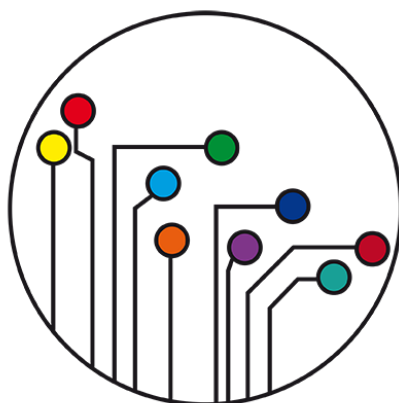


Posicionamiento, Búsqueda y Recuperación de la Información

Trabajo Búsqueda y Recuperación: Análisis de Información de Contenidos web/redes sociales

Máster Universitario en Ingeniería Informática

Curso 2020/2021



MÁSTER UNIVERSITARIO
INGENIERÍA INFORMÁTICA



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Luis Blázquez Miñambres

Contenido

Tabla de ilustraciones	3
1. Introducción	5
2. Identificación y selección de contenidos	7
3. Gestión de la información y resultado final	8
3.1. Metrica	8
3.1.1. Followers y bots.....	9
3.1.2. Otras redes sociales.....	9
3.1.3. Verificación	9
3.1.4. Engagement	9
3.1.5. Grupos de influencia.....	9
4. Conclusiones	10
Bibliografía	12

Tabla de ilustraciones

Ilustración 1: Arquitectura planteada del sistema	7
Ilustración 2: Métrica utilizada en la detección de influencers.....	8
Ilustración 3: Página principal de la vista del buscador en la web al detectar influencers para la marca 'Audi'	10
Ilustración 4: Resultado de los tres principales influencers detectados para la marca 'Audi' ..	11
Ilustración 5: Gráfica con el feedback (rts, favoritos e interacción) de los tweets de un influencer en el último mes.....	11

1. Introducción

En este documento se recoge el informe acerca de la práctica de usabilidad de la segunda parte de la asignatura “Posicionamiento, Búsqueda y Recuperación de la Información” del Máster en Ingeniería Informática por la Universidad de Salamanca en el Curso 2020-2021, relacionado con la aplicación de las técnicas de recuperación y análisis de contenidos web/redes sociales vistas en clase y justificarlas adecuadamente. Para ello se propone la realización de una plataforma cuya funcionalidad principal sea un detector de *influencers* que , en función de una serie de patrones (redes sociales en la que se encuentra el *influencer*, relevancia de los tweets, seguidores, etc.) y de herramientas de inteligencia artificial que he utilizado en mi trabajo, analizar distintas tendencias y ver cómo encajaría los *influencers* detectados para una marca o marcas determinadas en función de una métrica que utilizaré. Luego estos *influencers* serán visualizados en función de la tendencia que se escoja.

2. Identificación y selección de contenidos

En este apartado se presenta la arquitectura planteada para el sistema de detección de influencers en el que se pueden distinguir, a partir de la Ilustración 1, los siguientes subsistemas (cada uno de los cuales con un conjunto de herramientas y tecnologías a utilizar):

- **Subsistema de visualización:** en este primer subsistema se representará mediante una visualización web los resultados de los *influencers* o perfiles de usuarios resultantes del análisis y detección. Para el desarrollo de este subsistema se harán uso de Bootstrap para el diseño de la página web y la biblioteca AmCharts para la generación de gráficos y visualizaciones de representación de datos. Para la gestión o middleware entre la API de Python y el lado cliente de la web se usará Node.js con Express para la gestión de peticiones.
- **Subsistema de gestión de datos:** este subsistema servirá como nicho o punto de comunicación entre el subsistema de extracción de Twitter y la parte de inteligencia artificial con la parte web para servir los datos resultantes al lado del cliente con el fin de ser posteriormente visualizados. En este subsistema los datos serán tratados y formateados. Para el desarrollo de este subsistema se harán uso de Python como lenguaje de programación y de Flask RestXPlus para la elaboración de la API REST como punto centralizado entre el resto de los subsistemas.
- **Subsistema de extracción de Twitter:** este subsistema servirá para la comunicación con la API de Twitter para la extracción de información acerca de los perfiles de los usuarios utilizando las palabras clave adecuadas para su extracción.
- **Subsistema de análisis y detección de influencia:** este subsistema contendrá el cálculo por medio de la métrica de detección y análisis de influencia de los perfiles de los usuarios, junto con los modelos supervisados complementarios utilizados para amplificar o ayudar a concretar el resultado de la métrica. Para este subsistema también se hará uso de Python y de la biblioteca Scikit-Learn para la creación del modelo supervisado de inteligencia artificial.

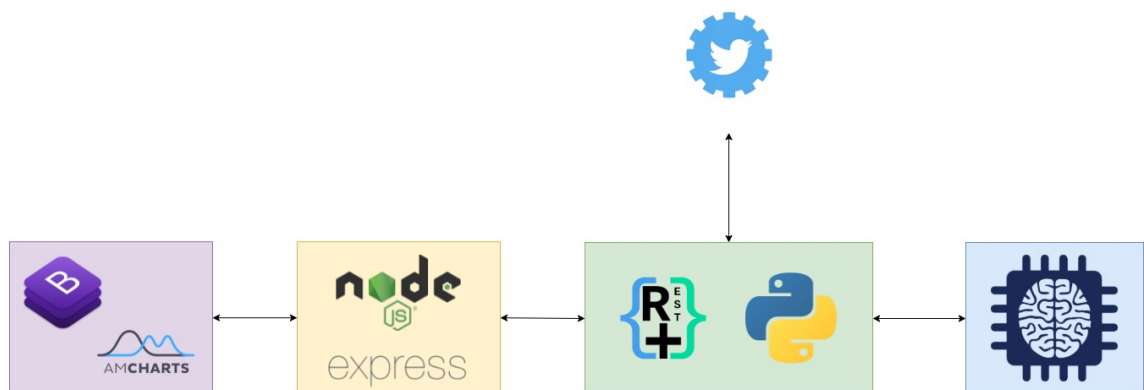


Ilustración 1: Arquitectura planteada del sistema

3. Gestión de la información y resultado final

En este apartado se explicará el cometido y funcionalidad principal de la aplicación con la que se pretende encontrar usuarios con una influencia suficiente en redes sociales como Twitter e Instagram. El objetivo es localizar posibles contratos publicitarios con la marca que use la aplicación.

Para ello, se utilizará la aplicación de ingesta de datos ya desarrollada. Gracias a dicha aplicación se buscarán perfiles de usuarios relacionados con las *keywords* que establezca la marca que use la aplicación.

A partir de ella se obtendrán un conjunto de perfiles de relacionados donde buscaremos los influencers. Obviamente no todos estos usuarios serán válidos, por lo que habrá que conseguir filtrar o cribar aquellos que tengas que ser mantenidos únicamente como influencers potenciales.

Para ello, se han establecido una serie de criterios usados en una métrica que indica el grado de influencia del usuario.

A continuación, se pretende explicar el código que permite sacar la lista con los datos de los posibles influencers. Se dividirá la explicación según los criterios o variables de la métrica utilizada.

3.1. Métrica

Una vez tenemos todos estos datos, se hace una métrica aplicando distintos pesos a los valores. En primer lugar, se ha dado **un peso de 0.20** a aquellas cuentas que tengan 0% de bots, variando dicho peso de forma inversamente proporcional respecto al porcentaje de bots.

Si el usuario posee un canal de YouTube, esto le aportará **un peso de 0.20**. (Esto se puede extender a varias redes, dando un mayor peso cuanto mayor sea el número de redes en las que se encuentre el *influencer*)

En cuanto a si la cuenta está verificada, se ha decidido darle **un peso de 0.25**, ya que es un criterio que da bastante certeza de que el usuario es alguien importante. Además, al ser un valor cualitativo, pues será cierto o no, provoca que se note más la diferencia entre aquellos usuarios no verificados y los que sí lo están.

Por último, se ha dado **un peso de 0.35** al *engagement*, pues es un criterio cuantitativo que aporta bastante información acerca de si llegan o no los mensajes del supuesto a *influencer* a sus seguidores que, al final, es el objetivo de esta funcionalidad.

La métrica se puede ver en la Ilustración 2.

$$\text{Engagement} = \left(\frac{\text{Average number of likes in posts 4-10}}{\text{Total number of followers}} \right) \times 100$$

Ilustración 2: Métrica utilizada en la detección de influencers

De forma que la formula para calcular la influencia de los perfiles quedaría como sigue:

$$\text{influencia} = 0.20 * 1(\text{si cuenta sin bot}) + 0.20 * 1(\text{si cuenta en otra RRSS}) + 0.25 * 1(\text{si cuenta verificada}) + 0.35 * 1(\text{engagement})$$

3.1.1. Followers y bots

El primer criterio establecido es el número de seguidores que tenga la cuenta. Obviamente cuanto más gente siga al usuario, más lejos llegará la publicidad que este haga. Sin embargo, hay que comprobar que dichos *followers* sean gente real. Esto se debe a que es una práctica común comprar bots para aumentar el número de seguidores.

Por tanto, la aplicación será capaz de filtrar, devolviendo un porcentaje respecto al número de seguidores, aquellas cuentas cuyo porcentaje de bots supera el valor de 40 %. Si esto ocurre se considerará el perfil como 'No Valido' y se desechará. Esto se conseguirá mediante un modelo de aprendizaje supervisado que detectará una serie de patrones para detectar cuentas fantasmas o cuentas *bot*.

3.1.2. Otras redes sociales

A la hora de buscar un *influencer*, se busca también que tenga influencia en otras plataformas. Entre estas plataformas, se valora especialmente aquellos usuarios que además tengan un canal de YouTube, pues es una plataforma con mucha interacción, donde muchas veces siguen más personas al *influencer* que en el propio Twitter

Otras plataformas como Blog spot o Tumblr también son bien valoradas positivamente dentro de esta variable en la métrica.

3.1.3. Verificación

Otro campo a tener en cuenta es si la cuenta está verificada. Normalmente se verifican cuentas que tienen gran presencia social o una credibilidad elevada. También aquellas con riesgo de suplantación.

Todo esto hace que una cuenta verificada tenga más probabilidades de ser catalogada como *influencer* que una que no

3.1.4. Engagement

El *engagement*, o participación, es una medida muy útil a la hora de calcular la influencia en las redes de un usuario. Dicho valor se obtiene al hacer la media del número de favoritos o RTs que tienen los posts del 4 al 10 y dividir el resultado entre el número de *followers*. Si además multiplicamos lo obtenido por 100, se saca el porcentaje de participación de los seguidores.

Se calcula en los posts del 4 al 10 ya que los primeros tres posts, al ser más nuevos, siguen evolucionando, y los posts a partir del décimo no son tan valiosos ya que nos interesa más el *engagement* actual que el pasado.

3.1.5. Grupos de influencia

Por último, aunque de una complejidad mayor, existen los llamados grupos de influencia. Estos son grupos de usuarios con acuerdos entre ellos, a veces económicos, para comentarse y darse *feedback*. Dicho *feedback* se realiza de manera muy rápida, lo que hace que los algoritmos de Twitter le den más visibilidad. Sin embargo, son comentarios vacíos y que no aportan nada. Además de que el *influencer* carece de un público real.

4. Conclusiones

En conclusión, tal y como se ha podido observar en el desarrollo de este informe, se ha conseguido plantear y desarrollar un sistema completo en el que, a partir de una tendencia, *topic* o palabra clave conocida conseguir una serie de perfiles de usuario que hablan del mismo en sus tweets y calcular, por medio de una métrica, la influencia o alcance que tienen estos perfiles dentro de la propia red social de Twitter y, aunque en menor medida, en otras redes sociales en general.

De esta forma se ha conseguido cumplir con uno de los objetivos de la asignatura, la detección de potenciales influencers mediante una plataforma sencilla y fácil de usar y que puedan ayudar a las empresas o marcas a potenciar el desarrollo y alcance de sus productos gracias al uso de redes sociales y del poder de comunicación de los influencers en Internet que, actualmente tienen más poder de venta que muchos de los medios tradicionales.

Finalmente, a continuación, en las Ilustraciones 3, 4 y 5 se puede observar un caso de uso del resultado de la web implementada para unificar la funcionalidad de la aplicación en la que, buscando una tendencia o palabra clave en Twitter (en este caso *influencers* para la marca de coches Audi), se extraerán el top 5 de los *influencers* que más puedan interesar a una marca o una tendencia en función de su influencia y su relación con dicha marca o tendencia.

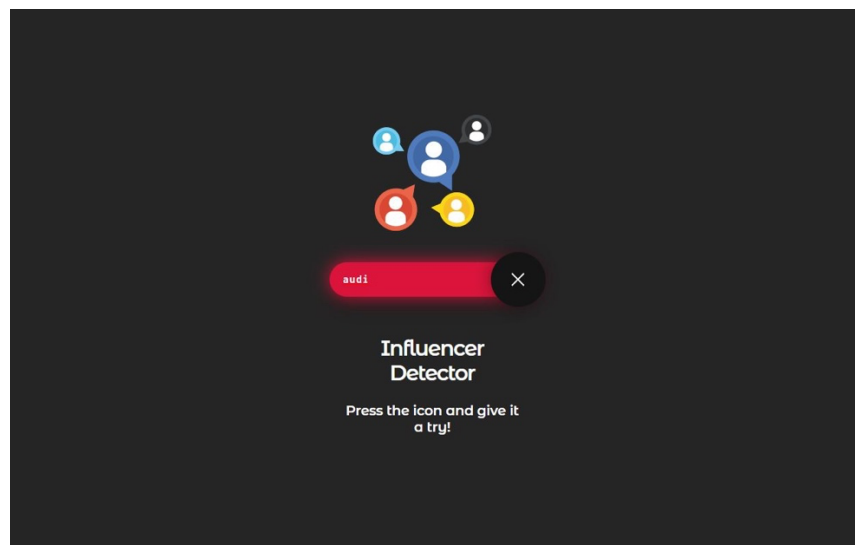


Ilustración 3: Página principal de la vista del buscador en la web al detectar influencers para la marca ‘Audi’

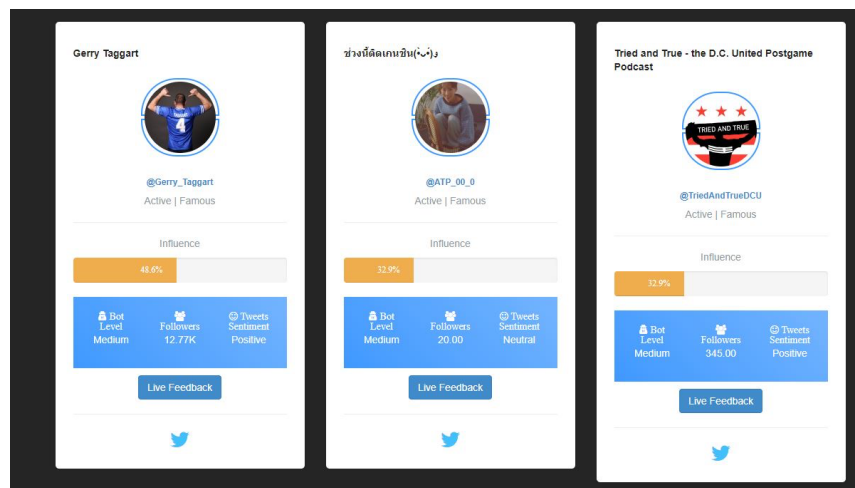


Ilustración 4: Resultado de los tres principales influencers detectados para la marca 'Audi'



Ilustración 5: Gráfica con el feedback (rts, favoritos e interacción) de los tweets de un influencer en el último mes

Bibliografia

Tools Pingdom. (s.f.). Obtenido de <https://tools.pingdom.com/>