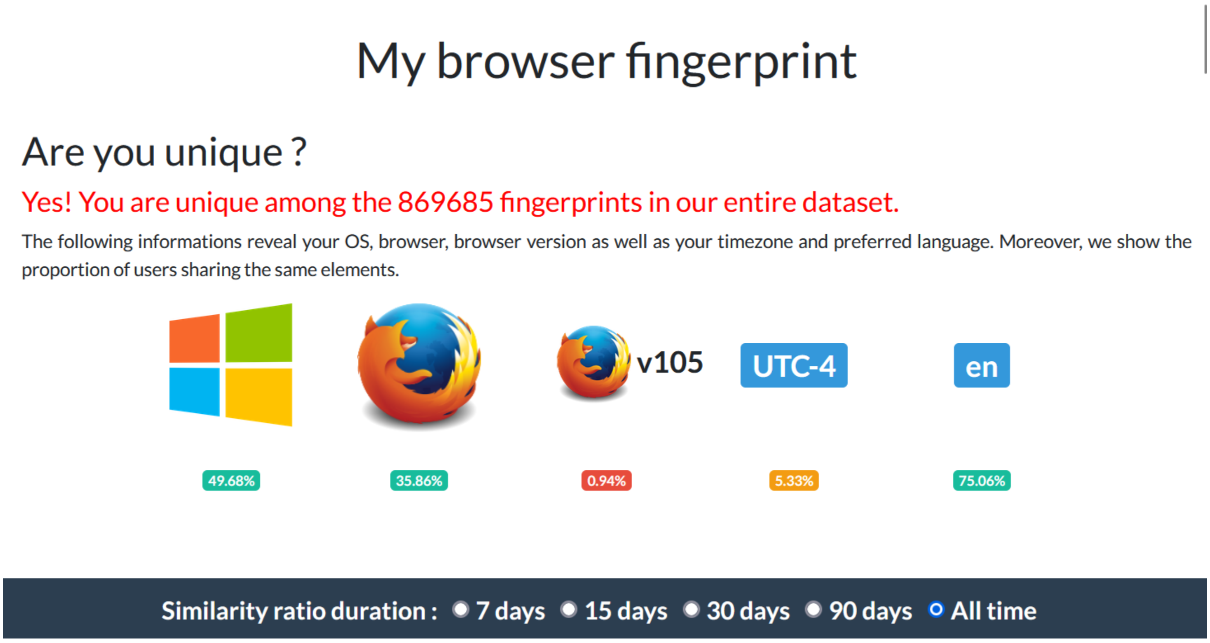
Homework-3

CSC 533 || Privacy in Digital Age

Question 1)

(a) **Major Differences between two finger browser-prints:** Here, the two browsers used are Firefox and Tor browser. The screenshots below highlight key differences between the two fingerprints. They are numbered in the same order, for both categories, that is, point (1) for Firefox and point (1) for Tor correspond to the same portion of the fingerprint, and so on, respectively.

- Mozilla Firefox:



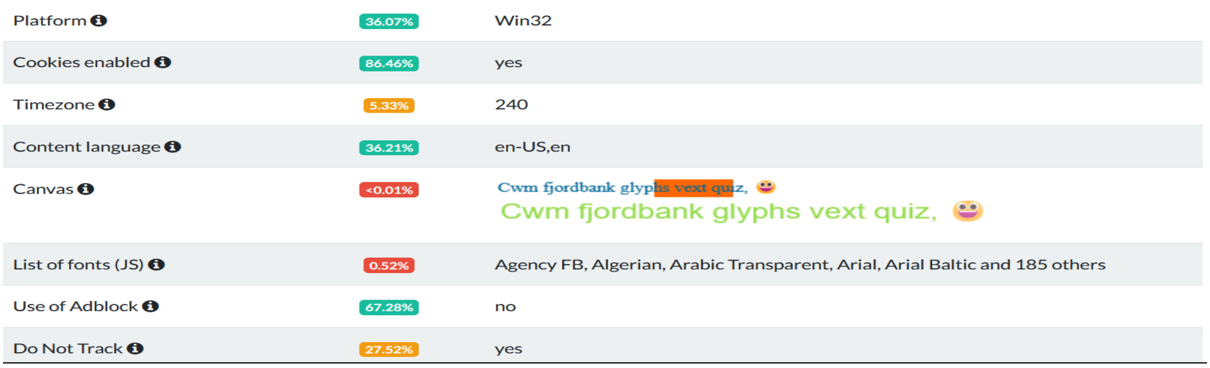
My browser fingerprint

Are you unique ?

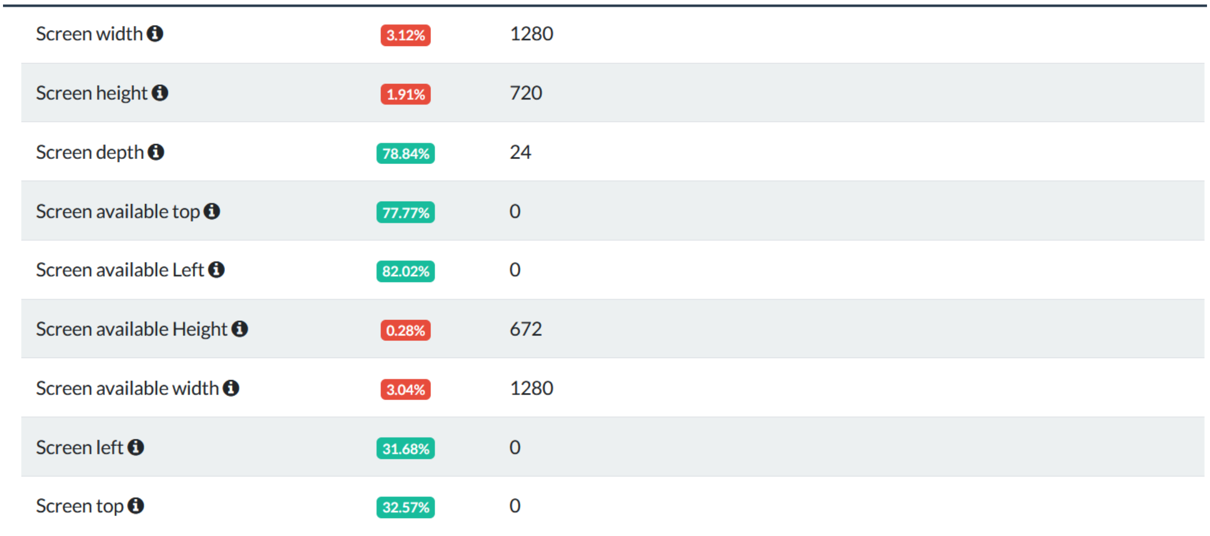
Yes! You are unique among the 869685 fingerprints in our entire dataset.

The following informations reveal your OS, browser, browser version as well as your timezone and preferred language. Moreover, we show the proportion of users sharing the same elements.

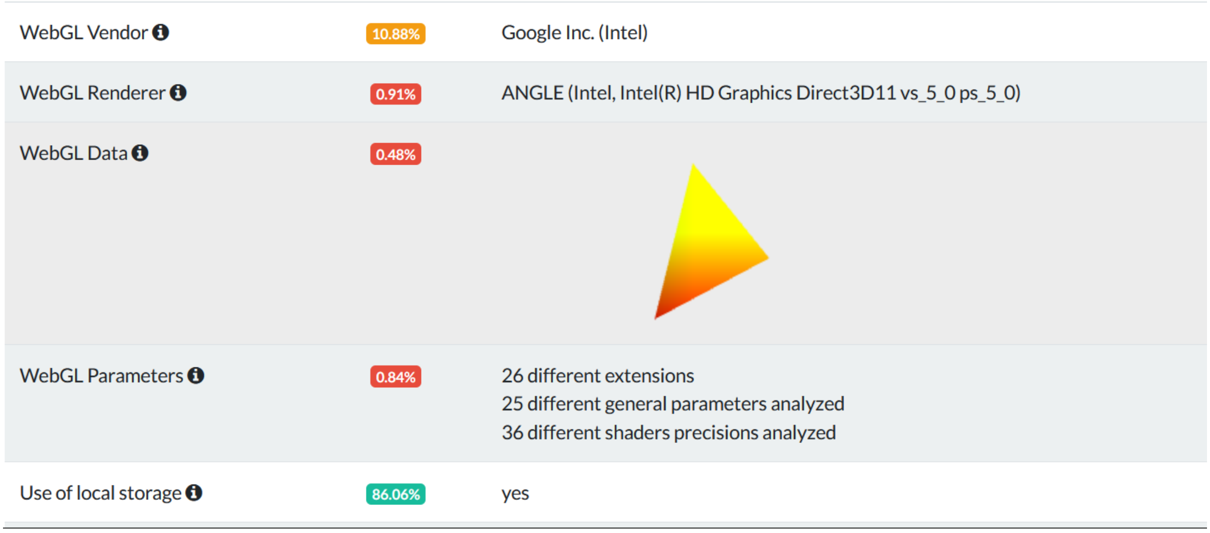
Similarity ratio duration : ○ 7 days ○ 15 days ○ 30 days ○ 90 days ● All time



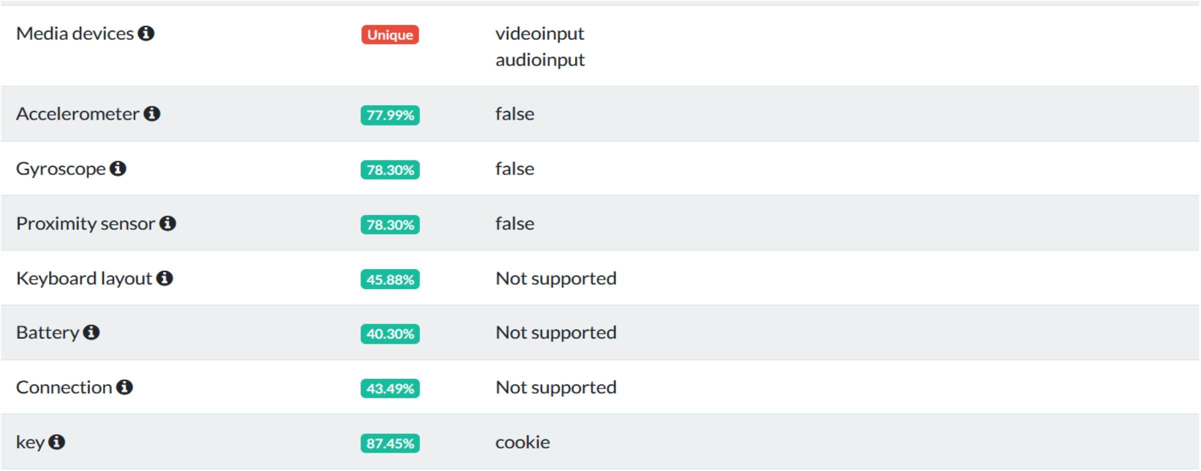
Platform	36.07%	Win32
Cookies enabled	86.46%	yes
Timezone	5.33%	240
Content language	36.21%	en-US,en
Canvas	<0.01%	Cwm fjordbank glyphs vext quiz, 😀 Cwm fjordbank glyphs vext quiz, 😀
List of fonts (JS)	0.52%	Agency FB, Algerian, Arabic Transparent, Arial, Arial Baltic and 185 others
Use of Adblock	67.28%	no
Do Not Track	27.52%	yes



Screen width ℹ	3.12%	1280
Screen height ℹ	1.91%	720
Screen depth ℹ	78.84%	24
Screen available top ℹ	77.77%	0
Screen available Left ℹ	82.02%	0
Screen available Height ℹ	0.28%	672
Screen available width ℹ	3.04%	1280
Screen left ℹ	31.68%	0
Screen top ℹ	32.57%	0

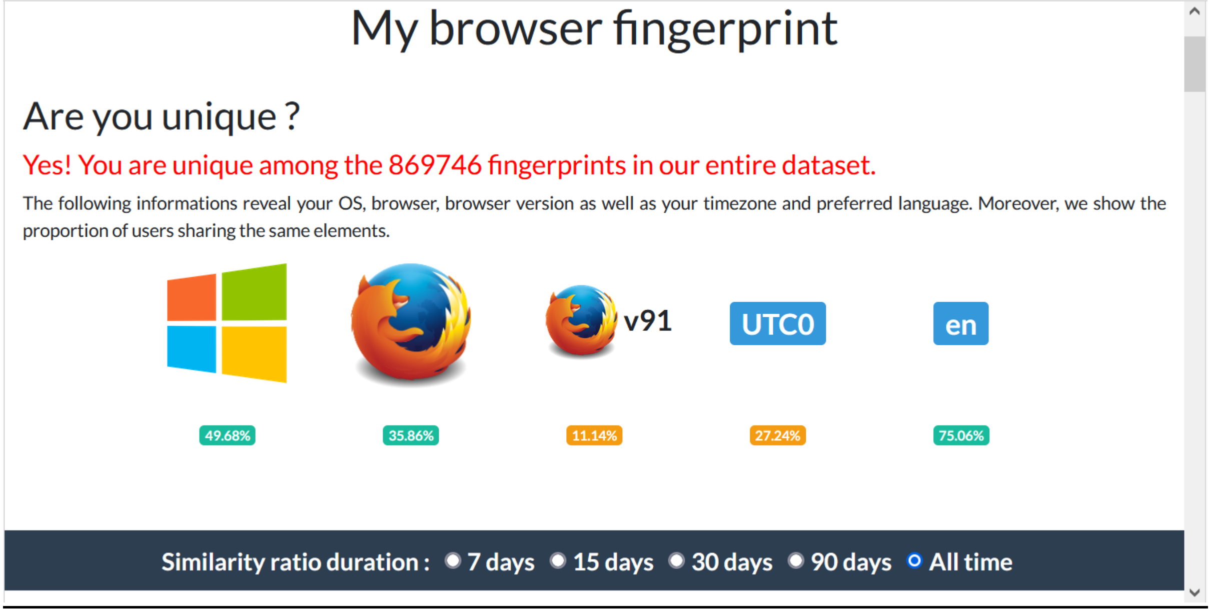


WebGL Vendor ℹ	10.88%	Google Inc. (Intel)
WebGL Renderer ℹ	0.91%	ANGLE (Intel, Intel(R) HD Graphics Direct3D11 vs_5_0 ps_5_0)
WebGL Data ℹ	0.48%	
WebGL Parameters ℹ	0.84%	26 different extensions 25 different general parameters analyzed 36 different shaders precisions analyzed
Use of local storage ℹ	86.06%	yes



Media devices ℹ	Unique	videoinput audioinput
Accelerometer ℹ	77.99%	false
Gyroscope ℹ	78.30%	false
Proximity sensor ℹ	78.30%	false
Keyboard layout ℹ	45.88%	Not supported
Battery ℹ	40.30%	Not supported
Connection ℹ	43.49%	Not supported
key ℹ	87.45%	cookie

- Tor:



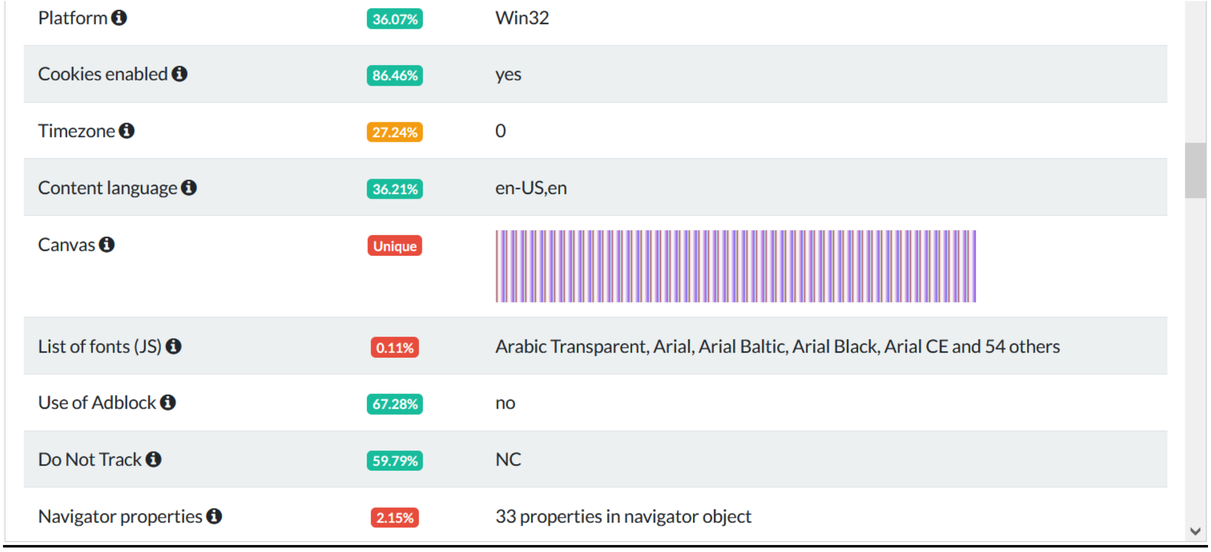
My browser fingerprint

Are you unique ?

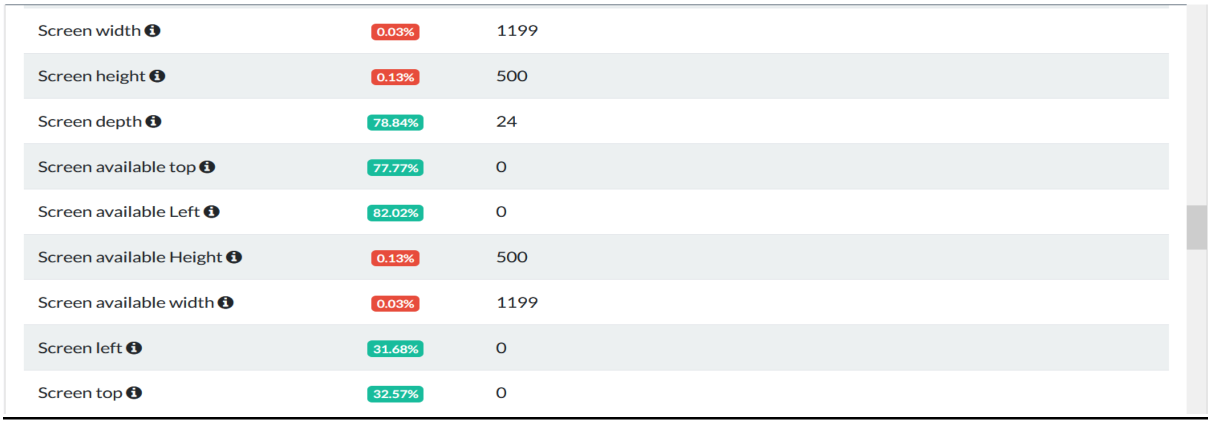
Yes! You are unique among the 869746 fingerprints in our entire dataset.

The following informations reveal your OS, browser, browser version as well as your timezone and preferred language. Moreover, we show the proportion of users sharing the same elements.

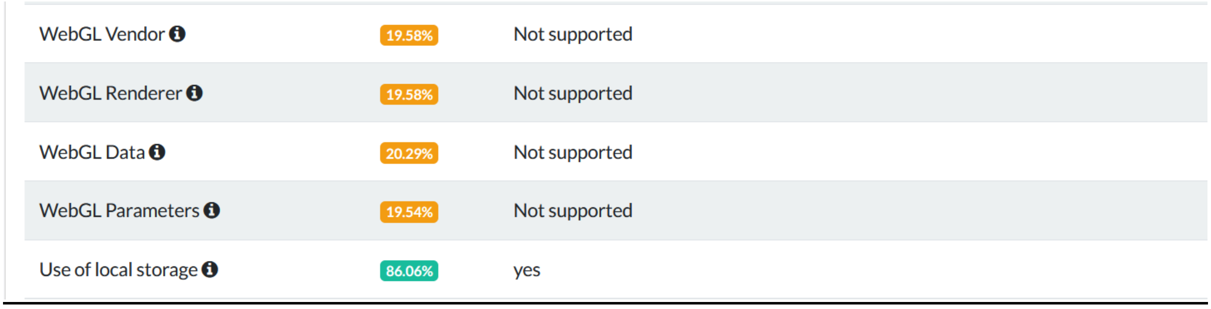
Similarity ratio duration : ○ 7 days ○ 15 days ○ 30 days ○ 90 days ● All time



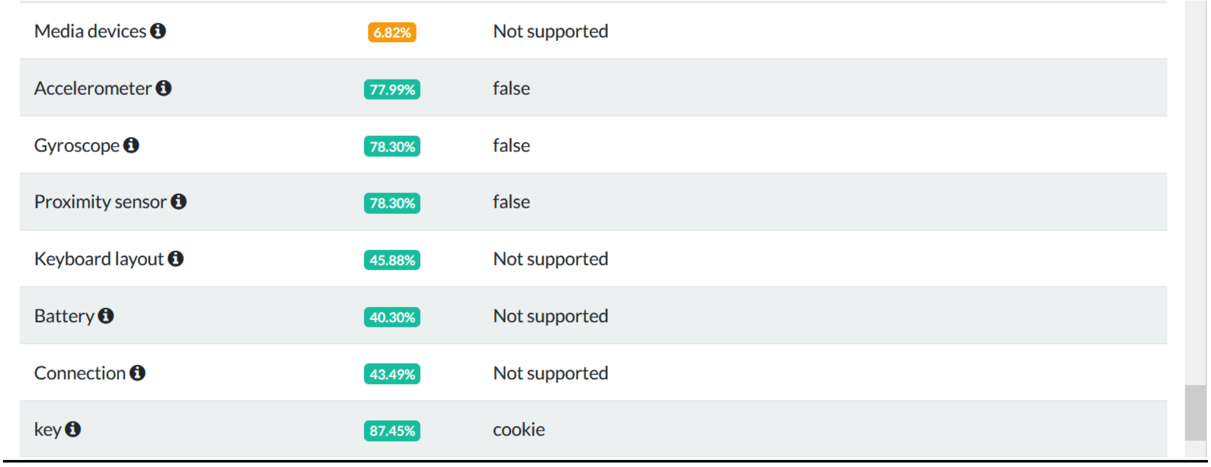
Platform	36.07%	Win32
Cookies enabled	86.46%	yes
Timezone	27.24%	0
Content language	36.21%	en-US,en
Canvas	Unique	
List of fonts (JS)	0.11%	Arabic Transparent, Arial, Arial Baltic, Arial Black, Arial CE and 54 others
Use of Adblock	67.28%	no
Do Not Track	59.79%	NC
Navigator properties	2.15%	33 properties in navigator object



Screen width	0.03%	1199
Screen height	0.13%	500
Screen depth	78.84%	24
Screen available top	77.77%	0
Screen available Left	82.02%	0
Screen available Height	0.13%	500
Screen available width	0.03%	1199
Screen left	31.68%	0
Screen top	32.57%	0



WebGL Vendor	19.58%	Not supported
WebGL Renderer	19.58%	Not supported
WebGL Data	20.29%	Not supported
WebGL Parameters	19.54%	Not supported
Use of local storage	86.06%	yes



Media devices	6.82%	Not supported
Accelerometer	77.99%	false
Gyroscope	78.30%	false
Proximity sensor	78.30%	false
Keyboard layout	45.88%	Not supported
Battery	40.30%	Not supported
Connection	43.49%	Not supported
key	87.45%	cookie

Analysis: There are five screenshots used that highlight major differences when it comes to fingerprinting, for which we have the following five points:

1. Browser Version and Time Zone (First Snapshot): The browser version for Firefox is 0.94% similar, while for Tor, it is 11.14% similar. This is a huge difference, which could imply many things. One conclusion is that a very few number of people use the current version of Firefox, making it highly identifiable, and lesser number of people use this particular version of Tor (built on top of Firefox, which is why it shows Firefox), which makes it less identifiable. Another can be that Tor is updated regularly, since it is used by people who are more concerned with privacy than Firefox, so using a newer version in Firefox might make one more identifiable. Without further data, these conclusions cannot be validated to a certainty.
   Time Zone has 5.33% similarity index for Firefox while a 27.24% for Tor. This difference is also significant, which can imply that the number of people using Firefox in this particular time zone is lesser than the percentage of people using Tor, for the same.

2. Canvas, List of Fonts, Do not Track (Second Snapshot): Canvas is highly unique for Firefox (<0.01%) and completely unique in Tor, which is not desirable. It could imply that from graphical rendering of html 5 canvas, one could completely identify Tor, and to a large degree, Firefox. List of Fonts are 0.52 and 0.11 on similarity scale, which means that JS could get more test attributes on Firefox, based on installed list

of fonts for fingerprinting in Firefox than on Tor. Also, list of fonts used in Firefox is higher, which could also be a factor for higher similarity. The “Do not track” attribute has 27.52 % and 59.79% similarity indices for Firefox’s DNT signal setting set to “Standard” by default, and Tor’s set to “NC”, by default, which means that it’s hard to fingerprint Tor based on its default setting than it is to fingerprint Tor. In this aspect, Tor seems to be a clear winner (in absence of more data).

3. Screen Width, Screen Height, Screen Available Width, Screen Available Height (Third Snapshot): The respective values are {3.12,1.91,0.28,3.04} for Firefox and {0.03,0.13,0.13,0.03} for Tor. Based on this, we can say that based on screen dimensions and pixels of free space available on window of browser, Tor is more identifiable or easily finger printable than Firefox. It is a common issue, which is why it’s always recommended to never maximize your Tor window.

4. WebGl Vendor,Renderor,Parameters, Data (Snapshot 4): The similarity indices for the four attributes for Firefox ={0.88,0.91,0.48,0.84} and Tor={19.58,19.58,19.54,20.29}. Based on this, it seems that it’s much more easier to fingerprint Firefox from the information regarding the graphics card, and for Tor, this information is not directly accessible (as seen from the right most column, value =”Not Supported”).

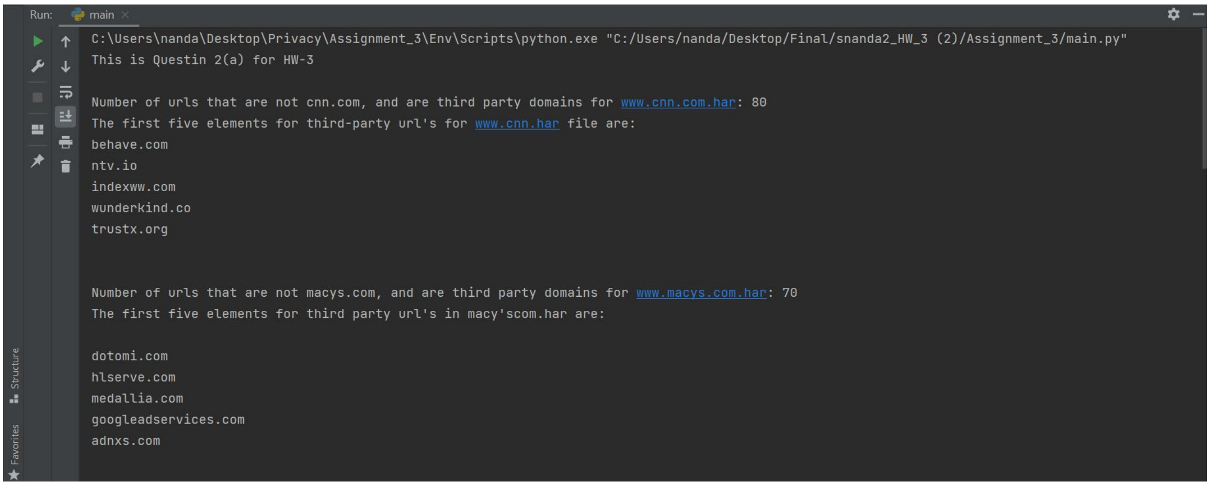
5. Media Devices (Snapshot 5): The similarity indices for media devices in Firefox and Tor are “unique” and 6.82%. This difference seems significant, and seems to indicate that it’s easier to fingerprint Firefox based on hardware attributes like cameras, microphone array, etc, than in a Tor browser.

(b) Basic Takeaway in terms of tracking capability of two browsers: From above, the takeaway seems to be, that while Tor is better at hiding the Graphic Card, Camera, Microphone and other hardware related information, it is not good at hiding away the rendering components of a page, including Screen Height,Width, and free space for available height and width. This is a major concern, and most online resources warn against fingerprinting based on this, and a common solution that’s offered is to not maximize the Tor window without some sort of a fix. With regards to page rendering, the HTML 5 canvas is completely unique in Tor, while almost unique in Firefox, which again means that while Tor is good at hiding away hardware based specifications, it has some shortcomings when it comes to page rendering. The values for other rendering attributes like fonts for JS are comparable, but as expected, more unique in Tor than in Firefox. Also, when it comes to default DNT settings, Tor fairs way higher than Firefox, in that it's much more common for DNT to be set to NC in Tor than in Firefox, which is expected, as Tor users are usually more concerned about being tracked and the default set to NC is expected among a large proportion of users. The data related to timezone could lead to a multitude of conclusions, one of which is that the given timezone has more number of Tor users or at least, more number of Tor users than Mozilla users who do not use VPN (plus those whose VPN points to the given destination). Finally, the browser versions for Tor are more similar, in that more number of people use the latest version. This could imply that Tor users update more regularly than Firefox users, as one conclusion, or that Tor users are more concerned about privacy, so they update frequently to avoid security concerns. These conclusions can be validated better with more data, where we can find out more or unexpected factors for the same.

<u>Question 2)</u>

<u>(a) Number of unique third party domains that are visited:</u>

From the code output, the number of unique third party domains visited while visiting www.cnn.com is 80, while that for www.macys.com is 70. The first five elements for the two sets (sets were used instead of lists for uniqueness, as python sets only allow unique elements) are also printed in the output below for reference:

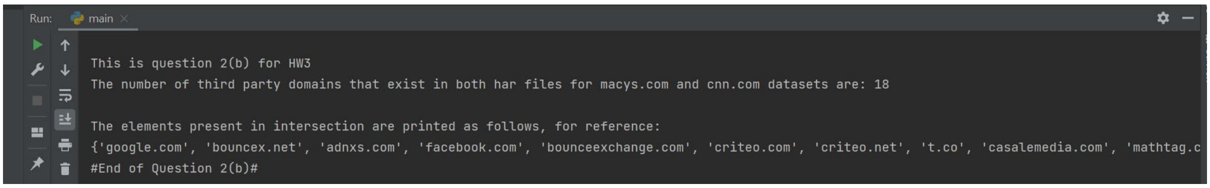


```
C:\Users\nanda\Desktop\Privacy\Assignment_3\Env\Scripts\python.exe "C:/Users/nanda/Desktop/Final/snanda2_HW_3 (2)/Assignment_3/main.py"
This is Questin 2(a) for HW-3

Number of urls that are not cnn.com, and are third party domains for www.cnn.com.har: 80
The first five elements for third-party url's for www.cnn.har file are:
behave.com
ntv.io
indexww.com
wunderkind.co
trustx.org


Number of urls that are not macys.com, and are third party domains for www.macys.com.har: 70
The first five elements for third party url's in macy'scom.har are:

dotomi.com
hlserve.com
medallia.com
googleadservices.com
adnxs.com
```

(b) Number of unique third party domains in both macys.com and cnn.com:

The number of unique third-party domains that occur in both www.macys.com and www.cnn.com from part (a), are calculated as 18, as per the output screen below.

The domains are also printed to the output, for reference.



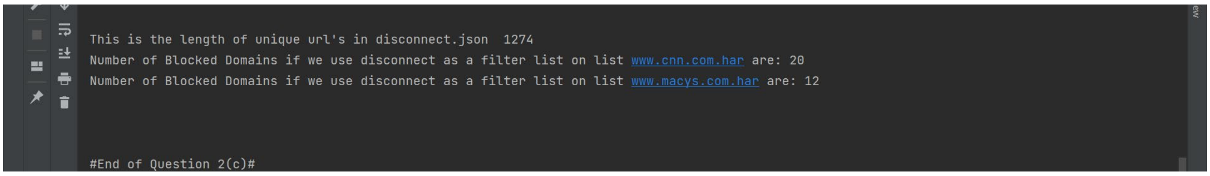
```
This is question 2(b) for HW3
The number of third party domains that exist in both har files for macys.com and cnn.com datasets are: 18

The elements present in intersection are printed as follows, for reference:
{'google.com', 'bouncex.net', 'adnxs.com', 'facebook.com', 'bounceexchange.com', 'criteo.com', 'criteo.net', 't.co', 'casalemedia.com', 'mathtag.c
#End of Question 2(b)#
```

(c) Number of Requests blocked if we were to use Disconnect.json:

The number of third-party domains that would be blocked, if we use disconnect.json as a filter list are:

1. 20 : For www.cnn.com
2. 12 : For www.macys.com

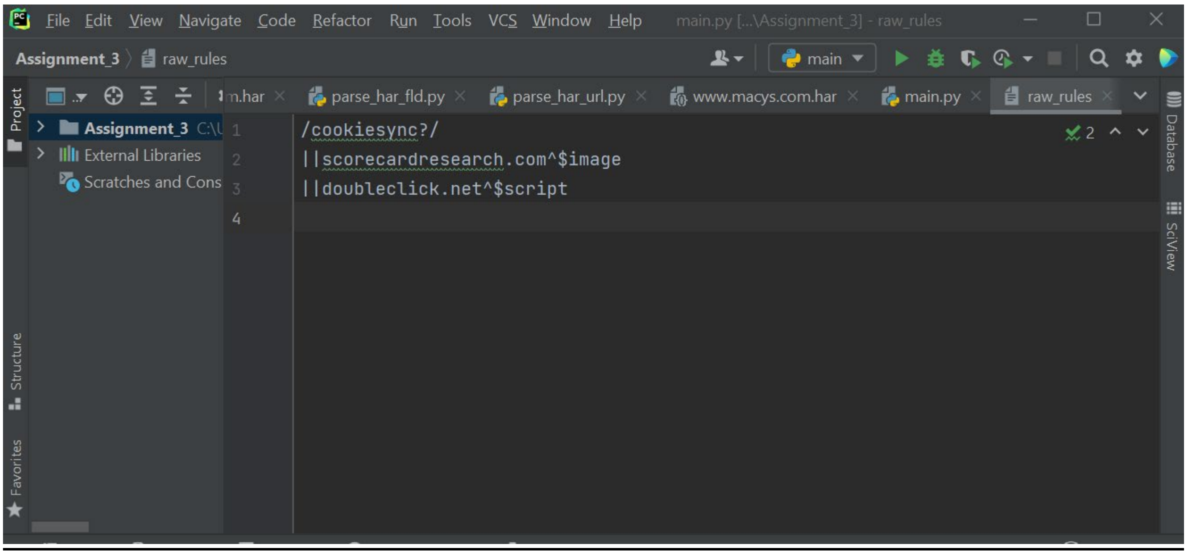
These are also produced to output, as shown below:



```
This is the length of unique url's in disconnect.json  1274
Number of Blocked Domains if we use disconnect as a filter list on list www.cnn.com.har are: 20
Number of Blocked Domains if we use disconnect as a filter list on list www.macys.com.har are: 12



#End of Question 2(c)#
```

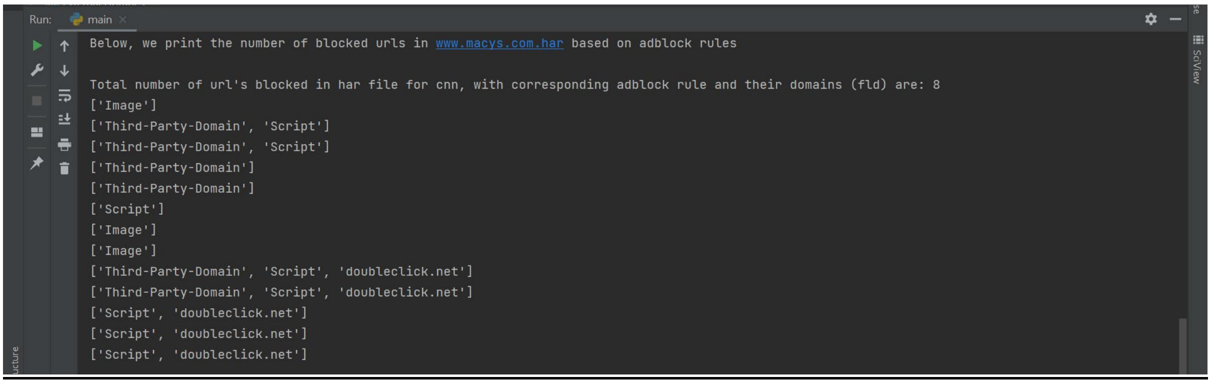
Question 3)

Writing and Testing adblock rules: We are given three adblock rules, which are used as 'raw-rules' in the files labelled raw-rules.txt (Also present in the zip file). Below is the screenshot for the three raw rules that were used:

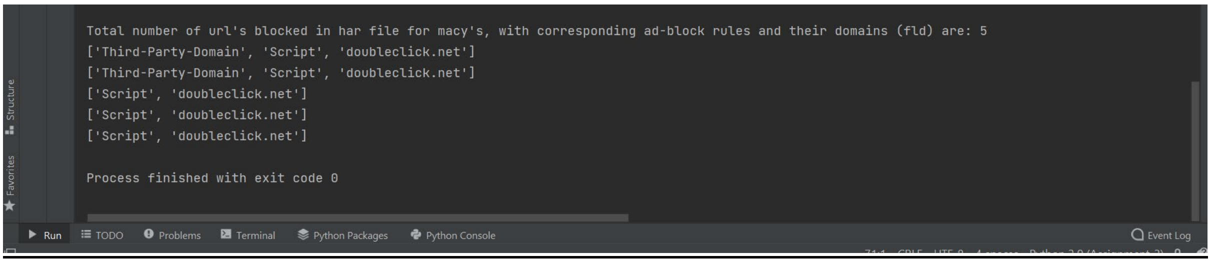


```
/cookiesync?/
||scorecardresearch.com^$image
||doubleclick.net^$script
```

Based on these raw rules, number of domains that were blocked in www.cnn.com.har file were 8, and these are also printed in the specified format mentioned in the question. Please note that repetition here does not denote the same url, since we are getting fld's for the corresponding url's (from get_fld() function in the tld package), and two different third party websites can have the same adblock rule and the same third-party domain, that is, there is no actual repetition.



```
Below, we print the number of blocked urls in www.macys.com.har based on adblock rules

Total number of url's blocked in har file for cnn, with corresponding adblock rule and their domains (fld) are: 8
['Image']
['Third-Party-Domain', 'Script']
['Third-Party-Domain', 'Script']
['Third-Party-Domain']
['Third-Party-Domain']
['Script']
['Image']
['Image']
['Third-Party-Domain', 'Script', 'doubleclick.net']
['Third-Party-Domain', 'Script', 'doubleclick.net']
['Script', 'doubleclick.net']
['Script', 'doubleclick.net']
['Script', 'doubleclick.net']
```



```
Total number of url's blocked in har file for macy's, with corresponding ad-block rules and their domains (fld) are: 5
['Third-Party-Domain', 'Script', 'doubleclick.net']
['Third-Party-Domain', 'Script', 'doubleclick.net']
['Script', 'doubleclick.net']
['Script', 'doubleclick.net']
['Script', 'doubleclick.net']

Process finished with exit code 0
```