provide users with rules and regulations of how two or more parties communicate, and people must read the agreements. However, people struggle to read the agreements because of factors such as time, word complexity, and length. In this paper, we propose a novel application which aims to create an outline-based summary for the terms of service agreements to encourage humans to read and help them to understand the content that is to be signed. The proposed model extracts and generates a summary for each paragraph in the agreement and creates a custom outline summary. Outlines help the user retain and understand the information clearly, so we hope this method enhances the user knowledge on Terms of Service Agreements.

## 1 Introduction

Over the past few decades, we have observed the advances in digital technology and applications in the world. Technology is the key in today's world, and it has its own vast benefits and consequences. According to the recent statistics [1], the number of internet users is approximately 4.388 billion with a 9.1 percent rate of increase and the number of social media users is 4.484 billion with a 9 percent rate of increase on a year-to-year basis. The Web has given us access to compelling services, especially social media which is pivotal when the question arises of data privacy and security. The users who don't read the TOS (Terms of Service) may not be aware of the legal consequences which is a major problem of our society as the user must acknowledge the legal information before signing the agreements.

Terms of Service govern the relationship between two parties, and it is important that people especially the young generation understand the significance of law and regulations. As a result of Deloitte's research, about 91 percent of consumers of the age 18-35 tend to ignore reading TOS because of factors such as time, length, and word difficulty [2], which exaggerates the importance of reading legal agreements.

One of the biggest challenges on the internet is "I have read and accept the Terms of Service". It does not reflect the reality and will remain as a critical challenge if there is no change. In [3], it has been stated that the only solution to get humans reading TOS is the simplification of it into plain and natural language.

Simplification is the approach taken in this project as an attempt to encourage users to read Terms of Service. The application solution will analyze the TOS of different service providers, contemplate the information and provide a summarized outline of such agreements. This approach is different since most TOS simplification provides consumers with a rating or grade, but the

concern of this study is to advise user to read, especially that of law and contracts. As the problem is with factors of time and length that restricts people from reading TOS, this model attempts to provide a solution.

The rest of the paper is organized as follows: in the next section we review related works, after that, the proposed model will be presented and finally, we will conclude.

**2 Related Works**

A number of researchers have been studying and implemented applications on simplification methods for Terms of Service for Social Media and software applications. Some methods include sentiment analysis, user-based review and self-reviewed certifications of such services.

Hugo Roy in [4] established the Terms of Service; Didn't Read application to help fix the biggest lie on the web, "I have read and agreed to the terms of service. This is a chrome-based extension application that provides ratings of popular social media websites ranging from Class A (green) to Class E (red).

The terms-service tracker in [5] is an organization that hosts a website that tracks changes made to TOS agreements and highlights the updated version along with reviews.

Alike Roy's application, Click Wrapped in [6] analyzes the TOS agreements and gives it a grade for each company in their databases.

Terms & Conditions checker in [7] is another chrome extension but unlike TOS;DR which presents user with self-reviewed summary, this application is built with a custom search engine for users to look up definitions and keywords felt suspicious.

## 3 Proposed Model

Extractive methods attempt to summarize the content by selecting a subset of words that represent the most important points in the text. The proposed system uses extractive summarization methods that consists of three parts: Headline, paragraph and sentence extraction, custom outline generator, and user interface. Each part is described in the following sections in detail. The overview of the model is illustrated in Figure 1.

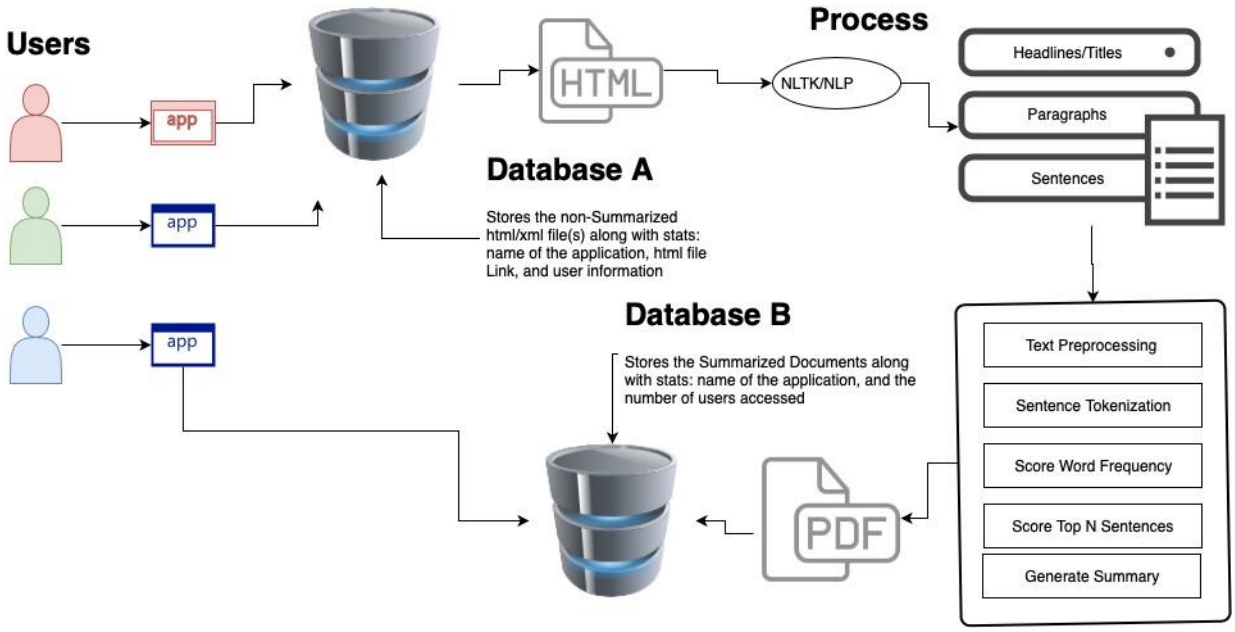


*Figure 1 Overview of the proposed model*

### 3.1 Methodology

In this section, we explain how to extract information and create the summarized outline using NLP based techniques and Python's NLTK library for text summarization. This process consists of five steps. The first step is for text extraction and paragraph splitting. The second step is text preprocessing of the paragraph(s). The third step is the word frequency score. The fourth step is sentence ranking and the fifth step prints the top N sentences to the outline document.

### 3.1.1 Text Extraction

Import or install the necessary NLTK and beautifulSoup4 libraries for text extraction from one of the HTML/XML files in Database A. Then we parse the data using the beautifulSoup4 method using the following tags <h> and <p> to split the headlines and paragraph into separate list (list[headlines] and list[paragraphs]).

### 3.1.2 Text Preprocessing

Using the sent_tokenize() split the paragraphs into sentences (list[sentences]) and split into words with the word_tokenize(). Then convert all text to lower case. Stop words, special characters, are filtered out by the nltk.corpus class.

### 3.1.3 Word Frequency Score

Find the weighted word frequency of each word in the list[sentences] divided by the highest number of word occurrence.

### 3.1.4 Sentence Ranking

Replace the weighted word frequencies in the list[sentences] with the corresponding words and find the sum. Now using the sum, orders the sentence from highest to lowest. Then choose the relevant and ‘n’ most important sentences to be extracted.

### 3.1.5 Outline Generator

Create an empty pdf document and input the list[title] and the relevant list[sentences] forming an outline of the entire HTML/XML text. The outline version does not contain any redundant or repetitive information, but useful information generated by the sentence ranking part of the model. The generated outline is stored in database B that holds all the summarized Terms of Service agreements of different applications which can be accessed easily by new users instantly.

**4. Conclusion.**

This paper has presented an outline-based summary system for Terms of Service agreements. The only solution in encouraging users to read legal agreements and avoid legal consequences is simplification of such agreements. Terms of Service agreements usually consist of large number of paragraphs. So, in order to effectively summarize the TOS an outline format is created by the model that serves as one of the solutions. Since an outline allows the writer to categorize the main points of a paragraph into bullet form that is effective in terms of comprehension of the content from the user side. It is estimated that the model can provide quality outline based on the agreement by extracting top and relevant sentences. This will not only reduce the length of the agreements but reduce the time of reading and complexity of words and ambiguity of the TOS agreements. Such, simplification will aid in effective communication between users and the service providers eliminating any redundant or unimportant legal concerns.

**References**


1. Chaffey, D. (2019, April 12). Global social media research summary 2019. Retrieved from https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/
2. Cakebread, C. (2017, November 15). You're not alone, no one reads terms of service agreements. Retrieved from https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11
3. Why Do We Blindly Sign Terms Of Service Agreements? (2014, September 01). Retrieved from https://www.npr.org/2014/09/01/345044359/why-do-we-blindly-sign-terms-of-service-agreements
4. Team, D. (n.d.). Terms of Service; Didn't Read. Retrieved from https://tosdr.org/
5. TOSBack. (n.d.). Retrieved from https://tosback.org/
6. Clickwrapped Who respects your rights online? (n.d.). Retrieved from https://www.clickwrapped.com/
7. Terms & Conditions Checker. (n.d.). Retrieved from https://chrome.google.com/webstore/detail/terms-conditions-checker/dpiobhofenkdpailcoedlnfkbiogmoap
8. Software Licenses in Plain English. (n.d.). Retrieved from https://tldrlegal.com/