

DATA SELFIE DISSECTED

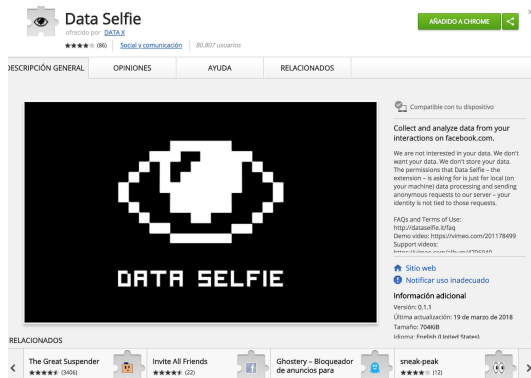


Seoul Artificial Intelligence

Emilio Jose Coronado Lopez

What it is.

Chrome / Firefox plugin (Install it at your own risk)



Creator + Developer + Co-Founder: Hang Do Thi Duc

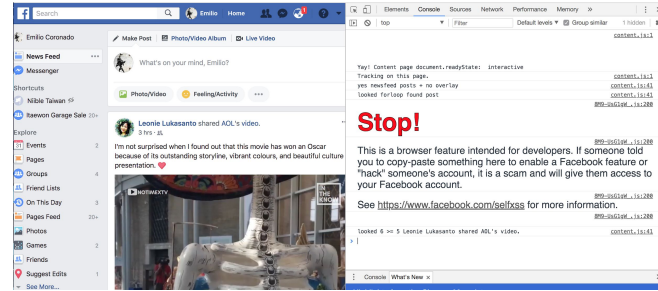
Co-Founder: Regina Flores Mir

Machine Learning Engineer: Leon Eckert

Currently the project is maintained by Hang.

<https://dataselfie.it/#/>

Web Crawler



Javascript plugin, activates and listens Facebook website HTML/JS events/activity, data crawler

- When do you scroll, where do you look, how long
- Which post, and what users belongs to.
- Some tags and images belonging each posts

Data is stored locally in IndexedDB, using Dexie
Check the browser console, to see Facebook activity log.

Not Facebook specific, can be adapted to other websites.

Analytics



Some self analytics about Facebook usage start appearing after a while:

- Top friends
- Activity
- Top pages
- Top likes
- Object detection (YOLO)
- Sentiment Analysis (Watson)
- Categories
- Personality prediction
- Religious Orientation
- Political Orientation
- Other Predictions
- Shopping Preferences
- Health + Activity + Other preferences

Live Demo

Implementation

The screenshot shows the GitHub profile for the organization DATA X, located in New York, NY. The profile has 4 repositories, 0 people, and 0 projects. A search bar and filters for repository type and language are visible. The repository list includes:

- data-selfie**: A browser extension to track yourself on Facebook and analyze your data. It is a JavaScript project with 877 stars, 165 forks, and a GPL-3.0 license. It was updated 11 days ago.
- facebook-cleaner**: A browser extension that cleans out a Facebook user's ad preferences, categories, and likes on a periodic basis. It is a JavaScript project with 7 stars, updated 26 days ago.
- data-selfie-image-classification**: A Python project with 3 stars and 1 fork, updated on Jan 18.
- node_apply-magic-sauce**: A JavaScript project with 39 stars and 11 forks, updated on Jun 16, 2017.

On the right side of the repository list, there are two boxes: 'Top languages' showing JavaScript and Python, and 'People' which notes that the organization has no public members.

Source code available here: <https://github.com/d4t4x>

Data-selfie is the JS plugin for data crawler.

Images are classified with Data-selfie-image-classification server, implements YOLO image analysis is implemented as a separate python server, it uses some Pyyolo and Darknet, precalculated weights from somewhere. (is in the source code)

Watson NLP API's are used for sentiment analysis

Node_apply-magic-sauce is the code implementing to traits personality ,orientation, etc.

Magic Sauce

Wonder about those personality prediction scores ?

https://applymagicsauce.com/documentation_traits.html

<https://www.psychometrics.cam.ac.uk/productsservices/apply-magic-sauce>



[You](#) [About Us](#) [Research](#) [Transparency](#) [Documentation](#) [Contact Us](#)

Predicted Traits

From Like IDs

Linear Traits

- The scores are expressed as percentiles (ranging from 0 to 1) relative to the general population. A percentile indicates the position of a given score in the context of other scores. For example, the Extraversion score of .2 (or the 20th percentile) indicates that .2 (or 20%) of the people in general population have a lower score and .8 (80%) have a higher score. Score of .5 (or 50th percentile) indicates an average score.
- Prediction accuracy is expressed as the correlation between the AMS prediction and the actual score. Accuracy of 1 indicates a perfect accuracy, whereas the accuracy of 0 indicates a random guess.
- If there are more traits in a group, using the group name as a Trait ID will calculate predictions for all traits in the group. E.g. BIG5, Religion, Politics.

More About Trait Predictions

Private traits and attributes are predictable from digital records of human behavior

<http://www.pnas.org/content/pnas/110/15/5802.full.pdf?with-ds=yes>

Private traits and attributes are predictable from digital records of human behavior

Michal Kosinski^{a,1}, David Stillwell^a, and Thore Graepel^b

^aFree School Lane, The Psychometrics Centre, University of Cambridge, Cambridge CB2 3RQ United Kingdom; and ^bMicrosoft Research, Cambridge CB1 2FB, United Kingdom

Edited by Kenneth Wächter, University of California, Berkeley, CA, and approved February 12, 2013 (received for review October 29, 2012)

We show that easily accessible digital records of behavior, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. The analysis presented is based on a dataset of over 58,000 volunteers who provided their Facebook Likes, detailed demographic profiles, and the results of several psychometric tests. The proposed model uses dimensionality reduction for preprocessing the Likes data, which are then entered into logistic/linear regression to predict individual psychodemographic profiles from Likes. The model correctly discriminates between homosexual and heterosexual men in 88% of cases, African Americans and Caucasian Americans in 95% of cases, and between Democrat and Republican in 85% of cases. For the personality trait "Openness," prediction accuracy is close to the test-retest accuracy of a standard personality test. We give examples of associations between attributes and Likes and discuss implications for online personalization and privacy.

social networks | computational social science | machine learning | big data | data mining | psychological assessment

A growing proportion of human activities, such as social interactions, entertainment, shopping, and gathering information, are now mediated by digital services and devices. Such digitally mediated behaviors can easily be recorded and analyzed, fueling the emergence of computational social science (1) and new

browsing logs (11–15). Similarly, it has been shown that personality can be predicted based on the contents of personal Web sites (16), music collections (17), properties of Facebook or Twitter profiles such as the number of friends or the density of friendship networks (18–21), or language used by their users (22). Furthermore, location within a friendship network at Facebook was shown to be predictive of sexual orientation (23).

This study demonstrates the degree to which relatively basic digital records of human behavior can be used to automatically and accurately estimate a wide range of personal attributes that people would typically assume to be private. The study is based on Facebook Likes, a mechanism used by Facebook users to express their positive association with (or "Like") online content, such as photos, friends' status updates, Facebook pages of products, sports, musicians, books, restaurants, or popular Web sites. Likes represent a very generic class of digital records, similar to Web search queries, Web browsing histories, and credit card purchases. For example, observing users' Likes related to music provides similar information to observing records of songs listened to online, songs and artists searched for using a Web search engine, or subscriptions to related Twitter channels. In contrast to these other sources of information, Facebook Likes are unusual in that they are currently publicly available by default. However, those other digital records are still available to numerous parties (e.g., governments, developers of Web browsers, search engines, or Facebook applications), and, hence, similar predictions are unlikely to be limited to the Facebook environment.

The design of the study is presented in Fig. 1. We selected traits and attributes that reveal how accurate and potentially intrusive

Final

Is a very simple, but very powerful and interesting tool.

Have fun with the source, the code is implemented as a concept for Facebook, but probably can be adapted to any website.