

Cálculo de la Reputación de Páginas Web

Emilio Bottino, Andrés Velásquez, Mariano Zibecchi
Estudiantes de la Maestría en Ingeniería de Software,
Universidad Nacional de la Plata
{emilio.bottino, andres.velazquez, mariano.zibecchi}@nuevobancosuquia.com.ar

Abstract

Una de las herramientas más utilizadas por los motores de búsqueda en la web para ordenar las páginas que se entregan como respuesta a una consulta, es el Análisis de Hipervínculos. Entre los métodos más conocidos para implementar dicho análisis encontramos a PageRank y HITS. Estos algoritmos permiten determinar la importancia de una página, definiéndola en función de la cantidad de páginas que tienen hipervínculos a ella. Mendelzon y Raffiei proponen un método, Medida de la Reputación (RM), que permite dada una página p determinar en que tópicos es esta página importante.

Mendelzon y Raffiei implementan un prototipo, TOPIC, que permite estimar, para una página dada, los tópicos en los cuales esta página tiene una alta reputación. En TOPIC, la determinación del conjunto inicial de páginas que apuntan a p es crucial para el cálculo de RM. De nuestras observaciones surge que S varía fuertemente según el motor. Proponemos entonces una mejora en el método de selección de páginas iniciales, direccionando dicha búsqueda inicial hacia algún lugar de donde podamos recuperar páginas rankeadas objetivamente.

Además, implementamos nuestro propio prototipo para analizar la correspondencia entre el ranking entregado por un motor de búsqueda, y el que surge de aplicar Penetración y Foco. Desarrollamos un prototipo que, si bien no implementa el algoritmo completo de Mendelzon y Raffiei, nos permite calcular dichos conceptos y lo utilizamos para realizar comparaciones contra los resultados entregados por un motor comercial de búsqueda..

Keywords: Bases de Datos; Análisis de hipervínculos; Búsqueda en la web; Reputación de Páginas.