

AlfaDatizando: Visualización de contenido generado por usuarios de redes sociales

Paladino Jeziel¹[0000-0002-5457-518X], Lliteras Alejandra Beatriz^{1,2}[0000-0002-4148-1299], Gardey Juan Cruz^{1,3}[0000-0002-1765-8189], Grigera Julián^{1,2,3}[0000-0002-7962-4312]

¹ UNLP, Facultad de Informática, LIFIA. 50 y 120. La Plata. Bs.As. Argentina

² CICPBA. Bs.As. Argentina

³ CONICET. Argentina

{lpaladino, lliteras, jcgardey, julian.grigera}@lifia.info.unlp.edu.ar

Abstract. La visualización de datos para desarrollar la habilidad de Pensamiento Computacional permite trabajar con los estudiantes secundarios diferentes tipos de datos y contenidos provenientes de diversas fuentes. En particular, los contenidos generados por los usuarios de las redes sociales, lo que podría ser usado en materias relacionadas a las humanidades digitales. En este trabajo se presenta un relevamiento de redes sociales que permiten acceder al contenido generado por sus usuarios a través de APIs. Para este trabajo, se realizó un trabajo en dos etapas, primero, el relevamiento bibliográfico de artículos que mencionan la posibilidad de acceder a contenido y luego, la validación analizando la disponibilidad de las APIs con el fin de consumir posteriormente, el contenido desde la plataforma de visualización AlfaDatizando.

Keywords: Redes Sociales, Contenido Generado por Usuarios, API, Visualización de Datos, Pensamiento Computacional, Humanidades Digitales, Ciencias Sociales

1 Motivación

El presente trabajo se enmarca en el Proyecto de Innovación con Alumnos (2022), de la Facultad de Informática, UNLP, llamado “Aprendo con Datos. Plataforma para la visualización de datos con fines educativos en nivel secundario para Ciencias Sociales y Humanidades”. La temática abordada es parte del proyecto de doctorado de la profesora Lliteras Alejandra.

Las nuevas formas de comunicación digital han permitido nuevos tipos de interacciones entre los miembros de los movimientos sociales, incluso permitiendo interacciones entre actores que no tenían relaciones previas y que nunca se han visto cara a cara. Los recientes desarrollos en datos, tecnología y métodos de análisis ofrecen oportunidades para que el análisis de redes sociales desempeñe un papel destacado en el nuevo mundo de la investigación de las Ciencias Sociales [1].

Por otro lado, es común el uso de contenido de la red social Twitter [2] mediante plataformas de visualización (por ejemplo, SocioViz¹) en actividades relacionadas a investigación en el área de Humanidades Digitales y de las Ciencias Sociales.

¹ <https://socioviz.net/>

AlfaDatizando²[3] es una plataforma de visualización de datos para desarrollar Pensamiento Computacional en estudiantes secundarios con foco en las Humanidades Digitales y las Ciencias Sociales, que en la actualidad cuenta con visualización de datos provenientes de la API de Twitter y de fuentes de datos en archivos csv. Sabiendo que las redes sociales de mayor alcance son [4]: YouTube, Facebook, Instagram, Pinterest, LinkedIn, Snapchat, Twitter, WhatsApp, TikTok, Reddit y Nextdoor, se analizarán en la bibliografía estudios que muestren que se usó alguna API oficial para acceder a contenido generado por los usuarios de alguna de las redes sociales.

En este trabajo se presenta un relevamiento y análisis de bibliografía para determinar las redes sociales que proveen de una API oficial para acceder al contenido generado por sus usuarios, para luego determinar, si en la actualidad, dichas redes sociales efectivamente cuentan con esa posibilidad, para posteriormente implementar su acceso desde AlfaDatizando.

2 Aporte

El trabajo consiste en dos etapas, por un lado, relevamiento y análisis bibliográfico respecto a las redes sociales sobre las cuales se accede al contenido generado por los usuarios y luego, conociendo las redes sociales de mayor alcance, cuáles de ellas de acuerdo con las APIs disponibles, era posible acceder al contenido generado por sus usuarios

La **metodología** empleada para este trabajo consistió en armar el siguiente *string* de búsqueda:

("social network" and "api" and "development" and "case study" and ("user content" OR "users data")) -"data security" -"CYBERATTACKS" -"blockchain" -"privacy" -"Domain Specific Language" -DSL

Como motor de búsqueda, se usó *Google Scholar* y se limitó la búsqueda entre los años 2018-2022 ya que es de suma importancia información reciente sobre las redes sociales. Nótese el caso de Meta (antes Facebook) que periódicamente fue añadiendo restricciones en cuanto al uso de sus API 's, provocando que muchos desarrollos e investigaciones, queden obsoletas.

A continuación, en la Tabla 1, se detallan los criterios de inclusión para los trabajos que retorne la búsqueda.

Tabla 1: Criterios de inclusión

#	Criterio
I1	Descripción sobre cómo usar la API de una red social.
I2	Presentación de la API de una red social.

² <http://www.alfadatizandonos.okd.lifia.info.unlp.edu.ar/>

I3	Utilización de los datos de los usuarios de una red social por medio de APIs oficiales provistos por la misma red social.
I4	Artículo en inglés.
I5	Muestra cómo obtener el contenido generado por los usuarios.

Mientras que los criterios de exclusión se detallan en la Tabla 2.

Tabla 2: Criterios de exclusión

#	Criterio
E1	La red social utilizada no se encuentra entre las de mayor alcance [4]
E2	No describe cómo usar la API de una red social.
E3	No presenta la API de una red social.
E4	No utiliza los datos de los usuarios de una red social por medio de APIs oficiales provistos por la misma red social.
E5	No está en inglés.
E6	No muestra cómo obtener el contenido generado por los usuarios.

Una vez *ejecutada la búsqueda* en Google Scholar usando el string de búsqueda, se obtuvieron 112 resultados. En base a la lectura del título, abstract y palabras claves de cada trabajo del conjunto de resultado, y aplicando los criterios de inclusión/exclusión, 33 trabajos fueron incluidos, mientras que 79 fueron excluidos. De los 33 trabajos que quedaron al cumplir alguno de los criterios de inclusión a partir de la lectura del título, abstract y palabras claves, se procedió a la lectura de trabajo completo y nuevamente se aplicaron los criterios de inclusión/exclusión. Luego de esta segunda fase, de los 33 trabajos, quedan 9 incluidos. La Fig. 1 muestra el proceso de análisis de los resultados.

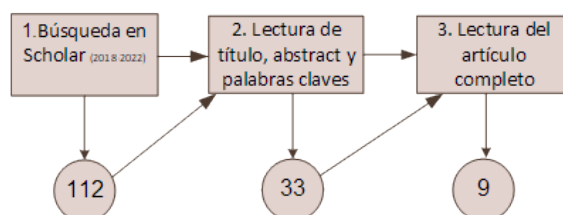


Fig. 1: Proceso de selección de artículos

De los 33 trabajos que quedaron al aplicar por primera vez los criterios de inclusión y exclusión, 17 eran artículos de revistas, conferencias y congresos, 14 responden a trabajos de tesis y 1 a un libro. La Tabla 3 muestra la distribución de los trabajos.

Los 9 trabajos resultantes son [9], [14], [16], [20], [22], [23], [25], [27] y [29]. De éstos el trabajo [20] usa la API de YouTube, mientras que los restantes la API de Twitter.

Tabla 3: Trabajos identificados por tipo

Artículo	[5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22]
Tesis	[23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36]
Libro	[37]

De acuerdo con el relevamiento bibliográfico, solo las redes sociales de Twitter y YouTube muestran casos de acceso al contenido generado por los usuarios. Como se mencionó anteriormente, las redes sociales más difundidas son YouTube, Facebook, Instagram, Pinterest, LinkedIn, Snapchat, Twitter, WhatsApp, TikTok, Reddit y Nextdoor. Por lo anterior, es que se analizó para cada una de estas redes si tenía disponible una API para acceder al contenido de los usuarios más allá de que no hayan aparecido en la búsqueda bibliográfica. La Tabla 4 muestra para cada red social, si tiene API disponible a la fecha para acceder a contenido generado por el usuario y la url de acceso a la API general de la red.

Tabla 4: Análisis de redes sociales

Red social	API Contenido	Url desde donde se la accede
YouTube	Si	https://developers.google.com/youtube/v3
Facebook	No	https://developers.facebook.com/
Instagram	No	https://developers.facebook.com/docs/instagram
Pinterest	No	https://developers.pinterest.com
Linkedin	No	https://developer.linkedin.com/
Snapchat	No	https://developers.snap.com/
Twitter	Si	https://developer.twitter.com/en
WhatsApp	No	https://developers.facebook.com/products/whatsapp/
TikTok	No	https://developers.tiktok.com/
Reddit	Si	https://www.reddit.com/dev/api/
Nextdoor	No	No posee API

Luego, y de acuerdo con lo relevado y analizado, sólo es posible acceder al contenido generado por los usuarios de las redes sociales Twitter, YouTube y Reddit

3 Líneas de Investigación Futura

A partir del relevamiento bibliográfico y análisis de APIs de redes sociales, se espera implementar en AlfaDatizando el acceso y visualización del contenido generado por los usuarios en las redes sociales YouTube y Reddit, considerando que la plataforma ya considera la visualización de contenido generado por los usuarios de Twitter.

Referencias

1. Tindall, D., McLevey, J., Koop-Monteiro, Y., & Graham, A. (2022). Big data, computational social science, and other recent innovations in social network analysis. *Canadian Review of Sociology/Revue canadienne de sociologie*
2. Yu, J., & Muñoz-Justicia, J. (2022). Free and low-cost twitter research software tools for social science. *Social Science Computer Review*, 40(1), 124-149.
3. Lliteras A., Artopoulos A., Fernandez A. & Huarte J. AlfaDatizando: a Data Visualization Platform to work Computational Thinking in Digital Humanities. Lacló 2022. In press
4. Pew Research. <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>
5. Hajikhani, A., & Porras, J. (2018). Advanced Methods: Operationalizing Social Network Services Data—Deep Content Analysis to Comprehend Brand Presence. In *Innovation Discovery: Network Analysis of Research and Invention Activity for Technology Management* (pp. 471-502).
6. Babvey, P., Lipizzi, C., & Ramirez-Marquez, J. E. (2019, December). Dissecting Twitter discussion threads with topic-aware network visualization. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1359-1364). IEEE.
7. Juric, T. (2022). Ukrainian refugee integration and flows analysis with an approach of Big Data: Social media insights. medRxiv.
8. Uyheng, J., & Carley, K. M. (2021, May). Computational Analysis of Bot Activity in the Asia-Pacific: A Comparative Study of Four National Elections. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 15, pp. 727-738)
9. Abu-Salih, B., Qudah, D. A., Al-Hassan, M., Ghafari, S. M., Issa, T., Aljarah, I., ... & Alqahtan, S. (2022). An Intelligent System for Multi-topic Social Spam Detection in Microblogging. arXiv preprint arXiv:2201.05203.
10. Rodríguez-Ibáñez, M., Gimeno-Blanes, F. J., Cuenca-Jiménez, P. M., Soguero-Ruiz, C., & Rojo-Álvarez, J. L. (2021). Sentiment Analysis of Political Tweets from the 2019 Spanish Elections. *IEEE Access*, 9, 101847-101862.
11. Chen, S., Chen, S., Wang, Z., Liang, J., Wu, Y., & Yuan, X. (2018). D-map+ interactive visual analysis and exploration of ego-centric and event-centric information diffusion patterns in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(1), 1-26.
12. Hussain, J., Satti, F. A., Afzal, M., Khan, W. A., Bilal, H. S. M., Ansaar, M. Z., ... & Lee, S. (2020). Exploring the dominant features of social media for depression detection. *Journal of Information Science*, 46(6), 739-759.

13. Masood, F., Almogren, A., Abbas, A., Khattak, H. A., Din, I. U., Guizani, M., & Zuair, M. (2019). Spammer detection and fake user identification on social networks. *IEEE Access*, 7, 68140-68152.
14. Tago, K., & Jin, Q. (2018). Influence analysis of emotional behaviors and user relationships based on twitter data. *Tsinghua Science and Technology*, 23(1), 104-113.
15. Burini, F., Cortesi, N., Gotti, K., & Psaila, G. (2018). The urban nexus approach for analyzing mobility in the smart city: towards the identification of city users networking. *Mobile Information Systems*, 2018.
16. Aggrawal, N., & Arora, A. (2019). Behaviour of viewers: YouTube videos viewership analysis. *International Journal of Business Innovation and Research*, 20(1), 106-128.
17. Saeed, M. U., & Hassan, T. U. (2020). Relationship Among the Attributes of World Countries and Their Coverage in Tweets of International News Agencies: 2010–2016. *Indian Journal of Science and Technology*, 13(08), 966-982.
18. Madanian, S., Airehrour, D., Samsuri, N. A., & Cherrington, M. (2021, October). Twitter Sentiment Analysis in Covid-19 Pandemic. In *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 0399-0405). IEEE.
19. Burini, F., Cortesi, N., Gotti, K., & Psaila, G. (2018). Research Article The Urban Nexus Approach for Analyzing Mobility in the Smart City: Towards the Identification of City Users Networking.
20. Ashraf, N., Zubiaga, A., & Gelbukh, A. (2021). Abusive language detection in youtube comments leveraging replies as conversational context. *PeerJ Computer Science*, 7, e742.
21. Wu, B., Cheng, W. H., Zhang, Y., Cao, J., Li, J., & Mei, T. (2018). Unlocking author power: On the exploitation of auxiliary author-retweeter relations for predicting key retweeters. *IEEE Transactions on Knowledge and Data Engineering*, 32(3), 547-559.
22. Shugars, S., Gitomer, A., McCabe, S., Gallagher, R. J., Joseph, K., Grinberg, N., ... & Lazer, D. (2021). Pandemics, protests, and publics: Demographic activity and engagement on Twitter in 2020. *Journal of Quantitative Description: Digital Media*, 1.
23. Haselwood, S. M. (2018). Evolution of educator professional development in the age of social media: A case study of the# Oklaed community of practice on Twitter (Doctoral dissertation, Oklahoma State University).
24. Krochmal, G. (2020). Sentiment of tweets and socio-economic characteristics as the determinants of voting behavior at the regional level. Case study of 2019 Polish parliamentary election. *arXiv preprint arXiv:2010.03493*.
25. Kouvela, M. (2020). Bot detective: explainable bot detection in twitter. A thesis submitted in fulfillment of the requirements for the degree of Master of Data & Web Science.
26. Binzagr, F. A. (2022). Social Intelligence Approach for Service-Oriented Software (Doctoral dissertation).
27. Vitali, L. (2020). Analysis and detection of social bots on Twitter mimicking human interests in people or contents.
28. Alduaiji, N. (2019). Edge Attribute-enhanced Community Discovery in Social Networks.
29. Rai, A. (2018). Inferring landscape preferences from social media using data science techniques (Doctoral dissertation, University of Illinois at Urbana-Champaign).
30. Vayansky, I. (2018). An evaluation of geotagged Twitter data during Hurricane Irma using sentiment analysis and topic modeling for disaster resilience (Doctoral dissertation, Coastal Carolina University).
31. Pitenis, Z. (2019). Detecting offensive posts in greek social media.
32. Rehman, F. U. (2018). Towards a Framework for Multiscale Social Event Extraction and Visualization (Doctoral dissertation, Université Grenoble Alpes).

33. Du, J. (2019). VaxInsight: an artificial intelligence system to access large-scale public perceptions of vaccination from social media.
34. Matamoros-Fernandez, A. (2018). Platformed racism: The Adam Goodes war dance and booing controversy on twitter, YouTube, and Facebook (Doctoral dissertation, Queensland University of Technology).
35. Apong, R. A. A. H. M. (2018). Mining negation and uncertainty in social healthcare networks. The University of Manchester (United Kingdom).
36. Saleem, H. M. (2022). Abusive language through the lens of online communities.
37. Nahili, W., Rezeg, K., & Miloudi, L. (2018). Towards better decision-making with twitter sentiment analysis. *Business Intelligence & Big Data*