

---

**1. INDICE**

---

<b>1. INDICE</b> .....	<b>2</b>
<b>2. RESUMEN</b> .....	<b>5</b>
<b>3. ABSTRACT</b> .....	<b>6</b>
<b>4. AGRADECIMIENTOS</b> .....	<b>7</b>
<b>5. INTRODUCCIÓN</b> .....	<b>8</b>
5.1 Planteo del problema .....	9
5.2 Objetivos planteados.....	12
<b>6. ANTECEDENTES Y SITUACIÓN ACTUAL</b> .....	<b>17</b>
<b>6.1 Introducción</b> .....	<b>17</b>
6.1.1 Motor de Búsqueda .....	17
6.1.2 Hipertexto .....	18
<b>6.2 Historia</b> .....	<b>20</b>
6.2.1 El principio de Internet .....	20
6.2.2 Motores de Búsqueda – Historia, efemérides e innovaciones .....	23
<b>6.3 Situación actual</b> .....	<b>31</b>
6.3.1 Organización .....	31
6.3.2 Evolución de los índices de los Motores de Búsqueda.....	32
6.3.3 Estadísticas.....	33
<b>7. IDENTIFICACIÓN Y ANÁLISIS DE LOS COMPONENTES</b> .....	<b>36</b>
<b>7.1 Indización</b> .....	<b>37</b>
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

---

<b>7.2</b>	<b>Recuperación</b> .....	<b>83</b>
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
<b>8.</b>	<b>EVALUACIÓN</b> .....	<b>109</b>
<b>8.1</b>	<b>Introducción</b> .....	<b>109</b>
<b>8.2</b>	<b>Resultados Experimentales</b> .....	<b>117</b>
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
<b>9.</b>	<b>PROTOTIPO</b> .....	<b>159</b>
<b>9.1</b>	<b>Introducción</b> .....	<b>159</b>
<b>9.2</b>	<b>Funcionamiento</b> .....	<b>159</b>
<b>9.3</b>	<b>Prueba Experimental</b> .....	<b>161</b>
<b>9.4</b>	<b>Resultados Numéricos</b> .....	<b>163</b>
<b>9.5</b>	<b>Conclusiones</b> .....	<b>167</b>
<b>10.</b>	<b>CONCLUSIONES</b> .....	<b>169</b>
<b>11.</b>	<b>BIBLIOGRAFÍA REFERENCIADA</b> .....	<b>173</b>
<b>12.</b>	<b>BIBLIOGRAFÍA CONSULTADA</b> .....	<b>176</b>

---

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