ДОСЛІДЖЕННЯ МЕТОДУ ЗАХИСТУ КОРПОРАТИВНОЇ ПОШТИ ВІД СПАМА

Подорожняк А.О., Діденко К.Ю.
Національний технічний університет «ХПІ», Харків, Україна

Розширення можливостей використання інформаційних ресурсів, доступних через мережу Інтернет, призвело до широкого поширення небажаної кореспонденції – так званого "спаму". Спам є однією з найбільш гострих проблем Інтернету. Поширення спаму пов'язано не тільки із втратами мережевих ресурсів, але і з часовими витратами, необхідними користувачеві мережі для обробки подібної інформації.

Поширення таких видів спаму як реклама, антиреклама, фішинг, листи релігійного змісту та ін. [1], небезпечно ще й тим, що найчастіше повідомлення, що розсилаються, містять комп'ютерні віруси. У такій ситуації особливої важливості набуває спосіб створення фільтрів, які перешкоджають поширенню небажаної електронної кореспонденції. Одним з таких способів є автоматична фільтрація – програмне забезпечення (так звані спам-фільтри).

Існує безліч алгоритмів пошуку спаму у вхідному потоці повідомлень. Найбільш часто використовуваним з існуючих є алгоритм на основі теореми Байєса [2]. В основі методу автоматичної фільтрації лежить механізм розбиття вхідних листів на умовні слова (так звані "токени"). На основі цих токенів складається частотний словник, і до отриманих наборів слів застосовується теорема Байеса. Далі, архів відсортованих повідомлень передається програмі навчання. Вона обчислює частотні словники для кожного типу повідомлень (папки: спам – не-спам): скільки разів певне слово зустрічалося в листах цієї папки. Коли словники заповнені, обчислення ймовірності приналежності конкретного нового листа до того чи іншого типу (папки) здійснюється за формулою Байеса для кожного слова нового листа. Підсумовуванням і нормалізацією ймовірностей визначення спаму в повідомленнях отримують загальну оцінку листів. Як правило, ймовірність приналежності повідомлення до одного з типів (до папки) набагато вища, ніж його приналежність до іншого типу.

Використання байєсівської теорії при створенні фільтрів, що перешкоджають поширенню спаму, дозволяє з досить великою ймовірністю визначати приналежність листів до спаму на основі аналізу його заголовка і тексту з урахуванням раніше отриманих конкретним користувачем повідомлень.

Список літератури

1. Dada E.G. Machine learning for email spam filtering: review, approaches and open research problems / E.G. Dada, J.S. Bassi, H. Chiroma, S.M. Abdulhamid, A.O. Adetunmbi, O.E. Ajibuwa – Heliyon: Elsevier, 2019 – 23 p. https://doi.org/10.1016/j.heliyon.2019.e01802

2. Bhagyashri G. Auto e-mails classification using bayesian filter / G. Bhagyashri, H. Pratap, D.Y. Patil // International Journal of Advanced Technology & Engineering Research (IJATER). – 2013, Vol. 3, Issue 4, pp. 19-24.