wanted to use LDAvis with output from the [topicmodels](http://cran.r-project.org/web/packages/topicmodels/index.html) package. It works really nicely with texts preprocessed using the [tm](http://cran.r-project.org/web/packages/tm/index.html) package. The trick is extracting the information LDAvis requires from the model and placing it into a specifically structured JSON formatted object.

To make the conversion from topicmodels output to LDAvis JSON input easier, I created a linking function called topicmodels\_json\_ldavis. The full function is below. To use it follow these steps:

1. Create a VCorpus object using the tm package's Corpus function.
2. Convert this to a document term matrix using DocumentTermMatrix, also from tm.
3. Run your model using topicmodel's LDA function.
4. Convert the output into JSON format using topicmodels\_json\_ldavis. The function requires the output from steps 1-3.
5. Visualise with LDAvis' serVis.

#' Convert the output of a topicmodels Latent Dirichlet Allocation to JSON

#' for use with LDAvis

#'

#' @param fitted Output from a topicmodels \code{LDA} model.

#' @param corpus Corpus object used to create the document term

#' matrix for the \code{LDA} model. This should have been create with

#' the tm package's \code{Corpus} function.

#' @param doc\_term The document term matrix used in the \code{LDA}

#' model. This should have been created with the tm package's

#' \code{DocumentTermMatrix} function.

#'

#' @seealso \link{LDAvis}.

#' @export

topicmodels\_json\_ldavis <- function(fitted, corpus, doc\_term){

# Required packages

library(topicmodels)

library(dplyr)

library(stringi)

library(tm)

library(LDAvis)

# Find required quantities

phi <- posterior(fitted)$terms %>% as.matrix

theta <- posterior(fitted)$topics %>% as.matrix

vocab <- colnames(phi)

doc\_length <- vector()

for (i in 1:length(corpus)) {

temp <- paste(corpus[[i]]$content, collapse = ' ')

doc\_length <- c(doc\_length, stri\_count(temp, regex = '\\S+'))

}

temp\_frequency <- inspect(doc\_term)

freq\_matrix <- data.frame(ST = colnames(temp\_frequency),

Freq = colSums(temp\_frequency))

rm(temp\_frequency)

# Convert to json

json\_lda <- LDAvis::createJSON(phi = phi, theta = theta,

vocab = vocab,

doc.length = doc\_length,

term.frequency = freq\_matrix$Freq)

return(json\_lda)

}