A catalog of quality criteria to guide the assessment of applications' privacy policies

Augusto Terra, Jéssyka Vilela, Mariana Peixoto

Centro de Informática, Universidade Federal de Pernambuco, Recife-PE, Brazil
{aht, jffv, mmp2}@cin.ufpe.br

Abstract. Context: The information about personal data processing must be described in the companies' privacy policies. Problem: privacy policies are long documents, full of jargon, and do not always comply with the current privacy law. Besides, the privacy policies should be consistent with requirements document and application behavior, hence, it is of paramount importance that stakeholders should be able to evaluate the quality of the privacy policies. Objective: This work proposes a catalog of criteria for assessing the quality of privacy policies. Method: the snowballing technique was performed to find relevant studies that evaluate privacy policies. Results: The proposed catalog, elaborated from the empirical results of 48 studies, has 29 different criteria grouped into five categories. Contributions: The developed catalog can help: (i) requirements engineers to check the consistency of the privacy policies content with the requirements document; (ii) writers to create more precise and complete documents regarding users' rights in accordance with the requirements document, (iii) analysts and developers to make them more straightforward which information must be properly documented about the practice of data collection; (iv) end-users to understand the content of the privacy policies.

Keywords: Privacy, Privacy Policy, Personal Data, Catalog, Data protection, Quality Criteria, Snowballing,.

1 Introduction

Users are increasingly concerned about companies that collect their personal information and the risks [12] of such information being shared inappropriately [6] or accessed by unauthorized persons. Considering the several cases of personal data leaks and the creation of privacy laws, users have sought to understand some issues before using a specific product or service [8]: *How and where is the information collected from users used*? *What happens to the information collected? Is the information shared with other websites or companies*? *Does the website install any software on the user's system*?. This information about data collection and processing must be described in the companies' privacy policies as required by regulations [9].

However, privacy policies have a history of being long, complex, full of jargon, and considered difficult for end-users to understand [6][8][9]. Previous research [7] reported that many developers do not have enough knowledge about how to develop privacy-sensitive software, and studies [6][10] demonstrated that many policies lack clarity and require a reading skill considerably higher than the average literacy level of the population that uses the internet daily.

Therefore, it is necessary to understand how these privacy policies have been developed and presented to users. In this context, this work seeks to investigate the quality of privacy policies by developing a catalog with criteria for evaluating privacy policies.

A catalog is defined as an organized or classified collection of objects or information that can be grouped. Catalogs are important in academia and industry because it is possible to present with clarity, fidelity and objectivity the points to be defended [13] and provide an organized collection of the main elements of an area.

The catalog proposed in this work constitutes a body of knowledge being a robust and straightforward way of exposing data. It is helpful to support, for example, requirements engineers and analysts/developers/testers to assess the consistency between the privacy policy, requirements document, and the application behavior since the treatment performed by the application must be consistent with the one described in the policy and in the requirements document; privacy policy writers to improve the completeness of the policy; and end-users to assess whether the policy follows good practices.

The quality catalog is important because it can be used to improve the privacy policy writing process; it would allow practitioners to be aware of the information that should be presented in a privacy policy; it could decrease the time for reflection of a privacy policy content; it may help practitioners that have a lack of knowledge in the area [7] to accurately write a privacy policy.

This paper is divided into the following sections. Section 2 describes the research method used. In Section 3, we present the proposed catalog of criteria to evaluate the quality of privacy policies. Finally, Section 4 presents contributions, limitations, and future work.

2 Methodology

This work is guided by the following Research Question (RQ): *RQ: What are the criteria used to evaluate the quality of privacy policies?* The research method adopted to answer this RQ was a Systematic Literature Study [1] as a means of following the evidence-based paradigm. Moreover, the technique for collecting data was snowballing [2], and the methods for analyzing quantitative data are frequencies and thematic analysis for qualitative data.

We opted to use snowballing because previous research [11] demonstrated that the conclusions and the patterns found in conducting database searches x snowballing are quite similar. We exhaustively performed backward and forward procedures following the procedures indicated by Wohlin [2].

We performed the snowballing process with the studies presented by Graber et al. [3], Krumay and Klar [4], and Zeadally and Winkler [5]. We selected these studies considering our previous knowledge and due to their relevance to the goal of our study, the diversity of authors and publication years of the paper and the number of citations. This strategy was adopted to obtain a high coverage of distinct papers reducing selection bias. We justify the choice of these papers based on Wohlin's [2] indication.

During snowballing interaction, both backward and forward, we apply inclusion and exclusion criteria in each study to select or discard. **Inclusion criteria:** 1. Studies written in Portuguese or English; 2. Studies that apply to the topic. **Exclusion criteria:** 1. Gray literature (books, websites, theses, dissertations, etc.); 2. Studies that do not perform policy assessments; 3. Studies that do not describe the criteria used to evaluate privacy policies. We present the detailed study selection process in Figure 1.

From the snowballing, we selected 59 studies. However, we observed 11 repeated studies indicating that we achieved saturation. Hence, we selected 48 papers in which several studies presented the same evaluation criteria.

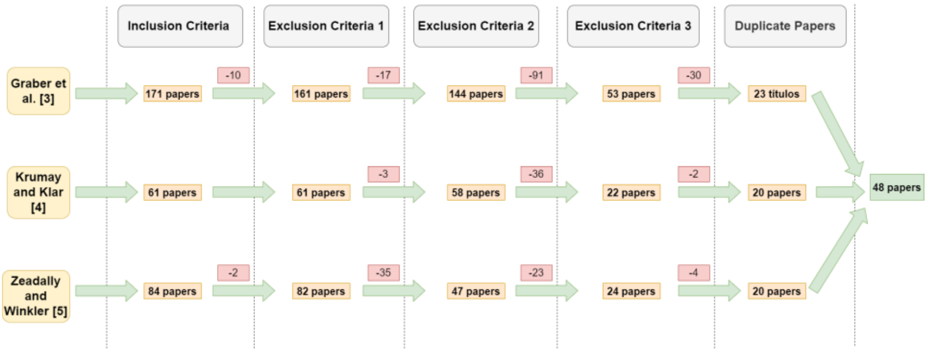

Fig. 1 Study selection process.

To reduce threats to the validity of this research, the selection criteria for initial studies proposed in [1] were used: The selected works are relevant to the scientific community, that is, published in conferences and journals recognized for their high quality; and that relate directly to the chosen theme. The initial articles also came from different scientific communities to avoid the risk of one work referencing the other, thus increasing the variety among the collected works. Finally, these articles also have a high number of citations, which confirms their relevance to the study.

3 Catalog

From the selected studies, twenty-nine criteria were extracted and grouped into five categories: **1. User Experience (7 criteria)**: Evaluates the user experience regarding accessing the document; **2. Privacy Policy Content (13 criteria)**: Evaluates the Privacy Policy in relation to its content; **3. Rights of the data subject (3 criteria)**: Evaluates if the rights of the data subject are clearly specified in the Privacy Policy; **4. Changes to the Privacy Policy (3 criteria)**: Evaluates how any changes to the Privacy Policy are handled; and, **5. User Consent and Permission (3 criteria)**: Evaluates how the user consent and permission are described in the Privacy Policy.

The criteria were grouped into these five categories considering thematic analysis based on our knowledge and information presented in the papers. We observed the criteria content and the keywords/synonyms. The developed catalog is presented in Table 2. In this table, each category is described along with its associated criteria. For traceability purposes, the article's reference in which the criteria were extracted can be found at https://www.cin.ufpe.br/~jffv/docs/TG_Augusto.pdf.

Table 2. Quality criteria, their description, and sources.

Category	**Criteria**	**Description**
User Experience	Does the application present the Privacy Policy when the user accesses the platform?	For these criteria, it is necessary to assess whether the user can access the Privacy Policy via an external link or a popup as soon as he enters the application.
	How easy is it for the user to find the Privacy Policy in the App?	The location where the link to the Privacy Policy is allocated and what visibility from him to the user.
	Is the document adequately translated to all the languages that the application supports?	The application privacy must be correctly written and translated to all languages that the application gives support.
	Is the Privacy Policy accessible to people with disabilities (PWD)?	The document text must present visual adjustments, like increasing and decreasing the text font or adjusting colors to aid users with low vision reading.
	Does the document present a readability level compatible with what the application users can read?	The Privacy Policy needs to be read and understood by users of the application; it must not be a long and tedious read. You should avoid the use of technical terms.
	Is the document responsive?	It is necessary that the Privacy Policy be available on devices with different screen sizes, whether desktop or mobile.
	Does the document present good usability in devices with different screen sizes?	The user experience of reading the Privacy Policy must be good, even on mobile devices.
Privacy Policy Content	Is the Privacy Policy based on the assumption that the visit to the application implies the user's consent to the Policy, independent of the user reading the document or not?	It is necessary that the user actively confirm that is in accordance with the practices of application data collection; it is not enough to consider that the user confirms terms of use if he only uses the application.
	Does the policy specify clearly what data is collected?	It is important that the Policy privacy detail clearly what data will be collected by the application.
	Does the Privacy Policy specify clearly how the data is collected?	The Policy needs to express clearly which tools the application used to collect data.

Category	Criteria	Description
	Does the Privacy Policy clearly specify if the application does use some tool or external service?	If any external services are used, it is necessary to have a link to the Privacy Policy of this third tool.
	Does the Privacy Policy specify how the company can use the collected data?	The Policy must indicate which purpose of collecting users' information.
	Does the Privacy Policy specify whether the information can be shared or sold to third parties?	If it involves third parties, it is necessary to describe that type of information that is shared, who the third parties are, and how the third parties can be classified and attached to the Privacy Policy of this third company.
	Does the Privacy Policy specify whether the data supply requested is voluntary or mandatory and the consequences of refusing to provide the requested information?	This criterion seeks to assess whether the Privacy Policy is flexible. that is, what happens if the user chooses not to provide certain information. For example, there are applications that use various tools of a smartphone such as microphone, GPS, and access to the list of contacts.
	Does the privacy policy specify the measures adopted by the application to ensure the confidentiality, integrity, and quality of Dice?	This criterion seeks to assess whether the application has some method to ensure the confidentiality and data integrity of the user. For example, if the data storage is encrypted or some IP mask is used.
	Does the privacy policy specify how data is stored?	They are informing how the data the company is stored spends more considerable credibility for your users.
	Does the privacy policy mentioned agree with current law?	The policy must bring explicitly if you are according to some law of privacy and indicate which law this is.
	Does the policy mention access for minors’ deity?	If the application allows access to minors of age, the privacy policy must address this topic.
	Does the policy address privacy issues related to children?	It is necessary to explain clearly how they raise questions related to privacy with children who access the application.
	Does the policy clearly explain what happens to the user's data if he deletes the account?	It is important that the policy describe what happens if the user unlinks from the application in the policy.
Rights of the data subject	Is the user free to access data about yourself even stored by application?	It is good practice to allow the user to view the data stored by the application. So, users may dispute the accuracy and the integrity of that data.

Category	Criteria	Description
	Does the privacy policy specify the user rights?	The privacy laws have rights that users have. It is good practice that policy describes these rights regarding personal data.
	Does the privacy policy report data to contact the company?	Ideally, there should be company area contact dealing with data privacy issues of your users.
Changes to the Privacy Policy	How are changes in policies handled?	After any change in the privacy policy, the users need to be informed and notified.
	What effect a change of privacy policy to an application imposes on the user?	Are users induced to read the new version of the policy of privacy? A good one practice is when the user enters the application it be redirected to a page clearly demonstrating the comparison between the versions.
	How is the frequency of modification of the policy privacy?	Constant changes in privacy policy do cause the company to lose user credibility.
User Consent and Permission	What is the method of choosing the user for consent or not with the policy of privacy?	The user must mark actively that is in accordance with the policy of Application privacy. Like in an opt-in format, for example.
	Does the user have the option of not agreeing with the applicable privacy policy?	The application must address the fact that the user possibly disagrees with the Privacy Policy.
	Is the user allowed to select what information it allows to be collected?	It is good practice that the user selects which information allows being collected and that the application handles each case the user does not accept that certain information is collected.

4 Conclusions and Future Work

This work presented a catalog of criteria for evaluating the quality of privacy policies. The construction of the catalog was performed through the application of the snowballing technique in which 48 studies were selected.

The privacy policy evaluation criteria catalog is a body of knowledge on the topic, and it brings together different views from different authors in the area. Therefore, it is expected to achieve different benefits when using the proposed catalog depending on the target audience: Privacy policy writers can base themselves on this work to create more complete and correct documents that present information regarding various aspects covered by privacy laws; Requirements Engineers and those responsible for the application also benefit from this work, as it becomes clearer for them which information should be properly documented about the data collection practice; Requirements Engineers can use it to check the consistency between the privacy policy, requirements document, and the application behaviour since the treatment performed by the application has to be consistent with the one described in the policy and in the requirements document; Users can also use the catalog created to check if the application has good practices when writing the privacy policy and answer any questions about the need to include certain information; and, The academic community can use the developed catalog to learn best practices related to the specification of personal data processing.

From the conduction of this work and the results obtained, the following directions for further work are proposed: Conduct validation of the catalog developed with privacy experts; Evaluate privacy policies of companies with the criteria described in the catalog; Expand the search for evaluation criteria for privacy policies, using other articles existing in the academic environment; and, Develop a tool for automatic evaluation of privacy policies with the criteria described in this catalog using machine learning and natural language processing techniques.

REFERENCES

1. Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. Technical report, 2007.
2. Claes Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th international conference on evaluation and assessment in software engineering. 2014. p. 1-10.
3. Mark Graber, Donna Alessandro, and Jill Johnson-West. Reading level of privacy policies on Internet health websites. In: Journal of Family Practice 51.7 (2002): 642-642.
4. Barbara Krumay and Jennifer Klar. Readability of privacy policies. In: IFIP Annual Conference on Data and Applications Security and Privacy. Springer, Cham, 2020.
5. Stephanie Winkler and Sherali Zeadally. Privacy policy analysis of popular web platforms. In: IEEE technology and society magazine 35.2 (2016): 75-85.
6. Ravi Inder Singh, Manasa Sumeeth, and James Miller. Evaluating the readability of privacy policies in mobile environments. In: International Journal of Mobile Human Computer Interaction (IJMHCI) 3.1 (2011): 55-78.
7. Mariana Peixoto, Dayse Ferreira, Mateus Cavalcanti, Carla Silva, Jéssyka Vilela, João Araújo, and Tony Gorschek. On understanding how developers perceive and interpret privacy requirements research preview. In: International Working Conference on Requirements Engineering: Foundation for Software Quality, 2020, pp. 116-123.
8. Alessandro Acquisti and Jens Grossklags. Privacy attitudes and privacy behavior. In: Economics of information security. Springer, Boston, MA, 2004. 165-178.
9. Welderufael Tesfay, Peter Hofmann, Toru Nakamura, Shinsaku Kiyomoto, Jetzabel Serna. PrivacyGuide: towards an implementation of the EU GDPR on internet privacy policy evaluation. In: Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics, 2018, pp. 15-21.
10. The New York times. We Read 150 Privacy Policies. They were an Incomprehensible Disaster. Available at: https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google privacy-policies.html. Accessed on 22/11/2021.
11. Samireh Jalali and Claes Wohlin. Systematic literature studies: database searches vs. backward snowballing. In: Proceedings of the ACM-IEEE international symposium on empirical software engineering and measurement, 2012, pp. 29-38.
12. Tobias Dienlin and Miriam Metzger. An extended privacy calculus model for SNSs: Analyzing self-disclosure and self-withdrawal in a representative US sample. In: Journal of Computer-Mediated Communication 21.5 (2016): 368-383.
13. Cristiane Rozeno Parangaba. Catálogo de dados dos trabalhos científicos de gestão ambiental e saúde da Escola Nacional de Saúde Pública Sérgio Arouca (ENSP/FIOCRUZ. In: Revista Informação na Sociedade Contemporânea 1 (2017): 1-19.