Preliminary Analyses RRI Tool Kit

Juan C. Correa

# Intro

This is a short preliminary analysis of the contents included in the sample of projects indexed by the *Responsible Research and Innovation* (RRI) toolkit. By employing web scraping techniques, we were able to retrieve a total of 1,414 documents from RRI, along with a brief document description, their titles, and the discipline they belong to.

As the initial data base contained several disciplines with quite similar categories (e.g., agrobiotechnology, agriculture, plant biology, smart farming), we decided to merge these categories under a single name. The resulting merged categories were re-named as follows:

* Agro = agribiotechnology + agriculture + plant biology + smart farming + Food + Food, agriculture, forestry, water and bio-economy + plant biology + rural environment
* Business & Industry
* Civil Society & Organization
* Climate = climate action + climate adaptation strategies + climate crisis + Climate, environment, and resources
* Energy = Energy + energy transition + renewable energy
* Ethics = ai ethics + Ethics + ethics of ai + responsible technology + security
* Health = health care + health technology + Health, demography & wellbeing + patient engagement + vaccine hesitancy + vaccines
* Technologies = Artificial intelligence + Digital technologies + emerging technologies + robotics
* Education = Education Community + Science education
* Govern = Governance + policy design + public engagement

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

Now, we have a clean data set to use for further analyses.

Disciplines <- Disciplines[c(3, 6,9,12,15,18,21,24,27, 30, 33)]  
variable.names(Disciplines)

## [1] "agro" "busi" "civsoc" "climate" "education"   
## [6] "energy" "ethics" "govern" "health" "policy"   
## [11] "technologies"

We start by estimating the mutual information between categorized documents as a first general analysis

library(infotheo)  
DIM <- mutinformation(Disciplines, method = "mm")  
library(corrplot)

## corrplot 0.89 loaded

library(RColorBrewer)  
whiteblack <- colorRampPalette(c("#FFFFFF", "#E5E5E5", "#F3F3F3", "#D8D8D8", "#B1B1B1", "#8A8A8A", "#5D5D5D", "#000000"))  
corrplot(DIM, is.corr = FALSE, cl.lim=c(min(DIM), max(DIM)), method = "number", type = "lower", col=whiteblack(6), tl.col = "black")

