TERMS OF SERVICE SIMPLIFIER (ToSS)

Official Statement of Hoshoku-chi
on the Document Rating and Document Evaluation
Modules of the Application

The initial proposed methodology in conducting Document Rating and Document Evaluation of the web application Terms of Service Simplifier (ToSS) was designed to have a quantitative approach with the use of category weights and thresholds. During the project presentation of Hoshoku-chi ("we" , "us" , "our" , "researchers" , "developers" , "project team") on April 17, 2021, an issue was raised regarding two modules of ToSS where the panel members noted that there is a lack of support for the measures and metrics designed in evaluating a Terms of Service (ToS) document, which still allowed the team to proceed with the development under the recommendation that additional references be cited to validate the claims of the researchers.

Before the project team provides updates on the developments of the aforementioned modules, a review of the previous approach is presented in the next sections of this communication. Some statements, tables, figures, and other necessary documents were retrieved from the initial project proposal to do this.

The researchers look forward that the following justification for the changes would be enough to consider ToSS as an application well backed-up by existing studies that also aim to assess contracts being issued by companies which touches on data collection, usage, and disposal, promoting awareness to the general public on the important roles that privacy plays in the safety and wellness of everyone.

Hoshoku-chi

STATEMENT CATEGORIZATION

Our approach defined four essential categories of terms where statements could be classified. Table 8 in the proposal identified these categories as "Collecting Data", "Using Data", "Sharing Data", and "Updating ToS Condition". However, the researchers did not include the reference for the basis of this categorization mechanics.

Sun and Xue (2020) conducted a study where they defined fundamental categories of statements that can be found in a privacy policy, which is shown below:

Table 2: Essential categories and items in privacy policies

Categories	Items
Basic information	Name of company, Address of company, Name of app, Contact info, and Date of last version.
Privacy Information Collected	Name of user, Email address, Phone number, Username/Account ID, IP address, Location information, Purchasing information, Content created by user, Information from third-parties, Device information, User activity, and Cookies.
Use of Information	For detecting abuse and illegal activity, For safety and security, For services improving, For service providing, Personalised advertising, Business analysis, Communication with users, Sharing with other users, Sharing with third-party, and Respond to legal requests.
Information Management	User rights, Data transfers, Protection of data, Data retention, Changes notifying, Business Transactions, Age. and Limits/Children's privacy

There are categories that the researchers made use of as they are from this study, and there were also those we improved by being more specific on the function of a statement. This solves the issue on the basis of the categorization. The table below shows how the researchers mapped these categories:

Existing Study	ToSS Categories
Privacy Information Collected	Collecting Data
Use of Information	Using Data
Information Management	Sharing Data
	Updating ToS

DETECTION OF UNLAWFUL STATEMENTS

Another concern was raised concerning the interpretation of the application to the extracted statements whether to be good or bad, which the panel members suggested to a.) consult other studies on their conduct of analysis or b.) consult a legal agency workforce (e.g., a lawyer) to provide insights on how to perform the process. We understand that this problem was raised because of the conclusive nature by which how the application was presented.

Section 1.3.4 of the project scope titled "Relevance of Simplified Text" discussed about the human aspect of these documents, where readers, in a general sense, still determine whether the extracted texts of the document are useful once analyzed by the system. Furthermore, the project team stated that "ToSS does not act as a replacement for ToS documents that are written in legal formats, but only as a tool to provide the abstract idea of the entire paper". That said, Hoshoku-chi does not claim that every statement extracted from the document is everything that the user needs to know, nor all of the terms identified are actually important; instead, we say that most of them are there, acknowledging the small error margin.

Furthermore, section 6.3.1 of the user classes and characteristics identified general consumers to be one of the target users, for reasons that they might just use the application for experimentation purposes, or for the simple reason of simplifying the document without regard to its analysis results. Our motivation of developing the application revolves around the idea of aiding the privacy needs of people as consumers, and not as a tool that can automate a function of any specific organization, such as the government.

The researchers still pursued the purpose statement of detecting unlawful statements, only that there is a need to change the manner by which how the results are being perceived. The proposal discussed the outcomes of this module to be binary, to classify statements whether lawful or unlawful, good or bad. This interpretation of the results sounded to be assertive and decisive, which caused uncertainty in its validity.

Which is why, we redefined the interpretation of the prediction of the application to be less determinative. Since the application is not acknowledge by any national organization that will validate the statement to be useful, we can only say at this point that there are possibilities of the extracted statements that could be beneficial or harmful to the user depending on how it is understood by the model. The following table presents the old prediction and interpretations of the predictions, followed by their restated meaning.

Old Prediction	Old Interpretation	New Prediction	New Description
Lawful	Statement is definitely within the laws of the Data Privacy Act of 2012, the legality basis for the statements.	Leaning towards good	Statement is not entirely treated as safe or friendly for the user, but there are only minimal possibilities by which the text can cause problems.
Unlawful	Statement goes outside the provisions of the Data Privacy Act of 2012, and is considered dangerous for the user.	Leaning towards bad	Statement is not entirely bad, but there are a lot of ways it can cause problems to the user due to its presentation being ambiguous, unclear, or skeptical.

For each category, we defined the semantic interpretation of the statements considered as leaning towards good or bad.

Category	Leaning Towards Good	Leaning Towards Bad
Collecting Data	Does not collect any data	Collects Personal Identifiable Information (PII)
Using Data	Does not require the processing of data	Needs the user to agree to data processing
Sharing Data	Does not share data	Involves the possibility of disclosing data to the government, third-parties, and other affiliated organizations outside the company
Updating ToS	Communicates all updates reflected in the agreement	Does not notify changes on the terms

The basis for the rules defined for each category were based on the RA 10173 or the Data Privacy Act of 2012. Chapter III Section 12 of the act discusses the lawful processing of personal information, and Section 13 for the privileged information. The researchers also checked international policies that par with the Philippine laws as an additional reference on how they treated privacy statements for evaluation.

A study was conducted to assess the privacy policies of Android diabetes apps, and how health information of patients is shared. The table below is retrieved from the paper published by S. Blenner et al., in 2016: Results show that approximately 81% of the apps with privacy policy provisions collected data, and 49% shared data even without the permission of the user. The conclusion from the study was that there are privacy risks posed for the users of the Android diabetes apps, and that there is a need for medical professionals to consider the possible implications on the privacy of the users if these apps are used. This is because at the time of the study, there were no laws that serve as protection against the inappropriate sale or disclosure of data to third parties.

A study from New Zealand also analyzed the contents of some privacy policies with the use of a questionnaire. Each question verifies whether a criterion from the privacy act of the nation is met, which is answerable by yes or no. It was found out that majority of the privacy statements failed to include the management and practices of the companies in managing companies. Tjhin et al. (2016) also identified other issues such as clauses that might allow third parties to use their information, as well as the lack of statements for notifying users when changing the policy.

There are other studies that brought up the issue on data sharing of companies to their affiliates, emphasizing that there is a lack of laws to protect privacy, and the need to reinforce the implementation of existing policies. These studies serve as a foundation in interpreting the statements per category. Since we have existing policies from the Data Privacy Act of 2012 that provide guidelines on maintaining ethical use of user data, the research team found the predefined rules ideal in determining a statement to be inclining towards good or bad.

DOCUMENT RATING

After we defined what would qualify statements to either be leaning towards good or bad, the researchers assessed the level of risks that can be implicated per category. When the risks are understood semantically, we can support the claim that high risks come from statements that allow the disclosure of data outside the company. We used this to come up with the document rating that made use of category weights. The idea was to get the percentage of statements from each category that were analyzed as statements leaning towards bad, and put weights on it. Our assignment of weights from the proposal is shown in the table below:

Criteria (Category)	Weight
Collecting Data	15%
Sharing Data	40%
Using Data	15%
Updating ToS	30%

Since this is a formula that was derived by the research team, panel members raised a question if the calculation should really be done in the manner it was presented. Other studies used a questionnaire that checks if the provisions of their respective privacy acts are complied, which made it difficult to insist on the mechanism initially presented by Hoshoku-chi. Their approaches in assessing and evaluating the documents are more syntactically done, and only uses the number of statements identified to be obscure as measures for quantifying the results. Below are some questions that were asked to assess the privacy policy documents, taken from one of the studies:

No.	Criteria	Yes	No
1.	Is the purpose of information collection outlined?	59	20
2.	Information is only collected from the individual not a third party (about the individual)?	54	35
3.	Are intended recipients listed?	10	69
4.	Is the name of collecting agency given?	3	76
5.	Is the address of the collecting agency given?	3	76
6.	Is the name of the holding agency given?	3	76
7.	Is the address of the holding agency given?	4	75
8.	Are consequences of not providing information outlined?	23	56
9.	Does the notice mention steps taken to secure transmissions between	56	23

That said, the project team disregarded the proposed formula and instead, made use of the values that can be quantified during the process, such as the total number of statements, number of ambiguous statements per category, etc. based on the rules we defined on the Detection of Unlawful Statements module. Then, we take note of the instances of the indefinite statements per category, but is presented in a general manner and not specifying the possible violation that was done (e.g., use of web cookies, sharing to third parties, disclosure to government, etc). Unlike other studies, we did not implement questions in consideration of the resources given for the development of the application, since it will require that we will expand the scope. If the developers followed the same procedures of existing literature by which the analysis of the documents are done, it would require the development of a model for every question, which will make the project infeasible for deployment if the time element is considered.

Instead, we checked if there is a presence of doubtful statements per category, which will be the basis of the document evaluation.

,

DOCUMENT EVALUATION

Originally, the values generated from the computation using the category weights from the previous module would be the basis which course of action to recommend to the user. For the final leg of an entire process cycle of ToSS, we declared value thresholds that will determine this. Retrieved from the proposal is the table that shows those limits:

Action	Maximum Score (less than or equal to)
Accept	10%
Accept with caution	20%
Reject document and warn user	21% above

Just like the other features of the application, the validity issues that come with these declarations were brought up, for reasons that a.) the recommendations were also assertive, implying that ToSS found statements definitely considered unlawful from the previous interpretation of the statements, and b.) the values were assigned by the research team without consulting other materials that could possible had used the same methodology as well. In order to improve this, we take a look on related studies to make a comparison of how they assess privacy policies.

When the researchers checked the discussion of the study on Android Diabetes Apps regarding the findings for the document, the salient points that were highlighted tackled more on the presence of certain alarming statements that were identified during the analysis phase. Unlike our approach, they did not do a conclusive recommendation whether to accept or reject the agreement, although the paper suggested to consider the use of these applications in the medicine field.

Meanwhile, the study on the privacy governance of New Zealand also concluded their study mentioning some provisions that were violated by the policies examined, and also statements that some companies lacked to include. The study also suggested ideas on how to improve the privacy policies for future revisions, like the elaboration of some statements, inclusion of process by which the company does a particular operation, etc.

In addition to the literature cited above, we found another study that assigned risk levels to the essential categories of a privacy policy statement. A study by Zaaem and Barber (2021) identified risk factors with three levels of risk. The following table shows these items, retrieved from the journal.

Privacy Factor	Green Risk Level	Yellow Risk Level	Red Risk Level
1. Email Address	Not asked for	Used for the intended service	Shared w/ third parties
2. Credit Card Number	Not asked for	Used for the intended service	Shared w/ third parties
3. Social Security Number	Not asked for	Used for the intended service	Shared w/ third parties
4. Ads and Marketing	PII not used for marketing	PII used for marketing	PII shared for marketing
5. Location	Not tracked	Used for the intended service	Shared w/ third parties
6. Collecting PII of Children	Not collected	Not mentioned	Collected
7. Sharing w/ Law Enforcement	PII not recorded	Legal docs required	Legal docs not required
8. Policy Change	Posted w/ opt out option	Posted w/o opt out option	Not posted
9. Control of Data	Edit/delete	Edit only	No edit/delete
10. Data Aggregation	Not aggregated	Aggregated w/o PII	Aggregated w/ PII

In order to make sense of the table above, we will include the other modules of the application.

First, our application classifies relevant statements in one of the four categories. As a simple recall, they are "Collecting Data", "Using Data", "Sharing Data", and "Updating ToS Condition". If we map each of these categories, we can come up with the following relationship:

Criteria (Category)	Risk Level
N/A	Green
Collecting Data	Yellow
Sharing Data	Red
Using Data	Yellow
Updating ToS	---

As for the Updating ToS, we considered it as a privacy factor similar to item 8 of the study above, which we cannot directly assign a risk level to, which we eliminated and will not be considered as a category in order to achieve two objectives: a.) align risk levels to statement categories and b.) align categories to the fundamental categories of a privacy policy from a related study. The categories will only now be the following:

Criteria (Category)
Collecting Data
Sharing Data
Using Data

The statements will be analyzed using the predefined rules when assessing whether a document is leaning towards good or bad, and if ToSS considers any statement for each category to

be ambiguous, we assigned the associated risk level for the category. A sample flow is demonstrated below:

Category	Presence of Bad Statements	Risk Level
Collecting Data	Yes	Yellow
	No	Green
Using Data	Yes	Yellow
	No	Green
Sharing Data	Yes	Red
	No	Green

The application will take the highest risk level that was associated with the privacy policy. The researchers also checked with the Data Privacy Act of 2012 if it regards the functions of these statement categories with the same risks, which Sections 12 and 13 supported. We removed the conclusive suggestions where the user is told to either accept or reject the agreement, and instead present the findings in a way that is not prescriptive, but evocative.

CLOSING STATEMENT

Hoshoku-chi hopes that this information resolved the arguments raised by the panelists, as well as other concerned individuals and group during the proposal presentation. The redesigned functionalities of the application transpired from the deliberation during the discussion between the team and the technical consultant. We presented existing studies that conducted the analysis and evaluation of the application. With that, the project team considers the issues addressed and resolved unless unrecalled concerns are identified.

REFERENCES

Blenner, Sarah R.; Köllmer, Melanie; Rouse, Adam J.; Daneshvar, Nadia; Williams, Curry; Andrews, Lori B. (2016). Privacy Policies of Android Diabetes Apps and Sharing of Health Information. JAMA, 315(10), 1051–. doi:10.1001/jama.2015.19426

Sun, R., & Xue, M. (2020). Quality Assessment of Online Automated Privacy Policy Generators: An Empirical Study. Proceedings of the Evaluation and Assessment in Software Engineering, 270–275. https://doi.org/10.1145/3383219.3383247

Tjhin, I., Vos, M., & Munaganuri, S. (2016). Privacy Governance Online: Privacy Policy Practices on New Zealand Websites. PACIS.