Undergraduate Thesis Prospectus

Social Networks and Archival Context OpenRefine Plugin

(technical research project in Computer Science)

Violent Video Games and Their Perceived Effects

(STS research project)

by

Mark Jeong

October 31, 2019

technical project collaborators:
Michael Chang
Sandy Gould
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: _____ date: _____

approved: _____ date: _____
Peter Norton, Department of Engineering and Society

approved: _____ date: _____
Ahmed Ibrahim, Department of Computer Science

General Research Problem

What are the implications of media applications of artificial intelligence? Social media grows every year. There are 3.48 billion social media users in 2019, an increase of 9% from the previous year; 45% of the world's population now uses social networks (Kemp, 2019), generating vast but unstructured data. With artificial intelligence (AI) tools, analysts can study user behaviors and search habits to distinguish target audiences.

AI can support automation that improves productivity, but at the cost of disruptions to employment and businesses. AI can encode biases (Frank, 2019) and influence which content we see, or which opportunities we encounter or miss.

Social Networks and Archival Context 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 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.

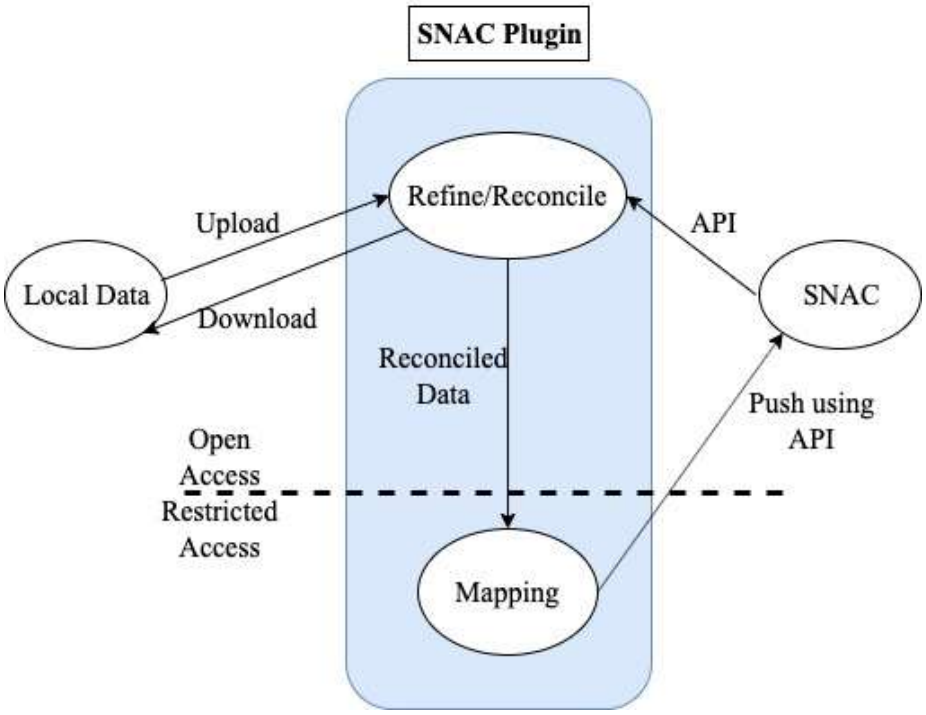

Figure 1: SNAC Plugin Model: An overview of the design of the plugin, depicting the different processes and functions that will be made available by the plugin (Xu, 2019).

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 interface 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.

Violent Video Games and Their Perceived Effects

How are defenders and critics of violent video games competing to influence its perceived responsibility for social pathologies? Nearly two-thirds of the U.S. population 13 years or older plays video games ("U.S. Games 360 Report," 2018). The video game industry is a growing field, generating a record $43.4 billion in revenue in 2018 in the U.S. alone, which was an 18% percent increase from the previous year (ESA, 2019). After recent shootings, some critics have blamed video games. According to Bushman (2015), 66% of researchers, 67% of parents, and 90.3% of pediatricians agreed that video games can increase aggressive behavior among children. Others find no such harm and, in fact, see benefits in video games.

Prescott et al. (2018) found that, after video gameplay, aggression is "most strongly observed among White participants, less strongly but reliably observed among Asian participants, and unreliably among Hispanic participants." The researchers did not consider culture or personality or type of violent game.

President Trump, among some other politicians, asserts that "we must stop the glorification of violence in our society" that "includes the gruesome and grisly video

games that are now commonplace" (Draper, 2019). Pete Hines, a father and a spokesman for Bethesda Softworks, contends that "parents need to be aware of what their kids are buying and playing" (D'Arcy, 2011).

To David Hogg, a gun control advocate and a survivor of the 2018 mass shooting in Parkland, Florida, "video games aren't to blame" (Romano, 2019). Renee Gittins, the director of the International Game Developers Association, claims that "blaming video games distracts from the broader issues at hand" (Parkin, 2019). The Entertainment Software Association, a major trade association, states that "video games do not cause real-world violence or aggression" and "blaming video games for real-world violence is no more productive than blaming other forms of media for the content they depict" (ESA, 2019). Many professional gamers have condemned claims that video games cause real violence. Professional gamer Mike Rufail says he "never felt the need to enact any kind of violence." Tim Olson, another professional gamer, says that he has "a really good work ethic from video games" and gaming has "done nothing but bring positivity" to his life (Austin, 2019).

References

Austin, P. L. (2019, August 9). Pro Gamers Say Guns, Not Games, Are to Blame for Shootings. *Time*. https://time.com/5648284/video-games-violence-shootings/.

Bertin, J. (2019, March 15). A Comment on Video Games and Violence. https://ncac.org/news/blog/a-comment-on-video-games-and-violence.

Bushman, B. J., Gollwitzer, M., Cruz, C. (2015, July). There Is Broad Consensus: Media Researchers Agree That Violent Media Increase Aggression in Children, and Pediatricians and Parents Concur. *Psychology of Popular Media Culture, 4*(3), 200-214.

D'Arcy, J. (2011, June 28). Violent video games and minors: Is it parents' fault? *Washington Post*. https://www.washingtonpost.com/blogs/on-parenting/post/violent-video-games-and-minors-is-it-parents-fault/2011/06/27/AGOCOMoH_blog.html?noredirect=on

Draper, K. (2019, Aug 5). Video Games Aren't Why Shootings Happen. Politicians Still Blame Them. *New York Times*. https://www.nytimes.com/2019/08/05/sports/trump-violent-video-games-studies.html?module=inline.

ESA (2019). Entertainment Software Association. Media Violence. https://www.theesa.com/policy/media-violence/

ESA (2019). Entertainment Software Association (Jan. 22). U.S. Video Game Sales Reach Record-Breaking $43.4 Billion in 2018. https://www.theesa.com/press-releases/u-s-video-game-sales-reach-record-breaking-43-4-billion-in-2018/.

Frank, M. R. (2019, March 12). The evolution of AI research and the study of its social implications. *Medium*. https://medium.com/mit-media-lab/the-evolution-of-ai-research-and-the-study-of-its-social-implications-4a9598b3d7db.

Ham, K. (2013). Free, Open-source Tool for Cleaning and Transforming Data. *Journal of the Medical Library Association*. 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. https://www.sciencedirect.com/journal/serials-review

Kemp, S. (2019, Jan. 30). Digital 2019: Global Digital Overview - DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2019-global-digital-overview.

Nielsen (2018, May 24). U.S. Games 360 Report: 2018.

https://www.nielsen.com/us/en/insights/report/2018/us-games-360-report-2018/.

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

Parkin, S. (n.d.). No, Video Games Don't Cause Mass Shootings. *Time*. https://time.com/5647659/no-video-games-dont-cause-mass-shootings-but-the-conversation-shouldnt-end-there/

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

Prescott, A. T., Sargent, J. D., & Hull, J. G. (2018, October 2). Metaanalysis of the relationship between violent video game play and physical aggression over time. *Proceedings of the National Academy of Sciences, 115*(40), 9882-9888. https://www.pnas.org/content/115/40/9882.

Romano, A. (2019, August 26). The frustrating, enduring debate over video games, violence, and guns. *Vox*. https://www.vox.com/2019/8/26/20754659/video-games-and-violence-debate-moral-panic-history

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

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