
1. INDICE

1. INDICE	2
2. RESUMEN	5
3. ABSTRACT	6
4. AGRADECIMIENTOS.....	7
5. INTRODUCCIÓN	8
5.1 Planteo del problema	9
5.2 Objetivos planteados.....	12
6. ANTECEDENTES Y SITUACIÓN ACTUAL	17
6.1 Introducción	17
6.1.1 Motor de Búsqueda.....	17
6.1.2 Hipertexto	18
6.2 Historia	20
6.2.1 El principio de Internet	20
6.2.2 Motores de Búsqueda – Historia, efemérides e innovaciones	23
6.3 Situación actual.....	31
6.3.1 Organización.....	31
6.3.2 Evolución de los índices de los Motores de Búsqueda.....	32
6.3.3 Estadísticas.....	33
7. IDENTIFICACIÓN Y ANÁLISIS DE LOS COMPONENTES.....	36
7.1 Indización	37
7.1.1 Introducción	37
7.1.2 Funcionamiento de los Robots.....	38
7.1.3 Métricas	45
7.1.4 Aspectos que hacen más compleja la Indización.....	60
7.1.5 Índice	72
7.1.6 Presentación de un ejemplo	74
7.1.7 Conclusiones.....	82

7.2	Recuperación	83
7.2.1	Introducción	83
7.2.2	Mecanismos de Recuperación.....	85
7.2.3	Algunas consideraciones.....	86
7.2.4	Métricas	90
7.2.5	Subjetividad	98
7.2.6	Aspectos que hacen más compleja la Recuperación.....	100
7.2.7	Presentación de un ejemplo	104
7.2.8	Conclusiones	107
8.	EVALUACIÓN	109
8.1	Introducción	109
8.2	Resultados Experimentales	117
8.2.1	Primer resultado experimental: “Prueba sobre el tamaño del Motor de Búsqueda”	119
8.2.2	Segundo resultado experimental: “Métodos para medir el rendimiento en el tiempo en los Motores de Búsqueda”	122
8.2.3	Tercer resultado experimental: “Métodos de prueba para Motores de Búsqueda de la WWW”	123
8.2.4	Cuarto resultado experimental: “Relevancia, variedad y dinamismo en los Motores de Búsqueda”	128
8.2.5	Conclusiones sobre los cuatro resultados experimentales	157
9.	PROTOTIPO	159
9.1	Introducción	159
9.2	Funcionamiento	159
9.3	Prueba Experimental	161
9.4	Resultados Numéricos	163
9.5	Conclusiones	167
10.	CONCLUSIONES	169
11.	BIBLIOGRAFÍA REFERENCIADA	173
12.	BIBLIOGRAFÍA CONSULTADA	176

13.	GLOSARIO	178
14.	ANEXO I.....	180
15.	ANEXO II.....	182
15.1.1	Prueba sobre el tamaño del Motor de Búsqueda.....	182
16.	ANEXO III.....	190
16.1.1	Métodos para medir el rendimiento en el tiempo en los Buscadores	190
17.	ANEXO IV	193
17.1.1	Métodos de prueba para Motores de Búsqueda de la WWW	193
18.	ANEXO V	207
18.1.1	Resultados experimentales.....	207
19.	ANEXO VI	208
19.1.1	Implementación Prototipo.....	208
19.1.2	Código de las rutinas.....	210

2. RESUMEN

El presente trabajo es un Análisis y Evaluación de los Motores de Búsqueda de Internet y cómo buscan, encuentran y califican la información encontrada.

Una de las causas por las cuales se abordó el problema del descubrimiento de recursos en Internet es debido al rápido crecimiento de dicha red en los últimos tiempos. Debido a esto, se vio la necesidad de tener algún tipo de orden o algún tipo de clasificación de los recursos de Internet (principalmente páginas, documentos e información). Esta función la comenzaron a cumplir los Robots, conocidos comúnmente como buscadores.

A lo largo de este documento, el lector podrá encontrar un análisis completo de los tres diferentes componentes que tiene un buscador de Internet. Los mismos son: el Robot, el cual está encargado de visitar las páginas y crear el índice; el Índice, el cual contiene la información recolectada por el Robot; el Motor de Búsqueda, el cual se encarga de buscar, clasificar y recuperar las páginas previamente indizadas.

Finalmente luego de analizar cada uno de los componentes, se expondrá un conjunto de diferentes pruebas realizadas por terceros y otras de nuestra propia autoría para evaluar el rendimiento de los diferentes sitios buscadores. Dichas pruebas, permitirán evaluar y comparar de forma general las dos tareas principales que un Motor de Búsqueda lleva a cabo (Indización y Recuperación) y cómo influyen las distintas métricas utilizadas por cada uno de ellos en el rendimiento global.

Palabras clave: Motor de Búsqueda, Internet, Robot

3. ABSTRACT

The present document is an Analysis and an Evaluation of Internet Search Engines and how they search, find and classify the information they find.

One of the reasons why we decided to tackle the problem of finding information on the Web, was due to the fast growth the Internet has had over the past few years. This is why, there was a need for having some sort of order or classification of these Internet resources (mainly pages, documents and information). Robots, normally known as searchers, started to solve this problem.

Throughout this document, the reader will find a thorough analysis of the three different components a Search Engine has. They are: a Robot, in charge of visiting the pages and creating the Index; the Index, which contains the information recovered by the Robot; the Search Engine, in charge of searching, classifying and recovering the pages previously indexed.

Finally, after analyzing each one of the components, we will present a set of tests carried out by other authors and tests created by ourselves to evaluate the performance of the different search sites. These tests, allow us to evaluate and compare in a generic way the two main tasks of a Search Engine (Index and Retrieval) and how the different metrics used by each one of them affects the overall performance.

Keywords: Search engine, Internet, Robot