

## TABLA DE CONTENIDOS

	página
<b>Dedicatoria</b>	<b>I</b>
<b>Agradecimientos</b>	<b>II</b>
<b>Tabla de Contenidos</b>	<b>IV</b>
<b>Índice de Figuras</b>	<b>VIII</b>
<b>Índice de Tablas</b>	<b>IX</b>
<b>Resumen</b>	<b>X</b>
<b>Abstract</b>	<b>XI</b>
<b>1. Introducción</b>	<b>2</b>
1.1. Presentación del tema . . . . .	2
1.2. Objetivos . . . . .	4
1.2.1. Objetivo general . . . . .	4
1.2.2. Objetivos específicos . . . . .	4
1.3. Descripción de los contenidos . . . . .	5
<b>2. Marco de Referencia</b>	<b>6</b>
2.1. Estructuras de datos y algoritmos . . . . .	6
2.1.1. Colas de prioridad . . . . .	6
2.1.2. Grafos . . . . .	7
2.1.2.1. Formas de representar un grafo . . . . .	8
2.1.2.2. Caminos más cortos . . . . .	9
2.1.2.3. Algoritmo de Dijkstra . . . . .	10
2.2. Espacios métricos . . . . .	12
2.2.1. Ejemplos de espacios métricos . . . . .	12
2.2.1.1. Espacios Vectoriales . . . . .	12
2.2.1.2. Modelo vectorial para documentos . . . . .	13
2.2.1.3. Diccionarios . . . . .	14

2.2.2.	Búsquedas en espacios métricos . . . . .	15
2.2.2.1.	Ejemplos de aplicaciones de búsquedas en espacios métricos . . . . .	15
2.2.2.2.	Consultas de interés en espacios métricos . . . . .	17
2.2.2.3.	Algoritmos de búsqueda en espacios métricos . . . . .	18
2.2.3.	La maldición de la dimensionalidad . . . . .	20
2.3.	Trabajos previos y relacionados . . . . .	21
2.3.1.	$k$ NNGs en baja y media dimensionalidad . . . . .	22
2.3.2.	Lista de clusters . . . . .	22
2.3.2.1.	Selección de centro . . . . .	25
2.3.2.2.	Selección de radio . . . . .	26
2.3.2.3.	Construcción . . . . .	26
2.3.3.	Lista de clusters recursiva . . . . .	27
2.3.3.1.	Definiciones . . . . .	27
2.3.3.2.	Análisis de $RLC$ . . . . .	29
2.3.4.	Algoritmos paralelos para la construcción de $k$ NNGs . . . . .	32
2.3.4.1.	Modelo . . . . .	32
2.3.4.2.	Consultas locales de vecinos más cercanos . . . . .	32
2.3.4.3.	Datos y flujos de control . . . . .	33
2.3.4.4.	Cálculos distribuidos . . . . .	33
2.3.5.	La Receta . . . . .	34
2.3.5.1.	Estructura de Datos Principal . . . . .	34
2.3.5.2.	Administración del $NHA$ . . . . .	35
2.3.5.3.	Uso del $NHA$ como un Grafo . . . . .	35
2.3.5.4.	Usando la Simetría de $d$ . . . . .	36
2.3.5.5.	$\mathbb{U}$ es Fijo . . . . .	36
2.3.5.6.	Check Order Heap . . . . .	37
2.3.6.	Algoritmos para la construcción de $k$ NNGs . . . . .	37
2.3.6.1.	Algoritmo básico para la construcción del $k$ NNG . . . . .	37
2.3.6.2.	Algoritmo basado en pivotes . . . . .	39
2.3.6.3.	Algoritmo basado en <i>Bisector Tree</i> . . . . .	40
<b>3.</b>	<b>Metodología</b>	<b>41</b>
3.1.	Implementación . . . . .	41

3.1.1.	Lenguaje C++ . . . . .	42
3.1.2.	Compilador . . . . .	42
3.1.3.	IDE . . . . .	43
3.2.	Planificación de pruebas . . . . .	43
3.2.1.	Espacios métricos de prueba . . . . .	44
3.2.1.1.	Espacio vectorial distribuido uniformemente bajo la distancia Euclidiana . . . . .	44
3.2.1.2.	Cadenas de caracteres bajo la distancia de edición . . . . .	44
3.2.1.3.	Documentos bajo la distancia coseno . . . . .	45
3.2.1.4.	Espacio vectorial con distribución Gaussiana bajo la distancia Euclidiana . . . . .	45
3.2.2.	Descripción de la estación de trabajo . . . . .	46
<b>4.</b>	<b>Construcción de <math>k</math>NNG con Lista de <i>Clusters</i></b> . . . . .	<b>47</b>
4.1.	Algoritmo de construcción del $k$ NNG usando Lista de <i>Clusters</i> . . . . .	47
4.2.	Etapa 1: Construcción del Índice . . . . .	48
4.3.	Etapa 2: Construyendo el $k$ NNG . . . . .	49
4.4.	Resultados Experimentales y Análisis . . . . .	53
4.4.1.	Sintonización de parámetros para <i>LC</i> y <i>RLC</i> . . . . .	53
4.4.1.1.	Mejorando <i>RLC</i> . . . . .	53
4.4.1.2.	Evaluando las actualizaciones de la base de datos . . . . .	54
4.4.1.3.	Comparando <i>RLC</i> con otras estructuras de datos . . . . .	56
4.4.2.	Espacio vectorial distribuido uniformemente bajo la distancia Euclidiana . . . . .	58
4.4.2.1.	Construcción . . . . .	58
4.4.3.	Cadenas de caracteres bajo la distancia de edición . . . . .	62
4.4.3.1.	Construcción . . . . .	62
4.4.4.	Documentos bajo la distancia coseno . . . . .	62
4.4.4.1.	Construcción . . . . .	62
4.4.5.	Espacio vectorial con distribución Gaussiana bajo la distancia Euclidiana . . . . .	64
4.4.5.1.	Construcción . . . . .	64

<b>5. Conclusiones y trabajo futuro</b>	<b>67</b>
5.1. Conclusiones . . . . .	67
5.2. Trabajo futuro . . . . .	68
<b>Bibliografía</b>	<b>70</b>

## ÍNDICE DE FIGURAS

	página
2.1. Grafo no dirigido . . . . .	9
2.2. Algoritmo de Dijkstra . . . . .	11
2.3. Búsqueda de los $k$ vecinos más cercanos y búsqueda por rango . . . .	18
2.4. Histogramas de distancias para un Espacio Métrico . . . . .	21
2.5. Algoritmo de construcción de la Lista de <i>Clusters</i> con radio fijo. . . .	24
2.6. Algoritmo de construcción de la Lista de <i>Clusters</i> con tamaño fijo. . .	24
2.7. Algoritmo de búsqueda en la Lista de <i>Clusters</i> . . . . .	25
2.8. Lista de <i>Clusters</i> Recursiva . . . . .	28
2.9. La receta de construcción de $k$ NNGs. . . . .	36
2.10. $\mathbb{U}$ es Fijo . . . . .	37
2.11. Algoritmo básico para la construcción del $k$ NNG . . . . .	38
2.12. <i>KNNpiv</i> . . . . .	39
2.13. <i>KNNrp</i> . . . . .	40
4.1. Algoritmo de $k$ NNG usando Lista de <i>Clusters</i> . . . . .	48
4.2. Algoritmo para el cálculo de distancia en el $k$ NNG usando Lista de Clusters . . . . .	49
4.3. Relación entre el <i>cluster</i> y la consulta . . . . .	50
4.4. Procedimiento auxiliar <b>extractFrom</b> . . . . .	52
4.5. Divisor de Radio vs Distancias Dimensión 20 . . . . .	55
4.6. Número promedio de distancias calculadas al borrar objetos . . . . .	56
4.7. Comparación de <i>RLC</i> con otras estructuras de datos . . . . .	57
4.8. Proceso de construcción de <i>RLC</i> versus <i>LC</i> . . . . .	58
4.9. Evaluación de la construcción de $k$ NNG en espacios vectoriales . . . .	61
4.10. $k$ NNGLC y $k$ NNGRLC en el espacio de las cadenas de caracteres . .	63
4.11. $k$ NNGLC y $k$ NNGRLC en el espacio de los documentos . . . . .	64
4.12. $k$ NNGLC y $k$ NNGRLC en espacios Gaussianos . . . . .	66

## ÍNDICE DE TABLAS

	página
2.1. Matriz de adyacencia . . . . .	9
4.1. Nivel máximo de los <i>clusters</i> . . . . .	54
4.2. <i>KNNGLC</i> Ajuste de mínimos cuadrados para espacios métricos de vectores . . . . .	59
4.3. <i>KNNGRLC</i> Ajuste de mínimos cuadrados para espacios métricos de vectores . . . . .	59
4.4. <i>KNNGLC</i> ajuste de mínimos cuadrados para espacios métricos Gaussianos . . . . .	65
4.5. <i>KNNGRLC</i> ajuste de mínimos cuadrados para espacios métricos Gaussianos . . . . .	65