Personalised Privacy Policies

Harshvardhan J. Pandit, Declan O'Sullivan, and Dave Lewis

ADAPT Centre, Trinity College Dublin, Dublin, Ireland
{harshvardhan.pandit|declan.osullivan|dave.lewis}@adaptcentre.ie

Abstract. Internet services have become an important part of the daily life for a large number of people, and often deal with varying amounts of personal information. A privacy policy is a legal document governed by territorial laws that outlines the collection, usage, storage, and sharing of personal data. A known problem with such documents is its ambiguity and difficulty in comprehension for end users. The General Data Protection Regulation (GDPR) requires transparency regarding the provision of such information to the data subject through its various obligations and rights. We propose a remodelling of the privacy policy based on provision of relevant information regarding personal data specific to the user. Such a policy will dynamically reflect the state of activities over personal data using a legal and comprehensive document, and can be used as a tool for the provision of rights and requests from data subjects. We support our discussion with an example use-case of a GDPR-based privacy policy adopted from online services. We present our analysis on identifying changes and our approach towards the representation and creation of such dynamic policies.

Keywords: privacy policy, personalisation, GDPR, metadata

1 Introduction

The internet has become an ubiquitous part of modern daily life by providing a plethora of services and content through more than a billion websites[1]. The use of personal data on such websites and services is governed by legal obligations and must adhere to their compliance. Privacy policies act as a form of legal agreement between the service providers and their users [15], and provide information on the collection, usage, storage, and sharing of personal information. A privacy policy is expected to change or update with changes in the underlying activities and their use of personal data. Therefore, it can be considered to be a dynamic document based on changes to the underlying system which it reflects.

The problem of privacy policies being difficult to read and comprehend is well known [6,8], and has seen several efforts to remedy this [2,9,13,14]. Additionally, a privacy policy is a common document for all users of the service, and therefore contains ambiguous legal language that is broad enough to capture all possible uses of the service. It does not contain any specifics and reflects only the

[1] http://www.internetlivestats.com/total-number-of-websites/

possibility of some action over data. For example, the sentence "*we may collect your email...*" informs about the action (collect) over data (email) but does not specify whether this *will happen* or has happened already.

Under the General Data Protection Regulation (GDPR) [1], data subjects are provided the right to information about their personal data. Service providers (*Controllers*) are required to provide this information to the users (*Data Subjects*) upon request, which necessitates some technical implementation capable of recording and providing the required information. Such an implementation must be capable of distinguishing individual requests from each data subject and providing only the required information pertaining to that particular individual.

We propose to remodel the privacy policy into a personalised document for providing information specific to the data subject. Such a privacy policy would be specific to the user, and would contain information about activities and data only for a particular data subject. The information provided can be used as part of provision of GDPR rights or Subject Access Requests, or combined with a more general privacy policy to better inform users about the use of their personal data.

In this paper, we present our discussion on the creation of personalised privacy policies. We start by discussing the relevant work in Section 2. The identification and representation of dynamic metadata specific to data subjects is presented in Section 3, with the creation of a personalised privacy policy discussed in Section 4. Potential applications are discussed in Section 5. The conclusion and future work are presented in Section 6.

2 Related Work

The related work is presented in two sections. The first, Section 2.1, presents work related to the systematic studies and categorisation of privacy policies. This work is relevant towards understanding the composition of information in privacy policies, and how it can be extracted and represented. The second, Section 2.2, presents work relevant to the visualisation of information associated with GDPR rights. This work is relevant towards understanding what information is required to be presented to the user and the various approaches associated with it.

2.1 Study of Privacy Policies

There have been several studies of privacy policies across a wide range of topics from readability to summarising. Automatic categorisation of privacy policies using machine learning [2,7,9] has been shown to be effective in annotating privacy policies with context. The UsablePrivacy Project[2] has used this method to categorise sentences in a privacy policy which can be queried and used in more complex systems for better representation of information [9,15]. PrivacyGuide

[2] `https://usableprivacy.org/`

[13,14] is a similar approach that uses machine learning to summarise privacy policies. It uses a risk-based approach based on GDPR to identify relevant information, and presents it in the visual form of a dashboard.

The work described above highlights the difficulty in human comprehension of privacy policies and the applicability of machine-based techniques to convert information into a more suitable format for end-users. The work also highlights the limitations of information that can currently be extracted automatically. Both the UsablePrivacy and PrivacyGuide projects can currently identify the context of a sentence, but do not work on extracting relevant metadata from it. This is in part due to the complex nature of such sentences as well as the ambiguity in the language (see example in Section 1). More information about data purpose and its specificity [3] is required for privacy policies to better inform users. A privacy policy personalised to the data subject would ideally contain lesser ambiguity and more specitivity, which can help such efforts (both manual and automated) to better extract the relevant information.

2.2 Visualising information for GDPR rights

Considering that the privacy policy exists as a mechanism to provide information about personal data to users, other approaches with similar aims will also use the same information. One such approach describes a visualisation of the privacy policy and consent forms as a decision tree [12] with the aim to provide better information about choices made by users. PrivacyGuide [13,14] provides a dashboard that contains a visual summary of privacy policy. Other approaches exist that use icons [5] or information flow diagrams [4,7] as graphical representations of privacy policies. The UsablePrivacy project shows a visual representation of categorisation of annotations using colours [9,15].

3 Dynamic Metadata in Privacy Policies

We focus on change in information describing personal data collection, storage, usage, sharing, and deletion. At the time of writing this paper, GDPR has not yet entered into force, and few organisations have public policies related to the provision of various rights. We discuss our work and approach using privacy policies publicly provided by Airbnb Ireland[3] and Twitter [4], with archived copies made available[5] in case of changes to the policy in future. We selected these examples due to their prominence as known commercial enterprises and their suitability for purposes of this research.

3.1 Structure of Information

The analysis of these policies requires identification of what information may change or is ambiguous and could be resolved using information provided from

[3] `https://www.airbnb.ie/terms/privacy_policy`
[4] `https://twitter.com/en/privacy`
[5] `https://opengogs.adaptcentre.ie/harsh/privacy-policy-dashboard/`

resolution of GDPR rights. For this purpose, the selected examples of privacy policies have a suitable structure which is ordered into contextual sections. Such organisation of information not only helps the reader better understand and navigate information, it also helps in categorising the different types of information represented within the policy. We primarily discuss this structure in relation to the policy provided by Airbnb Ireland, though the discussion is also applicable to the policy from Twitter.

The policy follows a very structured approach towards presenting information to the user. The broad sections of the policy provide information about data collection, usage, sharing, and rights. These are further classified based on the context of activity. We focus on the first section which deals with data collection (termed "Information we collect" in the policy). The policy provides two sources of data - collected directly from the data source (section 1.1 and 1.2) and obtained from third parties (section 1.3). The information collected from the data subject is further categorised based on whether it is necessary (section 1.1.1 and 1.1.3) or opt-in (section 1.1.2). Information about the nature of the data collection mechanism is also provided, whereby some of it is collected via automated systems (section 1.2). This structure is presented visually in Fig. 1.

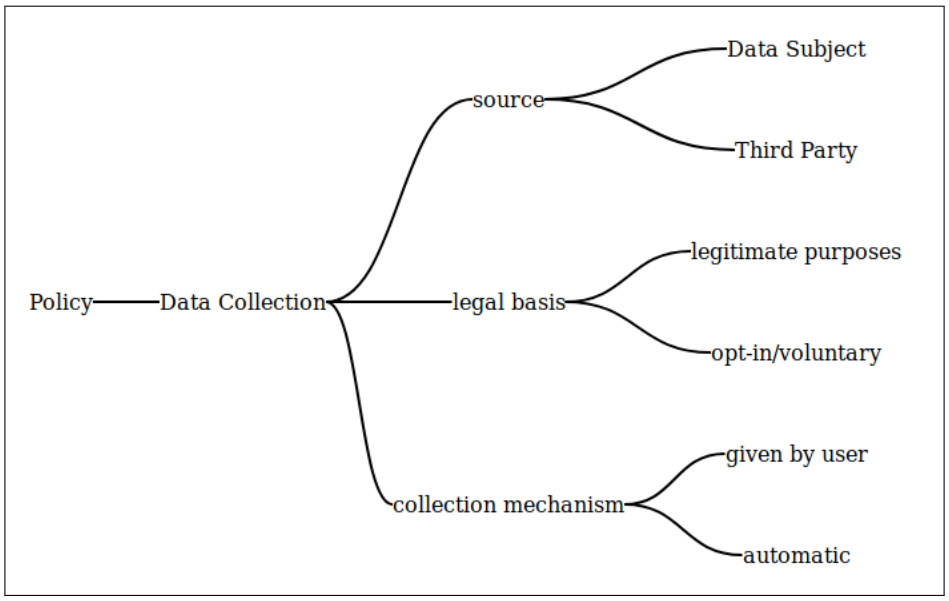

Fig. 1. Structuring of information related to data collection in policy

While the above information reflects the structure of the policy, the contents within each section provide information about the personal data involved. For example, the information in section 1.1.1 describes the categories of data involved. Each category is further described with the specific types of data that

fall under it. For example, *Account Information* is the first category within the section, which contains information about data types such as *first name, last name, email address*, and *date of birth*. Additionally, the sentence also mentions the specific process (*account sign-up*) used to collect this information.

This information is distinct from the earlier structuring of information in that it can change (is dynamic) based on the operation and provision of services. For example, it is possible that additional information such as nationality may be added as essential account information in the future. In such a case, it will be listed along with the other data types under the "*Account Information*" data category. Similarly, the mechanism for data collection may change as well to some other new or existing process or step.

3.2 Annotation Metadata

We distinguish between metadata representing the 'structure' of information and the representation of the underlying system. While the former will be common to all services and policies, the latter reflects information specific to organisation or service (and to the data subject). From the example, all privacy policies will have a section for describing the data categories, but the specific categories mentioned within the policy are unique and associated with the organisation and service it provides, and is updated based on changes to the system and operations. We term such information as 'dynamic metadata' to reflect this.

Based on this definition, we annotated the example privacy policy to visualise the different types of metadata, as presented in Fig. 2. The annotations are highlighted with different colours based on the context of the metadata, as shown in the legend in the picture. The figure reflects only a part of the annotated policy, which is available online[6]. The colours serve to visually represent the dynamic metadata, and help in understanding the different types of information and their context throughout the policy. This visual distinction of information is presented to view the different contexts within the privacy policy identified in our analysis. This follow a similar approach from the UsablePrivacy project [9] which use different colours to visually highlight the different types of information.

4 Implementation Approach

In this section we describe our approach towards the implementation of a personalised privacy policy using the dynamic metadata. The approach provides a general overview of how such policies can be implemented, and can be adopted to any set of technologies in practice. The first section, Section 4.1, describes a common template for a privacy policy which is then personalised using dynamic metadata specific to the user. The second section, Section 4.2, presents our approach towards representation and generation of dynamic metadata in policies.

[6] `https://openscience.adaptcentre.ie/projects/privacy-policy/personalise/`

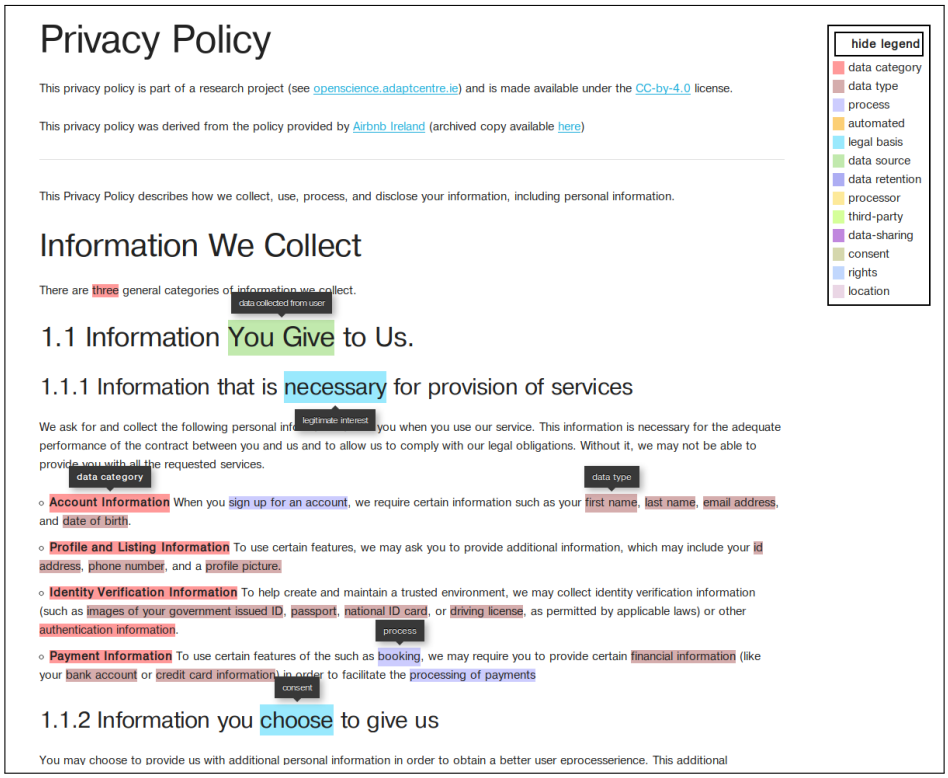

Fig. 2. Visualising annotations for dynamic metadata in privacy policy

4.1 Privacy Policy Template

The privacy policy itself can be represented as a template with information that is common to all policies being considered as static text, and that which is specific to the underlying processes or the data subject being considered as dynamic text. Based on this template, the information for the data subject or user is then used to populate and present a personalised privacy policy to the user. This approach allows reusing a common privacy policy layout and some part of the overall text for all users. In addition, it also allows personalisation of information specific to certain user-cases such as in case of minors or information pertaining to exercised rights.

Existing privacy policies (including the specified example) are monolithic documents composed primarily of text. They do not have any metadata that can describe the content or its context. We take this opportunity for remodelling the privacy policy to also annotate its contents with metadata that can assist in its interpretation and use in other tools and services.

Providing contextual information about the dynamic metadata can assist other systems and tools to interpret the results in order to assist the user, or for

research purposes. Since the privacy policy is inevitably served as a web-based document, the metadata too must be served in a compatible format such as Microdata[7] or RDFa[8] along with a suitable vocabulary such as schema.org[9]. Such formats and vocabularies must essentially be open in nature to foster interoperability.

The underlying contextual information about the dynamic metadata is largely abstracted by the displayed privacy policy as the user does not see or interact with it. However, it is of consequence to the organisation as they are required to maintain and provide it. It is therefore beneficial to store this information along with the relevant metadata in a form that assists in the creation of dynamic privacy policy. Such information is also required and is useful for compliance purposes as well as the provision of various rights. All of these can benefit from a structured method for representation of associated information along with the involved metadata. The ability of the underlying technology for expressing queries provides a means to efficiently retrieve information in a structured and relevant format.

4.2 Storage and Representation of Metadata

For our work, we focus on the use of semantic web technologies due to their open and extensible nature. For representing the metadata related to processes and the data they use, we use the GDPRov ontology [11] which extends PROV-O[10] and P-Plan[11]. PROV-O is a W3C recommendation, which provides interoperability of provenance information. P-Plan is an extension of PROV-O that allows representation of abstract workflows. For annotating information with concepts and terms from the GDPR, we use the GDPRtEXT resource [10].

Representing metadata using structured vocabularies such as GDPRov and GDPRtEXT allows querying for required information as well as annotating the policy with relevant metadata. We present our preliminary work towards using these to annotate privacy polices using RDFa. An example of this is presented in Listing 1, which shows the possible RDFa annotation of a personalised privacy policy continued from the previous example. The policy is provided as a HTML page and describes the data type "*first-name*" within the "*AccountInformation*" data category. *AccountInformation* is a defined as a subclass of *PersonalData* declared in GDPRov, and *first-name* is an instance of this class. This information stands to informs the data subject that their *first name* is being collected as part of the *Account Information*. The data category as well as data type is an example of dynamic metadata used to personalise the privacy policy for the data subject.

We describe here a more detailed technical description of the implementation of this system to demonstrate the particular use-case.

[7] `https://www.w3.org/TR/microdata/`
[8] `https://www.w3.org/TR/rdfa-primer/`
[9] `http://schema.org/`
[10] `http://www.w3.org/TR/prov-o/`
[11] `http://purl.org/net/p-plan`

```
<body
    vocab="http://example.com/use-case"
    prefix="gdprov:
        http://purl.org/adaptcentre/openscience/ontologies/gdprov#
        rdfs: http://www.w3.org/2000/01/rdf-schema#">
    <p resource="#AccountInfo">
        <span property="rdfs:label">Account Information</span></p>
    <ul>
        <li><label
            resource="#first-name"
            typeof="gdprov:PersonalData #AccountInfo">
            <span property="rdfs:label">First Name</span>
        </label></li>
    </ul>
</body>
```

Listing 1: Policy metadata described using GDPRov in RDFa

1. The model of the system is defined using GDPRov and GDPRtEXT to represent activities and how they interact with personal data. This is stored as RDF data in a triple store.
2. This is followed by the creation of a privacy policy template using a templating engine such as Jinja[12] that allows programmatically populating it with dynamic metadata.
3. As data subjects or users use and interact with the system, relevant metadata is stored using GDPRov in the triple store as RDF data.
4. When data subjects request to view a personalised privacy policy or exercise their right to retrieve information, the relevant data is retrieved using queries modelled using SPARQL[13].
5. The results are then used to populate the policy template to create a personalised privacy policy or information report that is annotated with RDFa.

5 Potential Applications

The work described in this paper has broader applications apart from personalising privacy policies such as addressing various rights and access requests (such as for GDPR) and to automate other similarly structured documents.

[12] http://jinja.pocoo.org/
[13] https://www.w3.org/TR/sparql11-query/

Address GDPR Rights and SARs

GDPR provides the data subjects with several rights through which an organisation is required to provide information about their activities over personal data. This can necessitate the creation of new technical measures to handle requests and to provide this information in a legally acceptable way. The use of a personalised privacy policy document can aid in the provision of this information as it uses legally relevant language and outlines the use of personal data in a structured way. Similarly, a Subject Access Report (SAR) can be created from the same mechanism used to implement the personalised privacy policy, as it largely operates on the same information.

Automate Reports and Documentation

Documentation related to compliance and other processes is often structured and refers to information in a specific way. A similar approach as the one described in this paper where stored metadata is used to dynamically populate a structured document can be used to automate this process. This can be used for generating reports that describe the various processes and how they relate with personal data based on the underlying model of the system. It can also be extended to create various technical reports regarding the use of internal processes.

6 Conclusion & Future Work

Through this paper, we presented our work on a personalised privacy policy that provides specific information about a data subject's personal data. We presented our analysis of existing real-world policies where we identified the structure of information and the dynamic metadata based on changes to the underlying system as well as the specific data subject. We presented our approach towards the representation of this metadata using a common and open format, and described our work towards creating such personalised policies using semantic web technologies. We also discussed how this work can be used as a tool for the provision of rights and requests from data subjects, with potential applications in similarly structured documents.

The primary future work is the implementation of such a personalised policy using the approaches and technologies described in this paper. With the advent of GDPR, we expect to see more examples of similarly structured privacy policies, which will need to be analysed to identify relevant metadata. This also presents an opportunity to assess the information provided by various organisations as part of the various rights and SARs; and to modify this work to better reflect real-world use-cases.

While the work presented in this paper presents the motivation for a personalised privacy policy, the generic privacy policy that is shown to all users must also be preserved to be displayed before any data or processes have been executed. Therefore, in effect, the organisation will have two privacy policies - one

generic and the other personalised, that will contain largely similar structures and metadata regarding the processes and data used. More work needs to be undertaken to distinguish the similarities between the two to take advantage of the similar structure and to also possibly generate such generic privacy policies in an automated manner.

Acknowledgements

This work is supported by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union **L119**, 1–88 (May 2016), http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC
2. Ammar, W., Wilson, S., Sadeh, N., Smith, N.A.: Automatic categorization of privacy policies: A pilot study (2012), http://repository.cmu.edu/lti/199/
3. Bhatia, J., Breaux, T.D.: A Data Purpose Case Study of Privacy Policies. In: Requirements Engineering Conference (RE), 2017 IEEE 25th International. pp. 394–399. IEEE (2017)
4. Bier, C., Kühne, K., Beyerer, J.: PrivacyInsight: The Next Generation Privacy Dashboard. In: Privacy Technologies and Policy. pp. 135–152. Lecture Notes in Computer Science, Springer, Cham (Sep 2016). https://doi.org/10.1007/978-3-319-44760-5_9, https://link.springer.com/chapter/10.1007/978-3-319-44760-5_9
5. Esayas, S., Mahler, T., McGillivray, K.: Is a Picture Worth a Thousand Terms? Visualising Contract Terms and Data Protection Requirements for Cloud Computing Users. In: Current Trends in Web Engineering. pp. 39–56. Lecture Notes in Computer Science, Springer, Cham (Jun 2016). https://doi.org/10.1007/978-3-319-46963-8_4, https://link.springer.com/chapter/10.1007/978-3-319-46963-8_4
6. Fabian, B., Ermakova, T., Lentz, T.: Large-scale Readability Analysis of Privacy Policies. In: Proceedings of the International Conference on Web Intelligence. pp. 18–25. WI '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3106426.3106427, http://doi.acm.org/10.1145/3106426.3106427
7. Fawaz, H.H.K., Schaub, R.L.F., Karl, K.G.S.: Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. Tech. rep., EPFL (2017), https://pribot.org/files/Polisis_Technical_Report.pdf
8. Jensen, C., Potts, C.: Privacy Policies As Decision-making Tools: An Evaluation of Online Privacy Notices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 471–478. CHI '04, ACM, New York, NY, USA (2004). https://doi.org/10.1145/985692.985752, http://doi.acm.org/10.1145/985692.985752

9. Oltramari, A., Piraviperumal, D., Schaub, F., Wilson, S., Cherivirala, S., Norton, T., Russell, N., Story, P., Reidenberg, J., Sadeh, N.: PrivOnto: A semantic framework for the analysis of privacy policies. Semantic Web **9**(2), 185–203 (Jan 2018). https://doi.org/10.3233/SW-170283, http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-170283
10. Pandit, H.J., Fatema, K., O'Sullivan, D., Lewis, D.: GDPRtEXT - GDPR as a Linked Data Resource. p. 14. Heraklion, Crete, Greece (2018)
11. Pandit, H.J., Lewis, D.: Modelling Provenance for GDPR Compliance using Linked Open Data Vocabularies. In: Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn) (2017), http://ceur-ws.org/Vol-1951/#paper-06
12. Rossi, A., Palmirani, M.: A Visualization Approach for Adaptive Consent in the European Data Protection Framework. In: 2017 Conference for E-Democracy and Open Government (CeDEM). pp. 159–170 (May 2017). https://doi.org/10.1109/CeDEM.2017.23
13. Tesfay, W.B., Hofmann, P., Nakamura, T., Kiyomoto, S., Serna, J.: I Read but Don'T Agree: Privacy Policy Benchmarking Using Machine Learning and the EU GDPR. In: Companion Proceedings of the The Web Conference 2018. pp. 163–166. WWW '18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2018). https://doi.org/10.1145/3184558.3186969, https://doi.org/10.1145/3184558.3186969
14. Tesfay, W.B., Hofmann, P., Nakamura, T., Kiyomoto, S., Serna, J.: 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. pp. 15–21. IWSPA '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3180445.3180447, http://doi.acm.org/10.1145/3180445.3180447
15. Wilson, S., Schaub, F., Dara, A.A., Liu, F., Cherivirala, S., Giovanni Leon, P., Schaarup Andersen, M., Zimmeck, S., Sathyendra, K.M., Russell, N.C., B. Norton, T., Hovy, E., Reidenberg, J., Sadeh, N.: 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). pp. 1330–1340. Association for Computational Linguistics, Berlin, Germany (Aug 2016), http://www.aclweb.org/anthology/P16-1126