

Usability of crypto API / online tracking

Mainack Mondal

CS 60081
Autumn 2021



Roadmap

- Passwords/multi factor authentications
- Usability for security developers
- Online tracking
- Privacy notices/dark patterns
- Temporal aspect of privacy

Roadmap

- Passwords/multi factor authentications
- Usability for security developers
- Online tracking
- Privacy notices/dark patterns
- Temporal aspect of privacy

Quick example of a study

Comparing the Usability of Cryptographic APIs

<https://www.cl.cam.ac.uk/~rja14/shb17/fahl.pdf>

Comparing the Usability of Cryptographic APIs

- <https://www.cl.cam.ac.uk/~rja14/shb17/fahl.pdf>
 - First a bit about crypto
 - Encryption
 - Decryption
 - Signatures
 - Hash
 - Now, almost no-one implement these, they use libraries
 - Library calls --> cryptographic APIs

Motivation

- Wanted to check if popular python crypto libraries are actually usable

		Sym		Asym						
		Key generation	Encryption	Key generation	Encryption	KDF	Digital sig.	X.509	Usability claims	Downloads
PyCrypto	[42]	●	●	●	●	●	●	●	○	25 149 446
cryptography.io	[8]	●	●	●	●	●	●	●	●	10 481 277
M2Crypto	[43]	●	●	●	●	●	●	●	○	2 369 827
Keyczar	[44]	●	●	●	●	○	●	○	●	595 277
PyNaCl	[45]	●	●	●	●	○	●	○	●	46 013

Recruitment

- Crawled all python repositories in Github
- Extracted emails
- Email them for taking part in the survey
- Got ~200 participants
- Ecological validity – why?

Contextualization

Asked participants to imagine they were developing code for an app called CitizenMeasure,

“a new global monitoring system that will allow citizen-scientists to travel to remote locations and make measurements about such issues as water pollution, deforestation, child labor, and human trafficking. Please keep in mind that our citizen-scientists may be operating in locations that are potentially dangerous, collecting information that powerful interests want kept secret. Our citizen scientists may have their devices confiscated and hacked.”

Methodology

- Randomly assigned tools to the developers
 - Between-subjects study
- Ask them to perform tasks online (py notebook)
 - Online study
 - Contextualization
- Qualitative analysis
 - Took the developer's solutions
 - Then two authors labeled them as functional, secure
 - Then they used statistics to measure usability!

Tasks

- Two symmetric encryption tasks
 - generating an encryption key and storing it securely in a password-protected file
 - using the key to encrypt and decrypt text
- Three asymmetric encryption tasks
 - generating a key pair and storing the private key securely
 - using the public key to encrypt and the private key to decrypt
 - validating an X.509 certificate.

Task example

Certificate validation

Goal: Verify that the SSL certificate from the central Citizen Measure server was issued by the Let's Encrypt Certificate Authority to ensure that citizen reports are not being intercepted. You have to validate the certificate's digital signature and common name. For your convenience, the SSL certificate from the Citizen Measure server is stored in `.citizenMeasureCertificate.pem` and the Let's Encrypt Certificate Authority certificate in `./leca.pem`. You can take also a look at the [Let's Encrypt X3 Root CA](#) and the [server certificate](#).

```
In [0]: 1 import nacl
2
3 def validate(certificate, root_certificate, hostname="citizen-measure.tk"):
4     """
5     Purpose:
6     Validate the given certificate's digital signature and common name.
7
8     Arguments:
9     certificate: The certificate to validate.
10    hostname: The server's hostname.
11
12    Return value:
13    validationresult: True if validating the certificate is correct, False otherwise.
14
15    Notes:
16    - The Citizen Measure server certificate can be found at ./citizenMeasureCertificate.pem
17    - The Let's Encrypt Certificate Authority certificate can be found at ./leca.pem
18    - If you used any other information source to solve this task than the linked documentation (e.g. a post on
19    StackOverflow, a blog post or a discussion in a forum), please provide the link right below:
20    - additional information sources go here (e.g. https://stackoverflow.com/questions/415511/how-to-get-current-time-in-
21    python)
22    """
23
24    # This is where your code goes
25    return False
26
27 # This is to test the code for this task.
28 certificate = open("./citizenMeasureCertificate.pem").read()
29 root_certificate = open("./leca.pem").read()
30 assert validate(certificate, root_certificate, "citizen-measure.tk"), "Certificate validation failed."
31 print "Task completed! Please continue."
```

Run and Test

Get unstick

NOT solved, Next Task

Solved, Next Task

Fig. 1. An example of the study's task interface.

Analysis: Regression

Factor	Description	Baseline
<i>Required factors</i>		
Library	The cryptographic library used.	PyCrypto
Encryption mode	Asymmetric or Symmetric	Symmetric
<i>Optional factors</i>		
Experienced	True if a programming in Python is part of participant's job, and/or if participant has been programming in Python for more than five years; otherwise false. Self-reported.	False
Security background	True or false, self-reported.	False
Library experience	Whether the participant has used the library before, seen code that used it but not used it themselves; or neither. Self-reported.	No experience
Copy-paste	Whether the participant pasted code during this task. Measured, per-task regressions only.	False
Library \times Mode	Interaction between the library and encryption mode factors described above.	cryptography.io :asymmetric

TABLE V

Factors used in regression models. Categorical factors are individually compared to the baseline. Final models were selected by minimum AIC; candidates were defined using all possible combinations of optional factors, with both required factors included in every candidate.

Result

Factor	O.R.	C.I.	<i>p</i>-value
M2Crypto	0.26	[0.09, 0.69]	0.007*
cryptography.io	1.68	[0.61, 4.61]	0.311
Keyczar	0.10	[0.04, 0.26]	< 0.001*
PyNaCl	1.58	[0.55, 4.56]	0.394
asymmetric	0.16	[0.07, 0.38]	< 0.001*
copy-paste	3.29	[1.97, 5.49]	< 0.001*
M2Crypto:asymmetric	8.14	[2.29, 28.95]	0.001*
cryptography.io:asymmetric	1.53	[0.4, 5.75]	0.532
Keyczar:asymmetric	1.50	[0.36, 6.22]	0.578
PyNaCl:asymmetric	0.49	[0.13, 1.86]	0.293

TABLE VIII

Results of the final logistic regression mixed model examining which factors correlate with task functionality. Odds ratios indicate relative likelihood of a task being functionally correct. Statistically significant values indicated with

*. See Section IV-B for further details.

Roadmap

- Passwords/multi factor authentications
- Usability for security developers
- **Online tracking**
- Privacy notices/dark patterns
- Temporal aspect of privacy

Behavioral targeting/tracking

Behavioral targeting/tracking

Scenario: You are visiting a website

- First party: the website your are visiting
- Second party: You
- Third party: Other sites the first site as a result of your visit to the site. Why will it happen?

Online tracking

- First party tracking
 - E.g., Google track your search results
- Solution: Use duckduckgo

- Stopping Third party tracking
 - Much harder...
 - But why would a third party track a user?

Do not track

- Proposed standard
- User checks a box
- Browser sends “do not track” to website
- Website stops “tracking”
 - What does that even mean?
 - cookies, javascript?
- Discontinued in apple, why?



Tools to stop tracking

- Browser privacy settings
 - Blocking cookies
 - P3P
- Browser extensions
- Opt-out cookies
- Digital Advertising Alliance (DAA) adchoices and associated opt-out pages

Extensions: Disconnect

The screenshot shows the Disconnect browser extension interface. At the top, it says "DISCONNECT" with "Help" and "Share" links. Below this are social media sharing buttons for Facebook (0), Google+ (1), and Twitter (1). The main interface is divided into several categories, each with a green icon and a dropdown arrow:

- Advertising**: 2 requests. Includes sub-items: Adobe (1 request) and Nielsen (1 request).
- Analytics**: 7 requests.
- Social**: 0 requests.
- Content**: 0 requests.

At the bottom, there are utility options: "Whitelist site" (with a link icon), "Visualize page" (with a bar chart icon), "Show counter" (checked), and "Cap counter" (checked). Below these are two bar charts labeled "Time saved" and "Bandwidth saved". At the very bottom is a green button that says "Get Mobile Protection".

The screenshot shows a popup window for the Disconnect extension on the website espn.com. The popup has a blue header with the site name "espn.com" and a close button (X). Below the header, it says "8 trackers blocked".

The main content area shows "Tracker blocking is **on** for this website" with a green toggle switch. Below this is a list of blocked trackers:














Google AdSense	blocked	🟢
Demdex	blocked	🟢
Twitter Badge	blocked	🟢
Omniture	blocked	🟢

Below the list is a red eye icon and a link: "see your tracker blocking stats and learn more about these companies".

A blue bar at the bottom of the popup displays "21 trackers blocked since Feb '17".





At the very bottom, there is a white bar with the BLUR logo, "Settings" and "Help" links, and a "Go Premium" button.

Extensions: Ublock origin

+ all	
	gizmodo.com ++ -
	d1alzvr683o8p8.cloudfront.net -
	doubleclick.net -
	google-analytics.com -
	google.com -
	indexww.com -
	kinja-img.com ++
	kinja-static.com ++
	kinja.com +
	media.net -
	privacy-mgmt.com +
	sp-prod.net +
	speedcurve.com -







gizmodo.com



   **Aa**⁹ ⁴²

Blocked on this page
13 (18%)

Domains connected
6 out of 13

Blocked since install
3.073M (8%)

More   Less

Browser fingerprinting

- Use features of your browser that are relatively unique to your machine
 - Fonts
 - GPU model anti aliasing (canvas fingerprinting)
 - User agent string
 - IP is often not used (why?)

Browser fingerprinting

- Use features of your browser that are relatively unique to your machine
 - Fonts
 - GPU model anti aliasing (canvas fingerprinting)
 - User agent string
 - IP is often not used (why?)

Check: <https://panopticlick.eff.org/>

Browser fingerprinting

Test	Result
Is your browser blocking tracking ads?	✓ yes
Is your browser blocking invisible trackers?	✓ yes
Does your browser unblock 3rd parties that promise to honor Do Not Track ?	✗ no
Does your browser protect from fingerprinting ?	✗ your browser has a unique fingerprint

Note: because tracking techniques are complex, subtle, and constantly evolving, Panopticlick does not measure all forms of tracking and protection.

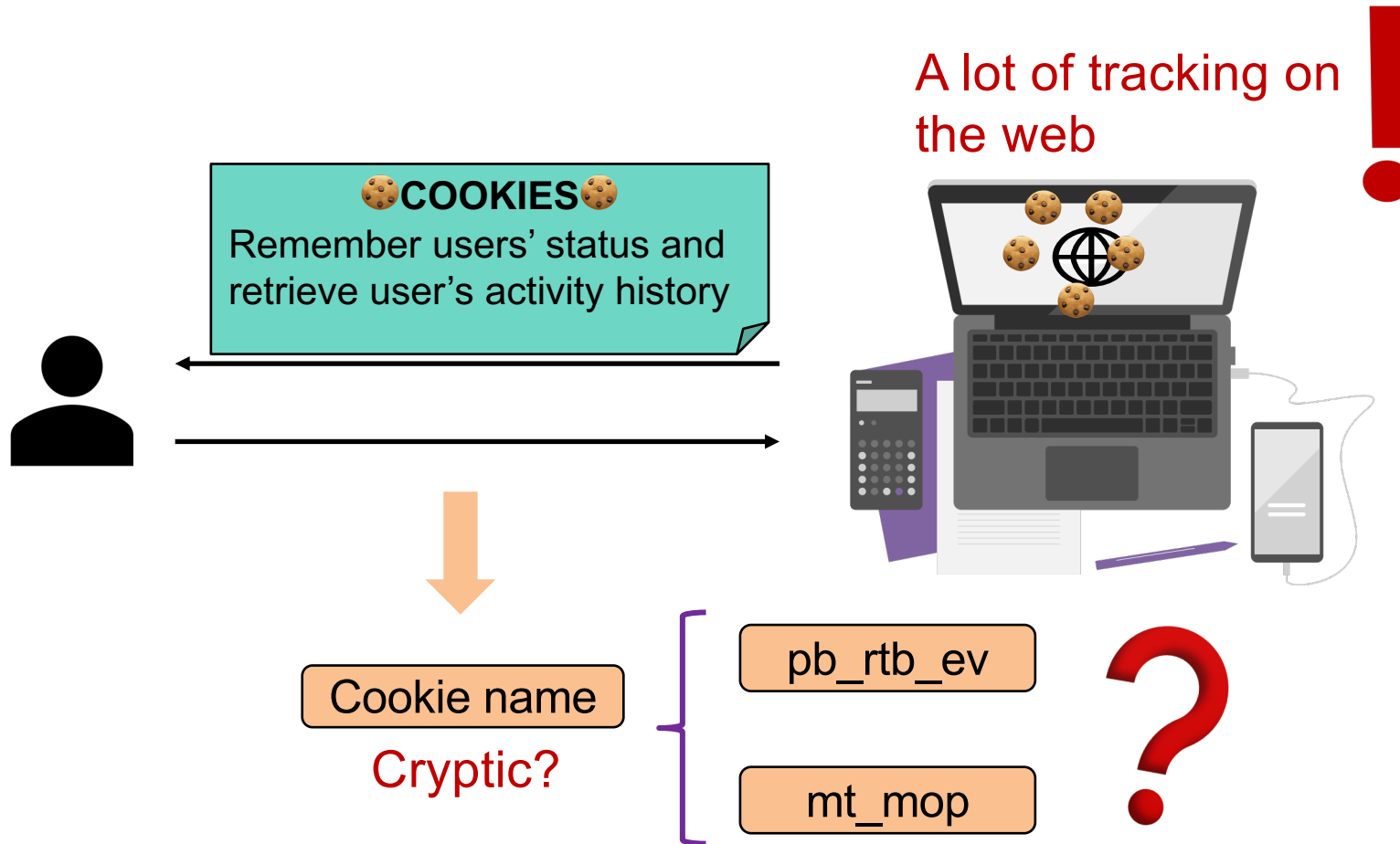
Your browser fingerprint **appears to be unique** among the 259,558 tested in the past 45 days.

Currently, we estimate that your browser has a fingerprint that conveys **at least 17.99 bits of identifying information**.

Browser fingerprinting

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	17.99	259558.0	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.113 Safari/537.36
HTTP_ACCEPT Headers	17.99	259558.0	text/html, */*; q=0.01 gzip, deflate, br en-US,en;q=0.9,bn;q=0.8,de;q=0.7
Browser Plugin Details	3.27	9.66	Plugin 0: Chrome PDF Plugin; Portable Document Format; internal-pdf-viewer; (Portable Document Format; application/x-google-chrome-pdf; pdf). Plugin 1: Chrome PDF Viewer; ; mhjfbmdgcfjbbpaeofofohoefglehjal; (; application/pdf; pdf). Plugin 2: Native Client; ; internal-nacl-plugin; (Native Client Executable; application/x-nacl;) (Portable Native Client Executable; application/x-pnacl;).
Time Zone Offset	4.58	23.99	-330
Time Zone	4.99	31.69	Asia/Calcutta
Screen Size and Color Depth	6.32	79.64	1280x800x24
System Fonts	8.12	279.09	Andale Mono, Arial, Arial Black, Arial Hebrew, Arial Narrow, Arial Rounded MT Bold, Arial Unicode MS, Book Antiqua, Bookman Old Style, Calibri, Cambria, Cambria Math, Century, Century Gothic, Century Schoolbook, Comic Sans MS, Consolas, Courier, Courier New, Geneva, Georgia, Helvetica, Helvetica Neue, Impact, Lucida Bright, Lucida Calligraphy, Lucida Console, Lucida Fax, LUCIDA GRANDE, Lucida Handwriting, Lucida Sans, Lucida Sans Typewriter, Lucida Sans Unicode, Microsoft Sans Serif, Monaco, Monotype Corsiva, MS Gothic, MS PGothic, MS Reference Sans Serif, Palatino, Palatino Linotype, Tahoma, Times, Times New Roman, Trebuchet MS, Verdana, Wingdings, Wingdings 2, Wingdings 3 (via javascript)
Are Cookies Enabled?	0.26	1.2	Yes
Limited supercookie test	1.53	2.89	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No, openDatabase: true, indexed db: true
Hash of canvas fingerprint	10.06	1068.14	e1cbad0c871db716d5068dc064815a2f
Hash of WebGL fingerprint	16.99	129779.0	5602af4402f28042575176f5bc1314a9

Tracking of Web Cookies



CCCC: Corralling Cookies into Categories with *CookieMonster*

Xuehui Hu

King's College London

Nishanth Sastry

University of Surrey

Mainack Mondal

Indian Institute of
Technology Kharagpur

Slides borrowed from: Xuehui Hu

Categories of cookies

Currently, the commonly used classification in the GDPR notification banner of the website is the one proposed by the UK International Chamber of Commerce (UK ICC)

useful for user

Strictly Necessary cookies e.g., login info

Functional cookies e.g., preferred language or location

Performance cookies e.g., count page visits, idle time on a page, bounce rates

Targeting/Advertising cookies e.g., profile users and serve them ads

useful for site

Categories of cookies

Currently, the commonly used classification in the GDPR notification banner of the website is the one proposed by the UK International Chamber of Commerce (UK ICC)

useful for user

Strictly Necessary cookies e.g., login info

Functional cookies e.g., preferred language or location

Performance cookies e.g., count page visits, idle time on a page, bounce rates

Targeting/Advertising cookies e.g., profile users and serve them ads

useful for site

cookies are used for targeted advertising and other purposes not useful to user

Categories of cookies

Currently, the commonly used classification in the GDPR notification banner of the website is the one proposed by the UK International Chamber of Commerce (UK ICC)

useful for user

Strictly Necessary cookies e.g., login info

Functional cookies e.g., preferred language or location

Performance cookies e.g., count page visits, idle time on a page, bounce rates

Targeting/Advertising cookies e.g., profile users and serve them ads

useful for site

cookies are used for targeted advertising and other purposes not useful to user

How many cookies of different kinds exist?

Our Goal

Accurate automatic classification of cookies in
the wild

Privacy Preference Center



Your Privacy

Strictly Necessary Cookies

Always Active

Strictly Necessary Cookies

which cookies work well for the user

These cookies are necessary for the website to function and cannot be switched off in our systems. They are usually only set in response to actions made by you which amount to a request for services, such as setting your privacy preferences, logging in or filling in forms. You can set your browser to block or alert you about these cookies, but some parts of the site will not then work. These cookies do not store any personally identifiable information.

Performance Cookies

Functional Cookies

which cookies for the benefit of the site operator

Targeting Cookies

Powered by **OneTrust**

Confirm My Choices

Allow All

Cookiepedia – Alexa global top20K

! Cookiepedia is incomplete!

COOKIE CATEGORY	#COOKIES	PROP.
Strictly Necessary	3,071	5.61%
Functionality	1,102	2.01%
Performance	3,025	5.53%
Targeting/Advertising	4,380	8.01%
Unknown	19,007	34.75%
Nonexistent	24,108	44.08%
Unknown+Nonexistent	43115	78.83%
Total	54,694	100%

<22% of cookies are in the database

<15% of the cookies set in the browsers of a set of real users can be categorised

Cookies Search Results: recent_history

About this cookie:
There is not yet any general information about this cookie based on its name only. If you have any information about this cookie, please get in touch.
The main purpose of this cookie is: **Unknown**

Cookiepedia—Alexa global top20K

Cookies Search Results: recent_history

About this cookie:

There is not yet any general information about this cookie based on its name only. If you have any information about this cookie, [please get in touch](#).

The main purpose of this cookie is: **Unknown**

COOKIE CATEGORY	#COOKIES	PROP
Unknown/Nonexistent	49,115	78.85%
Total	54,694	100%

can be categorised

the
ers

Cookiepedia – Alexa global top20K

! Cookiepedia is incomplete!

COOKIE CATEGORY	#COOKIES	PROP.
Strictly Necessary	3,071	5.61%
Functionality	1,102	2.01%
Performance	3,025	5.53%
Targeting/Advertising	4,380	8.01%
Unknown	19,007	34.75%
Nonexistent	24,108	44.08%
Unknown+Nonexistent	43115	78.83%
Total	54,694	100%

<22% of cookies are in the database

<15% of the cookies set in the browsers of a set of real users can be categorised

[Cookies Search Results: recent_history](#)

About this cookie:

There is not yet any general information about this cookie based on its name only. If you have any information about this cookie, [please get in touch](#).

The main purpose of this cookie is: **Unknown**

[Cookies Search Results: lms_analytics](#)

Sorry, your search returned no matches

Try entering a different search term or [return to the homepage](#).

Cookiepedia—Alexa global top20K

COOKIE CATEGORY

#COOKIES

PROP.

Cookies Search Results: Ims_analytics

Sorry, your search returned no matches

Try entering a different search term or [return to the homepage](#).

Nonexistent	24,108	44.08%
Unknown+Nonexistent	43115	78.83%
Total	54,694	100%

browsers of a set of real users
can be categorised

Cookiepedia—Alexa global top20K

COOKIE CATEGORY	#COOKIES	PROP.
-----------------	----------	-------

recent_history → useful to user (functional)
Lms_analytics → useful to website (performance)

Cookies Search Results

About this cookie:
There is not yet any general information about any information about this cookie, please get
The main purpose of this cookie is: **Unknown**

Cookies Search Results: lms_analytics

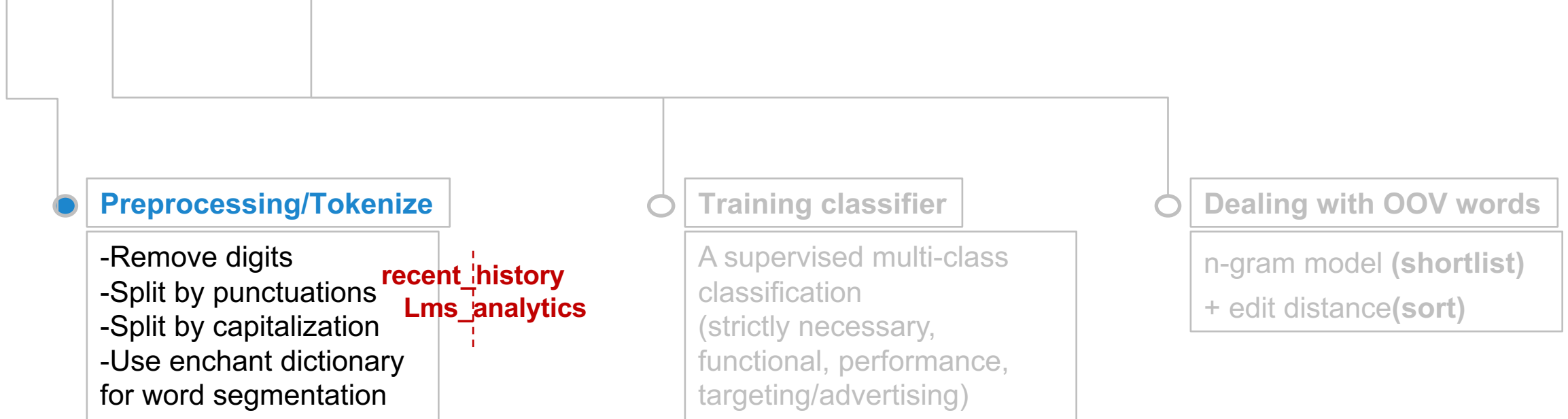
Sorry, your search returned no matches
Try entering a different search term or return to the homepage.

NONEXISTENT	24,108	44.08%
Unknown+Nonexistent	43115	78.83%
Total	54,694	100%

the browsers of a set of real users can be categorised

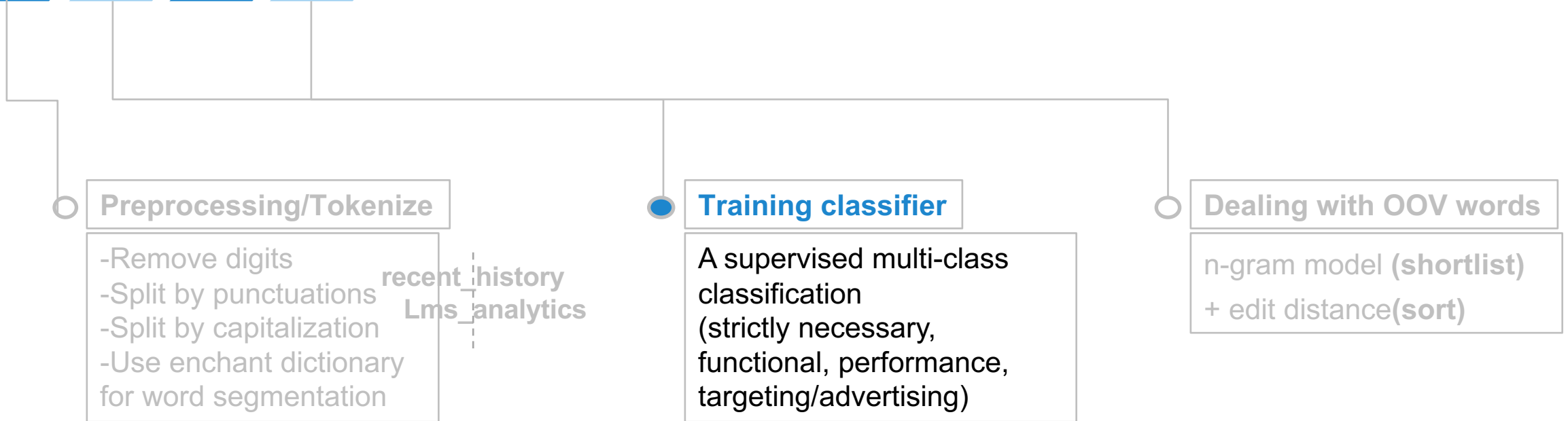
CookieMonster Design

Cookie Name



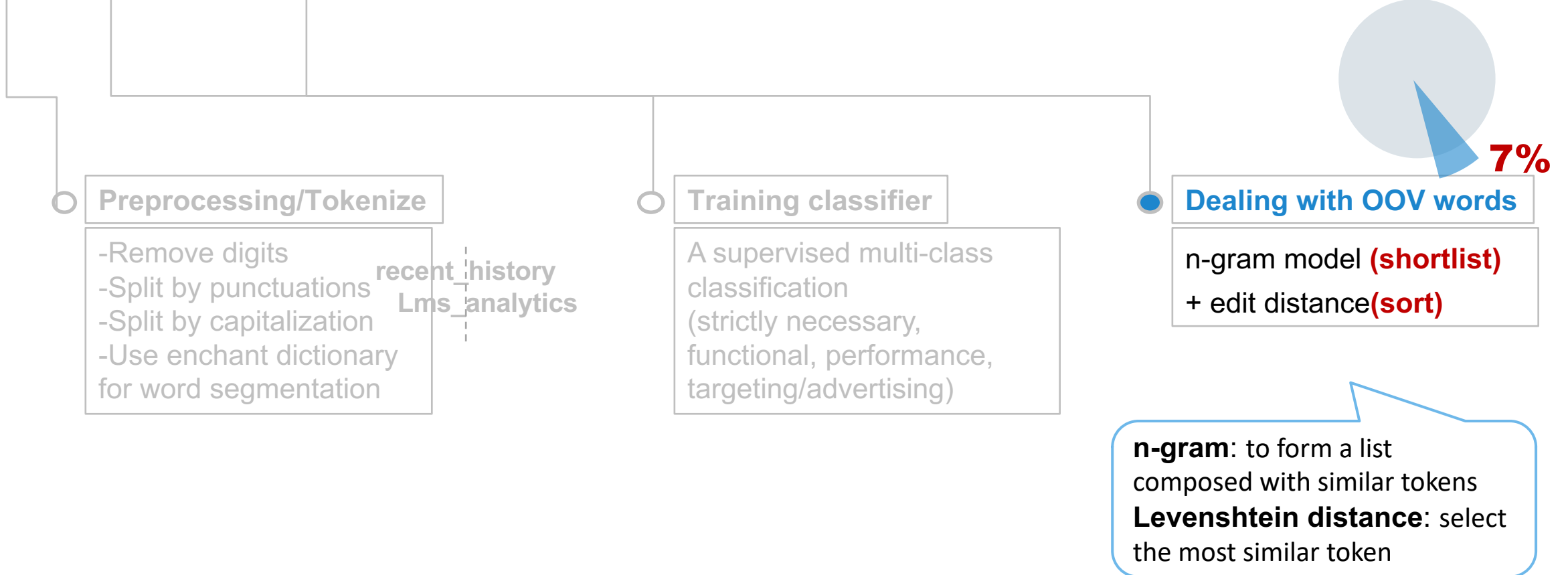
CookieMonster Design

Cookie Name



CookieMonster Design

Cookie Name



Performance (training dataset)

ALGORITHM	PRECISION	RECALL	F1	PREDICTION LATENCY (MS)
Multinomial Naïve Bayes (MNB)	0.951	0.94	0.9458	0.44
Softmax Regression (MLP)	0.944	0.948	0.9457	1.29
SVM	0.947	0.867	0.926	0.03
K-Nearest Neighbors (KNN)	0.929	0.907	0.916	3.23
Random Forest	0.886	0.77	0.778	9.73
Naïve Bayes	0.798	0.747	0.833	0.02
Binary Search Tree (BST)	0.649	0.461	0.409	0.05

Use pre-trained MNB model in *CookieMonster*

Tracking in social media

- Go to https://www.facebook.com/adpreferences/ad_settings
- Then “Categories used to reach you” → “Interest categories”

Tracking in social media

- Go to https://www.facebook.com/adpreferences/ad_settings
- Then “Categories used to reach you” → “Interest categories”

Removing yourself from an interest category prevents advertisers from reaching you by indicating that their ads should be shown to people in that specific interest category. It doesn't affect the number of ads you see overall. We may still show you ads related to these categories if we think these ads may be relevant to you.

Online degree	Remove
Entrepreneurship	Remove
Shapoorji Pallonji Group	Remove
Data science	Remove
Veganism	Remove
Netflix	Remove
Association football (Soccer)	Remove