

# НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 1. ОРГАНИЗАЦИЯ И МЕТОДИКА  
ИНФОРМАЦИОННОЙ РАБОТЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 11

Москва 2019

## ОБЩИЙ РАЗДЕЛ

УДК 34:004.774.056

Л.В. Астахова, И.А. Медведев

### Инструментальный мониторинг соответствия веб-сайтов требованиям законодательства в области персональных данных\*

*Показана актуальность разработки инструментального средства мониторинга веб-сайтов на соответствие требованиям законодательства в области персональных данных. Описан результат программной реализации такого средства. Отмечены преимущества разработанного продукта по сравнению с ручным мониторингом.*

**Ключевые слова:** информация, безопасность, защита информации, веб-сайт, персональные данные, мониторинг, законодательство, машинное обучение

#### ВВЕДЕНИЕ

Законодательство в области персональных данных и механизмы контроля за его исполнением в Российской Федерации активно развиваются после приня-

тия ФЗ №152 «О персональных данных»<sup>1</sup>, что отражено в деятельности органов государственной власти, образовании, здравоохранении и др. Так, согласно действующему законодательству, информация о деятельности государственных органов и органов ме-

\* Статья выполнена при поддержке Правительства РФ (Постановление №211 от 16.03.2013 г.), соглашение № 02. А03.21.0011

<sup>1</sup> Федеральный закон "О персональных данных" от 27.07.2006 № 152-ФЗ. – URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_61801/](http://www.consultant.ru/document/cons_doc_LAW_61801/) (дата обращения: 10.09.2019).

стного самоуправления, размещаемая ими в сети Интернет, должна содержать: сведения о руководителях названных органов, (фамилии, имена, отчества, а также при согласии указанных лиц иные сведения о них, должность); их структурные подразделения; информацию о должностном лице, ответственном за работу с обращениями граждан (фамилию, имя и отчество руководителя подразделения или иного должностного лица, номер телефона, по которому можно получить информацию справочного характера<sup>2</sup>.

На официальных сайтах медицинских организаций в информационно-телекоммуникационной сети Интернет (далее – сеть Интернет) должна размещаться информация о медицинских работниках, необходимая для проведения независимой оценки качества оказания ими услуг (фамилия, имя, отчество, занимаемая должность; сведения из документа об образовании (уровень образования, организация, выдавшая документ об образовании, год выдачи, специальность, квалификация); сведения из сертификата специалиста (специальность, соответствующая занимаемой должности, срок действия сертификата); график работы и часы приема медицинского работника); отзвы потребителей услуг<sup>3</sup>.

Сайт учреждения образования должен содержать следующую информацию: 1) о руководителе образовательной организации, его заместителях, руководителях филиалов образовательной организации (при их наличии): фамилия, имя, отчество (при наличии) руководителя, его заместителей; должность руководителя, его заместителей; контактные телефоны; адрес электронной почты; 2) о персональном составе педагогических работников: фамилия, имя, отчество (при наличии) работника; квалификация, опыт работы, занимаемая должность (должности); преподаваемые дисциплины; ученая степень (при наличии); ученое звание (при наличии); направление подготовки и (или) специальности; данные о повышении квалификации и (или) профессиональной переподготовке (при наличии); общий стаж работы; стаж работы по специальности<sup>4</sup>. Другие сведения, являющиеся

персональными данными, на сайтах названных организаций размещаются согласно ФЗ-152 «О персональных данных» (ст.8.1) только с письменного согласия субъекта персональных данных.

На официальном сайте финансово-кредитных организаций должна раскрываться информация о квалификации и опыте работы членов совета директоров, лиц, занимающих должности единоличного исполнительного органа, его заместителей и иных должностных лиц, в частности: фамилия, имя, отчество (при наличии), должность, квалификация, специальность, год окончания образовательной организации, ученая степень и ученое звание (при наличии), дата присуждения ученой степени (ученого звания), сведения о трудовой деятельности за 5 лет до назначения на должность и пр.<sup>5</sup>

Проведенный авторами настоящей статьи анализ российских веб-сайтов, показал, что некоторые сайты не соответствуют требованиям законодательства в области персональных данных. Это свидетельствует об острой необходимости постоянного мониторинга контента сайтов. В условиях цифровизации общества и больших объемов анализируемой информации ручной мониторинг не эффективен. Поэтому необходимы полифункциональные, оперативные, экономичные, удобные в использовании инструментальные средства мониторинга сайтов, созданные на основе новейших информационных технологий. Этим обусловлена цель статьи – обосновать методику и результаты разработки средства мониторинга соответствия контента веб-сайта требованиям законодательства и апробировать его на практике.

## ОБЗОР ЛИТЕРАТУРЫ ПО ПРОБЛЕМЕ

Стремительные темпы цифровизации информационных ресурсов в глобальном масштабе определили интенсивные процессы развития автоматизации *Open Source Intelligence (OSINT)* – разведки на основе открытых источников [1, 2]. Существуют приложения и веб-сервисы, которые автоматизируют процесс извлечения данных [3].

<sup>2</sup> Федеральный закон "Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления" от 09.02.2009 № 8-ФЗ (ст.13 п.1 ч.д.). – URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_84602/](http://www.consultant.ru/document/cons_doc_LAW_84602/) (дата обращения: 10.09.2019).

<sup>3</sup> Приказ Министерства здравоохранения РФ от 30 декабря 2014 г. № 956н "Об информации, необходимой для проведения независимой оценки качества оказания услуг медицинскими организациями, и требованиях к содержанию и форме предоставления информации о деятельности медицинских организаций, размещаемой на официальных сайтах Министерства здравоохранения Российской Федерации, органов государственной власти субъектов Российской Федерации, органов местного самоуправления и медицинских организаций в информационно-телекоммуникационной сети «Интернет» (Прил.1, п.6, п.12). – URL: <https://www.garant.ru/products/ipo/prime/doc/70777306/> (дата обращения: 10.09.2019).

<sup>4</sup> Постановление Правительства РФ от 10.07.2013 № 582 "Об утверждении Правил размещения на официальном сайте образовательной организации в информационно-телекоммуникационной сети «Интернет» и обновления

информации об образовательной организации" (ч. А, п.3). – URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_149242/](http://www.consultant.ru/document/cons_doc_LAW_149242/) (дата обращения: 10.09.2019).

<sup>5</sup> Указание Банка России от 19 мая 2015 г. № 3639-У "О порядке раскрытия кредитной организацией информации о квалификации и опыте работы членов совета директоров (наблюдательного совета) кредитной организации, лиц, занимающих должности единоличного исполнительного органа, его заместителей, членов коллегиального исполнительного органа, главного бухгалтера, заместителя главного бухгалтера кредитной организации, а также руководителя, главного бухгалтера филиала кредитной организации, на официальном сайте кредитной организации в информационно-телекоммуникационной сети "Интернет" (с изменениями и дополнениями). – URL: <https://base.garant.ru/71097560> (дата обращения: 10.09.2019); Федеральный закон "О банках и банковской деятельности" от 02.12.1990 N 395-1 (ст. 8). – URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_5842/](http://www.consultant.ru/document/cons_doc_LAW_5842/) (дата обращения: 10.09.2019).

Инструменты *OSINT* используются в сфере информационной безопасности. Специалисты разработали системы автоматизации поиска информации о людях в сети Интернет на основе использования открытых источников [4]; показали возможности фреймворка *DataSploit*, который автоматизирует сбор информации по технологиям *OSINT*, в частности, – скрипт *emailOsint.py* – для сбора информации об электронной почте [5]. На их основе может быть реализован процесс организации мониторинга деструктивного контента, циркулирующего в социальной сети: измерение его параметров – ареала распространения, периода активности; расчет вовлеченности пользователей в деструктивный контент; расчет риска заражения пользователей [6].

Инструменты *OSINT* играют все более важную роль в мониторинге и оценке рисков для предотвращения кибератак, разработке рекомендаций по созданию более надежной политики безопасности [7].

Мониторинг компьютерной сети имеет большое значение, помогая операторам понять ее текущее поведение. Поэтому точный и эффективный мониторинг жизненно важен при обеспечении работы сети в соответствии с предполагаемым поведением для устранения любых отклонений. Однако текущая практика сетевого мониторинга часто основана на ручных операциях, поэтому предприятия тратят значительную часть своих бюджетов на рабочую силу, чтобы контролировать свои сети.

Средство решения этой проблемы – инструментальный мониторинг сетей. Процесс мониторинга, его основные категории и классификации, а также сравнительная характеристика различных продуктов мониторинга компьютерной сети представлены в [8–10]. Однако, несмотря на бурное развитие этих средств, существует еще много проблем: (1) сбор данных с минимальными издержками для устройств сбора, (2) представление собранных данных в согласованных форматах, (3) эффективная передача собранных данных на управляющую станцию, (4) предоставление приложений более высокого уровня с разнообразными результатами анализа, (5) представление проанализированных данных операторам сети интуитивно понятным способом и др. [11].

Большое значение для информационной безопасности организации имеет мониторинг ее веб-сайта. В настоящее время существуют как ручной, так и автоматизированный мониторинг сайтов. Все множество систем инструментального мониторинга делится на типы, которые соответствуют функционалу, позволяющему осуществлять мониторинг определенных видов характеристик сайтов: работоспособность; посещаемость; SEO-оптимизация [12]. Стремительно развивается рынок инструментальных средств мониторинга работоспособности сайтов (*Ping-Admin.Ru*, *Host-tracker.com*, *Monitor.us*, *HyperSpin* и др. [13]. Активно используются на практике инструменты мониторинга посещаемости сайтов [14]. Поиск оптимизация веб-сайтов вызывает интерес в связи с усилением конкуренции на рынках товаров и услуг [15, 16].

В отечественной и зарубежной практике значительно меньше внимания уделяется инструментальному мониторингу соответствия веб-сайтов законодательству. Например, мониторинг контента сайтов

администраций проводился вручную студентами университета [17]. Однако в последние годы мониторинг соответствия – способ определения отклонения последовательности событий от ожидаемого поведения – все чаще становится объектом изучения. Так, установлено, что методы проверки соответствия могут показывать плохую производительность при работе с большими данными и потоками, что требует применения масштабируемых подходов. Поэтому зарубежные эксперты изучили возможность распространения мониторинга соответствия по сети вычислительных узлов для достижения желаемой масштабируемости и реализовали мониторинг проверки соответствия через существующую платформу *SCIFF*, которая обеспечивает язык логики высокого уровня для выражения контролируемых свойств и поддерживает разделение задач мониторинга. Распределенные вычисления достигаются с помощью подхода *MapReduce* и принятия существующего общего механизма для обработки потоков в больших масштабах [18].

В России разработана и утверждена Министерством экономического развития Российской Федерации Методика мониторинга официальных сайтов органов государственной власти и местного самоуправления [19] – автоматизированная информационная система «Мониторинг государственных сайтов», предназначенная для оценки открытости информации о деятельности органов государственной власти и доступности государственных информационных ресурсов для граждан в соответствии с требованиями действующего законодательства по обеспечению доступа к информации о деятельности государственных органов и органов местного самоуправления, а также для экспертного сообщества и конечных пользователей.

Механизм функционирования информационной системы для удаленного контроля качества образования и выявления нарушений требований законодательства в сфере образования представлен в [20], где описаны функциональные возможности системы и предложен ряд мероприятий по развитию этой системы.

Инструментальный мониторинг соответствия веб-сайтов законодательству в области персональных данных до сих пор не был предметом российских и зарубежных научных публикаций. Между тем, проведенный нами анализ российских официальных сайтов государственных, образовательных организаций и организаций здравоохранения показал неполное их соответствие требованиям законодательства в области персональных данных: отсутствуют обязательные сведения о персональных данных, неправомерно размещаются на сайте персональные данные и др. Это определяет актуальность разработки инструментального средства такого мониторинга для соблюдения требований и принятия правильных управленческих решений в области информационной безопасности. Для разработки такого средства большие возможности открывают использованные нами технологии машинного обучения (ML).

## РЕЗУЛЬТАТЫ РАБОТЫ

Для решения описанных проблем мы разработали веб-приложение для мониторинга сайтов на соответствие требованиям законодательства в области персональных данных. Обладая широкой совместимостью и низкими требованиями к ресурсам операционной сис-

темы, данный продукт сможет без проблем работать даже на ЭВМ с низкой вычислительной мощностью. Основные достоинства приложения – интуитивно понятный и современный интерфейс, легкость в освоении, а также большое количество опций для выполнения различных задач.

Веб-приложение написано на языке *Python* с использованием фреймворка *Flask*. Поскольку *Python* – это кроссплатформенный язык программирования, то он одинаково хорошо работает как на *Windows*, так и на *Unix* подобных системах. Сохранение данных реализовано на основе *MariaDB* – легковесной СУБД, что позволяет обеспечить высокую скорость взаимодействия веб-приложения и базы данных. Для анализа документов в сервис интегрирована текстовая модель машинного обучения *Word2Vec*, а для поиска имён и фамилий используется библиотека *py morphology2*. Инструментом сбора страниц сайта является утилита *wget*, которая позволяет рекурсивно обойти все страницы сайта и загрузить файлы *html*, *css*, *js*, а также все вложения на сайте. Обработчиком *html* кода на странице выступает *python* библиотека – *BeautifulSoup4*.

Минимальные требования к оборудованию для установки веб-приложения»:

- процессор *Pentium* с частотой 2ГГц или более быстрый;
- не менее 512 МБ оперативной памяти;
- не менее 2 ГБ свободного места на жестком диске;
- клавиатура, мышь *Microsoft Mouse* или совместимое указывающее устройство;
- видеокарта и монитор, поддерживающие режим *Super VGA* с разрешением не менее чем 1024x768 точек.

В процессе работы с программой пользователь должен:

- регистрироваться в веб-приложении. Это следует сделать единожды при первой работе с программой;
- входить в программу по своему уникальному логину/паролю;
- выбирать необходимый вид проверки (проверка на правильность использования персональных данных, проверка документов, проверка на наличие запрещенной информации);
- заполнять необходимые поля.

После обработки сайта пользователю выдается список найденных замечаний на экране или в файл, указанный в настройках задания на проверку.

В случае, если выбрана проверка персональных данных, то пользователю в результате работы программы будет выдана цитата из текста, где использовались персональные данные, точное расположение её на сайте с указанием ссылки на страницу, а также ссылки на фотографии с изображенными на них людьми для выявления возможных нарушений в размещении фотографий лиц, без их письменного согласия.

Результатом работы задания на поиск документов будет информация о наличии или отсутствии документов, регламентирующих действия с персональными данными, о наличии избыточных документов внутри архитектуры сайта, документов с грифом ДСП и других указанных в опциях задания документов.

Функция поиска запрещенной информации позволяет выявить размещение на сайте информации, содержащей негативный или запрещенный по закону контент (реклама, ссылки на сторонние ресурсы, призывы к протестной деятельности, оскорбление власти и т.д.) в комментариях, отзывах или на форуме.

Функционал данного сервиса не является конечным. В стадии разработки находятся такие сервисы как выявление:

- открытости или закрытости доступности к панели администратора сайта, внутренним папкам сайта;
- установки следящих программ на сайте, майнинга;
- ошибок в функционировании сайта, типовых ошибок в администрировании;
- вредоносного кода, ПО на сайте, вирусных атак и др.

Разработанный нами программный продукт был протестирован более чем на 50 сайтах государственных образовательных и медицинских учреждений. В процессе его работы постоянно улучшаются алгоритмы анализа и обработки данных сайта, совершенствуется внутренняя структура документов. Продукт позволяет обеспечить оперативность мониторинга соответствия контента сайтов законодательству в области персональных данных, визуализацию отчетов об этой информации для принятия управленческих решений; минимизировать ресурсное обеспечение мониторинга конфиденциального документооборота на предприятии.

## ЗАКЛЮЧЕНИЕ

Таким образом, по сравнению с ручным мониторингом, инструментальное средство, разработанное нами для целей мониторинга соответствия веб-сайтов требованиям законодательства в области персональных данных, имеет значительные преимущества: оперативность, многокритериальность поиска, техническая простота, удобство использования и др. Внедрение этого продукта в практику обеспечения безопасности персональных данных позволит существенно повысить эффективность этого процесса, что дает основание утверждать о необходимости его широкого использования в информационной деятельности.

## СПИСОК ЛИТЕРАТУРЫ

1. Ponder-Sutton Agate M. Chapter 1: The Automating of Open Source Intelligence // Automating Open Source Intelligence: Algorithms for OSINT / eds. R. Layton, P. A. Watters. 2016. – P. 1-20.
2. Open source intelligence in the twenty-first century: new approaches and opportunities (new security challenges) / eds. C. Hobbs, M. Moran, D. Salisbury. – London, UK: Palgrave Macmillan, 2014. – 191 p.
3. Sudhanshu Chauhan, Nutan Kumar Panda. Chapter 6: OSINT Tools and Techniques // Hacking Web Intelligence. – 2015. – P. 101-131.
4. Гольчевский Ю.В., Кузнецов Д.И. Автоматизация механизмов поиска информации на основе открытых источников в сети Интернет // Информация и безопасность. – 2017. – Т. 20, № 3. – С. 414-417.

5. Behrouz Mansoori. Сбор информации при помощи технологии OSINT в связке с DataSploit. – URL: <https://www.securitylab.ru/analytics/497348.php> (дата обращения: 10.09.2019).
6. Остапенко А.Г., Нежелский Е.Р., Степанов М.Н., Ружицкий Е., Заряев А.В. Организация мониторинга постов социальной сети вконтакте с помощью интерфейса VKAPI // Информация и безопасность. – 2018. – Т. 21, № 3. – С. 408-415.
7. Hayes Darren R., Carra Fr. Open-source intelligence for risk assessment // Business Horizons. – 2018. – Vol. 61, Issue 5 (September–October). – P. 689-697.
8. Нечволода В.Э., Смыкова В.Н., Рябцев С.С., Петренко П. Сравнительный анализ функциональных возможностей средств мониторинга компьютерной сети // Студенческая наука для развития информационного общества. Сб. материалов IX Всероссийской научно-технической конференции. – Ставрополь: Северо-Кавказский федеральный ун-т, 2019. – С. 69-78.
9. Hernantes J., Gallardo G., Serrano N. IT Infrastructure-Monitoring Tools // IEEE Software. – 2015. – Vol. 32(4). – P. 88-93. DOI: 10.1109/ms.2015.96
10. Marik O., Zitta S. Comparative analysis of monitoring system for data networks // 2014 International Conference on Multimedia Computing and Systems – ICMCS (Marrakech, 14-16 April 2014) // IEEE. – 2014. – P. 563-568. DOI: 10.1109/icmcs.2014.6911307.
11. Sihyung Lee, Kyriaki Levanti, Hyong S. Kim. Network monitoring: Present and future // Computer Networks. – 2014. – Vol. 65(2 June 2014). – P. 84-98.
12. Сирович Дж., Дари К. Поисковая оптимизация сайта (SEO) на PHP для профессионалов. Руководство разработчика по SEO = Professional Search Engine Optimization with PHP: A Developer's Guide to SEO. – М.: «Диалектика», 2008. – 352 с.
13. Романов М.В. Разработка автоматизированной системы мониторинга интернет-ресурсов // Молодой ученый. – 2014. – №12. – С. 53-61. – URL: <https://moluch.ru/archive/71/12210/> (дата обращения: 10.09.2019).
14. Вараксин М.А. Инструменты для мониторинга посещаемости сайтов // Информационные системы и технологии в образовании, науке и бизнесе (ИСИТ-2014). Материалы Всероссийской молодежной научно-практической школы. – Кемерово: Кузбасский гос. тех. ун-т имени Т.Ф. Горбачева, 2014. – С. 161-162.
15. Кривенцова П.С. Автоматизация процесса мониторинга поисковой оптимизации сайтов // Сб. трудов Международной конференции «Инновационные подходы к решению технико-экономических проблем». – М.: Национальный исследовательский ун-т "Московский институт электронной техники", 2016. – С. 153-157.
16. Кривенцова П.С. Реализация подсистемы мониторинга оптимизации сайтов // Сб. трудов Международной конференции «Инновационные подходы к решению технико-экономических проблем». – М.: Национальный исследовательский университет "Московский институт электронной техники", 2015. – С. 9-13.
17. Гасумова С.Е. Мониторинг соответствия интернет-ресурсов органов местного самоуправления законодательству РФ (на примере Пермского края) // Власть. – 2012. – № 1. – С. 55-60.
18. Loreti D., Chesani F., Ciampolini A., Mello P. A distributed approach to compliance monitoring of business process event streams // Future Generation Computer Systems. – 2018. – Vol. 82(May 2018). – P. 104-118.
19. Методика мониторинга официальных сайтов органов государственной власти и местного самоуправления. – М., 2014. – 151 с. – URL: [https://гтвм.мвд.рф/upload/site1/document\\_file/Uirsh4AwKN.pdf](https://гтвм.мвд.рф/upload/site1/document_file/Uirsh4AwKN.pdf) (дата обращения: 10.09.2019).
20. Кондакова Н.М., Тугой И.А. Информационная система мониторинга сайтов образовательных организаций: механизм, теги, функциональность // Профессиональное образование. Столица. – 2017. – № 3. – С. 32-35.

*Материал поступил в редакцию 10.09.19.*

#### **Сведения об авторах**

**АСТАХОВА Людмила Викторовна** – доктор педагогических наук, профессор, профессор кафедры защиты информации Южно-Уральского государственного университета (национального исследовательского университета), г. Челябинск  
e-mail: [astakhovalv@susu.ru](mailto:astakhovalv@susu.ru)

**МЕДВЕДЕВ Иван Алексеевич** – студент кафедры защиты информации Южно-Уральского государственного университета (национального исследовательского университета)  
e-mail: [ivanelgran@mail.ru](mailto:ivanelgran@mail.ru)