Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning

Hamza Harkous (EPFL)
Florian Schaub (U. Michigan)
Kassem Fawaz (U. Wisconsin)
Kang G. Shin (U. Michigan)
Rémi Lebret (EPFL)
Karl Aberer (EPFL)

pribot.org
hamzaharkous.com
Problem?
PRIVACY POLICIES ARE LONG AND COMPLEX

201 hours per year on average to read policies of services we encounter*

* (McDonald and Cranor -2008)
APPROACHES SO FAR?

Put more lawyers on the task.
Standardization

- A Nutrition Label for privacy
- Required providers to act
- **Surprise**: They didn’t.

Kelley et al., "A nutrition label for privacy." SOUPS’09
Crowdsourcing

- TOSDR.org
- Limited by volunteers’ availability
- Available for ~100 policies
- Unstructured → can only be used for limited automated labeling*

Manual work doesn’t scale.

Fails to cope with emerging technologies.
Voice-Activated Devices

- Read the whole policy?
- Show tables on small screens?
Regulation Compliance (e.g. GDPR)

Find Statements About Health Data Sharing

Get Segments such that

Category: third party sharing
d

personal information type: health information
Solution
Once we automate policies’ analysis, we can create a new interface for millions of policies with a single program.

- **Structured Queries**
- **Unstructured Queries**
Framework

Application Layer
- 1. privacy policy link

Data Layer
- Segmenter

ML Layer
- 2. policy segments

Query Module
- 1. user query

Class Comparison
- 3. query classes
- 4. classified segments

Query Classifier
- (ML Classifiers)

Segment Classifier
- 2. policy segments

Framework with ML Classifiers
Prepend the intro:

You don't need to register for a SoundCloud account in order to visit or use our Platform. However, certain Services require that you register for an account and, by doing so, you will need to provide us with certain information:

- **Account Information** - When you register any kind of SoundCloud account, you will need to provide your email address (except when you sign-up with Facebook or Google and don't give your permission to share it with SoundCloud), your age, and choose a password. In addition, if you purchase a Subscription, you will also need to provide your real name, billing address, and payment verification information.

- **Profile Information** - you may choose, at your discretion, to provide additional information for your public profile on SoundCloud® - for example:
  - your real name
  - a user name (which may be your real name or a pseudonym)
  - the city and country in which you live

Merge short lists

### HTML Preprocessing → Coarse Segmentation according to HTML Tags → Text Segmentation into semantically coherent segments*

*Glavas et al., "Unsupervised Text Segmentation Using Semantic Relatedness Graphs", ACL 2016*
Application Layer

Query Module

Class Comparison

Data Layer

Segmenter

ML Layer

Query Analyzer

Segment Classifier

1. user query
2. policy segments
3. query classes
4. classified segments

privacy policy link

1. policy segments

App

1. user query

Class Comparison

User query is classified into segments and policy segments are extracted. The policy segments are then analyzed to retrieve the privacy policy link.
We may need to retain certain information for recordkeeping purposes, as required under applicable legal obligations, and/or to complete any transactions that you began prior to requesting such change or deletion (...) Some of your information may remain within our systems and other records, in compliance with applicable law.
Online Privacy Policies Dataset

- 115 annotated policies
- 23K annotations
Hierarchical Data → Hierarchical Architecture
Hierarchical Architecture

Category Labels
- 1st Party Collection
- 3rd Party Sharing

Value-level Classifiers
- Info Type Classifier
- Purpose Classifier
- Collection Mode Classifier

Value Labels
- Financial
- Location
- Marketing
- Legal
- On-website
- From 3rd Party
Hierarchical Architecture

Similar architecture for the 21 classifiers

130,000 privacy policies from Play Store to train our custom word embeddings

Loss function: multi-label cross entropy

Embeddings size: 300, Number of filters: 200, Filter Size: 3, Dense Layer Size: 100, Batch Size: 40

Embeddings size:

19
Structured Querying
Privacy Icon Assignment as a Case Study
Automated Privacy Icons Assignment

TRUSTe & Disconnect Introduce Visual Icons to Help Consumers Understand Privacy Policies

June 23, 2014

Today, privacy innovators TRUSTe & Disconnect have launched Privacy Icons software to help consumers easily understand website privacy policies and how websites are handling their data.

Consumers want to know how websites are using their privacy and data, but they often do not have the time or patience to read existing privacy policies, which are typically quite long and complex. According to The TRUSTe Privacy Index, the average privacy policy is 2,464 words long and takes about 10 minutes to read.

Deployment of Disconnect Icons

- Chrome Extension
- Web App

Discontinued in 2017 😞
Disconnect Icons Description

**Expected Use**
Does this website’s privacy policy disclose whether data it collects about you is used in ways other than you would reasonably expect given the site’s service?
- **Red** = Yes, without choice to opt-out.
- **Or, undisclosed.**
- **Yellow** = Yes, with choice to opt-out.
- **Green** = No.
- **Gray** = Info unavailable.

**Expected Collection**
Does this website’s privacy policy disclose whether it allows other companies like ad providers and analytics firms to track users on the site?
- **Red** = Yes, without choice to opt-out.
- **Or, undisclosed.**
- **Yellow** = Yes, with choice to opt-out.
- **Green** = No.
- **Gray** = Info unavailable.

**Precise Location**
Does this website’s privacy policy disclose whether the site or service tracks a user’s actual geolocation?
- **Red** = Yes, possibly without choice.
- **Yellow** = Yes, with choice.
- **Green** = No.
- **Gray** = Info unavailable.

**Data Retention**
Does this website’s privacy policy disclose how long they retain your personal data?
- **Red** = No data retention policy.
- **Yellow** = 12+ months.
- **Green** = 0-12 months.
- **Gray** = Info unavailable.

**Do Not Track**
Does this website comply with a user’s Do Not Track browser preference?
- **Green** = Yes.
- **Gray** = Info unavailable.

**Children Privacy**
Has this website received TRUSTe’s Children’s Privacy Certification?
- **Green** = Yes.
- **Gray** = No.
Structured Query

Get Segments such that

\[ \langle \rangle \]

**Category:** third party sharing

**Purpose:** advertising

---

**Diagram Description:**

1. **Application Layer**
   - Privacy policy link

2. **Data Layer**
   - Segmenter

3. **ML Layer**
   - Query Analyzer
     - (Identity Function)
   - Segment Classifier
     - (ML Classifier)

4. **Query Module**
   - User query

5. **Class Comparison**
   - Query classes
   - Classified segments

---

**Structured Query Text:**

Get Segments such that

\[ \langle \rangle \]

**Category:** third party sharing

**Purpose:** advertising
Does this website’s privacy policy disclose whether it allows other companies like ad providers and analytics firms to track users on the site?

**Red** = Yes, without choice to opt-out. Or, undisclosed.

**Yellow** = Yes, with choice to opt-out.

**Green** = No.

**Gray** = Info unavailable.

---

**Structured Query**

```json
Expected Collection

Structured Query

color: yellow
category_filters:
  - fun: include_some_in_list
    lst: ['third-party-sharing-collection']

value_filters:
  - fun: include_some_in_list
    lst: ['purpose_advertising','purpose_analytics-research']

  - fun: include_some_in_list
    lst: ['action-third-party_track-on-first-party-website-app','action-third-party_collect-on-first-party-website-app']

  - fun: include_some_in_list
    lst: ['choice-type_opt-out-link','choice-type_opt-out-via-contacting-company']

decider:
  - fun: not_empty
```
Same rules, on 50 policies from OPP-115 dataset

Icon assignment based on law students’ labels

vs.

Icon assignment based on Polisis’ labels

<table>
<thead>
<tr>
<th>Icon</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. Use</td>
<td>92%</td>
</tr>
<tr>
<td>Exp. Collection</td>
<td>88%</td>
</tr>
<tr>
<td>Precise Location</td>
<td>84%</td>
</tr>
<tr>
<td>Data Retention</td>
<td>80%</td>
</tr>
<tr>
<td>Children Privacy</td>
<td>98%</td>
</tr>
</tbody>
</table>

88.4% accuracy
Unstructured Querying: Answer Selection as a Case Study
Do you share my **address** with other companies?

We will provide your **location** to third parties.
Unstructured Queries

Do you share my **address** with other companies?

**Application Layer**
- Privacy policy link

**Data Layer**
- Segmenter

**ML Layer**
- Query Analyzer
  - (ML Classifiers)
- Segment Classifier
  - (ML Classifiers)

**Query Module**
- User query

**Class Comparison**
- Query classes
- Classified segments

(Custom Class Matching)
In order to answer the question, the segments in the privacy policy constitute the pool of candidate answers.

Given the output of the Segment Classifier, the policy's link to the Data Layer. The ML layer processes on the input to QAPri and then describes how assistants (e.g., Amazon Alexa and Google Assistant) and free-form queries (in the form of user questions) on the new permissive strategy are significantly red-dominated.

Furthermore, by automatically generating icons, we do not hallucinate.

Our second application of free-form privacy QA and then describe how unrealistic the problem of free-form privacy QA and then describe how unrealistic interpretation of the icons would the distributions of candidate answers prioritize answers that include the question's classes with high probability (but not necessarily vice-versa).

\[ s(q,a) = \frac{\sum_i (\beta_i \times \text{min}(\beta_i, \alpha_i))}{\sum_i \beta_i^2} \times \text{cer}(a) \]

\[ \text{cer}(a) = 1 - \left( -\sum (p_n(c_i|a) \times \ln(p_n(c_i|a))/\ln(|C|)) \right) \]
Evaluation Baselines

1- BM25 Model (Term-matching state of the art)
   - Compute IDF values on the corpus of 130,000 policies

2- Semantic Vector Model
   - Flat hierarchy across all values, CNN classifier

3- Random answer Model
• **120 questions** about **102 companies:**
Evaluation Metrics

- **Predictive Accuracy** (compared to experts answers)
- **User-perceived Utility** (how users perceived the answers)
Predictive Accuracy

Segmenter

Two Experts

A_1, A_11

Ranking Model

How many questions have an expert answer in top-k?

A_4, A_11, A_28
Predictive Accuracy: top-k score

fraction of Qs with answer among top-k answers

Hierarchical: 82% of questions had accurate answers in top 3

Differences become less significant with higher k

Random Retrieval SemVec Hierarchical

<table>
<thead>
<tr>
<th>k</th>
<th>top-k score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.24</td>
</tr>
<tr>
<td>2.0</td>
<td>0.57</td>
</tr>
<tr>
<td>3.0</td>
<td>0.67</td>
</tr>
<tr>
<td>4.0</td>
<td>0.77</td>
</tr>
<tr>
<td>5.0</td>
<td>0.87</td>
</tr>
<tr>
<td>6.0</td>
<td>0.89</td>
</tr>
<tr>
<td>7.0</td>
<td>0.92</td>
</tr>
</tbody>
</table>
DEMO
What data does the company gather for its own use? 🤔

Types of info they collect:
- Computer information
  - Contact
  - Demographic
  - IP Address And Device IDs
  - Location
  - Social Media Data
  - User Profile
  - Generic Personal Information
  - Other Data

Collection reasons:
- Other Purposes
- Basic Service Feature
- Marketing
- Additional Service Feature
- Personalization Customization
- Analytics Research
- Advertising

What options do they give?
Cool. This session is all about https://www.khanacademy.org.

Don’t worry you can change this throughout.

What do you want to ask?

Type your question or select existing one

GO!  ANOTHER COMPANY?
**Confidence Threshold**: 0.2

| Good | | Bad |
|------|--------------------------|
| ![Green Checkmark] You can request access and deletion of personal data | ![Red X] Some data might be retained indefinitely. |
| ![Green Checkmark] In certain conditions, data is not shared. | ![Red X] Data might be shared in the case of a merger or acquisition. |
| ![Green Checkmark] Data is not shared with third parties for advertising purposes. | |
| ![Green Checkmark] The policy states that third parties do not receive personal information. | |
| ![Green Checkmark] The policy offers you clear links to control your data | |
| ![Green Checkmark] Some of the collected data is anonymized or aggregated. | |
Impact

Users of the app
>35,000

Minutes on our apps
>88,000

Websites analyzed
>21,000
Take-aways

• **Polisis:**
  - Unified framework for querying privacy policies
  - Assisting users, regulators, and researchers
  - Two applications:
    - Structured querying: privacy icons generation
    - Unstructured querying: question answering from the privacy policy.

• **Read more at:**
  - Our paper
  - **WIRED:** Polisis AI Reads Privacy Policies So You Don't Have To
  - **Fast Company:** This Data Viz Tool Explains Privacy Policies You're Too Lazy to Read
  - **WSJ:** Those Privacy Policies Flooding Your Inbox? Print Them Out and They Span a Football Field