

## Índice

Introducción . . . . .	1
1.1. Descripción del Contenido y Contribuciones de la Memoria . . . . .	2
Conceptos Básicos . . . . .	3
2.2. Strings . . . . .	3
2.3. Árboles trie . . . . .	3
2.4. Árbol de sufijos . . . . .	4
2.5. Arreglo de sufijos . . . . .	5
2.6. Entropía de un texto . . . . .	5
2.7. Codificaciones . . . . .	6
2.8. Bitmaps . . . . .	7
2.8.1. Munro y Clark . . . . .	8
2.8.2. Raman, Raman y Rao . . . . .	8
2.8.3. Okanohara y Sadakane . . . . .	8
2.8.4. Codificación delta para bitmaps poco densos . . . . .	9
2.9. <i>Directly addressable codes</i> . . . . .	10
2.10. Estructura de datos sucintas y autoíndices . . . . .	10
2.11. Arreglo de sufijos comprimido . . . . .	11
2.12. La Transformada de Burrows-Wheeler (BWT) y el FM-index . . . . .	11
2.13. Longest common prefix (LCP) . . . . .	13
2.14. Colecciones repetitivas . . . . .	13
2.15. Re-Pair: Compresión basada en gramáticas . . . . .	13
Representaciones Sucintas de Árboles de Sufijos . . . . .	16
3.16. Árboles sucintos . . . . .	16
3.16.1. Range min-max tree . . . . .	17
3.17. Árboles de sufijos comprimidos . . . . .	18
3.17.1. Árbol de sufijos de Sadakane . . . . .	19
3.17.2. Árbol de sufijos de Russo et al. . . . .	21
3.17.3. Árbol de sufijos de Fischer et al. . . . .	22
3.18. NPR: next smaller value, previous smaller value y range minimum query . . . . .	23
3.18.1. NPR de Fischer et al. (NPR-FMN) . . . . .	23
3.18.2. NPR de Cánovas y Navarro (NPR-CN) . . . . .	24
Nuestro Árbol de Sufijos Comprimido para Secuencias Repetitivas . . . . .	26
4.19. Arreglo de sufijos comprimido con run-length (RLCSA) . . . . .	26
4.19.1. Resultados Experimentales . . . . .	26
4.20. Representación del Arreglo <i>Longest Common Prefix</i> (LCP) . . . . .	27
4.20.1. Resultados Experimentales . . . . .	28
4.21. Una solución novedosa para NPR basada en Re-Pair . . . . .	30
4.21.1. Consultas <i>NSV</i> , <i>PSV</i> y <i>RMQ</i> . . . . .	31
4.21.2. Construcción del NPR . . . . .	33
4.21.3. Resultados experimentales . . . . .	36
Conclusiones . . . . .	43
5.22. Contribuciones de esta memoria . . . . .	43

5.22.1. Compresión del arreglo <i>LCP</i> . . . . .	43
5.22.2. Estructura que soporta la operaciones <i>NSV</i> , <i>PSV</i> y <i>RMQ</i> . . . . .	43
5.23. Trabajo Futuro . . . . .	44
Referencias . . . . .	45

## Lista de Figuras

2.1. Ejemplo de un trie y un árbol patricia. . . . .	3
2.2. Árbol de sufijos para el texto “alabar_a_la_alabarda\$”. . . . .	4
2.3. Ejemplo de un arreglo de sufijos. . . . .	5
2.4. Ejemplo de rank y select. . . . .	7
2.5. Ejemplo de la estructura <i>DAC</i> . . . . .	10
2.6. Ejemplo de transformada de Burrows Wheeler. . . . .	12
2.7. Ejemplo de compresión <i>Re-Pair</i> representado árbol . . . . .	14
3.8. Ejemplo de representaciones sucintas basadas en paréntesis balanceados de un árbol. . . . .	16
3.9. Ejemplo de un range min-max tree. . . . .	19
3.10. Árbol de sufijos de Sadakane para el texto “alabar_a_la_alabarda\$”. . . . .	20
4.11. Resultados de tiempo/espacio al acceder a posiciones al azar del arreglo LCP. . . . .	29
4.12. Ejemplo de estructura <i>NPR-RP</i> . . . . .	32
4.13. Resultados de Tiempo/Eso para <i>NPR</i> sobre el texto Einstein. . . . .	37
4.14. Resultados de Tiempo/Eso para <i>NPR</i> sobre el texto Escherichia Coli. . . . .	38
4.15. Resultados de Tiempo/Eso para <i>NPR</i> sobre el texto PARA. . . . .	39
4.16. Resultados de Tiempo/Eso para <i>NPR</i> sobre el texto Influenza. . . . .	40
4.17. Resultados de Tiempo/Eso para <i>NPR</i> sobre el texto DNA. . . . .	41

## Lista de Tablas

2.1. Operaciones sobre los nodos de un árbol de sufijos. . . . .	5
2.2. Ejemplo de las distintas codificaciones de enteros. . . . .	7
2.3. Espacio en bits ocupado y tiempo que toma rank y select para distintas estructuras de datos. . . . .	7
3.4. Operaciones sobre árboles, reducidas en tiempo constante sobre arreglo BP . . . . .	17
4.5. Textos usados para realizar todos los experimentos. . . . .	27
4.6. Arreglo de Sufijos RLCSA. . . . .	27
4.7. Compresión con técnica <i>Re-Pair</i> apilando. . . . .	34
4.8. Compresión con técnica <i>Re-Pair</i> encolando. . . . .	34