

# Tabla de contenidos

1	ANALÍTICA TEXTUAL .....	25
1.1	INTRODUCCIÓN .....	25
1.2	MINERÍA DE TEXTOS Y ANALÍTICA TEXTUAL .....	28
1.3	TAREAS Y APLICACIONES .....	30
1.4	EL PROCESO DE LA ANALÍTICA TEXTUAL .....	33
1.5	RESUMEN .....	36
1.6	PREGUNTAS .....	37
2	PROCESAMIENTO DEL LENGUAJE NATURAL .....	39
2.1	INTRODUCCIÓN .....	39
2.2	PROCESAMIENTO DEL LENGUAJE NATURAL .....	41
2.3	NIVELES Y TAREAS EN NLP .....	42
2.3.1	Fonología .....	43
2.3.2	Morfología .....	44
2.3.3	Léxico .....	45
2.3.4	Sintaxis .....	51
2.3.5	Semántica .....	55
2.3.6	Razonamiento y pragmática .....	60
2.4	RESUMEN .....	60
2.5	EJERCICIOS .....	62
2.5.1	Análisis morfológico .....	62
2.5.2	Análisis léxico .....	66
2.5.3	Análisis sintáctico .....	68
3	EXTRACCIÓN DE INFORMACIÓN .....	71
3.1	INTRODUCCIÓN .....	71
3.2	EXTRACCIÓN DE INFORMACIÓN BASADA EN REGLAS .....	75
3.3	EXTRACCIÓN DE ENTIDADES NOMBRADAS .....	76
3.3.1	Modelos de N-gramas .....	78
3.4	EXTRACCIÓN DE RELACIONES .....	81
3.5	EVALUACIÓN .....	86
3.6	RESUMEN .....	88
3.7	EJERCICIOS .....	90

3.7.1	Extracción vía expresiones regulares .....	90
3.7.2	Reconocimiento de entidades nombradas (NER).....	94
4	REPRESENTACIÓN DE DOCUMENTOS .....	97
4.1	INTRODUCCIÓN .....	97
4.2	INDEXACIÓN DE DOCUMENTOS .....	99
4.3	MODELOS DE ESPACIO VECTORIAL .....	101
4.3.1	Modelo de representación booleana .....	102
4.3.2	Modelo de frecuencia de términos .....	103
4.3.3	Modelo de frecuencia inversa de documentos .....	104
4.4	RESUMEN .....	106
4.5	EJERCICIOS.....	107
4.5.1	Modelo de representación TFxIDF .....	107
5	ANÁLISIS DE REGLAS DE ASOCIACIÓN .....	115
5.1	INTRODUCCIÓN .....	115
5.2	PATRONES DE ASOCIACIÓN.....	116
5.3	EVALUACIÓN .....	118
5.3.1	Support.....	118
5.3.2	Confidence .....	119
5.3.3	Lift.....	119
5.4	GENERACIÓN DE REGLAS DE ASOCIACIÓN .....	120
5.5	RESUMEN .....	124
5.6	EJERCICIOS.....	126
5.6.1	Extracción de reglas de asociación .....	126
6	ANÁLISIS SEMÁNTICO BASADO EN CORPUS .....	131
6.1	INTRODUCCIÓN .....	131
6.2	ANÁLISIS BASADO EN CORPUS .....	133
6.3	ANÁLISIS SEMÁNTICO LATENTE .....	135
6.3.1	Generación de vectores con LSA .....	136
6.4	WORD2VEC.....	140
6.4.1	Aprendizaje de embeddings en CBOW.....	143
6.4.2	Predicción e interpretación de embeddings .....	146
6.5	RESUMEN .....	148
6.6	EJERCICIOS.....	149
6.6.1	Análisis semántico latente (LSA).....	149
6.6.2	Modelo de Word embedding del tipo Word2Vec .....	156
7	AGRUPACIÓN DE DOCUMENTOS .....	161
7.1	INTRODUCCIÓN .....	161

7.2	CLUSTERING DE DOCUMENTOS.....	163
7.3	CLUSTERING K-MEANS.....	169
7.4	MAPAS AUTOORGANIZATIVOS.....	172
7.4.1	Aprendizaje de mapas topológicos.....	174
7.5	RESUMEN.....	178
7.6	EJERCICIOS.....	179
7.6.1	Clustering via K-means.....	179
7.6.2	Clustering vía mapas autoorganizativos.....	185
8	MODELAMIENTO DE TÓPICOS.....	188
8.1	INTRODUCCIÓN.....	189
8.2	MODELAMIENTO DE TÓPICOS.....	191
8.3	LATENT DIRICHLET ALLOCATION.....	193
8.4	EVALUACIÓN.....	200
8.5	RESUMEN.....	202
8.6	EJERCICIOS.....	203
8.6.1	Modelamiento de tópicos con LDA.....	203
9	CATEGORIZACIÓN DE DOCUMENTOS.....	209
9.1	INTRODUCCIÓN.....	209
9.2	MODELOS DE CATEGORIZACIÓN.....	211
9.3	CLASIFICACIÓN BAYESIANA.....	214
9.4	CATEGORIZACIÓN POR MÁXIMA ENTROPÍA.....	218
9.5	EVALUACIÓN.....	223
9.6	RESUMEN.....	225
9.7	EJERCICIOS.....	227
9.7.1	Categorización con Naïve Bayes.....	227
9.7.2	Categorización con Máxima Entropía.....	232
10	CONCLUSIONES.....	239
	Bibliografía.....	244
	Glosario.....	244
	Índice onomástico.....	244