

Minería de textos en la red social Twitter

Índice de contenidos

AGRADECIMIENTOS	III
RESUMEN	V
1 INTRODUCCIÓN	1
1.1 MOTIVACIÓN	1
1.2 OBJETIVOS	2
1.3 LÍMITES DE LA TESIS	3
1.4 VISTA GENERAL DE LA TESIS	4
2 PANORAMA GENERAL.....	7
2.1 MINERÍA DE TEXTOS.....	7
2.2 MINERÍA WEB	7
2.3 REDES SOCIALES.....	8
2.4 ACERCA DE TWITTER.....	9
3 MODELO DE ESPACIOS VECTORIALES	13
3.1 CONSTRUCCIÓN	14
3.2 PESO LOCAL.....	14
3.2.1 Binaria.....	15
3.2.2 Frecuencia del término.....	15
3.2.3 Frecuencia aumentada de términos normalizados	15
3.2.4 Logarítmico.....	16
3.3 PESO GLOBAL.....	16
3.3.1 Sin cambios	17
3.3.2 Frecuencia inversa de documentos (IDF)	17
3.3.3 Otros esquemas IDF.....	17
3.3.4 Entropía normalizada	18
3.4 NORMALIZACIÓN	18
3.4.1 Sin cambios	19
3.4.2 Normalización de coseno	19
3.4.3 Otros esquemas de normalización de coseno.....	19
3.4.4 Normalización del peso máximo.....	19
3.4.5 Normalización de pivote único	19
3.5 MÉTODOS MÁS UTILIZADOS.....	20

4	ANÁLISIS DE GRUPOS (CLUSTERING)	21
4.1	DISTANCIAS Y SIMILITUDES.....	21
4.1.1	Distancias para variables continuas	23
4.1.2	Similitudes para variables binarias	24
4.2	MÉTODOS JERÁRQUICOS.....	26
4.2.1	Dendograma y jerarquía indexada	26
4.2.2	Método del mínimo.....	27
5	APLICACIONES EMPRESARIALES CON TECNOLOGÍA JAVA EE.	31
5.1	INTRODUCCIÓN A LA TECNOLOGÍA JAVA EE.	33
5.1.1	Arquitectura	33
5.1.2	Componentes de aplicación	35
5.1.3	Los API de Java EE	36
5.1.4	Tecnologías de comunicación.....	38
5.1.5	¿Java EE para cualquier aplicación?.....	39
6	IMPLEMENTACIÓN Y RESULTADOS	40
6.1	PRESENTACIÓN GENERAL DEL SISTEMA	40
6.2	OBTENCIÓN DE LOS GORJEOS Y PREPROCESAMIENTO	41
6.2.1	Obtención de gorjeos	42
6.2.2	Preprocesamiento	43
6.2.3	Clasificación.....	44
6.3	PROCESAMIENTO DE LOS GORJEOS	45
6.3.1	Construcción del espacio vectorial	45
6.3.2	Cálculo de la energía textual.....	46
6.3.3	Agrupamiento de gorjeos.....	48
6.4	PRESENTACIÓN DE DATOS	50
6.5	CONTROL DE CRECIMIENTO.....	55
7	ANÁLISIS DE RESULTADOS	56
8	CONCLUSIONES Y TRABAJO FUTURO	66
9	REFERENCIAS	69
	APÉNDICE A	71
	APÉNDICE B	72
	APÉNDICE C	73

Índice de Diagramas

Figura 1 – Matriz documento-termino simple	14
Figura 2 – Dendograma correspondiente al ejemplo del vecino más próximo.....	30
Figura 3 – Arquitectura de Java EE	34
Figura 4 – Arquitectura del sistema propuesto	40
Figura 5 – Ejemplo de gorjeos	42
Figura 6 – Resultado de aplicar las transformaciones a los gorjeos	44
Figura 7 – Palabras tomadas en cuenta para el espacio vectorial	45
Figura 8 – Grupos generados para gorjeos de ejemplo.....	49
Figura 9 – Página de inicio	51
Figura 10 – Presentación de datos para el tema <i>Homicidios</i>	52
Figura 11 – Continuación de presentación de datos para el tema <i>Homicidios</i>	53
Figura 12 – Presentación de gorjeo original	53
Figura 13 – Presentación de gorjeo procesado.....	54
Figura 14 – Presentación de gorjeo vectorizado	54
Figura 15 – Resultados de los agrupamientos para los temas Tráfico, Homicidios y Asaltos	62
Figura 16 – Ejemplo del cálculo de una matriz cofenética.	64

Índice de Tablas

Tabla 1 – Fórmulas de pesos locales.....	15
Tabla 2 – Fórmulas de pesos globales	16
Tabla 3 – Fórmulas de normalización.....	18
Tabla 4 – Recuento de las coincidencias de n variables binarias definidas para dos casos i y j con $n = a+b+c+d$	25
Tabla 5 – Espacio vectorial para el ejemplo de 5 gorjeos.....	46
Tabla 6 – Temas y gorjeos recuperados para el análisis.	56
Tabla 7 – Resultados de coeficiente de correlación cofenético para los términos tráfico, homicidios y asaltos.....	65