CS 411W Lab 1 Ver. 1

The User Agreement Dilemma

Matthew Rojas

18 January 2022

Table of Contents

Listing of Figures

1 Introduction

Terms of service agreements are the legal jargon companies with any sort of software or

web service. With any terms of service, a user typically must agree to the agreement written by

the company of the current service trying to be used/signed up for. The reason companies require

every single user to agree before using their website, software or services is to protect the

company legally. This can include several different things from agreeing that a user's data can be

harvested, saved, stored, sold to advertising companies, or repurposed at the company's

discretion. To do this effectively the companies in question do not explicitly tell the users signing

up for their service out right that these things can/will happen. The company hides these

conditions in plain sight as legal jargon.

1.1 Problem Statement

Terms of Services on average 7,700 or more words for the 10 most popular applications

and services and are written in legal jargon that is difficult to understand. On average a chapter in

a novel can be any length of 1,500 to 5,000 words while the average is between 3,000 and 4,000.

A user can then expect a Terms of Service agreement to be the length of two chapters of a book.

On the surface this is seems like a small trial to go through to make sure that a user's data is

being handled in their favor. If a user were to only sign up for the top 5 most popular social

media websites the user can expect to read up to at least 10 chapters worth of reading material

and within that material the conditions are not explicit.

1.2 Societal Problem

1.2.1   TOS Length and Difficulty

As stated, before in section 1.1 the Terms of Service a company writes up is designed to

protect the company in question and gain the legal right to use the data of the user, accepting the

Terms of Service. On average a user can generate around $35 in United States dollars (USD) a

month, for a single user this does not seem reasonable to decline for a *free* service. There are an

estimated 2.89 billion users on Facebook equaling to a staggering $101 billion USD a month.

This possible monetary gain gives companies more of an incentive to hide their true intentions

with their users and the users' data. On average 91% of users do not read the Terms of Service

contracts when signing up for a new service or website [9]. Even though there are (according to

citation 9 in section 6) an average of 9% of users who read the Terms of Service. On average for

adults between ages 25 and 64 only 60% of adults understood the contents of the Terms of

Service [10]. For simplicity's sake of 1000 adults were a focus group ~54 people would

understand what the Terms of Service is asking users to agree too. Expanding this to literal

millions of people on a popular website such as Facebook or Instagram a lot of users end up

accepting to the terms of the website without understanding a single line of the contract each

willingly signed. Shown in figure 1 Terms of Service are riddled with long broad sections that

are designed to again protect the company in question legally rather than the users of

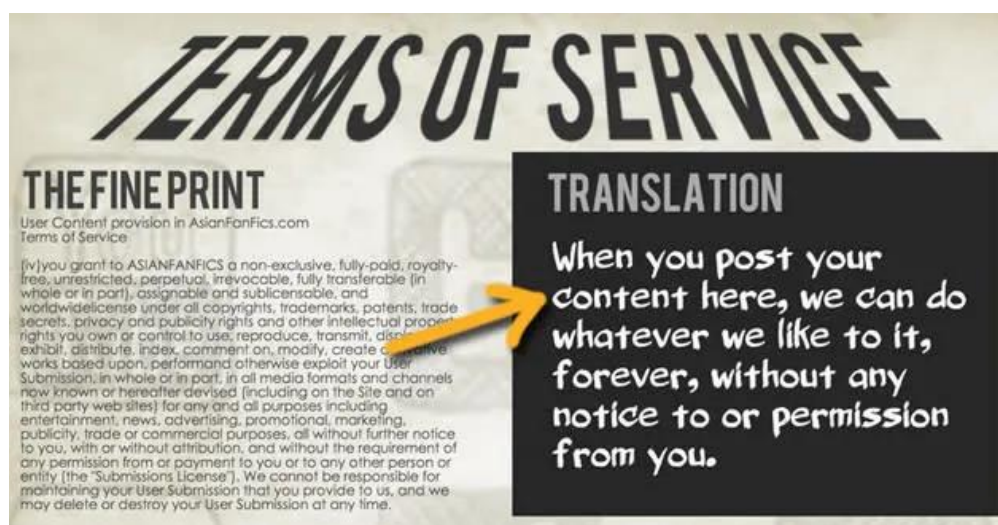application/website offered by the company.

Figure 1: The Fine Print

Figure 1 shows only a small snippet of Terms of Service for a company called ASIANFANFICS. The document contains the words/phrases, non-exclusive, fully paid, royalty-free, unrestricted, perpetual, irrevocable, etc.… Never straight out saying that ASIANFANFICS can and will use all user created content for anything. That includes possible advertisements (online and TV), which at the end of the day can give the company momentary gain that the users will possibly see a small, small percentage or nothing at all and users have willingly accepted these terms. As stated in sections 1.2.2 and 1.2.3 only 9% of users read the Terms of Service and only 60% of that understand the Terms of Service. This leaves many users not understanding the contents of the contract or not actually reading the contents.

1.3 Solution – Legal Lite

Legal Lite is the tool to help all users make a more informed decision in trek of understanding any Terms of Service contract. Legal Lite will be an additional tool for users to interact with for any website or service. Legal Lite will consist of users using the email used to sign up for the services in question to help keep all services linked to a single profile. This will allow Legal Lite to scrape websites with accounts connected to the email address. This will allow the user to quickly see what websites/services are connected to the email address. The overall goal of Legal Lite is to cut down the misunderstandings of Terms of Services. Remember is section 1.2.5 the vast majority of user do not read Terms of Service agreements. A smaller number of users do read but even fewer do not understand the Terms of Service agreements. This is where Legal Lite will assist users, Legal Lite will translate the legal jargon of a Terms of Service agreement into plain English. Currently Terms of Service agreements on average require a college sophomore reading level, the plain English Legal Lite is aiming for is a grade school level [10].

Legal Lite will require a vast amount of interaction from the community using Legal Lite.

Legal Lite will allow users to go through undocumented Terms of Service agreements and post

user generated translations for said website and services. Legal Lite will still need a moderator to

inspect these user generated posts before all users are allowed to read them. Legal Lite will

contain a ranking system of two categories, the first category will be based on the moderator's

opinion given after each translation specifically. While the second will be user based ranking

system, users may agree that personal data does not outweigh the technical benefits of the service

or software. Legal Lite will also be able to compare multiple services that are deemed similar to

one another (i.e., Dropbox, Google Drive, iCloud Drive, and OneDrive). This will allow users to

be able to see the different aspects of each website or service and where each rise and falls in

terms of privacy and ownership as shown in Figure 2.

Figure 2: Comparison Screen
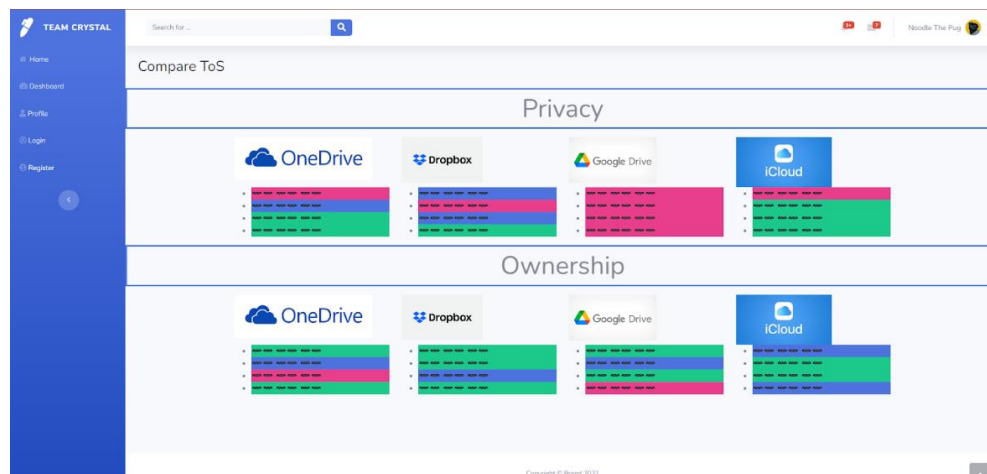
This will allow the user to come to quicker decision of whether to choose a specific service

based on the Terms of Service contract. Legal Lite will not go into different technologies within

each service.

Legal Lite hopes to gain the ability to automatically translate the legal jargon to plain

English, meaning that immediately after a user or an administrator uploads a new website Terms

of Service to Legal Lite, Legal Lite will be able to confidently publish the translation without the

intervention of an administrator for sanity checks. This will allow Legal Lite to become fully

community-based self-sufficient allowing administrators only to handle website tasks.

2 Product Description

Legal Lite is an online library that aims to educate users about Terms of Service that they

come across. Legal Lite stores user agreements and provides clear and concise versions to

significantly cut down the time it takes to become an informed user. With the constant additions

of new services and applications, not all Terms of Services are equal, and each have different end

goals when it comes to the data the users provide.**Error! Reference source not found.**

2.1 Key Product Features and Capabilities

Legal Lite aims to be a tool users will continuously use to determine what the true

intentions the company has over the users and the user's data. Legal Lite will aim to focus on

privacy and ownership of content within a Terms of Service, translate the lawyer jargon into

plain English. User accounts will not be mandatory for Legal Lite, anyone who is curious about a

website's/service's Terms of Service will be able to read and compare different services between

each other. If a user wants to be able to save/store the websites connected to their email address a

user will then need to create a user account this will give the user multiple abilities.

Once a user creates an account, a user of Legal Lite will be able to upload

website's/service's Terms of Service that may not be already on Legal Lite. Once uploaded the

user may also add their own translation if they believe Legal Lite has mistranslated or

completely missed a privacy or ownership section. As mentioned before in section 1.3 Legal Lite

will have two different ranking systems. One ranking system will be based on the grading criteria

of the administrators and the general public's ranking of whether the service itself outweighs the

possible negative implications of the Terms of Service. This will allow anyone who has never

visited the site before to understand how Legal Lite is set to work.

Legal Lite aims to be able to show users exactly how a privacy section may be broken

down. For example, Figure 3 shows a popular website Facebook and a small portion of its Terms

of Service. Legal Lite will break down each important section of the displayed portion for the

user and break down the legal jargon into plain English along with the website's privacy ranking.

This will allow users to quickly visit the site find the website or service in question and know

exactly what they are signing up for.

Figure 3: Privacy Visualized

Legal Lite will use an automated bot that will automatically scrape the internet whether it

be social media, video game launchers or e-commerce for Terms of Services, to allow a better

experience for its users. Initially Legal Lite will have to have Terms of Services uploaded

manually to its servers to get a baseline of the best ways of storing, translating, uploading, and

then ultimately updating and translating any updates to a Terms of Service to Legal Lite for users

to review.

The comparison of websites shown in Figure 3 will show the users the privacy and

ownership sections of each Terms of Service will not require users to create an account to use.

Legal Lite as explained with Figure 2 in section 1.3 will allow any user to compare multiple

websites and services within the same category to one another. For example, a user will be able

to compare Netflix, Hulu, and HBO-Max to one another but will not allow users to compare

Netflix to Amazon (The E-commerce website). Legal Lite will only list common websites when

users select the first website. Legal Lite will differentiate the multiple websites a certain

company may own for the benefit of the users who visit it.

With the ability to upload user created Terms of Service translations, users will gain the

ability to monitor each upload the user submits. With the ability to monitor user submissions this

allows users to know if the translations uploaded are correct or duplicates. This allows better

transparency between the user and Legal Lite allowing for a more trustworthy experience.

2.2 Major Components (Hardware/Software)

Legal Lite will be structured around the five major sections of Major Functional

Component Diagram in Figure 4. The web application will be utilized through a computer

(desktop/laptop MacOS/Windows) browser, mobile device (iOS/Android) or tablets (iPad

OS/Android). The databases will be used to store Terms of Services, user accounts, and

ultimately Legal Lite's translator bot which will be need access to external libraries such as Lex

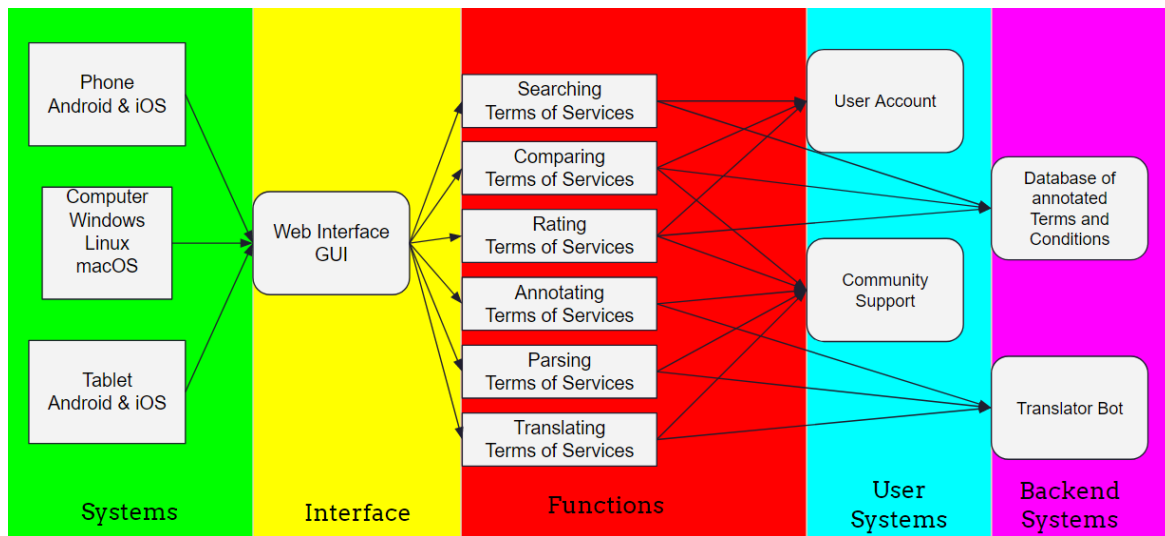NLP to translate the layer jargon quickly and easily into plain English.

Figure 4: Major Functional Component Diagram (MFCD)

The Legal Lite website itself will be a minimal interface allowing simple traversal throughout the website. Allowing users to create an account, comparison of different services, privacy rankings, community rankings, read simplified versions of a Terms of Service and if a user creates an account upload translation of a Terms of Service. Anyone looking to understand a Terms of Service will be allowed to use the website.

Legal Lite's front-end will be developed using HTML, CSS, and JavaScript for the minimal user experience. The back end will be developed using Python with the Django framework, using external libraries such as Lex NLP, Spacy, Pandas, BeautifulSoup and MatPlotLib. The databases will be managed with MySQL. Legal Lite's team will be using GitHub for source code management and VSCode for development.

3 Identification of Case Study

The intended user base of Legal Lite is for users who are trying to understand if any website or service is trying to misguide users into signing away the rights to the content users publish or the information of the users on said services. Legal Lite will not work to automatically accept or deny the Terms of Service of any website or service for a user. It will also not be able to alter the

contents of any Terms of Service to the benefit of the user. Legal Lite also will not act a lawyer

if a dispute rises between a user and a certain websites or service's Terms of Service. Legal Lite

hopes to work with those who are employed to write Terms of Service agreements to better

explain the smaller nuances of why companies choose to add certain things into their respective

Terms of Service agreements.

4 Product Prototype Description

4.1 Prototype Architecture (Hardware/Software)

4.1.1   Hardware Needed

4.1.1.1   A standard computer with internet access.

4.1.2   Software Needed

4.1.2.1   Source Code Management: GitHub

4.1.2.2   IDE: Visual Studio Code

4.1.2.3   Testing Framework: PyTest

4.1.2.4   Documentation: Pydoc/Sphinx

4.1.2.5   Testing Framework: PyTest

4.1.2.6   External Libraries: Django, Lex NLP, Spacy, Pandas, BeautifulSoup, MatPlotLib

4.1.2.7   Front End: Django, HTML, CSS, JavaScript

4.2 Prototype Features and Capabilities

Functionality/Feature	Real World Product	Prototype
Retrieve ToS/ToC	☑	☑
Automated Annotations	All Information in ToS	Privacy and Ownership Only

Site User Annotations	All Information in ToS	Privacy and Ownership Only
Website/ Service Privacy Comparisons	✓	✓
Website/Service Ranking	All Information in ToS	Privacy and Ownership Only
Retrieve/Update Out of Date ToS/ToC	Automatic Checking of WebPage	User-Initiated Only

Table 1: Features and Capabilities

4.2.1   Restrictions Planned

4.2.1.1   Annotations plan to be restricted to two sections (Privacy and Ownership Rights) to

   shorten the time spent researching user agreements and to maximize software

   development time.

4.2.1.1.1 Privacy and Ownership rights are sections of high interest from the average user.

4.2.1.1.2 User Annotations are also restricted due to research necessary to provide simplified

   attributes for every section of ToS.

4.2.1.1.2.1 User Annotation restrictions can and will be loosened based on the time needed to

   gather information needed.

4.2.1.2   Rankings plan to be restricted to two sections (Privacy and Ownership Rights).

4.2.1.2.1 Rankings can only be given for sections of ToS that are annotated.

4.2.1.2.2 As annotation restrictions are loosened, so will rankings.

4.2.1.3   Automated Checking of webpages for updates is restricted to the user-initiated flag for

   updates.

4.2.1.3.1 Unsure of the difficulty or time it takes to implement this feature.

4.2.1.3.2 Feature is a lower priority to demonstrate a prototype solution to the problem.

4.2.1.3.3 Feature will be implemented if time allows.

4.3 Prototype Developmental Challenges

4.3.1   Learning Tools Necessary for Development

4.3.1.1   Lack of familiarity with in-depth natural language processing libraries.

4.3.1.2   Lack of familiarity with Django and other webpage building tools.

4.3.2   No Expertise in Interpreting Lengthy Legal Documents

4.3.2.1   A significant amount of time will be needed in ensuring correct translations.

5 Glossary

ToS - Terms of Service

T&C - Terms and Conditions

NLP - Natural Language Processing

Annotation - Simplified, plain English summary of a portion of original TOS

Parsing - Analyzing a piece of the TOS into parts, and describing their syntactic roles

Translation - parsing an original unmodified TOS into plain English annotations

6 References

[1] Cakebread, C (2017). You're Not Alone, No One Reads Terms of Service Agreements.

BusinessInsider. https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11

[2] Hartzog, W (2018). User Agreements are Betraying You. *OneZero.*

https://onezero.medium.com/user-agreements-are-betraying-you-19db7135441f

[3] Karanicolas, M. (2021). Too Long; Didn't Read: Finding Meaning in Platform's Terms of

Service Agreements. *University of Toledo Law Review*, 52(1), 1 - 26.

http://dx.doi.org/10.2139/ssrn.3887753

[4] Lepan, N (2020) Visualizing the Length of the Fine Print, for 14 Popular Apps. VisualCapitalist.

https://www.visualcapitalist.com/terms-of-service-visualizing-the-length-of-internet-agreements/

[5] Maartje, E., Millie, E., ,Jorna, L., Marco, L., Joasia, L. (2016). Study on Consumers' Attitudes

Towards Terms and Conditions (2016). *European Commission.*

http://dx.doi.org/10.2818/950733

[6] *Terms of Service; Didn't Read*. (2012). Retrieved from https://tosdr.org

[7] *Agile as a software development process*. Software Engineering Authority. (2021, April 22).

Retrieved November 10, 2021, from https://ao.ms/agile-as-a-software-development-process/

[8] Pawtocol. "How Much Is User Data Worth?" Medium, Medium, 18 Mar. 2020,

https://pawtocol.medium.com/how-much-is-user-data-worth-f2b1b0432136#:~:text=A%20recent%20report%20estimated%20the,worth%20about%20%2435%20per%20month.

[9] Guynn, Jessica Usa Today. "What You Need to Know before Clicking 'I Agree' on That Terms

    of Service Agreement or Privacy Policy." USA TODAY, 29 Jan. 2020,

    eu.usatoday.com/story/tech/2020/01/28/not-reading-the-small-print-is-privacy-policy-

    fail/4565274002.

[10] Gilbert, B. "It's Not Just You: Terms of Service Agreements Really Are Confusing, Study

    Finds." Engadget, 17 Apr. 2014, www.engadget.com/2014-04-17-terms-of-service-study-

    georgia-tech.html.