

Homework- 1

(Kunal Saini, 2014053)

Q1:- There are 2 government url that provide information regarding any person with just 2-3 fields to fill. And my experiment to glean the PPI of faculty and staff uses these two links, Backpack, Facebook and Google Image Search.

Those two url are:-

- a- <https://incometaxindiaefiling.gov.in/eFiling/Services/KnowYourPanLink.html>

You just need to enter the Surname and Date of Birth and get the pan card number

- b- <http://electoralsearch.in/>

You just need to enter the name and you will get the voter id details of the person

1st Experiment

I will just enter the name and get the voter id detail and to get the pan card number I will enter the name and for the date of birth I need to try a combination of 12(months)*31(days)*X(year) ,and they might block me so if will keep on changing my IP too and for the year:-

- a. It can be estimated from the level of education that we can get from Facebook for instance:- for a professor who has done PhD we can calculate the year through the survey on internet that tells that :-

The average student takes 8.2 years to slog through a PhD program and is 33 years old before earning that top diploma.(Link:-

<http://www.cbsnews.com/news/12-reasons-not-to-get-a-phd/>)

Earning a **Ph.D.**, also known as a Doctor of Philosophy, regardless of the subject of study, requires a set of tasks that typically take 5-6 years to

complete(Link:-

http://study.com/how_long_does_it_take_to_get_a_phd.html)

- b. We can get the date of birth through Facebook (faculty and staff) or backpack (faculty) but for this they should have mentioned that there
- c. As I am an image analysis student there are algorithm that can estimate your age by looking at your image that I could get from Facebook or Google, and filtering is done on these website on the basics of IIITD as they are part of it.

2nd Experiment

Though this experiment might not be too successful but can fetch some good result, this I thought to try on PK sir too .I would call it **Spam Emails Test**

In this experiment I will make a Gmail account having a similar name of the organization to which that person (faculty or staff) is associated to or was previously associated ,and I will apply some Machine Learning or Data Mining algorithm to analyze the king of emails send by these organization and will try to make some same sort of email ,like some seminar is going on or we are thinking to hire you for some task and will ask for some of their PPI ,some identity proof type ,like in case of PK sir I have made a google account cmuniversity.edu@gmail.com resembles to that of CMU and I do have some mails of CMU send to its former students with me ,so I will also compose a mail for some seminar or guest invite to motivate students and ask for information .

Q2:-

- a. Measure the 'strength' of passwords:-

First of all there is no proper standard to measure the strength of the password, It is truly up to the organization or the study that yields what kind of passwords are more safe and protected vs what are unsafe and can be easily guessed by just knowing few information like the name, phone number, interest of the user (like animals(monkey),music etc.) or the names of some of his loved ones (his daughter ,mother or wife).

I would measure the strength of the password on the basics of some parameters namely:-

- Length of the password ,the more long is the password the more difficult it is the remember but more secure
- It should not contain the name/username or surname of the user
- It should have uppercase, lowercase letters, special symbols and digits though it might be complex but there are password manager to store them you don't need to remember
- No repetition of a character or number more than 3 time
- And lastly no dictionary words that re computationally easy to get

I would get some set of leaked passwords that we can find from torrent or other websites over the internet and based on the above mentioned password I will assign some weights to the passwords and from the data then I will do some clustering on the basics on these weights and rate which bunch of password are more secure and will also find some of the common password that people use mostly

- b. Some of the good practices are mentioned above too like long passwords and keeping in mind all the above mentioned parameters, using phrases and pronounceable passwords, some kind long sentence that might not relate to you directly but do have some relevance based on what all you have gone through in your life like your feelings and all and try to avoid those passwords that are very much common on the internet as we can easily get the list of those. People should not keep same password for all kind of platform and should not change password too often.
- c. To aware people we will create a training platform that will inform people regarding what is the advantage of strong passwords or what kind of information regarding you can be leaked once your password is cracked depending upon the way you are connected with the internet.

Q3:-

a. Privacy Policy of Myntra (<http://www.myntra.com/termsfuse>)

- I. Email Address
They ask for it but use it for intended services
- II. Credit Card Number/Home Address
They do not ask for it
- III. Social Security Number
They ask for it and use/sell/share it for other than the intended service
- IV. Marketing
They do use PII for ads and marketing but do not sell/share it to/with third party
- V. Location
They do not track it
- VI. Children under 13
They do not knowingly collect PII of children
- VII. Sharing with Law Enforcement
They do not require Warrant /Subpoena
- VIII. Privacy Policy Change
They may change it anytime without posting
- IX. Control of your data
You can't edit your information
- X. Aggregated data
They aggregate data but remove PII

b. Comparative study of Amazon ,Flipcart and Snapdeal

- XI. Email Address
Similar: - They ask for it but use it for intended services
- XII. Credit Card Number/Home Address
Amazon:-They do not ask for it
Flipcart & Snapdeal:-They ask for it but use it for intended services
- XIII. Social Security Number
Similar: - They do not ask for it

- XIV. Marketing
Similar:-They do use PII for ads and marketing but do not sell/share it to/with third party
- XV. Location
Amazon:-They track it but use only for intended services
Flipcart & Snapdeal:-They do not track
- XVI. Children under 13
Similar:-They do not knowingly collect PII of children
- XVII. Sharing with Law Enforcement
Amazon & Flipcart:-They do require Warrant /Subpoena
Snapdeal:-They do not require Warrant /Subpoena
- XVIII. Privacy Policy Change
Amazon:-They may change it anytime without posting
Flipcart & Snapdeal:- They post new policy but you can't opt out
- XIX. Control of your data
Similar:-You can't edit your information
- XX. Aggregated data
Similar:-They aggregate data but remove PII

Own Privacy Policy:-

- Inform the user for what purpose we are collecting your information and what all information we collect
- Decrypt user data i.e. ensure security of data to prevent privacy breach and in case it happens a mail will be send to the user within 24 hours.
- Do not store data (PII) of children below 13
- Inform the user what all data we are sharing with third party
- Do not sell any data
- Inform the user immediately as we change our privacy policy through mail or message
- Have both opt out and opt in(default)
- Store data but remove PII ,that data will be for analysis

- Allow to change some information like:-address,password,phone number etc.
- Track location only for company intended services
- They do use PII for ads and marketing but do not sell/share it to/with third party
- Require credit card but there is no compulsion to remember this information it depends on the user.

Q4:-

a. Similarity:-Both are Substitution Ciphers

Difference:-Short vs long key, in Vigenere cipher we can have long key whereas in Caesar cipher the key length is short.

The Caesar cipher has 25 possible shifts, A shift of 26 or more will simply repeat. Because there are only 25 possible shifts, anyone could test each possible shift to determine encoded message, so it is very easy computationally to find the key and decrypt the message. A system with more possibilities would be more difficult to crack .Now Vigenere cipher makes use of multiple shifts by employing a keyword (key), thus making it more difficult to decode. Substitution ciphers like the Caesar cipher have a one-to-one correspondence; whereas Vigenere cipher have a one-to-many correspondence. Using the Vigenere cipher will pose more decoding difficulty than a simple substitution cipher because of long length of the key and one-to-many correspondence. Instead of one shifted alphabet being used, multiple shifted alphabets are used so it is also more computationally difficult though it can be cracked too.

b. Feistel Cipher

It is a representation of DES (Data Encryption Standard) or we could say way of implementing DES. It is a block cipher; encrypting 64 bit blocks using a 64 bit key in 16 rounds, in each round a key of 48 bits is used that is

derived from the input 64 bit key. It is a product cipher; performs both substitution and transposition.

Q5:-

No, anyone can give any sort to information as there is no kind of verification regarding the information, it may be completely wrong and the person can easily enter the campus.

Yes, it is usable as it's just a 3-5 minute process. You just need to enter some text and get the entry; there is no pain in doing it.

Alternate system for entering the college -use of Biometrics:-

We can install a camera at the gate which will be connected to the database that have images of all students ,staff and faculty at IIITD ,this camera will detect the face of the person and verify whether that person matches with any image in the database. Now these algorithms don't provide 100% accuracy depending upon the light in the background. But we can train our system for some time period and then assign a threshold.

Usability issue:-IIITD is not too large family now but it may become more large in the coming days with increasing admissions and construction of new phase ,so due to which the database size will increase so those algorithm will take more time this means that the person need to wait for long ,and secondly the person might be correct but due to some bad light the algorithm might not give that threshold.

Secondly what we can do is we have scan the finger print of the person, but the usability issue here are also same the database could be large so more time or there might be dust on the finger ,or they might be wet. Though database size is more concerning issue.Similar if we consider other kind of biometrics like eye etc.

