

---

# MEMORIA FINAL

---



Título del proyecto:

**TwitterSI**

Título del documento

**Memoria Final**

Versión del documento:

**Final**

Grupo:

Jorge López Rosende  
Javier Pena Rodríguez  
Iván Humanes Cabañas

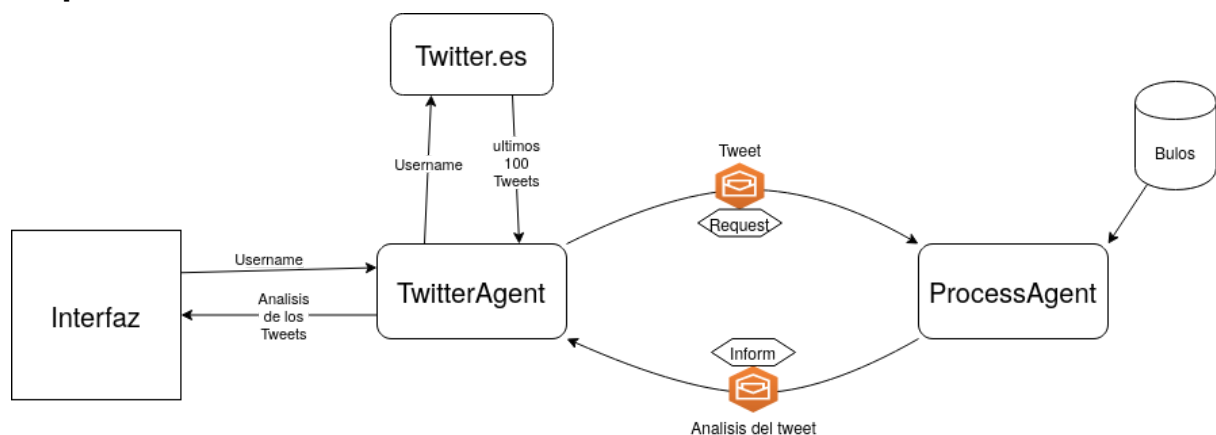
# Análisis de veracidad de una cuenta de Twitter

## Pasos a seguir en el análisis

- Obtención los últimos 100 tweets del usuario (tweet, retweet...), siempre y cuando la cuenta sea pública.
- Comparación con una base de datos de bulos. (<https://maldita.es>)
- Mostramos con un porcentaje entre 0 y 1 el índice de similitud de cada tweet del usuario, con la base de datos de bulos recopilados.



## Arquitectura



El proyecto cuenta con 2 agentes y una interfaz que describimos a continuación.

### Agente de percepción/adquisición

Este agente es el encargado de obtener los últimos 100 tweets de un usuario de twitter siempre y cuando este usuario tenga una cuenta pública.

### Agente inteligente

Este agente es el encargado de comparar los tweets obtenidos por el agente de percepción/adquisición con una API de bulos de <https://maldita.es> para llegar a la conclusión de si una cuenta de twitter distribuye o no bulos y en qué medida, a cada tweet se le da una puntuación entre 0 y 1, siendo 1 el índice de coincidencia mayor y 0 el menor, hemos considerado que cualquier valor por debajo de 0,25 el índice de coincidencia es demasiado bajo por lo que consideramos que no hay suficiente coincidencia para que sea un bulo.

En el algoritmo de análisis hemos usado el Tf-idf para crear un kernel de los bulos y generar un VSM (Vector Space Model).

Se ha pre procesando los bulos, eliminando enlaces e imágenes, caracteres especiales y stopwords.

Luego creamos un diccionario con las palabras resultantes y generamos los vectores que componen el kernel del VSM.

Por último a la query a comprobar se le aplica el mismo preprocesamiento y se calcula su índice de similitud, dándonos un ranking de similitud con todos los bulos.

## Interfaz gráfica

En la siguiente captura podemos ver como introduciendo un usuario de twitter nos muestra todos sus tweets con la puntuación obtenida al ser comparados con la API de bulos.



## Librerías utilizadas

### Python

- nltk
- operator
- gensim
- librerías core de python

### Java

- jade
- twitter4j
- swing
- awt
- librerías core de java

## Mejoras

Al usar jade y ser un plataforma de software distribuido es posible añadir nuevos agentes inteligentes que comprueben otros factores que puedan ayudar a descubrir cuentas falsas, bots o bulos en la aplicación de twitter.

Uso de algoritmos de data mining y análisis de grafos para determinar núcleos de cuentas falsas (bots).

Mejoras en la interfaz de usuario para permitir la selección y modificación de hiperparametros y ajustar la búsqueda y comprobación de los tweets.

- ej: aumentar o disminuir el índice de similitud por defecto en 0.25 en la versión inicial.