Construction of a Support Tool for User Reading of Privacy Policies and Assessment of its User Impact

Sachiko Kanamori[1], Hirotsune Sato[2,1], Naoya Tabata[3,1] and Ryo Nojima[1]
[1]*National Institute of Information and Communications Technology, Koganei, Tokyo, Japan*
[2]*Shinshu University, Matsumoto, Nagano, Japan*
[3]*Aichi Gakuin University, Nisshin, Aichi, Japan*

Keywords: Privacy Policy, Machine Learning, Natural Language Processing, Domain Adaptation.

Abstract: Today's service providers must notify users of their privacy policies and obtain user consent in advance. Frameworks that impose these requirements have become mandatory. Originally designed to protect user privacy, obtaining user consent in advance has become a mere formality. These problems are introduced by the gap between service providers' privacy policies, which prioritize the observance of laws and guidelines, and user expectations of these policies. In particular, users wish to easily understand how their data will be handled. To reduce this gap, we provide a tool that supports users in reading privacy policies in Japanese. We assess the effectiveness of the tool in experiments and follow-up questionnaires.

1 INTRODUCTION

It has long been considered mandatory for service providers to have frameworks for presenting privacy policies and obtaining prior consent. Posting a privacy policy has become a requirement for service providers. To ensure that users have at least looked at a policy, some service providers require that users scroll to the end of the privacy policy before they are allowed to click the consent button. Other service providers send a link to their privacy policies whenever a user navigates to their web page. Users are frequently requested to consent to the acquisition of personal data.

Presenting privacy policies and acquiring consent are intended to protect users, but they do not guarantee actual consent (McDonald and Cranor, 2008). Consent from users is now thought to be merely perfunctory; many users merely agree to the policy and use the service without reading it (Cate, 2010). Many users would rather use the service immediately than read a privacy policy (Kanamori et al., 2017).

Whereas service providers prioritize compliance with laws and guidelines, users are more interested in how their data will be handled (Reidenberg et al., 2014), (Rao et al., 2016). Privacy policies the service providers create resemble legal documents, and general users do not easily understand them (Proctor et al., 2008); they prefer information written in simple sentences.

For these reasons, signed consent has become a mere formality. Ensuring that users actually read and comprehend privacy policies will require reading assistance tools. In the White Paper 2020 (MIC, 2020), issued by the Ministry of Internal Affairs and Communications of Japan in December 2020, 89.8% of Japanese reported using the Internet, and approximately 110 million Japanese see privacy policies on a daily basis. Hence, privacy policy reading assistance tools in Japanese will be highly useful. In the present paper, we introduce our privacy policy user understanding support tool that will help Japanese users understand and recognize privacy and consent.

1.1 Our Contribution

We intend the privacy policy user understanding support tool for Japanese users who do not usually read Japanese privacy policies. To our knowledge, no such tool exists as yet. Unlike the words in English sentences, words in Japanese sentences are not separated by spaces, and this limitation demands morphological analysis by machine learning prior to data analysis. When privacy policies contain many legal terms, users can have difficulty interpreting them. We designed our tool to present users with separate unique expressions of information.

Kanamori, S., Sato, H., Tabata, N. and Nojima, R.
Construction of a Support Tool for User Reading of Privacy Policies and Assessment of its User Impact.
DOI: 10.5220/0010847500003120
In *Proceedings of the 8th International Conference on Information Systems Security and Privacy (ICISSP 2022)*, pages 412-419
ISBN: 978-989-758-553-1; ISSN: 2184-4356

1.2 Paper Organization

The remainder of the paper is organized as follows. Section 2 introduces related work, and Section 3 describes our constructed privacy policy user understanding support tool. In Sections 4 and 5, we describe how we created the machine learning training data for privacy policies and how we conducted the survey evaluation of the method, respectively. In Section 6, we analyze the results and discuss them in Section 7. Section 8 presents the conclusions and future work prospects.

2 RELATED WORK

To prevent the downgrading of consent acquisition to a mechanistic process, researchers have made many proposals (Kelley et al., 2009), (Cranor et al., 2006), which can be categorized into two groups: methods for service providers that display their privacy policies in plain language users can easily understand (Category 1,) and tools that help users understand the policies (Category 2). The tool we constructed is a Category 2 tool. In the existing literature, researchers have proposed Category 2 tools for privacy policies written in English, but to our knowledge, no scholars have developed any such tools in Japanese. We discuss Japanese privacy policies Section 2.2.1 of the present manuscript.

2.1 Proposals for Service Providers

Individual countries have begun mandating privacy policies and prior user consent acquisition by legal enforcement. Examples include the US Consumer Privacy Bill of Rights, protection regulations in Europe (EU, nd), and the revised Act on the Protection of Personal Information in Japan (PPC, nd). In Japan, the Personal Information Protection Committee supervises the handling of personal information. Under the revised Act, a business operator collecting personal information must publicly announce the purpose of use, management, and provision of that information to any third parties or pre-notify the person whose information is being collected.

However, although the system has improved, the acquisition of consent has lost its substance (Cate, 2010). Many proposals for service providers to make easily understandable privacy policies have been published (Kelley et al., 2009), (Cranor et al., 2006).

Some companies have revised their privacy policies for easier user interpretation. These revisions

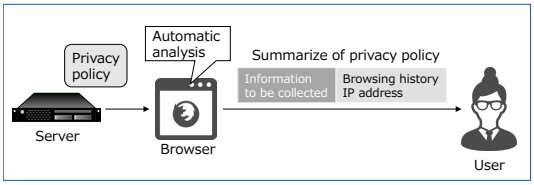

Figure 1: Overview of the Privacy Policy User Understanding Support Tool.

are sometimes accompanied by videos and illustrations. Nevertheless, because not all privacy policies are revised in this way, assistance tools are needed for users reading privacy policies.

2.2 Tools for the User

Several tools (Harkous et al., 2018), (H. Harkous and Aberer, 2016) can assist users in reading privacy policies in English. The authors of (Harkous et al., 2018) proposed a deep-learning auto-analysis framework for privacy policies known as "Polisis", and the authors of (H. Harkous and Aberer, 2016) introduced an interactive privacy bot called "Pribot" that discloses privacy policies and consent acquisition using a chatbot that mimics conversation with the user. However, both tools (Harkous et al., 2018), (H. Harkous and Aberer, 2016) interpret privacy policies in English and access the privacy policy database in (Wilson et al., 2016) for machine learning. Therefore, they are unavailable to Japanese users.

2.2.1 Tools for Reading Japanese Privacy Policies

Our proposed method extracts information from a privacy policy written in natural Japanese. Among the related studies on extraction methods, (Hasegawa et al., 2004) introduced a technology that extracts information of a predetermined event or matter from a large amount of text data such as newspaper articles and stores the extracted information in a database. Information extraction from general text uses only basic technologies such as syntactic analysis. Although the process is well researched, highly developed, and accurate, syntactic analysis loses accuracy when sentences contain many specific words (such as technical terms). Therefore, the authors of (Axelrod et al., 2011) combined a small number of specific field sentences with general sentences. For the present study, we designed privacy policy as a domain adaptation specialized for the information collected and constructed a user support tool using the training data.

3 PRIVACY POLICY USER UNDERSTANDING SUPPORT TOOL

3.1 Overview

Figure 1 shows the overall outline of our privacy policy user understanding support tool. The privacy policy displayed in the browser is automatically analyzed, and the result is presented to the user. The information collected from the user, the information used by the service providers, and the information provided to the third party are summarized and displayed.

The collected information is summarized and displyed in Figure 2[1]. The summary display screen presents the unique expressions in each privacy policy that related to a given type of information handling and a given tag.

3.2 Constructing Method

First, a user selects a privacy policy, and our constructed tool reads the policy in HTML. Then, the tool preprocesses the data, such as removing HTML tags and unnecessary characters, normalizing to prepare for Japanese natural language processing. This process requires morphological analysis to break down a sentence into its smallest units and divide it. For our study, we used MeCab (Liang, 2019): Yet Another Part-of-Speech and Morphological Analyzer, an open source program. To extract the unique expressions from the privacy policy sentences, we used long short-term memory with conditional random field (LEE, 2017).

4 CREATING MACHINE LEARNING TRAINING DATA FOR PRIVACY POLICY

To accurately extract the unique expressions from privacy policies, it is important to build precise training data. Here we describe how we created the training data for machine learning for the privacy policy user understanding support tool.

[1]Image indicates the screen capture of the tool (Summary of the privacy policy the user selected.).

Figure 2: The Privacy Policy Summary Display Screen.

4.1 Selecting Privacy Policies for Tagging

On June 8, 2017, we accessed the ALEXA TOP 500 site (Alexa Internet, 2017) and selected 94 privacy policies from the top 200 sites in Japan. We rejected the remaining 106 sites because they did not satisfy the specified privacy policy parameters[2].

Determining the Text to Be Tagged: Many sentences in privacy policies are irrelevant to what information is collected such as contact address. To obtain the appropriate information, we tagged words and phrases that indicated information handling. The 14 selected words and phrases are listed in Table 1. We selected the words following the process below:

- Among all selected privacy policies, 1,859 words appeared simultaneously with “personal information (2,504 occurrences)”.
- Among these 1,859 words, 14 related to the user information terms collect, use, and third party (Table 1), and we selected and divided these into three categories.

4.2 Select Tag Type

The tag type is required for generating the training data for sequence labeling. Many tag types are possible, as shown in the example below. Furthermore, because different people will interpret tag types differently, they must be carefully chosen.

4.2.1 Candidate of Tag Type

We collected and examined the data by comparing the tags on related projects.

IREX[3] Project: IREX (Sekine and Isahara, 2000) is a general expression extraction-tagging method.[3]

[2]The excluded sites included privacy policies in English (29 sites), Chinese (11 sites), and Spanish (21 sites) as well as policies that contained inappropriate content (32 sites) or were not downloadable (12 sites).

[3]IREX stands for Information Retrieval and Extraction Exercise.

Table 1: Selection of words and phrases for handling information.

No.	Phrase	Co-occurrence frequency	Handling category
1	Utilize	707	Use
2	Provide	476	Collect
3	Get	323	Collect
4	Collect	170	Collect
5	Use	140	Use
6	Consign	113	Third Party
7	Share	76	Third Party
8	Register	57	Collect
9	Process	53	Use
10	Hold	43	Use
11	Save	43	Use
12	Keep	37	Use
13	Maintain	36	Use
14	Transfer	28	Third Party

Scale of Privacy Information: (Sato and Tabata, 2013) defined the scale of privacy information and divided it into four categories: autobiographical information, attribute information, identification information,and password information.

Smartphone Privacy Initiative: The (SPI, 2012) committee categorized information provided by smart phone users in the following categories: user identification information, third-party information, action history of communication services, and user status.

4.2.2 Selected Tag Type

Comparing the frequency of mentions in (Sato and Tabata, 2013), (SPI, 2012) with the 94 privacy policies, we selected the following tag types:

Identification Information: Information that can identify the specific individual who requires the service.
Examples: Contractor information (name, address, etc.), payment information (credit card number, bank account number), personal identification code (ID, user name, e-mail address, password), mechanically allocated identification code (cookie, IP address, terminal identification information, application ID), information recorded in the service provider (usage history, communication history).

Personal Information: Various types of information related to individual users.
Examples: Special care-required personal information (such as race, ethnic information, health status), individual life (e.g., profile), list of friends/related parties (contact list, address book, friend list), history of online activities (search keywords, posted comments, customer reviews, current location), photos and videos (posted photos, posted videos, profile photos).

Abstract Information: a collection of multiple pieces of information (abstract words and phrases), a generic term for related phrases.
Examples: Personal information, statistical information, and personal data.

4.3 Tagging Work

Tagging was performed by six coauthors and two tagging rule-makers. The eight taggers were familiar with tagging rules and surveyed the privacy policies of many companies before selecting one. The eight taggers also repeatedly reviewed the policies and tagging rules as shown in Figure 3. The longest and shortest privacy policies were 13,481 characters and 1,413 characters in length, respectively. On average, one manual tagging required 30 minutes per 4,995 characters. Each of the eight workers tagged six privacy policies.

4.3.1 Tag Differences Caused by Differences among the Workers

Comparing the training data tagged by different workers, we found slight variations in both the length of the tagging range and the types of tags assigned.

Variation in Tagging Range: For instance, for the tagging target range, the same privacy policy statement was tagged with four ranges: “the nearest wireless network and base station for the mobile terminal,” “the nearest wireless network for the mobile terminal,” “wireless network and base station,” and “wireless network.”

Variations in Tag Types: The privacy policies were labeled with the tags described in subsection 4.2.2 (e.g., Identification Information). The selected tags varied among the workers. For example, what one worker tagged “location information as “identification information”, another might have tagged as “personal information.” Because it is possible to identify an individual by a combination of location information, it was sometimes difficult for the taggers to determine whether a location information tag was “identification information” or “personal information.” To resolve discrepancies, we conducted additional reviews among the workers.

5 SURVEY IMPLEMENTATION METHOD

We verified the effectiveness of the constructed tool in a user evaluation study in which we divided the evaluation subjects into two groups. One group was asked

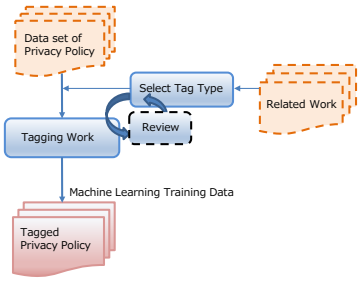

Figure 3: Tagging work procedure.

to simply read the privacy policy, and the other was given the tool we constructed. We measured the tool's effectiveness by comparing the differences between the groups.

Although the tool allows users to select a privacy policy, we created a sample policy for Internet shopping sites to ensure the same conditions for all subjects. Subjects who did not use the constructed tool were requested to read the full text of the sample policy and to answer the questions that followed the policy. The subjects who used the constructed tool were asked to display the unclear sample policy, examine the summary display the tool produced, and to answer the questions that followed.

5.1 Survey Method, Period, and Procedure

We administered the survey online to a panel of subjects who had registered with research companies. The survey included women and men aged 15 to 69 who owned personal computers. Among the 2,106 respondents whose surveys we collected, 516 said they did not wish to review privacy policies and 60 said they never shopped online. For the former group, we asked: "In this scenario, you are being asked to confirm the privacy policy on this shopping site. Do you want to confirm the privacy policy on this shopping site?" For the latter group, we asked, "Have you ever shopped on the Internet?" We then excluded respondents who answered no to either question. After we excluded these respondents' surveys, 1,530 remained for analysis: average age 44.0 years, 765 women (50.0 %), 765 men (50.0 %). For all indicators, we divided the respondents into two groups, those who did (n = 762) and those who did not (n = 768) use the privacy policy user understanding support tool. The group that did not use the tool were shown a privacy policy and asked about their subjective and objective understanding of the policy they read as well as their understanding of the service's risks. The group that could use the tool were shown an unclear privacy policy, requested the use of the tool from the monitor,

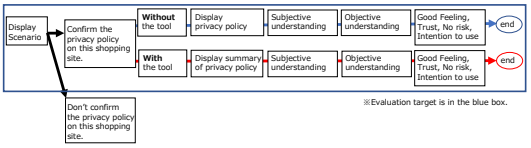

Figure 4: Survey procedures and evaluation targets.

and received a summary of the policy. They were also asked about their subjective and objective understanding and understanding of the service's risks.

Period: The web survey was administered from January 4 to 8, 2021.

Procedure: The survey procedure is shown in Figure 4. The tool is intended for users who believe it is important to read privacy policies before accepting them but find them too long or difficult to understand.

Ethics: The survey was approved by the Personal Data Handling Research and Development Council of the National Institute of Information and Communications Technology, Japan.

The research questions were as follows.

RQ1: Does the tool improve the user's understanding of privacy policies?

RQ2: Does the tool change the user's understanding of service risks?

5.2 Measuring Subjective Understanding

To measure the participants' subjective understanding, we evaluated readability, understandability, and plainness indices. Table 2 presents the question and response options for readability. The same 6-point Likert scale was used for the understandability and plainness evaluations.

5.3 Measuring Objective Understanding

To measure the degree of objective understanding, we asked three types of question about information the policy said would be collected. "Please select "Yes" for any information that you believe will be collected by the service provider under the privacy policy of this shopping site. Please select "No" for any information that you believe will not be collected". For the information to be used, the question changed to "Please select "Yes" for any information that you believe will be used by the service provider under the privacy policy of this shopping site." Same as the information to be provided to the third party. The question items are as Table 3. Five out of 10 choices provided the information to be collected, used, and provided to third parties. Each correct answer scored

Table 2: Question and answer options for subjective understanding: readability.

Question	Answer options
What did you think about the privacy policy of this shopping site? Please choose the one that you think is the closest.	1.Very difficult to read 2.Difficult to read 3.Somewhat difficult to read 4.Fairly easy to read 5.Quite easy to read 6.Very easy to read

Table 3: Answer item list for each information type.

information type	collect	used	provide
answer item list	date/time of use	name	cookie information
	cookie information	date of birth	device identifier
	IP address	mail address	location information
	location information	purchased goods	IP address
	purchased goods	credit card information	customer ID
	friend list	friend list	friend list
	face photo	face photo	face photo
	medical history	medical history	medical history
	interest	interest	interest
	voice sound	voice sound	voice sound.

Table 4: Question and answer options for service risk.

Question	answer option
When you read the privacy policy on this shopping site, were you satisfied with the service?	1.not favorable 2.rather unfavorable 3.rather favorable 4.very favorable
When you read the privacy policy on this shopping site, did you trust the service?	1.not trusted 2.rather did not trust 3.trusted 4.fully trusted
When you read the privacy policy on this shopping site, were you concerned that personal information could be leaked from this service?	1.dangerous 2.rather dangerous 3.rather safe 4.safe
After reading the privacy policy of this shopping site, would you consider using the service?	1.not at all 2.a little 3.mostly 4.certainly.

one point. We then determined the degree of objective understanding as the number of points out of 10. The first five items in each list were the correct answers, and participants who understood should have responded yes. The choices were randomly displayed to each respondent.

5.4 Measuring Understanding of Service Risk

In addition to assessing the content of the privacy policies, we evaluated service risks because this understanding can sometimes be crucial. We assumed that a better understanding of a privacy policy would alter user's positive feelings, trust, and risk aversion toward the service, and intentions to use the service. Properly understanding the risks described in a privacy policy is especially important when using a service.

We measured participants' levels of positive feelings, trust, and lack of risk degree as quantitative variables related to the presented services based on their answers to the questions in Table 4. Each item was worth 1 to 4 points.

Table 5: Subjective understanding (readability, understandability, plainness) without/with the tool AV (SD).

	without tool	with tool
readability	2.99	3.18
	(1.12)	(1.23)
understandability	3.17	3.29
	(1.03)	(1.14)
plainly	3.11	3.25
	(1.03)	(1.11)

Table 6: Objective understanding (collected information, used information, information provided to third parties) without/with the tool AV (SD).

	without tool	with tool
information to be collected	7.88	8.05
	(1.77)	(1.82)
information to be used	8.13	8.16
	(1.89)	(1.89)
information to be provided	7.16	7.38
for the third party	(1.94)	(1.98)

Table 7: Good feeling, trust, no risk, intention to use without/with tool AV(SD).

	without tool	with tool
good feeling	2.19	2.01
	(0.71)	(0.71)
Trust	2.23	2.06
	(0.71)	(0.71)
No Risk	2.09	1.88
	(0.66)	(0.67)
Will use	2.14	1.94
	(0.65)	(0.69)

6 ANALYSIS RESULTS

6.1 Subjective Understanding Results

As we noted earlier, we measured participants' subjective understanding of the privacy policies they viewed based on their ratings for the policies' readability, understandability, and plainness. Table 5 gives the results of the t test without correspondence. The readability, understandability, and plainness scores were significantly higher in the group that used the tool than in the group that did not have the tool $t(1375) = 2.56$ ($p < .01$), $t(1375) = 1.95$ ($p < .05$), and $t(1375) = 1.98$ ($p < .05$), respectively, confirming the effectiveness of the tool.

6.2 Objective Understanding Results

We measured participants' objective understanding by quantifying what they reported was the information to be collected, information to be used, and infor-

mation to be provided to third parties. Table 6 shows the results of the t test without correspondence.
The information the two groups understood would be collected and used did not differ significantly between groups $t(1375) = 1.53$ (ns) and $t(1375) = 0.10$ (ns), respectively, but the group that used the tool showed a better understanding of what information would be provided to third parties than the group without the tool $t(1375) = 1.99$ $(p < .05)$.

6.3 Service Risk Results

We measured survey respondents' positive feelings about the service, their trust in the service, their sense that there would no risk in using the service, and their intention to use the service in t tests with no correspondence. The results are given in Table 7. We found that the group that used the tool showed significantly lower ratings on positive feelings, trust, perception of risk, and intention to use the service than the group without the tool: $t(1375) = 4.98$ $(p < .01)$, $t(1375) = 4.79$ $(p < .01)$, $t(1375) = 5.78$ $(p < .01)$, and $t(1375) = 5.58$ $(p < .01)$, respectively.

7 DISCUSSION

Based on the results of our web survey on our privacy policy user understanding support tool, we answered the following research questions:

RQ1: Does the Tool Improve a User's Understanding of a Privacy Policy?

According to the subjective understanding results, the tool improved the readability, understandability, and plainness of the privacy policy. Therefore, the tool can improve the experience of reading privacy policies.

However, we could not identify any clear improvement in objective understanding, which we explain as follows. First, when summarizing and displaying the policy, the tool arranges a large number of unique expressions in tabular form that is not easy for viewers to remember in full. Second, some participants answered yes to every question about what information would be collected. Therefore, the results might reflect some users' concern that all information is collected whether or not it is related to the service.

RQ2: Does the Tool Change the User's Understanding of Service Risks?

The tool reduced the participants' positive feelings toward the service they were reading about, their trust in the service, their sense that there would be no risk in using the service, and their intention to use the service. We surmised that the tool enhanced users' recognition of the risks with their personal data, which suggests that it effectively raised users' risk awareness.

8 CONCLUSION AND FUTURE WORK

To prevent the acquisition of consent to companies' privacy policies from becoming mere formalities that carry no weight or understanding for users, we constructed a privacy policy user understanding support tool and verified its effectiveness. The tool enhanced users' subjective understanding of the services they read about and their awareness of the related risks. Service providers' privacy policies need to be easily understood by users, and providers should support this understanding. We expect that the tool we developed will help users better understand the content of the privacy policies they encounter and make decisions based on their understanding of how service providers intend to use their personal data.

One Limitation of this study is that we constructed our tool to target specifically privacy policies written in Japanese, and we conducted the survey and investigations on monitors registered with a Japanese survey company. Therefore, these results could be biased toward the Japanese context; it remains for future researchers to test whether our findings hold in other cultural contexts. Moreover, we constructed the tool solely for Japanese rather than any other languages. However, we believe that researchers who follow our procedure can construct appropriate privacy policy support tools in other languages as well. A second study limitation is that the tool is displayed on a personal computer screen, which limits the available monitors to the ones that belong to the computers' owners. Future researchers should study the tool's effectiveness with privacy policies displayed on a smart phone.

Artificial intelligence and other technological advancements will likely promote data usage in the future. Therefore, service providers should continue to implement appropriate and legally compliant measures to protect users' privacy. However, it is also important to increase users' own privacy awareness. We believe that in addition to general education about privacy and data collection, our privacy policy user understanding support tool can help in achieving that goal of greater consumer privacy awareness.

In this validation study, the privacy policy user understanding support tool enhanced the participants' understanding of the risks associated with using the

services they read about, which is one aspect of understanding privacy policies. Therefore, in future work, we must examine what types of user understanding would render a privacy policy a consensual agreement between users and the service provider. For this purpose, we should identify the factors related to users' understanding of and trust in a service by investigating the relationship between how service providers actually handle users' privacy information and how users think their data will be used. To improve user understanding, we must therefore explore additional functions and implementations of the privacy policy user understanding support tool.

ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to Mr. Yasushi Kasai and Mr. Takaomi Hayashi of the Institute of Future Engineering for their cooperation in preparing this paper. This work was supported by JST, CREST Grant Number JPMJCR21M1, Japan.

REFERENCES

Alexa Internet, I. (2017). Alexa top sites in Japan. https://www.alexa.com/topsites/countries/JP.

Axelrod, A., He, X., and Gao, J. (2011). Domain adaptation via pseudo in-domain data selection. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 355–362, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Cate, F. H. (2010). The limits of notice and choice. *IEEE Security & Privacy*, 8(2):59–62.

Cranor, L. F., Guduru, P., and Arjula, M. (2006). User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction*, 13(2):135—178.

EU (n.d.). Regulation (eu) 2016/679 of the european parliament and of the council. https://eur-lex.europa.eu/eli/reg/2016/679/oj. (Accessed on 06/08/2021).

H. Harkous, K. Fawaz, K. G. S. and Aberer, K. (2016). Pribots: Conversational privacy with chatbots. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, Denver, CO. USENIX Association.

Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K. G., and Aberer, K. (2018). Polisis: Automated analysis and presentation of privacy policies using deep learning. *Proceedings Of The 27Th Usenix Security Symposium*, pages 531–548.

Hasegawa, T., Sekine, S., and Grishman, R. (2004). Discovering relations among named entities from large corpora. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04, pages 415–es, USA. Association for Computational Linguistics.

Kanamori, S., Nojima, R., Iwai, A., Kawaguchi, K., Sato, H., Suwa, H., and Tabata, N. (2017). A study for reasons why users do not read privacy policies. In Proceedings of Computer Security Symposium 2017, pages 874-881. in Japanese.

Kelley, P. G., Bresee, J., Cranor, L. F., and Reeder, R. W. (2009). A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS 2009)*, pages 1–12, New York, NY, USA. Association for Computing Machinery.

LEE, C. (2017). Lstm-crf models for named entity recognition. *IEICE Transactions on Information and Systems*, E100.D(4):882–887.

Liang, X. (2019). Mecab usage and add user dictionary to mebab. https://towardsdatascience.com/mecab-usage-and-user-dictionary-to-mecab-9ee58966fc6. Accessed on 21/11/2021.

McDonald, M. A. and Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 4:543–568.

MIC (2020). Information and communications in Japan (white paper 2020). (Accessed on 06/08/2021).

PPC (n.d.). Personal information protection commission of Japan. https://www.ppc.go.jp/en/index.html. (Accessed on 06/08/2021).

Proctor, R. W., Ali, M. A., and Vu, K. P. L. (2008). Examining usability of web privacy policies. *International Journal of Human–Computer Interaction*, 24(3):307–328.

Rao, A., Schaub, F., Sadeh, N., Acquisti, A., and Kang, R. (2016). Expecting the unexpected: Understanding mismatched privacy expectations online. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 77–96, Denver, CO. USENIX Association.

Reidenberg, J. R., Russell, N. C., Callen, A., Quasir, S., and Norton, T. (2014). Privacy harms and the effectiveness of the notice and choice framework. *2014 TPRC Confarence Paper*.

Sato, H. and Tabata, N. (2013). Development of the multidimensional privacy scale for internet users (MPS-I). *Japanese Journal of Personality*, 21(3):312–315. in Japanese.

Sekine, S. and Isahara, H. (2000). IREX: IR & IE evaluation project in Japanese. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece. European Language Resources Association (ELRA).

SPI (2012). Smartphone privacy initiative. https://www.soumu.go.jp/menu_news/s-news/01kiban08_02000087.html.

Wilson, S., Schaub, F., Dara, A. A., Liu, F., Cherivirala, S., Leon, P. G., Andersen, M. S., Zimmeck, S., Sathyendra, K. M., Russell, N. C., Norton, T. B., Hovy, E., Reidenberg, J., and Sadeh, N. (2016). The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1330–1340, Berlin, Germany. Association for Computational Linguistics.