[ R 수행 평가(2) – 동적 크롤링과 스크래핑 ]

다음은 6월까지만 해도 스크래핑이 잘 됐던 yes24 사이트 명견만리 댓글을 읽어오는

소스이다...ㅜㅜ 그런데 지금은 안된다. 이유는 yes24 아이들이 뭔가를 변경했기 때문이다.

해결해보자... 변경해야 할것들을 반영하여

완성하면 yes24.R 과 yes24.txt 로 저장하여 메일로 제출한다.

건투를 빈다......





site <- paste("http://www.yes24.com/24/goods/40936880")

remDr$navigate(site)

webElem <- remDr$findElement("css", "body")

#remDr$executeScript("scrollTo(0, document.body.scrollHeight-200)", args = list(webElem))

remDr$executeScript("scrollBy(0, 2500)", args = list(webElem))

Sys.sleep(1)

remDr$executeScript("scrollBy(0, 2500)", args = list(webElem))

Sys.sleep(1)

remDr$executeScript("scrollBy(0, 2500)", args = list(webElem))

Sys.sleep(1)

remDr$executeScript("scrollBy(0, 1000)", args = list(webElem))

Sys.sleep(3)

repl\_v = NULL

endFlag <- FALSE

page <- 3

repeat {

for(index in 1:5) {

fullContentLinkCSS <- paste('#PartContent', index, sep='')

fullContentLink<-remDr$findElements(using='css', fullContentLinkCSS)

if (length(fullContentLink) == 0) {

endFlag <- TRUE

break

}

sapply(fullContentLink,function(x){x$clickElement()})

Sys.sleep(1)

fullContentCSS <- paste('#FullContent', index, sep='')

fullContent<-remDr$findElements(using='css', fullContentCSS)

print(fullContent)

repl <-sapply(fullContent,function(x){x$getElementText()})

print(repl)

repl\_v <- c(repl\_v, unlist(repl))

}

if(endFlag)

break;

page <- page+1;

if(page == 5)

page <- page+1;

nextPageCSS <- paste("#Div\_Review > p:nth-child(11) > a:nth-child(",page,")",sep="")

nextPageLink<-remDr$findElements(using='css',nextPageCSS)

if(page == 14)

page<-3

sapply(nextPageLink,function(x){x$clickElement()})

Sys.sleep(5)

remDr$executeScript("scrollBy(0, -3000)", args = list(webElem))

print(page)

}

write(repl\_v, "yes24.txt")