1. **Пример классификации текстовых документов**

**Классификация пресс-релизов британского правительства**

# <https://www.gov.uk/government/announcements?keywords=&taxons%5B%5D=all&subtaxons%5B%5D=all&announcement_filter_option=press-releases&departments%5B%5D=all&people%5B%5D=all&world_locations%5B%5D=all&from_date=01%2F09%2F2018&to_date=30%2F10%2F2018>

# # пакет для регулярных выражений

# install.packages("stringr")

# # основной пакет для веб-скрепинга

# install.packages("RCurl")

# # пакет для работы с XML, HTML, Xpath

# install.packages("XML")

# library("stringr")

# library ("RCurl")

# library ("XML")



**Проанализируем HTML–код. Каждый пресс-релиз содержится в гиперссылке, которая содержится в теге <li> с атрибутом id.**

**<li id**="news\_article\_560756" class="document-row">

<h3>**<a** href="[/government/news/foreign-secretary-congratulates-sierra-leone-on-eliminating-ebola](view-source:https://www.gov.uk/government/news/foreign-secretary-congratulates-sierra-leone-on-eliminating-ebola)"**>**Foreign Secretary congratulates Sierra Leone on eliminating Ebola**</a>**</h3>

<ul class="attributes">

<li><time class="public\_timestamp" datetime="2015-11-07T10:47:47+00:00"> 7 November 2015</time></li>

<li class="organisations"><abbr title="Foreign &amp; Commonwealth Office">FCO</abbr></li>

<li class="display-type">Press release</li>

</ul>

</li>

**Для перехода на следующую страницу нужно извлечь гиперссылку из тега <li class="next">, который находится в теге <nav id="show-more-documents">**

**<nav id="show-more-documents"** role="navigation">

<ul class="previous-next-navigation">

**<li class="next">**

<a href="[/government/announcements?announcement\_filter\_option=press-releases&amp;departments%5B%5D=all&amp;from\_date=01%2F01%2F2015&amp;keywords=&amp;page=2&amp;to\_date=&amp;topics%5B%5D=all](view-source:https://www.gov.uk/government/announcements?announcement_filter_option=press-releases&departments%5B%5D=all&from_date=01%2F01%2F2015&keywords=&page=2&to_date=&topics%5B%5D=all)">Next <span class="visuallyhidden">page</span> <span class="page-numbers">2 of 93</span></a>

</li>

</ul>

</nav>

# # Зададим временной интервал 1 сентября – 30 октября 2018 г.

# new\_results <-'government/announcements?keywords=&taxons%5B%5D=all&subtaxons%5B%5D=all&announcement\_filter\_option=press-releases&departments%5B%5D=all&people%5B%5D=all&world\_locations%5B%5D=all&from\_date=01%2F09%2F2018&to\_date=30%2F10%2F2018'

# # создаем объект для работы по протоколу SSL

# signatures = system.file("CurlSSL", cainfo = "cacert.pem",

# package = "RCurl")

# all\_links <- character()

# # Перебираем страницы

# while(length(new\_results) > 0)

# {

# new\_results <- str\_c("https://www.gov.uk/", new\_results)

# results <- getURL(new\_results, cainfo = signatures)

# results\_tree <- htmlParse(results)

# # берем со страницы все гиперссылки на пресс-релизы

# all\_links <- c(all\_links, xpathSApply(results\_tree,

# "//li[@id]//a", xmlGetAttr, "href"))

# # новая страница для загрузки

# new\_results <- xpathSApply(results\_tree,

# "//nav[@id='show-more-documents']//li[@class='next']//a",

# xmlGetAttr, "href")

# }

# dir.create("Press\_Releases")

# # Теперь загружаем все страницы на диск

# for(i in 1:length(all\_links)){

# url <- str\_c("https://www.gov.uk", all\_links[i])

# tmp <- getURL(url, cainfo = signatures)

# write(tmp, str\_c("Press\_Releases/", i, ".html"))

# Sys.sleep(1)

# }

# # Читаем первый пресс-релиз

# tmp <- readLines("Press\_Releases/1.html")

# # Убираем переносы строк

# tmp <- str\_c(tmp, collapse = "")

# # Строим HTML-объект

# tmp <- htmlParse(tmp)

**Текст пресс-релиза:**

<div class="gem-c-govspeak govuk-govspeak direction-ltr">

**<div class="govspeak">**

<p>Business Secretary Greg Clark and Greater Manchester’s Deputy Mayor for the Economy Richard Leese have today (30 October) set out progress towards a Greater Manchester Local Industrial Strategy, one of the UK’s first Local Industrial Strategies.</p>

<p>Developed in collaboration with regional partners, the Local Industrial Strategy for Greater Manchester will harness the regions distinctive strengths and build on them in areas like health research and advanced materials to unlock growth and boost people’s earning power across the region’s cities and places.</p>

<p>Once launched, the Local Industrial Strategy will capitalise on Greater Manchester’s long and established history of entrepreneurship and enterprise, its devolved health and social care budget of £6 billion per annum and its position as home to one of the largest graduate pools in Europe.</p>

<p>The <a href="[https://www.gov.uk/government/publications/local-industrial-strategy-greater-manchester-progress-statement](view-source:https://www.gov.uk/government/publications/local-industrial-strategy-greater-manchester-progress-statement)">progress statement</a> published today showcases the work that has been done so far by local leaders in the Northern Powerhouse who are working with the government and businesses to identify economic opportunities that the region can seize in the years to come.</p>

<p>Business Secretary Greg Clark said:</p>. . .

**Информация об организации:**

<div class="app-c-publisher-metadata\_\_other">

<dl data-module="track-click">

<dt class="app-c-publisher-metadata\_\_term">

From:

</dt>

<dd class="app-c-publisher-metadata\_\_definition" data-module="gem-toggle">

<**span class="app-c-publisher-metadata\_\_definition-sentence">**

<**a** href="[/government/organisations/ministry-of-housing-communities-and-local-government](view-source:https://www.gov.uk/government/organisations/ministry-of-housing-communities-and-local-government)">Ministry of Housing, Communities &amp; Local Government</a>, <a href="[/government/organisations/department-for-business-energy-and-industrial-strategy](view-source:https://www.gov.uk/government/organisations/department-for-business-energy-and-industrial-strategy)">Department for Business, Energy &amp; Industrial Strategy</a>, <a href="[/government/people/jake-berry](view-source:https://www.gov.uk/government/people/jake-berry)">Jake Berry MP</a>, and <a href="[/government/people/greg-clark](view-source:https://www.gov.uk/government/people/greg-clark)">The Rt Hon Greg Clark MP</a>

</span>

</dd>

</dl>

</div>

**Дата публикации:**

<**div class="app-c-published-dates "**>

Published 30 October 2018

</div>

release <- xpathSApply(tmp, "//div[@class='govspeak']", xmlValue)

organisation <- xpathSApply(tmp, "//span[@class='app-c-publisher-metadata\_\_definition-sentence']/a", xmlValue)

publication <- xpathSApply(tmp, "//div[@class='app-c-published-dates ']", xmlValue)

# # пакет Text mining

# install.packages("tm")

# library(tm)

# # Создаем объект корпус

# release\_corpus <- VCorpus(VectorSource(release))

# # Добавляем метаинформацию

# meta(release\_corpus[[1]], "organisation") <- organisation[1]

# meta(release\_corpus[[1]], "publication") <- publication

# meta(release\_corpus[[1]])

author : character(0)

datetimestamp: 2018-10-31 20:37:12

description : character(0)

heading : character(0)

id : 1

language : en

origin : character(0)

organisation : Attorney General's Office

publication : c(" Published 30 October 2018", " Published 30 October 2018")

# # Добавляем в корпус все остальные документы (выбрали 3 министерства)

n <- 1

for(i in 2:length(list.files("Press\_Releases/"))){

tmp <- readLines(str\_c("Press\_Releases/", i, ".html"))

tmp <- str\_c(tmp, collapse = "")

tmp <- htmlParse(tmp)

release <- xpathSApply(tmp, "//div[@class='govspeak']", xmlValue)

organisation <- xpathSApply(tmp, "//span[@class='app-c-publisher-metadata\_\_definition-sentence']/a", xmlValue)

publication <- xpathSApply(tmp, "//div[@class='app-c-published-dates ']", xmlValue)

if (length(release)!=0

&(organisation == 'Attorney General\'s Office' |

organisation == 'Highways England' |

organisation == 'Foreign & Commonwealth Office')

) {

n <- n + 1

tmp\_corpus <- VCorpus(VectorSource(release))

release\_corpus <- c(release\_corpus, tmp\_corpus)

meta(release\_corpus[[n]], "organisation") <- organisation[1]

cat("n=",n)

}

}

release\_corpus

<<VCorpus>>

Metadata: corpus specific: 0, document level (indexed): 0

Content: documents: 126

# Список министерств, выпустивших пресс-релизы

meta\_data<- data.frame()

for (i in 1:NROW(release\_corpus))

{

meta\_data [i, "organisation"] <- meta(release\_corpus[[i]], "organisation")

meta\_data [i, "num"] <- i

}

table(as.character(meta\_data[, "organisation"]))

Attorney General's Office Foreign & Commonwealth Office

20 75

Highways England

31

# Теперь выполняем преобразования данных (функция content\_transformer применяется для того,чтобы не менялся тип данных)

# Удаляем числа

release\_corpus <- tm\_map(release\_corpus, content\_transformer(removeNumbers))

# Знаки пунктуации заменяем на пробелы

release\_corpus <- tm\_map(release\_corpus,

content\_transformer(str\_replace\_all),

pattern = "[[:punct:]]", replacement = " ")

release\_corpus[[1]]$content

[1] "The visit will showcase the breadth and depth of the relationship between the UK and The Netherlands including through trade innovation and security The state visit is an excellent opportunity to celebrate and deepen our connections Our relationship may be changing but we are confident that the Netherlands and the UK will continue to have a close relationship and work together to tackle global issues The UK Foreign Secretary Jeremy Hunt will also use the opportunity to meet with his Dutch counterpart Stef Blok Foreign Secretary Jeremy Hunt said The UK and The Netherlands have a rich long history between our people and of course our Royal Families So it is a pleasure to welcome His Majesty King Willem Alexander and Queen Maxima to the United Kingdom reflecting our enduring commitment to this partnership Together we work to further peace and security across the globe tackling threats from terrorism and cyber and promoting values of freedom I look forward to celebrating o... <truncated>

# Удаляем стоп-слова

release\_corpus <- tm\_map(release\_corpus, content\_transformer(removeWords), words = stopwords("en"))

release\_corpus[[1]]$content

[1] "The visit will showcase breadth depth relationship UK The Netherlands including trade innovation security The state visit excellent opportunity celebrate deepen connections Our relationship may changing confident Netherlands UK will continue close relationship work together tackle global issues The UK Foreign Secretary Jeremy Hunt will also use opportunity meet Dutch counterpart Stef Blok Foreign Secretary Jeremy Hunt said The UK The Netherlands rich long history people course Royal Families So pleasure welcome His Majesty King Willem Alexander Queen Maxima United Kingdom reflecting enduring commitment partnership Together work peace security across globe tackling threats terrorism cyber promoting values freedom I look forward celebrating connections shared successes Foreign Minister Stef Blok looking ensuring continues years come From th century present day trade relations countries continue flo... <truncated>

# Преобразуем к нижнему регистру

release\_corpus <- tm\_map(release\_corpus, content\_transformer(tolower))

release\_corpus[[1]]$content

[1] "the visit will showcase breadth depth relationship uk the netherlands including trade innovation security the state visit excellent opportunity celebrate deepen connections our relationship may changing confident netherlands uk will continue close relationship work together tackle global issues the uk foreign secretary jeremy hunt will also use opportunity meet dutch counterpart stef blok foreign secretary jeremy hunt said the uk the netherlands rich long history people course royal families so pleasure welcome his majesty king willem alexander queen maxima united kingdom reflecting enduring commitment partnership together work peace security across globe tackling threats terrorism cyber promoting values freedom i look forward celebrating connections shared successes foreign minister stef blok looking ensuring continues years come from th century present day trade relations countries continue flo... <truncated>

# Проводим стемминг

release\_corpus <- tm\_map(release\_corpus, stemDocument, language = "english")

release\_corpus[[1]]$content

[1] "the visit will showcas breadth depth relationship uk the netherland includ trade innov secur the state visit excel opportun celebr deepen connect our relationship may chang confid netherland uk will continu close relationship work togeth tackl global issu the uk foreign secretari jeremi hunt will also use opportun meet dutch counterpart stef blok foreign secretari jeremi hunt said the uk the netherland rich long histori peopl cours royal famili so pleasur welcom his majesti king willem alexand queen maxima unit kingdom reflect endur commit partnership togeth work peac secur across globe tackl threat terror cyber promot valu freedom i look forward celebr connect share success foreign minist stef blok look ensur continu year come from th centuri present day trade relat countri continu flourish the netherland uk s third largest trade partner onli year heineken announc invest million creat job uk fellow dutch compani newcold invest million new distribut centr wakefield ad job next yea... <truncated>

# Создаем матрицу терминов-документов

tdm <- TermDocumentMatrix(release\_corpus)

tdm

<<TermDocumentMatrix (terms: 4067, documents: 126)>>

Non-/sparse entries: 17826/494616

Sparsity : 97%

Maximal term length: 24

Weighting : term frequency (tf)

# Можно фильтровать документы по ключевым словам

(court <- tm\_filter(release\_corpus,

FUN = function(x) any(str\_detect(x, "court"))))

<<VCorpus>>

Metadata: corpus specific: 0, document level (indexed): 0

Content: documents: 23

# Можно отыскивать взаимосвязанные термины

findAssocs(tdm, "court", 0.9)

$`court`

appeal sentenc crown

0.97 0.93 0.91

# Создаем матрицу документов-терминов

dtm <- DocumentTermMatrix(release\_corpus)

dtm

<<DocumentTermMatrix (documents: 126, terms: 4067)>>

Non-/sparse entries: 17826/494616

Sparsity : 97%

Maximal term length: 24

Weighting : term frequency (tf)

dtm <- removeSparseTerms(dtm, 0.9)

dtm

<<DocumentTermMatrix (documents: 126, terms: 348)>>

Non-/sparse entries: 8919/34929

Sparsity : 80%

Maximal term length: 19

Weighting : term frequency (tf)

# Дополнительный пакет для работы с текстами

install.packages("RTextTools")

library(RTextTools)

# Массив меток – названий министерств

org\_labels<-meta\_data[, "organisation"]

N <- length(org\_labels)

# Для применения методов классификации создаем контейнер

container <- create\_container(

dtm,

labels = org\_labels,

trainSize = 1:100,

testSize = 101:N,

virgin = FALSE

)

# Запускаем обучение моделей

# Support vector machine

svm\_model <- train\_model(container, "SVM")

# Random forest

tree\_model <- train\_model(container, "TREE")

# Maximal entropy

maxent\_model <- train\_model(container, "MAXENT")

# Запускаем классификацию

svm\_out <- classify\_model(container, svm\_model)

tree\_out <- classify\_model(container, tree\_model)

maxent\_out <- classify\_model(container, maxent\_model)

# Создаем общую матрицу меток

labels\_out <- data.frame(

correct\_label = org\_labels[101:N],

svm = as.character(svm\_out[,1]),

tree = as.character(tree\_out[,1]),

maxent = as.character(maxent\_out[,1]),

stringsAsFactors = F)

## SVM performance

table(labels\_out[,1] == labels\_out[,2])

TRUE

26

## Random forest performance

table(labels\_out[,1] == labels\_out[,3])

TRUE

26

## Maximum entropy performance

table(labels\_out[,1] == labels\_out[,4])

TRUE

26

# Запустим классификацию для 4 министерств:

'Attorney General\'s Office'

'Highways England'

'Department for International Development'

'Foreign & Commonwealth Office'

> ## SVM performance

> table(labels\_out[,1] == labels\_out[,2])

FALSE TRUE

1 26

>

> ## Random forest performance

> table(labels\_out[,1] == labels\_out[,3])

FALSE TRUE

2 25

>

> ## Maximum entropy performance

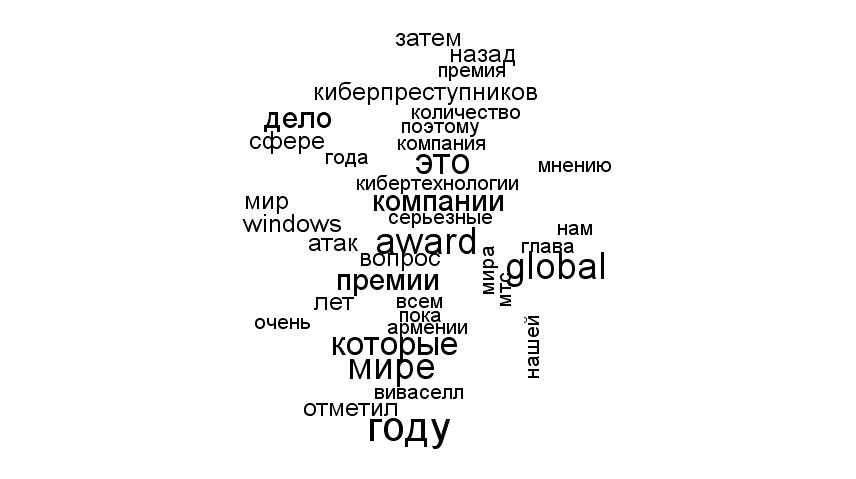
> table(labels\_out[,1] == labels\_out[,4])

FALSE TRUE

1 26

1. **Стемминг**

Создание облака слов без стемминга



Создание облака слов с обычным стеммингом из пакета tm (слова сокращаются)



Создание облака слов со стеммингом от Яндекс

<https://tech.yandex.ru/mystem/doc/index-docpage/>

<http://matrunich.com/blog/2014/05/12/wordcloud/>



Что такое «пок» и «армений»?

1. **Анализ тональности (sentiment analysis).**

**Примеры:**

[**https://www.paralleldots.com/sentiment-analysis**](https://www.paralleldots.com/sentiment-analysis)

[**https://www.csc2.ncsu.edu/faculty/healey/tweet\_viz/tweet\_app/**](https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/)

**С чего начать? При использовании подхода, основанного на словарях, загружаем словари, например:**

[**http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar**](http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar)

pos <- readLines("positive-words.txt")

# удаляем комментарии

pos <- pos[!str\_detect(pos, ";")]

pos <- pos[2:length(pos)]

neg <- readLines("negative-words.txt")

# удаляем комментарии

neg <- neg[!str\_detect(neg, ";")]

neg <- neg[2:length(neg)]

pos[1:100]

[1] "a+" "abound" "abounds" "abundance" "abundant"

[6] "accessable" "accessible" "acclaim" "acclaimed" "acclamation"

[11] "accolade" "accolades" "accommodative" "accomodative" "accomplish"

[16] "accomplished" "accomplishment" "accomplishments" "accurate" "accurately"

[21] "achievable" "achievement" "achievements" "achievible" "acumen"

[26] "adaptable" "adaptive" "adequate" "adjustable" "admirable"

[31] "admirably" "admiration" "admire" "admirer" "admiring"

[36] "admiringly" "adorable" "adore" "adored" "adorer"

[41] "adoring" "adoringly" "adroit" "adroitly" "adulate"

[46] "adulation" "adulatory" "advanced" "advantage" "advantageous"

[51] "advantageously" "advantages" "adventuresome" "adventurous" "advocate"

[56] "advocated" "advocates" "affability" "affable" "affably"

[61] "affectation" "affection" "affectionate" "affinity" "affirm"

[66] "affirmation" "affirmative" "affluence" "affluent" "afford"

[71] "affordable" "affordably" "afordable" "agile" "agilely"

[76] "agility" "agreeable" "agreeableness" "agreeably" "all-around"

[81] "alluring" "alluringly" "altruistic" "altruistically" "amaze"

[86] "amazed" "amazement" "amazes" "amazing" "amazingly"

[91] "ambitious" "ambitiously" "ameliorate" "amenable" "amenity"

[96] "amiability" "amiabily" "amiable" "amicability" "amicable"

neg[1:100]

[1] "2-faced" "2-faces" "abnormal" "abolish" "abominable"

[6] "abominably" "abominate" "abomination" "abort" "aborted"

[11] "aborts" "abrade" "abrasive" "abrupt" "abruptly"

[16] "abscond" "absence" "absent-minded" "absentee" "absurd"

[21] "absurdity" "absurdly" "absurdness" "abuse" "abused"

[26] "abuses" "abusive" "abysmal" "abysmally" "abyss"

[31] "accidental" "accost" "accursed" "accusation" "accusations"

[36] "accuse" "accuses" "accusing" "accusingly" "acerbate"

[41] "acerbic" "acerbically" "ache" "ached" "aches"

[46] "achey" "aching" "acrid" "acridly" "acridness"

[51] "acrimonious" "acrimoniously" "acrimony" "adamant" "adamantly"

[56] "addict" "addicted" "addicting" "addicts" "admonish"

[61] "admonisher" "admonishingly" "admonishment" "admonition" "adulterate"

[66] "adulterated" "adulteration" "adulterier" "adversarial" "adversary"

[71] "adverse" "adversity" "afflict" "affliction" "afflictive"

[76] "affront" "afraid" "aggravate" "aggravating" "aggravation"

[81] "aggression" "aggressive" "aggressiveness" "aggressor" "aggrieve"

[86] "aggrieved" "aggrivation" "aghast" "agonies" "agonize"

[91] "agonizing" "agonizingly" "agony" "aground" "ail"

[96] "ailing" "ailment" "aimless" "alarm" "alarmed"

**Проводим стемминг словарей и удаляем дубликаты**

pos <- stemDocument(pos, language = "english")

pos <- pos[!duplicated(pos)]

neg <- stemDocument(neg, language = "english")

neg <- neg[!duplicated(neg)]

pos[1:100]

[1] "a+" "abound" "abund" "access" "acclaim" "acclam"

[7] "accolad" "accommod" "accomod" "accomplish" "accur" "achiev"

[13] "acumen" "adapt" "adequ" "adjust" "admir" "ador"

[19] "adroit" "adul" "adulatori" "advanc" "advantag" "adventuresom"

[25] "adventur" "advoc" "affabl" "affect" "affection" "affin"

[31] "affirm" "affluenc" "affluent" "afford" "aford" "agil"

[37] "agreeabl" "all-around" "allur" "altruist" "amaz" "ambiti"

[43] "amelior" "amen" "amiabl" "amiabili" "amic" "amiti"

[49] "ampl" "ampli" "amus" "angel" "apotheosi" "appeal"

[55] "applaud" "appreci" "appropri" "approv" "ardent" "ardor"

[61] "articul" "aspir" "assur" "astonish" "astound" "astut"

[67] "attent" "attract" "attun" "audibl" "auspici" "authent"

[73] "authorit" "autonom" "avail" "aver" "avid" "award"

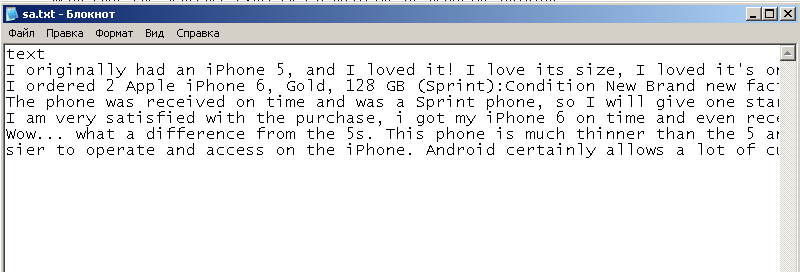
[79] "awe" "awesom" "awestruck" "awsom" "backbon" "balanc"

[85] "bargain" "beauteous" "beauti" "beautifullli" "beautifi" "beckon"

[91] "believ" "belov" "benefactor" "benefic" "benefici" "beneficiari"

[97] "benefit" "benevol" "benifit" "best"

Читаем данные из файла



data<-read.table("sa.txt", header=TRUE,sep=";", encoding="UTF-8", quote="")

# Создаем корпус

reviews <- Corpus(VectorSource(data$text))

reviews[[1]]$content

[1] "I originally had an iPhone 5, and I loved it! I love its size, I loved it's onductivity, but as soon as I came into contact with the iPhone 6, the game was over. I was worried hat it might be a little too big for my taste, but the more I hold it, the more I love it. The responsiveness of the screen, the reaction time of the fingerprint scanner, everything works like a dream. Also, after making multiple phone calls, I have noticed that there's a perfect clarity and that no one has trouble hearing me. You will not be sorry you picked this phone. "

# Выполняем преобразования

reviews <- tm\_map(reviews, content\_transformer(removeNumbers))

reviews[[1]]$content

[1] "I originally had an iPhone , and I loved it! I love its size, I loved it's onductivity, but as soon as I came into contact with the iPhone , the game was over. I was worried hat it might be a little too big for my taste, but the more I hold it, the more I love it. The responsiveness of the screen, the reaction time of the fingerprint scanner, everything works like a dream. Also, after making multiple phone calls, I have noticed that there's a perfect clarity and that no one has trouble hearing me. You will not be sorry you picked this phone. "

reviews <- tm\_map(reviews, content\_transformer(str\_replace\_all), pattern = "[[:punct:]]", replacement = " ")

reviews[[1]]$content

[1] "I originally had an iPhone and I loved it I love its size I loved it s onductivity but as soon as I came into contact with the iPhone the game was over I was worried hat it might be a little too big for my taste but the more I hold it the more I love it The responsiveness of the screen the reaction time of the fingerprint scanner everything works like a dream Also after making multiple phone calls I have noticed that there s a perfect clarity and that no one has trouble hearing me You will not be sorry you picked this phone "

reviews <- tm\_map(reviews, content\_transformer(tolower))

reviews[[1]]$content

[1] "i originally had an iphone and i loved it i love its size i loved it s onductivity but as soon as i came into contact with the iphone the game was over i was worried hat it might be a little too big for my taste but the more i hold it the more i love it the responsiveness of the screen the reaction time of the fingerprint scanner everything works like a dream also after making multiple phone calls i have noticed that there s a perfect clarity and that no one has trouble hearing me you will not be sorry you picked this phone "

reviews <- tm\_map(reviews, content\_transformer(removeWords), words = stopwords("en"))

reviews[[1]]$content

[1] " originally iphone loved love size loved s onductivity soon came contact iphone game worried hat might little big taste hold love responsiveness screen reaction time fingerprint scanner everything works like dream also making multiple ph one calls noticed s perfect clarity one trouble hearing will sorry picked phone "

reviews <- tm\_map(reviews, stemDocument, language = "english")

reviews[[1]]$content

[1] " origin iphon love love size love s onduct soon came contact iphon game worri hat might littl big tast hold love respons screen reaction time fingerprint scanner everyth work like dream also make multipl phone call notic s perfect clariti one troubl hear will sorri pick phone "

reviews

# Создаем матрицу терминов-документов

tdm <- TermDocumentMatrix(reviews, control = list(weighting = weightBin))

# Подсчитываем количество положительных и отрицательных терминов и разность этих значений

pos.mat <- tdm[rownames(tdm) %in% pos, ]

neg.mat <- tdm[rownames(tdm) %in% neg, ]

pos.out <- apply(pos.mat, 2, sum)

neg.out <- apply(neg.mat, 2, sum)

senti.diff <- pos.out - neg.out

senti.diff[senti.diff == 0] <- NA

senti.diff

1 2 3 4 5

6 4 5 6 15