library(rvest)

douban\_movie = data.frame()

for (i in 1:10) {

url <- paste("https://movie.douban.com/top250?start=",25\*(i-1),"&filter=",sep="")

web <- read\_html(url,encoding="utf-8")

movie\_title <- web %>%

html\_nodes(xpath = '//div[@class="hd"]/a/span[1]') %>%

html\_text()

link <- web %>%

html\_nodes(xpath = '//div[@class="hd"]/a') %>%

html\_attr('href')

movie\_detail <- web %>%

html\_nodes(xpath = '//div[@class="bd"]/p[1]') %>%

html\_text()

movie\_num <- web %>%

html\_nodes(xpath = '//div[@class="bd"]/div/span[2]') %>%

html\_text()

movie\_evaluate <- web %>%

html\_nodes(xpath = '//div[@class="bd"]/div/span[4]') %>%

html\_text()

douban\_movie <- rbind(douban\_movie,data.frame(movie\_title, movie\_detail,movie\_num,movie\_evaluate,link))

}

library(knitr)

kable(head(douban\_movie),format = "html")

write.table(douban\_movie, "D:/Documents/douban\_movie.csv", sep=",", row.names=FALSE)