



UNIVERSIDAD NACIONAL DE INGENIERIA
SEDE REGIONAL DEL NORTE
RECINTO UNIVERSITARIO AUGUSTO C. SANDINO
UNI-RUACS

INGENIERIA AGROINDUSTRIAL

ELABORADO POR:
Katherine Daniela cruz m.

HECHO EL 14 DE MARZO DEL 2012

Sistemas de buscadores avanzados de google.

Google es un poderoso sistema de búsqueda en Internet que permite un acceso muy fácil y rápido a cualquier tipo de información que haya disponible en la Gran Red, desde cualquier lugar del mundo.

Como curiosidad podemos decir que el nombre se creó permutando las letras de la palabra GOOGOL que fue inventada por Milton Sirotta cuando su tío, el matemático americano Edward Kasner, le pidió que "bautizase" a un número muy grande (10 elevado a 100: un 1 seguido de 100 ceros). La empresa tuvo sus orígenes en un garaje de California a partir de un proyecto universitario y contaba con cuatro empleados (1998). En la actualidad recibe más de 200 millones de consultas diarias y tiene más de 2000 trabajadores. Es posible efectuar búsquedas en más de 35 millones de documentos HTML.

Este y otros buscadores de la red son tan potentes que comportan ciertos problemas, ya que gracias a Google y a herramientas similares es casi imposible conservar intacta la privacidad en Internet. Basta, por ejemplo, con encontrar el nombre de una persona en la red y acceder a servicios como las paginas blancas para tener acceso a su teléfono y dirección.

Pero esta misma tecnología de búsquedas masivas permite cosas como por ejemplo navegar a través de un museo con la mayoría de las páginas web desde la fecha de su creación una dirección y curiosear las distintas versiones de la página desde su creación.

Google es un sitio que se ha convertido en sinónimo de búsquedas en Internet. Pero Google es mucho más que un simple sitio de búsqueda. Es una poderosa herramienta que si se utiliza adecuadamente puede ayudarnos mucho y ahorrarnos un tiempo precioso en la búsqueda de información en la red.

Cómo funciona Google

El funcionamiento de Google se basa en el cruce de diversas informaciones presentes en las páginas web registradas en la gigantesca base de datos del motor de búsqueda, y en la determinación de un orden jerárquico específico de las páginas obtenidas en cada búsqueda. El criterio de búsqueda de evaluación se apoya en la estructura de enlaces de la web. Mediante este mecanismo Google consigue rastrear e indexar eficientemente la web, y producir resultados más exactos que los de otros sitios de búsqueda. Esto es posible porque además de buscar por palabras, también atribuye un ranking o una "nota" para las páginas. Para determinar la "nota" de una página Google utiliza una función llamada PageRank. Un factor importante en esta función es el número de páginas

(referencias) que apuntan hacia la página que evalúa. Si tienes curiosidad por los fundamentos matemáticos de la función PageRank.

Para presentar los resultados al usuario, los programas de búsqueda trabajan básicamente de dos formas:

- por medio de un esquema de registro de páginas o sitios en el que las páginas que desean aparecer deben registrarse.
- Y por rastreo de la web, que es llevado a cabo por programas llamados robots que solicitan páginas a los servidores web y, además de almacenarlas para usos futuros, extraen sus enlaces en ellas contenidos a otras páginas. Estos enlaces se almacenan para ser visitados con posterioridad. De esta forma, si en una página almacenada hay una referencia a una página nueva, ésta será descubierta.

Pero el proceso no acaba cuando una página ha sido "descubierta". Es necesario almacenar el contenido de la página de forma ordenada, es decir, en forma de índices. Cada índice contiene un conjunto diferente de información sobre la página almacenada y, para cada página indexada, se almacenará información como: palabras contenidas en la página, título, enlaces... Con esta información el orden jerárquico comienza a calcularse.