Research reproducibility in football coach intervention: an approach for Systematic Literature Review

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. With the cleaning of a dataframe used in a Systematic Review. <https://cran.r-project.org/web/packages/bibtex/bibtex.pdf>

This was tested with bib2df version 1.1.1, installation instructions available [here](https://docs.ropensci.org/bib2df/), and one example [here](https://docs.ropensci.org/bib2df/articles/bib2df.html). Github project is available [here](https://github.com/ropensci/bib2df)

Is better to confirm that the files are in UTF-8 encoding. The bibtex files must be checked. There should be spaces between “=”:

@Article{Binmore2008,  
 Title = {Do Conventions Need to Be Common Knowledge?},  
 Author = {Binmore, Ken},  
 Journal = {Topoi},  
 Year = {2008},  
 Number = {1},  
 Pages = {17--27},  
 Volume = {27}  
}

# Import Libraries

# Setup library  
library(bib2df)  
library(bibtex)  
library(tibble)  
library(dplyr)  
library(lattice)

# get rmarkdown directory  
caminho<-getwd()  
# set working directory  
setwd(caminho)

This code correct that:

tx<-readLines(file)  
tx2<-gsub(pattern = "=\\{",replacement = " = {",x = tx)  
writeLines(text = tx2,con = 'corrected\_scopus.bib')

files<-list.files(path = '../articles/',pattern = '\\.bib$')  
files

## [1] "corrected\_scopus.bib" "data\_articles.bib" "scopus.bib"

# Reading bibtex source files

## ACM

# IEEE

# Science Direct

Incorrect data discarded 8 incomplete information tot

There are 6 entries without abstract.

The entries aren’t articles, we will remove them.

# Scopus

tx<-readLines('../articles/scopus.bib')  
tx2<-gsub(pattern = "=\\{",replacement = " = {",x = tx)  
writeLines(text = tx2,con = '../articles/corrected\_scopus.bib')  
  
  
scopus <- bib2df('../articles/corrected\_scopus.bib')  
dim(scopus)

## [1] 143 45

names(scopus)

## [1] "CATEGORY" "BIBTEXKEY"   
## [3] "ADDRESS" "ANNOTE"   
## [5] "AUTHOR" "BOOKTITLE"   
## [7] "CHAPTER" "CROSSREF"   
## [9] "EDITION" "EDITOR"   
## [11] "HOWPUBLISHED" "INSTITUTION"   
## [13] "JOURNAL" "KEY"   
## [15] "MONTH" "NOTE"   
## [17] "NUMBER" "ORGANIZATION"   
## [19] "PAGES" "PUBLISHER"   
## [21] "SCHOOL" "SERIES"   
## [23] "TITLE" "TYPE"   
## [25] "VOLUME" "YEAR"   
## [27] "DOI" "URL"   
## [29] "AFFILIATION" "ABSTRACT"   
## [31] "AUTHOR\_KEYWORDS" "CORRESPONDENCE\_ADDRESS1"  
## [33] "ISSN" "LANGUAGE"   
## [35] "ABBREV\_SOURCE\_TITLE" "DOCUMENT\_TYPE"   
## [37] "SOURCE" "ART\_NUMBER"   
## [39] "KEYWORDS" "PUBMED\_ID"   
## [41] "CODEN" "O"   
## [43] "EL" "METHODS."   
## [45] "RESULTS."

This code gets the abstract using the paper link We are going to use (rvest)[<http://rvest.tidyverse.org/>] The xpath was identified using Firefox: 1. Inspect in the abstract 2. Right click > Copy > XPath:

/html/body/div[1]/div[1]/div[1]/div[2]/div[1]/div[3]/div[3]/div[1]/div[1]/div[2]/div[2]/section[7]/p

# Springer

The bibtex file extracted from Springer don’t have abstract. We have to recreate that. We will use the same approach that was done to Scopus. Get the node with abstract using the css selector .c-article-section\_\_content or .Para. An easy way is to identify the css selector is to use the inspector.

Checking springer articles without abstract

# Web of Knowledge

# Create a dataframe for analysis

scopus2<-data.table::copy(scopus)  
data\_articles <- data.frame(scopus$CATEGORY,scopus$YEAR,scopus$TITLE,scopus$ABSTRACT,scopus$DOI,scopus$JOURNAL,scopus$BIBTEXKEY,  
 sapply(scopus$AUTHOR,FUN=toString))  
  
names(data\_articles)<- c('category','year','title','abstract','doi','journal','bibtexkey','author')  
data\_articles$source <- 'SCOPUS'  
  
  
#Transform the vector to string  
scopus$AUTHOR <- sapply(scopus$AUTHOR,FUN=toString)  
#ieee$AUTHOR <- sapply(ieee$AUTHOR,FUN=toString)  
#science$AUTHOR <- sapply(science$AUTHOR,FUN=toString)  
#springer$AUTHOR <- sapply(springer$AUTHOR,FUN=toString)  
#wos$AUTHOR <- sapply(wos$AUTHOR,FUN=toString)  
  
#SCOPUS  
if(!exists('scopus2')){  
 scopus2<-data.table::copy(scopus)   
 scopus<-scopus2 %>% select(CATEGORY,YEAR,TITLE,ABSTRACT,DOI,JOURNAL,BIBTEXKEY,AUTHOR )  
 names(scopus)<-c('category','year','title','abstract','doi','journal','bibtexkey','author')  
 scopus$source <- 'SCOPUS'  
   
 data\_articles<-rbind(data\_articles,scopus)  
}  
  
#WOS  
#if(!exists('wos2')){  
# wos2<-data.table::copy(wos) #Copy the dataset  
# wos<-wos2 %>% select(CATEGORY,YEAR,TITLE,ABSTRACT,DOI,JOURNAL,BIBTEXKEY,AUTHOR )  
# names(wos)<-c('category','year','title','abstract','doi','journal','bibtexkey','author')  
# wos$source <- 'WOS'  
#  
# data\_articles<-rbind(data\_articles,wos)  
#}  
  
str(data\_articles)

## 'data.frame': 143 obs. of 9 variables:  
## $ category : chr "ARTICLE" "ARTICLE" "ARTICLE" "ARTICLE" ...  
## $ year : num 2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 ...  
## $ title : chr "The creation of goal scoring opportunities at the 2015 womenâ\200\231s world cup" "A Model of Discipline: The Rule(s) of Midnight Football and the Production of Order in Subjects and Society" "Prevalence and Treatment Outcomes of Hand and Wrist Injuries in Professional Athletes: A Systematic Review" "Overweight and obese men's experiences in a sport-based weight loss intervention for men" ...  
## $ abstract : chr "Womenâ\200\231s international football is growing in terms of popularity, but to date there is limited publishe"| \_\_truncated\_\_ "This article explores the rationalities of social change of a sports-based intervention, midnight football, car"| \_\_truncated\_\_ "Background: Injuries to the hand and wrist constitute up to 25% of all athletic injuries, yet not much informat"| \_\_truncated\_\_ "In Western countries, such as Australia and the UK, a significantly greater proportion of men (relative to wome"| \_\_truncated\_\_ ...  
## $ doi : chr "10.1177/1747954120942051" "10.1177/0193723520919818" "10.1007/s11420-020-09760-w" "10.1016/j.psychsport.2020.101750" ...  
## $ journal : chr "International Journal of Sports Science and Coaching" "Journal of Sport and Social Issues" "HSS Journal" "Psychology of Sport and Exercise" ...  
## $ bibtexkey: chr "Scanlan2020803" "Ekholm2020450" "Lehman2020280" "Budden2020" ...  
## $ author : chr "Scanlan, M., Harms, C., Cochrane Wilkie, J., Maâ\200\231ayah, F." "Ekholm, D., Dahlstedt, M." "Lehman, J.D., Krishnan, K.R., Stepan, J.G., Nwachukwu, B.U." "Budden, T., Dimmock, J.A., Smith, B., Beauchamp, M., Rosenberg, M., Jackson, B." ...  
## $ source : chr "SCOPUS" "SCOPUS" "SCOPUS" "SCOPUS" ...

#Remove duplicates ## Check duplicates

## Remove { replacing by ’’

# Remove { or } from title  
# Escape metacaracthers in R is with \\  
data\_articles$title<-gsub("\\{|\\}","",data\_articles$title)  
data\_articles$author<-gsub("\\{|\\}","",data\_articles$author)

dim(data\_articles)

## [1] 143 9

length(unique(data\_articles$title))

## [1] 143

## Remove duplicates

data\_articles\_w\_duplicates<-data\_articles %>% distinct(title,.keep\_all = TRUE)  
dim(data\_articles\_w\_duplicates)

## [1] 143 9

# Articles by source

# Articles by year

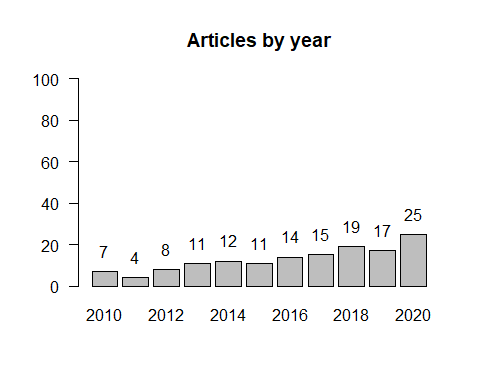
#Create contigency table for source column  
unique(data\_articles\_w\_duplicates$year)

## [1] 2020 2019 2018 2017 2016 2015 2014 2013 2012 2011 2010

tb<-table(data\_articles\_w\_duplicates$year)  
#Change names  
tb

##   
## 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020   
## 7 4 8 11 12 11 14 15 19 17 25

# barplot with articles by year  
bb<-barplot(height = tb,ylim = c(0,100),las=1, main = 'Articles by year')  
text(x = bb,y = tb+10,labels = tb)



# Export

library(writexl)  
  
write\_xlsx(data\_articles\_w\_duplicates,'../articles/data\_articles.xlsx')

df2bib(data\_articles\_w\_duplicates,file = '../articles/data\_articles.bib')

# ASReview processing

Data requirements for ASReview: \* title \* abstract

<https://asreview.readthedocs.io/en/latest/datasets.html#data-format>

1. Selected randomly some articles and are identified as relevant and irrelevant, at least 5 each. [Start reviewing](https://asreview.readthedocs.io/en/latest/quicktour.html#start-reviewing);
2. ASReview orders the publications in such a way that you see the most relevant publications first, simplifying the analysis of the abstracts;
3. all articles where screened. Although the last 178 didn’t identified any relevant articles and that is possible to stop the review when [stopping criterium](https://asreview.readthedocs.io/en/latest/quicktour.html#stop-reviewing) “ASReview orders the publications in such a way that you see the most relevant publications first. In other words, there is a point where you can be confident that you have seen (almost) all publications you need, and you can stop reviewing. When to stop is left to the user. A stopping criterium could be stopping after the last X presented abstracts were labeled irrelevant, or if your time is up. You can use the chart in the statistics panel to follow your progress”

#ASReview final dataset

Select only select articles

# Copy DOIs to clipboard to import to Zotero