УДК 004.056.5 + 656.2:004.056.5 Вестник УрФО № 4(26) / 2017, с. 16–20

Паршин К. А., Подгорный М. С.

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПРЕДПРИЯТИЯ ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА ПУТЕМ МОНИТОРИНГА ТЕКСТОВЫХ ПУБЛИКАЦИЙ В ОТКРЫТЫХ ИСТОЧНИКАХ ДАННЫХ

Статья посвящена рассмотрению проблемы распространения информации ограниченного доступа путем публикации текстовых данных в открытых источниках, таких как веб-порталы, интернет-сервисы и социальные сети. Основной предметной областью в статье является сфера железнодорожного транспорта. Рассмотрены методы работы с текстовыми данными, применяемые в системах предотвращения утечек информации, которые могут быть использованы в альтернативных программных продуктах. Акцент сделан на уникальность терминологии, используемой только в железнодорожной отрасли на всей территории Российской Федерации. Определен перечень признаков, характерных только к сокращениям железнодорожных объектов и субъектов.

Ключевые слова: *информационная безопасность, железнодорожный транспорт, мониторинг публикаций, открытые источники данных, методы анализа текстовых данных, терминология железнодорожной отрасли.*

ENSURING INFORMATION SECURITY OF RAILWAY TRANSPORT ENTERPRISE BY MONITORING TEXT PUBLICATIONS IN OPEN DATA SOURCES

The article is devoted to problems of dissemination limited access information by publishing text data in open sources, such as web portals, Internet services and social networks. The main subject area in the article is the sphere of railway transport. The article describes methods of working with text data used in Data leakage prevention systems that can be used in alternative software products. The emphasis of article is on the unique terminology used only in railway industry throughout the Russian Federation. A list of features characteristic only of abbreviations of railway objects and subjects was determined.

Keywords: *information security, railway transport, monitoring of publications, open data sources, methods for analyzing text data, railway industry terminology.*

Сфера железнодорожного транспорта Российской Федерации занимает существенное место в экономике страны. По отчету с официального сайта вклад ОАО «РЖД» составляет 2,5% при среднем обороте денежных средств, практически, в два триллиона рублей. Сеть железных дорог покрывает всю заселенную территорию страны и включает в себя 16 железных дорог. К территории Уральского федерального округа относятся Свердловская и Южно-Уральская железные дороги. Существенным показателем для ОАО «РЖД» является количество сотрудников предприятия – на 2017 год данный показатель достигает 890 тысяч человек[1].

Для любого предприятия с большим количеством сотрудников важным является вопрос обеспечения высокого уровня информационной безопасности для документов и данных, оборачиваемых внутри компании. Особо чувствительным моментом для подобного предприятия является публикация данных, содержащих информацию ограниченного доступа в сети Интернет, так как помимо самого распространения информации, происходит и негативное представление компании в открытой сети, что несет за собой различные экономические и социальные последствия. Примером подобных публикаций может быть случай закрепления на одном из публичных сайтов служебной переписки между поездным и станционным диспетчерами, а также машинистом локомотива после столкновения пассажирского и пригородного поездов на Московской железной дороге[2]. Статья на сайте содержала в себе практически всю текстовую запись разговора между причастными сотрудниками и была удалена лишь через неделю после публикации. Ключевым моментом в записи разговора было то, что в тексте использовались сокращения должностей, которые могут быть использованы только на железнодорожном транспорте (ДС, ДНЦ, ДСП).

Далее в работе рассматриваются методы по поиску текстовых записей на определенных публичных сайтах сети Интернет с целью уменьшения «времени жизни» подобных публикаций, что приведет к снижению негативного эффекта для компании.

Важным программным компонентом в области защиты информации на любом крупном предприятии являются системы предотвращения утечек информации *(Data Leak Prevention, DLP)*. Богатый функционал данных систем (методы работы с текстом, *OCR*, аудио-

и видеоанализ) позволяет определить возможную подобную запись еще на уровне закрепления текста публикации на том или ином форуме или профиле в социальной сети. Однако по большей части весь поиск и предотвращение возможны лишь в зоне действия внутренней сети предприятия (рис. 1). Статья может быть закреплена и со своего личного персонального компьютера или мобильного устройства через любую другую точку доступа. Именно поэтому, а также ввиду экономических ситуаций, предприятию требуются альтернативные методы и программные средства для поиска данных записей.

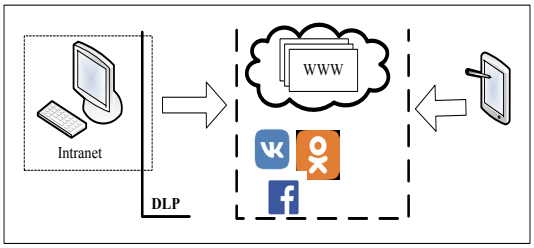

Рис. 1. Условная схема возможных публикаций

Однако стоит обратить внимание на программно-аналитические методы, которые используются в системах предотвращения утечек информации для работы с текстовыми данными:

– поиск по регулярным выражениям;

– предустановленные текстовые шаблоны;

– предустановленные тематические словари.

Именно на данных методах сделан упор в текущей работе ввиду особенности описания объектов и субъектов на железнодорожном транспорте. В отличие от систем предотвращения утечек информации местом поиска и анализа данных является не источник информации, а именно получатель информации, то есть заранее определенный пополняемый перечень веб-сайтов, интернет-сервисов, а также профилей пользователей и групп в социальных сетях. Целью мониторинга является именно уменьшение времени нахождения публикации в открытом доступе.

Как уже говорилось ранее, в сфере железнодорожного транспорта существует определенная терминология для описания тех или иных объектов или субъектов, единая на всей территории страны. Например, для описания должности поездного диспетчера используется сокращенное наименование ДНЦ. Термин не является какой-либо расшифровкой и имеет свои исторические корни[3].

Аналогичные сокращения имеют и объекты инфраструктуры на железнодорожном транспорте, например ДЦС или ВЧД. Уникальностью описания обладают и данные передаваемые в информационных системах. Любой документ, передаваемый по внутренним каналам связи, содержит как минимум телеграфный код причастных дирекций или служб, а также шифр исполнителя данного документа. Все это говорит о том, что предметная область в части железнодорожной терминологии заслуживает большого внимания при работе с текстовыми данными.

Регулярные выражения могут быть использованы при поиске и анализе следующих специфических элементов в общем тексте:

– телеграмма натурный лист грузового поезда (ТГНЛ) – уникальный цифровой код, описывающий содержание вагонов в грузовом поезде[4];

– сообщения системы АСОУП[5] – цифровой код, содержащий уникальные комбинации цифр и знаков пунктуации.

Кроме информационных систем, уникальностью и синтаксическими особенностями обладают и сами термины. Первой отличительной чертой железнодорожной терминологии является то, что объекты имеют определенную условную иерархичность (рис. 2).

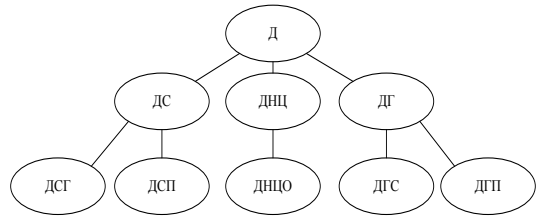

Рис. 2. Иерархичность в описании субъектов службы перевозок

Вторым признаком является условное наследование. Например, следующее описание должностей Службы перевозок:

– Д – Служба перевозок;

– ДС – начальник станции;

– ДСП – дежурный по станции;

– ДСПГ – дежурный по сортировочной горке;

– ДСПГО – оператор при дежурном по сортировочной горке.

Третьей особенностью является именно синтаксический состав и порядок букв в со-

кращении железнодорожных объектов и субъектов. При анализе выборки терминов[6], состоящей из 500-600 сокращений, была получена следующая статистика:

– общее количество символов в выборке равно 1583;

– общее количество согласных букв в выборке 81,81 %;

– количество терминов, начинающихся с гласной буквы 19,47 %;

– количество терминов, заканчивающихся гласной буквой 15,97 %.

Другими словами при текстовом анализе данных важно обращать внимание именно на наполнение и расположение в словах (токенах) согласных букв. Ниже представлена гистограмма (рис. 3) частоты встречи шаблонных терминов, содержащих согласные и гласные буквы («С» и «Г» соответственно).

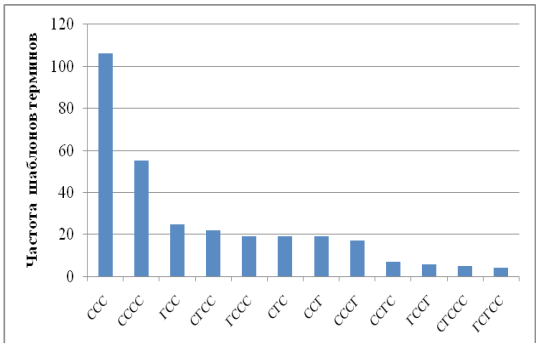

Рис. 3. Частотные показатели шаблонов терминов

Из графика видно, что наибольшей частотой обладают термины, состоящие из трех и четырех согласных букв («ССС» и «СССС» соответственно).

Из вышеописанных характеристик можно сделать вывод, что для поиска железнодорожных терминов могут быть использованы как заранее подготовленные словари, так и определенные подготовленные шаблоны.

При формировании тезауруса могут быть использованы следующие источники данных:

– специализированная железнодорожная литература;

– нормативные и правовые документы компании, а также дочерних и зависимых обществ;

– информационные системы;

– иные источники информации.

Синтаксические характеристики могут быть использованы при формировании заранее подготовленных текстовых шаблонов или «масок» поиска, аналогичных регулярным выражениям, но являющихся более гибкими при настройке. При работе с данными шаблонами важным моментом может быть ложное срабатывание в процессе определения термина, так как сокращение может относиться к сфере строительства, авиаперевозок или другой подобной сфере деятельности. Другими словами, в результате может быть получена неверная классификация термина.

Подведя итоги можно сказать, что уникальность терминологии железнодорожного транспорта содержит в себе большое количество признаков, которые могут быть использованы при мониторинге текстовых записей в сети интернет с использованием методов анализа текстовых данных. В дальнейших исследованиях планируется более подробно раскрыть тематику применения терминологии железнодорожного транспорта и через программные компоненты выполнить контрольные проверки методов при получении данных с профилей пользователей социальных сетей.

Литература

1. Показатели основной деятельности [Электронный ресурс] // официальный сайт, 2017. URL: http://ir.rzd.ru/static/public/ru?STRUCTURE_ID=63 (дата обращения: 10.11.2017).

2. Про столкновение электрички и поезда [Электронный ресурс] // форум, 2017. URL: http://www.yaplakal.com/forum15/st/175/topic1579951.html (дата обращения: 12.04.2017).

3. Общий курс железных дорог: Учебник для техникумов и колледжей ж.-д. транспорта / В.Н. Соколов, В.Ф. Жуковский, С.В. Котенкова, А.С. Наумов; Под редакцией В.Н. Соколова. — М.: УМК МПС России, 2002. С. 180—200.

4. Телеграмма-натурный лист поезда (ТГНЛ) [Текст] : методические указания / [С. А. Бессоненко и др.] ; Сибирский гос. ун-т путей сообщения (СГУПС).—2-е изд., измененное и доп. — Новосибирск: СГУПС, 2015. С. 30.

5. Санькова Г.В. Информационные технологии в перевозочном процессе: учебное пособие / Г.В. Санькова, Т.А. Одуденко. — Хабаровск: Изд-во ДВГУПС, 2012. — С. 64.

6. Железнодорожный словарь [Электронный ресурс] // форум, 2017. URL: http://rzd.me/inform-block/zhd-slovar/ (дата обращения: 20.10.2017).

References

1. Pokazateli osnovnoj deyatelnosti [Elektronnyy resurs] // oficialnyj sajt, 2017. URL: http://ir.rzd.ru/static/public/ru?STRUCTURE_ID=63 (data obrashcheniya: 10.11.2017).

2. Pro stolknovenie ehlektrichki i poezda [Elektronnyy resurs] // forum, 2017. URL: http://www.yaplakal.com/forum15/st/175/topic1579951.html (data obrashcheniya: 12.04.2017).

3. Obshchij kurs zheleznyh dorog: Uchebnik dlya tekhnikumov i kolledzhej zh.-d. transporta / V.N. Sokolov, V.F. ZHukovskij, S.V. Kotenkova, A.S. Naumov; Pod redakciej V.N. Sokolova. — M.: UMK MPS Rossii, 2002. S. 180—200.

4. Telegramma-naturnyy list poezda (TGNL) [Tekst] : metodicheskie ukazaniya / [S. A. Bessonenko i dr.] ; Sibirskyy gos. un-t putej soobshcheniya (SGUPS).—2-e izd., izmenennoe i dop. — Novosibirsk: SGUPS, 2015. S. 30.

5. Sankova G.V. Informacionnye tekhnologii v perevozochnom processe: uchebnoe posobie / G.V. Sankova, T.A. Odudenko. — Habarovsk: Izd-vo DVGUPS, 2012. — S. 64.

6. ZHeleznodorozhnyy slovar [Elektronyy resurs] // forum, 2017. URL: http://rzd.me/inform-block/zhd-slovar/ (data obrashcheniya: 20.10.2017).

ПАРШИН Константин Анатольевич, кандидат технических наук, доцент кафедры «Информационные технологии и защита информации», Уральский государственный университет путей сообщения, 620034, Свердловская обл., г. Екатеринбург, ул. Колмогорова, 66. E-mail: kparshin@usurt.ru

ПОДГОРНЫЙ Михаил Сергеевич, аспирант кафедры «Информационные технологии и защита информации», Уральский государственный университет путей сообщения, 620034, Свердловская обл., г. Екатеринбург, ул. Колмогорова, 66. E-mail: podgorny312@yandex.ru

PARSHIN Konstantin, PhD, associate professor of «Information technologies and information security», Ural State University of Railway Transport, 620034, Sverdlovsk region, Yekaterinburg, Kolmogorova St., 66. E-mail: kparshin@usurt.ru

PODGORNYY Mihail, graduate student of «Information technologies and information security», Ural State University of Railway Transport, 620034, Sverdlovsk region, Yekaterinburg, Kolmogorova St., 66. E-mail: podgorny312@yandex.ru