



РОЗПІЗНАВАННЯ ІНФОРМАЦІЇ РЕКЛАМНОГО ЗМІСТУ У ВХІДНИХ ЕЛЕКТРОННИХ ПОВІДОМЛЕННЯХ

Ільїнський Б. В.

*Національний технічний університет
«Харківський політехнічний інститут»
м. Харків, вул. Пушкінська, 79/2, тел. 707-63-60,
E-mail: bogdan.ilyinsky@gmail.com*

Останнім часом зростає використання інтернет-реклами як одного з найбільш дієвих способів залучення покупця. Незважаючи на те, що найбільшим попитом на сьогоднішній день користується просування товарів і послуг за допомогою соціальних мереж, існує шлях оповіщення споживача, який можна назвати класичним для мережі Інтернет – електронна пошта (*e-mail*).

Переваги використання *e-mail* для доставки рекламних повідомлень:

- електронна пошта є практично у всіх користувачів Мережі;
- електронна пошта являє собою *push*-технологію віщання;
- електронна пошта дає можливість персоніфікованого звернення.

Цікаве, з точки зору одержувача, повідомлення може бути поширене їм серед його колег і знайомих [3].

В Інтернеті існує безліч списків розсилки (*mailing lists*, "*opt-in*" *E-mail marketing*), які присвячені різним тематикам. Одержувачі подібних листів власноруч підписалися на розсилку, та в будь-який момент у них є право і можливість скасувати свою підписку.

Існують наступні види підписок *e-mail* розсилки:

- відкриті розсилки – доступні усім бажаючим;
- закриті – призначені для використання користувачам певного кола;
- безкоштовні – що існують за рахунок ентузіазму творців, спонсорської підтримки чи платних рекламодавців;
- платні – які дозволяють використання своїх послуг за передплату.

Оскільки список розсилки звичайно є засобом віщання для визначеної цільової групи та часто має тисячі абонентів, він є ефективним інструментом маркетингу. Ряд компаній на своїх офіційних сайтах пропонує відвідувачам підписатися на розсилку, що інформує про новини компанії та оновленнях сайту. Дана розсилка нагадує абонентам про сайт та діяльність його власника, інформуючи і стимулюючи повторні візити.

Деякі компанії, що займаються легальним бізнесом, рекламують свої товари або послуги за допомогою спаму. Спам (англ. *spam*) – розсилка комерційної та іншої реклами або інших видів повідомлень особам, які не виражають бажання їх отримувати. Привабливість такої реклами – низька вартість та (приблизно) велике охоплення потенційних клієнтів.

Слід зауважити, що повідомлення, які звично називають спамом не завжди несуть у собі маркетинговий характер. До інших видів несанкціонованих повідомлень належать: «нігерійські листи» (*scam*), «фішинг» (*phishing*), «лист щас-



тя», пропагандистські листи різного характеру, *DoS*- і *DDoS*-атаки, листи, що містять комп'ютерні віруси (*malware*) та ін. [4].

Багато досліджень базуються на класифікації пошти згідно з заданими категоріями, що в загальному випадку відповідає папкам у поштовому клієнті. Серед підходів, які були застосовані для такого типу класифікації, виділяють класифікацію за допомогою машинного навчання та видобування інформації (*machine learning and IR approaches*). До них належать: *MailCat*, *SVM (support vector machines)*, *Re:Agent*, *SpamCop* [1].

До підходів, які застосовувалися для фільтрування спаму, належать метод Баєса (*Bayesian approach*) та, як і у випадку класифікації поштових повідомлень, машинне навчання. Вони демонструють достатньо високу точність при визначенні спаму. Більшість систем, побудованих за таким зразком, використовують попереднє опрацювання даних, а саме, токенізацію та стемінг (*tokenization and stemming*). На етапі токенізації визначаються токени, себто елементи, які будуть використовуватися під час навчання. Ними зазвичай є слова. Тобто токенізація передбачає усунення пунктуації й екстрагування слів. Оскільки одне і те ж слово може використовуватися у різних формах (різних відмінках, різному числі тощо), необхідно здійснити стемінг, тобто виділити корінь/основу слова або представити слова лише у початковій формі: інфінітиві (для дієслів), називному відмінкові однини (для іменників, займенників) тощо. Майже в усіх дослідженнях використовувалося вже розроблене програмне забезпечення, що здійснює стемінг [2].

Результатом даного дослідження стало програмне забезпечення, що автоматично розподіляє вхідну кореспонденцію користувача по каталогах. Принципи роботи програми базуються на структурних властивостях маркетингових та інших видів повідомлень (усього розглянуто приблизно 1,5 тисячі повідомлень рекламного змісту) з підключенням створеної бази даних слів та словосполучень, характерних для декількох основних тематик кореспонденції (рекламні повідомлення, підписні розсилки-*newsletters*, персоналізовані повідомлення, несанкціоновані розсилки та ін.). Працюючи на основі баєсівського методу програма розподіляє вхідні повідомлення таким чином, що користувач, маючи можливість вносити зміни в налаштування, не відчуватиме незручностей, навіть не втрачаючи час на налаштування персоналізованих фільтрів.

Список літератури

1. Використання нейронної мережі кохонена для розпізнавання спаму – [Електронний ресурс]. – Режим доступу: http://pnzzi.kpi.ua/14/14_p106.pdf
2. Застосування методів навчання для розпізнавання спаму – [Електронний ресурс]. – Режим доступу: <http://ena.lp.edu.ua:8080/bitstream/ntb/9546/1/15.pdf>
3. Інтернет-реклама – [Електронний ресурс]. – Режим доступу: <http://ru.wikipedia.org/wiki/Интернет-реклама>
4. Терейковский И.А.Применение семантического анализа содержимого электронных писем в системах распознавания спама – [Електронний ресурс]. – Режим доступу: <http://do.gendocs.ru/docs/index-206082.html>