



UNIVERSIDAD DE GRANADA

RECUPERACIÓN DE INFORMACIÓN

# Práctica 1: Búsqueda de información en la web

---

*Miguel Ángel Torres López y Luis Balderas Ruiz*

September 18, 2018

# Contents

<b>1</b>	<b>Formatos de codificación.</b>	<b>2</b>
<b>2</b>	<b>SEO.</b>	<b>2</b>
<b>3</b>	<b>Detección de plagio.</b>	<b>2</b>
3.1	Definición. . . . .	2
3.2	Casos de plagio en la actualidad. . . . .	2
3.3	Métodos para detectar plagio. . . . .	2
3.4	Temporización de búsqueda. . . . .	4

# **1 Formatos de codificación.**

# **2 SEO.**

# **3 Detección de plagio.**

## **3.1 Definición.**

Según *plagiarism.org*[3], se considera plagio la acción de presentar o usar trabajo de otro autor sin citarlo y sin especificar la fuente de procedencia. Según la RAE[4], plagiar se define como la acción de copiar obras ajenas y darlas a conocer como propias.

En el área que nos compete, hablaremos de plagio en texto escrito, aunque también se considera plagio el uso de cualquier producción de un autor sin su debida mención. Cabe mencionar el caso de los archivos multimedia. Publicar imágenes, vídeos o fragmentos y composiciones de los mismos de otro autor sin su correspondiente fuente de procedencia también es motivo de conflicto.

## **3.2 Casos de plagio en la actualidad.**

En 2015 se abrió una página de Facebook llamada *Cabronazi* que empezó a publicar fotografías con mensajes graciosos. Tres años después cuenta con más de 12 millones de seguidores y factura cerca de 370000 euros al año. No obstante, detrás de este éxito muchos usuarios se quejan de que la mayoría de las publicaciones son plagio de otras en las redes sociales. Puede verse la discusión en el periódico *El Confidencial*[1] .

Un tema que genera más controversia en la actualidad es el de Pedro Sánchez, al que acusan de haber plagiado ciertas partes de su tesis doctoral. Tras haber publicado en internet dicha tesis, distintos medios han examinado el documento. Estos son los resultados según uno de esos medios[2].

## **3.3 Métodos para detectar plagio.**

Existen numerosos métodos para detectar plagio. Los más frecuentes por su rapidez son los software de detección de plagio, aunque para usarlos se

necesita los documentos en formato digital.

- **Cadenas de texto.** La mayoría del software comercial se basa en la comparación de cadenas de texto. Usan una base de datos de documentos para enfrentar el documento sospechoso. Esto plantea un inconveniente, puede que la base de datos usada no sea suficientemente extensa o no contenga el documento plagiado. Se podría intentar añadir la mayor cantidad de textos posibles para mejorar el contraste, pero esto supondría una penalización en tiempo de cómputo.
- **Bolsas de palabras.** Una forma de reducir el tiempo sería usar comparación de bolsas de palabras. Una bolsa de palabras es un vector que representa las palabras de un texto. Por tanto, al usar bolsas de palabras estamos comprobando si dos textos usan el mismo vocabulario, obviando el orden en el que las palabras aparecen. Esto produce una reducción de la eficacia del método.

Notar que sendos métodos pueden ser mejorados con el uso de diccionarios de sinónimos y traductores para evitar la reformulación de oraciones y las traducciones.

- **Analizador de estilo.** Existen otras técnicas que proveen un análisis más profundo del texto, por ejemplo los analizadores de estilo. Este tipo de detectores analizan distintos aspectos en el discurso de un autor para hacer un perfil de escritura. Este perfil se puede caracterizar por la longitud de las oraciones, el uso de muletillas o de reformuladores del discurso. Cuando se introduce un nuevo documento en el sistema, se realiza un análisis para encontrar estilos similares. Hay que notar que un analizador de estilo es sensible a idiomas, por tanto, las traducciones no literales de textos no serían detectadas por este método.

### 3.4 Temporización de búsqueda.

HORA	BUSCADOR	PALABRAS DE BÚSQUEDA
18:30	Google	Detección de plagio
2 minutos	<a href="https://www.whatsnew.com/2017/10/07/4-sitios-web-para-detectar-plagios/">https://www.whatsnew.com/2017/10/07/4-sitios-web-para-detectar-plagios/</a>	
18:32	Google	Plagiarism detection
11 minutos	<a href="https://en.wikipedia.org/wiki/Plagiarism\_%detection">https://en.wikipedia.org/wiki/Plagiarism\_%detection</a>	
18:43	Google	Plagiarism
8 minutos	<a href="https://www.plagiarism.org">https://www.plagiarism.org</a>	
18:51	Google	Bag of words
4 minutos	<a href="https://medium.com/greyatom/an-introduction-to-bag-of-words-in-nlp-ac967d43b428">https://medium.com/greyatom/an-introduction-to-bag-of-words-in-nlp-ac967d43b428</a>	
18:55	Google	Estilometria
2 minutos	<a href="https://www.estilometria.com/">https://www.estilometria.com/</a>	
18:57	Google	Rae
1 minuto	<a href="http://dle.rae.es/?id=TIZy4Xb">http://dle.rae.es/?id=TIZy4Xb</a>	
18:58	Google	El cabronazi plagio
6 minutos	<a href="https://www.elconfidencial.com/tecnologia/2018-08-09/cabronazi-cabroworld-carlos-soria-bernardo-memes\_%d1602201">https://www.elconfidencial.com/tecnologia/2018-08-09/cabronazi-cabroworld-carlos-soria-bernardo-memes\_%d1602201</a>	
19:04	Google	Plagio Sanchez
4 minutos	<a href="https://elpais.com/politica/2018/09/14/actualidad/1536938921\_%d232616.html">https://elpais.com/politica/2018/09/14/actualidad/1536938921\_%d232616.html</a>	

## References

- [1] Artículo sobre el plagio de cabronazi. 2018. ([https://www.elconfidencial.com/tecnologia/2018-08-09/cabronazi-cabroworld-carlos-soria-bernardo-memes\\\_1602201/](https://www.elconfidencial.com/tecnologia/2018-08-09/cabronazi-cabroworld-carlos-soria-bernardo-memes\_1602201/)).
- [2] Artículo sobre la tesis de pedro sánchez. 2018. ([https://elpais.com/politica/2018/09/14/actualidad/1536938921\\\_232616.html](https://elpais.com/politica/2018/09/14/actualidad/1536938921\_232616.html)).
- [3] Página web de plagiarism. 2018. (<https://www.plagiarism.org>).
- [4] Sitio web de la real academia espaola. 2018. (<http://dle.rae.es/?id=TIzy4Xb>).