



ANÁLISIS DE REDES SOCIALES PARA RECURSOS HUMANOS

Fabio Inui

Teresa Martínez

Javier Quintana

Silvia Santos

Versión preliminar September 22, 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	3
Executive summary	5
1 Planteamiento del proyecto	7
1.1 Descripción	7
1.2 Contexto de negocio	10
1.3 Objetivos	10
1.4 Hipótesis y limitaciones	10
1.4.1 La Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal	10
2 Planificación del proyecto	11
2.1 Equipo	11
2.2 Desarrollo temporal	11
3 Infraestructura	13
3.1 Repositorio GIT	13
3.2 Infraestructura para la obtención de datos	13
3.3 Desarrollo en la nube	13
4 Tratamiento inicial de los datos	15
4.1 Descripción de los datos	15
4.2 Obtención de los datos	15
4.3 Almacenamiento	15
4.4 Limpieza de los datos	15
5 Preparación de los datos	17
5.1 Extracción de la información relevante	17
5.2 Almacenamiento	17

6	Modelado de los datos	19
6.1	Principales hipótesis	19
6.2	Algoritmos	19
6.3	Almacenamiento	19
7	Visualización de los resultados	21
7.1	Herramientas	21
7.2	Acceso web	21
8	Áreas de mejora	23
	Bibliografía	25

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 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 currícula 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 de las redes más dinámicas, con millones de usuarios activos en todo el mundo, fácil de usar, rápida y divertida.

USO DE TWITTER / DATOS DE LA EMPRESA



Cifras aproximadas vigentes al 30 de Junio de 2016.

Figura 1.1.1: Twitter: uso 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 twitean 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 extraer de Twitter aquellos tuits cuyo contenido sea

Es decir, tras el proceso habitual de:

1. Apertura de la vacante 2. Recepción de CVs 3. Descarte y Selección previa de CV

.... aquí en este punto de la fase es donde creo que puede aportar información adicional que le sirva para hacer un pre-ranking de candidatos. . .

4. Entrevistas 5. “Short-list” y valoración final

-. . . y creo que en el punto final también puede ser un atributo determinante si en el proceso dudan entre dos candidatos finales, por ejemplo.

Y para tratar de plasmar de alguna forma a que me estoy refiriendo y como podemos enfocarlo, trato de explicarlo con un ejemplo.

Imaginemos que buscamos el mejor “data scientist” en Madrid y que ya tienen dos candidatos que les encajan y que han estado bien en las entrevistas. ¿cómo decidir?... pues con información complementaria de como están valorados en su propio sector.

Pongamos por caso que nuestros dos candidatos son:

Nota. Un problema a resolver es que no todo el mundo pone su perfil de twitter en la información de linkedin que es la fuente principal de información en un proceso de selección y de alguna forma hay que poder cruzar esos datos. En este ejemplo he tomado dos casos que si incluyen dicha información.

El primer paso es disponer de sus perfiles en twitter:

Y a partir de aquí hacer valoraciones que nos permitan crear una especie de ranking:

- Trabajan en lo que dicen?
- Siguen a perfiles del sector?
- Son seguidos por perfiles del sector?
- Hacen comentarios profesionales o mas bien personales?
- Cómo de valorados están estos comentarios?
- . . .

Por ejemplo, y sin ser una comparativa ni exhaustiva ni exacta:

Sixto Villalba Nuestro primer candidato parece tener un perfil más amplio y heterogéneo, y además, por el enlace que tiene a su “wordpress” podemos comprobar que está más centrado en “SEO/SEM y Marketing Digital” que en “Big Data” (hay forma de poder incluir esto en el modelo?...habrá que verlo) <https://sixtosevb.wordpress.com>

El tipo de tuits no se limitan solo al trabajo y sus seguidores también son heterogéneos..., aunque su actividad es más intensa..

Diego J. Bodas Perfil más reducido y menos intenso que el otro candidato, pero mucho más centrado en el “Big Data” (al menos en un primer vistazo jj)..

No se si con esto me explico...o aún os lio más...jejejejeje

Y también por lo que he podido ir leyendo sobre Twitter, cosas que en principio creo que se pueden obtener:

- Obtener tuits generales, trending topics, tuits que contienen un hashtag o una palabra en concreto
- Una vez obtenido se puede “limpiar” para quedarnos con lo que nos interese. Una de las cuestiones interesantes es por ejemplo la “geolocalización” de los tweets (no se de momento para que pero puede ser útil...)
- What entities are in a user’s tweets? What are the most frequently occurring entities that appear in a user’s tweets? Who does a given user retweet the most often? How many of a user’s tweets contain hashtags? How many of a user’s tweets get retweeted? How many of a user’s tweets contain at least one entity?
- Here, I will show you two cases:
- 1) What are the hashtags utilized by different users?
- 2) Who retweets whom?
- Quien retuitea y sigue a quien y completarlo con
- Analisis de sentimiento (en este enlace se explica...)

<http://www.slideshare.net/mcjenkins/how-sentiment-analysis-works>

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.).

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

Cómo nos afecta. Malamente.

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

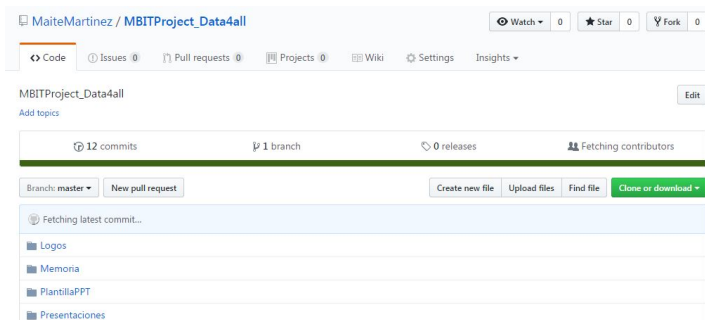


Figura 3.1.1: Repositorio del código del proyecto.

3.2 Infraestructura para la obtención de datos

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

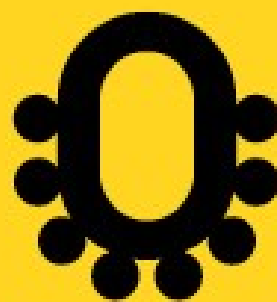
7.2 Acceso web

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/guiatecnicabuenaspracticass.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>.



OCTOPUS

DATA INSIGHTS