

## TABLA DE CONTENIDOS

	página
<b>Tabla de Contenidos</b>	<b>I</b>
<b>Índice de Figuras</b>	<b>III</b>
<b>Índice de Tablas</b>	<b>IV</b>
<b>Resumen</b>	<b>v</b>
<b>1. Introducción</b>	<b>6</b>
1.1. Objetivos . . . . .	8
1.1.1. Objetivo general . . . . .	8
1.1.2. Objetivos específicos . . . . .	8
1.2. Descripción de los contenidos . . . . .	8
<b>2. Antecedentes</b>	<b>10</b>
2.1. Espacios Métricos . . . . .	10
2.1.1. Espacios Vectoriales . . . . .	11
2.1.2. Tipos de distancias . . . . .	11
2.2. Consultas por proximidad . . . . .	12
2.3. Índices en espacios métricos . . . . .	13
2.4. Algoritmo basado en Permutaciones (ABP) . . . . .	13
2.4.1. Comparación entre permutaciones . . . . .	15
2.5. Método de compresión Re-Pair . . . . .	18
2.6. Heap . . . . .	19
2.7. Hash Tables . . . . .	20
2.7.1. Tablas de direccionamiento directo . . . . .	21
2.7.2. Hash Tables . . . . .	22
2.7.3. Resolución de Colisiones por Encadenamiento . . . . .	23
<b>3. Metodología</b>	<b>25</b>
3.1. Diseño . . . . .	26
3.2. Implementación . . . . .	27

3.2.1.	Lenguaje de programación C++ . . . . .	27
3.2.2.	IDE . . . . .	28
3.2.3.	Compilador . . . . .	28
3.3.	Pruebas . . . . .	28
3.3.1.	Espacios métricos de prueba . . . . .	28
3.3.2.	Estación de trabajo utilizada . . . . .	29
<b>4.</b>	<b>Compactación</b>	<b>30</b>
4.1.	Algoritmo de compactación ACT . . . . .	30
4.2.	Algoritmo de descompactación ACTD . . . . .	34
<b>5.</b>	<b>Compresión</b>	<b>38</b>
5.1.	Ejemplo práctico de compresión . . . . .	39
5.2.	Algoritmo de compresión AC . . . . .	45
5.3.	Algoritmo de descompresión ACD . . . . .	48
<b>6.</b>	<b>Compresión con compactación</b>	<b>53</b>
6.1.	Algoritmo de compresión más compactación versión 1 (ACC1): . . . . .	53
6.2.	Algoritmo de descompactación con descompresión versión 1 (ACCD1):	54
6.3.	Algoritmo de compresión más compactación versión 2 (ACC2): . . . . .	54
6.4.	Algoritmo de descompactación con descompresión versión 2 (ACCD2):	57
<b>7.</b>	<b>Experimentos y resultados</b>	<b>60</b>
7.1.	Descripción de los experimentos . . . . .	60
7.2.	Análisis del espacio utilizado . . . . .	61
7.2.1.	Análisis del espacio utilizado según cantidad de permutantes .	61
7.2.2.	Análisis del espacio utilizado según dimensionalidad . . . . .	63
7.2.3.	Espacio utilizado por el diccionario y la permutación comprimida	65
7.3.	Análisis del tiempo de creación de permutaciones . . . . .	67
7.4.	Análisis del tiempo de consultas . . . . .	68
<b>8.</b>	<b>Conclusiones y trabajo futuro</b>	<b>71</b>
8.1.	Conclusiones . . . . .	71
8.2.	Trabajo futuro . . . . .	72
	<b>Bibliografía</b>	<b>73</b>

## ÍNDICE DE FIGURAS

	página
2.1. Ejemplo de consultas en espacios métricos . . . . .	12
2.2. Fases de preprocesamiento y de consulta . . . . .	14
2.3. Ejemplo de permutación . . . . .	15
2.4. Ejemplo del ABP, en la fase de preprocesamiento y de consulta. . . .	17
2.5. Ejemplo de compresión aplicando Re-Pair . . . . .	18
2.6. Max-heap visto como un árbol binario y un arreglo . . . . .	20
2.7. Implementación de un diccionario mediante una tabla de direccionamiento . . . . .	22
2.8. Uso de una función de hash . . . . .	23
2.9. Manejo de colisiones por encadenamiento . . . . .	24
5.1. Iteración 1 para obtener los pares de símbolos de mayor frecuencia. .	41
5.2. Iteración 2 para obtener los pares de símbolos de mayor frecuencia. .	43
5.3. Iteración 3 para obtener los pares de símbolos de mayor frecuencia. .	44
7.1. Uso de memoria vs $n$ permutantes para dim medias . . . . .	61
7.2. Uso de memoria vs $n$ permutantes para dim altas . . . . .	62
7.3. Uso de memoria vs dim para permutantes entre $\{16, 32\}$ . . . . .	63
7.4. Uso de memoria vs dim para permutantes entre $\{64, 256\}$ . . . . .	64
7.5. Uso de memoria de las permutaciones y diccionario Re-Pair para permutantes entre $\{16, 32\}$ . . . . .	65
7.6. Uso de memoria de las permutaciones y diccionario Re-Pair para permutantes entre $\{64, 256\}$ . . . . .	66
7.7. Tiempo de creación de permutaciones versus tamaño de la base de datos	67
7.8. Tiempo de creación de permutaciones versus cantidad de permutantes	67
7.9. Tiempo de creación de permutaciones versus dimensionalidad . . . . .	68
7.10. Tiempo de consultas versus dimensionalidad para 16 permutantes . .	68
7.11. Tiempo de consultas versus dimensionalidad para 32 permutantes . .	69
7.12. Tiempo de consultas versus dimensionalidad para 64 permutantes . .	69
7.13. Tiempo de consultas versus dimensionalidad para 128 permutantes . .	70

## ÍNDICE DE TABLAS

	página
2.1. Ejemplo de cálculo de permutaciones (primera parte) . . . . .	17
2.2. Ejemplo de cálculo de permutaciones (segunda parte) . . . . .	18
4.1. Uso completo de los bits de la variable $a$ para 16 permutantes. . . . .	31
4.2. Uso completo de los bits de la variable $b$ para 16 permutantes . . . . .	31
4.3. Casos en donde el uso del espacio de las variables es completo. . . . .	33
4.4. Uso completo de los bits de la variable $a$ para 17 permutantes. . . . .	33
4.5. Uso completo de los bits de la variable $b$ para 17 permutantes. . . . .	33
4.6. Uso incompleto de los bits de la variable $c$ para 17 permutantes. . . . .	34
5.1. Conjunto de permutaciones para ejemplo de compresión . . . . .	39
5.2. Conjunto de permutaciones comprimidos después de la primera iteración	42
5.3. Conjunto de permutaciones comprimidos después de la segunda iteración	45
5.4. Ejemplo de pares símbolos con frecuencias . . . . .	48