

## Data Reading

Read IMDb reviews from a CSV file.  
The file is located in TheData/ SocialMedia

## Pre-processing

- Classic pre-processing of documents:  
Punctuation Erasure, Number Filter, N  
Chars Filter, Stop Word Filter, Case  
Converter  
Double-click the metanode to see the  
subworkflow

## Topic Detection

Build a list of topics of the pre-processed  
documents using the Topic-Extractor (Parallel  
LDA) node. Use 4 words for each topic and 8  
topics.

## Grouping

The GroupBy node concatenates  
the keywords for the identified  
topics.

### CSV Reader



Reading  
TheData/  
SocialMedia/  
IMDb-sample.csv

### Document Creation and document pre-processing



Transformation of  
strings to documents

### MISsing Topic Extractor (Parallel LDA)



4 words for  
8 topics

### Try this:

- 1) Go to the configuration window of the Topic Extractor (Parallel LDA)
- 2) Try to change the number of words and topics that you would like to detect in the document.

### GroupBy



Concat terms for  
topics