

Проблема намеренных искажений письменного текста в электронных рекламных рассылках (спаме)

А. Е. Власова
ЗАО «Ашманов и Партнеры»
anna@ashmanov.com

К. П. Зоркий
ЗАО «Ашманов и Партнеры»
zorky@ashmanov.com

В докладе рассматривается проблема автоматизированной обработки письменных текстов, в которые внесены намеренные искажения (от намеренных орфографических ошибок до вариативных фрагментов текста). Это специфическая проблема, непосредственно связанная с развитием и совершенствованием новых способов коммуникации, в частности, электронной почты.

В настоящее время электронная почта, являющаяся превосходным информационным каналом, активно используется для продвижения товаров и услуг. В результате, почтовые ящики конечного пользователя забиты электронными рекламными сообщениями или "спамом". В качестве меры противодействия разрабатываются технологии автоматической фильтрации спама. В свою очередь, спамеры вынуждены искать средства обхода подобных контентных фильтров. Другими словами, перед спамерами стоит задача создать N вариантов рекламного текста, которые программа фильтрации будет воспринимать как а) разные, б) не связанные с рекламой, - а человек, читающий письмо, будет воспринимать как а) одинаковые, б) очевидную рекламу.

Одним из распространенных способов решения этой задачи является намеренное искажение текста, позволяющее сообщению проходить через фильтр, не теряя полезной информационной составляющей. В докладе представлена классификация намеренных искажений на материале англоязычных и русскоязычных сообщений.

Спам как эффективный инструмент рекламы

Электронная почта представляет собой современное и высокотехнологичное средство коммуникации. В настоящее время этот коммуникационный канал активно используется не только для обмена информацией, но и для продвижения товаров и услуг, в том числе и для проведения массовых анонимных незапрошенных рекламных кампаний (другими словами - для рассылки *спама*).

Под *спамом* в настоящем докладе понимается исключительно анонимная незапрошенная массовая рассылка электронной почты. Прочие виды массовых почтовых рассылок спамом не

являются, в частности, не могут быть отнесены к спаму так называемые подписные листы, на получение которых пользователь дал согласие, т.е. рассылки, запрошенные пользователем. Рост доли спама в общем объеме электронной переписки объясняется его высокой эффективностью как средства рекламы:

1. спам охватывает огромную потребительскую аудиторию, т.к. количество активных почтовых ящиков в сети исчисляется миллионами;
2. спам обладает высокой скоростью распространения, т.е. рекламное предложение практически моментально доставляется потенциальным покупателям;
3. спамерская рассылка дешево стоит и быстро окупается.

В итоге мы получаем мощный и современный инструмент рекламы. Неудивительно, что с каждым годом объемы спама неуклонно растут.

Спам как угроза развитию электронной почты

Из побочного фактора почтовой коммуникации спам постепенно превратился в крупномасштабную угрозу электронной почте. За последние два года рост объемов спама превзошел самые смелые прогнозы. По данным лингвистической лаборатории "Спамтест" компании "Ашманов и партнеры", на конец 2003 года спам составлял около 75-80% всей входящей корреспонденции в публичных почтовых службах Рунета.

Поскольку потребители оплачивают доставку электронной почты (либо по времени пребывания в Интернете, либо по затраченному трафику), то спам наносит достаточно серьезные материальные убытки как отдельному пользователю, так и в масштабах всей интернет-индустрии. Ущерб от спама постепенно становится сравнимым с потерями, которые приносят пользователям атаки компьютерных вирусов.

Программы фильтрации - современное средство борьбы со спамом

В качестве меры противодействия разрабатываются технологии автоматической фильтрации спама. Современные программы фильтрации анализируют формальную и содержательную (контентную) часть письма и на основе этого анализа делают выводы об отнесении письма к спаму. Уже существует несколько эффективных методик и технологий фильтрации, которые к тому же постоянно совершенствуются. В данном докладе не будут рассматриваться методы обработки и анализа формальных характеристик (так называемых "заголовков") письма. Основное внимание здесь уделяется проблемам контентной фильтрации, т.е. анализу содержания письма.

Спамеры вынуждены искать средства обхода подобных фильтров. На уровне теоретической проблемы задача сводится к созданию нескольких вариантов одного и того же рекламного текста, которые будут удовлетворять следующим условиям:

1. программа фильтрации будет воспринимать эти варианты текста как
 - разные,
 - не связанные с рекламой;
2. человек, читающий письмо, будет воспринимать те же самые варианты текста как
 - одинаковые,
 - очевидную рекламу;
3. тексты, удовлетворяющие двум условиям, изложенным выше, должны порождаться автоматически, в противном случае существенно возрастут затраты на спамерскую рассылку.

Задача, как видим, совершенно небанальная. На первый взгляд она, возможно, даже покажется неразрешимой. Но в действительности уже существует несколько эффективных приемов, позволяющих спамерам автоматически генерировать тексты и их варианты, способные успешно обходить почтовые фильтры.

Намеренные искажения текста - эффективный способ обхода почтовых фильтров

Одним из распространенных способов решения этой задачи является намеренное искажение текста, позволяющее рекламному сообщению проходить через фильтр, не теряя полезной информационной составляющей. Намеренное искажение может существовать как на уровне слова, так и на уровне текста в целом. Типичным примером намеренного искажения может служить написание слова *viagra* с использованием символа @ вместо латинской буквы "a": *vi@gra*

Намеренное искажение - это абсолютно новый прием оформления текста, характерными особенностями которого являются:

1. использование исключительно в письменных текстах;
2. минимальное воздействие на читателя, в идеале подразумевается отсутствие воздействия, т.е. "незаметность" для читателя;
3. наличие разных уровней искажений (от слова до текста);
4. специфическое предназначение - для обмана автоматических проверяющих программ;
5. автоматическая (программная) генерация в текстах.

Использование искаженного написания слов как стилистический прием достаточно давно известно в теории и практике рекламного дискурса. Но если в традиционной рекламе искажение в написании слов используется для привлечения внимания адресата к той или иной коммуникативной составляющей текста, то в спамерской рекламе намеренные искажения преследуют совершенно иные цели. Они предназначены для решения конкретных практических задач: помешать автоматическому анализатору идентифицировать текст как относящийся к той или иной тематике, т.е. избежать распознавания ключевых слов конкретной тематики, или не допустить отождествления двух и более текстов как "одинаковых".

Намеренное искажение на уровне слова заключается в использовании одних схожих по начертанию букв и/или цифр вместо других, например, c1alis вместо cialis, использовании точек, пробелов и пр. для разбиения слова, например, "с_п_а_м" вместо "спам" и т.п. Намеренное искажение на уровне текста предполагает использование синонимичных вариантов фрагментов текста (фраз или предложений), из которых автоматически генерируется окончательный текст сообщения при сохранении его общего смысла и коммуникативной направленности.

Типология намеренных искажений на уровне слова

Типология намеренных искажений на уровне слова в самом общем виде сводится к трем крупным типам искажений:

- *замена* одних элементов слова (букв) на другие;
- *разбиение* слова небуквенными символами;
- *вставка* в слово "лишних" элементов.

При замене одних элементов слова другими могут использоваться буквы, принадлежащие языку, отличному от языка текста в целом, небуквенные символы и цифры. Например, на массиве из сотен тысяч англоязычных спамерских писем были зафиксированы следующие варианты написания слова *viagra* с использованием небуквенных символов, цифр и букв с надстрочными знаками (не используемыми в алфавите английского языка): *vi@gra*, *viagr@*, *vi@gr@*, *vi("gra*, *v!agra*, *v|agra*, *v\agra*, *viágra*, *viagrá*, *viágrá* и т.п.

Приведем пример спамерского письма, содержащего рекламу пищевой добавки для снижения веса. Практически все слова в тексте данного письма содержат замены букв английского алфавита на буквы с диакритиками, принадлежащие алфавитам других европейских языков:

On átkíns or the south beach díet, try our díet pátcĥ. A new cutting edge, advanced áppétíte sùpprèssant, mètabólism bôôster, and ènérgy ènháncer...all in one. The perfect supplémènt to ássist you in lôsíng those extrá pôũnds just in time for sùmmèr
Lèarñ the trùth about losing wèight.
All ordérs backéd by our nó rísk, monéy back Gmarántèe!
Shipped Discrèetly.

В русскоязычном спаме активно используется замена русских букв на совпадающие по начертанию латинские. Например, в слове *рассылка* могут быть использованы латинские буквы *p*, *a* и *c*, что, однако, остается незаметным для читателя. Также активно применяется замена на латинские буквы, сходные по начертанию, например, в слове *гоставка* использованы латинские буквы *g*, *m* и *k*, которые читатель легко определяет "на глаз" и классифицирует как буквы чужого алфавита, но при этом однозначно идентифицирует с русскими буквами *д*, *т* и *к*.

Еще одним способом замены символов является замена цифр на слова в контактных телефонах, которые спамеры размещают в своей рекламе. Например, Центр американского английского, пользующийся печальной славой "спамера № 1" в Рунете, использует в своей рекламе следующие варианты написания телефона: *1 0 пять пять 1 8 6, один 0 пять пять 1 8 шесть, 1 0 пять пять 1 восемь 6* и так далее.

Второй тип намеренного искажения слова - разбиение - предполагает деление слова на несколько частей таким образом, чтобы автоматические анализаторы не могли отождествить данное слово с одним из ключевых слов списка, но чтобы читаемость текста сохранялась. В качестве "разбивающих" символов чаще всего используются знаки препинания, которые для большинства программ автоматической обработки текста являются показателями границы слова. С точки зрения программы слово оказывается разбитым на несколько более коротких слов. Например, возможно следующее разбиение слова "рассылка": *р а с с ы л к а, р-а-с-с-ы-л-к-а, р.а.с.с.ы.л.к.а, р асс ы лка*.

Ниже помещен пример спамерского письма, в тексте которого использован прием разбиения. "Разбивающим" символом служит пробел, случайным образом расставленный в словах:

Ко неч но, это не совсем че ст но, но ка ко й му ж чи на не
ме чт ал тай но по дсы па ть в в ин о св ое й и зб ра нн иц ы
не ч то та кое, от че го он а бы на бр ос ил ась на не го в
со ст оян ии се кс уал ьн ого исс туп лен ия . . .
Пр ед ла га ем В а м, РЕ АЛ ЫН ОЕ, СИ ЛЬ НО ДЕ ЙС ТВ УЮ
ЩЕЕ СР ЕД СТ ВО...
...МО МЕ НТ АЛ ЫН О РА СТ ВО РЯ ЕТ СЯ В ЛЮ БО М НА
ПИ ТК Е, не ос та вля я ни ед ин ого пр ив ку са!

Третий тип намеренных искажений, - вставка "лишних" элементов, - встречается, в основном, в письмах формата HTML, так как этот формат позволяет оформлять письмо шрифтами разных гарнитур, использовать различные цветовые эффекты. Достаточно распространенным приемом является вставка букв того же цвета, что и фон письма, например, белым шрифтом по белому фону. Читатель воспринимает такие слова как набранные в разрядку. Например, если между буквами слова **VIAGRA** вставлена буква **W** (**VWIIWAWGWRWA**), и при этом каждой букве **W** задан атрибут "белый цвет", и общий фон письма тоже белый, то читатель увидит слово так, как будто оно набрано в разрядку: **V I A G R A**. При этом даже мысли о наличии лишних символов у пользователя не возникнет.

В русскоязычном спаме активно используется такая разновидность вставок, как удвоение букв. В спамерской терминологии этот прием носит название "дребезг клавиатуры". Текст, набранный с применением этой разновидности искажений может выглядеть следующим образом:

«Фрегат» доставит на доом, в офис, на банкет икру чеерную, красную, крабы по цеенам ниже рыноочных
--

Крайним случаем использования вставок является написание текста вообще без пробелов. Слова разделены символом или последовательностью символов, набранными шрифтом белого цвета по белому фону. Для читателя такое письмо выглядит как обычный текст, в котором между словами имеются белые промежутки, воспринимаемые как пробелы. На самом деле, это тоже буквы, только белого цвета. Для большинства антиспамерских фильтров это не текст, а одно длинное слово, причем в каждом письме такое слова будет разным т.к. "лишние" буквы вставляются в текст случайным образом.

Стремление спамеров обмануть контентные фильтры иногда приводит к анекдотическим результатам. Текст письма искажается настолько, что страдает его читаемость. Так произошло, например, со следующей рекламой сомнительных лекарств:

VI\GR/ \$0.95 /l)0SE
C /L S \$2.00 /l)0SE
} {E C/ L \$0.91 /l)()SE
PR()PECI GR^, GLUC()PH GR/ GE, V 0} {} {,
CELEBRE} {, V ERI)I ^, Z()LOFF, P/l} {IL, LIP T()R
ENTER

Поскольку этот текст получился практически нечитаемым, ниже дан перевод на "обычный" английский язык:

VIAGRA \$0.95 per dose
CIALIS \$2.00 per dose
XENICAL \$0.91 per dose
... etc.
ENTER

Перспективы развития антиспамерского программного обеспечения

Разработчики программ контентной фильтрации практически оказались вынужденными искать способы отождествления искаженного и "нормального" текстов. В настоящее время намечено как минимум два пути решения этой проблемы:

1. совершенствование алгоритмов "вычленения" текста письма, например, добавление анализа цветовых различий фона и шрифтов;
2. составление своеобразных вариативных словарей, где в явном виде перечисляются все варианты искаженного написания отдельно взятого слова.

Очевидно, каждый из этих путей имеет и недостатки, и преимущества. Наиболее оптимальное направление разработки состоит в том, чтобы учесть оба варианта. Естественно, это приведет к усложнению средств фильтрации, а это, в свою очередь, значит, что задача борьбы со спамом становится решаемой только с использованием серьезного профессионального антиспамерского ПО.