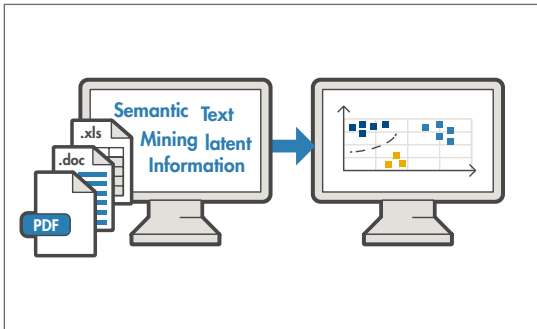# Get Started with Text Analytics Toolbox

Text Analytics Toolbox™ provides algorithms and visualizations for preprocessing, analyzing, and modeling text data. Models created with the toolbox can be used in applications such as sentiment analysis, predictive maintenance, and topic modeling.

Learn more at: *mathworks.com/products/text-analytics*

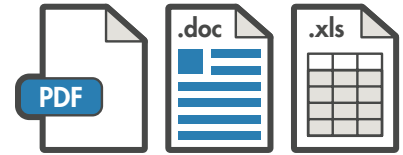| Function Name	Description
wordcloud	Create word cloud chart from bag-of-words or LDA model
wordCloudCounts	Count words for word cloud creation
textscatter	2-D scatter plot of text
textscatter3	3-D scatter plot of text
heatmap	Create heatmap chart
histcounts	Histogram bin counts
discretize	Group data into bins or categories

Visualize

Use word clouds and text scatter plots to summarize and validate results.

Model and Predict

Convert text into numeric representations using bag-of-words or pretrained word embedding models, and apply specialized machine learning algorithms for prediction and topic modeling.

Function Name	Description
readWordEmbedding	Read word embedding from text file
trainWordEmbedding	Train word embedding
word2vec/vec2word	Maps words to embedding vectors
ldaModel	Latent Dirichlet allocation (LDA) model
lsaModel	Latent semantic analysis (LSA) model
bagOfWords	Bag-of-words model
fitlda	Fit latent Dirichlet allocation (LDA) model
fitlsa	Fit a latent semantic analysis (LSA) model
predict	Predict top LDA topics of documents
fitdist	Fit probability distribution object to data
fitrlinear	Fit linear regression model to high-dimensional data
fitclinear	Fit linear classification model to high-dimensional data
fitcecoc	Fit multiclass models for classifiers

Function Name	Description
`extractFileText`	Read from PDF, Microsoft Word, and plain text
`textscan`	Read formatted data from text file or string
`readtable`	Create table from file
`compose`	Convert data into formatted string array
`xlsread`	Read Microsoft Excel spreadsheet file
`webread`	Read content from RESTful web service
`TabularTextDatastore`	Datastore for tabular text files
`FileDatastore`	Datastore with custom file reader
`SpreadsheetDatastore`	Datastore for spreadsheet files

Import

Extract text from Microsoft® Word® files, PDFs, text files, and spreadsheets.

~~Performed~~ preventive maintenance serviceing ~~on a~~ broken pump.

Preprocess

Remove less helpful artifacts such as common words, punctuation, and URLs and apply text normalization to stem words to their root word.

Function Name	Description
`tokenizedDocument`	Split documents into collections of words
`normalizeWords`	Remove inflections from words using the Porter stemmer
`bagOfWords`	Bag-of-words model
`stopWords`	Stop word list
`context`	Search documents for word occurrences in context
`removeWords`	Remove selected words from document or bag-of-words
`removeLongWords`	Remove long words from documents or bag-of-words
`removeShortWords`	Remove short words from documents or bag-of-words
`removeInfrequentWords`	Remove words with low counts from bag-of-words model
`erasePunctuation`	Erase punctuation from text and documents

Function Name	Description
`str = "Hello,world"`	Declare a string variable
`str = ["Hello", "World"]`	Declare a string array
`str = string( C )`	Convert a character vector C to a string
`str2double`	Convert a string to double numbers
`strlength`	Return the length of strings
`isstring`	Determine if input is string array
`join`	Combine strings
`split`	Split strings in string array
`splitlines`	Split string at newline characters
`replace`	Find and replace substrings in string array
`contains`	Determine if pattern is in string
`erase`	Delete substrings within strings
`extractBetween`	Extract substrings between indicators
`extractAfter`	Extract substring after specified position
`extractBefore`	Extract substring before specified position
`strcmp`	Compare strings
`regexp`	Match regular expression (case sensitive)

`"Hello,world"`

String

Manipulate, compare, and store text data efficiently.