# Zeit Index

*Dirk Ulbricht*

*Monday, January 05, 2015*

## Description

This file creates the Zeit indeces. This will be based either on the complete set of articles evaluated (politics, economics and opinion) or on the economics section only. In the latter case, economic_only should be given the value 'Yes' (the output csv will then be 'zeit economic.csv'). It is based on the evaluations of each article in "Ergebnis.csv"-Files containing the results of sentiment analysis.

```
economic_only='Yes'

# Load the necessary libraries and defining functions
library(zoo)
library(fBasics)

# Setting directories for storing files ------------------------------------------------------
DirRawTexts="H:/Zeit" # text files are stored here
DirCode='H:/git/zeit-2' # main directory
setwd(DirCode)

# Load register created by 'Getting_register.R' ---------------------------
load(paste(DirCode,"/register.RData",sep=''))
```

### Connecting sentiment evaluation and registry

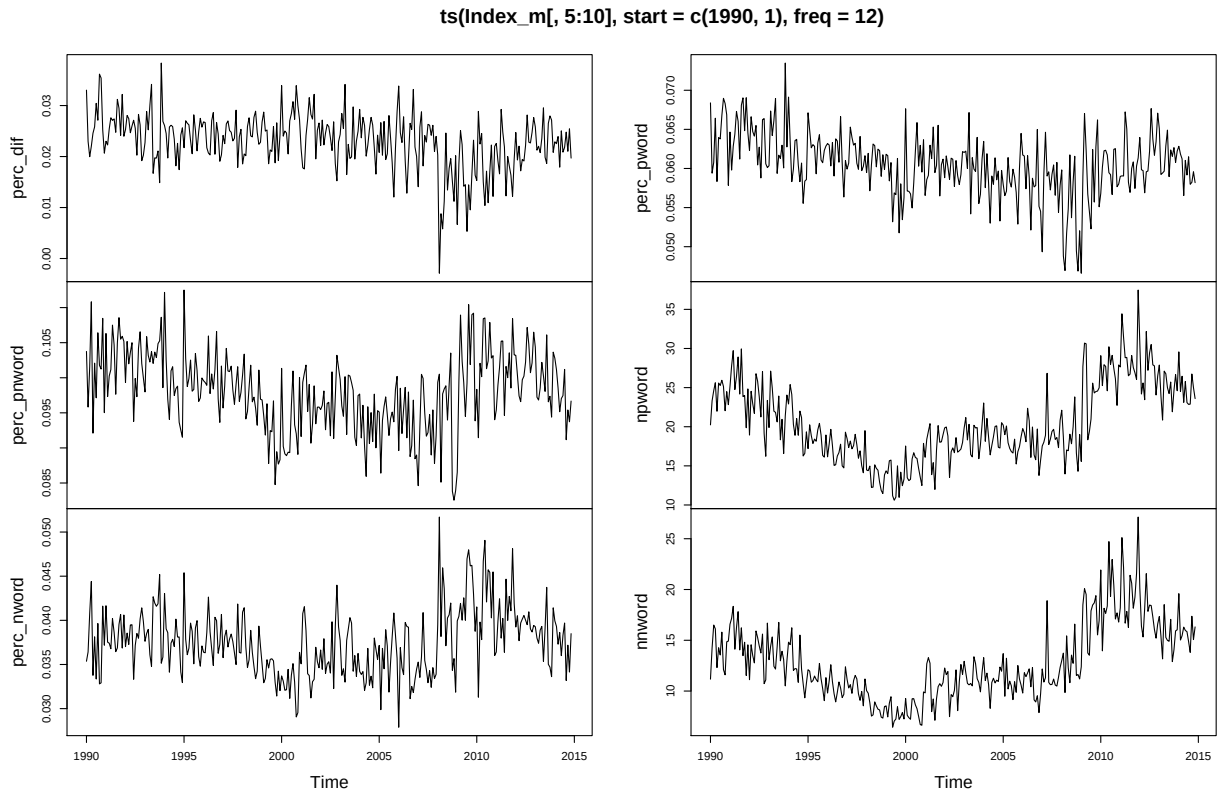
This is currently commented out, as it takes some time to compute.

```
# Getting "Ergebnis.csv" Files and integrate them into register -----------

# register$id=NA
# register$npword=NA
# register$nnword=NA
# register$nword=NA
# register$pvalue=NA
# register$nvalue=NA
# for (subd in listsubdirs){
#         if (subd=='1993.36'){next}
#         Ergebnis=read.csv(paste(DirRawTexts,'/',subd,'/','Ergebnis.csv',sep=''),row.names=1)
#         narticle=nrow(Ergebnis)
#         for (article in 1:narticle){
#                 register[Ergebnis$id[article],5:10]=Ergebnis[article,]
#         }
#
# }
# rm(Ergenbis)

# Adding dates ------------------------------------------------------------
```

```
#
# BegYear = 1990
# BegMonth = 1
# BegDay = 1
# EndYear = 2014
# EndMonth = 12
# EndDay = 31
#
# sBegDate = paste(BegDay, BegMonth, BegYear, sep=".")
# sEndDate = paste(EndDay, EndMonth, EndYear, sep=".")
# Date = seq(as.Date(sBegDate, format="%d.%m.%Y"), as.Date(sEndDate, format="%d.%m.%Y"), by="days")
# Tdays=Date[which(weekdays(Date)=='Thursday')]
#
# dates=data.frame(NA)
#
# dates=data.frame(year=format(Tdays,'%Y')
#                  ,month=format(Tdays,'%m')
#                  ,day=format(Tdays,'%d')
# )
#
#
# # check matches of thursdays and issues -----------------------------------
#
# nthurs_year=table(dates$year)
#
# aux=sapply(issues,strsplit,'\\.')
# aux=sapply(aux,function(x) x[1:2])
# aux=t(aux)
# aux=as.numeric(aux)
# aux=matrix(aux,ncol=2)
# nissues_year=table(aux[,1])
# compare_thur_issue=cbind(nissues_year,nthurs_year)
# compare_thur_issue_res=data.frame(year=c(1991,1993)
#                                   ,note=c('first issue missing',
#                                      '36. issue has no economic, politics or essay article'))
#
# # inserting dates ---------------------------------------------------------
#
# dates$issue=NA
# years=unique(dates$year)
# nyears=length(years)
# for (year in years){
#         nyissue=nrow(dates[which(year==dates$year),])
#         dates[which(year==dates$year),'issue']=1:nyissue
#
# }
# rm(year)
# dates$year.issue=NA
# register$day=NA
# register$month=NA
# for (i in 1:nrow(dates)){
#         ids=which(register$year==dates$year[i]&register$issue==dates$issue[i])
#         register[ids,'day']=dates$day[i]
```

```
#         register[ids,'month']=dates$month[i]
# }
# rm(dates,listsubdirs,Date,EndYear,i,BegDay,BegMonth,BegYear,EndDay,EndMonth,Tdays,ids,nyears,nyissue,

# Getting metadata out of the html files ---------------------------------

# register$title_in_text=NA
# register$date=NA
# register$keywords=NA
#
# for (i in 1:nrow(register)){
#         plainhtml <- read.csv(paste(DirRawTexts,'/',register$year[i],'.',register$issue[i],'/',i,'.tx
#         # plainhtml <-c(plainhtml)
#         plainhtml=apply(plainhtml,2,as.character)
#         plainhtml<-paste(plainhtml,sep="",collapse="")
#
#         title_index=regexec(paste('<title>','(.*)',' DIE ZEIT Archiv',sep=''),plainhtml)
#         register$title_in_text[i]=regmatches(plainhtml,title_index)[[1]][2]
#
#         date_index=regexec(paste('date" content="','([0-9]{4}-[0-9]{2}-[0-9]{2})',sep=''),plainhtml)
#         register$date[i]=regmatches(plainhtml,date_index)[[1]][2]
#
#         keywords_index=regexec(paste('keywords" content="','(.*)','\"><(meta property)=\"og:site_name
#         register$keywords[i]=regmatches(plainhtml,keywords_index)[[1]][2]
# }

# restriction to economic section ----------------------------------------

if (economic_only=='Yes'){
        load('e_register.RData')
        e_register=unique(e_register)
        e_index=match(e_register$link,register$link)
        register=register[e_index,]
        }

# Getting subdirectories -------------------------------------------------
listsubdirs=list.files(DirRawTexts)

# calculating relative values --------------------------------------------
register$perc_dif=(register$npword-register$nnword)/register$nword
register$perc_pword=register$npword/register$nword
register$perc_nword=register$nnword/register$nword
register$perc_pnword=(register$npword+register$nnword)/register$nword
register$rpvalue=register$pvalue/register$npword
register$rpvalue[register$npword==0]=0
register$rnvalue=register$nvalue/register$nnword
register$rnvalue[register$nnword==0]=0
register$rvalue=(register$pvalue+register$nvalue)/(register$npword+register$nnword)
register$yearissue=paste(register$year,register$issue,sep='.')
register$yearmonth=paste(register$year,register$month,sep='.')
register=register[-which(register$link=='http://www.zeit.de/1993/36/ein-ganz-legaler-nepp'),] # has not
```

```r
# aggregating to issues ----------------------------------------------------

Index=aggregate(register[,c("year","month","day")],list(register$yearissue),mean,na.rm=T)
paste0=function(x){if (nchar(x)==1){y=paste('0',x,sep='')
                                    return(y)}else{return(x)}}
Index$Month=sapply(Index$month,paste0)
Index$Day=sapply(Index$day,paste0)
Index$YearMonthDay=as.Date(paste(Index$year,Index$Month,Index$Day,sep='/'),'%Y/%m/%d')
Index$yearmonth=paste(Index$year,Index$month,sep='.')
val_gr=c("perc_dif", "perc_pnword","perc_nword","perc_pword","npword","nnword","nword","pvalue","nvalue
Index=cbind(Index,aggregate(register[,val_gr],list(register$yearissue),mean,na.rm=T)[,-1])




# aggregating over month ---------------------------------------------------

Index_m=suppressWarnings(aggregate(Index,list(Index$yearmonth),mean,na.rm=T))
# Index_m=Index_m[order(Index_m$yearmonth),]
Index_m$yearmonth=NULL
Index_m[,c(2)]=NULL
Index_m$Month=NULL
Index_m$Day=NULL
Index_m$day=NULL



# plot(ts(Index_m$kur_perc_dif,start=c(1990,1),freq=12),type='l')

# Variation within one month -----------------------------------------------

Index_m$var_perc_dif=aggregate(register$perc_dif,list(register$yearmonth),sd,na.rm=T)[,2]
Index_m$var_perc_pword=aggregate(register$perc_pword,list(register$yearmonth),sd,na.rm=T)[,2]
Index_m$var_perc_nword=aggregate(register$perc_nword,list(register$yearmonth),sd,na.rm=T)[,2]
Index_m$var_perc_pnword=aggregate(register$perc_pnword,list(register$yearmonth),sd,na.rm=T)[,2]
Index_m$var_rpvalue=aggregate(register$rpvalue,list(register$yearmonth),sd,na.rm=T)[,2]
Index_m$var_rnvalue=aggregate(register$rnvalue,list(register$yearmonth),sd,na.rm=T)[,2]
Index_m$var_rvalue=aggregate(register$rvalue,list(register$yearmonth),sd,na.rm=T)[,2]

plot(ts(Index_m[,5:10],start=c(1990,1),freq=12))

# Kurtosis within one month ------------------------------------------------

Index_m$kur_perc_dif=aggregate(register$perc_dif,list(register$yearmonth),kurtosis,na.rm=T)[,2]
Index_m$kur_perc_pword=aggregate(register$perc_pword,list(register$yearmonth),kurtosis,na.rm=T)[,2]
Index_m$kur_perc_nword=aggregate(register$perc_nword,list(register$yearmonth),kurtosis,na.rm=T)[,2]
Index_m$kur_perc_pnword=aggregate(register$perc_pnword,list(register$yearmonth),kurtosis,na.rm=T)[,2]
Index_m$kur_rpvalue=aggregate(register$rpvalue,list(register$yearmonth),kurtosis,na.rm=T)[,2]
Index_m$kur_rnvalue=aggregate(register$rnvalue,list(register$yearmonth),kurtosis,na.rm=T)[,2]
Index_m$kur_rvalue=aggregate(register$rvalue,list(register$yearmonth),kurtosis,na.rm=T)[,2]

plot(ts(Index_m[,5:10],start=c(1990,1),freq=12))
```




```
# Rolling Means ------------------------------------------------------------------------
Index_m$ym=as.yearmon(Index_m$YearMonthDay,'%Y/%m/%d')
Index_m=Index_m[order(Index_m$ym),]
Index_m=Index_m[-nrow(Index_m),]
zeit=zoo(Index_m[,5:30],Index_m$ym)

zeit3=rollmean(zeit,k=3,align='right')
zeit=cbind(zeit,zeit3)
zeit_df=as.data.frame(zeit)
for (i in 1:nrow(zeit_df)){Index_m$Month[i]=paste0(as.character(Index_m$month[i]))}
row.names(zeit_df)=paste(Index_m$year,Index_m$Month,sep='/')

# zeit_df_window=zeit_df[which(row.names(zeit_df)=='Jan 2001'):nrow(zeit_df),]
zeit_df=zeit_df[order(row.names(zeit_df)),]
# write.csv(zeit_df_window,'zeit.csv')
if (economic_only=='Yes'){write.csv(zeit_df,'zeit_economic.csv')}else{
    write.csv(zeit_df,'zeit.csv')
}
```