

# Supplementary Material for Unsupervised Alignment of Privacy Policies using Hidden Markov Models

Rohan Ramanath Fei Liu Norman Sadeh Noah A. Smith

School of Computer Science

Carnegie Mellon University

Pittsburgh, PA 15213, USA

{rrohan, feiliu, sadeh, nasmith}@cs.cmu.edu

## 1 Human QA

Below are the nine multiple choice questions that were presented to the expert annotators.

1. Does the policy state that the website might **collect contact information** about its users?
  - (a) **No** - the policy states that the website will not collect contact information.
  - (b) **Yes** - the policy states that the website might collect contact information.
  - (c) **Unclear** - the policy makes statements that could mean the website might collect contact information but is not clear about it.
  - (d) **Not applicable** - the policy does not appear to address the question.
2. Does the policy state that the website might **collect current location** of a user when he/she is using their service?
  - (a) **No** - the policy states that the website will not collect current location information of its user.
  - (b) **Yes** - the policy states that the website might collect current location information of its user.
  - (c) **Unclear** - the policy makes statements that could mean the website might collect current location but is not clear about it.
  - (d) **Not applicable** - the policy does not appear to address the question.
3. Does the policy state that the website might **collect information relating to health** of its users. Health Information ?
  - (a) **No** - the policy states that the website will not collect health information.
  - (b) **Yes** - the policy states that the website might collect health information.
  - (c) **Unclear** - the policy makes statements that could mean the website might collect health information but is not clear about it.
  - (d) **Not applicable** - the policy does not appear to address the question.
4. Does the policy state that the website might **collect financial information** about its users?
  - (a) **No** - the policy states that the website will not collect financial information.
  - (b) **Yes** - the policy states that the website might collect financial information.
  - (c) **Unclear** - the policy makes statements that could mean the website might collect financial information but is not clear about it.
  - (d) **Not applicable** - the policy does not appear to address the question.
5. Does the policy state that the website might **share contact information** with third parties?
  - (a) **No sharing** the policy states that the website will not share contact information with third parties.
  - (b) **Sharing for core service only** - the policy states that the website might share contact information with third parties, but only for the purpose of providing a core service.
  - (c) **Sharing for other purpose (with consent)** - the policy states that the website might share contact information with third parties for other purposes, but only if the website has explicit permission/consent from the user.
  - (d) **Sharing for other purpose (without consent)** - the policy states that the website might share contact information with third parties for other purposes and might do so without explicit permission/consent from the user.
  - (e) **Unclear** - the policy makes statements that could mean the website might share contact information with third parties but is not clear about it.
  - (f) **Not applicable** - the policy does not appear to address the question.
6. Does the policy state that the website might **share current location information** with third parties?
  - (a) **No sharing** the policy states that the web-

- site will not share current location information with third parties.
- (b) **Sharing for core service only** the policy states that the website might share current location information with third parties, but only for the purpose of providing a core service.
  - (c) **Sharing for other purpose (with consent)** - the policy states that the website might share current location information with third parties for other purposes, but only if the website has explicit permission/consent from the user.
  - (d) **Sharing for other purpose (without consent)** - the policy states that the website might share current location information with third parties for other purposes and might do so without explicit permission/consent from the user.
  - (e) **Unclear** - the policy makes statements that could mean the website might share current location information with third parties but is not clear about it.
  - (f) **Not applicable** - the policy does not appear to address the question.
7. Will the website **share health information** with third parties?
- (a) **No sharing** the policy states that the website will not share health information with third parties.
  - (b) **Sharing for core service only** the policy states that the website might share health information with third parties, but only for the purpose of providing a core service.
  - (c) **Sharing for other purpose (with consent)** - the policy states that the website might share health information with third parties for other purposes, but only if the website has explicit permission/consent from the user.
  - (d) **Sharing for other purpose (without consent)** - the policy states that the website might share health information with third parties for other purposes and might do so without explicit permission/consent from the user.
  - (e) **Unclear** - the policy makes statements that could mean the website might share health information with third parties but is not clear about it.
  - (f) **Not applicable** - the policy does not appear to address the question.
8. Will the website **share financial information** with third parties?
- (a) **No sharing** the policy states that the website will not share financial information with third parties.
  - (b) **Sharing for core service only** the policy states that the website might share financial information with third parties, but only for the purpose of providing a core service.
  - (c) **Sharing for other purpose (with consent)** - the policy states that the website might share financial information with third parties for other purposes, but only if the website has explicit permission/consent from the user.
  - (d) **Sharing for other purpose (without consent)** - the policy states that the website might share financial information with third parties for other purposes and might do so without explicit permission/consent from the user.
  - (e) **Unclear** - the policy makes statements that could mean the website might share financial information with third parties but is not clear about it.
  - (f) **Not applicable** - the policy does not appear to address the question.
9. What is the website's policy about letting its user **delete their personal data**?
- (a) **No removal** - the policy states that the user will not be allowed to delete their personal data.
  - (b) **Full removal** - the policy states that users may delete their personal data and that no data will be retained for any purpose.
  - (c) **Retention for legal purposes** - the policy states that users may delete their personal data but some/all of the data might be retained for legal purposes only.
  - (d) **Retention for other purposes** - the policy states that users will be allowed to delete their personal data but some/all of the data might be retained for purposes other than legal obligations.
  - (e) **Unclear** - the policy makes statements that could mean that users may delete their personal data but is not clear about it.
  - (f) **Not applicable** - the policy does not appear to address the question.

## 2 Example Answers

The first question in §1 was answered by five annotators for each of thirty policies. As an

example, we consider the privacy policy of `huffingtonpost.com`, for which all five annotators selected “**Yes** - the policy states that the website might collect contact information” as the answer. The sections from the privacy policy that were highlighted by the annotators are as follows.

*and mobile number;*

- A<sub>1</sub> However, when you use certain Site functions, such as when you register for certain services, access certain content or features, or directly contact the Site, we may ask you to provide information, including contact information, such as name, email address, postal address, phone number, and mobile number*
- A<sub>2</sub> You are not required to provide information directly to us in order to view the Site. However, when you use certain Site functions, such as when you register for certain services, access certain content or features, or directly contact the Site, we may ask you to provide information, including contact information, such as name, email address, postal address, phone number, and mobile number;*
- A<sub>3</sub> Information you provide directly to us: You are not required to provide information directly to us in order to view the Site. However, when you use certain Site functions, such as when you register for certain services, access certain content or features, or directly contact the Site, we may ask you to provide information, including contact information, such as name, email address, postal address, phone number, and mobile number; username and password; information posted in community discussions and other interactive online features; search queries conducted on the Site; and correspondence you send to us*
- A<sub>4</sub> You are not required to provide information directly to us in order to view the Site. However, when you use certain Site functions, such as when you register for certain services, access certain content or features, or directly contact the Site, we may ask you to provide information, including contact information, such as name, email address, postal address, phone number, and mobile number; username and password;*
- A<sub>5</sub> when you use certain Site functions, such as when you register for certain services, access certain content or features, or directly contact the Site, we may ask you to provide information, including contact information, such as name, email address, postal address, phone number,*