

TABLA DE CONTENIDOS

	página
Dedicatoria	I
Agradecimientos	II
Tabla de Contenidos	III
Índice de Figuras	V
Índice de Tablas	VII
Resumen	VIII
Abstract	IX
1. Introducción	1
1.1. Descripción del problema	1
1.2. Alcance del proyecto	3
1.3. Objetivos	4
1.3.1. Objetivo general	4
1.3.2. Objetivos específicos	4
2. Antecedentes	5
2.1. Recuperación de la información	5
2.1.1. Introducción	5
2.1.2. Definición	6
2.1.3. Características	6
2.1.4. El proceso de recuperación de información	6
2.1.5. Arquitectura general	8
2.1.6. Modelos de recuperación de información	10
2.1.7. Comportamiento morfológico	15
2.2. Indexación y almacenamiento de información	16
2.2.1. Tipos de métodos de indexación para Base de Datos	16
2.2.2. Gestión de memoria en JAVA	17

2.3.	Listas invertidas	18
2.3.1.	Introducción	18
2.3.2.	Indices invertidos	19
2.3.3.	Tablas de dispersión	21
3.	Metodología	23
3.1.	Introducción	23
3.1.1.	Descripción detallada de la problemática	23
3.2.	Fases de la metodolodía de trabajo	24
3.2.1.	Descripción de las fases de la metodología	24
4.	Desarrollo	26
4.0.2.	Descripción y arquitectura general de la aplicación	26
4.0.3.	Módulo de indexamiento	27
4.1.	Fase I: Configuración de un esquema de Base de Datos relacional . . .	31
4.1.1.	Definición de niveles de estructura de tablas	31
4.1.2.	Colección de prueba y aplicación de índices	33
4.1.3.	Comportamiento tamaño de tablas	34
4.1.4.	Comportamiento del vocabulario	34
4.2.	Fase II: Construcción del índice invertido	36
4.2.1.	Diseño de estructuras de datos de índices	36
4.2.2.	Comparación de estructuras de índices	39
5.	Resultados Experimentales	40
5.1.	Introducción	40
5.2.	Evaluaciones	40
5.2.1.	Evaluación de métodos de indexación para Base de Datos . . .	40
5.2.2.	Evaluación del índice en RAM	49
5.3.	Comparación entre índice RAM v/s índice Base de Datos	51
6.	Conclusiones y Trabajo Futuro	62
6.1.	Conclusiones	62
6.2.	Trabajo futuro	64
7.	Anexos	66

ÍNDICE DE FIGURAS

	página
2.1. Interacción del usuario con sistema de <i>recuperacion de información</i>	7
2.2. Vista lógica de documentos	8
2.3. El proceso de recuperación de información.	10
2.4. Modelo de ejecución de JAVA	17
2.5. Heap de JAVA	18
2.6. Índice invertido	20
2.7. Ejemplo de Tabla Hash	21
4.1. Prototipo de refinamiento de traducciones automáticas.	27
4.2. Diagrama de bloques del módulo de indexamiento.	29
4.3. Diagrama de clases del módulo de indexamiento.	30
4.4. Configuración I de implementación de Base de Datos.	31
4.5. Configuración II de implementación de Base de Datos	32
4.6. Configuración III de implementación de Base de Datos	32
4.7. Configuración IV de implementación de Base de Datos	33
4.8. Comportamiento empírico del vocabulario de palabras	35
4.9. Estructura del <i>idx-ram</i> usando tablas auxiliares.	37
4.10. Estructura de <i>idx-ram</i> sin usar tablas auxiliares.	38
5.1. Tiempo de respuesta de intersección <i>tres</i> palabras.	42
5.2. Tiempo de respuesta de intersección <i>tres</i> palabras usando Hash y B-tree	42
5.3. Tiempo de respuesta de intersección de <i>cuatro</i> palabras.	43
5.4. Tiempo de respuesta de intersección <i>cuatro</i> palabras usando Hash y B-tree	43
5.5. Tiempo respuesta por niveles para <i>cuatro</i> intersecciones.	44
5.6. Tamaño (MB) de la tabla que guarda las <i>frases</i>	46
5.7. Tamaño (mb) de la tabla que guarda las <i>sub-frases</i>	47
5.8. Tamaño (mb) de la tabla que guarda las <i>palabras</i>	48
5.9. Tamaño (mb) tabla <i>phrase</i>	50
5.10. Modelo de comparación RAM y Base de Datos (tiempo respuesta) . .	52
5.11. Niveles definidos de tamaños posting list.	53

5.12. Tiempo de respuesta de búsqueda de sub-frases por conjunto de palabras	57
5.13. Tiempo respuesta búsqueda sub-frases RAM	58
5.14. Tiempo respuesta búsqueda de <i>frases</i>	59
5.15. Tiempo respuesta búsqueda de <i>frases</i> (con más detalle)	60
7.1. Tiempo de respuesta de <i>dos</i> a <i>cinco</i> palabras por Base de Datos. . . .	66
7.2. Tiempo de respuesta de <i>seis</i> a <i>ocho</i> palabras por Base de Datos. . . .	67

ÍNDICE DE TABLAS

	página
5.1. Rangos definidos de tamaños posting list a testear	54
7.1. Posting list tamaño uno.	67