

TABLA DE CONTENIDOS

	página
Tabla de Contenidos	I
Índice de Figuras	IV
Índice de Tablas	VIII
1. Introducción	9
1.1. Presentación del tema	9
1.2. Objetivos	10
1.2.1. Objetivo general	10
1.2.2. Objetivos específicos	11
1.2.3. Alcances	11
1.3. Descripción de los contenidos	11
2. Antecedentes	13
2.1. Estructuras de datos y algoritmos	13
2.1.1. Grafos	13
2.1.1.1. Conceptos y terminologías de grafos	16
2.1.1.2. Representaciones de grafos	18
2.1.1.3. Algoritmos de recorrido en grafos	20
2.2. Espacios métricos	21
2.2.1. Consultas por proximidad	22
2.2.2. Maldición de la dimensionalidad	25
2.2.3. Ejemplo de espacios métricos	26
2.2.3.1. Espacios vectoriales	26
2.2.3.2. Diccionarios	27
2.3. Estrategias utilizadas en espacios métricos	28
2.3.1. Métodos basados en particiones	28
2.3.1.1. Particionado por zonas esféricas (ball partitioning)	29
2.3.1.2. Particionado por hiperplanos	30
2.4. Soluciones existentes en espacios métricos	31

2.4.1.	Métodos basados en particiones	31
2.4.1.1.	Bisector tree (BST)	31
2.4.1.2.	Generalized hyperplane tree (GHT)	33
2.4.2.	Otros métodos	35
2.4.2.1.	Método basado en permutaciones	35
3.	Metodología	39
3.1.	Análisis y diseño	39
3.2.	Implementación	40
3.3.	Experimentos y análisis de los resultados	42
3.3.1.	Diccionarios de búsqueda	42
3.3.2.	Descripción de la máquina de trabajo	42
4.	Construcción de los índices de firmas binarias	44
4.1.	Construcción primer índice (Firma1)	44
4.2.	Construcción segundo índice (Firma2)	46
5.	Implementación de búsquedas por similitud	52
6.	Resultados experimentales	58
6.1.	Tiempo de construcción de índice Firma1 y Firma2	58
6.2.	Búsqueda de k -vecinos	59
6.2.1.	Experimentos en diccionarios	59
6.2.1.1.	Resultados índice Firma1	60
6.2.1.2.	Resultados índice Firma2	61
6.2.1.3.	Resultados índice Permutante	63
6.2.1.4.	Permutantes vs Firma1 y Firma2	63
6.2.2.	Experimentos en vectores	65
6.2.2.1.	Resultados índice Firma1	65
6.2.2.2.	Resultados índice Firma2	66
6.2.2.3.	Resultados índice Permutante	69
6.2.2.4.	Permutantes vs Firma1 y Firma2	73
6.2.2.5.	Permutantes vs Firma1 a igual espacio, 10% base de datos revisada	79
6.3.	Búsqueda por rango	79

6.3.1.	Experimentos en diccionarios	81
6.3.1.1.	Resultados índice Firma1	81
6.3.1.2.	Resultados índice Firma2	81
6.3.1.3.	Resultados índice Permutante	84
6.3.1.4.	Permutantes vs Firma1 y Firma2	84
6.3.2.	Experimentos en Vectores	86
6.3.2.1.	Resultados índice Firma1	86
6.3.2.2.	Resultados índice Firma2	86
6.3.2.3.	Resultados índice Permutante	88
6.3.2.4.	Permutantes vs Firma1 y Firma2	88
7.	Conclusiones y trabajo futuro	92
7.1.	Conclusiones	92
7.2.	Trabajo Futuro	93
	Bibliografía	95
	A. Publicación generada	97

ÍNDICE DE FIGURAS

	página
2.1. Grafo no dirigido.	14
2.2. Grafo dirigido.	15
2.3. Grafo no dirigido ponderado	19
2.4. Lista de adyacencia	20
2.5. Algoritmo de recorrido (o búsqueda) en profundidad.	21
2.6. Algoritmo de recorrido (o búsqueda) en amplitud.	21
2.7. Búsqueda por rango.	22
2.8. Búsqueda 5-Vecinos más cercanos.	23
2.9. Ejemplo de indexación y consultas.	24
2.10. Histogramas de distancia con dimensión intrínseca baja(izquierda) y alta (derecha).	26
2.11. Particionado por zonas esféricas.	29
2.12. Criterio de descarte en zonas esféricas.	30
2.13. Particionado por hiperplanos.	31
2.14. Criterio de descarte en Hiperplanos.	32
2.15. Particionado por hiperplanos (BST).	32
2.16. Índice del método (BST).	33
2.17. Fase inicial del proceso de búsqueda (BST).	34
2.18. Fase inicial del proceso de búsqueda (BST).	36
2.19. Índice de permutantes.	37
2.20. Ejemplo de consulta (índice de permutantes).	38
3.1. Espacio métrico dividido a través de 2 hiperplanos.	40
4.1. Arreglo de elementos y arreglo de firmas de la Figura 3.1.	45
4.2. Algoritmo de construcción del índice básico de firmas binarias.	46
4.3. Espacio métrico con 2 hiperplanos y firmas etiquetadas (Índice Firma2).	48
4.4. Grafo generado por las firmas de la Figura 4.3 (Índice Firma2).	48
4.5. Espacio métrico con 4 hiperplanos y firmas etiquetadas (Índice Firma2).	49
4.6. Grafo generado por las firmas de la Figura 4.5 (Índice Firma2).	49
4.7. Algoritmo de construcción del índice Firma2.	50

4.8. Algoritmo de que ingresa una nueva firma si es necesario y guarda las posiciones de los elementos cargados en memoria en la Firma que corresponde.	51
5.1. Algoritmo de búsqueda de los k -vecinos más cercanos para el índice Firma1.	54
5.2. Algoritmo de búsqueda por rango para el índice Firma1.	55
5.3. Algoritmo de búsqueda de los k -vecinos más cercanos para el índice Firma2.	56
5.4. Algoritmo de búsqueda por rango para el índice Firma2.	57
6.1. Tiempo de construcción variando cantidad de hiperplanos (índice Firma2).	59
6.2. Tiempo de construcción variando cantidad de elementos (índice Firma2).	60
6.3. Porcentaje recuperación en diccionarios dado $k=1$ (índice Firma1).	61
6.4. Porcentaje recuperación en diccionarios dado $k=20$ (índice Firma1).	62
6.5. Porcentaje recuperación en diccionarios dado $k=1$ (índice Firma2).	62
6.6. Porcentaje recuperación en diccionarios dado $k=20$ (índice Firma2).	63
6.7. Porcentaje recuperación en diccionarios dado $k=1$ (índice Permutante).	64
6.8. Porcentaje recuperación en diccionarios dado $k=20$ (índice Permutante).	64
6.9. Porcentaje recuperación en diccionarios $k=1$ y $k=20$ Permutantes vs F1 y F2.	65
6.10. Porcentaje de recuperación en vectores dado $k=1$ y $dim=32$ (índice Firma1).	66
6.11. Porcentaje de recuperación en vectores dado $k=20$ y $dim=32$ (índice Firma1).	67
6.12. Porcentaje de recuperación en vectores dado $k=1$ y $dim=64$ (índice Firma1).	67
6.13. Porcentaje de recuperación en vectores dado $k=1$ y $dim=64$ (índice Firma1).	68
6.14. Porcentaje de recuperación en vectores dado $k=1$ y $dim=128$ (índice Firma1).	68
6.15. Porcentaje de recuperación en vectores dado $k=1$ y $dim=128$ (índice Firma1).	69

6.16. Porcentaje de recuperación en vectores dado $k=1$ y $dim=32$ (índice <i>Firma2</i>).	70
6.17. Porcentaje de recuperación en vectores dado $k=20$ y $dim=32$ (índice <i>Firma2</i>).	70
6.18. Porcentaje de recuperación en vectores dado $k=1$ y $dim=64$ (índice <i>Firma2</i>).	71
6.19. Porcentaje de recuperación en vectores dado $k=20$ y $dim=64$ (índice <i>Firma2</i>).	71
6.20. Porcentaje de recuperación en vectores dado $k=1$ y $dim=128$ (índice <i>Firma2</i>).	72
6.21. Porcentaje de recuperación en vectores dado $k=20$ y $dim=128$ (índice <i>Firma2</i>).	72
6.22. Porcentaje de recuperación en vectores dado $k=1$ y $dim=32$ (índice <i>Permutante</i>).	73
6.23. Porcentaje de recuperación en vectores dado $k=1$ y $dim=32$ (índice <i>Permutante</i>).	74
6.24. Porcentaje de recuperación en vectores dado $k=1$ y $dim=64$ (índice <i>Permutante</i>).	74
6.25. Porcentaje de recuperación en vectores dado $k=1$ y $dim=64$ (índice <i>Permutante</i>).	75
6.26. Porcentaje de recuperación en vectores dado $k=1$ y $dim=128$ (índice <i>Permutante</i>).	75
6.27. Porcentaje de recuperación en vectores dado $k=1$ y $dim=128$ (índice <i>Permutante</i>).	76
6.28. Porcentaje recuperación en vectores $dim=32$, $k=1$ y $k=20$, <i>Permutantes vs F1 y F2</i>	77
6.29. Porcentaje recuperación en vectores $dim=64$, $k=1$ y $k=20$ <i>Permutantes vs F1 y F2</i>	77
6.30. Porcentaje recuperación en vectores $dim=128$, $k=1$ y $k=20$ <i>Permutantes vs F1 y F2</i>	78
6.31. Recuperación porcentaje revisión fijo, vectores $dim=32$, <i>Permutantes vs Firma1</i>	80
6.32. Recuperación porcentaje revisión fijo, vectores $dim=64$, <i>Permutantes vs Firma1</i>	80

6.33. Recuperación porcentaje revisión fijo, vectores $dim=128$, Permutantes vs Firma1.	81
6.34. Porcentaje recuperación en String rango = 1 (índice Firma1).	82
6.35. Porcentaje recuperación en String rango = 1 (índice Firma1).	82
6.36. Porcentaje recuperación en String rango = 1 (índice Firma2).	83
6.37. Porcentaje recuperación en String rango = 1 (índice Firma2).	83
6.38. Porcentaje recuperación en String rango = 1 (índice Permutante).	84
6.39. Porcentaje recuperación en String rango = 1 (índice Permutante).	85
6.40. Resumen recuperación búsqueda por rango (índice Permutante vs Firma1 y Firma2), Strings.	85
6.41. Porcentaje recuperación en vectores donde el rango permite recuperar 100 respuestas (índice Firma1).	86
6.42. Porcentaje recuperación en vectores donde el rango permite recuperar 2000 respuestas (índice Firma1).	87
6.43. Porcentaje recuperación en vectores donde el rango permite recuperar 100 respuestas (índice Firma2).	87
6.44. Porcentaje recuperación en vectores donde el rango permite recuperar 2000 respuestas (índice Firma1).	88
6.45. Porcentaje recuperación en vectores donde el rango permite recuperar 100 respuestas (índice Permutante).	89
6.46. Porcentaje recuperación en vectores donde el rango permite recuperar 2000 respuestas (índice Permutante).	89
6.47. Resumen recuperación búsqueda por rango (índice Permutante vs Firma1 y Firma2), Vectores $dim=32$	90
6.48. Resumen recuperación búsqueda por rango (índice Permutante vs Firma1 y Firma2), Vectores $dim=64$	91
6.49. Resumen recuperación búsqueda por rango (índice Permutante vs Firma1 y Firma2), Vectores $dim=128$	91

ÍNDICE DE TABLAS

	página
2.1. Eficiencia matriz de adyacencia y lista de adyacencia	19
2.2. Matriz de adyacencia del grafo de la Figura 2.3	20
6.1. Tabla de firmas binarias para mostrar error al buscar para el índice Firma2.	78
6.2. N° permutantes y hiperplanos a mismo espacio.	79