USABLE PRIVACY POLICY AND PERSONALIZED PRIVACY ASSISTANT PROJECTS

An Overview of Usable Privacy Technologies, Tools and Findings Coming Out of Recent Research at CMU

Instructors:

Anupam Das, Martin Degeling, Norman Sadeh,
Sebastian Zimmeck

Carnegie Mellon University

usableprivacy.org   privacyassistant.org

explore.usableprivacy.org

Instructors/Moderators

Anupam Das	**Post-doctoral Fellow, School of Computer Science, CMU Member, Personalized Privacy Assistant Project PhD, Univ. of Illinois, Urbana-Champaign**	
Martin Degeling	**Post-doctoral Fellow, School of Computer Science, CMU Member, Personalized Privacy Assistant Project PhD, Univ. of Duisburg-Essen**	
Norman Sadeh	**Professor, School of Computer Science, CMU Principal Investigator, Usable Privacy Policy Project & Personalized Privacy Assistant Project PhD, CMU**	
Sebastian Zimmeck	**Post-doctoral Fellow, School of Computer Science, CMU Member of Usable Privacy Policy Project & Personalized Privacy Assistant Project PhD, Columbia University**	

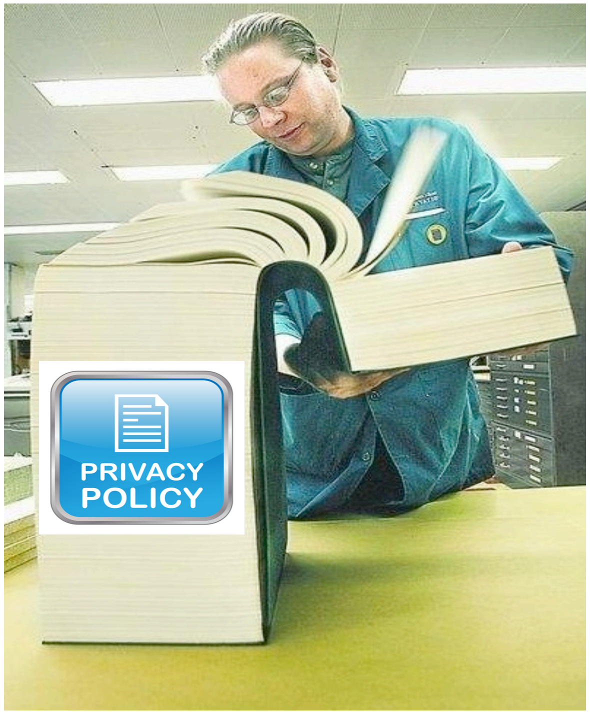
A Word About this Tutorial

- This tutorial is intended to be **interactive**. Feel free to interrupt us at anytime. We have also carved out specific times for discussion in each session – typically at the end
- **Disclaimer:** As its title implies, this tutorial focuses on research at CMU
  - We have built on the work of many others and aim to always acknowledge everyone in our publications
  - For the sake of maintaining a fluid narrative, we will be focusing solely on work at CMU. **Please refer to our publications for a proper set of citations.**

Privacy in the Age of IoT

- Data-centric economy
- **Notice and choice** in its current implementation is **not working/practical**
- **91%** of people report feeling they **have lost control over their information** -

Pew Survey 2014 http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/



Mobile and IoT: A Number of Complicating Factors

- A typical mobile phone user with 50 mobile apps each requesting 3 permissions would have to **configure over 100 settings**
- IoT: Technology is often **"invisible"**
- **Reading policies is even less practical**
- Explosion in the number of apps and devices: Developers often **lack the necessary sophistication**

"Modeling Users' Mobile App Privacy Preferences: Restoring Usablility in a Sea of Permission Settings", J. Lin, B. Liu, N. Sadeh, J. Hong, Proc. of the USENIX Symposium on Usable Privacy and Security, SOUPS 2014, Jul. 2014

What If….

- **Computers understood privacy policies?**
  - Machine-readable policies have been proposed but have not gained traction
- **Computers understood** what we **care about** and what we **already know/expect**

We Could Develop…

- **UI's (e.g. Personal Privacy Assistants)** that:
  - selectively **inform us** about practices we care about/don't expect - **NOTICE**
  - Help us discover and **configure available settings** – **CHOICE**
- Tools to **help developers** avoid being in violation of relevant laws
- Tools to **help app stores and regulators identify potential violations** of relevant laws
- Monitor **privacy policy trends over time**

This Tutorial: Three Sessions

- Session I (1-2:15pm) : **Semi-automated extraction of data practice statements from natural language privacy policies**
  - **Instructors/Moderators: Sadeh and Zimmeck**
- Session II (2:30-3:45pm) : **Mobile App Privacy Compliance Analysis**
  - **Instructors/Moderators: Zimmeck and Sadeh**
- Session III (4-5:15pm) : **Personalized Privacy Assistants for Mobile and IoT**
  Instructors/Moderators: Sadeh, Das and Degeling

USABLE PRIVACY POLICY AND PERSONALIZED PRIVACY ASSISTANT PROJECTS

Session I: Semi-automated extraction of data practice statements from natural language privacy policies

Instructor:

Norman Sadeh

Carnegie Mellon University

usableprivacy.org   privacyassistant.org

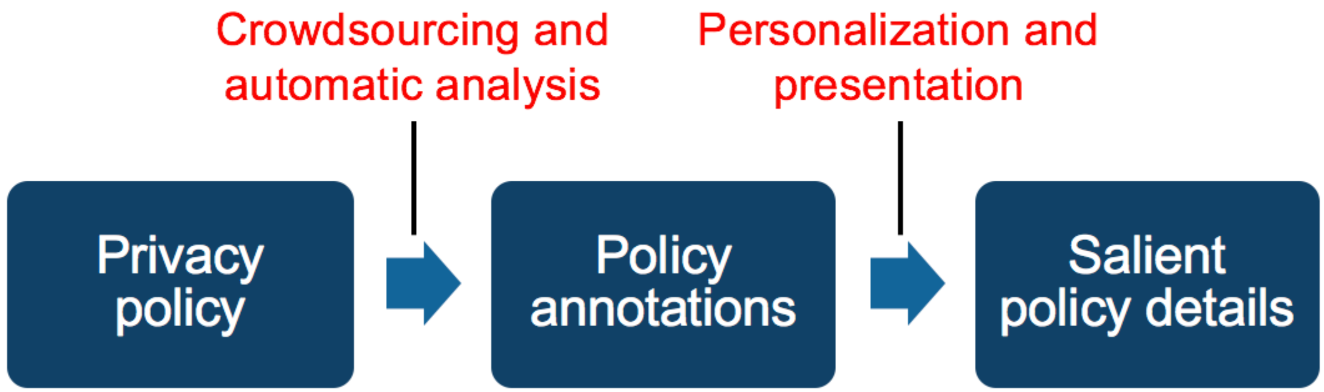
explore.usableprivacy.org

Session I: Outline

- Crowdsourcing Privacy Policy Annotations
- Automating the Extraction of Privacy Policy Annotations
- Existing Results and Tools
  - Including hands-on evaluation and discussion
- Semantic Reasoning (time permitting)

The Usable Privacy Policy Project

Approach: Use crowdsourcing, machine learning, and NLP techniques to automatically (or semi-automatically) extract salient details from privacy policies.



www.usableprivacy.org

"The Usable Privacy Policy Project", N. Sadeh et al., CMU Technical Report, CMU-ISR-13-119, 2013

Can We Use Crowdworkers to Annotate Policies?

Crowdsourcing Experiment #1

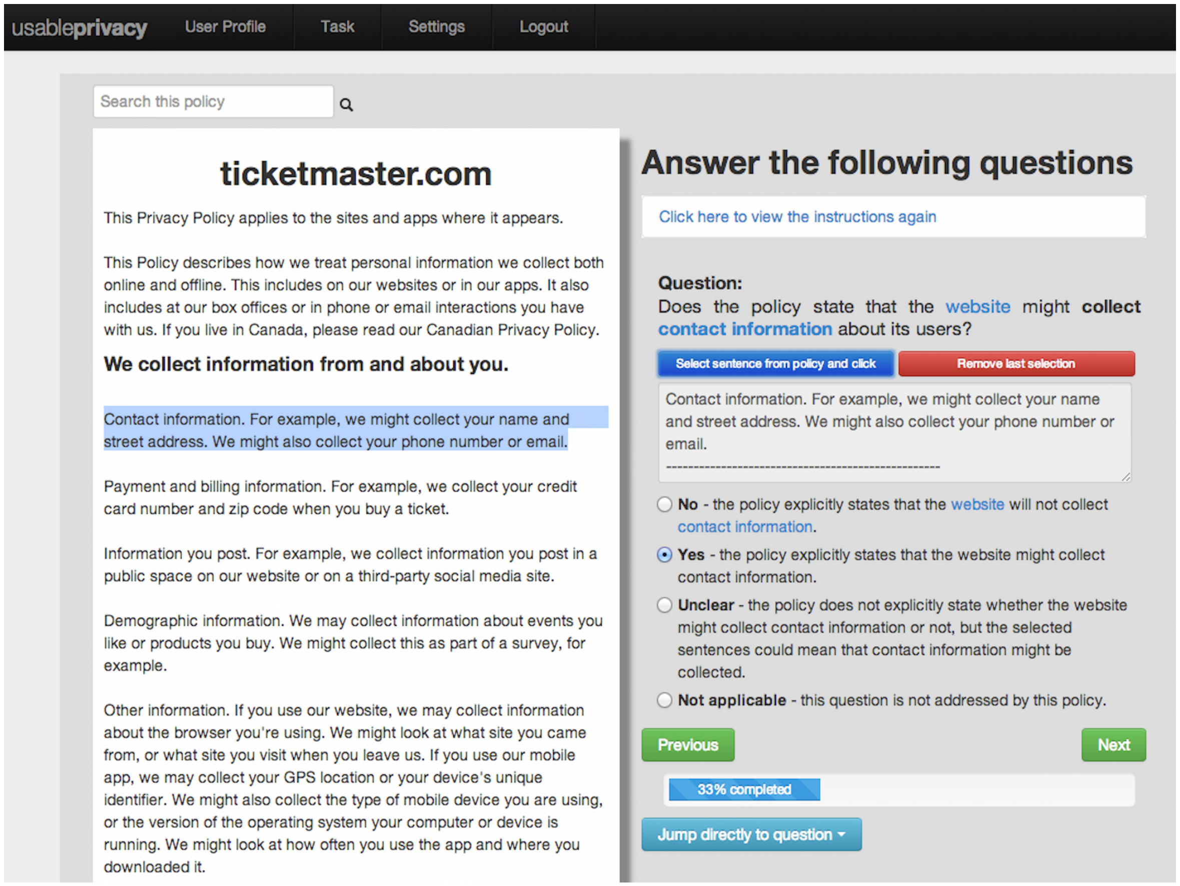
- 26 website privacy policies
- 9 questions

Annotators per policy:

- 10 crowdworkers (Mechanical Turk)
- 5 skilled annotators (law students or equivalent) – used as gold standard

A Crowdsourcing Task

14

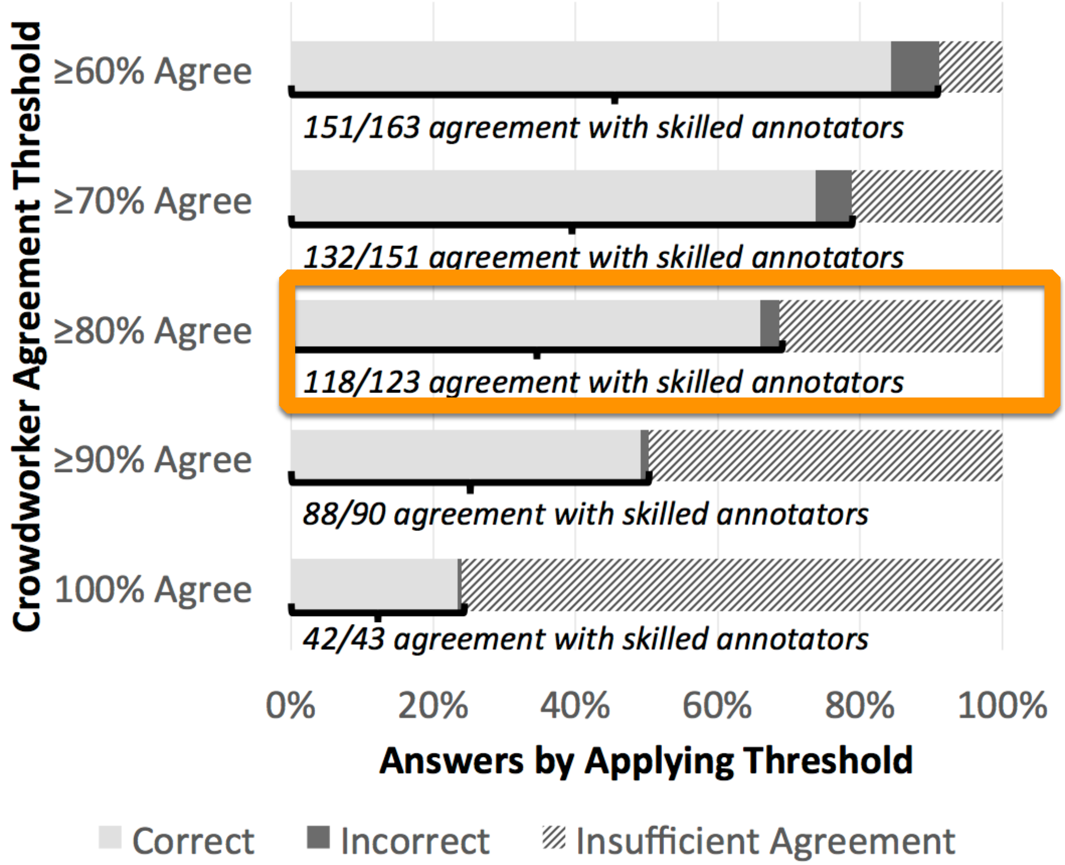


Crowdworkers Can Actually Be Good at This

with sufficiently high **agreement threshold:**

crowdworkers agree on **same interpretation as skilled annotators**

or **do not converge**



Annotation of 26 policies

Wilson, S., Schaub, F., Ramanath, R., Sadeh, N., Liu, F., Smith, N., and Liu, F. Crowdsourcing Annotations for Websites Privacy Policies: Can It Really Work? WWW Conference, May 2016

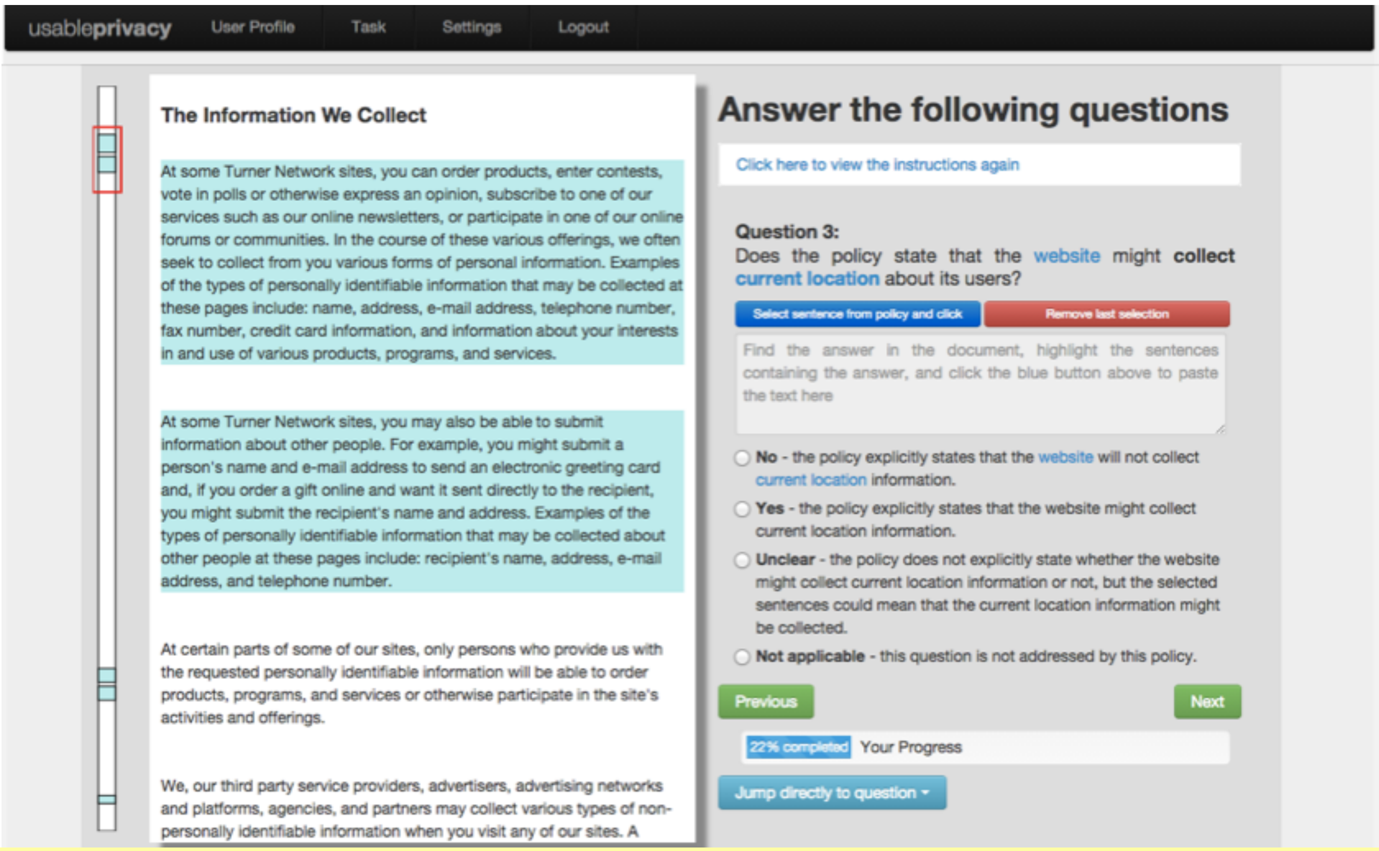
Can We Help the Crowdworkers?

On average, crowdworkers took 24 minutes to answer all nine questions about a privacy policy.

We wanted to make the task less difficult and help crowdworkers read more efficiently.

To do this, we built relevance models for each of the nine questions and highlighted highlighted ragraphs that were relevant for each question.

An Improved Crowdsourcing Task



Highlighting based on handcrafted regular expressions and some machine learning

Wilson, S., Schaub, F., Ramanath, R., Sadeh, N., Liu, F., Smith, N., and Liu, F. Crowdsourcing Annotations for Websites Privacy Policies: Can It Really Work? WWW Conference, May 2016

Crowdsourcing Experiment #2

12 website privacy policies

9 questions

108 question-policy pairs

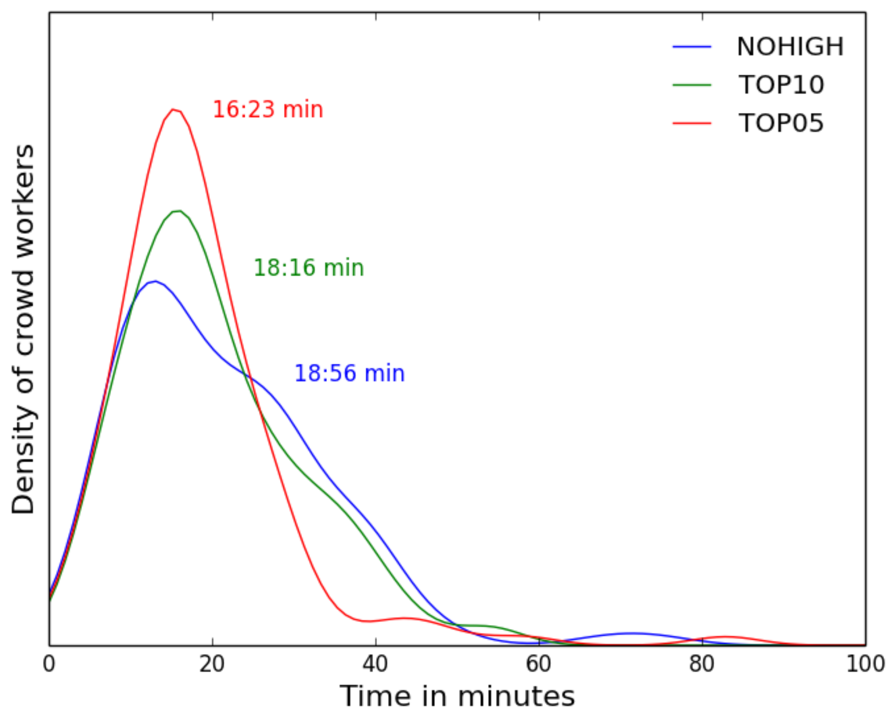
Three conditions:

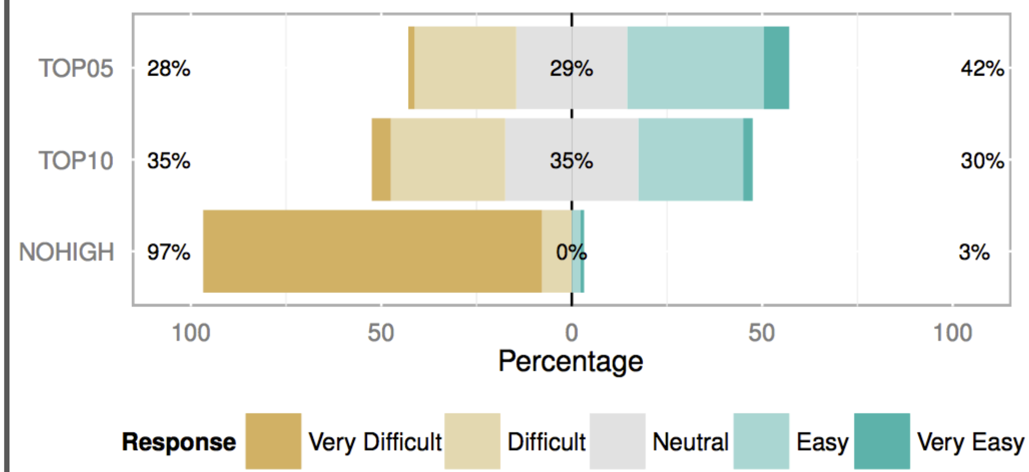
NOHIGH, TOP05, TOP10

10 crowdworkers in each condition

All crowdworkers were unique

Crowdworkers Can Be Helped

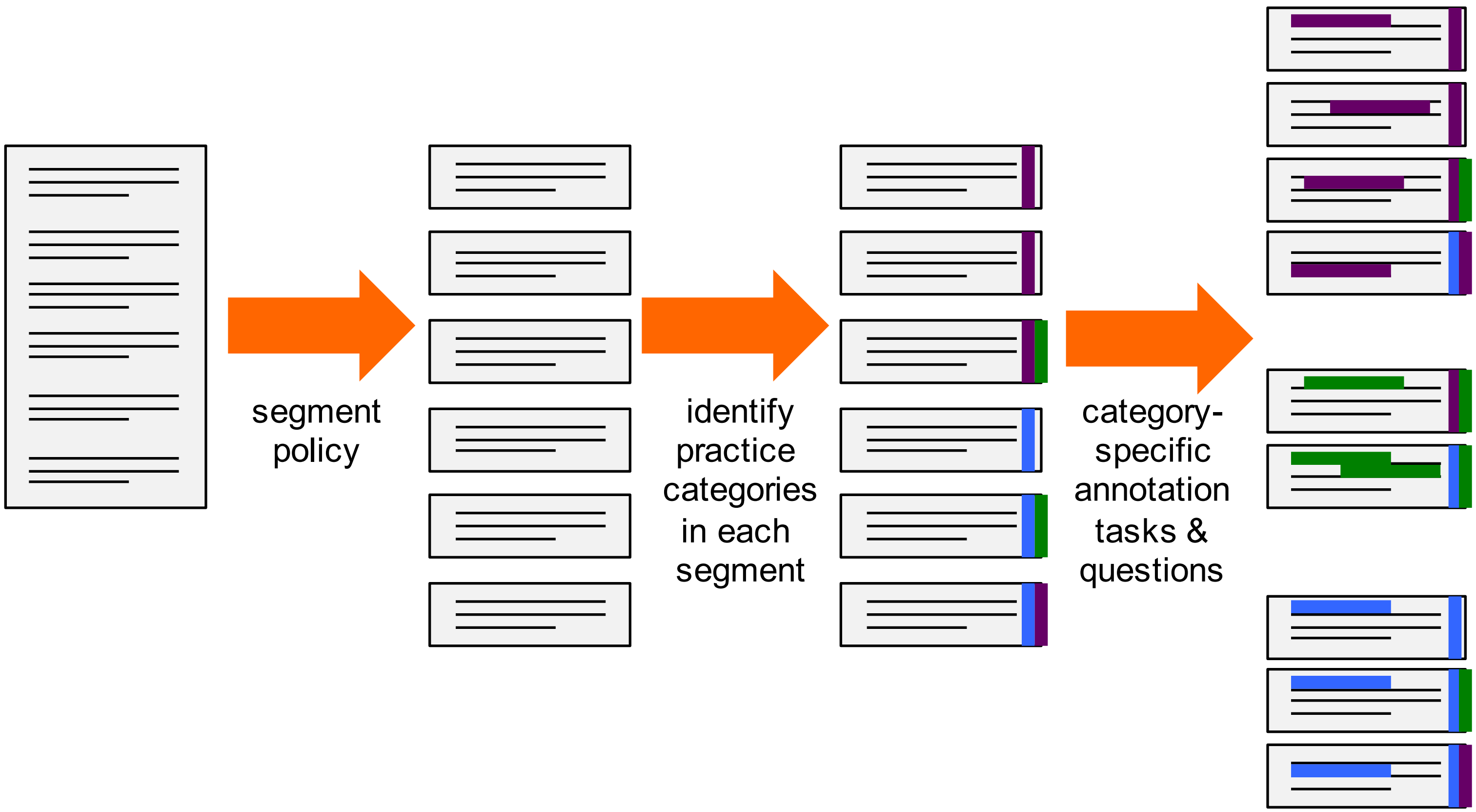




Observations

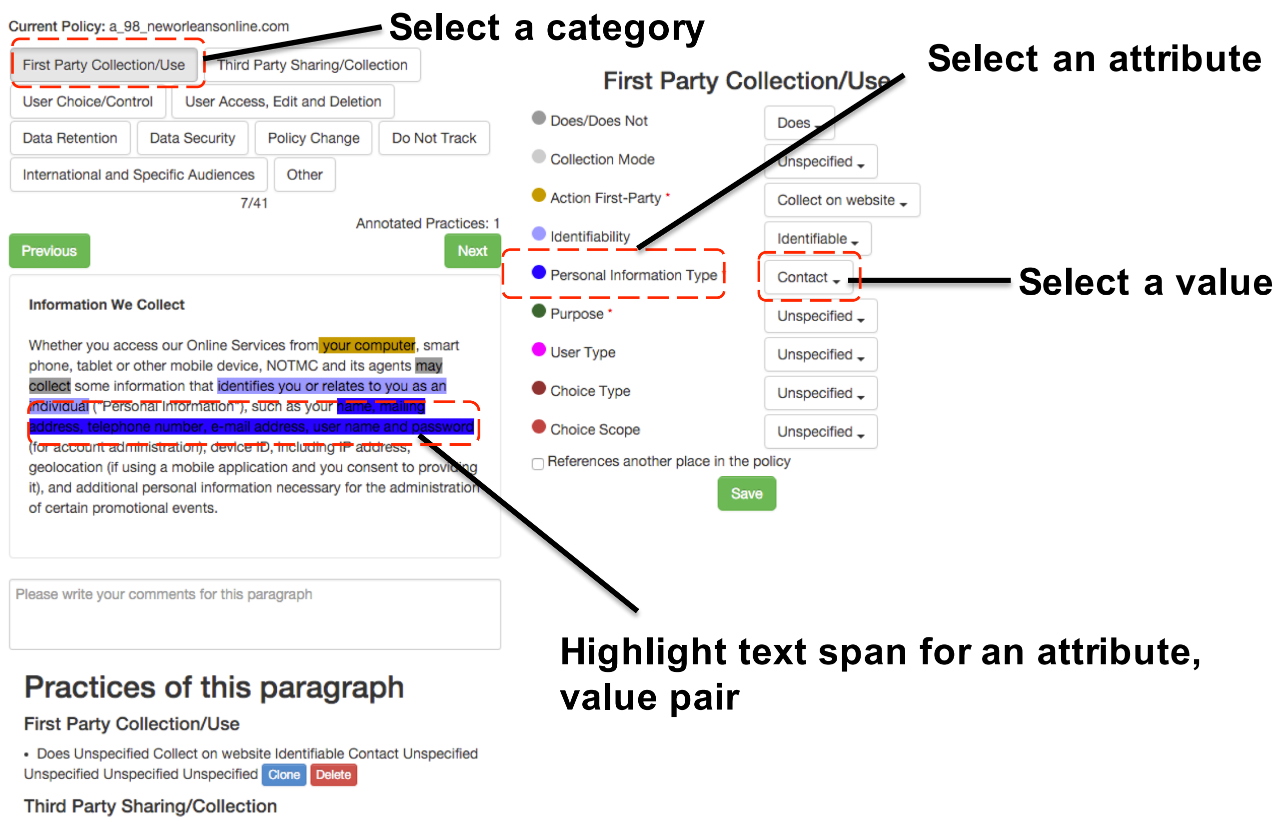
- Aggregating crowdworkers' answers to questions about privacy policies produces fairly accurate results – **crowdworkers often converge on the correct answers**
- **Highlighting** relevant paragraphs for each question:
  - **Does not negatively impact crowdworker accuracy**
  - Shows (mild) **indications of speeding up the task**
  - Makes **crowdworkers more confident** about reading privacy policies
- **But the tasks are still too long for this approach to really scale**

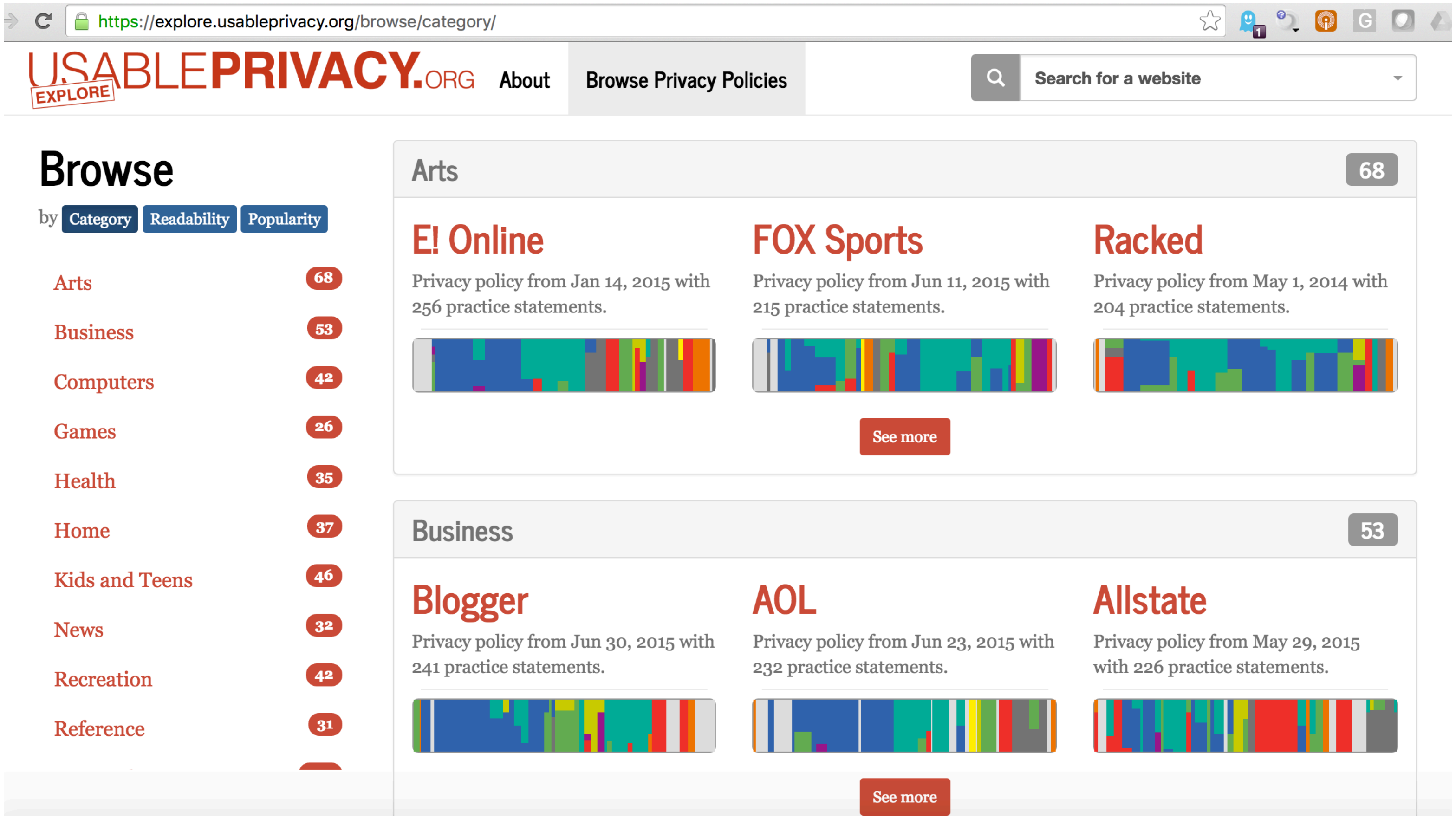
Multi-step annotations



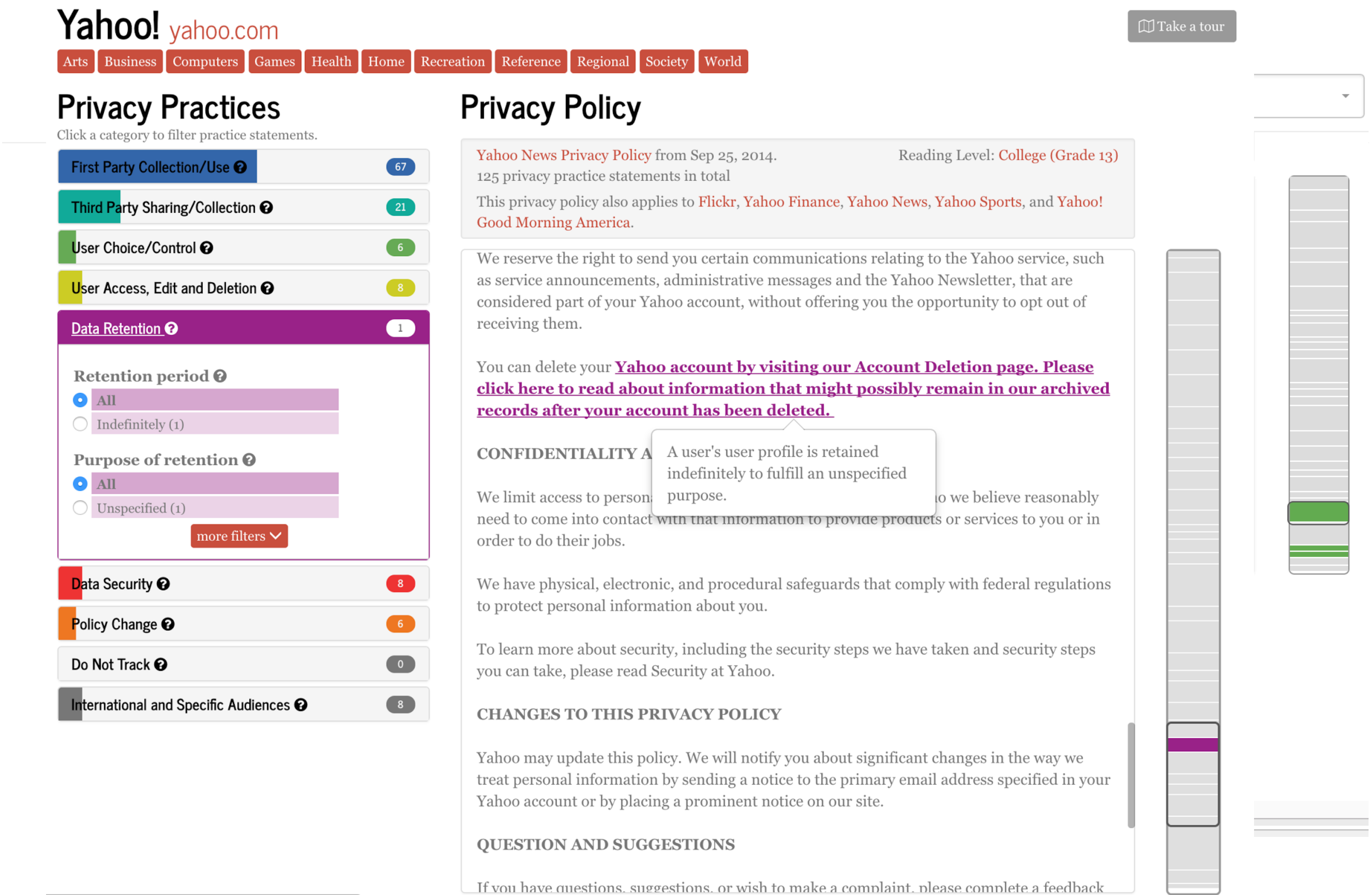
F. Schaub, T. Breaux, N. Sadeh, “Crowdsourcing Privacy Policy Analysis: Potential, Challenges and Best Practices,” in *Information Technology*, Vol. 58, 2016

Annotation Tool





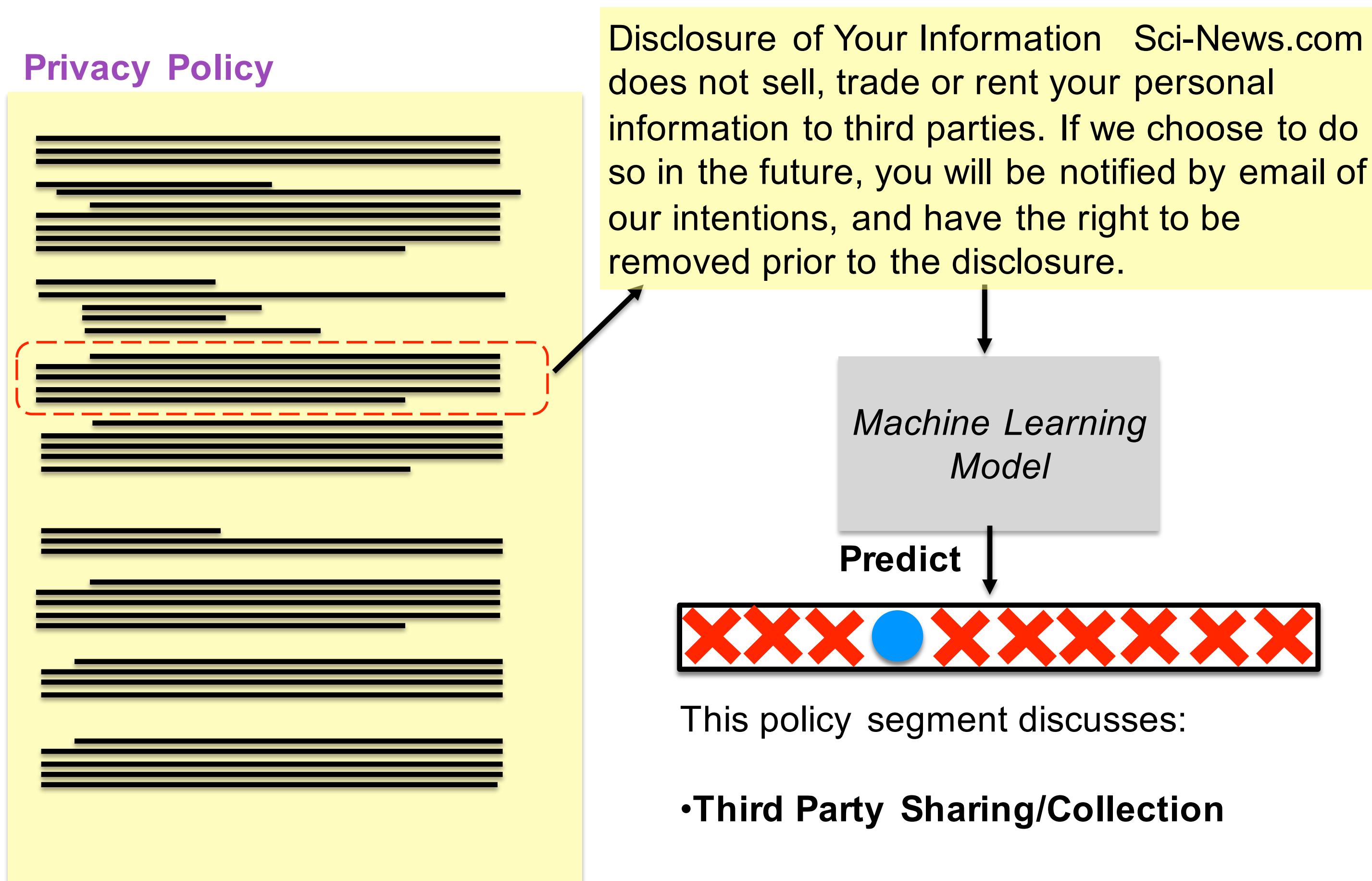
S. Wilson, F. Schaub, A. Dara, F. Liu, S. Cherivirala, P.G. Leon, M.S. Andersen, S. Zimmeck, K. Sathyendra, N.C. Russell, T.B. Norton, E. Hovy, J.R. Reidenberg, N. Sadeh, "The Creation and Analysis of a Website Privacy Policy Corpus", ACL '16: Annual Meeting of the Association for Computational Linguistics, Aug 2016


Yahoo! yahoo.com
Take a tour
Arts Business Computers Games Health Home Recreation Reference Regional Society World
Privacy Practices
Click a category to filter practice statements.
First Party Collection/Use 67
Third Party Sharing/Collection 21
User Choice/Control 6
User Access, Edit and Deletion 8
Data Retention 1
Retention period
All
Indefinitely (1)
Purpose of retention
All
Unspecified (1)
more filters
Data Security 8
Policy Change 6
Do Not Track 0
International and Specific Audiences 8
Privacy Policy
Yahoo News Privacy Policy from Sep 25, 2014.
Reading Level: College (Grade 13)
125 privacy practice statements in total
This privacy policy also applies to Flickr, Yahoo Finance, Yahoo News, Yahoo Sports, and Yahoo! Good Morning America.
We reserve the right to send you certain communications relating to the Yahoo service, such as service announcements, administrative messages and the Yahoo Newsletter, that are considered part of your Yahoo account, without offering you the opportunity to opt out of receiving them.
You can delete your Yahoo account by visiting our Account Deletion page. Please click here to read about information that might possibly remain in our archived records after your account has been deleted.
A user's user profile is retained indefinitely to fulfill an unspecified purpose.
We have physical, electronic, and procedural safeguards that comply with federal regulations to protect personal information about you.
To learn more about security, including the security steps we have taken and security steps you can take, please read Security at Yahoo.
CHANGES TO THIS PRIVACY POLICY
Yahoo may update this policy. We will notify you about significant changes in the way we treat personal information by sending a notice to the primary email address specified in your Yahoo account or by placing a prominent notice on our site.
QUESTION AND SUGGESTIONS

Observations & Question

- Crowdsourcing has scaling issues
- Could we automate parts of this process/the entire process?

A First Task: Segment Annotation



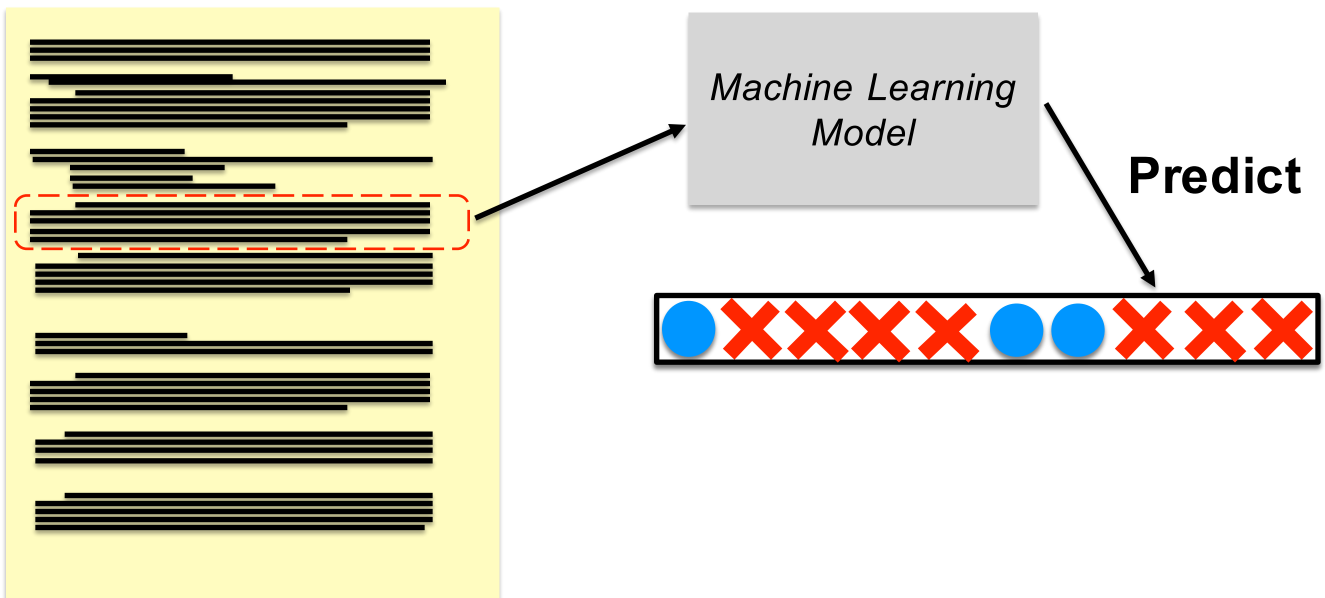
Dataset

- Number of policy segments: 3792
- Number of categories: 10
- The golden standard is the aggregation of three annotators (e.g. 2 out of 3)

Approach

- One classifier per data practice

Unannotated Privacy Policy



F. Liu, S. Wilson, F. Schaub, N. Sadeh.. *Analyzing Vocabulary Intersections of Expert Annotations and Topic Models for Data Practices in Privacy Policies* AAAI Fall Symposium on Privacy and Language Technologies. 2016.

Multiple Possible Classifiers

- **Traditional Methods**
  - Bag of N-grams as features
  - Multinomial Naive Bayes
  - Logistic regression
  - Support Vector Machines
- **Neural Methods**
  - One-hot vector as input
  - Recurrent Neural Networks
  - Convolutional Neural Networks

Training

- We split the 115 policies of the OPP-115 corpus into 80% (92 polices) training and 20% (23 policies) for testing.
- Built binary classifiers for each category.
- We used a unigram, bigram term frequency--inverse document frequency (tf--idf) for traditional methods. The parameters for each model are tuned with 5-fold cross validation.
- The parameters for the Neural Models use 10% of the training set as a held-out development set to pick the best models.

Number of instances

- Our dataset consists of 3,792 instances at the segment level, and 11,033 at the sentence level extracted from the 115 policies by setting an instance as positive if 2 or more annotators agree that the instance contains information about the specific category.
  - *Note:* results with sentence-level predictions are not as good. Here we focus on segment-level predictions

Performance (Precision/Recall/F1)

	Segment					
	RNN	CNN	LR	SVM	MNB	NBLR
First Party Collection/Use	0.80/0.74/0.80	0.79/0.85/0.82	0.79/0.84/0.81	0.78/0.84/0.81	0.73/0.84/0.78	0.83/0.72/0.77
Third Party Sharing/Collection	0.72/0.70/0.72	0.84/0.69/0.76	0.79/0.81/0.80	0.79/0.81/0.80	0.78/0.78/0.78	0.82/0.72/0.77
User, Choice/Control	0.79/0.55/0.65	0.80/0.49/0.61	0.68/0.66/0.67	0.71/0.64/0.67	0.68/0.49/0.57	0.77/0.50/0.61
User, Access, Edit &Deletion	0.75/0.40/0.52	0.83/0.50/0.62	0.71/0.67/0.69	0.76/0.63/0.69	0.69/0.37/0.48	0.83/0.50/0.62
Data Retention	0.00/0.00/0.00	1.00/0.12/0.21	1.00/0.41/0.58	1.00/0.35/0.52	0.80/0.24/0.36	0.86/0.35/0.50
Data Security	1.00/0.49/0.66	0.97/0.59/0.73	0.85/0.67/0.75	0.89/0.63/0.74	0.92/0.67/0.77	0.94/0.63/0.75
Policy Chang	0.83/0.50/0.62	0.92/0.60/0.73	0.85/0.85/0.85	0.83/0.75/0.79	0.64/0.90/0.75	0.88/0.70/0.78
Do Not Track	1.00/0.75/0.86	1.00/0.75/0.86	1.00/0.75/0.86	1.00/0.75/0.86	1.00/0.75/0.86	1.00/0.75/0.86
International & Specific Audiences	0.91/0.80/0.85	0.86/0.80/0.83	0.87/0.87/0.87	0.85/0.87/0.86	0.85/0.85/0.85	0.92/0.85/0.88
Micro Average	0.81/0.65/0.72	0.84/0.70/0.75	0.79/0.78/0.78	0.80/0.77/0.78	0.76/0.73/0.74	0.84/0.68/0.75

Classification results (precision/recall/F1-score)

- ❖ Segment labeling beats sentence labeling
- ❖ There are differences in performance but many techniques are pretty close
- ❖ Selecting techniques just based on F1 scores is probably simplistic
  - ❖ Need to think about one's objective (e.g. precision might be more important than recall)
- ❖ Note: Performance is also a **reflection of the number of available training instances in each category**

Another Task: User Choice Instance Extraction

Choice Instance !!!
If you do not want us to use personal information that we gather to allow third parties to personalize advertisements we display to you, please adjust your Advertising Preferences .

- Users choices often buried deep in the text of long policies
- Is it possible to **automatically extract information** about such “choice instances” from privacy policies?
- Use Natural Language Toolkit tokenizer to subdivide segments into sentences & build classifiers

K.M. Sathyendra, F. Schaub, S. Wilson, N. Sadeh. *Automatic Extraction of Opt-Out Choices from Privacy Policies.* AAAI Fall Symposium on Privacy and Language Technologies. 2016.
K.M. Sathyendra, S. Wilson, F. Schaub, S. Zimmeck, N. Sadeh. *Identifying the Provision of Choices in Privacy Policies, EMNLP Conference, 2017 (accepted for publication)*

Privacy Choices

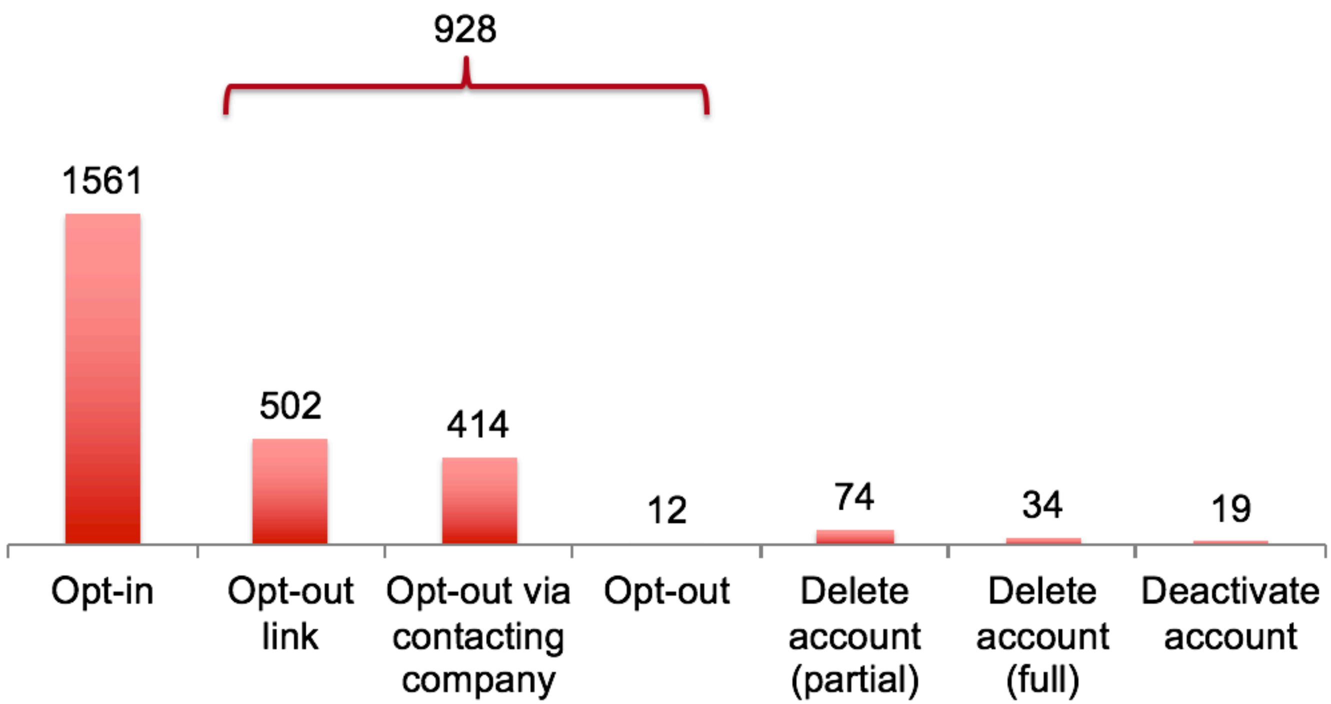
- Privacy choices include choices such as Deactivate Account, Delete Account, Opt-In, Opt-Out, Opt-Out Hyperlink, Opt-Out via contacting company

Opt Out Choices

↑

First Party Collection/Use
Third Party Sharing/Collection
User Choice/Control
User Access, Edit, & Deletion
Data Retention
Data Security
Policy Change
Do Not Track
International & Specific Audiences
Other

Privacy Choice Distribution in OPP 115 Corpus



Machine Learning Models Summary

Feature Set	Model	Precision	Recall	F1	Accuracy
Unigram	**Logistic Regression**	**0.574**	**0.493**	**0.530**	**0.987**
	SVM	0.417	0.493	0.452	0.982
	Naïve Bayes	0.263	0.634	0.372	0.967
	Random Forest	0.667	0.254	0.367	0.987
Unigram + bigram Bag of words	**Logistic Regression**	**0.565**	**0.549**	**0.557**	**0.987**
	SVM	0.537	0.507	0.522	0.986
	1 Nearest Neighbor	0.542	0.451	0.492	0.986
	4-NN with 1000 features	0.581	0.352	0.439	0.986
	Naïve Bayes	0.324	0.662	0.435	0.974
	4 NN	0.571	0.338	0.425	0.986
	Random Forest	0.645	0.282	0.392	0.987
	5 NN	0.543	0.268	0.358	0.985
Custom Feature: Unigram and Bigram bag of words + Modal Verbs and opt-out specific phrases	**Logistic Regression**	**0.614**	**0.605**	**0.609**	**0.988**
Custom Feature and Phrase Inclusion Model 1	**Combination Model: Logistic Regression and Phrase Inclusion Model 1**	**0.689**	**0.591**	**0.636**	**0.989**

Best results today: Precision: 0.926; Recall: 0.641; F1: 0.758

Fine Grained Classification

	FI	TH	BR	
AD	**15**	**52**	0	67
SH	**6**	**2**	0	8
AN	0	**4**	0	4
CK	**1**	**1**	**2**	4
CM	**19**	0	0	19
	42	59	2	101

	True Positives	True Negatives	False Positives	False Negatives	Precision	Recall
FI	21	102	0	1	1	0.954545455
TH	100	21	1	2	0.99009901	0.980392157
AD	87	26	8	3	0.915789474	0.966666667
CM	18	103	1	2	0.947368421	0.9
CK	2	122	0	0	1	1
Other Tags	0	0			0	0

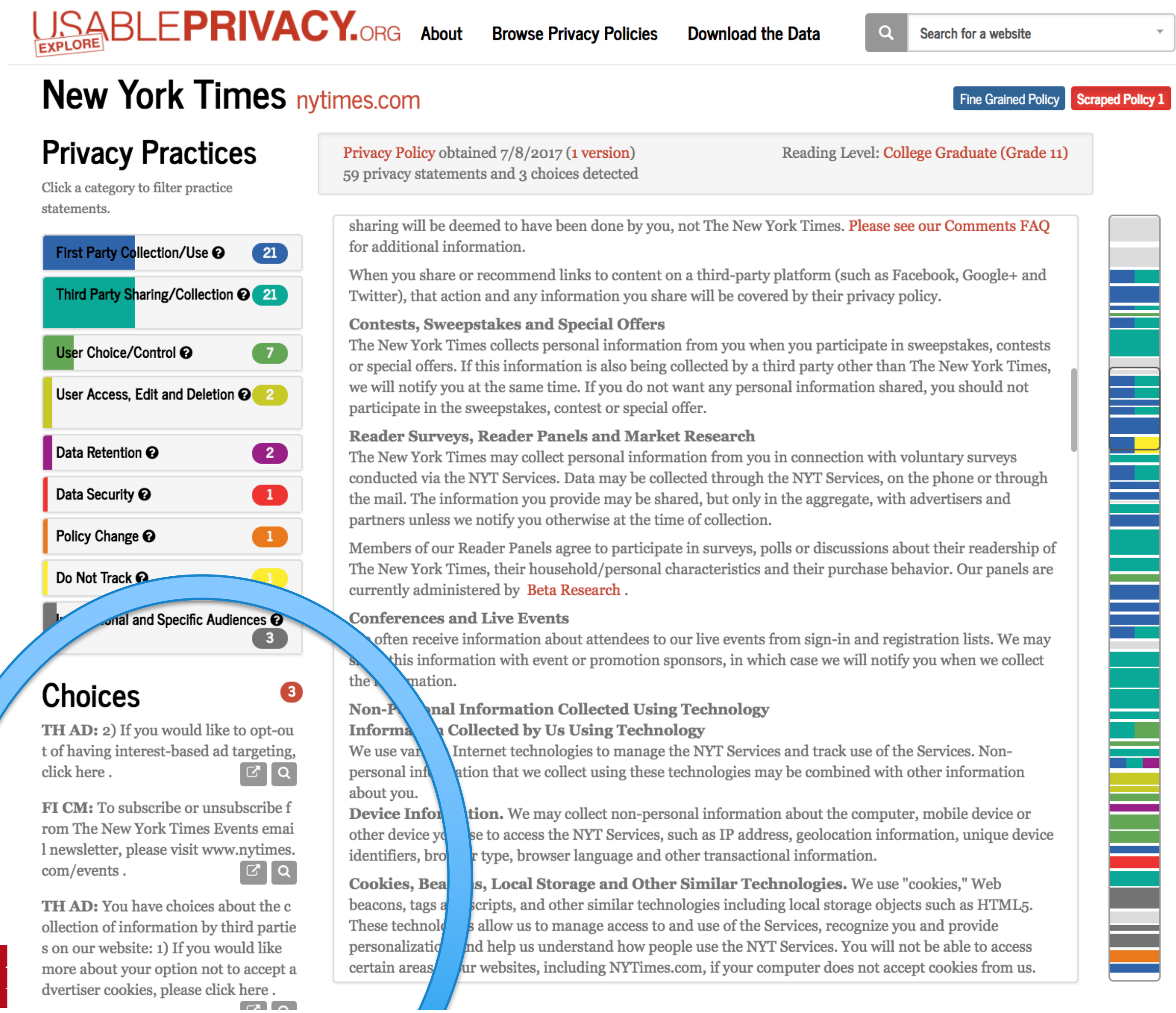
K.M. Sathyendra, S. Wilson, F. Schaub, S. Zimmeck, N. Sadeh. *Identifying the Provision of Choices in Privacy Policies, EMNLP Conference, 2017 (accepted for publication)*

Observation & Question

- Automated understanding is not beyond reach but we will have to do with less than 100% accuracy….at least for a while
- Could this be enough to automatically analyze privacy policies and identify compliance issues?

Hands-On with Automated Annotations

- Let's start by looking at sites we have automatically annotated
- Policies automatically annotated over the past week --- this is **still in beta**.
  - **Please do not share with others right now**: the site is still under development and will likely be down/unstable over the weeks ahead
  - We expect to have an official launch later this summer

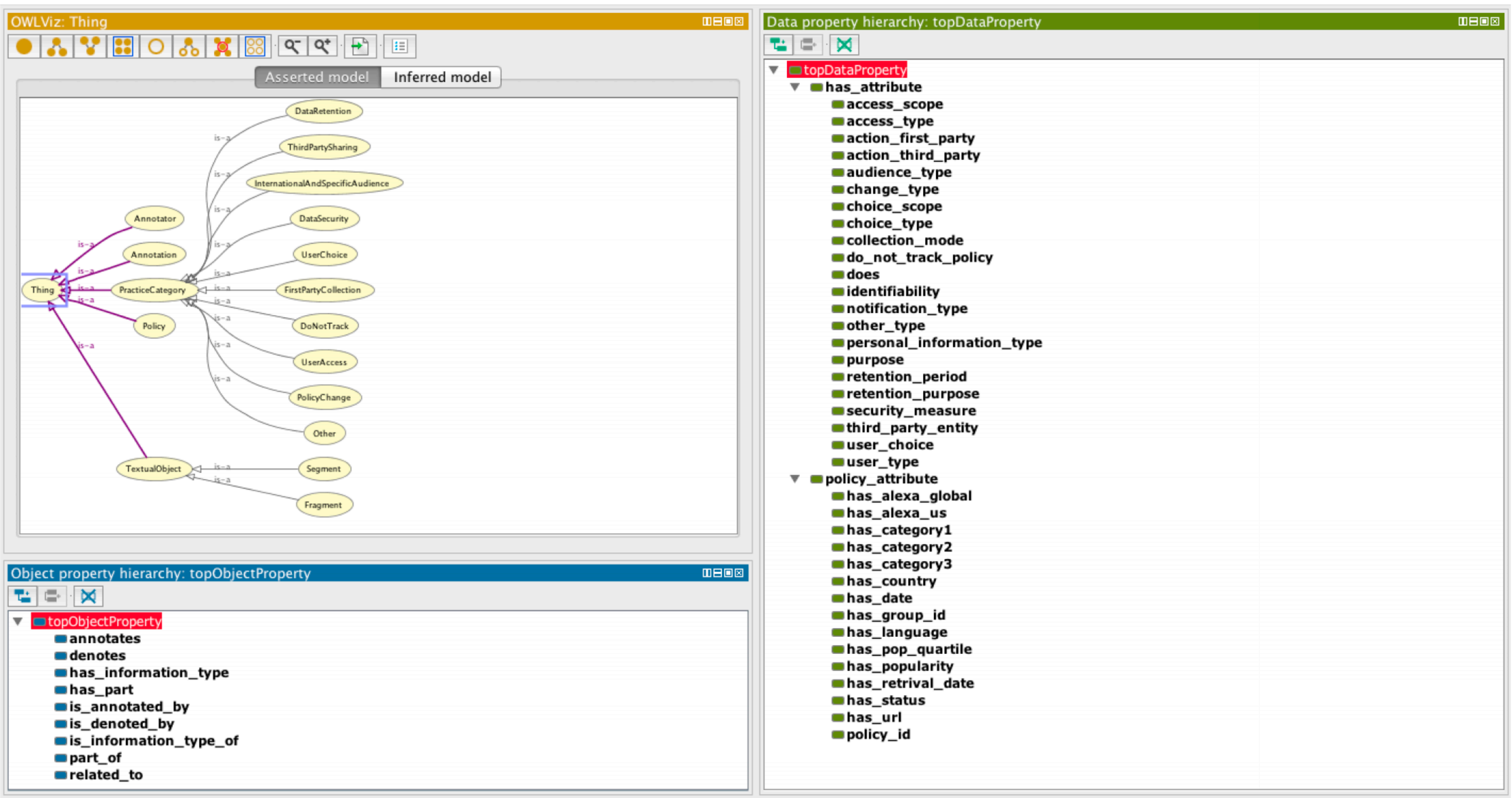

USABLEPRIVACY.ORG
EXPLORE
About
Browse Privacy Policies
Download the Data
Search for a website
New York Times nytimes.com
Fine Grained Policy
Scraped Policy 1
Privacy Practices
Click a category to filter practice statements.
First Party Collection/Use 21
Third Party Sharing/Collection 21
User Choice/Control 7
User Access, Edit and Deletion 2
Data Retention 2
Data Security 1
Policy Change 1
Do Not Track
onal and Specific Audiences 3
Choices 3
TH AD: 2) If you would like to opt-out of having interest-based ad targeting, click here .
FI CM: To subscribe or unsubscribe from The New York Times Events email newsletter, please visit www.nytimes.com/events .
TH AD: You have choices about the collection of information by third parties on our website: 1) If you would like more about your option not to accept advertiser cookies, please click here .
Privacy Policy obtained 7/8/2017 (1 version)
59 privacy statements and 3 choices detected
Reading Level: College Graduate (Grade 11)
sharing will be deemed to have been done by you, not The New York Times. Please see our Comments FAQ for additional information.
When you share or recommend links to content on a third-party platform (such as Facebook, Google+ and Twitter), that action and any information you share will be covered by their privacy policy.
Contests, Sweepstakes and Special Offers
The New York Times collects personal information from you when you participate in sweepstakes, contests or special offers. If this information is also being collected by a third party other than The New York Times, we will notify you at the same time. If you do not want any personal information shared, you should not participate in the sweepstakes, contest or special offer.
Reader Surveys, Reader Panels and Market Research
The New York Times may collect personal information from you in connection with voluntary surveys conducted via the NYT Services. Data may be collected through the NYT Services, on the phone or through the mail. The information you provide may be shared, but only in the aggregate, with advertisers and partners unless we notify you otherwise at the time of collection.
Members of our Reader Panels agree to participate in surveys, polls or discussions about their readership of The New York Times, their household/personal characteristics and their purchase behavior. Our panels are currently administered by Beta Research .
Conferences and Live Events
often receive information about attendees to our live events from sign-in and registration lists. We may this information with event or promotion sponsors, in which case we will notify you when we collect the mation.
Non-P nal Information Collected Using Technology
Informa n Collected by Us Using Technology
We use va Internet technologies to manage the NYT Services and track use of the Services. Non-personal inf ation that we collect using these technologies may be combined with other information about you.
Device Infor tion. We may collect non-personal information about the computer, mobile device or other device y se to access the NYT Services, such as IP address, geolocation information, unique device identifiers, bro r type, browser language and other transactional information.
Cookies, Bea s, Local Storage and Other Similar Technologies. We use "cookies," Web beacons, tags a scripts, and other similar technologies including local storage objects such as HTML5. These technol s allow us to manage access to and use of the Services, recognize you and provide personalizati nd help us understand how people use the NYT Services. You will not be able to access certain areas ur websites, including NYTimes.com, if your computer does not accept cookies from us.

PrivOnto: Semantic Reasoning

- Based on RDF/OWL language
  - Annotations mapped onto privacy ontology based on underlying taxonomy of practices
- "Query-able" through SPARQL
- Supports automatic inferences

Oltramari, Piraviperumal, Schaub, Wilson,Cherivirala, Norton, Russel, Story, Sadeh, Reidenberg, "PrivOnto: A Semantic Framework for the Analysis of Privacy Policies", Semantic Web Journal, 2017

PrivOnto Hierarchies of Classes



Sample Queries

Targeted information and related query types.

Targeted Information	Query example
Percentage	What percentage of policies apply to websites and mobile apps?
Count on Practices	How many practice statements per policy are unclear about where information are collected from users?
True or False	Is information shared or collected as part of a merger or acquisition?
Count on Policy	How many policies have statements on user choice?
Count on distribution of policies across values in practice category	For each of the security-measure values, how many websites mention them?

Queries on Info Collected from Users

Question	First Party Collection	% Policies	Third Party Collection	% Policies
Fragments that collect/share location information and for what purpose?	265	59.13	61	26.09
Fragments that collect/share contact information and for what purpose?	736	90.43	246	57.39
Fragments that collect/share device identifier and for what purpose?	319	76.52	75	25.22
What kind of Fragments are especially negated	199	67.83	313	78.26
Fragments that collect/share finance info and for what purpose?	231	63.48	102	35.65
Fragments that collect/share user's online activities info and for what purpose?	559	87.83	294	66.96
Fragments that collect/share user's general personal information info and for what purpose?	587	88.70	730	91.30
Fragments that collect/share user's unspecified info and for what purpose?	936	85.22	820	88.70

PrivOnto Demo

- Peter Story, PhD Student, School of Computer Science, CMU

Session I Recap

- Crowsourcing privacy policies is feasible but does not scale well
- NLP/ML can be used to improve crowd worker productivity but also has its limitations
- Automated extraction of privacy policies shows promise but is not 100% accurate
- These technologies open the door to new applications – from browser plug-ins to mobile app compliance tools (Session II)

- The **Usable Privacy Policy Project** and the **Personalized Privacy Assistant Project both** involve collaborations with a number of individuals.
- See **usableprivacy.org** and **privacyassistant.org** for additional details incl. lists of collaborators, publications, sponsors and recent news
- Subscribe to our mailing lists to stay up to date - https://usableprivacy.org/contact and https://www.privacyassistant.org/contact

Q&A

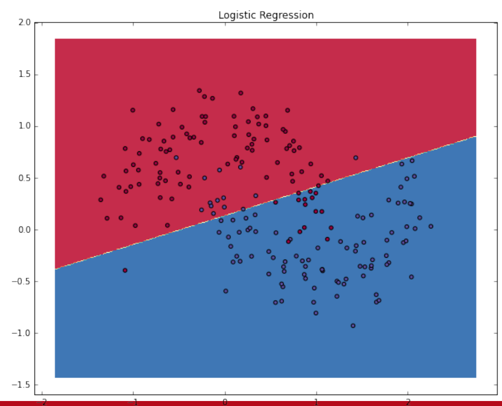
NLP Bonus Slides – Session I

Multinomial Naive Bayes

- Given an instance represented as a feature vector $\mathbf{x} = (x_1, \cdots, x_n)$, where x_i is the number of times the i th vocabulary occurs in the instance.
- Let $p_{ki} = \dfrac{N_{ki} + \alpha}{N_k + n\alpha}$ be the probability that the ith vocabulary occurs in class k. (alpha is the smoothing term to avoid zero probability)
- The label of the instance is set to $\arg\max_k p(C_k|\mathbf{x}) \propto \arg\max_k p(C_k) \prod_{i=1} p_{ki}^{x_i}$

Logistic Regression

- Given a set of instances (training data) and each instance is represented as a feature vector $\mathbf{x} = (x_1, \cdots, x_n)$, where x_i is the tf-idf of the i th vocabulary of the instance.
- We try to find a vector $\mathbf{w}$ such that it best separates the data.



Naive Bayes Logistic Regression

- Wang et al. (2012) showed that integrating Naive Bayes word counts into discriminative classifiers boost performance 1-2% in various datasets.
- Given a set of instances (training data) and each instance is represented as a feature vector $\mathbf{x} = (x_1, \cdots, x_n)$, where x_i is the binarized count of the i th vocabulary of the instance.
- Weight the features with a log-count ratio as new features

$$\mathbf{p} = \alpha + \sum_{i:y^{(i)}=1} \mathbf{x}^{(i)}$$

$$\mathbf{q} = \alpha + \sum_{i:y^{(i)}=-1} \mathbf{x}^{(i)}$$

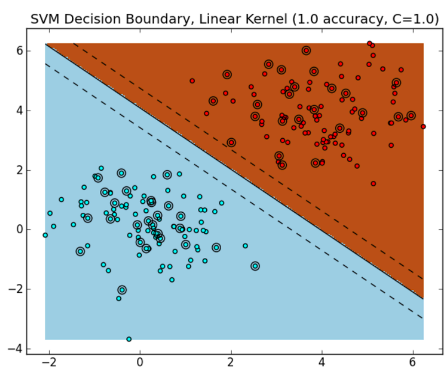
$$\mathbf{r} = \log\left(\frac{\mathbf{p}/||\mathbf{p}||_1}{\mathbf{q}/||\mathbf{q}||_1}\right)$$

$$\hat{\mathbf{x}} = \mathbf{r} \cdot \mathbf{x}$$

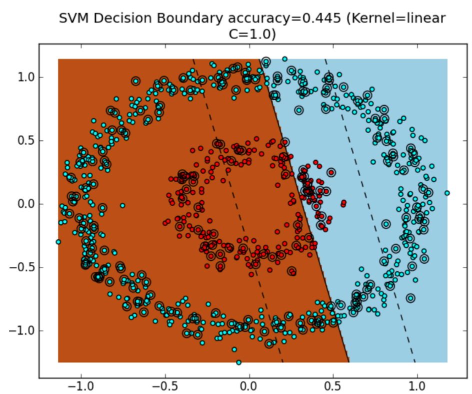
Wang et al. 2012

Support Vector Machines

- Similar to logistic regression
- Difference:
  - Max margin
  - Feature transformation (kernels)

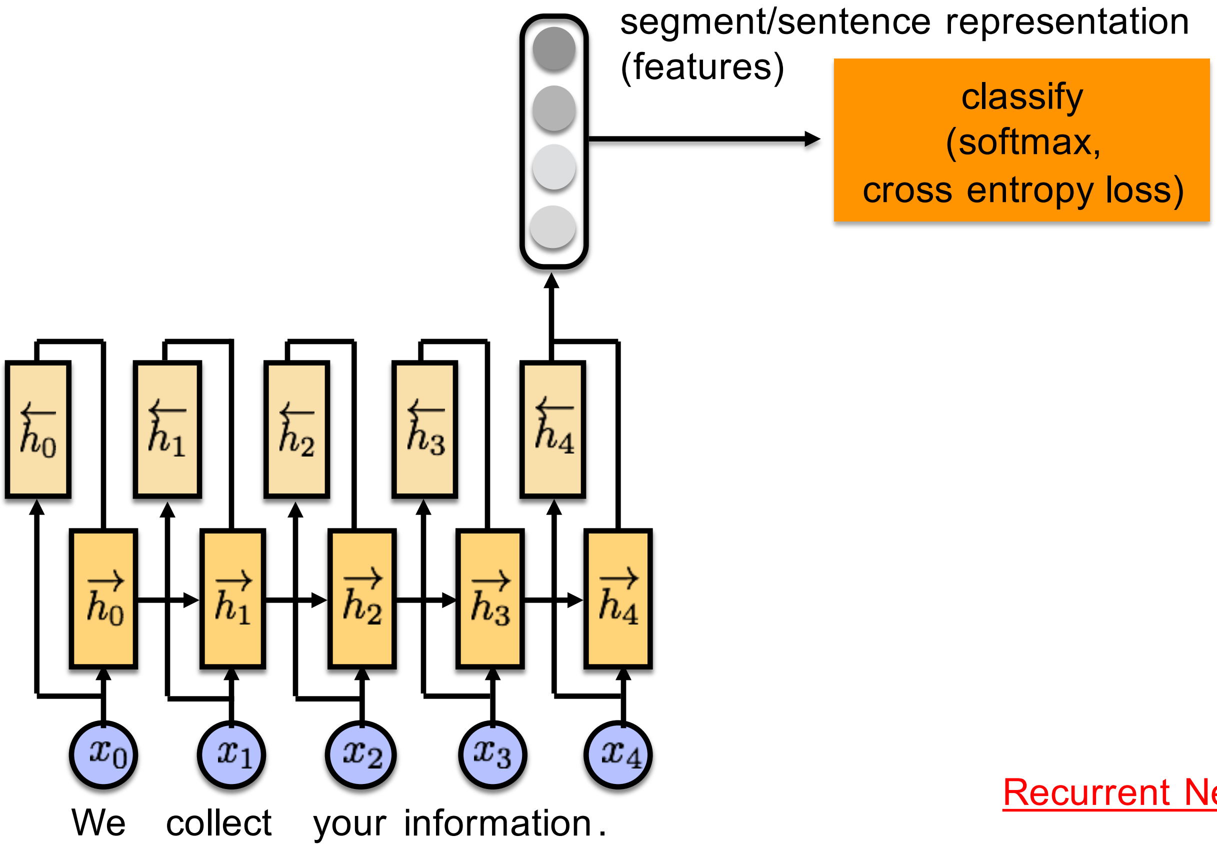


Max margin



kernel trick

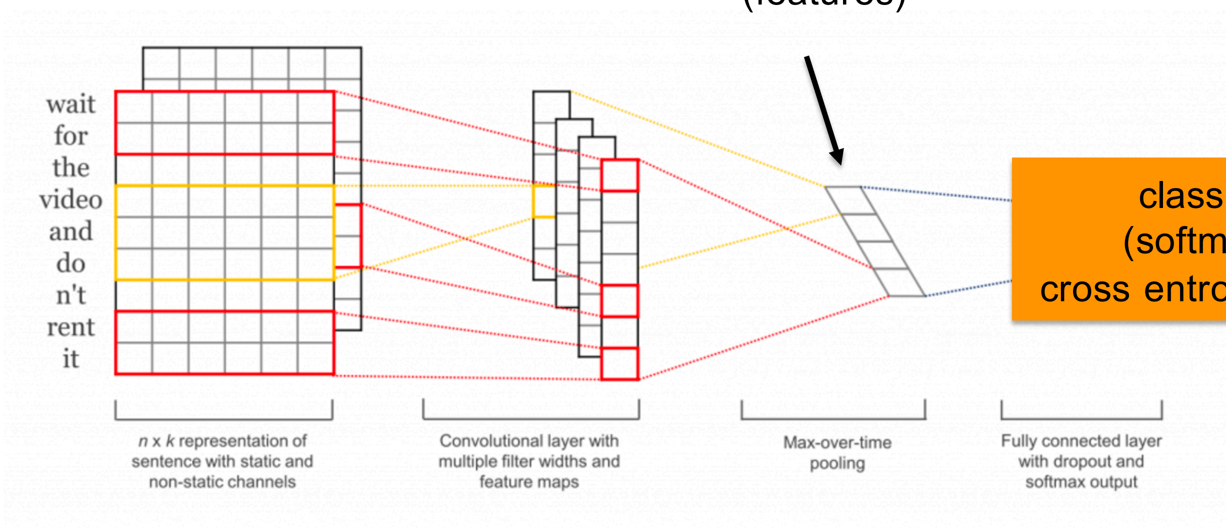
Recurrent Neural Networks



[Recurrent Nets]

Convolutional Neural Networks

segment/sentence representation (features)



classify (softmax, cross entropy loss)

Kim et al. 2014