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

Hamza Harkous, EPFL
Kassem Fawaz, UW-Madison
Rémi Lebret, EPFL
Florian Schaub, UMich
Kang G. Shin, UMich
Karl Aberer, EPFL

Published at USENIX Security 2019

PRIVACYCON

Problem?

Privacy Policies, *As the de facto Privacy Notice Mechanism*, are long and complex

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

*(McDonald and Cranor -2008)

PRIVACYCON

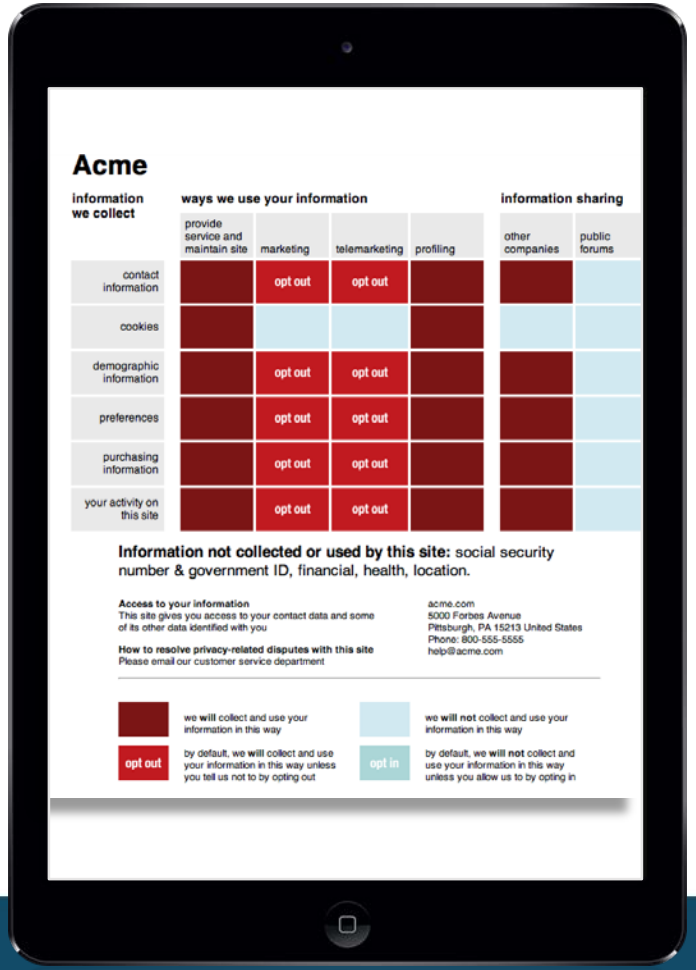
Approaches So Far?

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

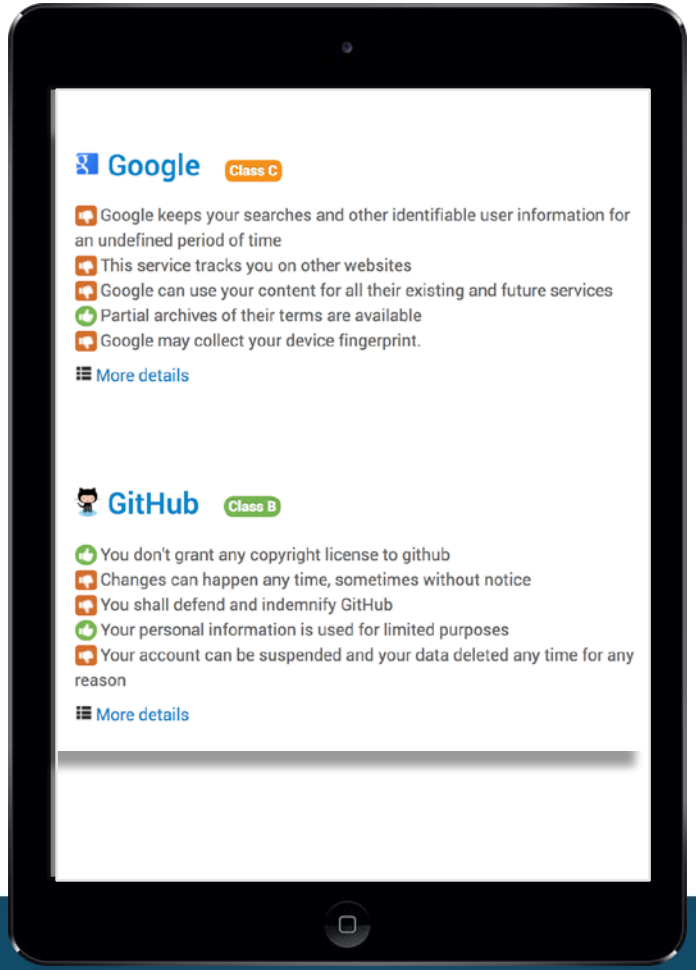


PRIVACYCON

Crowdsourcing

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

*Zimmeck and Bellovin , "Privee: An Architecture for Automatically Analyzing Web Privacy Policies". USENIX Security 2014



Manual work doesn’t scale.

...

Fails to cope with emerging technologies.

PRIVACYCON

Voice-Activated Devices

- Read the whole policy?
- Show tables on small screens?

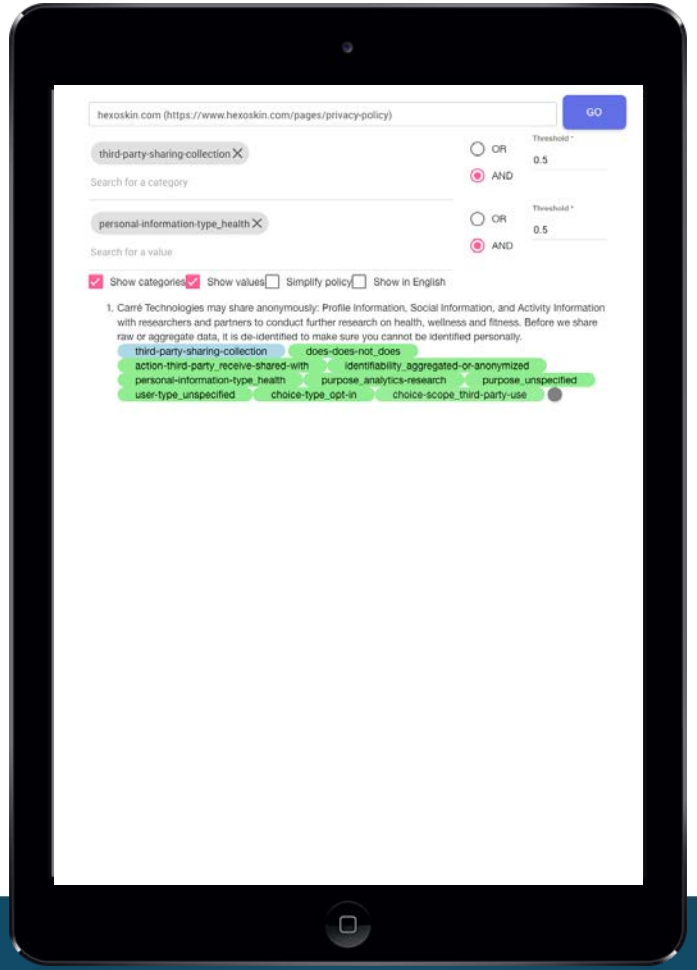
Unstructured Query (User Questions)

Regulation Compliance (e.g. GDPR)

Find Statements About Health Data Sharing

```
Get Segments such that
Category: third party sharing
personal information type: health information
```

Structured Query



Solution?

POLISIS

Unified Framework for Privacy Policies Analysis

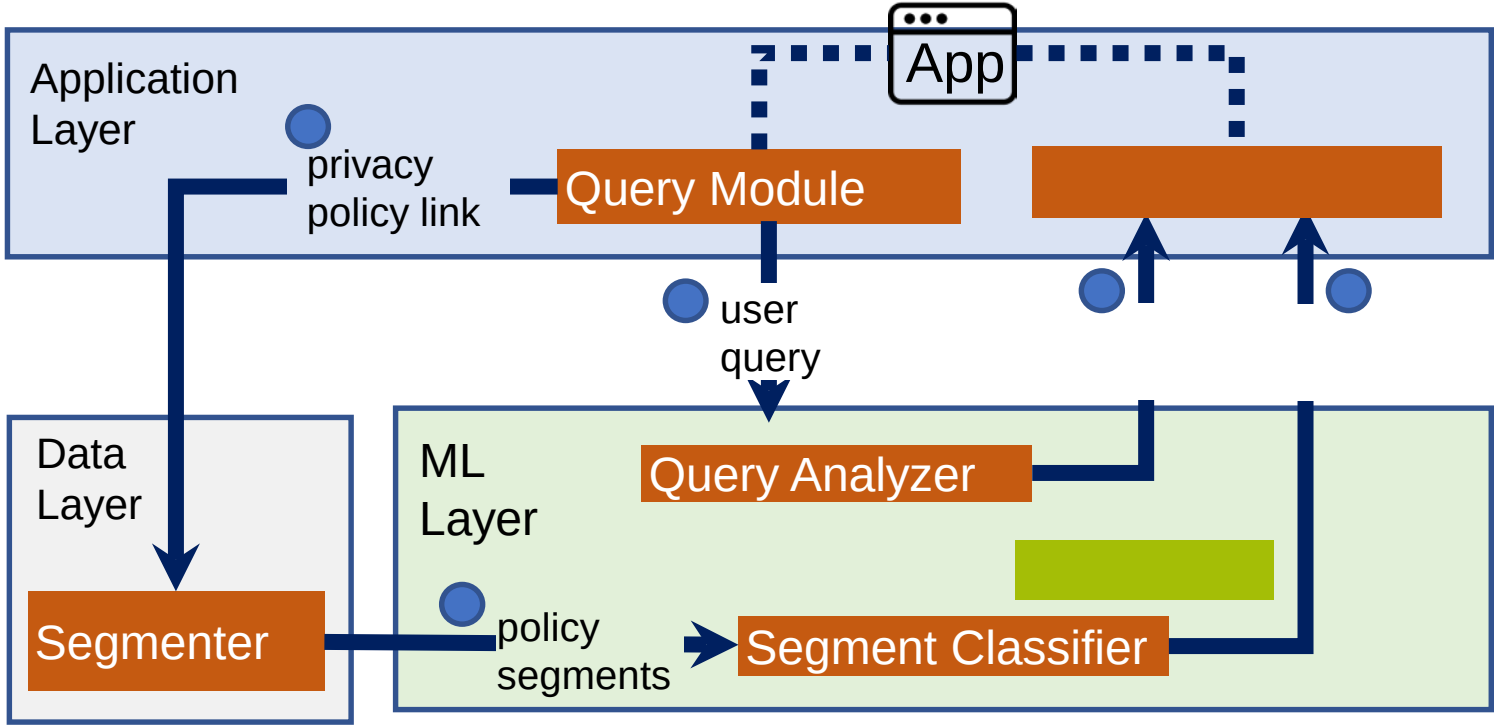
Once we automate policies' analysis, we can create a new interface for millions of policies with a single program.

Structured Queries

Unstructured Queries

pribot.org

Framework

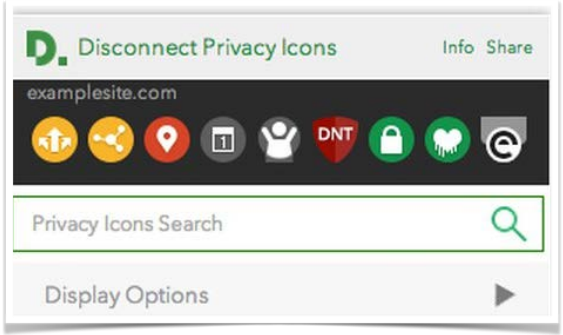

Application Layer
App
privacy policy link
Query Module
user query
Data Layer
ML Layer
Query Analyzer
Segmenter
policy segments
Segment Classifier

Structured Querying

Privacy Icon Assignment as a Case Study

PRIVACYCON

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.

Structured Query Example

Expected Collection

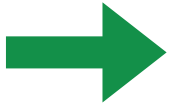
Does this website's privacy policy disclose whether it allows other companies like ad providers and analytics firms to track users on the site?

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

Yellow = Yes, with choice to opt-out.

Green = No.

Gray = Info unavailable.



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
```

Unstructured Querying

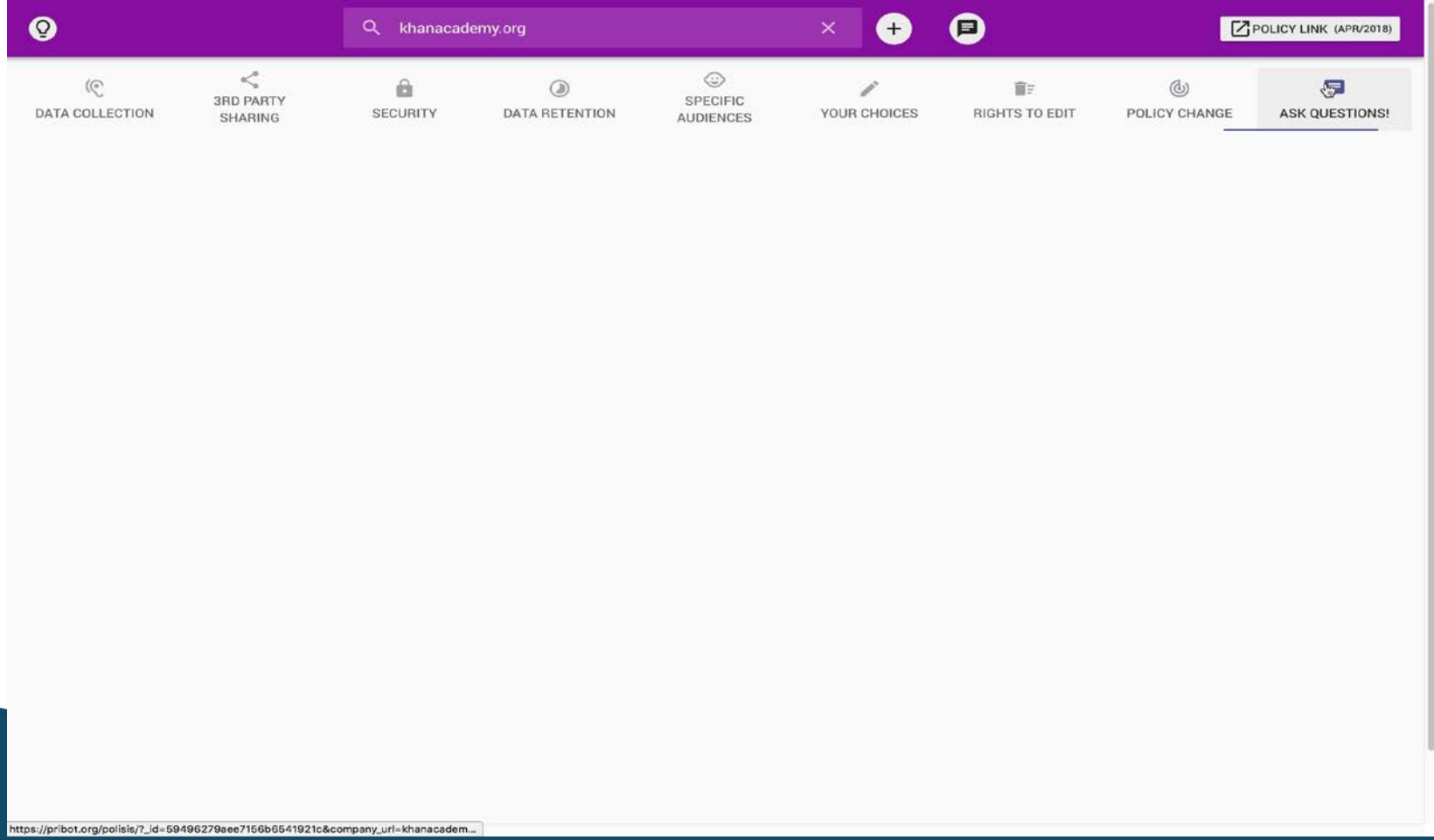
Answer Selection as a Case Study

PRIVACYCON

Answer Selection

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

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


khanacademy.org
POLICY LINK (APR/2018)
DATA COLLECTION
3RD PARTY SHARING
SECURITY
DATA RETENTION
SPECIFIC AUDIENCES
YOUR CHOICES
RIGHTS TO EDIT
POLICY CHANGE
ASK QUESTIONS!
https://pribot.org/polisis/?_id=59496279aee7156b6541921c&company_url=khanacadem...

Impact

Users of the app
>45,000

MINUTES ON OUR APPS
>111,000

WEBSITES ANALYZED
>21,000

NVIDIA · FAST COMPANY · Product Hunt · WSJ · WIRED · WORLD ECONOMIC FORUM

Take-aways

- **Polisis**:
  - Unified framework for querying privacy policies
  - Assisting users, regulators, and researchers
  - **Applications**: Structured and Unstructured Queries
- **Read more at:**
  - Our Polisis paper and GDPR paper
    - **Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning**
    - **The privacy policy landscape after the GDPR**
  - **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