

Asignatura: Tipología y ciclo de vida de los datos

Ejercicio: Práctica web scraping

Alumno: Alejandro Belchí Hernández

Práctica

1. Título del dataset:

Repercusion-del-campeonato-nacional-de-kendo-en-Twitter

2. Subtítulo del dataset:

Repercusión en Twitter las semanas previas al campeonato nacional de kendo y después de la celebración del mismo.

4. Contexto:

El conjunto de datos extraídos de Twitter, mediante el programa R, corresponde a todos los tuits relacionados con el 65º campeonato nacional de Kendo, en Japón, que tuvieran incluidos los hashtags #kendo, #AJKC, #ajkendof, #martialarts, #japan. La intención es analizar los comentarios previos a dicho campeonato y los comentarios después del mismo, ya que no se ha limitado la fecha de finalización de extracción de los datos.

5. Contenido:

El periodo de tiempo seleccionado es entre el 1 de octubre de 2017, a falta de un mes del campeonato hasta la fecha actual de la extracción, ya que se quiere ver los tweets anteriores y posteriores a la celebración del mismo.

- campos

<i>text</i>	<i>favorited</i>	<i>favoriteCount</i>
<i>replyToSN</i>	<i>created</i>	<i>truncated</i>
<i>replyToSID</i>	<i>id</i>	<i>replyToUID</i>
<i>statusSource</i>	<i>screenName</i>	<i>retweetCount</i>
<i>isRetweet</i>	<i>retweeted</i>	<i>longitude</i>
<i>latitude</i>		

6. Agradecimientos:

Los datos han sido extraídos de Twitter desde R. Los datos seleccionados están bajo la licencia de Twitter.

7. Investigación:

Entiendo que el conjunto de datos no puede ser muy interesante al resto de la comunidad, se trata de un deporte minoritario, pero por el cual tengo interés, ya que lo práctico. La intención de los datos es observar el impacto en el resto de países, no en Japón porque el hashtag se debería de haber puesto en japonés (alfabeto kanji). Revisar donde, geográficamente, se han producido el mayor tráfico de mensajes y con ellos compararlo con la repercusión que tienen esos países dentro de este deporte. Durante el 2018 se realizará el

Asignatura: Tipología y ciclo de vida de los datos

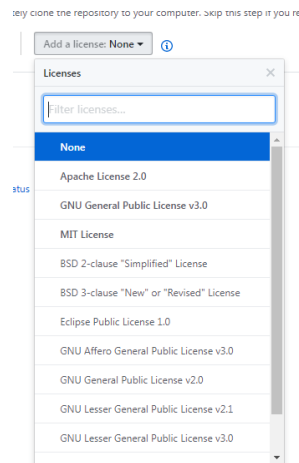
Ejercicio: Práctica web scraping

Alumno: Alejandro Belchí Hernández

campeonato del mundo y tal vez se pueda realizar una campaña, en los países con menos repercusión, para conocer más este deporte.

8. Licencia:

Salvo error por mi parte, a la hora de crear el dataset no me han aparecido las opciones de licencias que se indican en la práctica.



Free as in Freedom Me he decantado por la licencia GNU General Public License v3. Leyendo en diferentes foros y Wikipedia, me he decidido por ella al ser la más popular dentro de las licencias de software libre y por la protección que realiza sobre los contenidos.

La principal característica es esta, identificar que se trata de un software libre, con las opciones que esto te permite realizar como copiar, modificar y compartir, y lo más importante a mi parecer, que la distribución se ha de realizar bajo la misma licencia original, evitando que se realice un uso diferente al original.

Solo pondría una pega y es la posibilidad de poder solicitar un pago por el uso de los proyectos publicados con esta opción, pero queda a disposición del autor el solicitar o no pago alguno.