

# Trabajo 13

## Analisis de Sentimientos

Facultad De Ingeniería, Universidad De Cuenca  
TEXT MINING

Freddy L. Abad L.

[freddy.abadl@ucuenca.edu.ec](mailto:freddy.abadl@ucuenca.edu.ec)

### PRIMERA PARTE

En la primera parte la meta es recopilar y analizar tweets utilizando el "Análisis de texto de AYLIEN" para RapidMiner. Para ejecutar este proceso es necesario tener una cuenta en AYLIEN.



Figura 1: Credenciales de AYLIEN necesarios para el analisis de sentimientos

Para crear una cuenta por favor acceda al siguiente enlace  
<https://developer.aylien.com/signup?source=rapidminer>

Las tareas a ejecutar son las siguientes

- Recoger tweets usando el operador de **búsqueda de Twitter**



Figura 1: Operador Search Twitter

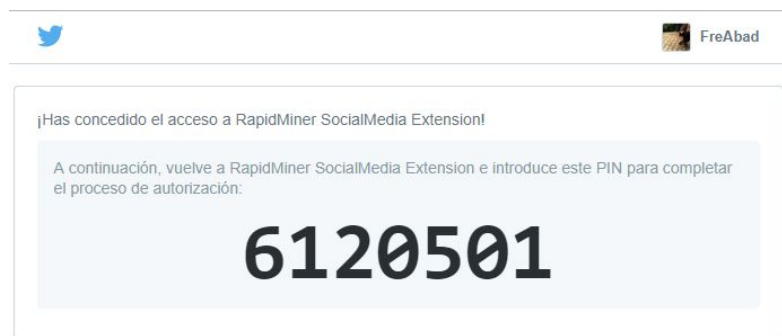
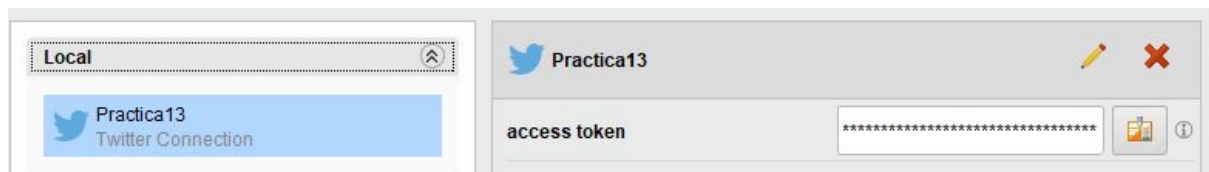


Figura 1: Credenciales de Twitter necesarias para conexion en rapidminer












*Figura 1: Conexion establecida con Twitter, es necesaria la validacion con una cuenta personal de TWITTER*

Analizar Sentimiento utilizando el Operador **Analizar Sentimiento**

- Asignar los tweets a diferentes categorías usando el operador de categorizar
- Visualizar los resultados y hacer los datos consumibles y comprensibles

**Paso 1.** Recopilación de tweets Usando el operador Search Twitter cree un proceso de búsqueda como lo haría con la API de búsqueda de Twitter. Usando los parámetros ejecute un proceso de búsqueda, por ejemplo buscando tweets que contengan la palabra clave "Samsung". Elimine algunas búsquedas quitando retweets (-rt) y enlaces (-http).

Para facilitar el proceso restrinja el número de tweets y coloque el idioma en el que desea recibir las respuestas. Recuerde además que el parámetro Result type permite devolver solo tweets recientes o populares.

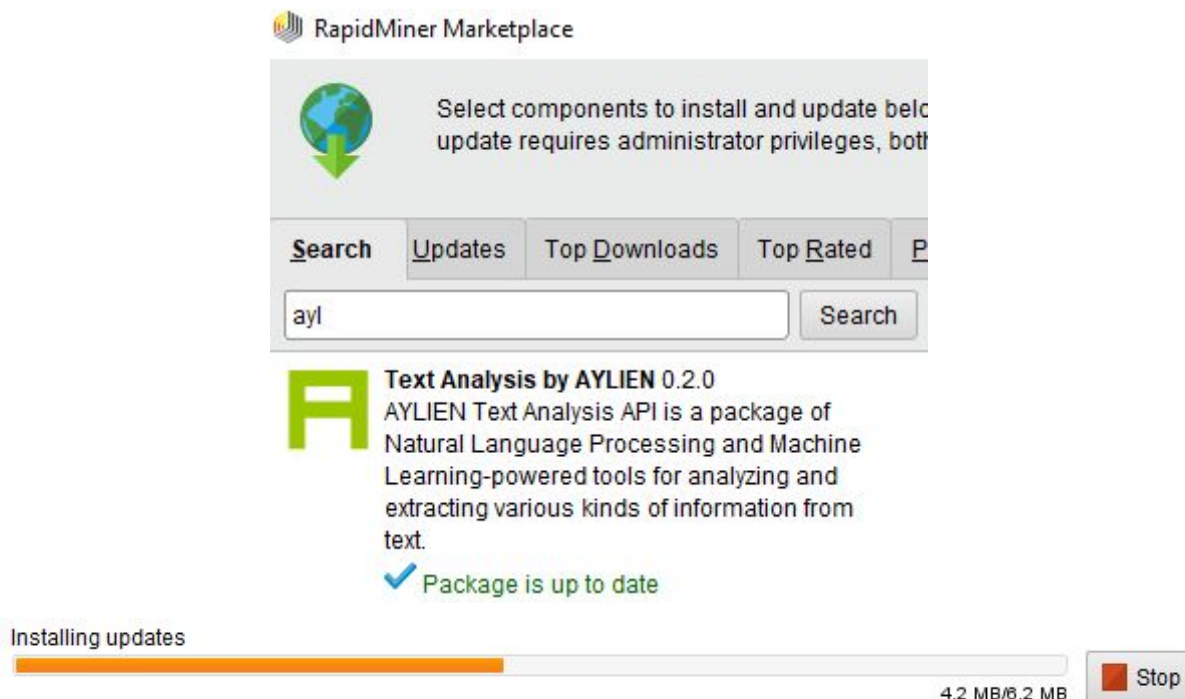
Search Twitter	
connection	Practica13  
query	#10YearsWithoutMichaelJackson -rt 
result type	recent or popular  
limit	100 
since id	<input type="text"/> 
max id	<input type="text"/> 
language	en 

*Figura 1: Configuración de Operador Search Twitter, es necesario eliminar los retweets (rt) y los enlaces (-https), y configurar el idioma según el ISO 639-1, en este caso escogemos **en** debido a que el análisis de sentimientos en Inglés tiene más soporte dentro de los desarrolladores.*

From-User	From-User-Id	To...	T...	Lang...	Source	Text
Genius	72073289	?	-1	en	<a ...	10 years ago today, the world lost the king of pop #10YearsWithoutMic...
Taj Jackson	6148462	?	-1	en	<a ...	Number 1 trending in the states 🇺🇸
Taj Jackson	6148462	?	-1	en	<a ...	Going to sleep...feeling blessed that my uncle has such amazing fans...
Geraldo Rivera	246500501	?	-1	en	<a ...	#10YearsWithoutMichaelJackson Call me crazy but I still think he was ...
Taj Jackson	6148462	?	-1	en	<a ...	#10yearswithoutMichaelJackson has been trending in the top 5 here i...
Complex	13049362	?	-1	en	<a ...	10 years ago today, we lost the King of Pop. 🕯️ #10YearsWithoutMicha...
La Toya Jack...	156818636	?	-1	en	<a ...	We Love You! We Miss You!
JayEss	28435382	?	-1	en	<a ...	Current Mood 🕯️ #HonorMJ #10YearsWithoutMichaelJackson https://t...

*Figura 1: Resultados del Operador Search Twitter sin nign operador adicional*

**Paso 2.** Analizando los tweets Con la colección de tweets almacenados en un ExampleSet el objetivo ahora es verificar el sentimiento de cada tweet, es decir, si son Positivos, Negativos o Neutros. Una opción para ejecutar este proceso es usar el operador **Analyze Sentiment Operator**. Los parámetros necesarios son seleccionar como atributo de entrada, la opción texto. Con esta configuración se tiene un proceso de análisis de opinión de Twitter relativamente simple que recopila tweets sobre el tema especificado y los analiza para determinar la polaridad (es decir, positiva, neutra o negativa) y la subjetividad (es decir, subjetiva u objetiva) de cada tweet. Si se ejecuta el proceso, entonces podemos visualizar el conjunto de datos que contiene los resultados no solo con los tweets que se extrajeron, sino también su polaridad y subjetividad correspondientes, así como un puntaje de confianza para ambos:




*Figura : Instalacion de TEXT ANALYSIS BY AYLIEN*



*Figura : Operador de Aylien: Analisis de Sentimientos*



*Figura : Nueva conexion con AYLIEN, necesaria para utilizar estos recursos*

connection	Practica13Aylien ▼	 ⓘ
input attribute	Text ▼	ⓘ
sentiment mode	tweet ▼	ⓘ
<input type="checkbox"/> Is input URL ⓘ		

*Figura : Configuracion de operador Analyze Sentiments, es necesario escoger la columna a analizar (TEXT), y el modo de sentimiento (TWEET)*

ExampleSet (Search Twitter) X								
Open in		Turbo Prep		Auto Model		Filter (15 / 15 examples): all		
Row No.	Id	polarity_con...	subjectivity_...	polarity	subjectivity	Created-At	From-User	From
1	1143418882...	0.539	1	neutral	subjective	Jun 25, 2019 ...	Taj Jackson	614
2	1143517605...	0.857	0.987	neutral	objective	Jun 25, 2019 ...	Genius	720
3	1143540282...	0.827	1	neutral	subjective	Jun 25, 2019 ...	Complex	130
4	1143539316...	0.987	1	positive	subjective	Jun 25, 2019 ...	La Toya Jack...	156
5	1143454706...	0.668	0.991	neutral	objective	Jun 25, 2019 ...	Taj Jackson	614
6	1143456663...	1.000	1.000	positive	subjective	Jun 25, 2019 ...	Taj Jackson	614
7	1143517249...	0.736	1	negative	subjective	Jun 25, 2019 ...	Geraldo Rivera	246
8	1143450771...	0.987	1	positive	subjective	Jun 25, 2019 ...	Taj Jackson	614
9	1143504836...	0.827	0.994	neutral	objective	Jun 25, 2019 ...	JayEss	284
10	1143425513...	0.796	1	negative	subjective	Jun 25, 2019 ...	REY	167
11	1143465077...	0.525	1	neutral	subjective	Jun 25, 2019 ...	Sudheer Babu	482
12	1143496138...	0.574	0.976	neutral	subjective	Jun 25, 2019 ...	MOBO	296
13	1143508939...	0.498	1.000	negative	subjective	Jun 25, 2019 ...	Elgin Charles	100
14	1143514962...	0.886	1.000	neutral	objective	Jun 25, 2019 ...	billboard	969

Figura : Resultado de proceso de analisis de sentimientos

## SEGUNDA PARTE

En la segunda parte se busca analizar el sentimiento en las reseñas en línea de la película Star Wars usando el método basado en léxico.

En este taller se describe dos técnicas de análisis de sentimientos disponibles en RapidMiner: basado en léxico y basado en modelos, que utiliza aprendizaje automático.

### Técnica basada en léxico

Para ejecutar el proceso de análisis de opinión a los Tweets relacionados con la película de Star Wars se recomienda ejecutar una búsqueda en la red social twitter

Averigüe el operador dentro de rapidminer que permite ejecutar búsquedas.

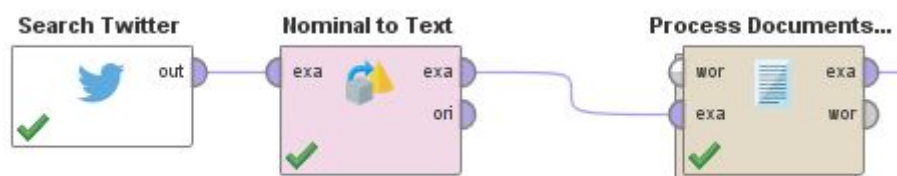
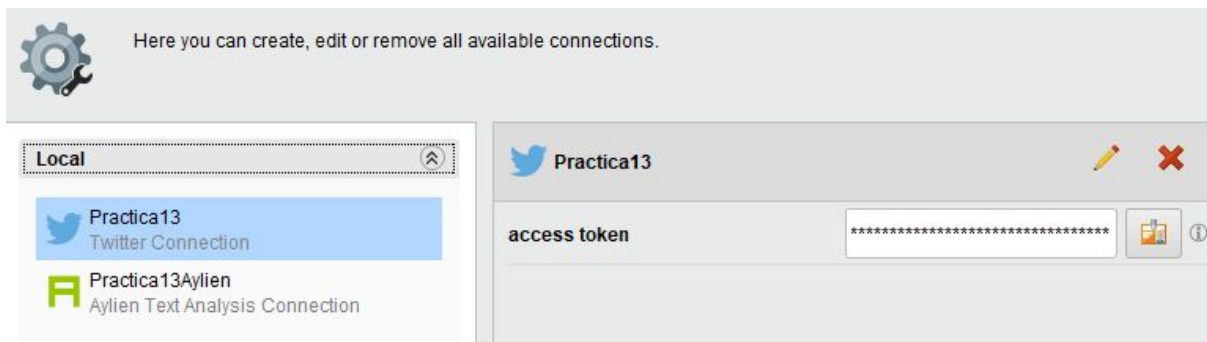


Figura: Operadores utilizados en la practica 2

Primero se debe configurar una conexión a Twitter.



*Figura : Conexion establecida con Twitter*

Busque los 1000 tweets más recientes o populares para el hashtag de consulta "#starwars". Luego de probar el funcionamiento de este operador seleccione el atributo "Text", que incluye el contenido de los tweets recuperados (es decir, ignoramos el resto de los metadatos).

*Figura : Configuración de Operador Search Twitter, es necesario variar el límite de tweets a analizar y su lenguaje*

Procese cada uno de los documentos (tweets) usando alguna representación como TF-IDF para la creación de un espacio vectorial.



*Figura : Operador Process Documents from Data*



**Process Documents from Data**

- ☒ create word vector
- vector creation ☒ TF-IDF
- ☒ add meta information
- ☐ keep text ☒
- prune method ☒ none
- data management auto
- ☐ select attributes and weights

Figura : Configuración de Process Documents from Data

Previo a ello se recomienda ejecutar algunas tareas de preprocesamiento como: **Tokenización** para extraer las palabras individuales de cada tweet y, así, representar un tweet con una "bolsa de palabras". También se debería **filtrar las palabras vacías (stop words)**.

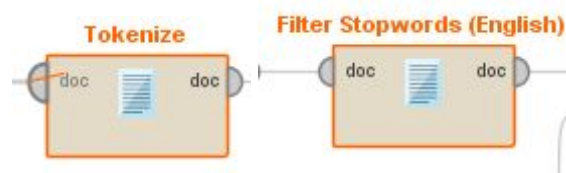


Figura : Operador Tokenize (Bolsa de Palabras) y Filter Stopwords (palabras vacías)

Se pide investigar la forma de extraer la raíz disponible de una palabra usando la extensión **Wordnet**, que encuentra el lema de cada palabra.

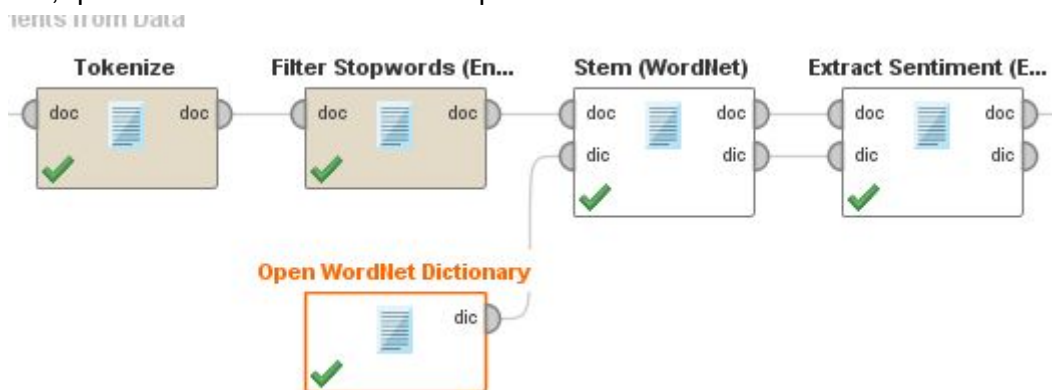


Figura : Operadores del subproceso Process Documents from Data



Figura : Operador Open WORDNET, es necesario descargarse de Internet el Diccionario Wordnet, y descomprimirlo

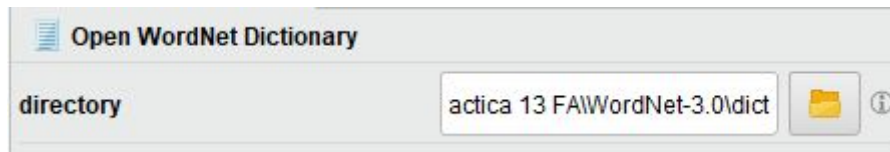


Figura : Configuración del PATH hacia el diccionario de OPENWORDNET, es necesario que se ubique dentro de la carpeta /dic/



Figura : Operador Stem(WORDNET) que devuelve la raíz de las palabras según el Diccionario de Open wordnet

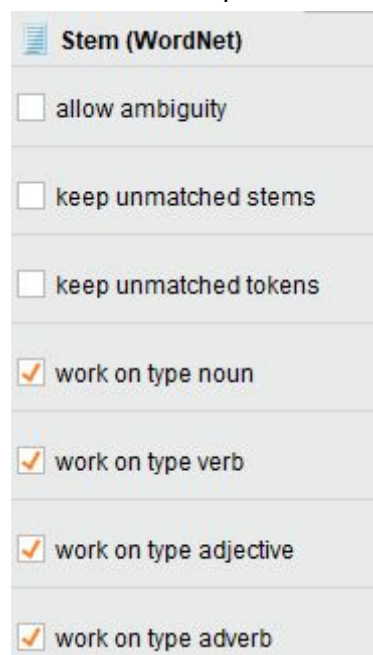


Figura : Configuración del Operador Stem (Wornet), utilizando sustantivos, verbos, adjetivos, adverbios

Finalmente, se solicita aplicar el operador "**Extract Sentiment**" (nuevamente desde la extensión de **Wordnet**). Se recomienda usar todas las entidades disponibles de voz parcial, como sustantivos, verbos, adjetivos y adverbios. Como tarea adicional de esta práctica se pide que la salida del proceso se pueda representar gráficamente en RapidMiner con un histograma de los valores de sentimiento calculados (atributo "sentimiento"). Un ejemplo de resultado que se espera se muestra a continuación.





Figura : Operador Extract Sentiment (English)

Figura : Configuración de Extract Sentiment (English)

El **histograma** ofrece ideas interesantes; por ejemplo, podemos detectar en la parte derecha de la distribución la existencia de un subgrupo de usuarios de Twitter que expresa un sentimiento positivo relativamente alto, que probablemente serán fieles seguidores de la serie Star Wars.

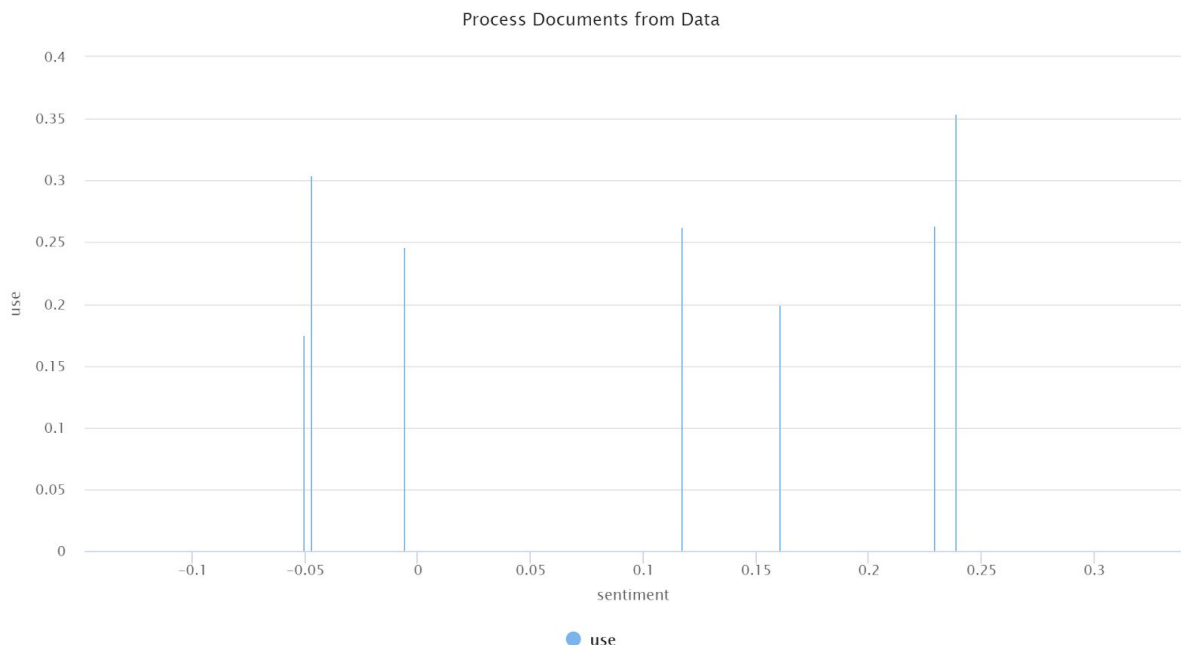


Figura : Histograma USE vs Sentimiento