

2019/02

Lenguajes y Paradigmas de Programación

Proyecto

En esta clase continuaremos el proyecto a realizar en el semestre

Proyecto

Realizar un web scraper que permita:

- Extraer noticias de sitios nacionales (emol, latercera, ...)
- Procesar la información de las noticias
- Comparar con los trends de Twitter

Proyecto

- Comparar con los trends de Twitter, por ejemplo (responda tres al menos):
 - Cuánto demora una noticia en aparecer en el sitio versus el trend
 - Cuantos trends ya tienen noticias publicadas
 - Cuántas noticias no son trends
 - Qué porcentaje de trends son noticias
 - Dado un conjunto de keywords, establecer si aparece en las noticias o en un trends o en ambos
 - Dado dos palabras (por ejemplo “Chile” y “Garin”), establecer un ranking de importancia (usando trends o keywords de noticias o ambas)

Proyecto

- Comparar con los trends de Twitter
 - Se requiere una cuenta de Twitter
 - Tener acceso a la API de Twitter
 - <https://www.youtube.com/watch?v=XQ6X5DYcdzA>
 - Consumir el servicio trends/place

Uso de API de Twitter

```
from TwitterAPI import TwitterAPI
from twitter import *

import sys
sys.path.append(".")

twitter = Twitter(auth = OAuth("access_key",
                                "access_secret",
                                "consumer_key",
                                "consumer_secret"))
```

Uso de API de Twitter

```
results = twitter.trends.place(_id = 349859)
```

```
print("Chile Trends")
```

```
for location in results:  
    for trend in location["trends"]:  
        print(" - %s" % trend["name"])
```

Idea de estructura

- Scraping: paradigma procedural
- Estructura de noticias: orientado a objetos
- Procesamiento de datos: funcional

En esta clase

Extraigan los trends desde Twitter

Extraigan los keywords de las noticias de los sitios web

Realizar la comparación

¿Preguntas?