Abstract

Text mining, at times alluded to as content information mining, is harshly equal to content investigation, which alludes to the procedure of determining astounding data from content. RapidMiner is unquestionably the world-leading open-source system for data mining. It is available as a stand-alone application for data analysis and as a data mining engine for the integration into own products. Tokenization is the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called tokens. The word frequency counter allows you to count the frequency usage of each word in your document. Applying tokenization and word frequency counter for a text document (resume in this case) helps us find
out occurrence of each word in a document but there is no provision to find a particular word frequency occurrence according to user choice.

References


Index Terms

Computer Science Speech Processing

Keywords

Rapidminer rapidminer Text Processing Rapidminer Process Document From File
Rapidminer Transform Case
Rapidminer Tokenize
Operator
Operator
Operator.