Проектирование базы данных реестра фейковых сайтов организаций

В. А. Смирнов, email: v.a.d.i.m@bk.ru [1]
А. Н. Привалов, email: privalov.61@mail.ru [2]

[1] Ивановский государственный университет (Шуйский филиал)
[2] Тульский государственный педагогический университет им. Л.Н. Толстого

Аннотация. *В данной работе анализируется возникновение угроз информационной безопасности, связанное с распространением QR-кодов. Как средство борьбы с фейковыми QR-кодами предлагается внедрение списков запрещенных ресурсов, основанных на загрузке информации из реестров. Рассматривается процесс создания базы данных для реестра фейковых сайтов организаций, в том числе: анализ требований, создание структуры и построение SQL-запросов.*

Ключевые слова: *реестр запрещенных сайтов, информационная система, структура базы данных, SQL-запросы, анализ требований.*

Введение

Повсеместным стало распространение передачи информации при помощи QR-кодов (двухмерных штрих-кодов (бар-кодов), предоставляющих информацию для быстрого ее распознавания с помощью камеры на мобильном телефоне [1]). В настоящее время QR-коды предоставляют возможность оплаты счетов по определенным реквизитам при помощи системы быстрых платежей [2]. Активное внедрение QR-кодов для оплаты товаров и услуг предназначено для снижения вероятности ошибки из-за человеческого фактора, связанного с некорректным вводом платежных данных, суммы покупки и т. д. В последнее время QR-коды активно используются и как средство аутентификации [3].

Данные опроса, проведенного в сентябре 2020 года [4], о том, когда респонденты в последний раз сканировали QR-код, представлены на рис. **Ошибка! Источник ссылки не найден.**. Подобное распределение говорит о высокой распространенности QR-кодов в бытовой жизни пользователей смартфонов.

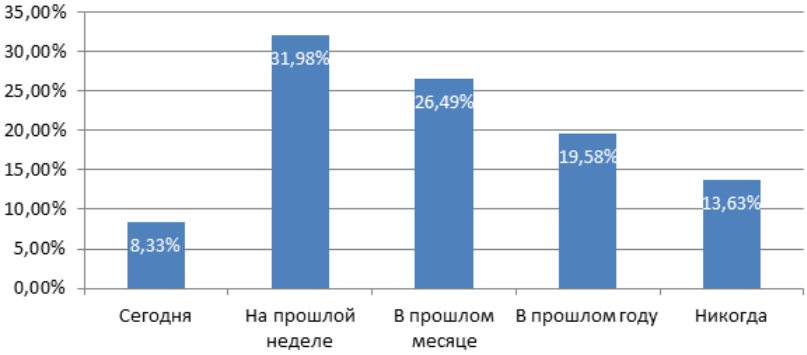

Рис. 1. Доля респондентов, давших указанный ответ на вопрос о QR-кодах

Обратной стороной распространения QR-кодов является возникновение новых угроз информационной безопасности. При вводе какой-либо ссылки вручную или при переходе по ней на компьютере пользователь может обратить внимание на опечатки, которые служат признаками поддельности ресурса [5]. Исходя из приведенного выше определения, QR-коды предназначены для чтения их со смартфона, адресная строка которых может вместить существенно меньшее количество текста, поэтому оценка подлинности ресурса стала сложнее, что приводит к возникновению нового способа информационной атаки, связанной с применением QR-кодов.

В китайской социальной сети Weibo хакеры распространяли фишинговые QR-коды под предлогом раздачи бесплатных игровых аккаунтов [6]. В дальнейшем, после ввода пользователем по закодированной ссылке своих логина и пароля, скомпрометированные аккаунты использовались для распространения рекламы с непристойными материалами. Другой вариант использования фейковых QR-кодов стал популярен во время пандемии. Созданные мошенниками QR-коды были предназначены для перехода на фейковый сертификат о вакцинации [7].

Защита от фейковых QR-кодов различается в зависимости от целей мошенника. Например, если его целью является попытка продать поддельный продукт вместо легального, то производителем может наноситься на подлинный продукт QR-код со встроенным цифровым водяным знаком или шаблоном обнаружения копирования [8].

Одним из более универсальных методов защиты от поддельных ресурсов является внедрение в пользовательские устройства механизма

блокировки при помощи «черных списков». В этом случае Интернет-ресурс, включенный в перечень запрещенных ресурсов, станет недоступен на устройстве пользователя. Поскольку поддельные ресурсы создаются ежедневно, данный перечень должен иметь возможность обновления из открытых источников. Такими источниками могут служить реестры фейковых сайтов, основанные на базах данных, которые позволяют хранить, обрабатывать и предоставлять по запросу информацию об Интернет-ресурсах. В связи с этим, актуальной становится научная задача проектирования базы данных реестра фейковых сайтов организаций.

1. Проектирование структуры базы данных

Созданию структуры базы данных предшествует построение требований к функционалу реестра фейковых сайтов. Одним из средств формализации требований является UML-диаграмма вариантов использования. Построенная в процессе работы диаграмма для реестра фейковых сайтов организаций представлена на рис. 2.

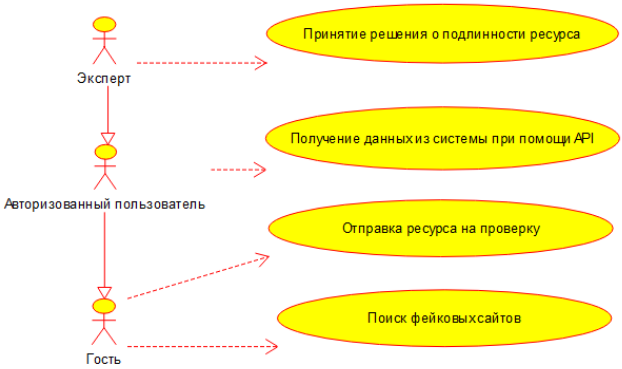

Рис. 2. Диаграмма вариантов использования реестра

Исходя из предположенных вариантов использования системы и на основе анализа сходных Интернет-ресурсов (например, PhishTank) была разработана структура базы данных реестра. Реляционная модель базы данных реестра фейковых сайтов организаций представлена на рис. 3.

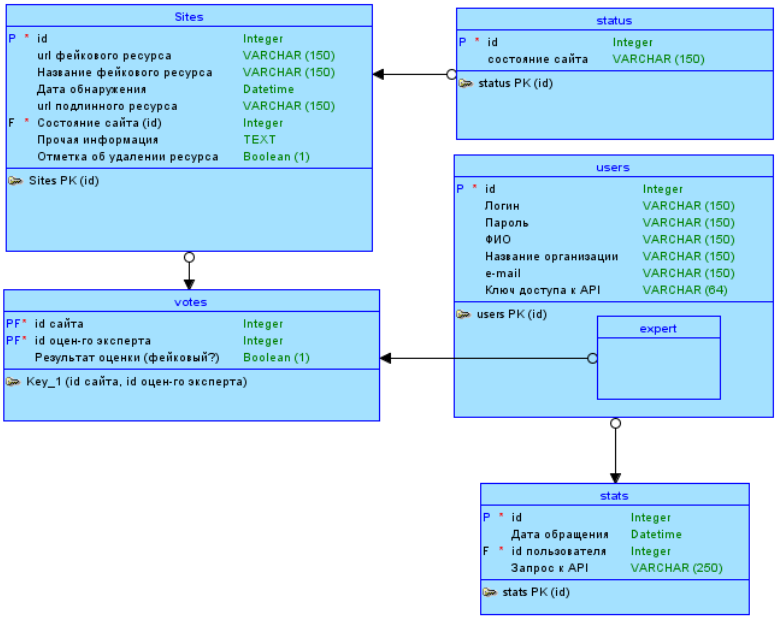

Рис. 3. Структура базы данных реестра

В этой базе предусмотрено хранение аккаунтов пользователей (users), данных о фейковых сайтах в реестре (sites), о результатах оценки сайтов экспертами (votes), текущем состоянии сайта – работает, заблокирован хостинг-провайдером и др. (status), статистике обращений к реестру при помощи ключа API (stats). Последняя таблица необходима для более детального анализа объема использования реестра в будущих исследованиях.

2. Построение запросов на языке SQL

Следующим необходимым этапом для использования спроектированной базы данных является создание представлений и построения запросов манипулирования данными. При этом различные команды для получения данных на языке SQL связаны с применением операций реляционной алгебры. В частности, запрос для получения списка сайтов, признанных фейковыми, на языке реляционной алгебры будет состоять из двух частей и выглядеть следующим образом:

1) Представление, позволяющее выделить идентификаторы сайтов, которые не менее трех экспертов признали фейковыми:

$$FAKE\_SITES = \rho_{site\_id,id}\left(\sigma_{SUM(result)\geq 3}\left(\gamma_{site\_id,\ SUM(result)}(\sigma(VOTES))\right)\right) \quad (2)$$

2) Основной запрос:

$$\rho_{\substack{STATUS.name,\\ status}}\left(\pi_{\substack{SITES.url,\\ SITES.name,\\ STATUS.name,\\ SITES.dt,\\ SITES.original}}\left(\sigma_{FAKE\_SITES.id = SITES.id}\left(\left(\sigma_{STATUS.id = SITES.status\_id \wedge SITES.visible = 1}\begin{pmatrix} SITES \\ \times \\ STATUS \end{pmatrix}\right) \times FAKE\_SITES\right)\right)\right) \quad (3)$$

Как показано выше, при получении данной выборки используется представление для получения id сайтов, которые были признаны фейковыми как минимум тремя экспертами. После этого отбираются данные сайтов, входящих в указанное множество идентификаторов. Ограничение количества выводимых сайтов и постраничный вывод данных обеспечивается при помощи оператора LIMIT.

Основываясь на полученных выражениях реляционной алгебры, можно построить соответствующие команды на языке SQL. Таким образом, выборка полного списка сайтов может быть осуществлена с помощью запросов на рис. 4-5.

```
CREATE VIEW fake_sites AS (
        SELECT `site_id` AS `id`
                FROM `votes` GROUP BY `site_id`
                HAVING SUM(`result`) >= 3)
```

Рис. 4. Представление для получения идентификаторов сайтов, которые не менее трех экспертов признали фейковыми

```
SELECT `url`, `sites`.`name`,
        `status`.`name` AS `status`, `dt`,
        `original`, `other`
        FROM `sites`, `status`, `fake_sites`
        WHERE (`status`.`id` = `sites`.`status_id`)
            AND (`sites`.`visible` = '1')
            AND (`sites`.`id` = `fake_sites`.`id`)
```

Рис. 5. Запрос для получения списка сайтов, признанных фейковыми

На странице поиска сайтов в SQL-команду добавляются дополнительные условия, построенные с использованием MySQL-команды INSTR. При помощи этого обеспечивается фильтрация выводимых ресурсов по введенным пользователем URL-адресу (или части адреса) фейкового сайта, названию, URL-адресу (или части адреса) подлинного сайта.

Другой задачей при организации взаимодействия приложения с базой данных реестра фейковых сайтов организаций является построение запроса для вывода перечня сайтов в личном кабинете эксперта. Результаты построения целесообразно представить на языке реляционной алгебры в трех частях:

1) Вложенный подзапрос – идентификаторы сайтов, за которые уже проголосовал эксперт:

$$VOTES\_EXPERTS = \rho_{site\_id,id}\left(\pi_{site\_id}\left(\sigma_{user\_id="\{\$id\}"}(VOTES)\right)\right) \quad (4)$$

2) Вложенный подзапрос – идентификаторы сайтов, за которые ещё не проголосовал эксперт:

$$UNVOTES\_EXPERTS = \pi_{id}\left(\sigma_{visible="1"}(SITES)\right) - VOTES\_EXPERTS \quad (5)$$

3) Основной запрос:

$$\left(\rho_{"0",\ verified}\left(\pi_{SITES.id,\ SITES.url,\ SITES.name,\ SITES.status\_id,\ SITES.dt,\ SITES.original,\ SITES.other}\left(\sigma_{SITES.id=UNVOTES\_EXPERTS.id}\left(SITES \times UNVOTES\_EXPERTS\right)\right)\right)\right) \cup \left(\rho_{"1",\ verified}\left(\pi_{SITES.id,\ SITES.url,\ SITES.name,\ SITES.status\_id,\ SITES.dt,\ SITES.original,\ SITES.other}\left(\sigma_{SITES.id=VOTES\_EXPERTS.id}\left(SITES \times VOTES\_EXPERTS\right)\right)\right)\right) \quad (6)$$

Реализация в базе данных MySQL основного запроса представлена на рис. 6. При этом стоит учитывать, что MySQL не поддерживает оператор MINUS.

```
SELECT `sites`.`id`, `sites`.`url`,
       `sites`.`name`, `sites`.`status_id`, `sites`.`dt`,
       `sites`.`original`, `sites`.`other`, 0 as `verified`
     FROM sites, (
          SELECT `id` FROM `sites`
          WHERE (`visible`='1') AND
              (`id` NOT IN (
                  SELECT `site_id` AS `id`
                      FROM `votes` WHERE `user_id`='{$id}')
              )
          ) AS unvotes_experts
     WHERE `sites`.`id` = unvotes_experts.id
UNION
SELECT `sites`.`id`, `sites`.`url`, `sites`.`name`,
       `sites`.`status_id`, `sites`.`dt`,
       `sites`.`original`, `sites`.`other`, 1 as `verified`
     FROM sites, (
          SELECT `site_id` AS `id` FROM `votes`
          WHERE `user_id`='{$id}'
          ) AS votes_experts
     WHERE `sites`.`id` = votes_experts.id
```

Рис. 6. Запрос для получения списка сайтов для оценки экспертом

После упрощения данного запроса за счёт использования встроенных средств MySQL для вывода перечня сайтов в личном кабинете эксперта используется запрос на рис. 7.

```
SELECT `id`, `url`, `name`, `status_id`,
       `dt`, `original`, `other`,
       IF (`id` NOT IN (SELECT `site_id` AS `id`
                        FROM `votes`
                        WHERE `user_id`='{$id}'
                       ),
           '0', '1') AS `verified`
           FROM `sites`
           WHERE  (`visible` = '1')
           ORDER BY `verified`, `id`
```

Рис. 7. Сокращенный запрос для получения списка сайтов для оценки экспертом

В зависимости от значения столбца verified сайт будет отображаться в личном кабинете эксперта как проверенный сайт или как ресурс, требующий оценки. Дальнейший вывод и форматирование данных осуществляется средствами приложения, работающего с базой данных. В случае реестра фейковых сайтов организаций наиболее подходящей формой реализации такого приложения является web-сайт, то есть использование языка разметки HTML, языка оформления CSS, языков программирования JavaScript и PHP.

Заключение

Предложенная структура базы данных и запросы к ней могут быть использованы при создании реестра фейковых сайтов организаций. Использование такого реестра в качестве источника информации для защитных антифишинговых решений позволит повысить степень защиты пользователей от информационных угроз фейковых сайтов организаций.

Список литературы

1. Рыжко К. В. Использование информационно-коммуникационных технологий на уроке географии // От цифровизации к цифровой трансформации : Материалы VI Международной научно-практической конференции, Миасс, 28 января 2022 года. – Челябинск: Челябинский институт развития профессионального образования, 2022. – С. 229-232.

2. Оплата по QR-коду: как это работает и кому подходит [Электронный ресурс]. – Режим доступа: https://sbis.ru/articles/retail/oplata_po_qr_kodu_kak_eto_rabotaet

3. Вход с двухфакторной аутентификацией QR код - QR code [Электронный ресурс]. – Режим доступа: https://yandex.ru/support/id/authorization/twofa-login.html

4. When was the last time you scanned a QR code? [Электронный ресурс]. – Режим доступа: https://www.statista.com/statistics/199334/us-qr-code-scanners-last-time-scanned/

5. Привалов А. Н., Смирнов В. А. Метод нечеткого сравнения строк для обнаружения фейковых сайтов // Известия Тульского государственного университета. Технические науки. – 2022. – № 2. – С. 184-191. – DOI 10.24412/2071-6168-2022-2-184-191.

6. Китайский техногигант Tencent подвергся атаке с использованием фишинговых QR-кодов [Электронный ресурс]. – Режим доступа: https://www.securitylab.ru/news/532539.php

7. Штраф или срок грозит за фейковый QR-код. Нарушителей спасет только человеческий фактор [Электронный ресурс]. – Режим

доступа: https://360tv.ru/tekst/obschestvo/shtraf-ili-srok-grozjat-za-fejkovyj-qr-kod-narushitelej-spaset-tolko-chelovecheskij-faktor/

8. QR код - QR code [Электронный ресурс]. – Режим доступа: https://ru.zahn-info-portal.de/wiki/QR_code#Counterfeit_detection