

Investigating Privacy Complaints

Christopher Castillo¹, Prof. Deirdre Mulligan², Jen King², Nick Doty² ¹Loyola Marymount University, ²University of California, Berkeley, School of Information

Overview

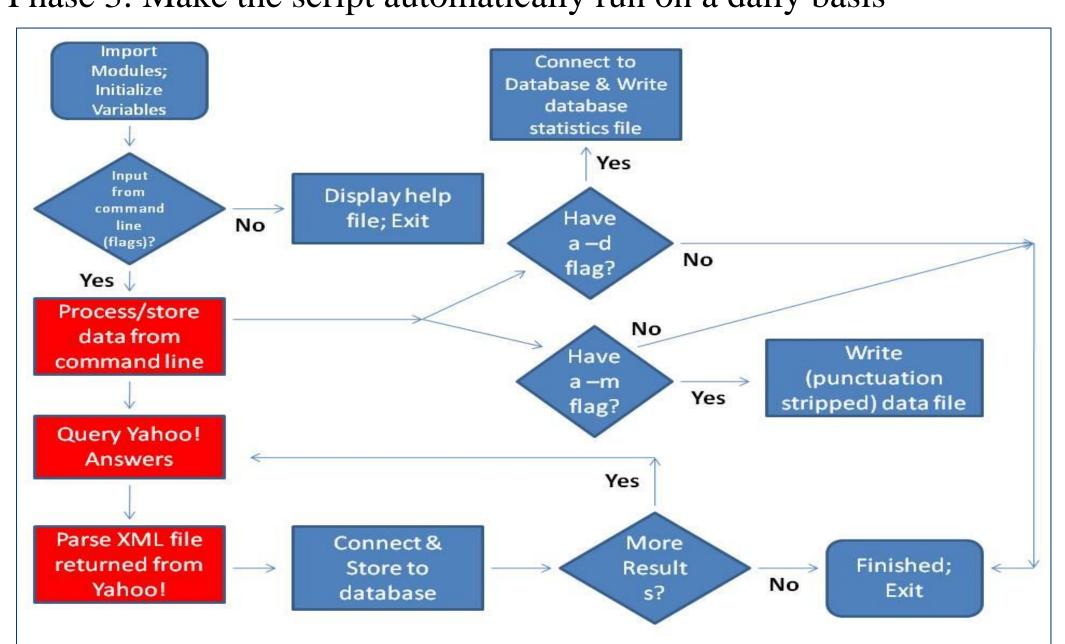
With recent issues of personal information sharing from both online files (config.xml & categories.xml) and works well with minidom and offline sources making headlines, individuals are becoming more aware of information privacy. These concerns often come from popular what online users are participating in. However, little has been attempted to acknowledge these concerns and implement some type of legal protection for our online personas. There must be some response to individuals' privacy concerns and how the users conceptualize Parsing the Data privacy.

In our research, we came up with a basis that allows us to review theories of internet privacy. We created a command-line tool in python that queries Yahoo! Answers database for relevant questions about privacy. We will use this data in order to understand existing Chosen Answer, Category, and Link. conceptions of privacy and the methods used by individuals to conserve and protect their personal information.

Research Process

In order create our command-line tool, we followed this overall process:

- •Phase 1: Create a script to manually query Yahoo! Answers' database with a single term that returns results
- •Phase 2: Create a database using MySQL to send and store results
- •Phase 3: Make the script automatically run on a daily basis



Outline of the entire code with specific contribution to the code highlighted in red

Methods

Why Yahoo! Answers?

YAHOO! ANSWERS

Yahoo! Answers enables its users to ask and answer questions about issues or concerns of interest. Within this database we want to gather people's real world questions based on internet and information privacy. This data is well structured, open to the public, and available through the Yahoo! Answers Application Programming Interface (API). The API is a tool for developers to use Yahoo! Answers' content and services for the development of their own program.

Connecting to Yahoo!

- > XML: Allows user created tags, which enabled us to create two XML
- > Urllib: Preinstalled library containing predefined functions for use with URLs. Used two primary functions: urlopen(), which opens the Request social networking sites and new location aware services, the bulk of URL, and urlencode(), which converts the parameters to a url-encoded string capable of passing through urlopen().
 - > Parameters: Used the following API specified parameters: query, search_in, sort, appid, start, and results.

- > Reasons: Convert from XML to plain text format and get rid of information unnecessary to our research.
- > Necessary Information: Subject, Question ID, Timestamp, Content,
- ► Minidom: Enables searching through parent and children nodes of the XML formatted data in order to search for the desired element. Extracted the data and converted it to a plain text format. More efficient for a large amount of query terms.

Results

I work for an insurance company and we do not record conversations with our clients unless needed, a nd when we do, we ask for their consent. As employees, our calls are sometimes monitored by the comp any. A client recently advised that he has recorded every conversation I had with him and he never g

Can someone please explain what rights I have as an employee? This client invaded my privacy!

Timestamp : 1219386953

Chosen Answer : The vast majority of states are "one-party" states, which means that only one of the participants of the conversation has to know it's being recorded (without a wiretap warrant). Most companies volunteer this information to their customers as a matter of courtesy.

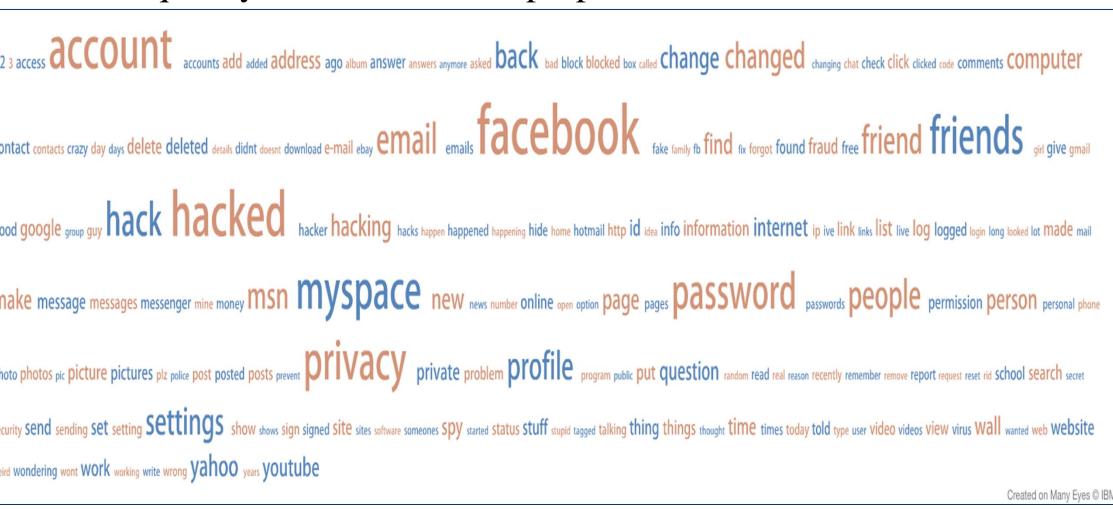
Only California, Connecticut, Delaware, Florida, Massachusetts, Maryland, Montana, New Hampshire, Pe nnsylvania, Washington, and Wyoming are "two-party" states. Federal wiretap laws, which would apply to calls recorded across state lines, address law enforcement use, and whether or not this particul ar instance is illegal would probably be up to a judge.

Personally, I'm wondering why you'd even care. If these are just calls about insurance, and you did your job correctly and with professionalism, how could recording the calls harm you in any way? Pr osecuting the person criminally would only make it look like you have something to hide, and you cou ldn't win a civil suit without damages. If the guy gives you the creeps, just pass him off to anoth

Example of parsed results in the terminal

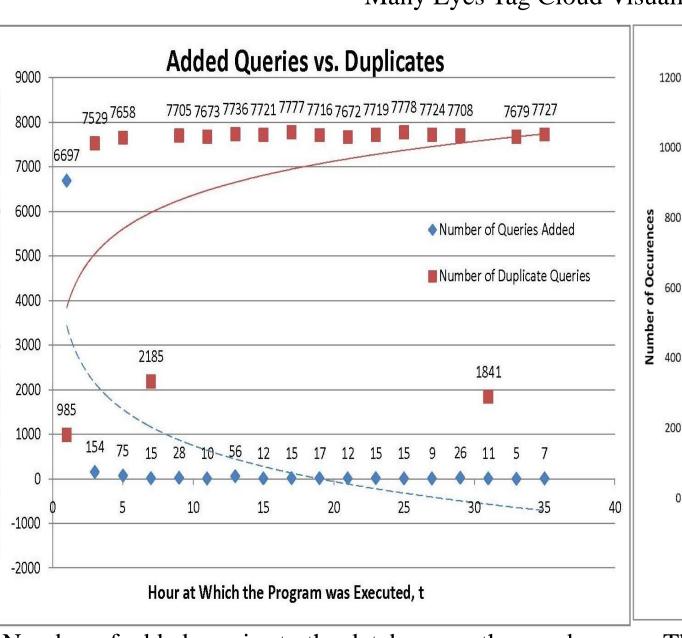
Many Eyes Tag Cloud Visualization

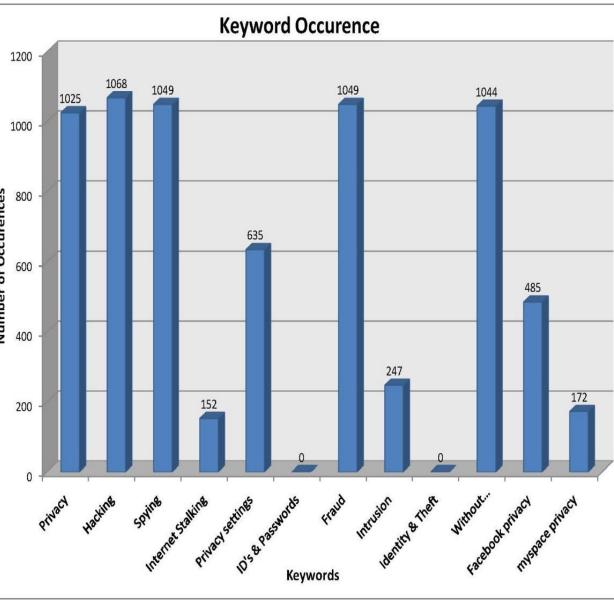
Many Eyes is a data visualization tool that allows users to upload data and create a graphical representation. The tag cloud visualization is based on how often words appear in a text. The relationship between a word and its frequency within the text is proportional to the size of the word.



Many Eyes Tag Cloud Visualization: One Word

facebook account 6 er log back long story long time mafia wars messenger account months ago msn account msn hacked msn hacking msn hotmail msn messenger msn password msn spy mutual friend mutual friends myspace 20 myspace accounts myspace hack myspace hacked myspace hacking myspace page myspace password myspace privacy myspace profile new account new email new facebook new myspace new password new privacy ttings news feed news feeds password back password reset people hack personal info personal information personal i time ago tool bar Wall posts wall privacy web cam web page Web site weeks ago Windows live wireless internet wont work wrong password yahoo account yahoo answers yahoo email yahoo id yahoo mail Many Eyes Tag Cloud Visualization: Two Words





Number of added queries to the database vs. the number of duplicates found

The frequency of specific keywords that were found to yield relevant results

Conclusion & Next Steps

Once connecting to the Yahoo! Answers database, we were able to gather some results and parse those results to gather information relevant to our research. By doing so we were able to analyze the questions using Many Eyes and come up with a few conclusions. By the types of keywords we used, the data had a lot of information linked with Facebook. This is surprising since this is a Yahoo! web service, so we were expecting more questions concerning Yahoo!. This might have been from people actually concerned about privacy issues over their Facebook accounts, or an overspill of questions directed to Facebook that were not dealt with at their site.

For future research we plan on creating a taxonomy of privacy terms and phrases. I would also try to improve the code to run more efficiently by gathering the data much faster. Also, I would like to join the graduate mentors on continuing to analyze people's concerns and issues about internet privacy, to review and possibly come up with some theories to protect online users' privacy.

Acknowledgments

I would like to thank the Team for Research in Ubiquitous Secure Technology (TRUST), the National Science Foundation (NSF), and Dr. Kristin Gates, for giving me the opportunity to conduct this research. I would also like to thank my advisors, Prof. Deirdre Mulligan, Jen King, and Nick Doty, for their guidance in conducting this research. Finally, I would like to thank my colleagues, Jennifer Felder, German Gomez, Rafael Negron, and Anand Sonkar, for all their hard work and dedication to this project.









