В результате работы алгоритма Дамерау-Левенштейна составляется матрица преобразования символов, где ее значения - это расстояния для преобразования символов. Итоговым значением преобразования одной строки в другую будет элемент с максимальными индексами. Например, расстояние между словами «дагестан» и «арестант» будет равно 3. Данный алгоритм требует довольно значительных затрат памяти и времени выполнения. Однако он может быть упрощен путем сокращения количества вычислений. Сравнение строк может происходить с учетом морфем, так могут подвергаться трансформации только суффиксы слов или окончания. Другим вариантом снижения затрат может быть использование ограничений по количеству сравнений.

Ключевым недостатком данных методов является время выполнения поиска. При работе с мобильными устройствами время является важным критерием работы приложения. Однако данные методы можно оптимизировать за счет их комбинаций или внесения дополнительных условий, задаваемых пользователем или разработчиком.

ЛИТЕРАТУРА

1. Adobe Experience Design [Электронный ресурс]. – Режим доступа: http://www.adobe.com/ru/products/experience-design.html.

2. Google Material Design [Электронный ресурс]. – Режим доступа: https://material.io/guidelines/#introduction-principles.

3. Нечеткий поиск в тексте и словаре [Электронный ресурс]. – Режим доступа: https://habrahabr.ru/post/114997.

УДК 519.256

Студ. Д.А. Радиванович

Науч. рук. ст.преп. Ю.О. Герман

(кафедра информационных систем и технологий, БГТУ)

ПРИЛОЖЕНИЕ ДЛЯ ФИЛЬТРАЦИИ ЭЛЕКТРОННОЙ КОРРЕСПОНДЕНЦИИ

Спам — массовая рассылка сообщений различного содержания, пользователям, не выражающим желания их получать. Согласно определению «Лаборатории Касперского» спам — это анонимная массовая непрошенная рассылка. В этом определении важно каждое включенное в него слово. Анонимная: все страдают именно от автоматических рассылок со скрытым или фальсифицированным

обратным адресом. Массовая: эти рассылки именно массовые, и только они являются настоящим бизнесом для спамеров и настоящей проблемой для пользователей. Непрошенная: очевидно, подписные рассылки и конференции не должны подпадать под наше определение (хотя условие анонимности и так в значительной мере это гарантирует) Спам является довольно серьёзной проблемой нашего времени, создающей множество неудобств большинству пользователей. Мошенничество, вымогательство, захламление почтовых ящиков, распространение вирусов и даже доведение человека до нервного срыва – всё это и многое другое реализуется при помощи спама. На самом деле понятие спам, в котором понимает его большинство, не очень корректно. В зависимости от содержания таких сообщений, принято понимать два обобщённых типа: UCE - «коммерческий» спам — «unsolicited commercial e-mail» и (UBE) «некоммерческий» — «unsolicited bulk e-mail». По статистике на апрель 2015 года, собранной Лабораторией Касперского, спам составил 71,1% глобального почтового траффика. К тому же это на 7,6% больше чем в марте месяце. Причиной возросшего количества сообщений стали пасхальные праздники. Собственно, практически каждый значимый праздник, становится «праздником» для спамеров. Такое количество бесполезных сообщений создаёт проблему для компаний, предоставляющих услуги электронной почты, ведь это в разы увеличивает нагрузку на серверы и заставляет тратить большее количество денег на оборудование. В роли источника такой рассылки могут выступать как сами спамеры, так и ни о чём не подозревающий пользователь, чей IP адрес или адрес электронной почты, был захвачен злоумышленниками и используется для рассылки [1].

Методов борьбы со спамом существует огромное количество, что, во-первых, показывает, что это именно проблема для Интернета, а, во-вторых, – что ею активно занимаются. Но никто не стал бы заниматься этим просто так, что наталкивает на мысль – спам приносит хорошие доходы своим организаторам, что говорит о том, что несмотря ни на что, пользователь всё-таки пользуется тем, что предлагают спам-сообщения, и до тех пор, пока доходы превышают расходы, спам будет существовать. Тогда можно с уверенностью сказать, что главным методом борьбы со спамом является полный отказ от услуг, предлагаемых таким образом, и максимально возможное игнорирование подобных сообщений. В этом случае лица, организующие и использующие такие методы воздействия на обывателей, потеряют всякий доход с этого предприятия. Но,

несмотря на идейные, есть также и технические методы борьбы со спамом [2].

Превентивные методы – методы предварительной защиты, направленные на предотвращение попыток использования, к примеру, электронной почты. Здесь многое зависит от человеческого фактора. Если вы и в правду хотите сохранить свою электронную почту в безопасности, то можно предпринять целый ряд мер. Не публиковать свой адрес на общедоступных сайтах. Если вдруг это действительно необходимо, то постараться сделать так и написать его таким образом, чтобы он был не читаем для программ-харвестров, например, поставить нижние подчёркивания и пробелы в словах, заменить буквы на цифры, где возможно. Для того, кто должен прочитать это сообщение, оставить комментарии, хотя в некоторых случаях программы могут распознать закодированный текст.

Публичные сайты, должны быть организованы так, чтобы электронный адрес не был виден, а сообщения было бы возможно отправлять, но никнэйму. Адрес можно предоставить в виде картинки, или QR-кода, при этом не стоит доверять онлайн-службам, делающим это, так как известны случаи продажи таких баз данных с адресами спамерам. Лучше разобраться самому в графическом редакторе, или в крайнем случае написать на бумаге и сфотографировать. На web-страницах адреса можно кодировать с помощи JavaScript. Можно разделить всю свою почту на два ящика, один, используя для обмена личными и корпоративными сообщениями, а второй для регистрации и авторизации на различных сайтах и в разных службах, не внушающих доверия. Естественно, никогда не открывать и не читать сообщения подозрительного содержания, и тем более не отвечать, если вы на все сто не уверены в том, от кого это сообщение, не загружать никаких изображений, приложенных в письме.

При создании почтового ящика, постараться придумать сложный, трудноподбираемый и не имеющий никакого конкретного смысла адрес. Не использовать имена или простые распространённые слова, ведь спамеры люди подкованные и с лёгкостью, пользуясь «особенными словарями» и методом перебора, подберут ваш адрес. Также постараться сделать имя длинным, лучше всего, если это будет простой набор букв, не имеющий смысла. Использование, например, букв русского и латинского алфавита, а также цифр, изрядно усложнит задачу спамерам, и практически исключит случайную атаку. Однако все эти попытки избежать угадывания, несут один значительный минус. Делая email сложным для спамеров, мы также

усложняем жизнь себе. Особенно неудобно, если это адрес какой-нибудь фирмы, и он должен быть понятен и прост для чтения.

На сегодняшний день наиболее действенным методом является фильтрация. Автоматическая фильтрация – принцип, при котором специализированное программное обеспечение (ПО) сканирует сообщения и принимает решение спам/не спам. Предназначено, как для простых пользователей, так и для использования на серверах. Использует два основных подхода.

Первый – это анализ текста сообщения, которое опираясь на известные приёмы спамеров, методы байесовской фильтрации, а также поиск по ключевым словам принимает решение, является ли сообщение спамом или нет и помешает в соответствующую папку. Естественно, прежде чем запускать подобную программу, нужно провести её предварительное обучение. Однако при работе такого ПО не у клиента, а на сервере, возникает риск, что фильтр ошибочно пометит сообщение как спам, и оно не дойдет до пользователя. Здесь требуются большое количество статистических данных для увеличения вероятности правильного распознавания.

Второй подход заключается в том, чтобы, не изучая текст письма, опознать самого отправителя как спамера и заблокировать. Такое программное обеспечение может работать только на сервере. Но опять же, не исключены вероятности ошибки, и что простая массовая рассылка людям, которые друг с другом знакомы, будет предпринята как попытка распространения спама. Существуют также специализированные online-сервисы, например, «Лаборатория Касперского», Outcom «СПАМОРЕЗ», ИНКАП «Антиспам-Пост», ContrSpam, Антиспамус, предоставляющие платную защиту от спама. Изменение MX-записи в доменном имени предприятия особым образом позволяет перенаправить почту для защищаемого домена на специализированный почтовый сервер, где она очищается от спама и вирусов, а затем направляется на корпоративный почтовый сервер. Неавтоматическая фильтрация – метод позволяющий пользователю самостоятельно создавать фильтры, состоящие из слов или выражений. Однако это слишком трудоёмко и практически не используется. Сформировав свой собственный фильтр, пользователь может быть уверен, что нужные ему сообщения не будут отсеяны [3].

Данное программное приложение позволяет фильтровать письма и делать вывод о спаме, опираясь на известную теорему Байеса [4]:

$$P(c/d) = \frac{P(d/c)}{P(d)}, \quad (1)$$

где $P(c|d)$ — вероятность, что сообщение d принадлежит классу c, именно её нам надо рассчитать;

$P(d|c)$ — вероятность встретить сообщение d среди всех сообщений класса c;

$P(c)$ — безусловная вероятность встретить сообщение класса c в базе сообщений;

$P(d)$ — безусловная вероятность сообщения d в базе сообщений.

Приложение реализовано на платформе .Net с помощью языка C#.

Для верной классификации писем необходимо собрать статистическую информацию по предыдущим данным, которые заносятся в подключаемую к модулю базу данных. Любое новое письмо будет проверено на наличие спама, используя указанную теорему.

ЛИТЕРАТУРА

1. Спам [Электронный ресурс] / Википедия. – Режим доступа: https://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B0%D0%. – Дата доступа: 01.04.2017.

2. Борьба со спамом: история и методы [Электронный ресурс] – Режим доступа: https://mipt.ru/dmcp/student/diff_articles/no_spam.php. – Дата доступа: 03.04.2017.

3. Методы борьбы со спамом [Электронный ресурс] – Режим доступа: http://www.securelist.com/ru/threats/spam?chapter=157 – Дата доступа: 03.04.2017.

4. Герман, О. В. Экспертные системы: лабораторный практикум / О. В. Герман, Н. В. Батин. – Минск : БГУИР, 2003. – 75 с.

УДК 004.4

Студ. А.С. Бубель
Науч. рук. доц. Н.Н. Буснюк
(кафедра информационных систем и технологий, БГТУ)

АЛГОРИТМ ДЛЯ НАХОЖДЕНИЯ ВСЕХ ВОЗМОЖНЫХ РЕШЕНИЙ В СЕТЕВОМ ГРАФИКЕ, ЗАДАННОМ В ДЕКАРТОВОЙ СИСТЕМЕ КООРДИНАТ

Цель и задачи. Разработать алгоритм, который будет находить максимальное и минимальное расстояния между начальным узлом