

TABLA DE CONTENIDOS

	página
Dedicatoria	I
Agradecimientos	II
Tabla de Contenidos	III
Indice de Figuras	VI
Indice de Algoritmos	VII
Resumen	VIII
1. Introducción	9
1.1. Contexto	9
1.2. Áreas de aplicación	10
1.3. Descripción del problema	11
1.4. Objetivos	11
1.4.1. Objetivo general	12
1.4.2. Objetivos específicos	12
1.5. Organización del documento	12
2. Antecedentes	14
2.1. Espacios métricos	14
2.1.1. Espacios vectoriales	15
2.1.2. Medidas de distancia	16
2.1.3. Dimensión intrínseca	20
2.2. Consultas por proximidad	22
2.2.1. Consultas por rango	22
2.2.2. Consulta de vecinos cercanos	23
2.2.3. Unión por similitud	25
2.3. Indexación	25
2.3.1. Técnica de indexación basada en pivotes	26

2.3.2.	Esquema de particiones compactas	28
2.3.3.	Indexación por permutaciones	29
2.4.	Permutantes	30
2.4.1.	Midiendo similitud entre permutaciones	31
2.4.2.	Búsqueda aproximada usando permutaciones	32
3.	Metodología	34
3.1.	Desarrollo y proceso	34
3.2.	Obtención de los datos	34
3.3.	Cálculo de distancias	35
3.4.	Presentación e implementación de heurísticas	37
3.5.	Pruebas experimentales	37
3.5.1.	Bases de datos sintéticas	37
3.5.2.	Base de datos reales	39
3.6.	Análisis de los resultados	39
4.	Heurísticas de selección	41
4.1.	Selección aleatoria	41
4.2.	Selección por promedios de distancias	42
4.3.	Selección por varianza de distancias	43
4.4.	Selección espacial de permutantes dispersos	45
4.5.	Selección por cuantiles	46
5.	Refinación de permutantes	49
5.1.	Propuesta de refinación	50
5.2.	Criterios de evaluación	53
6.	Resultados experimentales	54
6.1.	Descripción de los experimentos	54
6.2.	Espacio métrico uniforme	54
6.3.	Espacio métrico con datos agrupados	60
6.4.	Imágenes como espacio métrico	63
6.5.	Análisis del proceso de refinación	67

7. Conclusiones	71
7.1. Trabajos futuros	73
Bibliografía	74

INDICE DE FIGURAS

	página
2.1. Métricas L_p en dos dimensiones	17
2.2. Histogramas de distancias para baja y alta dimensionalidad usando la norma L_2	21
2.3. Histograma de distancias para un diccionario usando distancia de edición.	22
2.4. Búsqueda por rango en un plano de dos dimensiones.	23
2.5. Búsqueda kNN en el plano, con $k=3$	24
2.6. Etapas de una búsqueda por similitud.	26
3.1. Construcción del vector de distancias	36
6.1. Resultados de la heurística “Random”	55
6.2. Recuperación de respuestas en un espacio métrico uniforme con 256 permutantes.	56
6.3. Recuperación de respuestas en un espacio métrico uniforme con 32 permutantes.	57
6.4. Tiempos de ejecución de selección de permutantes y consulta.	58
6.5. Tiempos de ejecución de selección de permutantes y consulta para las heurísticas de tipo SSS.	59
6.6. Tiempos de ejecución de selección de permutantes y consulta para las heurísticas de tipo Permutantil.	60
6.7. Recuperación de respuestas en un espacio métrico con clústers.	62
6.8. Recuperación de respuestas en un espacio métrico con clústers.	63
6.9. Recuperación de respuestas en imágenes.	65
6.10. Recuperación de respuestas en imágenes usando 32 permutantes.	66
6.11. Cantidad de permutantes reemplazables con 33.5 % de umbral	68
6.12. Cantidad de permutantes reemplazables con 34 % de umbral	69
6.13. Búsqueda refinada.	70

INDICE DE ALGORITMOS

	página
2.1. <i>Distancia Levenshtein</i>	19
2.2. <i>Creación del índice usando pivotes</i>	27
2.3. <i>Proceso de consulta usando pivotes</i>	28
2.4. <i>Búsqueda aproximada usando permutaciones</i>	33
2.5. <i>Selección de permutantes</i>	33
3.1. <i>Construcción del vector de distancias</i>	36
4.1. <i>Selección aleatoria de permutantes</i>	41
4.2. <i>Selección por promedio ascendente de distancias</i>	43
4.3. <i>Cálculo de promedios de distancias</i>	43
4.4. <i>Selección por varianza ascendente de distancias</i>	44
4.5. <i>Cálculo de varianza por objeto de la base de datos.</i>	44
4.6. <i>Selección de permutantes usando SSS</i>	46
4.7. <i>Selección de permutantes usando cuantiles</i>	47
5.1. <i>Refinación de permutantes</i>	52