

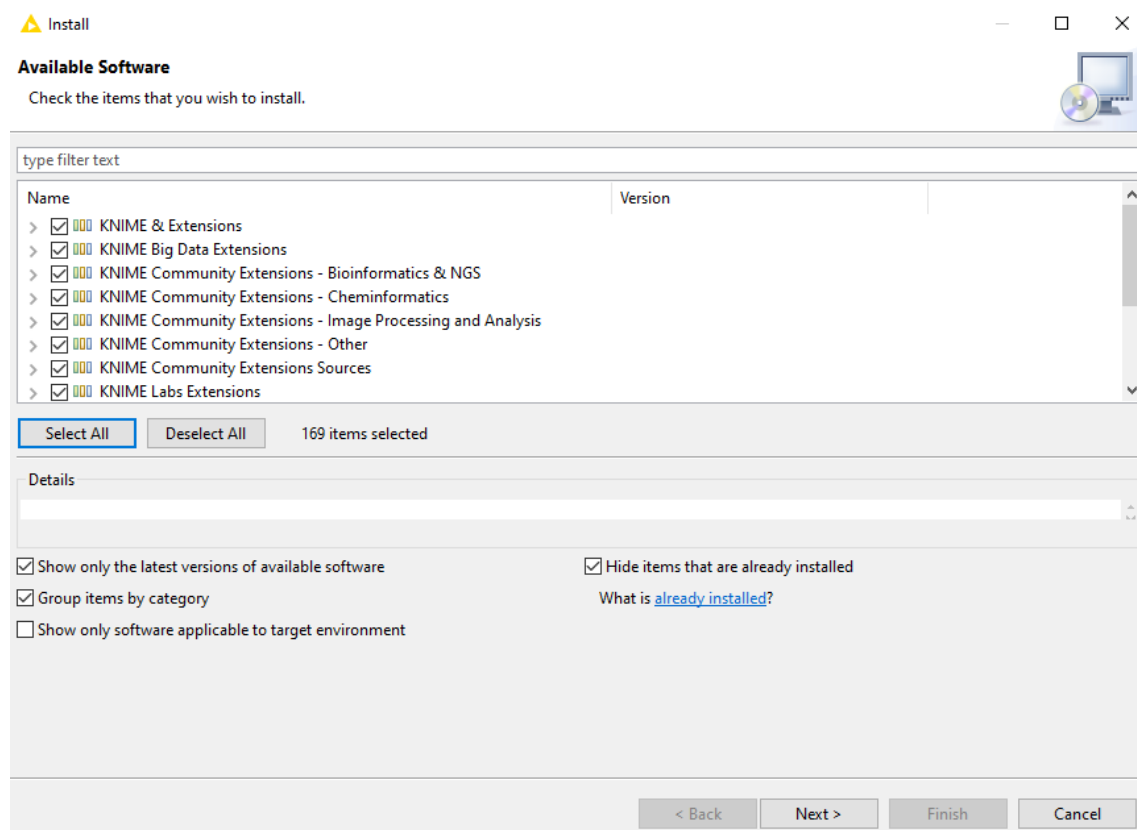
## 23 Junio 2020. Text Mining. Knime Parte 2

Construya un tutorial en que detalle el desarrollo de la tarea.

1. File Install Knime Extensions
2. Cargar el ejemplo Social Media
3. Ejecutar cada uno de los nodos, uno a uno e ir pasando al siguiente nodo después de verificar que confirmó la ejecución exitosa de cada nodo.
4. ¿Con qué datos se trabaja?
5. ¿Porqué se consideran módulos?
6. ¿Qué tareas (nodos) comprende el módulo posterior a cargar los datos?
7. ¿Cuáles son los principales descriptores de cada tópico?

1.

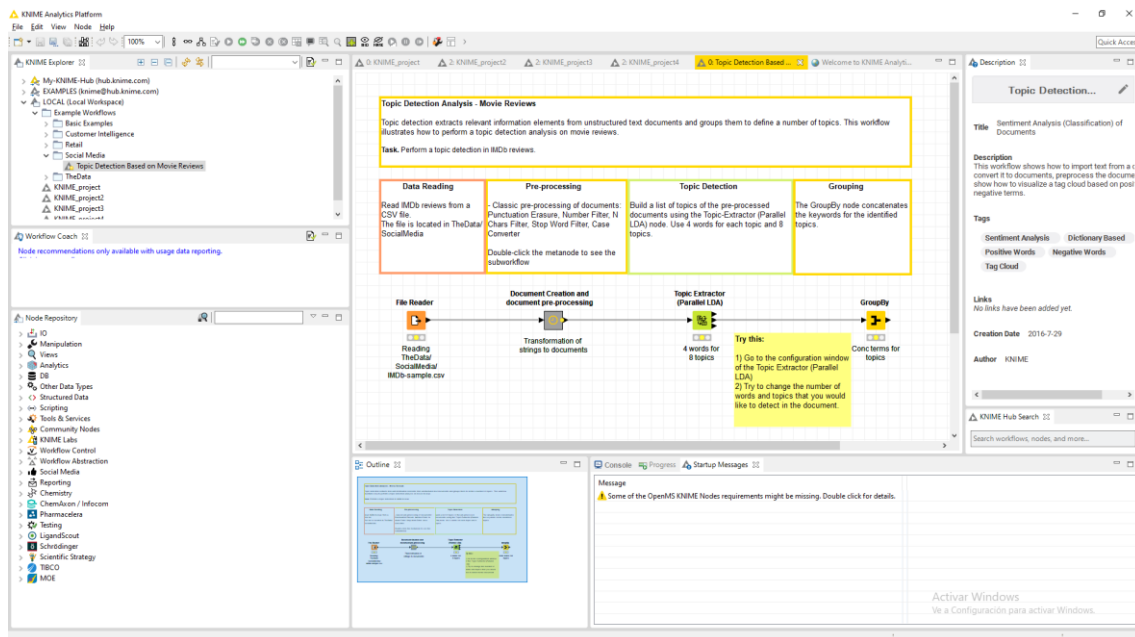
Abrimo Knime. En File→Install Knime Extension. En la siguiente ventana seleccionamos todas las extensiones y continuamos la instalación con normalidad.



2.

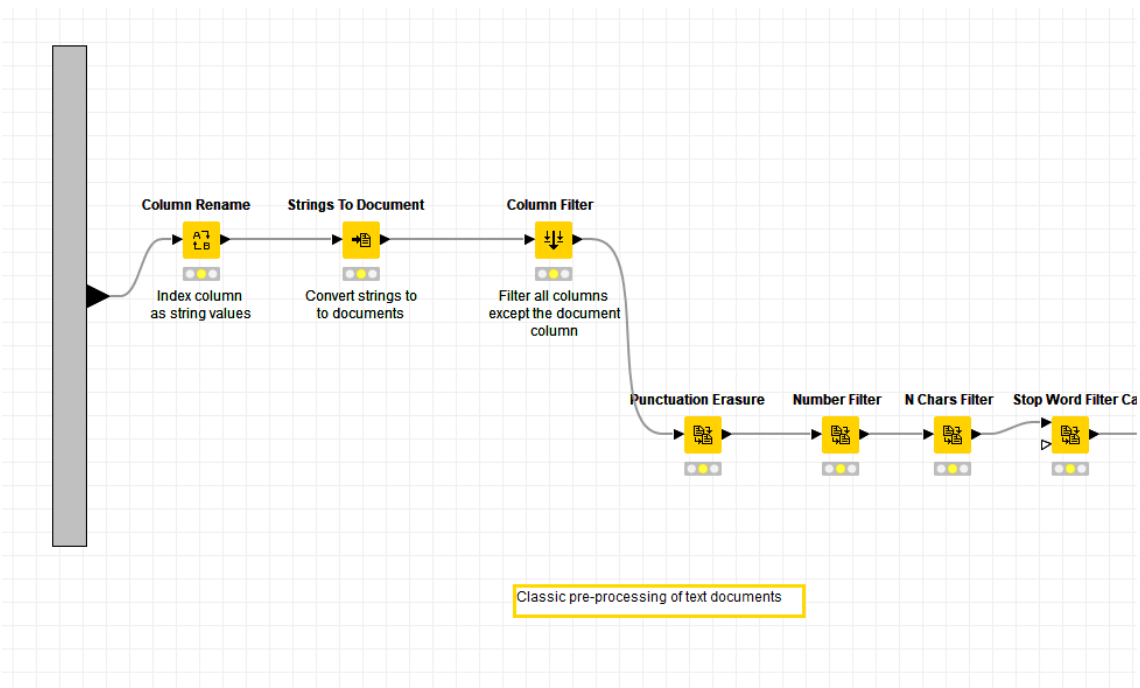
Buscamos en Knime Explorer→social Media→Topic Detection Based on movie reviews, y cargamos el ejemplo.

Este ejemplo contiene críticas a películas. Cuando abrimos el ejemplo que vienen en Knime, nos aparecerá esto:



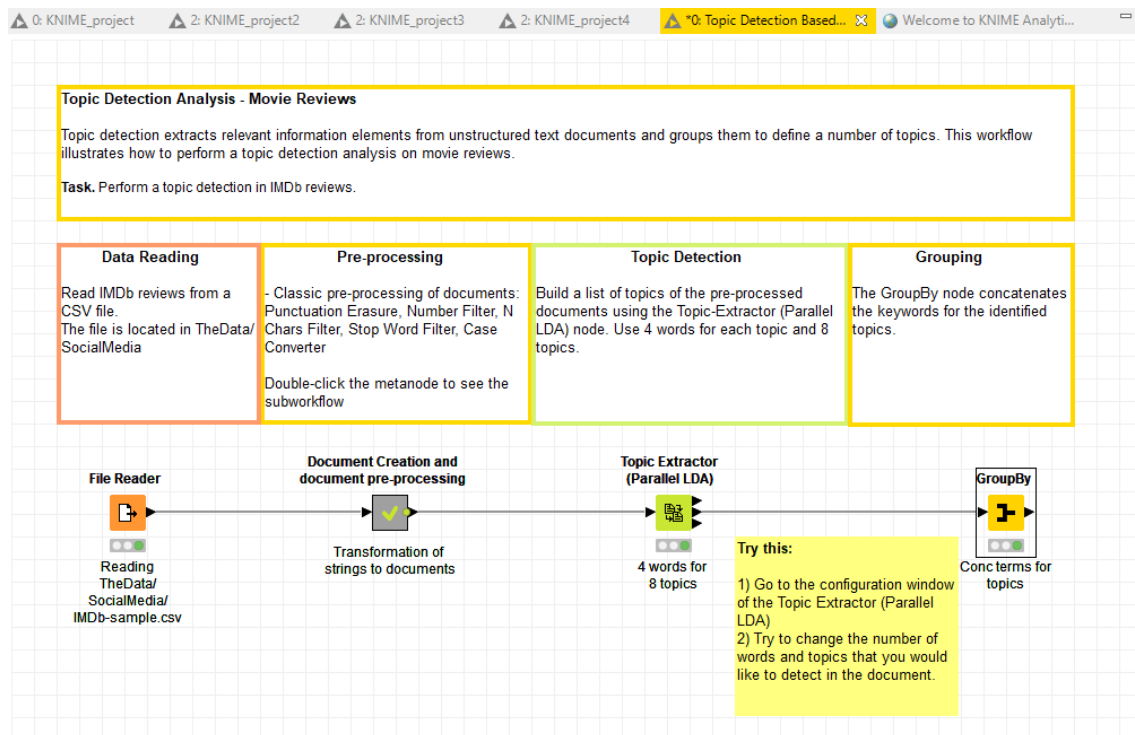
Aquí podemos ver como están conectados los nodos y los módulos de nodos entre ellos, así como una breve explicación de qué funcionalidad tiene cada uno. Si entramos a abrir el nodo file reader, vemos que los datos con los que estamos trabajando son textos completos.

Si entramos en el segundo módulo, vemos que este está compuesto internamente de varios nodos.



### 3.

Vamos a ir seleccionando uno a uno los nodos, y pulsando F7 para ejecutarlos. La aparición de un semáforo en verde o de un tick, nos confirmará la correcta ejecución de los nodos.



4.

Estamos trabajando con textos, tal y como comentamos anteriormente. Son opiniones de películas.

5.

Se usan esos módulos para encapsular muchos nodos en uno solo. De esta manera ocupan menos espacio en el workflow.

6.

En este módulo, se preprocesan los datos para que se puedan crear modelos a partir de ellos. Para preprocesar los datos se usarán nodos que eliminen los signos de puntuación, filtre los números, palabras “vacías” que no representan nada útil, y se convierten todas las letras a minúsculas o mayúsculas.

7.

Para saber que palabras describen cada tópico, nos vamos al último de los nodos y hacemos click con el botón derecho. Entre las opciones habrá una denominada Group table. La seleccionamos y nos aparecerá una ventana con los descriptores de cada tópico:

Row ID	S Topic id	S Concatenate(Term)
Row0	topic_0	his, and, best, who
Row1	topic_1	horror, killer, her, effects
Row2	topic_2	the, and, this, that
Row3	topic_3	the, was, were, war
Row4	topic_4	the, and, that, film
Row5	topic_5	her, she, with, and
Row6	topic_6	the, and, show, for
Row7	topic_7	the, his, and, who

Sin embargo, no es difícil observar que en alguna parte del proceso ha fallado algo, ya que aparecen entre los descriptores palabras que consideramos vacías, pues no aportan ninguna definición clara, tal como “the, his, she, and...”.

