Natural Language Processing for sensory and consumer scientists. A tidy introduction in R.

Instructors: Ruben Rama, Symrise Ltd, UK

Katja Tiitinen, Symrise, Germany

This tutorial offers a step-by-step guide to help sensory and consumer scientists start their journey on Natural Language Processing (NLP) analysis using R, a free software environment for statistical computing and graphics. Natural Language Processing (NLP) is a field of Artificial Intelligence that makes human language intelligible to machines, combining the power of linguistics and computer science to study the rules and structure of language, and create intelligent systems capable of understanding, analyzing, and extracting meaning from text.

Text has quickly become a key tool for sensory and consumer scientists to understand several aspects of consumer behavior, including open ended questions, customer online reviews, consumer opinions, social media, etc.

During this workshop, participants will get hands-on experience coding in R, gaining familiarity with tidy techniques and packages for NLP. This tutorial is not intended as a statistics course and participants will not need to have a deep technical understanding of NLP, or of specific analyses, but by the end participants will be able to conduct Sentiment Analysis and graph the results on consumer responses. This method will allow participants to pre-process, analyze and visually represent text data from different sources, such as surveys, MROCs, websites, Twitter, online reviews, etc.

In this workshop we will introduce the audience to the R statistical programming environment and the RStudio Interactive Development Environment (IDE) with the aim of developing sufficient basic skills to conduct NLP analyses on sensory and consumer text sets, including:

- the creation of a corpus or tidytext alternative,
- text pre-processing using tidy principles (removal of punctuation, numbers, word case and stopwords), and
- performing sentiment analysis.

A suitable dataset for learning and practicing will be provided, together with an R script covering the full workflow.

By the end of the tutorial, participants will have gained familiarity with basic Natural Language Processing techniques, have a comprehensive understanding of the R programming environment and its major packages for NLP, and acquire the necessary skills to analyze and visualize basic text.

Duration	3 hours
Audience	Sensory and consumer scientists interested in using Natural Language Processing (NLP) tools in their research activities.
Background:	No statistics knowledge, but useful to be comfortable with R/RStudio
Laptop:	Participants are encouraged to bring their personal computers with an up-to-date version of R/RStudio installed.