ANÁLISIS DE REDES SOCIALES PARA RECURSOS HUMANOS

Fabio Inui **Teresa Martínez** **Javier Quintana** **Silvia Santos**

Versión preliminar September 26, 2017

Dedicamos esta memoria a nuestras familias, sin cuya paciencia sin límites, no hubiera sido posible su elaboración. Agradecemos a nuestros profesores (**a unos más que a otros**) su dedicación y ayuda.

Contenidos

Resumen ejecutivo

Executive summary

Capítulo 1

Planteamiento del proyecto

En este capítulo vamos a describir las ideas y contexto en el que vamos a desarrollar el contenido del proyecto.

1.1 Descripción

Cuando un departamento de Recursos Humanos o una empresa de reclutamiento se enfrenta a una petición para cubrir un puesto vacante o de nueva creación, el proceso suele llevarse a cabo en diversas fases, que podríamos describir del siguiente modo [1]:

1. **Preselección**: etapa inicial en la que se detectan candidatos adecuados para el perfil buscado, bien recurriendo a anuncios en portales especializados, bien con búsquedas personalizadas de perfiles. En esta etapa se elabora una lista de candidatos que pasarán a las siguientes fases del proceso, descartando aquellos cuyas competencias no sean las adecuadas para el puesto.

2. **Entrevista inicial**: en esta etapa los candidatos seleccionados en la etapa anterior son contactados para conseguir ampliar la información de la que se dispone sobre ellos (por ejemplo sobre las aptitudes particulares y experiencias previas consignadas en el CV), y verificar el interés y compromiso del candidato con respecto a la oferta.

3. **Informe**: tras la entrevista inicial, se seleccionan los mejores candidatos para el puesto, y se realiza un informe donde se consignan los datos originales (el CV, por ejemplo) y los datos añadidos en el curso de la entrevista inicial.

4. **Presentación de candidatos**: el empleador recibe el informe elaborado en el punto anterior, y selecciona aquellos que mejor se ajusten a sus necesidades, muy habitualmente realizando nuevas entrevistas con ellos.

5. **Decisión**: es el momento en que se elige el candidato al que se se le va a ofrecer el puesto, etapa en la que puede complementarse la información recogida hasta el momento con referencias recabadas de anteriores empleadores.

6. **Oferta**: etapa en la que la empresa presenta al candidato la oferta en firme, habitualmente por escrito, consignando la voluntad de la empresa de incorporar al candidato y los detalles económicos.

7. : **Seguimiento**: para comprobar que una vez incorporado a la empresa, tanto empleado como empleador están conformes con el resultado del proceso.

Tradicionalmente, el comienzo de este proceso, la detección de candidatos, se realizaba en numerosas ocasiones a través de anuncios en prensa, bases de datos de candidatos construidas a lo largo del tiempo, y la explotación de la red de contactos personales del entorno del empleador. Hoy por hoy, estos métodos tradicionales han sido complementados, y algunos dirían que prácticamente suplantados, por métodos que explotan la información contenida en la web.

Los técnicos de selección se enfrentan a un mundo muy diverso donde tanto la difusión de los posibles puestos como la información sobre los candidatos para los mismos está diseminada en numerosos formatos, teniendo un papel preponderante diversas plataformas o portales web (InfoJobs, Monster, etc.) y redes sociales en general (LinkedIn, Twitter, Facebook, Instagram, etc.). Desde el punto de vista del técnico de selección, las primeras contienen mucha información sobre las aptitudes de los posibles candidatos, sus conocimientos, formación y experiencia, ya que son portales donde los propios usuarios consignan sus curricula vitae, y también sobre su situación laboral actual y expectativas. En el segundo grupo de fuentes, las redes sociales, hay algunas que tienen el carácter específico de las primeras (LinkedIn es el ejemplo más claro), y hay otras en las que se consigna información diversa, llamémoslas de propósito general, tal vez en mayor medida personal que profesional.

El objetivo de nuestro proyecto es complementar el trabajo habitual de un departamento de Recursos Humanos o de un seleccionador de personal en los portales y redes sociales dedicados al mundo laboral, con información laboral extraída de fuentes menos estándar, como son las redes sociales de propósito general. Estas redes son a menudo aprovechadas por los usuarios para difundir mensajes relacionados con su actividad laboral, y una descripción de su actividad en las redes es relevante desde el punto de vista de un reclutador, en la medida que da información del compromiso de la persona con su actividad, su valoración por parte de otros usuarios, su proactividad, etc.

En este trabajo hemos elegido la red social Twitter por diversos motivos: es una red muy dinámica, fácil de usar, rápida y divertida, que involucra cientos de millones de usuarios activos en todo el mundo: 328 millones según la web de la red. El crecimiento del número de usuarios fue casi exponencial entre 2010 y 2014, si bien últimamente la velocidad a la que adquiere nuevos usuarios ha perido intensidad. Es también una red que, desde sus orígenes, ha puesto a disposición de los interesados los mecanismos necesarios para acceder a la información que atesora, con ciertas limitaciones, pero de forma relativamente sencilla.

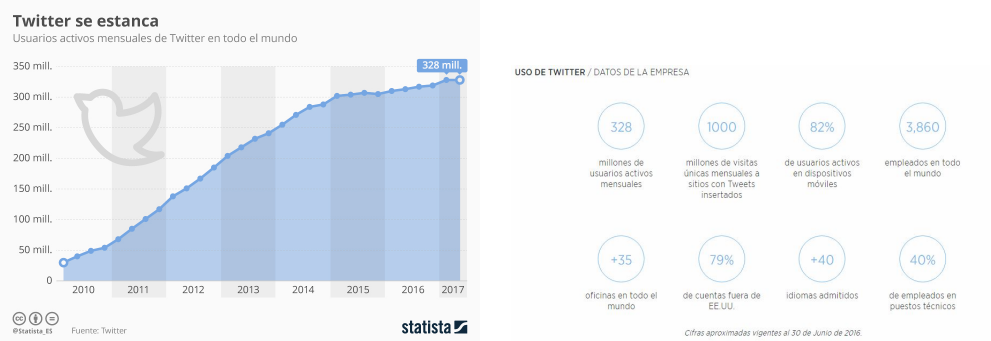


Figura 1.1.1: Twitter: evolución del número de usuarios (Statista 2017, https://es.statista.com/grafico/10476/el-numero-de-usuarios-de-twitter-se-estanca) y datos de la empresa, https://about.twitter.com/es/company.

Esta red da cabida a relaciones diversas, entre usuarios de variada índole. Dado que muchos de los usuarios publican información relacionada con su ocupación laboral, es natural esperar que en Twitter se formen comunidades de individuos que comparten interés en diferentes aspectos de dicho ámbito. Nuestro propósito es definir e implementar un proceso que permita agregar información referente a esas comunidades a un determinado proceso de selección.

Observemos las dos siguientes ofertas de trabajo aparecidas recientemente (Septiembre 2017) en LinkedIn, incluyendo los requisitos solicitados a los posibles candidatos:

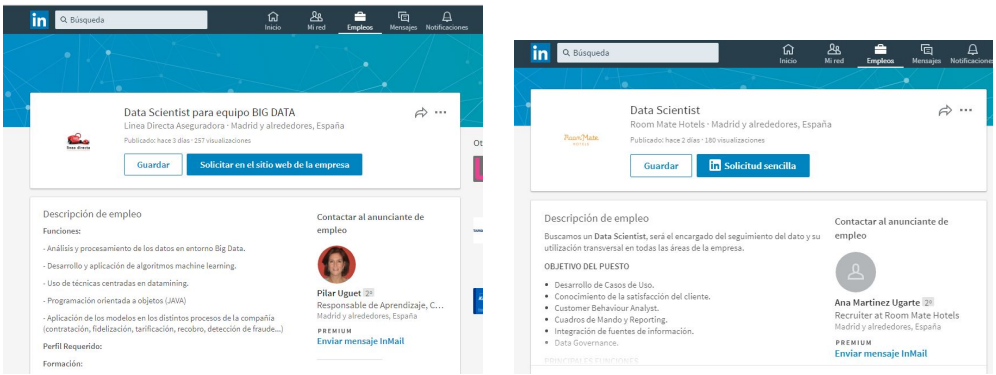


Figura 1.1.2: Dos ofertas de empleo.

Figura 1.1.3: Requisitos de las dos ofertas de empleo.

En ambos casos, entre los requisitos se encuentran conocimientos sobre Python, R, SQL, machine learning y data mining. Un reclutador probablemente usará esas palabras clave para buscar

los perfiles adecuados para alguno de los dos puestos, y construirá un conjunto de posibles candidatos (el primer paso en nuestra descripción del proceso de contratación). En esta fase, y gracias a Octopus Data Insights, nuestro reclutador contará con una ayuda extra. Octopus Data Insights le proporcionará una lista de usuarios de Twitter que hayan publicado contenido en el que aparezcan esas palabras clave, que complementarán el resultado que el reclutador haya obtenido por sus propios medios. La información proporcionada por Octopus Data Insights resultará relevante también más adelante en el proceso, cuando haya que tomar una decisión entre varios candidatos para determinar cuáles son los más adecuados para el puesto: usando la información de Twitter, los usuarios de la lista estarán ordenados según diversos criterios de relevancia.

El proceso para producir la información que ayudará al reclutador en el proceso es el siguiente:

1. Identificar los vocablos que determinan las habilidades que ha de poseer cualquier candidato para la oferta en cuestión y extraer de Twitter aquellos tuits con contenido relacionado con ellos.

2. Dados esos tuits, construir un conjunto de usuarios, que entendemos como posibles candidatos a la oferta.

3. Usar la información publicada por los usuarios para determinar el grado de adecuación a la oferta (serán más adecuados aquellos que hayan publicado información sobre todos los conocimientos requeridos que aquellos que solo hayan publicado sobre alguno de ellos, y más relevantes aquellos más activos, según el número de tuits publicados sobre cada área).

4. Estudiar la relación entre los usuarios de este conjunto, y determinar los más relevantes en sentido relativo (en términos de actividad en la red, cuáles son los más "retuiteados", cuáles los más seguidos, etc.).

1.2 Contexto de negocio

Lo que hay hecho y lo que no.

1.3 Objetivos

Lo que queremos conseguir, qué significa que lo hayamos conseguido.

1.4 Hipótesis y limitaciones

Aquí todo lo que asumamos al plantear el proyecto, y hasta dónde puede llegar. Límites del uso de la información de las redes sociales, límites del proceso en sí (ventana temporal, no detección de todos los candidatos, etc.).

La hipótesis fundamental que estamos haciendo al iniciar este proceso, es que la actividad en Twitter acerca de un determinado tema (por ejemplo, publicar algo relacionado con Python), supone

que el usuario en cuestión tiene conocimientos sobre dicho tema (en nuestro caso, entenderíamos que ese usuario posee conocimientos de Python). Esto es cuestionable, por supuesto, pero también ponderable si tenemos en cuenta que la actividad no sea esporádica. Si un usuario publica sobre un tema en numerosas ocasiones, la hipótesis de que ese tema no le resulta ajeno, va cobrando fuerza.

Entre las limitaciones de las que adolece el proceso definido para llevar a cabo el proyecto, se encuentran las siguientes:

1. En general, no todos los posibles candidatos tienen por qué usar Twitter, y por tanto habrá muchos que queden directamente fuera de nuestro proceso.

2. Los tuits utilizados en el proceso están sujetos a una ventana temporal. Habrá muchos candidatos, usuarios de Twitter, que no aparezcan en nuestros registros, por no presentar actividad durante ese tiempo.

3. Twitter impone limitaciones en la cantidad de información a la que deja acceder, y por ello, también es posible que los usuarios pierdad visibilidad en este proceso, porque el contenido publicado por ellos no se encuentre entre el proporcionado por la red social durante el proceso de extracción de datos.

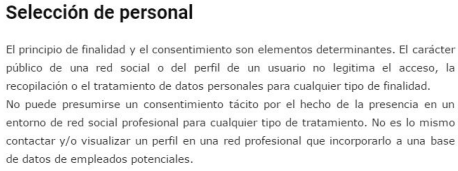
Otra limitación del proceso es que la información que obtenemos de la red es a nivel de usuario de Twitter. La dirección de correo o el nombre verdadero de la persona en cuestión, o cualquier dato que pudiera identificarla no está necesariamente disponible en la aplicación, salvo que el usuario lo haya querido hacer público explícitamente. Esta información, y la forma en que se utilice, es clave para la usabilidad del resultado del proyecto, en dos aspectos principales:

1. para que el reclutador pueda hacer uso de la información, la persona ha de estar identificada, lo suficientemente como para abrir un canal de comunicación entre el reclutador y el posible candidato.

2. desde el punto de vista de la comercialización del resultado del proyecto, el hecho de identificar usuarios en una red social y usar esa información con fines lucrativos, ha de ser implementado de forma muy cuidadosa. El impacto de la Ley Orgánica de Protección de Datos de Carácter Personal (LOPD) es muy relevante en nuestro proyecto, y merece un apartado especial. Nos ocupamos de ello en la sección 1.4.1.

En relación al primer punto, evidentemente proporcionar un usuario de Twitter ya es abrir un canal de comunicación. Sin embargo, solo la información de las publicaciones del usuario no es suficiente para incluirlo en un proceso de selección, incluso antes del primer contacto entre reclutador y candidato, y el primero probablemente necesitará más información (por ejemplo un CV) para considerar al segundo. Una forma de solventarlo sería cruzar la información de Twitter (el nombre de usuario) con la contenida en otros portales (como LinkedIn, Facebook, Academia.edu, ResearchGate, Glassdoor, etc.), ya que a menudo el usuario de Twitter es parte de los datos consignados en los CV. Esta extensión del proyecto la hemos dejado deliberadamente fuera del planteamiento de este proyecto, aunque sería *conditio sine qua non* para una implementación comercializable del proyecto.

1.4.1 La Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal



Selección de personal

El principio de finalidad y el consentimiento son elementos determinantes. El carácter público de una red social o del perfil de un usuario no legitima el acceso, la recopilación o el tratamiento de datos personales para cualquier tipo de finalidad.
No puede presumirse un consentimiento tácito por el hecho de la presencia en un entorno de red social profesional para cualquier tipo de tratamiento. No es lo mismo contactar y/o visualizar un perfil en una red profesional que incorporarlo a una base de datos de empleados potenciales.

Figura 1.4.1: Ayuda Ley de Protección de Datos https://ayudaleyprotecciondatos.es/2010/09/16/redes-sociales-empresas-y-proteccion-de-datos/

Otra forma, sería implementar un sistema que cuando un usuario de Twitter fuera a ser incluido en uno de los procesos

Capítulo 2

Planificación del proyecto

2.1 Equipo

El equipo de Octopus Data Insights está formado por cuatro personas, con perfiles multidisciplinares y complementarios:

- Fabio Inui:
- Teresa Martínez: matemática, con diez años de experiencia en investigación y docencia a nivel universitario, ocho en construcción de modelos de valoración de derivados en empresa financiera de primer nivel, y cuatro de gestión de fondos en una de las principales gestoras españolas.
- Javier Quintana:
- Silvia Santos:

2.2 Desarrollo temporal

Capítulo 3

Infraestructura

En esta sección describiremos la infraestructura que hemos construido para el desarrollo del proyecto.

3.1 Repositorio GIT

El código del proyecto, así como las presentaciones y memoria de este proyecto, está almacenado en el repositorio `https://github.com/MaiteMartinez/MBITProject_Data4all`

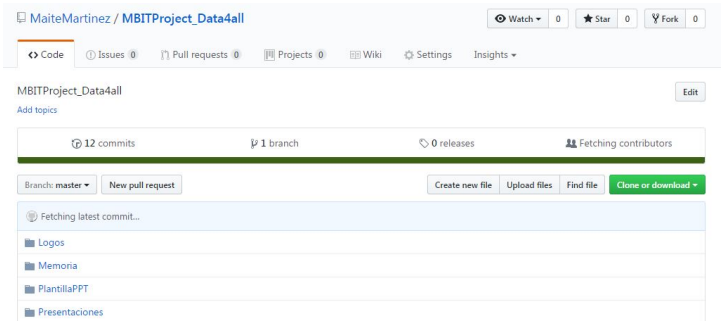


Figura 3.1.1: Repositorio del código del proyecto.

3.2 Infraestructura para la obtención de datos

Como muchas redes sociales, Twitter ofrece acceso a la información que generan sus usuarios a través de un API (*Application Programming Interface*)[7]. El API de Twitter ofrece diversas opciones: Webhook APIs, ADS API, REST APIs y Streaming APIs. La primera está enfocada a generar notificaciones instantáneas a partir de detección de eventos y la segunda a la integración de aplicaciones con la plataforma de publicidad de Twitter. Para nuestro proyecto, solo son relevantes por tanto las dos segundas:

- El API Rest (*Representational State Transfer*) permite realizar consultas puntuales con los parámetros de búsqueda indicados, a través de una componente denominada Search API. El Search API funciona de manera similar, aunque no exactamente igual, a la búsqueda en

la página web de Twitter. El Search API realiza la búsqueda entre una muestra de tuits publicados en los últimos siete días, y las búsquedas están limitadas a 180 peticiones cada ventana temporal de 15 minutos.

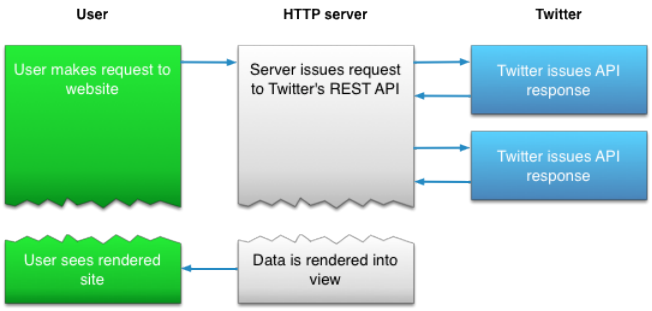


Figura 3.2.1: Funcionamiento del API Rest. https://dev.twitter.com/streaming/overview

La búsqueda realizada por este API está centrada en la relevancia y no en la completitud, lo que quiere decir que algunos tuits y usuarios podrían quedarse fuera.

- El API Streaming permite un acceso con baja latencia al flujo global de tuits de la aplicación, y requiere una conexión HTTP continua. Entre los tipos de flujos disponibles, en la web de Twitter para desarrolladores, se recomienda usar los flujos públicos para realizar minería de datos (en dichos flujos aparecen muestras de los datos públicos de Twitter).

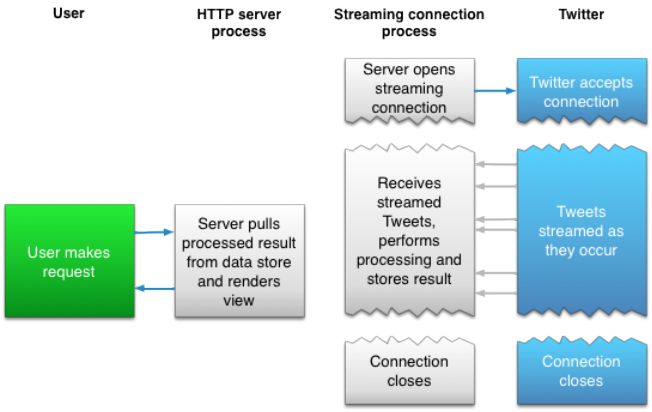


Figura 3.2.2: Funcionamiento del API Streaming. https://dev.twitter.com/streaming/overview

El acceso a ambas versiones de API está gobernado por la autentificación mediante el protocolo OAuth (*Open Authorization*), lo que implica que para cada aplicación deben obtenerse los tokens necesarios de la sección de desarrolladores de Twitter, estableciéndose un número máximo de 7 por usuario. También para ambas versiones del API existen restricciones de acceso. Estas restricciones solo afectan a las versiones gratuitas de los APIs. Hay una versión de pago de este acceso (Twitter

Firehose) que garantiza como respuesta el 100% de los tuits que cumplan los criterios de la búsqueda. Para el desarrollo de este proyecto nos hemos servido del API gratuito, por limitación de costes.

Entre los API REST y Streaming, hemos decidido utilizar el API REST con procesos planificados de actualización frecuente como suficiente aproximación al tiempo real sin necesidad de usar un API Streaming que requeriría una arquitectura más compleja y robusta.

3.3 Desarrollo en la nube

Capítulo 4

Tratamiento inicial de los datos

4.1 Descripción de los datos

Estructura general de los tuits y campos necesarios para nuestro objetivo.

4.2 Obtención de los datos

Cómo hemos hecho para bajarlos, parámetros de la búsqueda de twits.

4.3 Almacenamiento

4.4 Limpieza de los datos

Capítulo 5

Preparación de los datos

5.1 Extracción de la información relevante

5.2 Almacenamiento

Capítulo 6

Modelado de los datos

6.1 Principales hipótesis

6.2 Algoritmos

6.3 Almacenamiento

Capítulo 7

Visualización de los resultados

7.1 Herramientas

7.2 Acceso web

Capítulo 8

Áreas de mejora

Bibliografía

[1] María Gloria Castaño collado, Gerardo de la Merced López Montalvo, José María Prieto Zamora, *Guía técnica y de buenas prácticas en reclutamiento y selección de personal (R& S)*. Documento aprobado por la Junta de Gobierno del Colegio Oficial de Psicólogos de Madrid, Febrero de 2011. http://www.copmadrid.org/webcopm/recursos/guiatecnicabuenaspracticas.pdf

[2] *Selección de personal para no especialistas*. Andalucía Emprende, Fundación Pública Andaluza. Consejería de Economía y Conocimiento. https://www.andaluciaemprende.es/wp-content/uploads/2015/02/guia\discretionary{-}{}{}seleccion-personal.pdf

[3] *Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal*. Jefatura del Estado BOE núm. 298, de 14 de diciembre de 1999 Referencia: BOE-A-1999-23750 http://www.agpd.es/portalwebAGPD/canaldocumentacion/legislacion/estatal/common/pdfs/2014/Ley_Organica_15-1999_de_13_de_diciembre_de_Proteccion_de_Datos_Consolidado.pdf

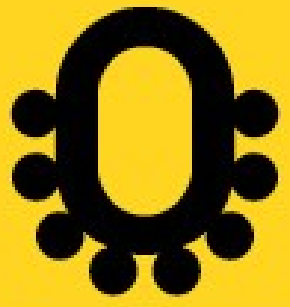
[4] María Luz Congosto Martínez, *Caracterización de usuarios y propagación de mensajes en Twitter en el entorno de temas sociales*. Tesis doctoral.

[5] "Twitter". Wikipedia. https://es.wikipedia.org/wiki/Twitter.

[6] Kumar, Shamanth; Morstatter, Fred y Liu, Huan. *Twitter Data Analytics*. Springer (2013).

[7] Twitter Developer Dpcumentationhttps://dev.twitter.com/

Documento producido con LaTeX.



OCTOPUS

DATA INSIGHTS