testing

Vorisek\_CN

20 4 2022

First, load relevant packages for the network analysis:

library(tidyverse)  
library(igraph)  
library(ggraph)  
library(visNetwork)  
library(readxl)

# Analysis of Excel file

## Data import and cleaning

Import Excel file with included studies (“IncludedStudies2.xlsx”):

excel\_data <- read\_excel("IncludedStudies2.xlsx")

Clean data (keep only variables relevant for analysis, rename variables, reformat, …):

excel\_data\_clean <- excel\_data %>%   
 select(Author = `Author (First)`,  
 Title,  
 Year,  
 Country = `Country (according to 1st author affiliation)`,  
 Item\_mapped\_keyword = `Item mapped\_keyword`,  
 Goal\_keyword,  
 Other\_standards\_keyword = `Other standards\_keyword`,  
 Other\_software\_keyword = `Other Software\_keyword`,  
 FHIR\_resource\_used = `FHIR Resource used`,  
 FHIR\_extension\_used = `FHIR extension used`,  
 FHIR\_version = `FHIR version`,  
 Patients,  
 Variables,  
 Variables\_count,  
 Research\_category = `Research Category`,  
 Medical\_area = `Medical Area`,  
 Journal,  
 Impact\_factor = `Impact Factor Research)`,  
 Research\_area = `Research Area`,  
 Clinical\_trials = `Clinical Trials`,  
 generic,  
 SNOMED\_CT = `Other standards - SNOMED CT`,  
 LOINC = `Other standards - LOINC`,  
 ICD\_10 = `Other standards - ICD 10`,  
 OMOP = `Other standards - OMOP`,  
 Other = `Other standards - Other`,  
 None = `Other standards - None (nicht angegeben oder nicht verwendet)`) %>%   
 mutate(Journal = str\_trim(Journal)) %>%   
 mutate(Journal = ifelse(Journal %in% c("Studies in Health technology and informatics", "Studies in Health Technology and Informatics"),  
 "Studies in health technology and informatics", Journal)) %>%   
 mutate(Journal = ifelse(Journal == "JMIR medical informatics",  
 "JMIR Medical Informatics", Journal)) %>%  
 mutate(Impact\_factor = ifelse(Impact\_factor %in% c("None", "-"), NA, Impact\_factor)) %>%   
 mutate(Impact\_factor = str\_replace(Impact\_factor, ",", ".")) %>%   
 mutate(Impact\_factor = round(as.numeric(Impact\_factor), 2)) %>%   
 mutate(Country = ifelse(Country == "USA / Switzerland", "USA", Country)) %>%   
 mutate(Research\_area = ifelse(Research\_area == "Clinical trials",  
 "Clinical Trials", Research\_area))

### Goals

This section analyzes the most common keywords describing the goals of the articles.

Load tidytext package and get word frequencies of all words in Goal\_keyword variable:

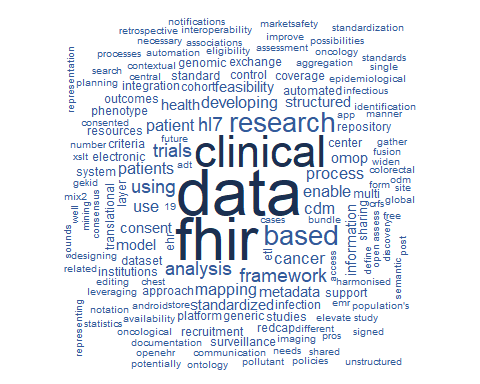
library(tidytext)  
  
# define some stop words to be excluded  
stop\_words <- c("of", "and", "for", "from", "a", "an", "between",  
 "across", "on", "the", "to", "in", "into", "with",  
 "as", "or", "other", "such", "via")  
  
goal\_keywords <- excel\_data\_clean %>%   
 select(Goal\_keyword) %>%   
 unnest\_tokens(word, Goal\_keyword, to\_lower = TRUE) %>%   
 filter(!(word %in% stop\_words)) %>%   
 group\_by(word) %>%   
 summarize(freq = n()) %>%   
 arrange(desc(freq))

The wordcloud shows the most common words mentioned as keywords:

library(wordcloud)

## Loading required package: RColorBrewer

wordcloud(words = goal\_keywords$word, freq = goal\_keywords$freq, min.freq = 1, max.words=150, random.order=FALSE, rot.per=0.10,   
 colors=c("#335a99", "#1b3051"))



Note that there are still some typos shown in the wordcloud. (Correct all typos in the final Excel file before analysis.)