



**UNIVERSIDAD
DE GRANADA**

MINERÍA DE MEDIOS SOCIALES
MÁSTER CIENCIA DE DATOS E INGENIERÍA DE COMPUTADORES

MINERÍA DE TEXTO

EJERCICIO CON KNIME

Autor

Ignacio Vellido Expósito
ignaciove@correo.ugr.es



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

CURSO 2020-2021

1. Experimento

Para el ejercicio se ha usado un conjunto de tweets extraídos de la API gracias a la consulta **#Oscars2021 lang:en**. El resultado obtenido es un conjunto de 977 tweets sobre el cuál se calcula una tag-cloud, un clustering jerárquico y un conjunto de reglas difusas.

La lista de técnicas utilizadas en el preprocesamiento son:

1. Eliminación de URLs y símbolos especiales.
2. Eliminación de tweets excesivamente cortos.
3. Filtrado de sustantivos.
4. Stemming.

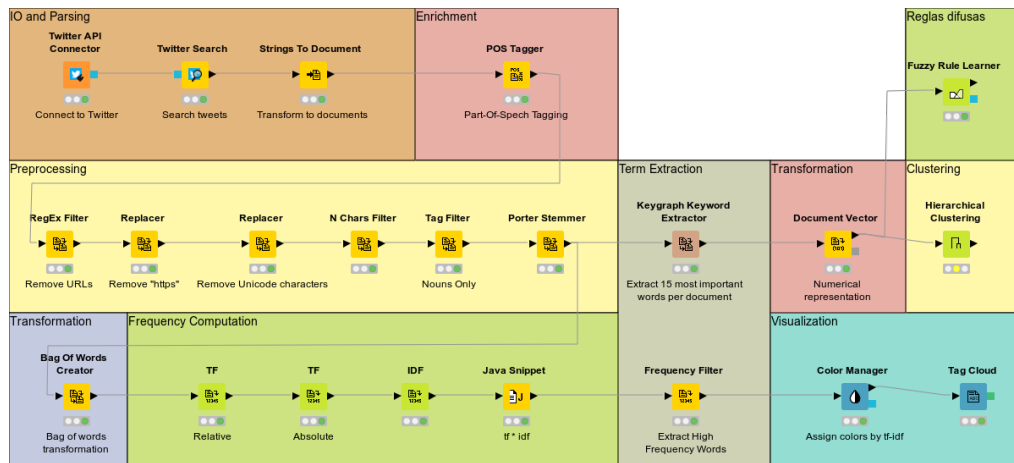


Figura 1: Workflow general del experimento.

#farmageddon
#soundofmetal
@miznulligan Road
Spielberg chicken
ABC
Work
@leveragetwo
consider

Figura 2: Tag cloud.

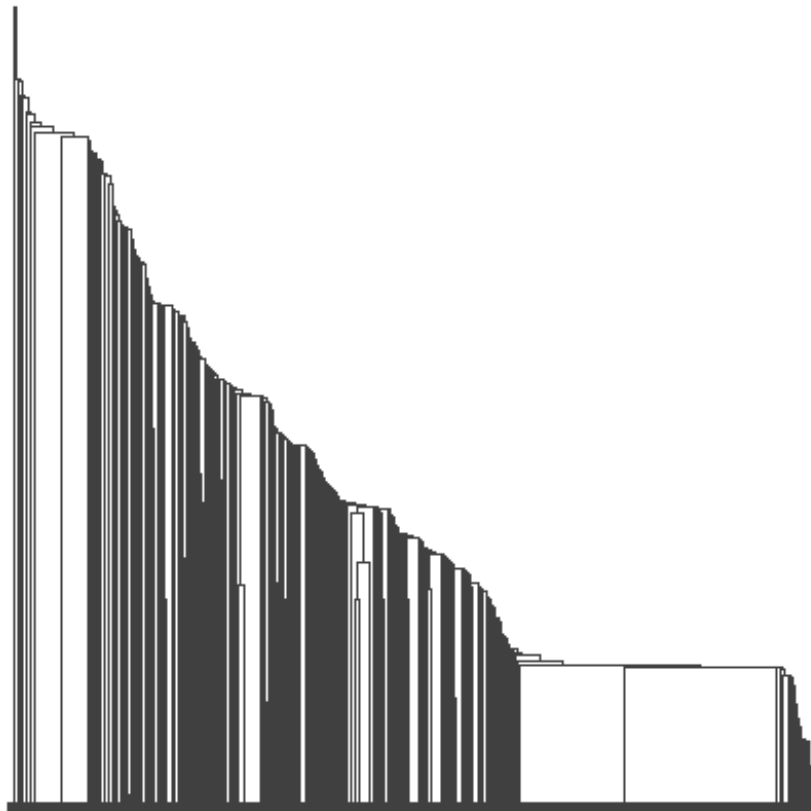


Figura 3: Dendrograma del análisis jerárquico.

Columns: 1013	Column Type	Column Index	Color Handler	Size Handler	Shape Han...	Filter Handler	Lower Bound	Upper Bound
Document	Text document	0					?	?
ultra	Number (dou...	1					0	7
#oscars	Number (dou...	2					0	6
dai	Number (dou...	3					0	22
#cdnscreenawards	Number (dou...	4					0	3
#thejunaoawards	Number (dou...	5					0	3
#sagawards	Number (dou...	6					0	6
#cinema	Number (dou...	7					0	3
film	Number (dou...	8					0	24
art	Number (dou...	9					0	13
#farmageddon	Number (dou...	10					0	3
Road	Number (dou...	11					0	3
Oscars2021	Number (dou...	12					0	21
#film	Number (dou...	13					0	21
#director	Number (dou...	14					0	13
#producer	Number (dou...	15					0	13
#peace	Number (dou...	16					0	7
#love	Number (dou...	17					0	7
#actor	Number (dou...	18					0	15
@SkyCinemaUK	Number (dou...	19					0	3
Brilliant	Number (dou...	20					0	1
#movie	Number (dou...	21					0	1
OliviaColman	Number (dou...	22					0	4
OscarNom	Number (dou...	23					0	17
#soundofmetal	Number (dou...	24					0	3
movi	Number (dou...	25					0	19
Sundai	Number (dou...	26					0	26
coverag	Number (dou...	27					0	25
morn	Number (dou...	28					0	16
Red	Number (dou...	29					0	16
Carpet	Number (dou...	30					0	16
freedom	Number (dou...	31					0	21

Figura 4: Algunos de los clusters obtenidos.

Row ID	[_] ultra	[_] #oscars	[_] dai	[_] #cdnsree...	[_] #thejunoa...	[_] #sagawards	[_] #cinema	[_] film	[_] art	[_] #farmaged...
Rule_1	<0.0,7.0,7.0,7....	<6.0,6.0,6.0,6....	<6.0,6.0,6.0,6....	<3.0,3.0,3.0,3....	<3.0,3.0,3.0,3....	<3.0,3.0,3.0,3....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....
Rule_2	<0.0,0.0,0.0,7....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....
Rule_3	<0.0,0.0,0.0,7....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....
Rule_4	<0.0,0.0,0.0,7....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....
Rule_5	<0.0,0.0,0.0,7....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....
Rule_6	<0.0,0.0,0.0,7....	<0.0,4.0,4.0,4....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....
Rule_7	<0.0,0.0,0.0,7....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....	<0.0,0.0,0.0,0....

Figura 5: Tabla de reglas.