

## **Помилки адресації в електронних мережах та методи їх коректування**

Северин М.В.

Одеська національна академія зв'язку ім. О.С. Попова, [n\\_severin@ukr.net](mailto:n_severin@ukr.net)

Розглянуто питання обробки помилок адресації в е-мережах. Запропоновано методи автоматичної корекції помилок в електронних адресах.

У мережі Інтернет досить поширені випадки помилкового набору електронної адреси. Причинами таких помилок можуть бути як неуважність користувача, так і труднощі запам'ятовування електронної адреси. Адже електронна адреса, сформована з букв латиниці та службових символів, може містити в собі набори символів, які є транслітераціями «іноземних» слів, скорочення слів, або взагалі неосмислені послідовності символів. Помилки такого роду у записях URL призводять до втрати потенційних клієнтів сайту, а при пересиланні листів (e-mail) - до недоствалення листа адресатові.

Дослідження проблеми «помилки набору» у мережі може, у певних випадках, містити і «комерційну складову». Так, у [1] розглядаються питання, пов'язані з залученням відвідувачів на сайт (розкручування сайту) за рахунок використання друкарських помилок при ручному наборі URL-адреси у браузері користувача. Для залучення «чужих» клієнтів на свій сайт здійснюється реєстрація «хибного» сайту. Тобто реєструється домен, близький за назвою до популярного домену (у якості імені домену вибирається запис, який містить найбільш ймовірну помилку при наборі згаданої назви). Такий метод отримав назву «туре-in-traffic», і його застосування найбільш характерне для відомих і легко запам'ятовуваних сайтів, введення адреси яких не викликає великих труднощів і тому часто виконується безпосередньо у рядку браузера.

Не менш розповсюджені і помилки запитів в пошукових системах. Вони також досліджуються, і за аналогією з помилками адресації використовуються для залучення клієнтів на свій сайт [2]. І хоча у запитах, на відміну від електронних адрес, використовується осмислена інформація, статистика пошукових систем свідчить про велику кількість орфографічних помилок або описок на зразок: А) спотворення символу (кількох символів), В) пропуску символу (ів), С) вставки «зайвого» символу (ів), D) переставлення кількох (часто - сусідніх) символів. Сьогодні у пошукових системах застосовуються заходи, спрямовані на усунення помилок подібного роду. Так, при введенні «спотвореного» тексту, система пропонує (підказує) користувачеві можливі варіанти «правильних» слів зі словника. Використання сервісу пошуку і виправлення помилок дає позитивні результати як у плані зниження навантаження на пошуковий сервер, так і в плані підвищення якості надавання ним послуг.

У представленій доповіді пропонуються методи корекції помилок для вирішення проблем з помилковим введенням електронних адрес. Ці

методи ґрунтуються на так званій «теорії подібності скінченних послідовностей» (ТПСкП) [3], у якій розроблені чисельні міри оцінки «подібності» символічних рядків, що характеризуються, зокрема, спотвореннями видів A) - D). Застосовуючи методи ТПСкП, веб-сервер, у випадках отримання помилкової URL-адреси, міг би виконати пошук «близької» (можливої) адреси і перенаправити запит на неї, або ж разом з повідомленням про відсутність шуканого сайту запропонувати список сайтів, адреса яких «близька» за написанням.

Такий підхід прийнятний і для електронної пошти. Наразі тут застосовують наступні прийоми обробки «неправильних» адрес. Агент передачі повідомлень МТА (Mail Transfer Agent) у випадках відсутності зазначеного відправником адресата формує повернений лист (bounce message). Це повідомлення відсилається назад відправникові, коли поштова скринька одержувача не існує або недоступна. (Існують різні шаблони таких листів, і вміст повернутого листа може відрізнятись у залежності від налаштувань поштового сервера). У тексті повернутого листа вказується текст помилки, адреса поштової скриньки, список кодів помилок, і причина, за якою лист не було доставлено. Також до bounce message може бути додано оригінал листа або деяка його частина.

Таким чином, відправлена користувачем термінова і важлива кореспонденція може не досягти адресата через помилку (помилки) у записі електронної адреси. А у разі незбереження ним копії листа - до його втрати. Якщо ж повторне відправлення листа можливе, його відправнику необхідно буде з'ясувати правильну адресу, і у разі складності такого з'ясування він може намагатися підібрати її методом перебору. Це призводить як до збільшення навантаження на мережу, так і до зниження якості надання послуг.

У традиційному поштовому зв'язку існує вимога прийняття всіх заходів щодо доставлення листа адресатові, у тому числі і щодо з'ясування нового місця перебування адресата. Електронна пошта є альтернативою традиційній, і якість надання нею сервісу має бути принаймні «не гірше». Враховуючи досвід, отриманий у традиційній пошті, перш ніж відправити e-mail-повідомлення про неможливість доставлення листа адресату, слід вжити всіх заходів із його «відшукування».

Широке поширення в мережі серверів, побудованих на Linux / Unix-платформах з використанням вільного програмного забезпечення [4], дає можливість розробникам не тільки вивчати вихідні коди, але й виконувати їх модифікацію. Найбільшу популярність серед МТА, що використовуються на Linux / Unix-платформах, набули: Sendmail, Postfix, Exim та Qmail. З урахуванням цього, можна запропонувати - як один з методів боротьби з помилками адресації - впровадження у МТА модуля, що дозволяє виконувати корекцію помилкової або ж пошук «близької» адреси. У разі можливості однозначного визначення адресата, модуль дозволить виконати коригування адреси і доставлення листа. При неможливості однозначно визначити адресата у повернений лист може бути вміщено перелік «близьких» адрес. Автор розробляє модуль корекції

«помилкових» адрес для MTA Postfix та Qmail.

Розглянемо приклади оцінок близькості електронних адрес, отриманих при використанні методів ТПСкП. У якості «шаблону» візьмемо електронну адресу автора: n\_severin@ukr.net. Оцінки близькості будуть виконані тільки для першої частини адреси (до символу @), оскільки ми тут вважаємо, що повідомлення знаходиться на сервері адресованого домену, і на поштового агента покладено тільки завдання пошуку адресата у своєму домені.

У таблиці справа від помилкової адреси показано значення ТПСкП-міри його «близькості» до правильної адреси.

nseverin	0.889	n_sveerin	0.889
m_severin	0.889	n_siverin	0.889
n-severin	0.889	n_sivirin	0.778
nv_severin	0.900	n_civerin	0.778

З прикладу видно, що незначні спотворення (помилки в одному з символів, перестановки символів) «мало» спотворюють електронну адресу. Встановивши поріг міри близькості, при якому адреси, «близькі» до правильної, припустимо вважати результатами «друкарських помилок», рівним 0.88, шість з восьми наведених вище адрес можна автоматично «виправити». Звичайно, на практиці всі адреси у наведеному списку могли б належати іншим користувачам мережі. Якщо така ситуація вважається високо-ймовірною, згадані шість адрес можна включити в bounce message. Отже, навіть у випадках неможливості або неприпустимості автоматичної корекції адреси, запропоновані методи роблять bounce message більш інформативним. Таким чином, застосування методів ТПСкП дозволить як підвищити якість роботи сервісу, так і знизити «помилковий» трафік мережі.

Література

1. С. Жарков [Кривые пальцы, приносящие трафик](http://www.linkz.ru/promotion/krivuee_palmztcue_prinosyascie_trafik.html) // URL: [http://www.linkz.ru/promotion/krivuee\\_palmztcue\\_prinosyascie\\_trafik.html](http://www.linkz.ru/promotion/krivuee_palmztcue_prinosyascie_trafik.html)
2. С. Кокшаров Роль опечаток в SEO // URL: <http://devaka.ru/articles/seo-misprints>
3. Леоненко Л.Л., Поддубный Г.В. Теория подобия конечных последовательностей и ее приложения к распознаванию образов // Автоматика и телемеханика. - 1996. - №8. - с.119–131.
4. Security Lab Исследование: «Идеальный почтовый сервер» // URL: <http://www.securitylab.ru/analytics/309889.php>