Better Privacy Indicators: A New Approach to Quantification of Privacy Policies

Manar Alohaly
Department of Computer Science and Engineering
University of North Texas
ManarAlohaly@my.unt.edu

Hassan Takabi
Department of Computer Science and Engineering
University of North Texas
Takabi@unt.edu

ABSTRACT

Privacy notice is the statement that contains all data practice of a particular app. Presenting privacy notice as a lengthy text has not been successful as it imposes reading fatigue. Therefore, several design proposals that substitute the classic privacy notice have been employed to different audience and in different contexts as a means to enhance user's awareness. However, there is still a shortage in having a notice display that helps users shape a coherent idea about app's data gathering practice and seamlessly allowing them to compare different application alternatives based on their data gathering practices. In this work, we propose an approach to quantify the amount of data collection of an application by analyzing its privacy policy text using natural language processing (NLP) techniques. There are in fact numerous use cases for such a quantitative measure, one of which is designing a visceral notice that relies on an experiential approach to communicate privacy information to users. The results show that our quantification approach holds promise. Using our quantification measure, we propose a new display for nano-sized visceral notice in which we leverage user's familiarity with pie chart as a data measuring tool to communicate information about an app's data collection practice.

General Terms

Human Factors, Privacy.

Keywords

Usable Privacy, Privacy Notice, Privacy Notice, Natural Language Processing

1. INTRODUCTION

Surveys have proved that users are concerned about their online privacy. Studies have shown that enhancing users' awareness about data practice over users' personal information affect their application installation behavior [1, 2]. It also plays an active role in making an informed decision about what app to use. The impact of users' awareness goes beyond the individuals to reach the market. That is a user who is well aware of policy content and its privacy implications would act as a pushing factor for apps' developers/owners to provide a good data practice and avoid the bad ones. By "data practice" we refer to what type of information an app will access, how it will be used and with whom it will be shared. Privacy notice is the statement that contains all data practice of a particular app. In this work, we will use "privacy notice", "privacy policy" and "privacy terms" interchangeably. Presenting privacy notice as a lengthy text has not been successful as it imposes reading fatigue. However, it is considered an acceptable regulatory mechanism that gives apps developers/owners the legal power to use or even abuse users' personal information. Several alternatives to the classic privacy notice have been employed to different audience and in different contexts as a means to enhance user's awareness. However, the semantic complexity of the privacy terms, the length of the text and the fact that these terms are application dependent paired with the inherent constraints in smartphone e.g. small screen size impose challenges on communicating this kind of information to a user. Despite the aforementioned challenges, privacy notice design has witnessed significant improvements in terms of the comprehensibility of privacy notice display of an individual app; yet, there is still a shortage in having a notice design that helps the users to form a coherent idea about app's data gathering practice ***and seamlessly allows them to compare different applications alternatives based on their data gathering practices***. Indeed, the hope of a notice design that perfectly derives user decision has not turned to reality yet. A user study of over 860 participants showed that 83% of participants have reported that at installation time "app features" is what really counted toward their installation decision [5]. However, when about half of the participants were asked if they would be "surprised" if the app has reached some unexpected data that was not intended for the app primary purposes, 80% confirmed that they would be surprised. This supports the fact that while users report having privacy concerns, they may not actively consider privacy while downloading apps from smartphone application marketplaces. This does not contradict by any means the fact that information contained in privacy policies are, and meant to be, relevant to the decisions users make. However, the most persistent question is how could we deliver this information in an easy to digest way? Several contributions have been made to improve the display of privacy notice in mobile apps [2, 3, 8, 9, 11]. However, each proposed design has improved the display of privacy notice of an app with an implicit assumption that a user will go several steps ahead and check the privacy notice. But in actuality the majority of the users will most likely focus on the primary task, namely completing the setup process to be able to use the system, and fail to pay attention to notices [6]. In other words, using current notice displays, users are still required to go back and forth between several alternatives, applications with similar features, to be able to

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee. *Symposium on Usable Privacy and Security (SOUPS) 2016*, June 22-24, 2016, Denver, Colorado.

compare and choose the app that offers the most conservative data practices over users' personal information. This imposes extra burden on the users' side which in turn leads to little or no actual benefit of these clearly displayed notice. In this work, we try to bridge this gap by quantifying the amount of data collection practice of an application using natural language processing. There are in fact numerous use cases for such a quantity, one of which is designing a visceral notice that leverages users' experience to seamlessly communicate privacy information. For instance, a study suggested designing a notice as eyes that appear and grow on a smartphone's home screen in proportion to how often the user's location has been accessed [13]. One can also imagine designing the notice as a pie chart where the shaded area represents the amount of collected data. ***Such a notice design not only allows the users to easily understand the notice, but also enables them to effectively compare different applications based on their data gathering practices.***

The remainder of this paper is organized as follows. In Section 2, we review related work. In Section 3, we introduce our scoring method to quantify app's data gathering practice along with our "information types" extraction method. In Section 4, we report our results and discussed limitations and lessons learned. In Section 5, we propose a use case for quantifying app's data gathering practice, and we conclude with the future work in Section 6.

2. RELATED WORK

We discuss the related work in three different categories: different proposals for better privacy design, research in using quantified information disclosure as a privacy indicator, and the use of NLP techniques for more usable privacy policies.

2.1 Privacy Notice Design

Earlier studies have researched privacy policy interfaces to improve the way in which information about app's data collection practices are delivered to a user. Kelley et al. leveraged user's familiarity with "nutrition label" to design the policy terms as nutrition label filled with just enough amount of information about data practice [3]. Kelley et al. also have shown that including privacy facts in an app's description in the app store, effectively enables users to take into account privacy considerations prior to making installation decision [2]. Reeder et al. examined the usability of the Expandable Grid interface for presenting online privacy policies [11]. Choe et al. suggested that the framing effect can be used to nudge people away from privacy invasive apps [8]. The National Telecommunications and Information Administration (NTIA) published guidelines for a short-form mobile friendly privacy notice in July 2013, aiming to supply app users with clear information about the way their personal data are collected, used and shared by apps [9].

2.2 Quantifying Information Disclosure and Notice Design

The concept of quantifying information disclosure of an application is not a newly emerging concept. Schlegel et al. have proposed a quantification model, applied on location and context sharing systems [13]. This model was based on counting the number of aggregate access requests made by *quarries* and targeted toward a *provider* within a certain time interval. The quantification measure was used to adjust the size of a visual metaphor of eyes that provides users with feedback about their exposure (i.e. the size of the eyes is relative to the number of access made by queries). We envision that having a quantified data disclosure integrated with the UI design of privacy policy will take the transparency over the utilization of users' personal information to the next level. While quantifying information disclosure is anything but a new concept, at the best of our knowledge, using NLP techniques to analyze app's privacy policy text aiming to quantify the amount of data gathering practice has not been done yet. In this paper we proposed a scoring method to quantify the data gathering practice of an app using NLP techniques.

2.3 NLP and Privacy Policy

Sadeh et al. suggested using NLP techniques in preprocessing stage of crowdsourcing to filter out the irrelevant text fragments e.g. advertisement from the core of privacy policy aiming to reduce the amount of work to be crowdsourced, and enable crowd workers to zoom in on potentially relevant text segments in a privacy policy [14]. They also suggested that the crowdsourcing results can be augmented with machine learning and NLP techniques to develop tools for automatic extraction of answers to privacy terms questions. Explore Privacy Policies website [15], originated from of Usable Privacy Policy Project at Carnegie Mellon University, leverages crowdsourcing, machine learning along with NLP techniques to semi-automatically analyze a privacy policy to extract and summarize key features from natural language website privacy policies.

3. THE PROPOSED QUANTIFICATION APPROACH

To quantify data collection practice of a particular app, we use NLP techniques to analyze its privacy policy, extract *potentially* collected "information types" or "data items", which are noun phrases associated with collection practice, and then compare the extracted data items (i.e. noun phrases) against all possible information types mentioned in Information Type Lexicon [18] aiming to identify which of these extracted noun phrases are indeed information types. The resulted subset of matching-data items are added up using a simple sum which can then be normalized by dividing to the total number of items listed in lexicon. This normalized score depicts the amount of data collection practice of an application. Our proposed quantification framework consists of four parts; analyzing the privacy policy to locate text fragments that are relevant to data collection practice, followed by the task of extracting collected data items i.e. "information types", then matching extracted items with the ones in Information Type Lexicon to find similar pairs, and finally computing collection score/rate. Figure 1 shows an overview of our framework, where squares depict the four main steps and arrows point in the direction of data flow.

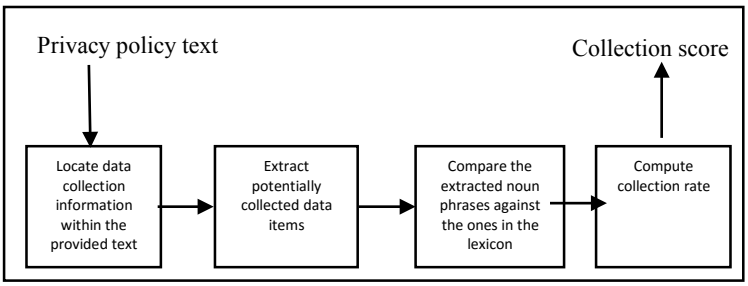

Figure 1: High Level Overview of the Proposed Quantification Approach

3.1 First Step: Locating Data Collection Practice in Privacy Policy Text

Our system takes the privacy policy of an app x as an input, searches through the text to locate the sentences that discuss data collection practices. To identify the presence of data collection practice in a particular text fragment, we use simple rule base classifier that analyzes all sentences to detect the ones that contain term "collect" or one of its synonyms. While simply searching through the text to find "collect" or its synonyms works well in identifying relevant text fragments, it does not suffice. Meaning that further text analysis is required to filter out the irrelevant passages while ensuring that each paragraph's context is kept unaffected. To demonstrate this issue with an illustrative example, we show the following text quoted from a real privacy policy.

"What information we collect

1. Personal Information

We do NOT collect any Personal Information about you. "Personal Information" means personally identifiable information, such as your name, email address, physical address, calendar entries, contact entries, files, photos, etc.

2. Non-Personal Information

We collect non-personal information about your use of our Apps and aggregated information regarding the usages of the Apps. "Non-Personal Information" means information that is of an anonymous nature, such as the type of mobile device you use, your mobile devices unique device ID, the IP address..."

As shown in above mentioned example, privacy policy explicitly specifies what kind of data/information types that the system collects and what it does not. Thus, we cannot solely rely on identifying sections that discuss data collection issues to extract the data items that are of interest of the system. Using CoreNLP [20], we partially resolve this issue by analyzing the semantic relations associated with occurrence of data collection practice in privacy policy, to filter out those that come in negative context e.g. sentences similar to "we don't collect ..."

3.2 Second Step: Extracting Potentially Collected Data Items

Using the text segments resulted from the previous stage, we extract data items that are of interest to an app x. We assume that all noun phrases that come in a "collection" context are possibly collected data items/information types. To evaluate whether or not a noun phrase is actually a data item, we compare it against the list of items in Information Type Lexicon. For instance, if privacy policy states that," *We collect email address, ip address and physical address",* then *"we", "email address*", "*ip address*" and "*physical address*" are the noun phrases that might refer to collected items. To zoom our focus on the data items and filter out other noun phrases, we compare the extracted chunk of text against the list of items in Information Type Lexicon as explained in step 3.

3.3 Third Step: Comparing the Extracted Items Against Information Types

In this work, we use Information Type Lexicon [18] to further support our analysis of privacy policies. This lexicon was constructed from 3850 annotations obtained from crowd workers who were asked to analyze 15 privacy policies. It basically contains noun phrases that describe the kind of information that is being collected, used, shared, maintained or manipulated by a system. Such information is referred to as "Information Type." Originally the lexicon contains 840 information types. After removing redundancies and some vague or general terms that are far from being collected data items e.g. "change", "third party" etc. we end up having 763 items.

To compare a noun phrase that constitutes a potential item of interest (i.e. collected item by system x) against items in Information Type Lexicon, we used WordNet similarity measures. For this purpose, we basically match each noun phrase extracted from the previous stage with all the data items listed in the lexicon. The pair that achieves the highest matching score denotes potentially similar items, if the matching score hits or goes beyond a certain threshold, which is 0.7 in our experiments. Noun phrases that satisfy the previous condition are indeed data items.

Based on empirical results, we found that a similarity score within the range of [0.5, 0.7) is obtained in two extreme cases:

1- When a pair of phrases have similar words in common, but they are semantically irrelevant.
2- When a pair of phrases are semantically close to each other but they share few or no terms in common.

We resorted to query expansion to resolve this vocabulary mismatch issue. Similarity measures along with query expansion technique that we adopted are detailed in following section.

3.3.1 Similarity Measures and Query Expansion

WordNet similarity measures can be classified into four main classes [12]: path length based measures, information content based measures, feature based measures, and hybrid measures. Path based measures express the semantic similarity as length of the path linking the underlying concepts. Information content (IC) based measures are based on the assumption that the more common information two concepts share, the more similar the concepts are. Feature based measure associates each concept with set of terms indicating its properties or features. Concepts pair with more common features/terms and less non-common features/terms are more similar. Feature measure does not work appropriately with the absence of a complete feature set. Hybrid method combines the idea of the previously mentioned measurements. Empirically, we tested the path based measures and the IC and the results were almost the same. Therefore, we used the path measure for its simplicity.

These measures are not readily available for longer text comparison as they were originally developed for measuring word to word similarity or relatedness. Therefore, we resorted to greedy-matching for sentence to sentence comparison [19], that was built upon the principle of compositionality. This principle states that the meaning of long text is determined by its constituent words

3.3.1.1 Greedy Matching

In this approach each word W_i in the first phrases P_1 is paired with every word Vi in the second one P_2 to enumerate all possible combinations. The highest score obtained by Wi determines its best match regardless of the best matching score of W_{i+1} …. W_n in P_1. The similarity score of word to word matching is added up to denote the phrase to phrase similarity measure. In greedy matching the similarity scores fall in the range [0,1], where 1 is the highest

3.3.1.2 Applying Query Expansion

The fact that we are dealing with short text fragments [16] imposes a challenge when using greedy semantic similarity to match the extracted noun phrases against the list of information types. This is

because the underlying measures rely heavily on terms occurring in both phrases. If these phrases (data item in the lexicon and data item in the privacy policy) do not have any terms in common, then they receive a relatively low similarity score, regardless of how semantically related they actually are. This is well-known as the vocabulary mismatch problem. This problem occurs if we attempt to use these measures to compute the similarity of two short text fragments. For example, the closest match of an extracted noun phrase "personally-identifying information" was "personal information" as suggested by our matching approach. While these two phrases are semantically similar, they obtained a low similarity score of 0.5. Such a score could be obtained by semantically irrelevant phrases as well. Therefore, within a certain range of similarity scores, it is desirable to generate an extended version for the short text segments that include contextually relevant information, and then compare the similarity of the extended versions of the phrases. Many techniques have been proposed to overcome the vocabulary mismatch problem, including stemming, latent semantic indexing (LSI), and query expansion [17]. Stemming is the process of reducing words to their base or root form [16]. It partially helps in resolving the vocabulary mismatch problem by using all the synonyms of the base form of words in the query to expand the query. However, it does not effectively handle the shortcomings of matching short text segments. LSI assumes that words that share similar meaning will occur in similar pieces of text [16]. This does not suit our need as we want to discriminate words that often occur in similar text segments. Thus, we resort to query expansion technique which is the best fit for our needs. It is a technique used to convert a typically short text segment into a richer representation of the information [17]. One possible external source of information related to the phrases include web (or other) search results returned by issuing the short text segment, i.e. the extracted noun phrase and possibly matching information type, as a search query. Search results provide a set of contextual text that can be used to expand the original sparse text representation.

3.4 Fourth Step: Computing Collection Rate

The resulted subset of matching pairs is added up using a simple sum which can then be normalized by dividing to the total number of items listed in Information Type Lexicon. This normalized score depicts the amount of data collection practice of an application. An obvious limitation of this approach is that we consider an item is collected if it comes in an affirmative collection context, regardless of whether or not the collection practice is conditioned upon certain attributes e.g. time, location or user's utilization pattern. A more precise quantification measure should reflect such situation. For instance, our current quantification score considers IP address is collected when privacy policy states that "we *occasionally* collect an IP address ...", while more precise quantification should weight the likelihood of this collection practice.

4. EVALUATION AND LIMITATIONS

We tested our app's data collection quantification approach on 10 different "flash light" applications. Such apps would typically not have a need to collect data, and therefore are not expected to be privacy invasive apps [5]. However, while testing, we have shown that different apps offering similar functionalities, e.g. flashlight, exhibit different data collection practices. Our data collection quantification approach has successfully captured these differences. Thus, it can be used to communicate them to the user. Table 1 shows the testing results of extracting collected data items from policy text of one flash light application, and then matching them with the appropriate information type from the Lexicon. The first column lists noun phrases that were extracted from the policy text and successfully recognized as data items using our matching approach. The second column shows different information types, from the lexicon, that were assigned as the closest match to the corresponding noun phrase in the first column. The third column presents the similarity score for each pair. The fourth lists the similarity score of the expanded version of the pair if the original version obtained a similarity score in the range [0.5, 0.7). In this case, we identified this pair of phrases as similar if the similarity score did not drop by more than 0.1 after expansion. We empirically tested these thresholds and the experiment results have shown that we were able to capture on average 68% of collected items, considering the 10 tested privacy policies. It is worth mentioning that there is no feasible loss of generality using these values, when one considers the average number of words in a noun phrase that constitute a potential data item. Meaning that applying our matching and scoring approach on an unseen policy will yield similarity scores that have three possible interpretations; a score above the threshold indicates a matching phrase, immediately below the threshold, within a range of [0.5, 0.7) as per our experiment, requires further investigation e.g. query expansion, or a low score indicates irrelevant phrases. Finally, the last column tells which of these items are actually collected by the application as resulted from our manual inspection of policy text

Table 1: The results of step 2, step 3, similarity scores before and after expansion (if needed), and whether the item is actually collected by the app.

Data Items	Information Type	Similarity Score	Similarity Score After Expansion	Collected Items
information	personal information	0.66	0.80	No
that information		0.5	0.40	
advertisements	ads	1.0	-	No
Geolocation	geolocation	1.0	-	Yes
your persistent identifiers	persistent identifiers	0.80	-	Yes
a cookie	cookie	0.66	0.61	Yes
IP	ip address	0.66	1.0	Yes
a mobile device	mobile device	0.80	-	Yes
your computer	computer data	0.5	0.40	Yes
system	operating system	0.66	0.80	Yes
application	application	1.0	-	Yes

The first three items, shown in Table 1, were the false positive phrases that match particular information types. However, these were not actually collected by the application. The false positive and false negatives are mainly attributed to two main factors; the precision of identifying the fragments of policy text that cover data collection practices and thoroughly filtering out the irrelevant text.

The second factor is the nature of the information type in the lexicon. That is the lexicon was meant to include information types that are subject to any data practice including sharing, collection, retention, etc. Thus, for our purpose, some items in the lexicon appear to be misleading e.g. "ads."

For quantification, we then add up the different information types that are captured by our approach. Table 2 shows the quantification measure for the 10 flash light application as computed by our approach, the actual amount as resulted from the manual policy review and number of false positives, respectively.

4.1 Limitations

Our current quantification approach has some limitations. We heavily rely on information types reported in the lexicon to identify a noun phrase in policy text as a collected data item, based on phrases comparison and similarity scores. Thus, our approach cannot capture items that do not have a close-match in the lexicon. Extending the lexicon will certainly improve the results of data extraction stage. In fact, our approach of extracting data based on similarity measures resolves limitations related to extracting data based on pattern [18]. Consider for example this statement from policy text "we collect *information related to browsing behavior*." The information type clause might be rewritten as "*user's browsing behavior information"*. Using similarity measures will make up for such cases where it is difficult to define common pattern. Another limitation is that we quantify data collection practice by counting the number of collected items. However, the category of the information type and the provided level of details differs among policies. For instance, one policy might mention that "we collect contact information" while other policy might say something like "we collect email address and phone number." Thus, a weighted sum based on the sensitivity, clarity or vagueness of the collected data items will be more representative.

Table 2: The number of collected items using our approach (step 4), the actual number of collected items and the number of false positives.

Application	Our Approach	Actual Number	False Positives
Flash light1	8	12	3
Flash light2	7	11	2
Flash light3	4	6	2
Flash light4	7	10	4
Flash light5	4	6	2
Flash light6	9	12	5
Flash light7	13	17	1
Flash light8	10	12	6
Flash light9	3	6	4
Flash light10	7	11	0

5. USE CASE

When smartphone users search for apps in app marketplace, applications that meet users' query appear in the search results for users to select. A study [2] has shown that by the time users select an app to proceed with the installation process, they have already made their purchase decision, without actively considering differences in data practice and privacy issues of app alternatives. Since the first thing a user will see of an app is its launcher icon, we conjecture that taking the transparency over data collection practice of an app to this early stage of "users' interaction with apps", and displaying it in a way that facilitates the comparison between different, yet comparable or possibly competitive, apps will certainly affect users' first impressions and most likely will empower the role of privacy information in decision making. To achieve this level of transparency, we propose a new notice icon that leverages users' familiarity with pie charts as a data measuring tool to build a Nano-sized notice that represents the amount of data collected by an app. The pie chart icon is divided into two sectors, possibly red and green. Red portion depicts the proportion of the data that are gathered by an app as measured by our data collection quantification approach. This icon is suggested to be displayed on top of an app's launcher icon. In fact, several pioneering studies have proposed different notice designs to encapsulate policy information in concise privacy indicators. *Privacy Grade,* for example, was designed to communicate information as of to what extent app's policy meets user's expectation. The Grades, as shown in Figure 2, are on a scale from A to D, where A denotes that app's data practice perfectly conforms with user's expectation, and the grade degrades as the gap between app's behavior and user's expectation increases [21, 22]. Indicators that are similar to *Privacy Grade* convey general concept about a policy, for example, how far a policy matches user's expectation or preference. Hence, they can be used to compare and choose among different applications. However, they only communicate information about a single concept, say user's privacy preference, and cannot be scaled to communicate information about other aspects of app's policy e.g. amount of shared and retained data. Consider, for example, a user who is interested in knowing what data an app uses and how much of this data are shared. In such a case it is better to use an indicator that is concise, yet has enough capacity to communicate information about several aspects of an app's policy.



Figure 2: Privacy grade of three flash light applications as assigned by *PrivacyGrade*

Another approach of using privacy icons is to visualize core aspects of the policy using visual metaphors of real world objects e.g. lock icon that depicts sensitive data [23]. While these indicators have shown to influence user's comprehension, they are not actively employed in app choice and installation decisions. Owing to the shortage of above mentioned indicators, we propose using a pie chart as a Nano-sized notice, as shown in Figure 3. The intuition behind this design decision is that pie chart is concise, has been known as a data measuring tool, meaning that it can be used to convey information about quantified data practices, and can be scaled to communicate information on more than one dimension e.g. not only the amount of collected data but also how much of these data are shared and retained etc. Hence we envision that it facilitates the comparison among different apps based on their data practices at the right time, namely prior to installing an app, so users can compare and choose an app that offers the most conservative data collection practice



Figure 3: Apps' launcher icons, with the pie chart that plots the amount of collected data.

6. CONCLUSION AND FUTURE WORK

In this work we mainly focused on quantifying data collection practice by analyzing the policy text. The same approach is also applicable for other data practices e.g. sharing, retention etc.

Our quantification approach consists of four phases: locating the text segments that are relevant to collection practices, extracting noun phrases that are potentially collected items, comparing the extracted noun phrases with the information types in the lexicon, using similarity measures, to filter out noun phrases that are not data items, and finally counting the number of collected items. Our experimental results show that we are able to capture on average 68% of collected items. Improving the precision of phase 1 will definitely improve the results.

Our future work will mainly focus on computing weighted sum that better reflects the vagueness and sensitivity of the collected data items, considering the main purpose of an app. We also would like to investigate how to discriminate absolute from conditioned data collection practice and how to reflect that in the scoring model. Finally, we will conduct a usability study to investigate the usability and usefulness of our notice design.

7. REFERENCES

1. Wang, N., Zhang, B., Liu, B.,& Hongxia, J.(2105). "Investigating Effects of Control and Ads Awareness on Android Users' Privacy Behaviors and Perceptions," In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services: MobileHCI '15.*

2. Kelley, P.G., Cranor, L.F., & Sadeh, N. (2013). "Privacy as part of the app decision-making process," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI '13.*

3. Kelley, P.G., Bresee, J., Cranor, L.F., & Reeder, R.W. (2009). "A "nutrition label" for privacy," In *Proceedings of the 5th Symposium on Usable Privacy and Security: SOUPS '09.*

4. RBoyles, J.L., Smith, A., & Madden, M. (2012). "Privacy and data management on mobile devices," *Pew Internet & American Life Project 4.*

5. McDonald, A.M., & Lowenthal, T. (2013). "Nano-Notice: Privacy Disclosure at a Mobile Scale," *Journal of Information Policy, 3(2013), 331-45.*

6. Schaub, F., Balebako, R., Durity, A.L., Cranor, L.F. (2015). "A Design Space for Effective Privacy Notices" In *Proceedings Symposium on Usable Privacy and Security: SOUPS.*

7. "Statistic Brain." *Statistic Brain*. Web. 11 May 2016. <http://www.statisticbrain.com/mobile-phone-app-storestatistics/>.

8. Choe, E.K., & Lowenthal, Jung, J., Lee, B., & Fisher, K. (2013). "Nudging People Away from Privacy-Invasive Mobile Apps through Visual Framing," *Human-Computer Interaction – INTERACT 2013, 8119, 74-91.*

9. Hfederman. "NTIA User Interface Mockups | Application Privacy." N.p., June-July 2013. Web. 11 May 2016. <http://www.applicationprivacy.org/2013/07/25/ntia-userinterface-mockups/>.

10. Balebako, R., Shay, R., & Cranor, L.F. (2013). "Is Your Inseam a Biometric? Evaluating the Understandability of Mobile Privacy Notice Categories"

11. Reeder, R.W., Kelley, P.G., McDonald, A.M., & Cranor, L.F. (2008). "A user study of the expandable grid applied to P3P privacy policy visualization," In *Proceedings of the 7th ACM workshop on Privacy in the electronic society: WPES '08.*

12. Lingling, M., Runqing, H., Junzhong, Gu.(2013). "A review of semantic similarity measures in wordnet." *International Journal of Hybrid Information Technology*, *6*(1), 1-12

13. Schlegel, R., Kapadia, A., & Lee, A.J. (2011). "Eyeing your exposure: quantifying and controlling information sharing for improved privacy," In *Proceedings of the Seventh Symposium on Usable Privacy and Security: SOUPS '11.*

14. Sadeh, N., Acquisti, A., Breaux, T.D., Cranor, L.F., McDonalda, A.M., Reidenbergb, J.R., Smith, N.A., Liu, F., Russellb, N.C., Schaub, F., & Wilson, S. (2013). "The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About," CMU-ISR-13-119.

15. "Explore Privacy Policies Join Our Mailing List!" *Usable Privacy*. NWeb. 11 May 2016. <https://explore.usableprivacy.org/>.

16. Metzler, D., Dumais, S., & Meek, C. (2007). "Similarity Measures for Short Segments of Text," *Advances in Information Retrieval, 4425(2013), 16-27.*

17. Lavrenko, V., & Croft, W.B. (2001). "Relevance based language models," In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval: SIGIR '01.*

18. Bhatia, J., & Breaux, T.D. (2015). "Towards an information type lexicon for privacy policies" *Requirements Engineering and Law :RELAW, IEEE Eighth International Workshop on. IEEE*

19. Rus, V., & Lintean, M. (2012, June). "A comparison of greedy and optimal assessment of natural language student input using word-to-word similarity metrics". In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP* (pp. 157-162). Association for Computational Linguistics

20. "Stanford CoreNLP." – *a Suite of Core NLP Tools*. Web. 16 May 2016<http://stanfordnlp.github.io/CoreNLP/>..

21. Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., & Zhang, J. (2012, September). “Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing”. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 501-510). ACM.

22. "Privacy Grade." *PrivacyGrade*. Web. 4 June 2016. <http://www.privacygrade.org/>.

23. Holtz, L. E., Nocun, K., & Hansen, M. (2010). “Towards displaying privacy information with icons”. In *Privacy and Identity Management for Life* (pp. 338-348). Springer Berlin Heidelberg.