ДОДАТОК А
АЛГОРИТМ БЛОКУВАННЯ САЙТІВ НА ПК

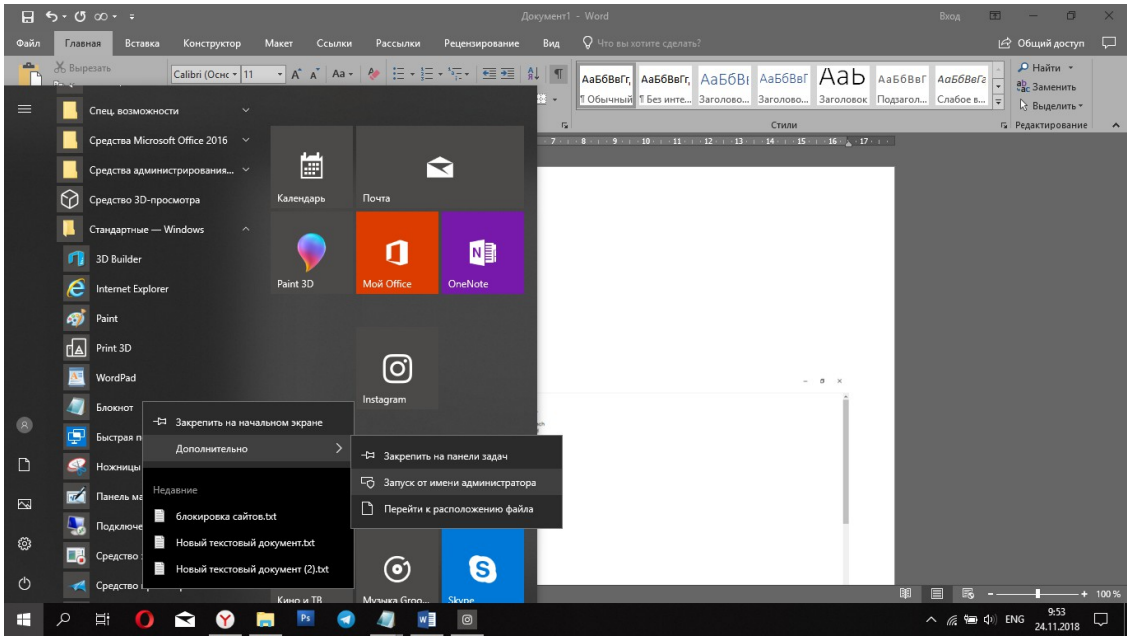


Рисунок А.1- Інтерфейс «Відкриття блокноту з правами Адміністратора в ОС Windows 10»

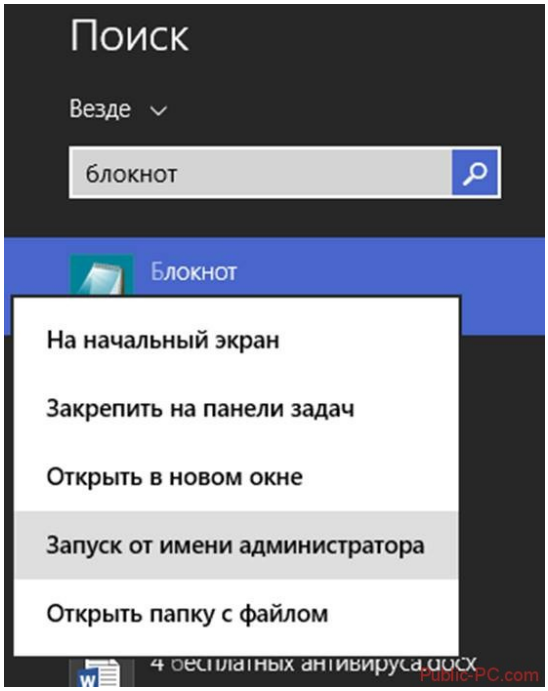


Рисунок А.2- Інтерфейс «Відкриття блокноту з правами Адміністратора в ОС Windows 8/8.1»

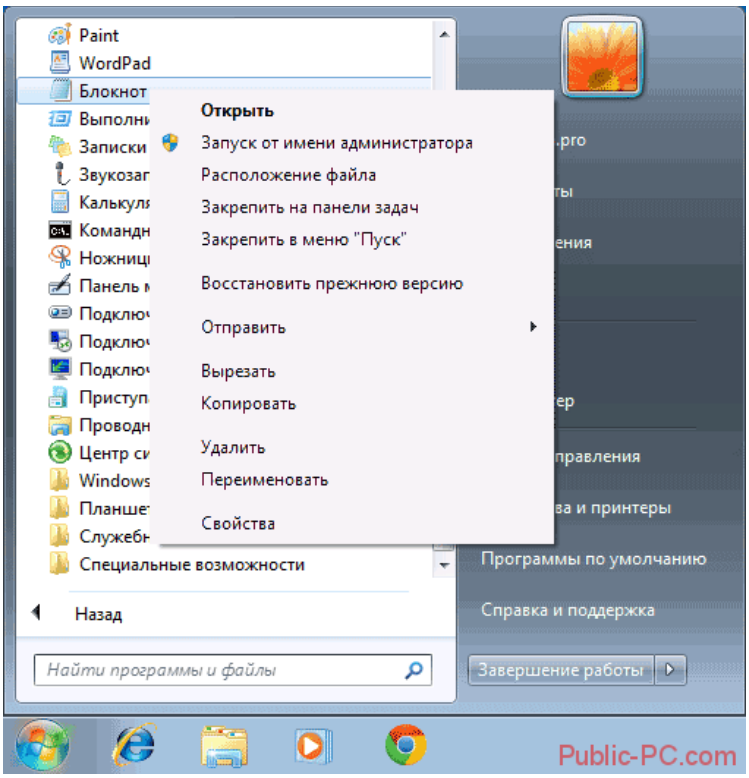


Рисунок А.3- Інтерфейс «Відкриття блокноту з правами Адміністратора в ОС Windows 7»

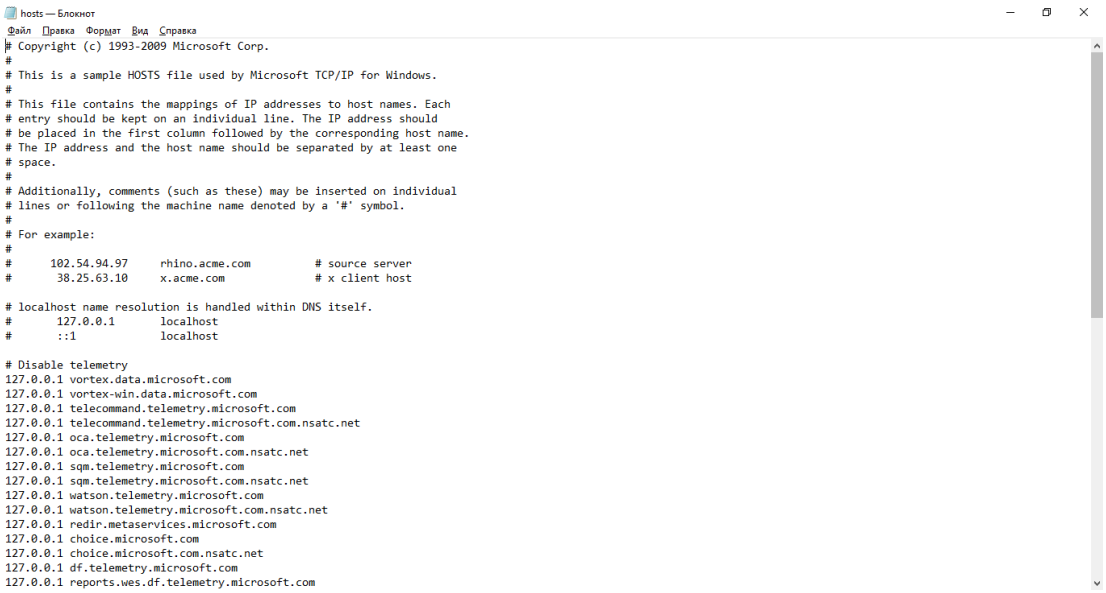


Рисунок А.4- Інтерфейс «файл «host»»

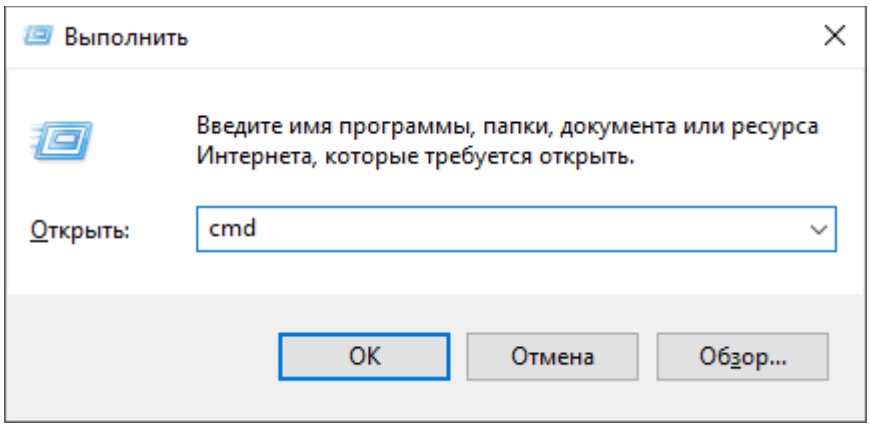


Рисунок А.5 – Інтерфейс «строка «Виконати»»

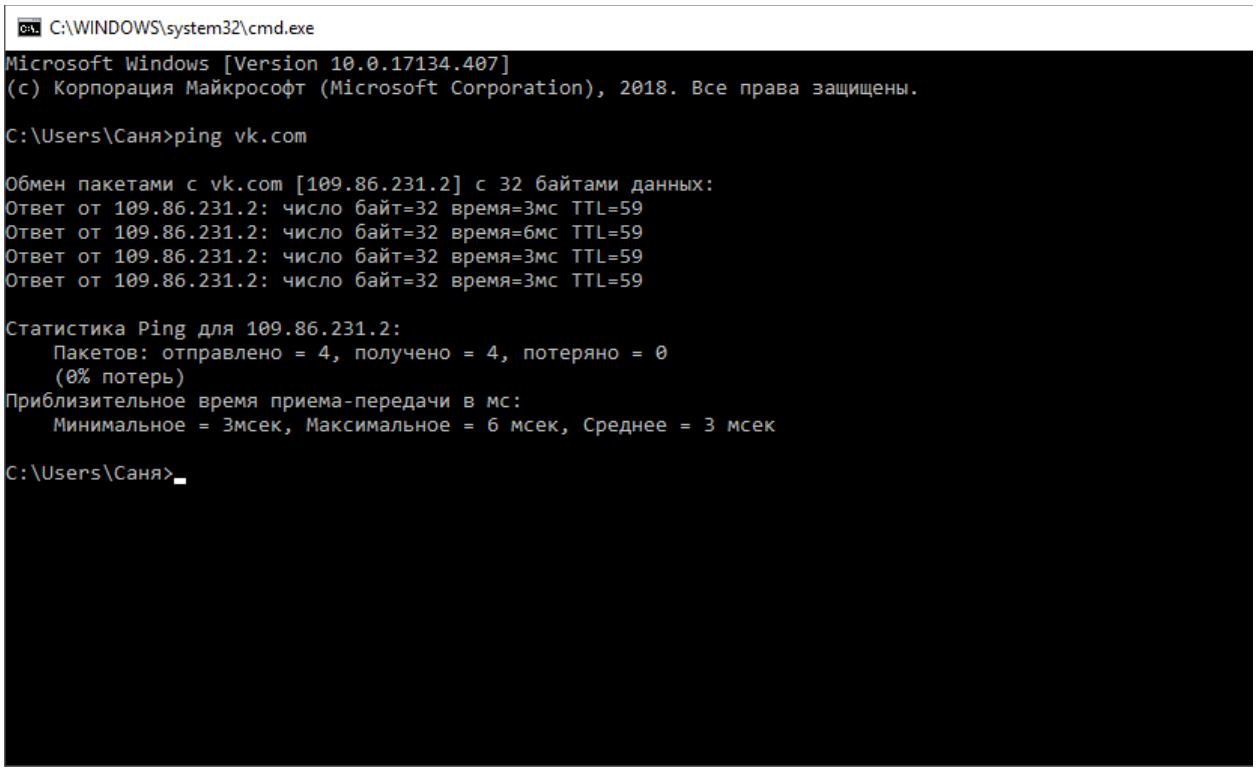


Рисунок А.6 – Інтерфейс «строка «Виконати»»

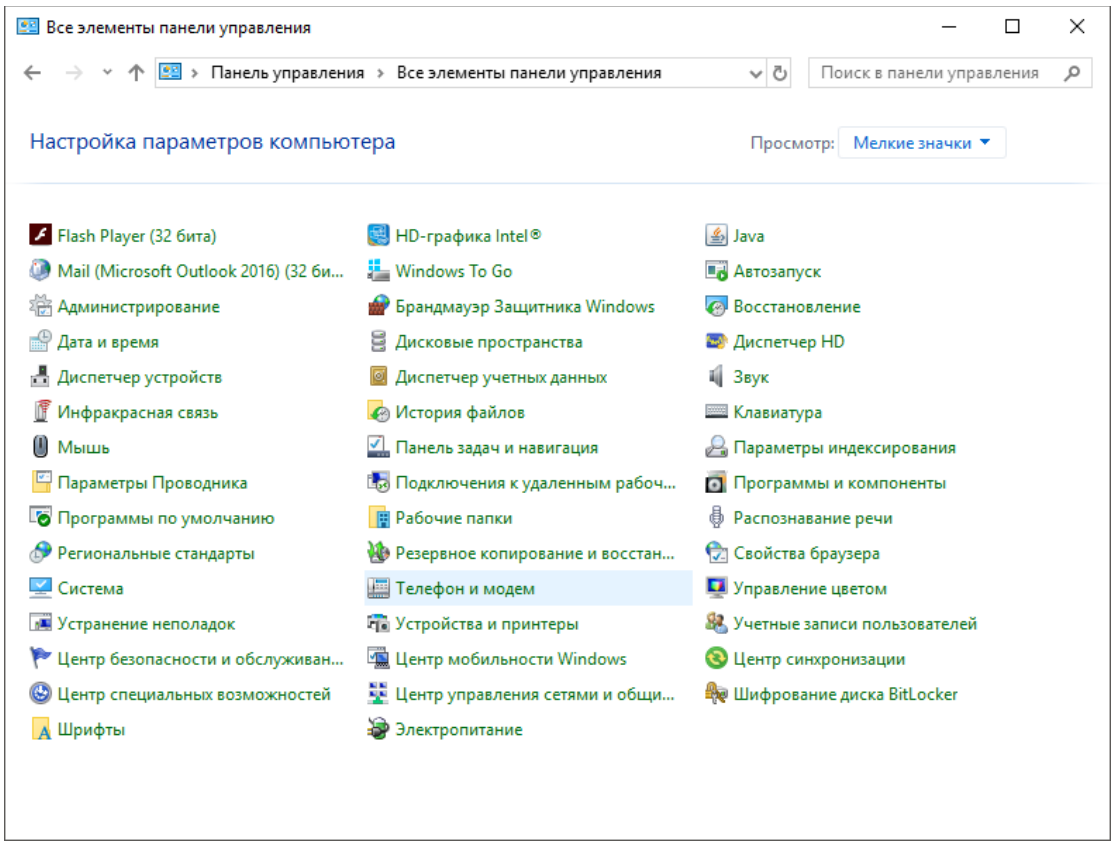


Рисунок А.7 - Інтерфейс «Панель управління»

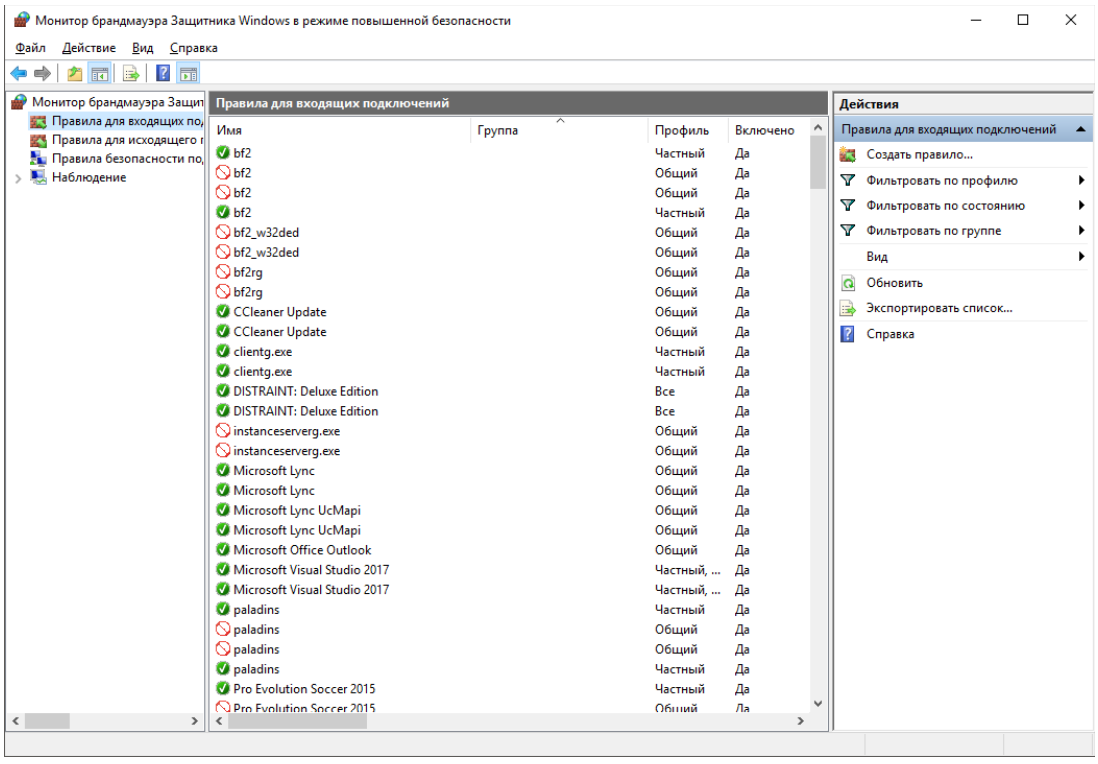


Рисунок А.8 - Інтерфейс «Брандмауер Windows»

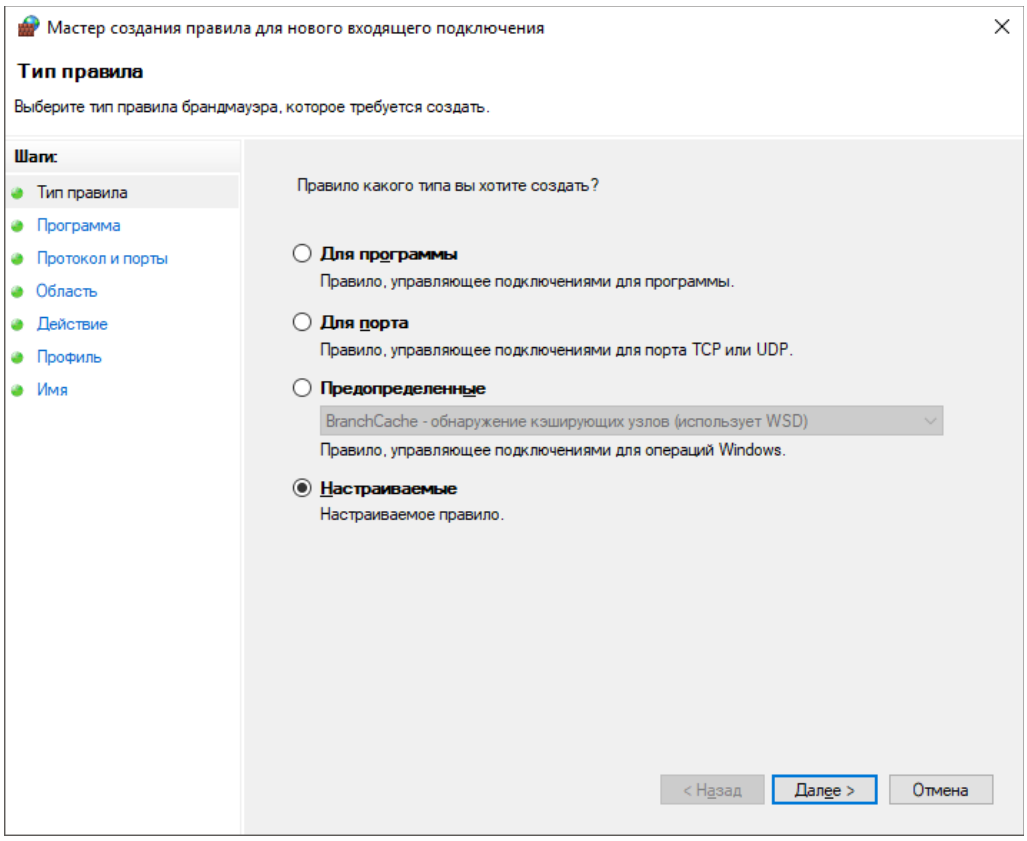


Рисунок А.9- Інтерфейс «Додача правила для брандмауера»

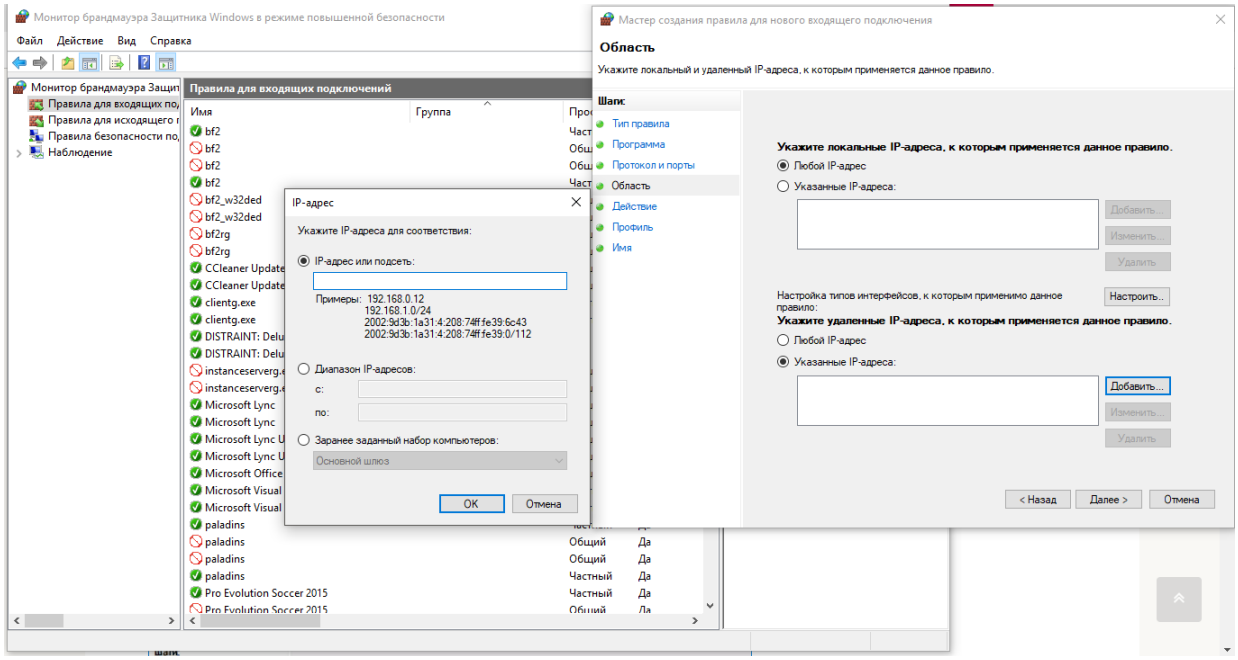


Рисунок А.10- Інтерфейс «Вказання IP – адресу Інтерфей»

Додаток Б

КОД ПРОГРАМИ ДЛЯ АНАЛІЗУ ТЕКСТУ

```
> Sys.setlocale(locale = "Russian")
[1]
> cname <- file.path("~", "Desktop", "texts")
> cname
[1] "~/Desktop/texts"
> dir(cname)
character(0)
> cname <- file.path("C:", "texts")
> cname
[1] "C:/texts"
> dir(cname)
[1] "Информационная безопасность.txt"    "История.txt"
[3] "Объём понятия.txt"                  "Принципы.txt"
[5] "Программно-аппаратное средство.txt"
> cname <- file.path("C:", "texts")
> cname
[1] "C:/texts"
> dir(cname)
[1] "Информационная безопасность.txt"    "История.txt"
[3] "Объём понятия.txt"                  "Принципы.txt"
[5] "Программно-аппаратное средство.txt"
> library(tm)
Загрузка требуемого пакета: NLP
> docs <- VCorpus(DirSource(cname))
> summary(docs)
                                Length Class            Mode
Информационная безопасность.txt    2      PlainTextDocument list
История.txt                        2      PlainTextDocument list
Объём понятия.txt                  2      PlainTextDocument list
Принципы.txt                       2      PlainTextDocument list
Программно-аппаратное средство.txt 2      PlainTextDocument list
> inspect(docs[1])
<<VCorpus>>
Metadata:  corpus specific: 0, document level (indexed): 0
Content:  documents: 1

[[1]]
<<PlainTextDocument>>
Metadata:  7
Content:  chars: 9645

> writeLines(as.character(docs[1]))
list(list(content = c("Информационная безопасность (англ. Information Security, а также — англ. InfoSec) — практика предотвращения несанкционированного доступа, использования, раскрытия, искажения, изменения, исследования, записи или уничтожения информации. Это универсальное понятие применяется вне зависимости от формы, которую могут принимать данные (электронная или, например, физическая). Основная задача информационной безопасности — сбалансированная защита конфиденциальности, целостности и доступности данных[1], с учётом целесообразности применения и без какого-либо ущерба
```

производительности организации[2]. Это достигается, в основном, посредством многоэтапного процесса управления рисками, который позволяет идентифицировать основные средства и нематериальные активы, источники угроз, уязвимости, потенциальную степень воздействия и возможности управления рисками. Этот процесс сопровождается оценкой эффективности плана по управлению рисками[3].",
"", "Для того, чтобы стандартизовать эту деятельность, научное и профессиональное сообщества находятся в постоянном сотрудничестве, направленном на выработку базовой методологии, политик и индустриальных стандартов в области технических мер защиты информации, юридической ответственности, а также стандартов обучения пользователей и администраторов. Эта стандартизация в значительной мере развивается под влиянием широкого спектра законодательных и нормативных актов, которые регулируют способы доступа, обработки, хранения и передачи данных. Однако внедрение любых стандартов и методологий в организации может иметь лишь поверхностный эффект, если культура непрерывного совершенствования[en] не привита должным образом[4].",
"Общие сведения", "В основе информационной безопасности лежит деятельность по защите информации — обеспечению её конфиденциальности, доступности и целостности, а также недопущению какой-либо компрометации в критической ситуации[5]. К таким ситуациям относятся природные, техногенные и социальные катастрофы, компьютерные сбои, физическое похищение и тому подобные явления. В то время, как делопроизводство большинства организаций в мире до сих пор основано на бумажных документах[6], требующих соответствующих мер обеспечения информационной безопасности, наблюдается неуклонный рост числа инициатив по внедрению цифровых технологий на предприятиях[7] [8], что влечёт за собой привлечение специалистов по безопасности информационных технологий (ИТ) для защиты информации. Эти специалисты обеспечивают информационную безопасность технологии (в большинстве случаев — какой-либо разновидности компьютерных систем). Следует отметить, что под компьютером в данном контексте подразумевается не только бытовой персональный компьютер, а цифровые устройства любой сложности и назначения, начиная от примитивных и изолированных, наподобие электронных калькуляторов и бытовых приборов, вплоть до индустриальных систем управления и суперкомпьютеров, объединённых компьютерными сетями. Крупнейшие предприятия и организации, в силу жизненной важности и ценности информации для их бизнеса, нанимают специалистов по информационной безопасности, как правило, себе в штат. В их задачи входит обезопасить все технологии от вредоносных кибератак, зачастую нацеленных на похищение важной конфиденциальной информации или на перехват управления внутренними системами организации.",
"", "Информационная безопасность, как сфера занятости, значительно развилась и выросла в последние годы. В ней возникло множество профессиональных специализаций, например, таких, как безопасность сетей и связанной инфраструктуры, защиты программного обеспечения и баз данных, аудит информационных систем, планирование непрерывности бизнеса, выявление электронных записей и компьютерная криминалистика[en]. Профессионалы информационной безопасности имеют весьма стабильную занятость и высокий спрос на рынке труда. Масштабные исследования, проведённые организацией (ISC)? показали, что на 2017 год 66 % руководителей информационной безопасности признали острую нехватку рабочей силы в своих подразделениях, а по прогнозам к 2022 году недостаток специалистов в этой области составит по всему миру 1 800 000 человек[9].",
"", "Угрозы и меры противодействия", "Угрозы информационной безопасности могут принимать весьма разнообразные формы. На 2018 год наиболее серьёзными считаются угрозы связанные с «преступлением как услугой» (англ. Crime-as-a-Service), Интернетом вещей, цепями поставок и усложнением требований регуляторов[10]. «Преступление как услуга» представляет собой модель предоставления зрелыми преступными сообществами пакетов криминальных услуг на даркнет-рынке по доступным ценам начинающим киберпреступникам[en][К 1]. Это позволяет последним совершать хакерские атаки, ранее

недоступные из-за высокой технической сложности или дороговизны, делая киберпреступность массовым явлением[12]. Организации активно внедряют Интернет вещей, устройства которого зачастую спроектированы без учёта требований безопасности, что открывает дополнительные возможности для атаки. К тому же, быстрое развитие и усложнение Интернета вещей снижает его прозрачность, что в сочетании с нечётко определёнными правовыми нормами и условиями позволяет организациям использовать собранные устройствами персональные данные своих клиентов по собственному усмотрению без их ведома. Кроме того, для самих организаций проблематично отслеживать, какие из собранных устройствами Интернета вещей данных передаются во вне. Угроза цепей поставок состоит в том, что организации, как правило, передают своим поставщикам разнообразную ценную и конфиденциальную информацию, в результате чего теряют непосредственный контроль над ней. Таким образом, значительно возрастает риск нарушения конфиденциальности, целостности или доступности этой информации. Всё новые и новые требования регуляторов значительно осложняют управление жизненно-важными информационными активами организаций. Например, введённый в действие в 2018 году в Евросоюзе Общий регламент защиты персональных данных (англ. General Data Protection Regulation, GDPR), требует от любой организации в любой момент времени на любом участке собственной деятельности или цепи поставок, продемонстрировать, какие персональные данные и для каких целей имеются там в наличии, как они обрабатываются, хранятся и защищаются. Причём эта информация должна быть предоставлена не только в ходе проверок уполномоченными органами, но и по первому требованию частного лица — владельца этих данных. Соблюдение такого комплаенса требует отвлечения значительных бюджетных средств и ресурсов от других задач информационной безопасности организации. И хотя упорядочение обработки персональных данных предполагает в долгосрочной перспективе улучшение информационной безопасности, в краткосрочном плане риски организации заметно возрастают[10].",
"", "Большинство людей так, или иначе испытывают на себе воздействие угроз информационной безопасности. Например, становятся жертвами вредоносных программ (вирусов и червей, троянских программ, программ-вымогателей)[13], фишинга или кражи личности. Фишинг (англ. Phishing) представляет собой мошенническую попытку[К 2] завладения конфиденциальной информацией (например, учётной записью, паролем или данными кредитной карты). Обычно пользователя Интернета стараются заманить на мошеннический веб-сайт, неотличимый от оригинального сайта какой-либо организации (банка, интернет-магазина, социальной сети и т. п.)[14][15]. Как правило, такие попытки совершаются с помощью массовых рассылок поддельных электронных писем якобы от имени самой организации[16], содержащих ссылки на мошеннические сайты. Открыв такую ссылку в браузере, ничего не подозревающий пользователь вводит свои учётные данные, которые становятся достоянием мошенников[17]. Термин Identity Theft с англ.?—?«кража личности» появился в английском языке в 1964 году[18] для обозначения действий, в которых чьи-либо персональные данные (например, имя, учётная запись в банковской системе или номер кредитной карты, часто добытые с помощью фишинга) используются для мошенничества и совершения иных преступлений[19][20]. Тот, от чьего имени преступники получают незаконные финансовые преимущества, кредиты или совершают иные преступления, зачастую сам становится обвиняемым, что может иметь для него далеко идущие тяжёлые финансовые и юридические последствия[21]. Информационная безопасность оказывает непосредственное влияние на неприкосновенность частной жизни[22], определение которой в различных культурах может весьма разниться[23].",
"", "Органы государственной власти, вооружённые силы, корпорации, финансовые институты, медицинские учреждения и частные предприниматели постоянно накапливают значительные объёмы конфиденциальной информации о своих сотрудниках, клиентах, продуктах, научных исследованиях и финансовых результатах. Попадание такой информации в руки конкурентов или киберпреступников может повлечь для организации и

её клиентов далеко идущие юридические последствия, невосполнимые финансовые и репутационные потери. С точки зрения бизнеса информационная безопасность должна быть сбалансирована относительно затрат; экономическая модель Гордона-Лоба[en] описывает математический аппарат для решения этой задачи[24]. Основными способами противодействия угрозам информационной безопасности или информационным рискам являются:",
"", "снижение — внедрение мер безопасности и противодействия для устранения уязвимостей и предотвращения угроз;", "передача — перенос затрат, связанных с реализацией угроз на третьих лиц: страховые или аутсорсинговые компании;", "принятие — формирование финансовых резервов в случае, если стоимость реализации мер безопасности превышает потенциальный ущерб от реализации угрозы;", "отказ — отказ от чрезмерно рисковой деятельности[25]."), meta = list(author = character(0), datetimestamp = list(sec = 22.9859669208527,
    min = 2, hour = 19, mday = 14, mon = 0, year = 121, wday = 4, yday = 13, isdst = 0), description = character(0), heading = character(0), id = "Информационная безопасность.txt", language = "en", origin = character(0))))
list()
list()
> docs <- tm_map(docs,removePunctuation)
> writeLines(as.character(docs[1]))
list(list(content = c("Информационная безопасность англ Information Security а также — англ InfoSec — практика предотвращения несанкционированного доступа использования раскрытия искажения изменения исследования записи или уничтожения информации Это универсальное понятие применяется вне зависимости от формы которую могут принимать данные электронная или например физическая Основная задача информационной безопасности — сбалансированная защита конфиденциальности целостности и доступности данных1 с учётом целесообразности применения и без какоголибо ущерба производительности организации2 Это достигается в основном посредством многоэтапного процесса управления рисками который позволяет идентифицировать основные средства и нематериальные активы источники угроз уязвимости потенциальную степень воздействия и возможности управления рисками Этот процесс сопровождается оценкой эффективности плана по управлению рисками3",
"", "Для того чтобы стандартизовать эту деятельность научное и профессиональное сообщества находятся в постоянном сотрудничестве направленном на выработку базовой методологии политик и индустриальных стандартов в области технических мер защиты информации юридической ответственности а также стандартов обучения пользователей и администраторов Эта стандартизация в значительной мере развивается под влиянием широкого спектра законодательных и нормативных актов которые регулируют способы доступа обработки хранения и передачи данных Однако внедрение любых стандартов и методологий в организации может иметь лишь поверхностный эффект если культура непрерывного совершенствованияen не привита должным образом4",
"Общие сведения", "В основе информационной безопасности лежит деятельность по защите информации — обеспечению её конфиденциальности доступности и целостности а также недопущению какойлибо компрометации в критической ситуации5 К таким ситуациям относятся природные техногенные и социальные катастрофы компьютерные сбои физическое похищение и тому подобные явления В то время как делопроизводство большинства организаций в мире до сих пор основано на бумажных документах6 требующих соответствующих мер обеспечения информационной безопасности наблюдается неуклонный рост числа инициатив по внедрению цифровых технологий на предприятиях78 что влечёт за собой привлечение специалистов по безопасности информационных технологий ИТ для защиты информации Эти специалисты обеспечивают информационную безопасность технологии в большинстве случаев — какойлибо разновидности компьютерных систем

Следует отметить что под компьютером в данном контексте подразумевается не только бытовой персональный компьютер а цифровые устройства любой сложности и назначения начиная от примитивных и изолированных наподобие электронных калькуляторов и бытовых приборов вплоть до индустриальных систем управления и суперкомпьютеров объединённых компьютерными сетями Крупнейшие предприятия и организации в силу жизненной важности и ценности информации для их бизнеса нанимают специалистов по информационной безопасности как правило себе в штат В их задачи входит обезопасить все технологии от вредоносных кибератак зачастую нацеленных на похищение важной конфиденциальной информации или на перехват управления внутренними системами организации",
"", "Информационная безопасность как сфера занятости значительно развилась и выросла в последние годы В ней возникло множество профессиональных специализаций например таких как безопасность сетей и связанной инфраструктуры защиты программного обеспечения и баз данных аудит информационных систем планирование непрерывности бизнеса выявление электронных записей и компьютерная криминалистикаen Профессионалы информационной безопасности имеют весьма стабильную занятость и высокий спрос на рынке труда Масштабные исследования проведённые организацией ISC показали что на 2017 год 66 руководителей информационной безопасности признали острую нехватку рабочей силы в своих подразделениях а по прогнозам к 2022 году недостаток специалистов в этой области составит по всему миру 1 800 000 человек9",
"", "Угрозы и меры противодействия", "Угрозы информационной безопасности могут принимать весьма разнообразные формы На 2018 год наиболее серьёзными считаются угрозы связанные с «преступлением как услугой» англ CrimeasaService Интернетом вещей цепями поставок и усложнением требований регуляторов10 «Преступление как услуга» представляет собой модель предоставления зрелыми преступными сообществами пакетов криминальных услуг на даркнетрынке по доступным ценам начинающим киберпреступникаменК 1 Это позволяет последним совершать хакерские атаки ранее недоступные изза высокой технической сложности или дороговизны делая киберпреступность массовым явлением12 Организации активно внедряют Интернет вещей устройства которого зачастую спроектированы без учёта требований безопасности что открывает дополнительные возможности для атаки К тому же быстрое развитие и усложнение Интернета вещей снижает его прозрачность что в сочетании с нечётко определёнными правовыми нормами и условиями позволяет организациям использовать собранные устройствами персональные данные своих клиентов по собственному усмотрению без их ведома Кроме того для самих организаций проблематично отслеживать какие из собранных устройствами Интернета вещей данных передаются во вне Угроза цепей поставок состоит в том что организации как правило передают своим поставщикам разнообразную ценную и конфиденциальную информацию в результате чего теряют непосредственный контроль над ней Таким образом значительно возрастает риск нарушения конфиденциальности целостности или доступности этой информации Всё новые и новые требования регуляторов значительно осложняют управление жизненноважными информационными активами организаций Например введённый в действие в 2018 году в Евросоюзе Общий регламент защиты персональных данных англ General Data Protection Regulation GDPR требует от любой организации в любой момент времени на любом участке собственной деятельности или цепи поставок продемонстрировать какие персональные данные и для каких целей имеются там в наличии как они обрабатываются хранятся и защищаются Причём эта информация должна быть предоставлена не только в ходе проверок уполномоченными органами но и по первому требованию частного лица — владельца этих данных Соблюдение такого комплаенса требует отвлечения значительных бюджетных средств и ресурсов от других задач информационной безопасности организации И хотя упорядочение обработки персональных данных предполагает в долгосрочной перспективе

улучшение информационной безопасности в краткосрочном плане риски организации заметно возрастают10",
"", "Большинство людей так или иначе испытывают на себе воздействие угроз информационной безопасности Например становятся жертвами вредоносных программ вирусов и червей троянских программ программвымогателей13 фишинга или кражи личности Фишинг англ Phishing представляет собой мошенническую попыткуК 2 завладения конфиденциальной информацией например учётной записью паролем или данными кредитной карты Обычно пользователя Интернета стараются заманить на мошеннический вебсайт неотличимый от оригинального сайта какойлибо организации банка интернетмагазина социальной сети и т п1415 Как правило такие попытки совершаются с помощью массовых рассылок поддельных электронных писем якобы от имени самой организации16 содержащих ссылки на мошеннические сайты Открыв такую ссылку в браузере ничего не подозревающий пользователь вводит свои учётные данные которые становятся достоянием мошенников17 Термин Identity Theft с англ—«кража личности» появился в английском языке в 1964 году18 для обозначения действий в которых чьилибо персональные данные например имя учётная запись в банковской системе или номер кредитной карты часто добытые с помощью фишинга используются для мошенничества и совершения иных преступлений1920 Тот от чьего имени преступники получают незаконные финансовые преимущества кредиты или совершают иные преступления зачастую сам становится обвиняемым что может иметь для него далеко идущие тяжёлые финансовые и юридические последствия21 Информационная безопасность оказывает непосредственное влияние на неприкосновенность частной жизни22 определение которой в различных культурах может весьма разниться23",
"", "Органы государственной власти вооружённые силы корпорации финансовые институты медицинские учреждения и частные предприниматели постоянно накапливают значительные объёмы конфиденциальной информации о своих сотрудниках клиентах продуктах научных исследованиях и финансовых результатах Попадание такой информации в руки конкурентов или киберпреступников может повлечь для организации и её клиентов далеко идущие юридические последствия невосполнимые финансовые и репутационные потери С точки зрения бизнеса информационная безопасность должна быть сбалансирована относительно затрат экономическая модель ГордонаЛобаеn описывает математический аппарат для решения этой задачи24 Основными способами противодействия угрозам информационной безопасности или информационным рискам являются",
"", "снижение — внедрение мер безопасности и противодействия для устранения уязвимостей и предотвращения угроз", "передача — перенос затрат связанных с реализацией угроз на третьих лиц страховые или аутсорсинговые компании", "принятие — формирование финансовых резервов в случае если стоимость реализации мер безопасности превышает потенциальный ущерб от реализации угрозы", "отказ — отказ от чрезмерно рисковой деятельности25"), meta = list(author = character(0), datetimestamp = list(sec = 22.9859669208527,
    min = 2, hour = 19, mday = 14, mon = 0, year = 121, wday = 4, yday = 13, isdst = 0), description = character(0), heading = character(0), id = "Информационная безопасность.txt", language = "en", origin = character(0))))
list()
list()
> for (j in seq(docs)) {
+   docs[[j]] <- gsub("/", " ", docs[[j]])
+   docs[[j]] <- gsub("@", " ", docs[[j]])
+   docs[[j]] <- gsub("\\|", " ", docs[[j]])
+   docs[[j]] <- gsub("\u2028", " ", docs[[j]]) # This is an ascii character that did not translate, so it had to be removed.
+ }

```
> docs <- tm_map(docs, removeNumbers)
> docs <- tm_map(docs, tolower)
> docs <- tm_map(docs, PlainTextDocument)
> DocsCopy <- docs
> docs <- tm_map(docs, removeWords, stopwords("russian"))
> docs <- tm_map(docs, PlainTextDocument)
> docs <- tm_map(docs, removeWords, c("случай", "понятие"))
> for (j in seq(docs))
+ {
+   docs[[j]] <- gsub("Информационная безопасность", "Информационная безопасность",
docs[[j]])
+   docs[[j]] <- gsub("компьютерные сети", "компьютерные сети", docs[[j]])
+   docs[[j]] <- gsub("безопасность сети", "безопасность сети", docs[[j]])
+ }
> docs <- tm_map(docs, PlainTextDocument)
> docs_st <- tm_map(docs, stemDocument)
> docs_st <- tm_map(docs_st, PlainTextDocument)
list()
list()
> docs_stc <- tm_map(docs_st, stemCompletion, dictionary = DocsCopy, lazy=TRUE)
> docs_stc <- tm_map(docs_stc, PlainTextDocument)
list()
list()
list(index = TRUE, maps = list(function (x)
FUN(x, ...)))
> docs <- tm_map(docs, stripWhitespace)
> docs <- tm_map(docs, PlainTextDocument)
> dtm <- DocumentTermMatrix(docs)
> dtm
<<DocumentTermMatrix (documents: 5, terms: 2297)>>
Non-/sparse entries: 2890/8595
Sparsity           : 75%
Maximal term length: 33
Weighting          : term frequency (tf)
> tdm <- TermDocumentMatrix(docs)
> tdm
<<TermDocumentMatrix (terms: 2297, documents: 5)>>
Non-/sparse entries: 2890/8595
Sparsity           : 75%
Maximal term length: 33
Weighting          : term frequency (tf)
> freq <- colSums(as.matrix(dtm))
> length(freq)
[1] 2297
> ord <- order(freq)
> m <- as.matrix(dtm)
> dim(m)
[1]    5 2297
> dtms <- removeSparseTerms(dtm, 0.2) # This makes a matrix that is 20% empty space,
maximum.
> dtms
<<DocumentTermMatrix (documents: 5, terms: 19)>>
```

```
Non-/sparse entries: 95/0
Sparsity           : 0%
Maximal term length: 14
Weighting          : term frequency (tf)
> freq <- colSums(as.matrix(dtm))
> head(table(freq), 20)
freq
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   19   20
1712  307  103   62   27   23    8    9   10    5    6    5    3    2    1    3    1    1    1
  22
   1
> tail(table(freq), 20)
freq
 7  8  9 10 11 12 13 14 15 16 17 19 20 22 23 34 42 65 69 97
 8  9 10  5  6  5  3  2  1  3  1  1  1  1  2  1  1  1  1  1
> freq <- colSums(as.matrix(dtms))
> freq
  безопасности   безопасность        данных   деятельности        защиты     информации
            97             17             19             11             42             69
информационной информационных        которые         могут    обеспечения   организации
            65             16              9              7             20             34
      средств    технических         угроз     управление     управления     целостности
            16             11             12              8             15             11
          это
            13
> freq <- sort(colSums(as.matrix(dtm)), decreasing=TRUE)
> head(freq, 14)
  безопасности     информации информационной         защиты    организации           англ
            97             69             65             42             34             23
       системы         систем    обеспечения         данных   безопасность информационных
            23             22             20             19             17             16
      политика        средств
            16             16
> findFreqTerms(dtm, lowfreq=5)
 [1] "cia"             "активов"         "англ"            "аудит"
 [5] "безопасности"    "безопасность"    "бизнеса"         "внутренних"
 [9] "года"            "году"            "данные"          "данных"
[13] "деятельности"    "деятельность"    "документов"      "документы"
[17] "должна"          "должны"          "доступа"         "доступности"
[21] "доступность"     "защита"          "защиты"          "иб»"
[25] "изменения"       "информацией"     "информации"      "информационная"
[29] "информационной"  "информационных"  "использования"   "компьютерных"
[33] "контроля"        "конфиденциальной" "конфиденциальности" "конфиденциальность"
[37] "которая"         "которые"         "которых"         "лишь"
[41] "мер"             "меры"            "могут"           "наиболее"
[45] "например"        "необходимо"      "обеспечение"     "обеспечения"
[49] "области"         "обработки"       "обычно"          "определения"
[53] "определить"      "организации"     "организаций"     "организацию"
[57] "органов"         "основе"          "оценка"          "оценки"
[61] "подобные"        "политика"        "политики"        "правило"
[65] "предотвращения"  "предприятия"     "программного"    "процедур"
[69] "процедуры"       "процесс"         "работы"          "реализации"
```

```
 [73] "риска"            "рисками"          "россии"           "российской"
 [77] "своих"            "свойство"         "связи"            "систем"
 [81] "системы"          "следующие"        "служба"           "соиб"
 [85] "специалистов"     "средств"          "средства"         "стандартов"
 [89] "стандарты"        "также"            "такие"            "технических"
 [93] "технологий"       "тому"             "требований"       "требования"
 [97] "угроз"            "угрозы"           "уничтожения"      "управление"
[101] "управления"       "уровня"           "уязвимости"       "федерации"
[105] "хранения"         "цели"             "целостности"      "целостность"
[109] "шифрования"       "этих"             "это"              "является"
[113] "являются"
> wf <- data.frame(word=names(freq), freq=freq)
> head(wf)
                    word freq
безопасности     безопасности   97
информации        информации   69
информационной информационной   65
защиты               защиты   42
организации       организации   34
англ                 англ   23
> library(ggplot2)

Присоединяю пакет: ‘ggplot2’

Следующий объект скрыт от ‘package:NLP’:

    annotate

> p <- ggplot(subset(wf, freq>50), aes(x = reorder(word, -freq), y = freq)) +
+   geom_bar(stat = "identity") +
+   theme(axis.text.x=element_text(angle=45, hjust=1))
> p
> findAssocs(dtm, c("безопасности" , "информации"), corlimit=0.85)
$безопасности
     информационной           политику              работ              меры
               0.95               0.94               0.94               0.93
         федерации         информацию          следующие          способы
               0.91               0.90               0.90               0.90
       требованиям        требования             уровня           включая
               0.90               0.89               0.89               0.86
               гост           доктрина        допустимого            защиту
               0.86               0.86               0.86               0.86
        излучений         инструкции             исомэк            каждая
               0.86               0.86               0.86               0.86
         объектов       организация             правил           правила
               0.86               0.86               0.86               0.86
программноаппаратные             рамках         содержание          создания
               0.86               0.86               0.86               0.86
    соответствовать             уровне     электромагнитных      использования
               0.86               0.86               0.86               0.85

$информации
```

```
  требованиям       возможно        отнести   соответствия    внутреннего     выполнения
         0.92           0.90           0.90           0.90           0.88           0.88
     каждого      множества  необходимости       обороны         право      процессов
         0.88           0.88           0.88           0.88           0.88           0.88
  руководства         своей        систему   составляющих стандартизации
         0.88           0.88           0.88           0.88           0.88

Предупреждение:
В temp$cex : достигнуто ограничение времени операции
> findAssocs(dtms, "информационной", corlimit=0.70)
$информационной
безопасности       защиты
        0.95         0.81

> library("wordcloud")
Загрузка требуемого пакета: RColorBrewer
> set.seed(142)
> wordcloud(names(freq), freq, min.freq=10)
> set.seed(142)
> wordcloud(names(freq), freq, max.words=100)
> set.seed(142)
> wordcloud(names(freq), freq, min.freq=5, scale=c(5, .1), colors=brewer.pal(6, "Dark2"))
> set.seed(142)
> dark2 <- brewer.pal(6, "Dark2")
> wordcloud(names(freq), freq, max.words=100, rot.per=0.2, colors=dark2)
> dtmss <- removeSparseTerms(dtm, 0.15) # This makes a matrix that is only 15% empty space,
maximum.
> dtmss
<<DocumentTermMatrix (documents: 5, terms: 19)>>
Non-/sparse entries: 95/0
Sparsity           : 0%
Maximal term length: 14
Weighting          : term frequency (tf)
> library(cluster)
> d <- dist(t(dtmss), method="euclidian")
> fit <- hclust(d=d, method="complete")   # for a different look try substituting: method="ward.D"
> fit

Call:
hclust(d = d, method = "complete")

Cluster method   : complete
Distance         : euclidean
Number of objects: 19

> plot(fit, hang=-1)
> plot.new()
> plot(fit, hang=-1)
> groups <- cutree(fit, k=6)   # "k=" defines the number of clusters you are using
> rect.hclust(fit, k=6, border="red")
```

ДОДАТОК В
СХЕМИ РІЗНИХ ВИДІВ БЛОКУВАННЯ

Рисунок В.1 – Блокування по IP- адресу і протоколу

Рисунок В.2 – Блокування за допомогою технологій DPI

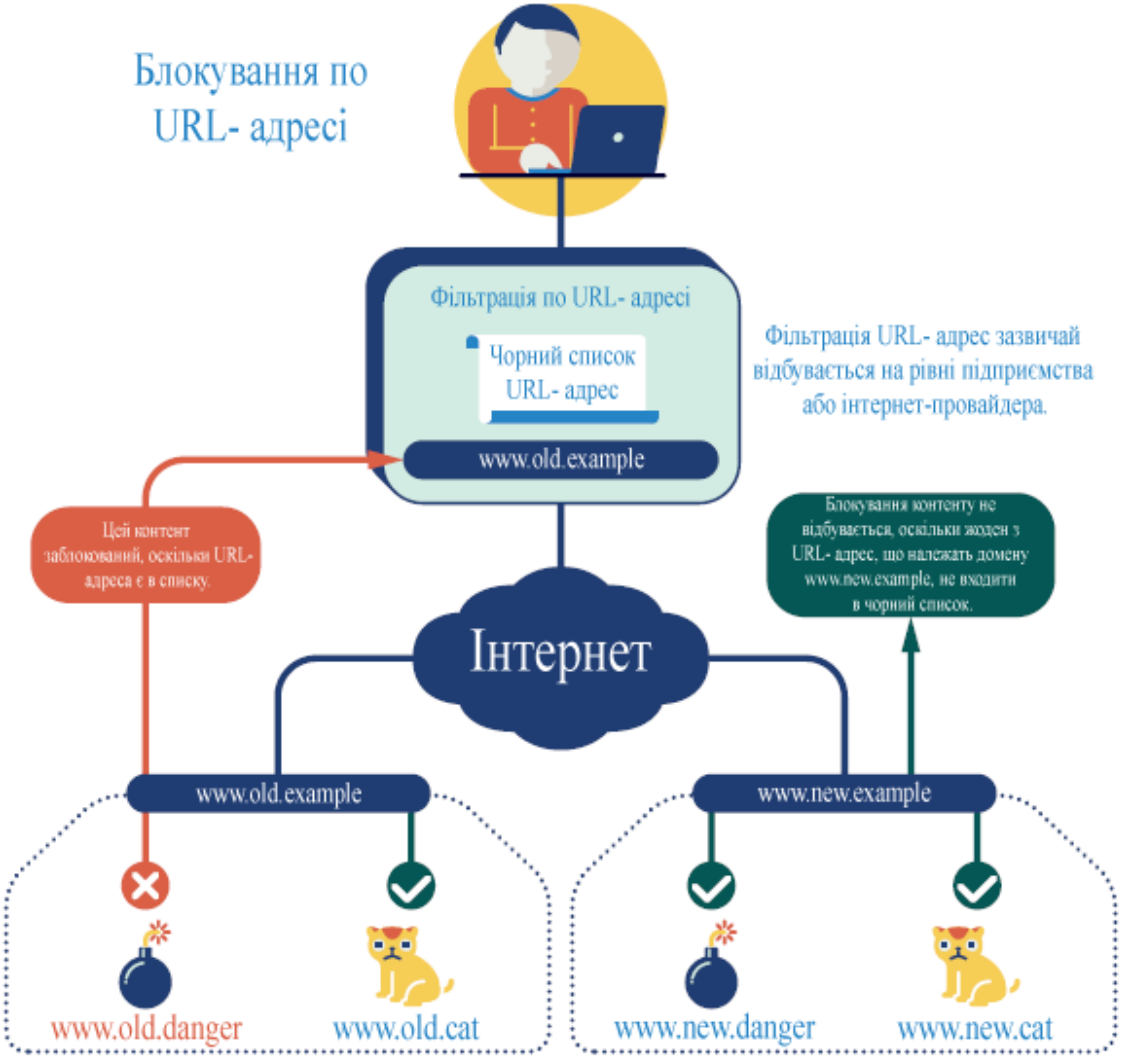


Рисунок В.3 – Блокування по URL- адресу

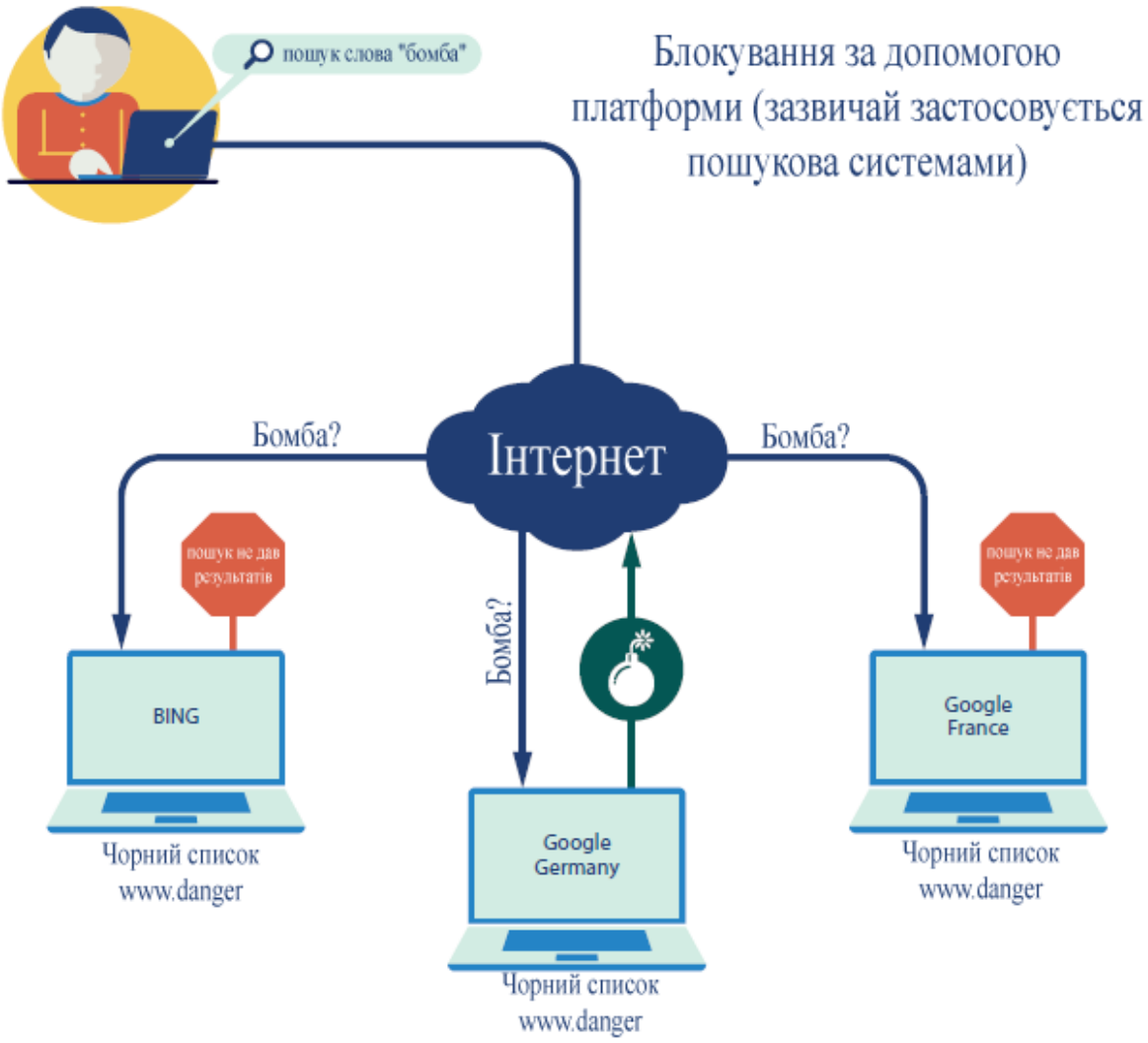


Рисунок В.4 – Блокування за допомогою платформи

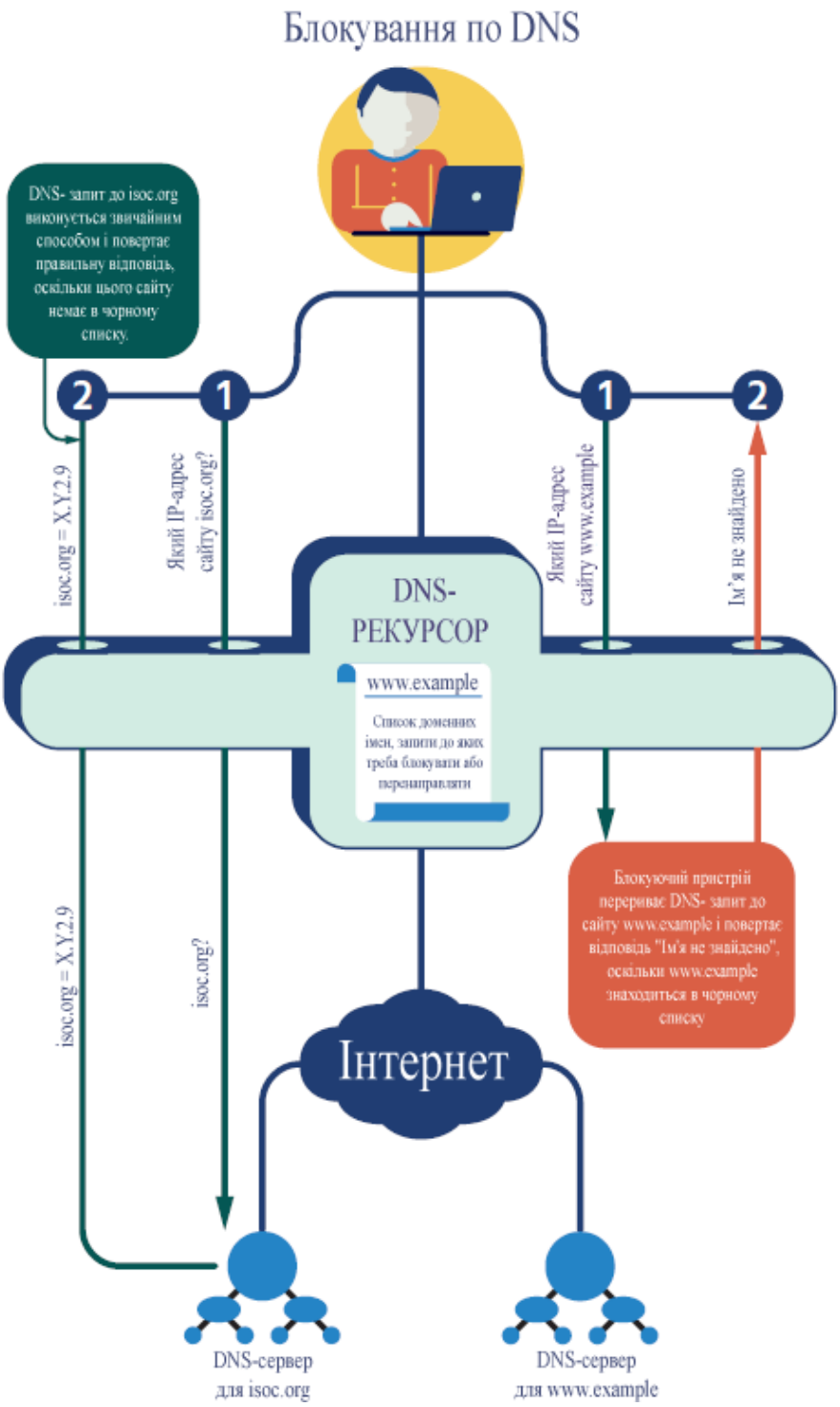


Рисунок В.5 – Блокування по DNS

ДОДАТОК Г
Публікація за темою роботи

МАТЕРІАЛИ 25-го МІЖНАРОДНОГО МОЛОДІЖНОГО ФОРУМУ

«РАДІОЕЛЕТРОНІКА І МОЛОДЬ У XXI СТОЛІТТІ»

20 – 22 квітня 2021 р.

Том 4

КОНФЕРЕНЦІЯ

«ПЕРСПЕКТИВИ РОЗВИТКУ ІНФОКОМУНІКАЦІЙ ТА ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНИХ ТЕХНОЛОГІЙ»

БЛОКУВАННЯ САЙТІВ З ВИКОРИСТАННЯМ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Семенченко О. А.
Науковий керівник – к.т.н. доцент Омельченко А. В.
Харківський національний університет радіоелектроніки
(61166, Харків, пр. Науки,14, каф. Інформаційно-мережна інженерія,
тел. (057) 702-13-06)
e-mail: oleksandr.semenchenko@nure.ua, +380507683086

The rapid development of information technology is gradually transforming the world. Open and free cyberspace expands the freedom and opportunities of people, enriches society. But unfortunately, not all information is beneficial fo person. Therefore, it is necessary to monitor such resources and block. In this we will help data mining, namely Data Mining.

With the help of Data Mining technologies it becomes possible to solve many problems the analyst faces. The main ones are: classification, regression, search for associative rules and clustering.

Стрімкий розвиток інформаційних технологій поступово трансформує світ. Відкритий та вільний кіберпростір розширює свободу і можливості людей, збагачує суспільство. Але не вся інформація може нести користь людині.

Тому необхідно відстежувати шкідливі ресурси і блокувати їх. Широкі можливості з автоматизації цих процесів з'являються внаслідок використання засобів інтелектуального аналізу даних (Data Mining) [1-5], зокрема Text Mining та Web Mining.

Маючи на руках засоби Text Mining та Web Mining можна проаналізувати матеріал на наявність шкідливого або небезпечного матеріалу.

До шкідливого матеріалу можна віднести [1]: ненормативну лексику; заклики до суїциду; утиску прав віруючих; екстремістські матеріали; використання образ та матеріали, що сіють ворожнечу за расовою, національною, релігійною або статевою ознакою.

Метою роботи є розвиток методів і засобів виявлення шкідливого контенту (ненормативної лексики та спроб торгівлі органами) у текстових даних, для подальшого блокування пов'язаних з ними ресурсів.

Для розв'язання задач Text Mining існують програмні засоби на таких мовах програмування як: Python, R, MatLab, SQL, Java, Scala, Julia, C++, JavaScript, Ruby, Perl.

У даній роботі для розв'язання задач виявлення шкідливої інформації у текстах використано мову програмування R, яка є широко розповсюдженою, має у своєму розпорядженні прикладні пакети практично для будь-якого застосування, зокрема стосовно задач Text Mining.

Додаткова зручність програмування мовою R забезпечується завдяки використанню середовища розробки програмного забезпечення RStudio.

У практичній частині роботи проводиться аналіз декількох текстів на наявність ненормативної лексики з попереднім пошуком сленгових слів. Пошук проводиться в створеній програмі, яка за заданими параметрами знаходить слова чи частину слів, що викликають підозру.

Спочатку програма присвоює кожному слову порядковий номер і після чого вказує, де саме у тексті знаходиться це слово або його частина, яка може бути замаскованою (додаванням зайвих літер або написанням слів разом).

Отримавши результат, аналітик може визначити, наскільки слово несе загрозу. Відносно тексту з небезпечним контентом можна поступити наступним чином: винести попередження, при якому власник сайту повинен знищити шкідливий матеріал або блокувати ресурс.

Розглянуто методи блокування ресурсів: блокування по IP-адресу, за допомогою технології DPI, блокування по URL-адресу, блокування за допомогою платформи та DNS блокування.

Виконано багатокритеріальний вибір найкращого методу блокування ресурсу за сукупністю показників якості, що враховують складність програмного забезпечення, затрати на апаратуру, умови блокування. Встановлено, що блокування за допомогою DNS є найкращим методом блокування.

Література:

5. Закон України «Про основні засади забезпечення кібербезпеки України» // (Відомості Верховної Ради (ВВР), 2017, № 45, ст.403
6. Конвенція Ради Європи «Конвенція про кіберзлочинність» // http://zakon.rada.gov.ua/laws/show/994_575
7. А. А. Барсегян, М. С. Куприянов, І. І. Холод, М. Д. Тесс, С. І. Елізаров. «Аналіз даних і процесів: навч. Посібник» - 3-е вид., Перераб. і доп. - СПб .: БХВ-Петербург, 2009. - 512 с .
8. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta «Practical Data Science Cookbook»
9. Ingo Feinerer «Introduction to the tm Package Text Mining in R» // https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf

ДОДАТОК Д
СЛАЙДИ ПРЕЗЕНТАЦІЇ

Харківський національний університет радіоелектроніки

Кафедра Інформаційно-мережної інженерії

Кваліфікаційна робота на тему:

Блокування сайтів з використанням методів інтелектуального аналізу даних

Студент:	Семенченко Олександр Андрійович
Група:	ІМІм-19-2
Керівник:	доц. Омельченко Анатолій Васильович

Харків - 2021

1

Мета роботи

1. Дослідження важкості порушення
2. Аналіз тексту на наявність порушення
3. Оцінювання можливостей блокування (порівняльний аналіз) цих методів
4. Повторне блокування/розблокування

2

Актуальність проблеми

Стрімкий розвиток інформаційних технологій поступово трансформує світ. Відкритий та вільний кіберпростір розширює свободу і можливості людей, збагачує суспільство, створює новий глобальний інтерактивний ринок ідей, досліджень та інновацій, стимулює відповідальну та ефективну роботу влади і активне залучення громадян до управління державою та вирішення питань місцевого значення, забезпечує публічність та прозорість влади, сприяє запобіганню корупції.

Водночас переваги сучасного цифрового світу та розвиток інформаційних технологій обумовили виникнення нових загроз національній та міжнародній безпеці. Поряд із інцидентами природного (ненавмисного) походження зростає кількість та потужність кібератак, вмотивованих інтересами окремих держав, груп та осіб.

Поширюються випадки незаконного збирання, зберігання, використання, знищення, поширення, персональних даних, незаконних фінансових операцій, крадіжок та шахрайства у мережі Інтернет. Кіберзлочинність стає транснаціональною та здатна завдати значної шкоди інтересам особи, суспільства і держави.

3

Цензура в Європі

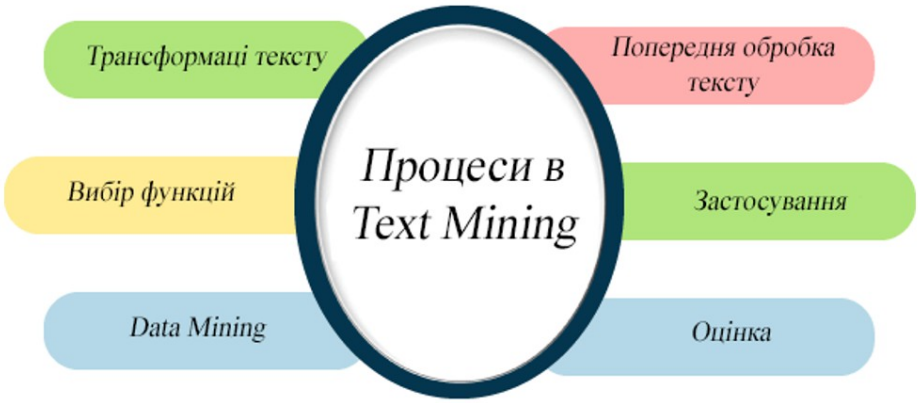
4

Розподіл матеріалів по важкості порушення

Одразу заблокувати	Обмежети доступ	Увідомити про необхідність видалення
Заклик до єкстремізму або єкстремістські висловлювання	Заклики в участі в неузгоджених мітінгах	Поширення фейков
Розповсюдження дитячої порнографії	Розповсюдження інормації о способах здійснення самовбивства або схиляння до самогубства	Неповага до держави та приниження влади
Розголошення персональних даних	Пропаганду вживання наркотиків	Порушення авторських прав
Розголошення персональних даних	Оправдання тероризму	Розповсюдження матеріалів небажаних організацій
Торгівля органами		
Торгівля забороненими товарами		
Розміщення нелегальних онлайн-казино та фінансових пірамід		

5

Процеси в Text Mining

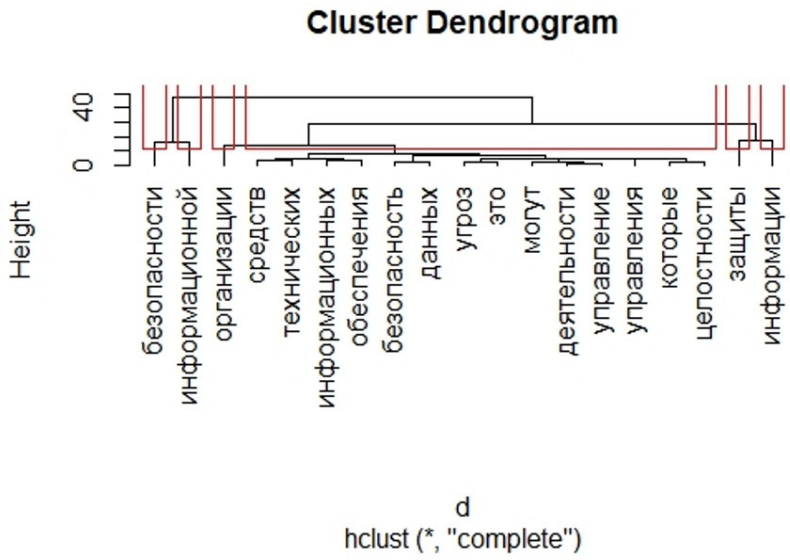


6

Хмара слів

7

Дендрограма слів

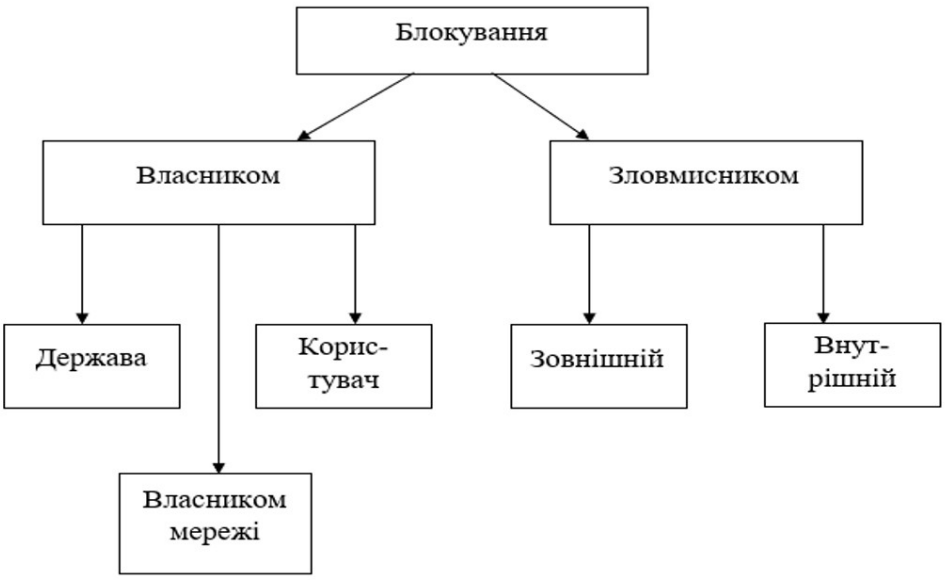


8

Пошук слів по тексту

```
 [15] "хате"          "и"          "думает"      "две"        "мысли."     "Первая"     "мысль:"
 [22] "о,"            "ништяк."    "Ну,"         "это"        "чисто"      "абстрактная" "мысль,"
 [29] "это"           "он"         "по"          "сезону"     "всегда"     "так"        "думает,"
 [36] "как"           "проснётся:" "о,"          "ништяк."    "Потому"     "что"        "ништяк"
 [43] "в"             "натуре."    "тело"        "как"        "перышко,"   "крыша"      "как"
 [50] "друшляк,"      "внутри"     "желудка"     "пустота."   "А"          "вот"        "вторая"
 [57] "мысль,"        "он"         "думает:"     "а"          "неплохо"    "бы"         "вот"
 [64] "подняться"     "и"          "что-нибудь"  "из"         "ништяков"   "вчерашних"  "заточить"
 [71] "неплохо"       "бы."        "потому"      "что"        "там"        "ништяков"   "нормально"
 [78] "осталось,"     "типа"       "банка"       "тушонки,"   "булка"      "хлеба,"     "картошки"
 [85] "пол-казана,"   "короче"     "ни"          "фига"       "себе"       "ништяков"   "осталось."
 [92] "и"             "вот"        "он"          "встаёт"     "и"          "идёт"       "их"
 [99] "заточить."     "А"          "ништяков,"   "короче,"    "нету."      "Пустой"     "казан"
[106] "стоит,"        "и"          "всё."        "даже"       "хлеба"      "не"         "осталось."
[113] "Нету"          "вобще"      "ничего,"     "короче."    "и"          "вот"        "растаман"
[120] "громко"        "думает:"    "а"           "кто"        "это"        "мои"        "ништяки"
[127] "всё"           "захавал?"   "А"           "из-под"     "шкафа"      "отзывается" "стрёмный"
[134] "загробный"     "голос:"     "Это"         "я"          "ништяки"    "твои"       "ЗАХАВАЛ!!»"
> # txta <- tm_map(txta,removePunctuation)
> for (j in seq(txta)) {
+   txta[[j]] <- gsub("/", " ", txta[[j]])
+   txta[[j]] <- gsub("@", " ", txta[[j]])
+   txta[[j]] <- gsub("\\|", " ", txta[[j]])
+   txta[[j]] <- gsub("\u2028", " ", txta[[j]])
+ }
> # txta <- tm_map(txta, tolower)
> # txta <- tm_map(txta, PlainTextDocument)
> # TxtaCopy <- txta
> grep("ништяк",txta)
[1]  23  39  42  68  76  90 101 126
> grep("заточить",txta)
[1] 70 99
> j<- grep("ништяк",txta)[1]
> j
[1] 23
> gregexpr("захавал",txta)[j]
[[1]]
```

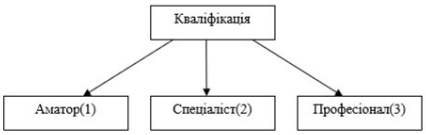

Схема різновиду типів блокування



Загальна схема пошуку і блокування інформації в Інтернеті

11

Схема різновидів критеріїв блокування/розблокування сайтів



12

Формули розрахунку ймовірності найкращого блокування

$$V = \frac{\sqrt[n]{\prod_{j=1}^{n} w_{ij}}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} w_{kj}}}, \quad (4.1)$$

де $\prod_{j=1}^{n} w_{ij}$ – добуток всіх елементів рядка;

$\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} w_{kj}}$ – загальна сума всіх добудків кожної строчки матриці ймовірностей;

V_i – ймовірність обходу блокування.

$$\bar{V}_i = \frac{\sqrt[n]{\prod_{j=1}^{n} w_{ij} \times k_j}}{\sum_{k=1}^{n} \sqrt[n]{\prod_{j=1}^{n} w_{kj} \times k_j}}, \quad (4.2)$$

де k_j – коефіцієнт важливості.

13

$\bar{V}$

Результат розрахунку ефективності блокування

Вид блокування	Блокування державами		Блокування підприємством		Блокування користувачем	
	V	$\bar{V}$	V	$\bar{V}$	V	$\bar{V}$
Блокування по IP-адресу і протоколу	0.1488	0,149	0,1418	0,141	0,1829	0,1182
Блокування за допомогою технології DPI	0,281	0,28	0,2678	0,268	0,2414	0,241
Блокування по URL-адресу	0,2446	0,245	0,2332	0,233	0,210	0,21
Блокування за допомогою платформи	0,1963	0,196	0,215	0,215	0,1829	0,182
Блокування по DNS	0,1295	0,1295	0,1418	0,141	0,1829	0,182

14

Результат розрахунку ефективності обходу блокування

Метод обходу блокування	V	$\bar{v}$
Коэфіцієнт	0.1532	0.15328
CGI-proxy	0.1288	0.12889
VPN	0.1532	0.15329
Cash Google	0.1532	0.15328
IP-address	0.1532	0.15328
Translator	0.1288	0.12889
DNS	0.1532	0.15328
TOR	0.1288	0.12889

15

Висновки

В роботі розглянуті засоби виявлення шкідливого контенту в текстових даних в Інтернеті: використання ненормативної лексики.

Використовуючи методи Data Mining, а саме Text Mining ми розробили програму за допомогою якої іде перевірка матеріалу на наявність шкідливого матеріалу. Також беручи заздалегіть небезпечну інформацію навчаємо програму для подальшого кращого аналізу.

Виходячи з мультипікативного аналізу багатоконтурного блокування була проведена оцінка різних видів блокування і вибраний кращий результат.

В ході виконання роботи було узято два коротких текста та проаналізовано їх на наявність ненормативної лексики (для даної роботи шукав сленгові слова).

16

Як результат в першому тексті немає заборонених слів, а у другому є. Відносно цього результату для другого тексту ми можемо поступити так:

-винести попередження, при якому ресурс повинні знищити матеріал;

- блокувати ресурс.

Проаналізовані основні способи блокування веб-сайтів і за кількісним критерієм виконано їх порівняння. Було виявлено, що найкращим видом блокування по DNS.

В подальшому потрібно буде набрати базу даних небезпечних слів за допомогою яких буде проводитись аналіз текстів, а також створити більш кращий інтерфейс для користувачів.

Позначення	Найменування	Дод. відомості
	Текстові документи	
1.	Пояснювальна записка	93 с.
	Графічні документи	
2.	Слайди презентації	18

Змін.	Арк.	Номер докум.	Підпис	Дата	Блокування сайтів з використанням методів інтелектуального аналізу даних	Літ.	Аркуш	Аркушів
Розроб.		Семенченко				У	1	1
Перевір.		Омельченко			Відомість кваліфікаційної роботи	ХНУРЕ кафедра ІМІ		
Н. контр.		Омельченко						
Затв.		Безрук						