

# Índice general

<b>1. Introducción</b>	<b>2</b>
1.1. Aplicaciones . . . . .	4
1.2. Aporte de la Tesis . . . . .	6
1.3. Organización de la Tesis . . . . .	7
<b>I CONCEPTOS BÁSICOS</b>	<b>10</b>
<b>2. Conceptos Básicos</b>	<b>11</b>
2.1. Consultas por Proximidad . . . . .	12
2.1.1. Consultas de los $K$ Vecinos más Cercanos . . . . .	13
2.2. Espacios Vectoriales . . . . .	14
2.2.1. Tipos de Distancias . . . . .	14
2.2.2. Índices . . . . .	15
2.3. Efectos de la Dimensión Intrínseca Alta . . . . .	16
2.4. Concepto de Índice Métrico . . . . .	17
2.5. Definiciones . . . . .	18
2.5.1. Tipos de Relaciones . . . . .	19
2.5.2. Cadenas y Árboles (Tries) . . . . .	20
2.6. Bases de Datos de Prueba . . . . .	21
2.6.1. Vectores en el Cubo Unitario . . . . .	21
2.6.2. Diccionarios . . . . .	21
2.6.3. Documentos . . . . .	22

<b>3. Estado del arte</b>	<b>25</b>
3.1. Clasificación de Algoritmos . . . . .	25
3.1.1. Algoritmos Basados en Pivotes . . . . .	25
3.1.2. Algoritmos Basados en Particiones Compactas . . . . .	28
3.2. Descripción de los Algoritmos . . . . .	30
3.3. Comparación del Estado del arte . . . . .	39
3.4. Algoritmos Inexactos . . . . .	42
<b>II NUESTRA TÉCNICA</b>	<b>48</b>
<b>4. Técnica Basada en Inversiones</b>	<b>49</b>
4.1. Definiciones . . . . .	50
4.2. Comparación de las Inversiones contra las Familias Existentes . . . . .	53
4.2.1. Inversiones frente a los Algoritmos de Pivotes . . . . .	53
4.2.2. Inversiones frente a las Particiones Compactas . . . . .	54
4.3. Uso de la Técnica Basada en Inversiones . . . . .	56
4.3.1. Indexación . . . . .	56
4.3.2. Proceso de Búsqueda . . . . .	58
4.4. Experimentación . . . . .	59
4.4.1. Diccionario . . . . .	60
4.4.2. Cubo unitario . . . . .	60
<b>5. Algoritmo Basado en Permutaciones</b>	<b>63</b>
5.1. Comparación entre Permutaciones . . . . .	64
5.2. Buscando las Permutaciones más Similares . . . . .	65
5.2.1. Empleando un Trie . . . . .	66
5.2.2. Ordenando el Conjunto . . . . .	68
5.3. Experimentación . . . . .	70
5.3.1. Documentos TREC . . . . .	72
5.3.2. Comparación entre Exactos y Probabilísticos . . . . .	73

5.4. Selección de Permutantes . . . . .	73
<b>III APLICACIONES</b>	<b>82</b>
<b>6. Nuevo Algoritmo Exacto: iAESA</b>	<b>83</b>
6.1. AESA . . . . .	83
6.1.1. Proceso de Búsqueda de AESA . . . . .	84
6.2. iAESA . . . . .	86
6.2.1. Proceso de Búsqueda de iAESA . . . . .	86
6.2.2. Comparando AESA con iAESA . . . . .	86
6.2.3. Combinando AESA e iAESA . . . . .	88
6.3. Versiones Probabilísticas . . . . .	89
6.3.1. Limitando el Trabajo . . . . .	90
6.3.2. Relajando las Condiciones de Búsqueda . . . . .	90
6.4. Resultados Experimentales . . . . .	90
6.4.1. iAESA Exacto . . . . .	91
6.4.2. Algoritmos Probabilísticos . . . . .	92
6.5. Cálculo Dinámico del Footrule . . . . .	97
6.5.1. Experimentación . . . . .	107
<b>7. Aplicaciones: Identificación de Caras</b>	<b>111</b>
7.1. Reconocimiento de Caras . . . . .	111
7.1.1. Transformación de Objetos . . . . .	112
7.1.2. Reducción Dimensional . . . . .	113
7.2. Descripción de las Bases de Datos . . . . .	114
7.2.1. Descripción de la Base de Datos <i>BD-762</i> . . . . .	115
7.2.2. Descripción de la Base de Datos <i>BD-7327</i> . . . . .	115
7.3. Resultados en <i>BD-762</i> . . . . .	115
7.3.1. Experimentos . . . . .	118
7.4. Resultados en <i>BD-7327</i> . . . . .	119

7.4.1. Experimentos . . . . .	119
<b>8. Bases de Datos No Métricas</b>	<b>125</b>
8.1. Norma Fraccionaria $L_p$ . . . . .	125
8.1.1. Experimentos . . . . .	126
<b>9. Conclusiones</b>	<b>132</b>
9.1. Trabajo Futuro . . . . .	134
<b>A. Nubes de Puntos</b>	<b>144</b>
A.1. Cubo Unitario . . . . .	145
A.1.1. Dimensión 8 . . . . .	145
A.1.2. Dimensión 128 . . . . .	145
A.2. Documentos . . . . .	145
A.3. Base de datos de Caras FERET . . . . .	151
A.4. Diccionario . . . . .	151
A.5. Espacios no Métricos . . . . .	151
A.5.1. $L_p$ con $p = 0,2$ . . . . .	151
A.5.2. $L_p$ con $p = 0,8$ . . . . .	151

# Índice de figuras

1.1. Proyección de un objeto complejo. . . . .	3
2.1. Tipos de consultas por proximidad. . . . .	13
2.2. Ejemplo de histogramas de distancias. . . . .	17
2.3. Proceso de consulta en la búsqueda por similitud. . . . .	18
2.4. Ejemplo de un trie. . . . .	21
2.5. Histograma de distancias del cubo unitario. . . . .	22
2.6. Histograma de distancias de las palabras. . . . .	23
2.7. Histograma de distancias de los documentos. . . . .	24
3.1. Ejemplo de un algoritmo basado en pivotes. . . . .	27
3.2. Ejemplo de un algoritmo basado en particiones compactas. . . . .	29
3.3. Familias de algoritmos del estado de arte. . . . .	31
3.4. Ejemplo de AESA. . . . .	32
3.5. Ejemplo del GNAT. . . . .	33
3.6. Ejemplo de BKT, FQT, FHQT Y FQA. . . . .	35
3.7. Ejemplo de un BKT. . . . .	36
3.8. Ejemplo de un VPT. . . . .	37
3.9. Ejemplo de un BST. . . . .	39
3.10. Ejemplo de un SAT. . . . .	40
3.11. Estado de arte en el tiempo. . . . .	41
3.12. Estado de arte. Dimensión 8 y 20. . . . .	44
3.13. Diferencia entre algoritmos exactos y aproximados. . . . .	46

4.1.	Ejemplo de órdenes en un espacio $\mathbb{R}^2$ . . . . .	50
4.2.	Ejemplo de las inversiones en $\mathbb{R}^2$ . . . . .	51
4.3.	Comparación entre los algoritmos basados en particiones compactas y los basados en inversiones. . . . .	55
4.4.	Comparación entre las particiones de Voronoi frente a los permutantes. . . . .	55
4.5.	Trie de permutaciones. . . . .	57
4.6.	Ejemplo del proceso de búsqueda en un trie basado en permutaciones. . . . .	58
4.7.	Desempeño de la técnica de las inversiones en un diccionario de palabras. . . . .	61
4.8.	Desempeño de la técnica de las inversiones en el cubo unitario. . . . .	62
5.1.	Ejemplo de la técnica basada en permutaciones. . . . .	65
5.2.	Tiempo de procesamiento de la técnica de las permutaciones con un Trie. . . . .	69
5.3.	Tiempo de procesamiento de la técnica basada en permutaciones usando qsort e IQS. . . . .	75
5.4.	Tiempo de procesamiento para la técnica de las permutaciones (usando PowerPC). . . . .	76
5.5.	Tiempo de procesamiento para la técnica de las permutaciones (usando Intel). . . . .	77
5.6.	Desempeño de los algoritmos probabilísticos en dimensiones altas. . . . .	78
5.7.	Comparando diferentes medidas de similitudes entre permutaciones. . . . .	79
5.8.	Comparación de las métricas $L_1$ y $L_\infty$ como base de ordenamiento. . . . .	79
5.9.	Desempeño de la técnica basada en permutaciones en un espacio de documentos. . . . .	80
5.10.	Comparación del desempeño entre algoritmos exactos e inexactos. . . . .	80
5.11.	Diferentes heurísticas para la selección de permutantes. . . . .	81
6.1.	Ejemplo del proceso de AESA e iAESA. . . . .	89
6.2.	Comparando iAESA contra AESA en distintas dimensiones. . . . .	91
6.3.	Comparación de AESA e iAESA para documentos. . . . .	92
6.4.	Comparación en tiempo entre AESA e iAESA. . . . .	93
6.5.	Desempeño de AESA, iAESA e iAESA2 probabilísticos en dimensión 128. . . . .	94
6.6.	Error relativo de las distancias en las consultas infructuosas. . . . .	95
6.7.	Comparación de AESA, iAESA e iAESA2 en dimensión 128. . . . .	95
6.8.	Comparación de iAESA, iAESA2 y ordenamiento usando permutaciones en dimensión 128. . . . .	96

6.9. Comparación de AESA, iAESA e iAESA2 probabilístico en dimensión 8. . . . .	98
6.10. Comparación de AESA, iAESA e iAESA2 aproximado en dimensión 16. . . . .	99
6.11. Desempeño de la técnica basada en permutaciones. Dimensiones 8 y 16. . . . .	100
6.12. Comparación entre la métrica dinámica del Footrule y la secuencial. . . . .	108
6.13. Comparación de tiempos de procesamiento de iAESA. . . . .	110
7.1. Sistema general de Reconocimiento de Caras. . . . .	113
7.2. Histogramas de la Base de Datos FERET (real y proyectado). . . . .	116
7.3. Histograma de la Base de Datos BD-7327 (real y proyectado). . . . .	117
7.4. Comparación del reconocimiento de caras sobre la base de datos FERET. . . . .	120
7.5. Tiempo de procesamiento para la base de datos de caras FERET. . . . .	121
7.6. Comparación del reconocimiento de caras sobre la base de datos FERET. . . . .	123
7.7. Tiempo de procesamiento para la base de datos de caras FERET. . . . .	124
8.1. Histograma de la desigualdad triangular. . . . .	126
8.2. Desempeño de las permutaciones bajo una norma fraccionaria, $p = 0,2$ y $p = 0,8$ dimensión 8. . . . .	128
8.3. Desempeño de las permutaciones bajo una norma fraccionaria, $p = 0,2$ . . . . .	129
8.4. Desempeño de las permutaciones bajo una norma fraccionaria, $p = 0,8$ . . . . .	130
8.5. Desempeño de las permutaciones bajo una norma fraccionaria. . . . .	131
A.1. Nubes de puntos: Cubo unitario, dimensión 8. . . . .	146
A.2. Nubes de puntos: Cubo unitario, dimensión 128. . . . .	147
A.3. Acercamiento en nubes de puntos: Cubo unitario, dimensión 128. . . . .	148
A.4. Nubes de puntos: Base de datos de documentos. . . . .	149
A.5. Acercamiento en nubes de puntos: Base de datos de documentos. . . . .	150
A.6. Nubes de puntos: Base de datos de caras. . . . .	152
A.7. Acercamiento en nubes de puntos: Base de datos de caras. . . . .	153
A.8. Nubes de puntos: Diccionario. . . . .	154
A.9. Acercamiento en nubes de puntos: Diccionario. . . . .	155
A.10. Nubes de puntos: Cubo unitario usando una norma fraccionaria $L_p$ con $p = 0,2$ . . .	156

A.11. Acercamiento nubes de puntos: Cubo unitario usando una norma fraccionaria  $L_p$  con  $p = 0,2$ . . . . . 157

A.12. Nubes de puntos: Cubo unitario usando una norma fraccionaria  $L_p$  con  $p = 0,8$ . . . 158

A.13. Acercamiento nubes de puntos: Cubo unitario usando una norma fraccionaria  $L_p$  con  $p = 0,8$ . . . . . 160

# Índice de cuadros

3.1. Comparación de los algoritmos para la búsqueda por proximidad en espacios métricos. . . . .	43
--	----

# Lista de algoritmos

1.	Algoritmo-pivotero-indexado . . . . .	27
2.	Algoritmo-pivotero-consulta . . . . .	27
3.	Búsqueda-trie( $T$ trie, $d_{max}$ , $nivel$ ) . . . . .	59
4.	Probar-hoja( $T$ trie, $d_{max}$ , $nivel$ ) . . . . .	60
5.	Trie_búsqueda(Trie R, Consulta $q$ ) . . . . .	67
6.	Sort-Búsqueda( $\Pi_q$ , $q$ ) . . . . .	67
7.	AESA . . . . .	85
8.	iAESA . . . . .	87
9.	Footrule-dinámico . . . . .	105
10.	Footrule-dinámico-estimado() . . . . .	109