

GUAITA: Monitorización y análisis de redes sociales para la ayuda a la toma de decisiones

GUAITA: Monitoring and analysis of social media to help decision making

Ferran Pla¹, Lluís-F. Hurtado¹, José-Á. González¹, Vicent Ahuir¹, Encarna Segarra¹, Emilio Sanchis¹, María-José Castro¹ and Fernando García¹

¹VRAIN: Valencian Research Institute for Artificial Intelligence, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

Abstract

El proyecto GUAITA tiene como objetivo extraer grandes cantidades de información proveniente de las redes sociales y proporcionar herramientas de análisis de dicha información que puedan ser útiles para la toma de decisiones de las organizaciones. En ese sentido, instituciones y empresas pueden beneficiar de los avances logrados. Fruto de este proyecto en la actualidad se dispone de un prototipo software que integra diferentes modelos basados en redes neuronales para la monitorización y análisis multilingüe de Twitter. Aunque ya existen en el mercado herramientas de recogida de datos y de análisis, un factor diferencial de GUAITA es el uso de modelos de estado del arte en el análisis del lenguaje natural en redes sociales. Esto permite ampliar la funcionalidad básica de análisis de sentimiento, detectando, por ejemplo, el lenguaje inapropiado, el discurso del odio o el nivel de toxicidad presente en los mensajes. Además, GUAITA tiene en consideración idiomas habitualmente no considerados por este tipo de herramientas como es el caso del catalán.

English translation. The GUAITA project aims to extract large amounts of information from social media and provide tools for the analysis of this information that may be useful for decision-making in organizations. In this sense, institutions and companies can benefit from the progress achieved. As a result of this project, there is currently a software prototype that integrates different models based on neural networks for multilingual monitoring and analysis of Twitter. Although there are already data collection and analysis tools on the market, a differentiating factor of GUAITA is the use of state-of-the-art models in the analysis of natural language in social networks. This allows the analysis to be extended not only to sentiment analysis but also to go further and be able to detect, among others, inappropriate language, hate speech or the level of toxicity present in the messages. In addition, GUAITA takes into account languages that are not usually considered by this type of tool, such as Catalan.

Keywords

Social Media, Natural Language Processing, Neural Networks.

1. Introducción

La tecnología actual ha permitido que la información disponible para la toma de decisiones sea cada vez más abundante y oportuna; esto, unido a nuevos desarrollos en ciencia de datos, ha ayudado a mejorar la velocidad de reacción y la calidad de dichas decisiones por parte de las empresas y organizaciones. Uno de los grandes apoyos en

esta nueva toma de decisiones es el desarrollo de plataformas, servicios y modelos de analítica avanzada y visualización de datos. Se hace necesario desarrollar herramientas que presenten una interfaz amigable para el usuario y que no impliquen un alto coste de implantación, tanto económico como temporal.

Aunque ya existen en el mercado herramientas de recogida de datos y de análisis, estas herramientas están todavía en un nivel de desarrollo muy inicial cuando la fuente de datos son las redes sociales. El procesamiento automático del lenguaje natural utilizado en este tipo de redes constituye un problema abierto dentro de la comunidad científica, por lo que la transferencia al mercado de herramientas de las tecnologías del habla logradas dentro del área de investigación del procesamiento del lenguaje natural es de gran interés.

En la actualidad existe una gran cantidad de compañías que ofrecen herramientas destinadas a

SEPLN-PD 2022. Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations, September 21-23, 2022, A Coruña, Spain

✉ fpla@dsic.upv.es (F. Pla); lhurtado@dsic.upv.es (Lluís-F. Hurtado); jogonba2@inf.upv.es (José-Á. González); viahes@eui.upv.es (V. Ahuir); esegarra@dsic.upv.es (E. Segarra); esanchis@dsic.upv.es (E. Sanchis); mcastro@dsic.upv.es (M. Castro); fgarcia@dsic.upv.es (F. García)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CEUR Workshop Proceedings (CEUR-WS.org)



WS.org

dar soporte a los Community Manager (CM) en su labor de gestionar la presencia corporativa en redes sociales. Típicamente permiten la agregación en una única plataforma o aplicación de múltiples cuentas de usuario en varias redes sociales.

Además del soporte a la gestión del CM, muchas de las herramientas permiten el análisis de la actividad de usuarios en redes sociales, fundamentalmente en Twitter. Entre algunas de estas herramientas que permiten monitorizar la actividad de usuarios podemos destacar las siguientes: Twitter Analytics, Followerwonk, Twitonomy, Rebold, Brandwatch Consumer Research, Buffer, BuzzSumo, Klear, Union Metrics, Mentionmapp, Foller.me, PressClipping.

De entre las principales características comunes a estas herramientas podemos destacar las siguientes. Todas ellas suelen ser herramientas de pago y que se basan esencialmente en la web con gran relevancia del componente gráfico. En algunos casos permiten exportar la información utilizando formatos estándar.

La mayoría de las herramientas están diseñadas para la monitorización de la red social Twitter aunque algunas contemplan otras redes sociales.

La información proporcionada por estas herramientas se basa principalmente en el análisis de los metadatos proporcionados por las redes. Esta información consiste en datos agregados y la distribución temporal de estos: número de seguidores, repercusión de un post a lo largo del tiempo (número de me gusta, número de retweets, número de replicas). En algunos casos se incluye la información geográfica de los tweets, basada en la geolocalización de los usuarios o en su perfil. En general, no se realiza un análisis profundo del contenido textual.

En este trabajo se presenta una demostración de los principales logros obtenidos en el proyecto GUAITA: Monitorización y análisis de redes sociales para la ayuda a la toma de decisiones subvencionado por la Agencia Valenciana de la Innovació (AVI) de la Generalitat Valenciana. Se describe el sistema desarrollado incluyendo su arquitectura y principales funcionalidades.

El sistema integra diferentes modelos basados en redes neuronales para la monitorización y análisis multilingüe de la red social Twitter. Un factor diferencial de GUAITA es el uso de modelos de estado del arte en el análisis del lenguaje natural en redes sociales. Esto permite ampliar el análisis no solo al análisis de sentimientos sino ir más allá y ser capaz de detectar, entre otros, el lenguaje inapropiado, el discurso del odio o el nivel de toxicidad presente en los mensajes. Asimismo, GUAITA incluye el

catalán entre los idiomas soportados; idioma que no suele estar considerado por otras herramientas de esta índole.

2. Descripción del sistema

El sistema GUAITA está concebido como una herramienta software que permita el seguimiento de acontecimientos, personas o cualquier tema de interés para el usuario en la red social Twitter.

Las principales funcionalidades del sistema son las siguientes:

- *Seguimiento de redes sociales.* Permite realizar la monitorización de la red social Twitter para la obtención y almacenamiento de la información relacionada con el tema de interés. Para ello, nos permite definir tareas y programarlas en el tiempo. En cada tarea se pueden definir capturas (búsquedas de Twitter) siguiendo los criterios que se consideren oportunos.
- *Obtención de modelos específicos de análisis de textos para diferentes lenguas.* El sistema también permite la recolección de textos que sean útiles para aprender modelos específicos en una lengua en concreto o dominio. La herramienta dispone de modelos para el español, inglés y catalán que permiten el procesado y etiquetado de corpus en estas lenguas.
- *Visualización de resultados.* El sistema presenta gráficamente los resultados de los análisis desarrollados mediante una serie de interfaces web. También dispone de un generador de informes, que de forma automática, elabora un dossier de toda la información relacionada con una tarea definida por el usuario. Dichos informes pueden ser de gran utilidad para su análisis con el fin de determinar la reputación de una institución o compañía.
- *API REST.* Permite la comunicación de nuestra aplicación con aplicaciones de terceros.

3. Arquitectura del sistema

En la Figura 1 se muestra la arquitectura general de la aplicación y las interconexiones entre los distintos módulos que la componen. Como se puede observar, el sistema GUAITA está compuesto por cuatro módulos principales que se describirán de forma sucinta en esta sección.

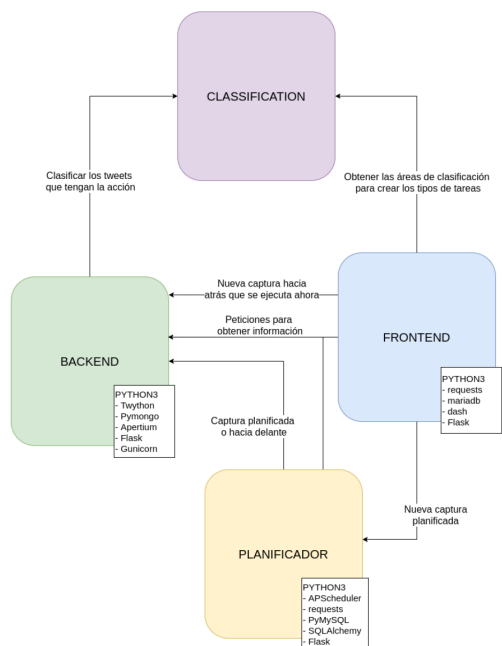


Figure 1: Arquitectura del Sistema GUAITA.

3.1. Backend

El Backend es el núcleo del sistema GUAITA. Este módulo es el responsable de gestionar el comportamiento de toda la aplicación. Fundamentalmente realiza la descarga de contenido de redes sociales y las peticiones al motor de clasificación y perfilado de usuario, genera las estadísticas y datos necesarios para la presentación gráfica de la información; información que será accesible mediante la API REST del módulo.

El Backend consta de una capa de persistencia dividida en dos. Por un lado se utiliza una instancia distribuida de una base de datos orientada a documentos MongoDB donde se guarda toda la información referente al análisis de cada captura. Por otro lado, se utiliza MariaDB para la persistencia relacionada con los procesos de captura. Esta capa es la responsable del almacenamiento en memoria secundaria de toda la información necesaria para el funcionamiento global de la aplicación desde el contenido descargado de Twitter o la información generada por el motor de clasificación.

3.2. Planificador

Este módulo se encarga de enviar capturas periódicas o futuras al Backend. En el contexto de la monitorización de redes sociales es habitual querer planificar tareas con antelación, por ejemplo para seguir el impacto de una nueva campaña publicitaria. También es habitual querer hacer consultas recurrentes en el tiempo, por ejemplo, para seguir un programa de televisión que se emite siempre a la misma hora un día determinado de la semana.

3.3. Motor de Clasificación

GUAITA permite obtener información del contenido textual de los tweets descargados mediante el uso de modelos de clasificación basados en redes neuronales. Todos estos modelos están aprendidos utilizando arquitecturas neuronales del estado del arte y corpus de múltiples competiciones internacionales. En la Sección 4 se describen los modelos y corpus utilizados.

3.4. Frontend

El sistema GUAITA está dotado de una interfaz de programación de aplicaciones (API REST) que le permite ser utilizado e integrado en software de terceros. Sin embargo, también existe una versión web que facilita el uso de la herramienta a usuarios humanos. El Frontend es el encargado de gestionar los formularios y demás páginas de la aplicación web y realizar las peticiones al Backend utilizando la API. En la Figura 2 se muestra parte de la salida gráfica proporcionada por la aplicación para la consulta: #SagitarioA OR #SagittariusA OR #BlackHole.

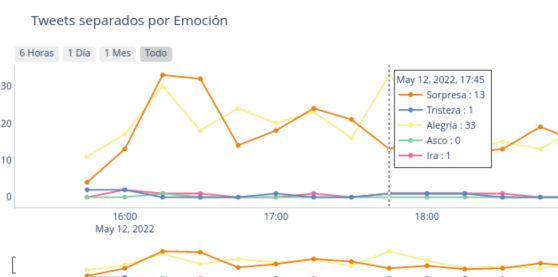


Figure 2: Salida del sistema, detección de emociones.

4. Modelos de clasificación

Para el motor de clasificación se han utilizado tres *encoders* : TWILBERT[1], BETO[2] y XLM-T[3]. Se han utilizado corpus de múltiples competiciones para aprender diversos modelos de clasificación. GUAITA utiliza cada uno de los tres *encoders* dependiendo de la tarea y en función del rendimiento. Los corpus utilizados han sido los que se enumeran a continuación. TASS 2019[4] para los modelos de polaridad, Irosva 2019[5] para los modelos de detección de ironía, EmoEvalEs 2021[6] para la detección de emociones y lenguaje ofensivo, HateEval 2019[7] para los modelos de detección de lenguaje del odio y agresividad, HaHa 2019[8] para la detección de humor presente en los tweets y Detoxis 2021[9] para varios modelos: lenguaje impropio, sarcasmo, toxicidad, etc.

5. Conclusiones y trabajos futuros

En este trabajo se ha presentado el sistema desarrollado en el proyecto GUAITA: Monitorización y análisis de redes sociales para la ayuda a la toma de decisiones. Se ha descrito su arquitectura y principales funcionalidades actuales. No obstante GUAITA es una herramienta en constante crecimiento y mejora. Entre las ampliaciones que se pretenden incorporar podemos destacar la detección de aspectos y las alertas automáticas ante mensajes que fomenten el odio durante el seguimiento de algún evento.

Acknowledgments

Este trabajo ha sido parcialmente subvencionado por la Agencia Valenciana de la Innovación (AVI) de la Generalitat Valenciana, proyecto GUAITA (INNVA1/2020/61), el Vicerrectorado de Investigación de la Universitat Politècnica de València (PAID-11-21) y por el Ministerio de Ciencia e Innovación y fondos de la Unión Europea con el proyecto BEWORD PID2021-126061OB-C41.

References

- [1] J.-Á. González, L.-F. Hurtado, F. Pla, Twilbert: Pre-trained deep bidirectional transformers for spanish twitter, *Neurocomputing* 426 (2021) 58–69. URL: <https://www.sciencedirect.com/science/article/pii/S0925231220316180>. doi:<https://doi.org/10.1016/j.neucom.2020.09.078>.
- [2] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: *PML4DC at ICLR 2020*, 2020.
- [3] F. Barbieri, L. E. Anke, J. Camacho-Collados, Xlm-t: A multilingual language model toolkit for twitter, 2021. URL: <https://arxiv.org/abs/2104.12250>. doi:10.48550/ARXIV.2104.12250.
- [4] e. a. Manuel Carlos Díaz-Galiano, Overview of TASS 2019: One More Further for the Global Spanish Sentiment Analysis Corpus, in: *Proceedings of the Iberian Languages Evaluation, IberLEF@SEPLN 2019*, Bilbao, Spain, volume 2421 of *CEUR Workshop Proceedings*, 2019, pp. 550–560.
- [5] R. O. Bueno, F. M. R. Pardo, D. I. H. Farías, P. Rosso, M. M. y Gómez, J. E. Medina-Pagola, Overview of the task on irony detection in spanish variants, in: *IberLEF@SEPLN*, 2019.
- [6] F. M. Plaza-del Arco, S. M. Jiménez Zafra, A. Montejo Ráez, M. D. Molina González, L. A. Ureña López, M. T. Martín Valdivia, Overview of the emoevals task on emotion detection for spanish at iberlef 2021, 2021.
- [7] V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. Rangel Pardo, P. Rosso, M. Sanguinetti, SemEval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter, in: *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019, pp. 54–63.
- [8] L. Chiruzzo, S. Castro, A. Rosá, HAHA 2019 dataset: A corpus for humor analysis in Spanish, in: *Proceedings of the 12th Language Resources and Evaluation Conference*, 2020, pp. 5106–5112.
- [9] M. Taulé Delor, A. Ariza, M. Nofre, E. Amigó Cabrera, P. Rosso, Overview of detoxis at iberlef 2021: Detection of toxicity in comments in spanish, 2021-09.