LEVEL: Intermediate (3)
DURATION: 1/2 Day

Prerequisites

- A good working knowledge of R programming is assumed, attendees should have complete Mango's Programming in R for Analysts or similar
- Familiarity with basic analytic techniques and linear modelling is required.

Overview

Many companies have a large amount of data stored as text that is not being used effectively. In this introductory workshop we will show how you can get started with analysing text data, from simple manipulation through to sentiment analysis. By the end of the course attendees will have a good understanding of the techniques as well as how to implement them in R.

Teaching Approach

This will be a hands-on workshop taught using the RStudio IDE with exercises throughout. All attendees will need access to a computer and will need to have pre-installed a recent version of R and RStudio and will need to be able to install R packages. The course will be taught by Mango Solutions consultants.

Details

Simple Text Manipulation

- Regular expressions
- Tidy text format
- Tokenisation and n-grams
- Removing stop words and stemming

Word and Document Frequency

- Term Frequency
- Inverse Document Frequency

Sentiment Analysis

- Sentiment Lexicons
- Joining sentiments to documents

Trainers

Duncan Leng and Nicolas Attalides, Mango Solutions