```
#========================================================================

#姓名：郭维娟  学号：SY1608230
#时间：5/12/2017
#作业：抓取TripAdviser上的信息
#========================================================================

### 提取一页信息
library(rvest)
url <- "https://www.tripadvisor.cn/Hotels-g294212-Beijing-Hotels.html"
web <- read_html(url)

## 抓取酒店名字
hotel_title_nodes <- html_nodes(web, xpath =
"//div[@class=\"listing_title\"]/a")
hotel_title <- html_text(hotel_title_nodes)

## 抓取酒店排名
hotel_rank_nodes <- html_nodes(web, xpath =
"//div[@class=\"popRanking\"]")
hotel_rank <- html_text(hotel_rank_nodes)
hotel_rank <- gsub("在北京市 6,454 家酒店中排名第", "", hotel_rank)

## 抓取最新评论 这个在最后无法输出，说： 参数值意味着不同的行数：30, 0。在输出标签
时遇到相同问题（标签相关代码已删）
newest_review_nodes <- html_nodes(web, xpath = "//div[@class=\"listing
easyClear
p13n_imperfect\"]/div/div/div/div/div[1]/div[4]/ul/li[1]/span[2]/a")
newest_review <- html_text(newest_review_nodes)

## 抓取酒店评论链接
hotel_reviewlink_nodes <- html_nodes(web, xpath =
"//div[@class=\"rating\"]/span/a")
hotel_reviewlink <- html_attr(hotel_reviewlink_nodes, "href")
link <- sprintf("https://www.tripadvisor.cn/%s", hotel_reviewlink)

TripAdviser_hotels <- data.frame(名字 = hotel_title, 排名 = hotel_rank,
更多评论 = link)
TripAdviser_hotels
getwd()
write.table(TripAdviser_hotels, file = "TripAdviser_hotels.txt")

###存在以下问题：
#1.在输出data.frame时提示行数不一致，导致不能输出，不知该如何解决，这是遇到最多的
问题。
#2.翻页之后，网站的地址没有发生变化，所以不知道如何写循环来实现翻页，从而输出更多页
的内容。
```