

Práctica 8 de Sistema de Información para la web

Ángel García Menéndez

November 11, 2019

0.1 Ejercicio 1

Se ha escogido la siguiente [noticia](#) del medio británico *BBC* . Tras pasarla por Open Calais y añadir las entidades y tipos correspondientes el resultado ha sido el siguiente:

- THOMSON REUTERS CORPORATION
-
- Europe
- Russia
- United States
- Ukraine
- gas
- impeachment
- Central Intelligence Agency
- House Intelligence Committee
- Judiciary Committee
- National Security Council
- Senate
- US House of Representatives
- White House
- Bill Taylor
- Charles Cooper
- Donald John Trump
- Joe Biden
- John Bolton
- Nancy Pelosi
- Rudy Giuliani
- Tim Morrison
- Volodymyr Zelensky
- presidential election

- adviser
- conservative author
- diplomat
- lawyer
- National Security Adviser
- officer
- official
- President
- Press Secretary
- Speaker
- White House Press Secretary
- Thursday
- Republican
- Democrat
- articles of impeachment
- 25 of July
- whistleblower
- Wednesday
- 7 of November

De forma ideal, todos los elementos debería tener un *itemid*, precisamente para poder explotar el potencial de enlazamiento de datos. Por ejemplo, si se puede establecer un identificados común para *Donald Trump* es posible encontrar conceptos o información relacionada con el presidente norteamericano. De lo contrario se estaría dependiendo de comparación de cadenas de caracteres, que al margen de cuestiones de optimización y escalabilidad, pueden varias entre idiomas. Un claro ejemplo de esto último puede verse en la empresa *Burisma Holdings Ltd.*, cuyo nombre en alfabeto latino sería *Burisma Holdings Ltd.* De no dársele un identificador único, ambos conceptos no tendrían por qué estar relacionados, perdiéndose información por el camino.

Empero, esto produce el siguiente dilema, ¿cuál debería ser el identificador? Si bien en lo que atañe a libros o empresas existen identificadores únicos, al tratarse de personas u otras entidades no necesariamente tienen por qué existir. En este caso, se ha de recurrir a soluciones alternativas, como pueden ser un

recurso que los identifique (la página web personal) o entradas en clasificaciones de amplio uso (como Wikidata). Aunque todas estas alternativas pueden estar sujetas a debate, el empleo de los identificadores dados por plataformas como Wikidata resulta práctico por varios motivos:

- Uso extendido de dicho modelo.
- Los identificadores han sido diseñados para este propósito.

Por dichas razones se procederá al empleo de los mismos.

La cuestión de establecer una norma general puede resultar contenciosa. Ante la falta de una estandarización propiamente dicha, siempre queda la opción de recurrir a aquello que más se utilice, por mera y simple conveniencia. A pesar de todo, y pudiendo alargarse el debate hasta la infinitud, lo mínimo que sí se puede elegir es consistencia: una vez escogido un sistema de referencia, ceñirse al mismo, no acabar generando un *popurrí* de nula utilidad.