

Información extraída de:

**::Recuperación y organización de la información. Plagio. Algoritmos
para detectarlo::**

<http://www.geocities.com/algoritmosdeplagio>

© 2009 Vanesa A.

PLAGIO. ALGORITMOS DE PLAGIO

Plagio. Algoritmos de Plagio. Introducción

La aparición de las Tecnologías de la Información y de la Comunicación (TIC) ha aportado muchas ventajas, pero sin duda alguna también han sido muchos los inconvenientes que han generado, siendo uno de los más relevantes, la existencia de una mayor facilidad a la hora de plagiar documentos.

El plagio electrónico

Lo primero que se hará en estas líneas, para evitar posibles confusiones, es definir el concepto de plagio y detección de plagio.

- Plagio: Tomar ideas propias de otros y hacerlas pasar como propias.
- Detección de Plagio: Proceso mediante el cual se detecta que un documento ha sido plagiado.

Si extrapolamos estas definiciones al mundo electrónico, estamos por tanto tratando con la copia de documentos en cualquier formato electrónico (.doc, .pdf, .ppt), la copia del contenido de páginas Web y los procesos existentes en el ámbito de la informática que permiten detectarlo.

Distintas formas de plagio

Existen diversas formas de realizar plagios. A continuación se mencionan y explican brevemente las más importantes:

- Corta – Pega: Se realiza una copia textual de un documento y se incluye otro como si fuera propio.
- Modificaciones puntuales: Se modifica algunas partes del texto para intentar que el plagio pase desapercibido.
- Mosaico: El texto plagiado se encuentra desordenado a lo largo de todo el documento.
- Múltiples fuentes: Se combinan varios fragmentos plagiados de procedencia distinta.

También existen otras formas de plagio que se encuentran en el limbo de lo considerado como un documento elaborado correctamente y un documento plagiado. La más común es el parafraseo, que consiste en la interpretación de un documento original.

DetECCIÓN DEL PLAGIO

Actualmente se puede encontrar varios sistemas, herramientas y portales (tanto gratuitos como de pago) que se encargan de analizar documentos para detectar si el contenido de este ha sido plagiado.

Algunas de las aplicaciones más conocidas son:

- Copyscape
- CopyTracker
- Ephorus
- iThenticate
- Jplag
- MOSS
- MyDropBox
- Plagiarismdetect
- Plagiarism-detector
- PlagiarismscannerSafeAssign
- Plagium
- Scanmyessay
- SeeSources
- Turnitin
- Urkund

ALGORITMOS PARA DETECTAR EL PLAGIO

Si por algo se caracterizan las aplicaciones que se acaban de mencionar es por la utilización de una serie de algoritmos matemáticos que permiten detectar el plagio.

Tras realizar un proceso de investigación acerca de esta materia, se puede observar que existe una clara reticencia a mostrar estos algoritmos de plagio, y es que no hay que olvidar, que como le ocurre a los algoritmos de posicionamiento que utiliza Google, una vez descubierto, pierden todo su valor, ya que los usuarios no tardan en descubrirlos.

Durante el estudio mencionado anteriormente se ha podido obtener dos algoritmos para detectar el plagio, que están recogidos en el artículo "Similaridad y Plagio en la Web" de Francisco Claude Faust (ver referencias) y que han obtenido buenos resultados tras su aplicación en un experimento realizado por dicho autor.

Estos algoritmos de plagio están basados en la subsecuencia común más larga de dos textos y cuenta con dos aproximaciones:

La primera aproximación de este algoritmo es:

$$s = \frac{100 \times n_1}{|\mathbf{lcs}(T_1, T_2)|}$$

Y la segunda, que es una mejora del anterior es:

$$s = 100 \times \left(1 - \frac{\text{mín}(\text{ed}(T_1, T), \text{ed}(T_2, T))}{|T|} \right)$$

Donde:

- T1 es el texto 1
- T2 es el texto 2
- n1 es la longitud del texto 1
- n2 es la longitud del texto 2

Aunque no se haya podido visualizar de manera explícita los algoritmos de plagio que se emplean para la detección del plagio electrónico si que se ha podido llegar a la conclusión de que los factores principales que priman en ellos son:

- Frecuencia de aparición de términos idénticos en dos documentos.
- Posición que ocupan dichos términos dentro de los documentos.

A pesar de que la efectividad de la aplicación de estos algoritmos de plagio, como los empleados en otro tipo de recursos, es bastante certera, hay que tener en cuenta que la acusación de plagio es un hecho grave, por lo que siempre se debe de contar con una supervisión humana, antes de tomar medidas legales.

Utilidades de la detección del plagio electrónico

Una de las principales utilidades de la detección del plagio, en el entorno en el que nos estamos moviendo (recuperación y organización de la información), es sin duda su aplicación en los algoritmos de posicionamiento de los diversos motores de búsqueda.

Y es que como bien es sabido, por todos los webmaster que ambicionan con estar entre los primeros puestos en las páginas de resultados, los motores de búsqueda penalizan a sus sitios cuando detectan que se ha producido plagio.

Referencias Bibliográficas

Los documentos que se han utilizado para la elaboración de esta página web dedicada al plagio y a los algoritmos de plagio que existen para poder detectarlo son:

- CLAUDE FAUST, Francisco. Similaridad y detección del plagio en la Web [en línea]. Disponible en: <http://www.dcc.uchile.cl/~fclaude/docs/CopyDetect.pdf>
- Colaboradores de Wikipedia. Plagiarism detection [en línea]. Wikipedia, La enciclopedia libre, 2008 [fecha de consulta: 20 de noviembre del 2009]. Disponible en: http://en.wikipedia.org/wiki/Plagiarism_detection
- PARKER, Alan; HAMBLEEN, James. “Computer Algorithms for Plagiarism Detection”. Ieee Transactions on Education. Mayo 1989. vol. 32, núm. 2, p. 94 - 99
- SANABRIA FERNÁNDEZ, Javier. “Plagio e Internet” Director: Julio César Hernández Castro. Universidad Carlos III, 2007