Prospectus

Social Network and Archival Context (SNAC) OpenRefine Plugin
(Technical Topic)

How has the publics' perception of security affected the widespread use, development, and availability of biometrics in mobile devices?
(STS Topic)

By

Charles Michael Chang

October 30th, 2019

Technical Project Team Members:

Sandy Gould, Mark Jeong, John Perez, Victor Shen, Peter Tran, Grace Wu, Jessica Xu

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: Michael Chang

Approved: _____ Date 12/5/2019
Rider W. Foley, Department of Engineering and Society

Approved: Ahmed Ibrahim Date 11/25/19
Ahmed Ibrahim, Department of Computer Science

Introduction

Already under development, Social Networks and Archival Context (SNAC) is a free, online resource that aids users in discovering biographical and historical information about people, organization, and groups that are created or documented in historical resources and their connections to one another. Users are able to locate these collections and related resources held at worldwide cultural heritage institutions. Many libraries and archival institutions would like to contribute their records to SNAC, but figuring out which of their records already exist in SNAC (and should be updated in SNAC) and which are completely new (and need to be inserted) is a challenge.

The problem is that there is no easy way to normalize the data being passed in and allow people to align their schema to SNAC's schema. To address this problem, my team is developing an OpenRefine plugin to help compare, share, and create data between SNAC and other archival organizations. The plugin is a good way for libraries or data centers from all over the world to collectively import their information into website that is accessible to the public. Ideally, this plugin will allow data to be reconciled and imported into SNAC. The plugin will allow users to load their archival records of persons, families, and corporations as spreadsheets and match their columns to the corresponding data fields in SNAC. Users could then match their records to SNAC, inserting new records and updating existing records via our Application Program Interface (API). This plugin will involve both back-end and front-end work to build the OpenRefine user interface. It should be noted that WikiData, an existing plugin, will be used as a resource to design our plugin for SNAC. My team will create a plugin for the SNAC interface, and when finished, the research and development done by our team will provide users with a means to reconcile and create archival data.

Though my technical research focuses on developing a plugin, I will investigate biometrics, a completely different topic. Sparked by my interest in Apple's iPhone 5s and other phones with biometrics, there are a lot of potential issues surrounding this technology. With people storing more sensitive information on their mobile devices and increased vulnerabilities, biometrics and good mobile security implementations are a must (Lerner, 2019). I will be exploring the publics' perception of security and how it has affected the widespread use and availability of biometrics in mobile devices.

Technical Topic – SNAC OpenRefine Plugin

Social Networks and Archival Context (SNAC) is a free, online resource that allows users to discover information about the people and organizations that are documented in primary source documents and the connections between them (Social Networks and Archival Context [SNAC], n.d.). SNAC is used to locate archived collections as well as related resources held around the world. As an international cooperative, SNAC works to "build a corpus of reliable descriptions" of people and artifacts that link to and "provide a contextual understanding" of historical records (SNAC, n.d., para. 1). In order to create these contextual connections, SNAC sources its information from many different libraries and archival institutions. SNAC cooperates with over 4000 institutions to gather and reconcile data (SNAC, n.d.). Each of these institutions has a different structure for storing records. Relationships between different entities, labels for certain types of data, and the hierarchy of the data itself are inconsistent from each outside institution. SNAC needs to reconcile the differences between the outside data and its own data storage structure before importing the data into its database. It is extremely impractical to clean up the data manually or with simple tools (Ham, 2013). The reconciliation of this data is vital to

the functionality of an archival organization such as SNAC because it is crucial for efficient and accurate querying (Park, 2008).

The technical project seeks to develop a standalone plugin for Social Networks and Archival Context (SNAC) using OpenRefine. OpenRefine is an open source software that is community-maintained designed specifically for data normalization, transformation, and cleaning (Hill, 2016). It allows users to import and normalize data with a series of pre-existing default user interfaces after connecting to a target resource. OpenRefine provides a "powerful yet user-friendly interface" for experimenting with and querying data (Hill, 2016, p. 228).

With over 700 edits occurring to its data schema in week, Social Networks and Archival Context (SNAC) is no small data archive (SNAC, n.d.). The current workflow for refining and updating data in SNAC is quite difficult and inaccessible to inexperienced users. It involves users hitting SNAC's APIs for refining data on their server from the user's local machine. The technical project aims to greatly simplify this process by creating a streamlined plugin that will have all the functionalities needed to refine and upload data in one location. The logical flow and components needed for the project are illustrated in Figure 1. The plugin

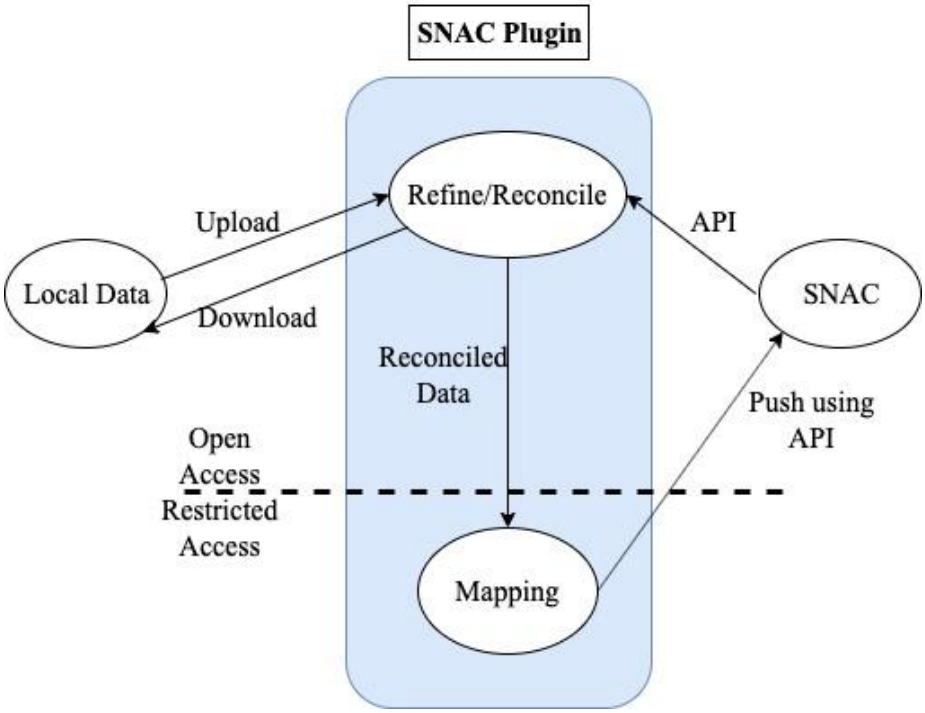

Figure 1. SNAC Plugin Model: An overview of the design of the plugin, depicting the different processes and functions that will be made available by the plugin (Xu, 2019).

will serve as a connection between the user's local data and SNAC's server. It will allow users to import external data in the form of comma-separated values (CSV) files and make use of APIs provided by SNAC to reconcile and refine that data with SNAC's unique JavaScript Object Notation (JSON) data structure. The plugin will have two main user groups: privileged and unprivileged users. Both types of users will be able to use the plugin to format any data ported in using SNAC's organizational schema. Only privileged users will be able to then push the formatted data into SNAC's own database utilizing the APIs provided by SNAC. The technical project will provide an easy way to reconcile outside data with SNAC's existing data in addition with an improved user interfaced for an enhanced user experience.

The development will conduct biweekly customer meetings with the client in order to gather system requirements and get feedback about ongoing work. The minimum requirements for the plugin to be completed by the end of this semester include:

- Allowing users to import CSV data into the plugin
- Connect the data fields with different SNAC IDs
- Search for constellations in SNAC and match them to the imported data
- Allow a human editor to choose from several options to match for when the plugin is unsure
- Reconcile the imported changes based on the connection and matches
- Download the data that is now reconciled with SNAC's structure
- Users with privileges will be able to publish the data to SNAC

Desired requirements include:

- Users will be able to reconcile more complex data items like relationships and geolocations

- Users will be able to edit already existing resources and constellations

So far, no optional requirements have been specified by the client.

The technical project will be developed over the course of the two-semester capstone series led by Professor Ahmed Ibrahim from the Computer Science department, and will result in a technical report. To create this plugin, OpenRefine will be used, as it is a powerful tool for working with disorganized data that can "[transform] it from one format into another; extending it with web services and external data" (OpenRefine, n.d., para. 1). A similar project exists already for WikiData, but the technical project will create a new implementation specifically for Social Networks and Archival Context (SNAC). The plugin will hopefully provide a faster and more intuitive way for SNAC users to reconciliate and update data.

STS Topic

Giving an alternative to one-way, deterministic ideologies, Hughes (2003, p. 287) proposes the theory of technological momentum, finding that "social development shapes and is shaped by technology". He posits that technological systems obtain momentum as the system passes through various phases and evolves, resulting in both technology and society changing. These phases are invention, development, innovation, transfer, and growth, competition, and consolidation. Though these phases ideally follow this order, problems arise, in which a component falls behind or is out of phase with the rest of the system. This is called a reverse salient (Hughes, 1987, p. 73). Analyzing the progression of these phases, I will be using Hughes' framework to look at biometrics as a technology, its link to mobile phones along with the resulting growth, and the publics' perception biometric security and adoption over time.

Biometrics, as defined by Pocovnicu (2009, p. 57), is "technology used to uniquely identify individuals based on their physical, chemical or behavioral traits." Biometrics provide convenience, which comes in the form of benefits that include automatic cooperation between user and device, emergency identification, higher throughput, uniqueness of biometric traits, and cost. This usability and convenience drive the adoption of biometric features. In terms of public recognition, biometrics has seen some incredible growth. Just a decade ago, people were not really familiar with the concept of biometrics nor did they care enough to be interested in it, but Bhagavatula (2015) credits this awareness to mobile phones implementing "cool" features. Data from a 10-person lab study and 198-person online survey show the relationship between human interaction and that fingerprint authentication method was most favorable, so it is no surprise that fingerprint recognition is the most common type of biometric (Bhattacharyya, 2009, p. 15).

While there are many benefits, biometrics are not without drawbacks. Social criticisms about the effects of the differences in the human body on biometrics have been a hot topic for discussion. Recent social science studies credit the ineffectiveness of biometrics to race and gender. A few examples include visual imaging of individuals being calibrated and optimized for "white-skinned subjects" (Koppenburg, 2018, p. 7), iris-scanning technology failing more than 6 times to get an image quality of 90% for an Asian American woman because her eyes were not opened wide enough, or a Somalian woman being accused of mutilating her fingerprints after a device is only able to produce a 40% match. These studies are brought up not to show racism from biometric failure, but rather the unintended side effects of calibrating biometric technology to a preset lightness of skin or certain body-feature shape. This hiccup in biometric technology would be considered a reverse salient. While fingerprint and iris scan biometrics are expanding and becoming popular, imaging techniques are at a standstill, going back into a development

phase. Problems arise in any technology however, not just in biometrics, forcing inventors and developers to go back to the drawing board if the technology is to progress.

One way that biometrics has become widespread is through its integration with smartphones and other mobile devices (Pocovnicu, 2009). Acuity Market Intelligence forecasts that in 2020, 100% of 2.7+ billion smartphones will incorporate biometrics, an increase from the 40% in 2016 (Armstrong, 2017). Whether buyers are aware or not, biometrics have been almost completely embraced by consumers by taking root in the smartphone market. In establishing itself in a profitable market, biometric technology has gained and is continuing to gain considerable momentum. Looking at just iPhone users, 2018 had 101 million users. This 56-million-user increase since 2012, and it is projected to increase another 9 million by 2021 (Holst, 2019). Smartphones are very popular because of their versatility. Ranging from accessing the internet for entertainment to storing sensitive information, people need to be able to trust phones to keep their individual data confidential. Thus, the security that biometrics offers is an attractive feature that could improve its public perception in securing people's personal data. Could the increased security potentially shift society to eradicate mobile passcodes in support of biometric authentication? For this to be possible, biometrics would have to be more widely favored.

Pocovnicu (2009) argues that biometrics are a better and more secure alternative to the traditional pin number since it forgoes the memorization step. Guessing a 4-digit PIN is one in one thousand whereas tricking Apple's face recognition is one in one million. Given this, privacy is one of the main reasons people resist implementation of biometrics (Des, 2016). On top of the general weariness that humans have about things they have no control over, like getting their fingerprints or irises scanned, concern stems from where the data is being stored. Passcodes can easily be changed, but the same cannot be said for faces or fingerprints. But seeing that the

average iPhone user unlocks their device 80 times a day (Kelly, 2019), the instantaneous unlock from biometric authentication may seem like a better alternative than punching in a 6-digit code. People are willing to overlook some of the negative traits, like privacy violations, in exchange for convenience (Thakkar, 2018). The benefits of new biometrics far outweigh consequences, achieving efficiency and heightened security (Crego, 2012). In 2018, a survey showed that of 4,000 adults, even though 67% of people said they were comfortable using biometrics at the time, 87% said they would be comfortable with them in the future (Kelly, 2019). This increased percentage could be due to society anticipating results and proof of better security. If the government were to deliver solutions to the public, biometrics could cater to the people in a way that not only boosts not only morale in security, but also a trusting bond between governments and citizens (Crego, 2012). Gaining public trust of biometrics leads to continued support giving rise to innovate using new methods (i.e. vein recognition, using other body parts, etc.). This relationship between the people and the technology gives way biometrics gaining momentum now and in the coming years.

Research Question and Methods

How has the publics' perception of security affected the widespread use, development, and availability of biometrics in mobile devices? Seeing so many people have smartphones these days, I can hardly think of anyone who has a phone without biometric security features. Thus, I want to see how many people actually use it, how effective people think it is, and if it is one of the deciding factors when choosing a certain phone.

I plan on conducting this research through case studies between three different phone brands, Samsung, Apple, and Huawei. First, I will be analyzing both popularity and technical

features, seeing if there is any correlation between security features and usage. This will require researching mobile hardware, biometric capabilities like fingerprint and facial recognition, feature preference, trends in sales, and user satisfaction. The information will likely come from prior literature on case studies already carried out.

Next, I will supplement the research found in the previous method with primary data gathered through both surveys and interviews. This data will ideally verify and support what is found in the case studies. With the case studies, surveys, and interviews, I will apply technological momentum by using the concept of phase analysis to assess the current stage of biometrics, how it fits in with biometrics as a whole, and project a potential path of growth. Taking all of this information, I will draw conclusions on the current state of biometrics (how much it has advanced), quantify the effect that the mobile platform has had on its development, as well as decide if a certain type of biometric is likely to dominate the industry in the future.

Conclusion

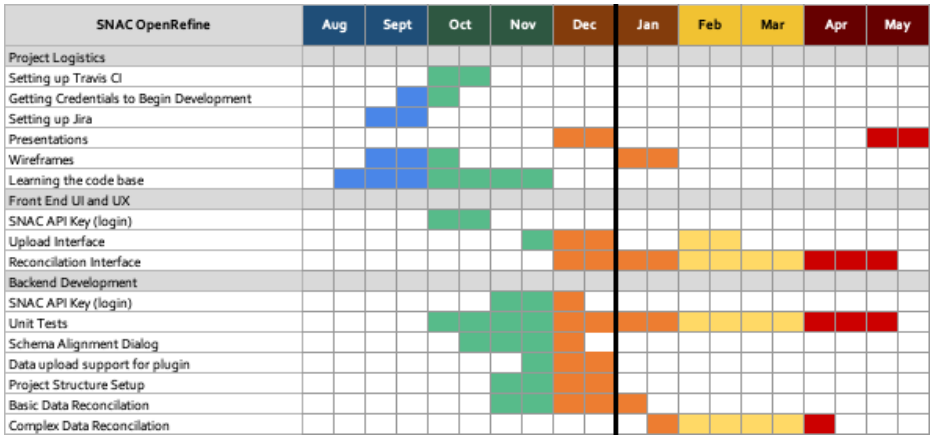

Figure 2 Gantt Chart for Capstone Project. This figure illustrates a timeline showing when the team plans to develop and finish component features.

My technical topic for my capstone is to create a plugin extension for the OpenRefine application to work with SNAC so that it can reconcile user data with pre-existing SNAC data or create new data and push it into SNAC. The tentative deliverable is to implement the second part, the functionality in which the plugin will compare the schemas of user data and SNAC– making sure that profile's name matches with name data, age matches with age data, and so on– to create new profiles. The working application will provide users with a means to archive data for convenient access. An in-progress timeline in the form of a Gantt Chart provided in Figure 2 shows the team's plan to complete each step of the project. Once finished, I plan on adding it in my thesis. Though I do not plan on publishing it as a scholarly article, I do want this to be a paper I am proud of. However, I do hope that should anyone continue developing the open-source code for the plugin, the technical portion will be of use to help understand the function of the technology.

My STS topic explores the spread and advancement of biometrics. Namely, why is it popular, or how sophisticated is it now. Brought on by my interest in these security features in phones, I hope to learn how the mobile platform has induced changes in this industry. Personally, I would use this information to educate me on security trends to be equipped with devices with strong security features. As for the results of research, I would love for it to be resource to help people understand what biometrics are and how readily available it is for everyone, in hopes that the public is more security conscious of devices they use every day.

References

Armstrong, M., & Richter, F. (2017). Infographic: The Future of Mobile Biometrics. Retrieved from https://www.statista.com/chart/11122/the-future-of-mobile-biometrics/.

Bhagavatula, R. Ur, B. Iacovino, K. Kywe, S. Cranor, L. Savvides, M. (2015). Biometric authentication on iPhone and Android: Usability, perceptions, and influences on adoption. *USEC '15: Workshop on Usable Security*. 1-10.

Bhattacharyya, D., Ranjan, R., Alisherov, F., & Choi, M. (2009). Biometric Authentication: A Review. *International Journal of u- and e- Service, Science and Technology*, *2*(3), 1–28.

Cockburn, A., & Highsmith, J. (2001). Agile software development, the people factor. *Computer*, *34*(11), 131–133. doi: 10.1109/2.963450

Crego, M., & Bataller, C. (2012). Biometrics boost business outcomes and security in public services. *Biometric Technology Today*, *2012*(3), 8–11. doi: 10.1016/s0969-4765(12)70073-5

Des, R. (2016, November 14). The Social Impacts of Biometric Technology. Retrieved from https://resources.infosecinstitute.com/the-social-impacts-of-biometric-technology/.

Ham, K. (2013). Free, Open-source Tool for Cleaning and Transforming Data. *Journal of the Medical Library Association*. Retrieved from https://www.ncbi.nlm.nih.gov/

Hill, K. M. (2016). In Search of Useful Collection Metadata: Using OpenRefine to Create Accurate, Complete, and Clean Title-Level Collection Information. *Serials Review*, 42(3), 222-228. Retrieved from https://www.sciencedirect.com/journal/serials-review

Holst, A. (2019). iPhone users in the US 2012-2021. Retrieved from https://www.statista.com/statistics/232790/forecast-of-apple-users-in-the-us/.

Hughes, T.P. (1987). The Evolution of Large Technological Systems. In W.E. Bjiker, T.P. Hughes, and T. Pinch (Eds). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. MIT Press: Cambridge, MA. pp. 51-82.

Hughes, T.P. (2003). Technological Momentum. In Katz, Eric, et al. *Controlling Technology: Contemporary Issues. Amherst*, NY: Prometheus Books. pp. 287-296.

Kelly, H. (2019). Fingerprints and face scans are the future of smartphones. These holdouts refuse to use them. Retrieved from https://www.washingtonpost.com/technology/2019/11/15/fingerprints-face-scans-are-future-smartphones-these-holdouts-refuse-use-them/.

Kloppenburg, S., Ploeg, I. V. D. (2018). Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences. *Science as Culture*, 1–20. doi: 10.1080/09505431.2018.1519534.

Lerner, S. (2019). 11 Best Practices To Improve Mobile Device Security. Retrieved from https://www.enterprisemobilityexchange.com/eme-security/whitepapers/mobile-device-security-best-practices.

OpenRefine. (n.d.). *Introduction to OpenRefine*. Retrieved from http://openrefine.org/

Park, J-R. (2008). Metadata Quality in Repositories: a Survey of the Current State of the Art. *Cataloging & Classification Quarterly*, 47(3), 213-228. doi:10.1080/01639370902737240

Pocovnicu, A. (2009). Biometric Security for Cell Phones. *Informatica Economică*, 13, 57-63.

Social Networks and Archival Context. (n.d.). *About SNAC*. Retrieved from https://portal.snaccooperative.org/about

Thakkar, D. (2018, August 8). How Does Biometric Technology Impacts Society? Retrieved from https://www.bayometric.com/biometric-technology-impacts-society/.

Xu, Jessica (2019). Figure 1: SNAC Plugin Model.