

## **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 for 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](http://zakon.rada.gov.ua/laws/show/994_575)
7. А. А. Барсегян, М. С. Купріянов, І. І. Холод, М. Д. Тесс, С. І. Єлізаров. «Аналіз даних і процесів: навч. Посібник» - 3-є 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>