ch03

2023-10-08

## chapter 03

### 웹문서 읽어오기(rvest 및 dplyr 패키지 이용)

# install.packages("rvest")  
# install.packages("dplyr")  
  
library(rvest)

## Warning: package 'rvest' was built under R version 4.3.1

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

exurl = "https://ko.wikipedia.org/wiki/%EB%B9%84%EC%A0%95%ED%98%95\_%EB%8D%B0%EC%9D%B4%ED%84%B0"  
html\_ex = read\_html(exurl, encoding = "UTF-8")  
html\_ex %>% html\_nodes(".mw-parser-output p") %>% html\_text()

## [1] "\n"   
## [2] "비정형 데이터 (unstructured data, unstructured information, 비정형 정보), 비구조화 데이터, 비구조적 데이터는 미리 정의된 데이터 모델이 없거나 미리 정의된 방식으로 정리되지 않은 정보를 말한다. 비정형 정보는 일반적으로 텍스트 중심으로 되어 있으나 날짜, 숫자, 사실과 같은 데이터도 포함할 수 있다. 이로써 변칙과 모호함이 발생하므로 데이터베이스의 칸 형식의 폼에 저장되거나 문서에 주석화된(의미적으로 태그된) 데이터에 비해 전통적인 프로그램을 사용하여 이해하는 것을 불가능하게 만든다.\n"  
## [3] "1998년 메릴린치는 잠재적으로 이용 가능한 모든 비즈니스 정보 중 약 80~90% 정도가 비정형 형식에서 기원한 것으로 보는 경험 법칙을 언급하였다.[1] 이 경험 법칙은 1차 연구나 양적 연구에 근간을 두지 않지만 그럼에도 일부 받아들여지고 있다.[2]"   
## [4] "비즈니스 인텔리전스에 대한 최초의 연구는 수치 데이터가 아닌 비정형 텍스트 형태의 데이터에 초점을 두었다.[1] 1958년 초에 H. P. Luhn 등의 컴퓨터 과학 연구원들은 특히 비정형 텍스트의 추출과 분류에 관심을 가졌다.[1] 그러나 세기가 바뀐 뒤에서야 비로소 기술이 연구적 관심을 따라잡을 수 있게 되었다. 2004년, SAS 인스티튜트는 더 효율적인 분석을 위하여 특이값 분해(SVD)로 초차원적 텍스트 공간을 더 작은 차원으로 줄이기 위해 사용되는 SAS 텍스트 마이너를 개발하였다.[2]"

## 웹문서 읽어오기 연습

library(rvest)  
library(dplyr)  
  
# 2023-10-08 06:21   
# 업데이트 정보만 추출하였습니다.  
  
ex2url = "https://www.leagueoflegends.com/en-us/news/game-updates/"  
html\_ex2 = read\_html(ex2url, encoding = "en-us")  
html\_ex2 %>% html\_elements("div") %>% html\_nodes(".style\_\_Title-sc-1h41bzo-8") %>% html\_text()

## [1] "Worlds 2023 Event Trailer"   
## [2] "Patch 13.20 Notes"   
## [3] "Teamfight Tactics patch 13.20 notes"   
## [4] "Fluft of Poros | Little Legend Showcase"  
## [5] "Teamfight Tactics patch 13.19 notes"   
## [6] "Patch 13.19 Notes"

α = 4.5  
β = 11.5  
m = 1000  
(ex3.theta = rbeta(m, α, β))

## [1] 0.35514727 0.36559188 0.37879412 0.32277317 0.40079158 0.18485522  
## [7] 0.35841144 0.28059145 0.35451862 0.21191445 0.27467286 0.21270322  
## [13] 0.30470692 0.19818813 0.49982394 0.10314760 0.28754539 0.27184786  
## [19] 0.17581019 0.33432893 0.23523348 0.11831378 0.25696260 0.24582244  
## [25] 0.09924549 0.44486351 0.23052810 0.21613259 0.21313377 0.02745624  
## [31] 0.29238633 0.21246852 0.56384028 0.24028956 0.55508537 0.10527612  
## [37] 0.24282969 0.35699042 0.18915392 0.37329007 0.14383157 0.17452463  
## [43] 0.50807677 0.17115354 0.30057843 0.22713907 0.27600268 0.12891261  
## [49] 0.34900560 0.37229103 0.33456268 0.05514273 0.08234253 0.20504823  
## [55] 0.32432570 0.35151327 0.27629414 0.47817962 0.27486820 0.20539449  
## [61] 0.23357364 0.26779706 0.34051140 0.26581790 0.38959706 0.32842321  
## [67] 0.08563801 0.38537446 0.36037519 0.21172989 0.44494046 0.40964108  
## [73] 0.27628091 0.30018975 0.20008331 0.26650769 0.29576078 0.19257814  
## [79] 0.16030043 0.35147494 0.19380794 0.33040052 0.50654537 0.39663399  
## [85] 0.37218256 0.28979773 0.45773616 0.13974780 0.29317951 0.31183146  
## [91] 0.15929800 0.33370765 0.19883061 0.10688469 0.23430809 0.31205956  
## [97] 0.11301028 0.32244704 0.34578547 0.41187823 0.08640648 0.15038158  
## [103] 0.42842765 0.12186091 0.25985370 0.30668061 0.13432880 0.17282404  
## [109] 0.11078109 0.30578124 0.08130141 0.11611083 0.33653137 0.22671933  
## [115] 0.37553175 0.28732722 0.36403880 0.41228173 0.31066575 0.31459156  
## [121] 0.22026693 0.39370264 0.44876081 0.25340881 0.27943388 0.27593112  
## [127] 0.14828795 0.15144351 0.14180571 0.38764008 0.21314282 0.32843769  
## [133] 0.50729236 0.08354547 0.23196500 0.29604007 0.28561482 0.36371596  
## [139] 0.50380905 0.22895266 0.30012145 0.26010492 0.60102471 0.41614171  
## [145] 0.37177919 0.33228364 0.12287327 0.24329683 0.26077551 0.29333370  
## [151] 0.17375395 0.18259029 0.19448933 0.20846334 0.40539215 0.30553301  
## [157] 0.22807132 0.35919756 0.44164403 0.06810261 0.28219318 0.50290033  
## [163] 0.24640110 0.55612606 0.11770421 0.20264475 0.13322096 0.23283151  
## [169] 0.26535584 0.16178440 0.33398557 0.30074671 0.23474608 0.19236674  
## [175] 0.09139618 0.32316919 0.29174905 0.34688110 0.29409488 0.72694016  
## [181] 0.49515657 0.32285685 0.24730454 0.27733089 0.24262342 0.28454483  
## [187] 0.53766966 0.35899644 0.23724735 0.31079870 0.37307047 0.33490965  
## [193] 0.25991065 0.17775530 0.22861688 0.22387002 0.21107264 0.22900426  
## [199] 0.12938738 0.31589041 0.26282398 0.33836753 0.32748547 0.36325712  
## [205] 0.24072267 0.28751701 0.23174833 0.20895850 0.26908366 0.27761331  
## [211] 0.40800308 0.17715585 0.27076543 0.13507462 0.34611960 0.33090054  
## [217] 0.13421067 0.34681401 0.16717688 0.55350155 0.12877258 0.20210105  
## [223] 0.15167534 0.34359352 0.37379754 0.34859155 0.35980687 0.15080548  
## [229] 0.30748195 0.24847481 0.32673967 0.40317717 0.46366436 0.16244208  
## [235] 0.38779699 0.34309094 0.21629654 0.25211126 0.21299751 0.23571785  
## [241] 0.05494106 0.21120539 0.37526931 0.42586981 0.29468439 0.33308382  
## [247] 0.25063491 0.10921295 0.35585820 0.09621215 0.45442429 0.23339968  
## [253] 0.40413927 0.24150411 0.31046156 0.37600762 0.24134188 0.34525493  
## [259] 0.37294279 0.44383614 0.30050437 0.06248785 0.22654325 0.37843561  
## [265] 0.37547358 0.25492943 0.19953124 0.26188225 0.18000857 0.15394308  
## [271] 0.54516615 0.45993576 0.25160767 0.26214763 0.34533929 0.19258142  
## [277] 0.27943067 0.31525268 0.21213594 0.37674569 0.25476422 0.13045178  
## [283] 0.37455568 0.37760219 0.27651715 0.37016106 0.38469768 0.21302273  
## [289] 0.20759128 0.27803438 0.40499511 0.14792135 0.29166707 0.23938302  
## [295] 0.34665367 0.38201426 0.22528984 0.52271341 0.22552419 0.24042211  
## [301] 0.20411395 0.32029191 0.09915237 0.31127147 0.35257793 0.42365177  
## [307] 0.39666200 0.54017881 0.25891865 0.19742498 0.32664194 0.27546133  
## [313] 0.06391404 0.31789085 0.32497695 0.23418833 0.39651144 0.11979875  
## [319] 0.29332837 0.34362212 0.27169318 0.16835902 0.24025625 0.13183745  
## [325] 0.24941381 0.26484359 0.22388262 0.27505912 0.10543679 0.40348327  
## [331] 0.27521728 0.33511340 0.50579749 0.19942644 0.25873197 0.50449528  
## [337] 0.13818481 0.49585732 0.33416144 0.34113435 0.21839860 0.23961904  
## [343] 0.32635699 0.54716075 0.17197771 0.03728712 0.14835796 0.36641220  
## [349] 0.10210611 0.38159551 0.13760501 0.25438203 0.18386912 0.23093161  
## [355] 0.30038550 0.24053430 0.27870363 0.48424148 0.29862029 0.33415064  
## [361] 0.09894831 0.31916307 0.43681332 0.28942298 0.51322832 0.28651365  
## [367] 0.37525008 0.31524933 0.21081759 0.27892904 0.10937908 0.28045541  
## [373] 0.21629667 0.41463945 0.26148649 0.22433463 0.21263104 0.28389944  
## [379] 0.28385445 0.18829077 0.34604403 0.26869898 0.30519795 0.29984310  
## [385] 0.33610293 0.19042211 0.16672784 0.38020504 0.43044223 0.32103988  
## [391] 0.27407068 0.30294120 0.15212180 0.24268869 0.25879852 0.34724140  
## [397] 0.26999027 0.22751494 0.24572162 0.24777541 0.52494968 0.16346970  
## [403] 0.24630013 0.14431153 0.40022073 0.31487635 0.11682679 0.32773922  
## [409] 0.53062751 0.19496173 0.07775294 0.42774455 0.40575957 0.11542794  
## [415] 0.18539623 0.42794633 0.25914013 0.37232386 0.35974550 0.14576789  
## [421] 0.23256589 0.31745271 0.42339580 0.49594281 0.42342056 0.22062935  
## [427] 0.37067599 0.55247575 0.32283670 0.37107273 0.14489811 0.53278283  
## [433] 0.22932451 0.09363151 0.47347295 0.40605533 0.21670734 0.18046723  
## [439] 0.44734040 0.48982355 0.34938578 0.33346618 0.46889904 0.28890807  
## [445] 0.26179782 0.53295996 0.23915328 0.09571887 0.33818976 0.51957612  
## [451] 0.18624958 0.35834571 0.13517179 0.35377654 0.40723473 0.35710821  
## [457] 0.33462989 0.31625962 0.26329278 0.60548059 0.35271574 0.43205087  
## [463] 0.23320031 0.38492256 0.27928596 0.53248303 0.40077030 0.19495761  
## [469] 0.43662195 0.28439489 0.18442339 0.15788066 0.40037776 0.62818178  
## [475] 0.38655539 0.21339136 0.22708368 0.33796659 0.30307662 0.19561285  
## [481] 0.25264870 0.40765655 0.18728608 0.47012443 0.48052845 0.18641565  
## [487] 0.10347379 0.36827383 0.36829260 0.20841713 0.43394240 0.30432667  
## [493] 0.34406179 0.32258784 0.37614757 0.38267234 0.27110508 0.30999775  
## [499] 0.21733073 0.17694749 0.31479982 0.38417191 0.45676537 0.20861788  
## [505] 0.27054271 0.26778803 0.37589173 0.16932102 0.51889369 0.15939043  
## [511] 0.35571482 0.21244460 0.25158587 0.24406147 0.33802279 0.24982791  
## [517] 0.36763944 0.22445815 0.24309519 0.30760155 0.24820277 0.19901355  
## [523] 0.48031561 0.36670041 0.44276132 0.19387690 0.40660117 0.29243075  
## [529] 0.18742106 0.36186890 0.34773897 0.07942689 0.33570708 0.34350194  
## [535] 0.23554409 0.16219987 0.19174104 0.12115202 0.25149016 0.25282623  
## [541] 0.27153645 0.19726738 0.53962844 0.32280105 0.32308862 0.16150650  
## [547] 0.41401876 0.20501015 0.40149430 0.23807643 0.29200481 0.31912786  
## [553] 0.20815068 0.21482644 0.25168061 0.19931321 0.38149904 0.08054962  
## [559] 0.26680597 0.21611825 0.21408980 0.47646704 0.20153421 0.19519652  
## [565] 0.33221320 0.15230244 0.22626038 0.49032600 0.19801655 0.36347653  
## [571] 0.19070841 0.14265351 0.14492368 0.13036039 0.28828228 0.27832916  
## [577] 0.34129974 0.14730518 0.35025846 0.45318354 0.26236879 0.27299727  
## [583] 0.12898495 0.35528489 0.30203421 0.20935267 0.16743007 0.30650135  
## [589] 0.41998827 0.45087644 0.24546519 0.41759056 0.36016028 0.38701998  
## [595] 0.26898482 0.20071700 0.41095024 0.06034479 0.46537014 0.32426727  
## [601] 0.19877629 0.35089194 0.41410916 0.35405327 0.35879271 0.29368198  
## [607] 0.24536013 0.27910031 0.19979461 0.23337911 0.37259124 0.26500534  
## [613] 0.27044031 0.15303404 0.37730451 0.41613530 0.17651835 0.24741395  
## [619] 0.18596356 0.23949612 0.30789164 0.03850104 0.53590492 0.18192132  
## [625] 0.32571394 0.43619263 0.17849393 0.23828109 0.24558450 0.24022136  
## [631] 0.26913616 0.13944190 0.17356546 0.29360868 0.31185291 0.09110745  
## [637] 0.18660408 0.25523601 0.14745966 0.33927406 0.17399729 0.33845731  
## [643] 0.13713525 0.21654904 0.12790984 0.25634636 0.36486068 0.46391989  
## [649] 0.27736913 0.19589585 0.19482040 0.22345160 0.13207581 0.31495785  
## [655] 0.22712463 0.22076411 0.31259445 0.28016494 0.40739327 0.37412942  
## [661] 0.32706209 0.21782440 0.20719686 0.37880796 0.25026170 0.40787872  
## [667] 0.27714095 0.34661677 0.24525589 0.38000492 0.25962696 0.23709102  
## [673] 0.23703525 0.34758120 0.32372280 0.35652536 0.28198333 0.11851584  
## [679] 0.39102822 0.20478370 0.35121677 0.24575732 0.08034798 0.17863945  
## [685] 0.27312543 0.16214426 0.45287146 0.19266200 0.16573258 0.39286214  
## [691] 0.48885033 0.67079359 0.28469448 0.27441031 0.39351515 0.43216866  
## [697] 0.20815197 0.34960782 0.13733647 0.25125935 0.17667309 0.43007136  
## [703] 0.15610538 0.17974026 0.36271735 0.38721435 0.19742929 0.41815213  
## [709] 0.40384168 0.21346064 0.29303093 0.44067437 0.15082568 0.40839519  
## [715] 0.23356294 0.41635705 0.17622506 0.34111525 0.34578658 0.44686705  
## [721] 0.19754851 0.27669769 0.36930637 0.38886834 0.38345052 0.48531637  
## [727] 0.13817980 0.10749850 0.25663865 0.30843004 0.40536507 0.20972321  
## [733] 0.23610087 0.18841356 0.23393971 0.18508873 0.45068424 0.31050129  
## [739] 0.28212303 0.58720239 0.28048494 0.15611863 0.21179418 0.36544413  
## [745] 0.14539709 0.18949734 0.14469842 0.33108549 0.25280647 0.31589772  
## [751] 0.28238634 0.12371029 0.37942046 0.18891648 0.17058709 0.40322575  
## [757] 0.26912696 0.29205594 0.17531033 0.41601685 0.37022212 0.22925116  
## [763] 0.24662483 0.30148914 0.23661489 0.31003711 0.21803051 0.16679683  
## [769] 0.40136212 0.18511985 0.35189150 0.34057978 0.15907492 0.41270204  
## [775] 0.51656070 0.25088433 0.19991955 0.49739098 0.17495851 0.36070902  
## [781] 0.23138076 0.21567947 0.21476592 0.25154062 0.32875023 0.17870661  
## [787] 0.20444818 0.13570812 0.27938250 0.17316767 0.33916829 0.56834922  
## [793] 0.24081303 0.46831684 0.30990956 0.15856265 0.43154932 0.23532077  
## [799] 0.24619259 0.18866230 0.46339740 0.09028219 0.28816365 0.24663985  
## [805] 0.30960634 0.28999907 0.37508577 0.15833140 0.21909353 0.33012056  
## [811] 0.31400508 0.25476965 0.17660484 0.32016141 0.30558198 0.31640747  
## [817] 0.42489867 0.31854571 0.28798263 0.30005770 0.17711578 0.24100430  
## [823] 0.14909563 0.22689509 0.09330154 0.25446936 0.39223536 0.36651230  
## [829] 0.52181164 0.30599192 0.14758608 0.29798960 0.26223883 0.23423534  
## [835] 0.08948487 0.14583357 0.34092106 0.48734409 0.38910697 0.34550135  
## [841] 0.21170180 0.25576308 0.19589069 0.24123034 0.33015422 0.17417720  
## [847] 0.35691890 0.26541982 0.27457518 0.27981715 0.26081702 0.30366373  
## [853] 0.37417089 0.28650949 0.15728117 0.40228296 0.30037216 0.29620894  
## [859] 0.12129682 0.33549605 0.32259289 0.18037272 0.23765045 0.36156405  
## [865] 0.25848991 0.44682478 0.37291880 0.13317098 0.45804349 0.21340244  
## [871] 0.06869128 0.24147102 0.32272971 0.30630622 0.44353952 0.11502588  
## [877] 0.28833800 0.13598126 0.46004203 0.20126547 0.28290773 0.25589258  
## [883] 0.33349957 0.26184166 0.31343131 0.35898720 0.23643694 0.34551004  
## [889] 0.31262583 0.32313444 0.23708394 0.25925190 0.32037475 0.15298468  
## [895] 0.27182924 0.20941059 0.17122822 0.24512251 0.27018531 0.30730412  
## [901] 0.40409475 0.30946824 0.36763085 0.30863210 0.17019433 0.22949152  
## [907] 0.30071914 0.23447262 0.32676227 0.32825340 0.38739271 0.24497199  
## [913] 0.47151982 0.24668471 0.27370965 0.16077333 0.21954571 0.34106041  
## [919] 0.28489187 0.27163004 0.39391235 0.30184897 0.48652731 0.33918403  
## [925] 0.22523535 0.18666219 0.13613198 0.42244337 0.29690446 0.25574360  
## [931] 0.29132288 0.26397982 0.11472027 0.15239468 0.17716902 0.28218456  
## [937] 0.53184636 0.31414267 0.30795907 0.22189462 0.43425817 0.43178794  
## [943] 0.41402095 0.27679414 0.28917451 0.18691885 0.25469372 0.65795269  
## [949] 0.30285627 0.32657103 0.17188011 0.27043368 0.36770602 0.39880112  
## [955] 0.14168699 0.24991269 0.17699126 0.36383498 0.30093209 0.20193975  
## [961] 0.13483297 0.28134954 0.13781680 0.40817042 0.44368180 0.26699570  
## [967] 0.31999373 0.28128318 0.39244474 0.14169364 0.29235775 0.21989928  
## [973] 0.17310907 0.51542576 0.33757550 0.24723461 0.13600732 0.31898244  
## [979] 0.26814719 0.22442140 0.22794007 0.11998729 0.16206664 0.36365467  
## [985] 0.17173715 0.28836526 0.13952453 0.36876821 0.49615028 0.39608722  
## [991] 0.10653077 0.40670521 0.26547662 0.17807653 0.20696289 0.25043291  
## [997] 0.22068775 0.17237037 0.19391953 0.36938061