

УДК 004.021

АЛГОРИТМ АНАЛИЗА ИНТЕРНЕТ-СТРАНИЦ ИНФОРМАЦИОННОГО РЕСУРСА

Быстрицкий Н.Д.

*ФГАОУВО «Московский государственный университет имени М.В. Ломоносова», Москва,
e-mail: fastnika@yandex.ru*

В соответствии с требованиями нормативно-методических документов РФ и международных стандартов ISO 8879, W3C и ECMA 262 к базовым государственным информационным ресурсам в данной работе автором предложен алгоритм анализа исходного кода интернет-страниц, основанный на составляющих информационных технологий веб-ресурса (HTML, CSS и JavaScript), который позволит повысить достоверность определения недопустимых функционально-некорректных ошибок. Главный цикл процесса анализа состоит из трех ключевых этапов: загрузка, анализ содержимого и анализ найденных интернет-ссылок для продолжения исследований. Такое ключевое разделение с применением параллельных технологий на пользовательском уровне позволяет более эффективно использовать возможности не только процессорной системы, но и существующего канала связи, что дает перспективы для проведения более быстрого анализа больших информационных ресурсов за обозримое время.

Ключевые слова: интернет-безопасность, аспект, информация, сетевое пространство, код

ANALYZING ALGORITHM OF INFORMATION RESOURCE WEB PAGES

Bystritsky N.D.

*Federal Autonomous Educational Institution of Higher Education Moscow state University
name after Lomonosov, Moscow, e-mail: fastnika@yandex.ru*

This article is dedicated to the study of the influence of esters and acids content on anti-friction properties of n accordance with the requirements of normative-methodical documents of the Russian Federation and international standards ISO 8879, W3C, and ECMA 262 to the basic state information resources in this work the author of an algorithm for the analysis of the source code of web pages, mainly in the information technology components of Web resource (HTML, CSS, and JavaScript), which will improve the accuracy of determining the harmful functionally incorrect errors. The main loop of the analysis process consists of three key steps: loading, analysis and content analysis found Internet links for further research. This is a key division with parallel technologies at the user level allows for more efficient use of the processor, not only the system but also the existing communication channel that provides opportunities for a more rapid analysis of large information resources for the foreseeable future.

Keywords: Internet security, aspect, the information, network space, code

Реализовав в 1990 году первый в мире веб-браузер WorldWideWeb [1], Tim Berners-Lee заложил основной принцип организации гипертекстовых документов посредством коммуникационного взаимодействия пользователя с информационной системой через сеть Интернет. Такие достоинства, как структуризация информации, простота и привычность интерфейса, возможность удаленной работы и быстрота разработки веб-приложения, позволили веб-обозревателю стать одним из обязательных самостоятельных приложений в составе большинства операционных систем, а интернет-ресурсам – одним из стратегически важных и динамически развивающихся видов информационных ресурсов. Современный информационный ресурс сегодня представляет собой не просто статичный набор веб-страниц, а многофункциональный портал с использованием различных средств и технологий, в том числе и применением различных шаблонов для разных уровней вложенности.

Вместе с этим неотъемлемой частью полноценного безопасного функциониру-

ния информационного ресурса является его корректность. Это выражено не только тем, что «браузерный движок» веб-обозревателя для преобразования содержимого веб-страниц в интерактивное отображение использует собственную трактовку интернет-спецификаций, но и возможность присутствия на информационном ресурсе функционально-некорректных конструкций. Таким образом, простая проверка того, что интернет-ресурс приемлемо функционирует в нескольких веб-браузерах, только показывает удовлетворяемое качество разработанного кода на текущий момент времени и не дает абсолютно никакой гарантии его правильного отображения в следующий раз. Существование такой проблемы также подтверждают внесенные в 2014 году при разработке спецификации HTML5 консорциумом W3C предложения по анализу структуры HTML-документа [9, п.п. 8.2.8].

Отчасти, это связано как с постепенно возрастающей сложностью веб-приложения, которая не позволяет разработчикам информационного ресурса своевременно контролировать качество

написанного кода, так и отсутствием разработанных методов и программных инструментов для проведения такого комплексного анализа. Разработанный консорциумом W3C Validator Suite [12] позволяет проводить анализ корректности исходного кода информационного ресурса только по стандартам HTML и CSS [9, 10], что не позволяет полностью охватить все используемые технологии информационного ресурса. Более того, данное программное средство производит слишком строгий анализ на соответствие исследуемым стандартам. Другие веб-анализаторы, такие как Rational AppScan (IBM), Web Vulnerability Scanner (Acunetix), NTOSpider (NT Objectives, Inc.), NetSparker (**NetSparker Ltd.**), WebInspect (HP), Application Inspector (PT), SkipFish (Google) и др. направлены, в первую очередь, на получение оценки защищенности информационного ресурса.

Тем самым на текущий момент не предоставляется возможным своевременно отлавливать критические ошибки информационного ресурса. Это значит, что невозможно получить его достоверную оценку корректности функционирования и провести комплекс мероприятий, направленных на повышение общей безопасности интернет-ресурса. Корректность функционирования таких источников информации необходима и важна не только государственным, но и коммерческим организациям, чья деятельность подразумевает гарантированное предоставление услуг и информации в сети Интернет. Именно такие задачи при эксплуатации государственных информационных ресурсов общего пользования определены законодательными и нормативно-методическими документами Российской Федерации [3–6]. Это обуславливает необходимость проведения периодического мониторинга качества кода информационного ресурса при его создании (разработке) и эксплуатации. Поэтому задача исследования состоит как в разработке эффективной проверки функционирования исходного кода интернет-страниц информационного ресурса, так и в легкости применения разработанных средств. Разрабатываемый алгоритм требуемого программного обеспечения должен:

- иметь гибкую структуру для его модификации и расширения возможностей;
- использовать параллельные технологии для проведения эффективного анализа;
- проводить анализ доступности всех компонентов интернет-ресурса;
- учитывать основные составляющие технологии информационного ресурса;

– исследовать не только функционально-корректные конструкции, но и допустимость задания в них значений атрибутов, селекторов и т.д.;

– учитывать современные тенденции к формированию исходного кода интернет-страниц информационного ресурса с учетом использования системы управления содержанием (CMS).

В соответствии со сформированными условиями, алгоритм логично разделить на две составляющие:

– алгоритм анализа всех составляющих частей информационного ресурса (основной алгоритм проведения анализа);

– алгоритм анализа веб-страниц информационного ресурса (алгоритм работы веб-анализатора).

Главный цикл процесса анализа состоит из трех ключевых этапов: загрузка, анализ содержимого и анализ найденных интернет-ссылок для продолжения исследований [9]. Такое ключевое разделение с применением параллельных технологий на пользовательском уровне позволяет более эффективно использовать возможности не только процессорной системы, но и существующего канала связи, что дает перспективы для проведения более быстрого анализа больших информационных ресурсов за обозримое время.

Содержимое интернет-страницы в общем случае представляет собой совокупность нескольких веб-технологий. Основной технологией такого документа является HTML. Выпущенный в 2014 году стандарт HTML5, как приложение стандартного обобщенного языка разметки SGML, содержит только принципы для построения синтаксической схемы [9, п.п. 8.1]. Основными вспомогательными технологиями HTML-документа являются CSS и JavaScript. Стандарт HTML5 устанавливает взаимно-однозначное использование между данными технологиями.

Язык стиля CSS определен принятой в 2011 году спецификацией CSS2.1, в которой содержатся правила построения селекторов и свойств [4, п. 4, п. 10, Приложение G]. Разрабатываемая спецификация CSS3 и CSS4 существенно расширяет функциональность текущего стандарта, однако до сих пор (с 29 сентября 2011 года) не является рекомендованной консорциумом W3C.

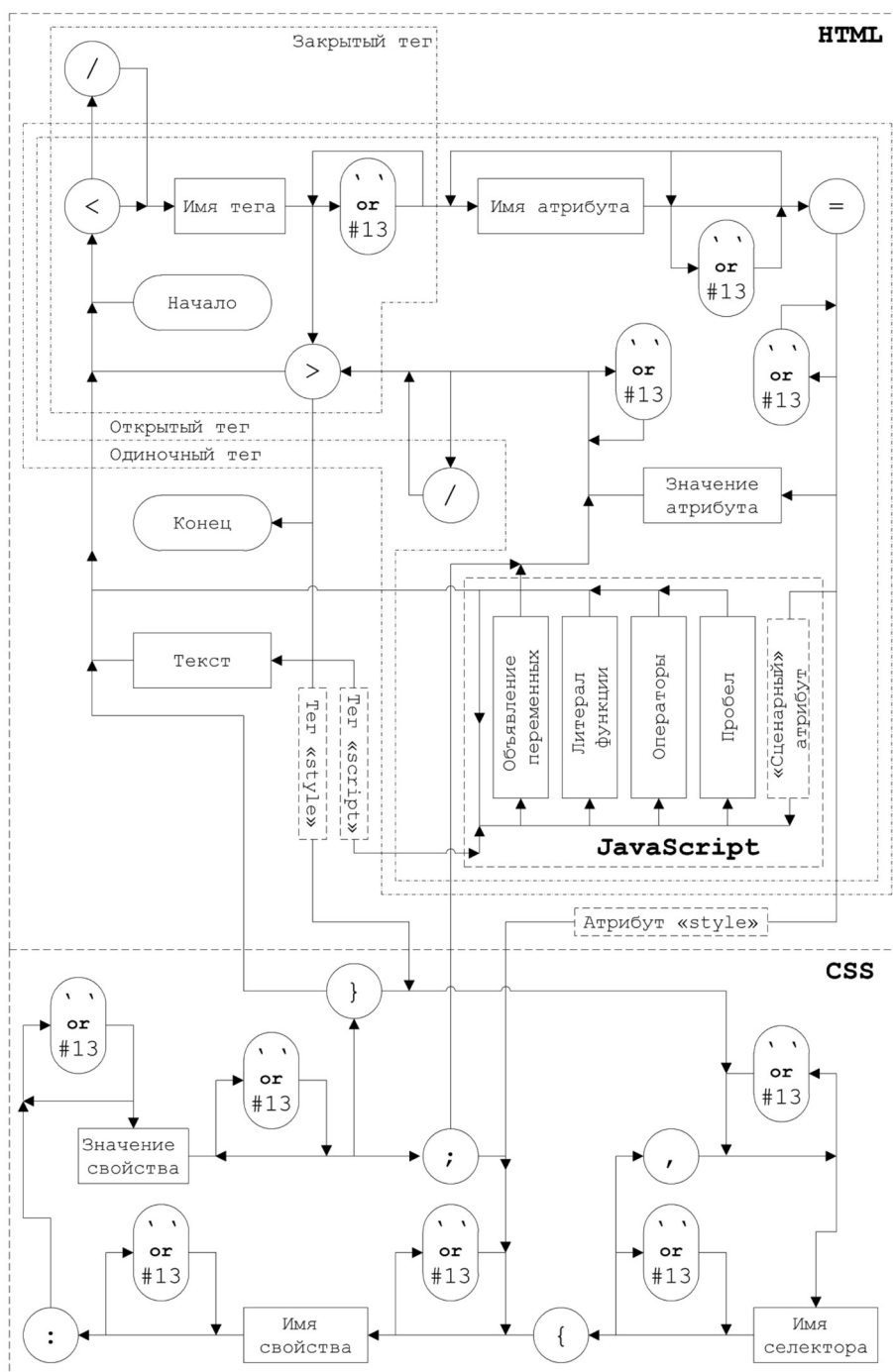
Язык JavaScript является расширением языка ECMAScript 5.1 [10]. Учитывая тот факт, что каждый из веб-обозревателей является независимой разработкой, данный стандарт рассматривается как основа для построения скриптовых языков. Таким образом, основными составляющими языка JavaScript являются [11]:

- «клиентский» JavaScript (интерпретатор JavaScript, встраиваемый в веб-браузер);
- «базовый» JavaScript (язык JavaScript, определяемый спецификациями);
- Document Object Model (спецификация W3C DOM [12]).

И хотя спецификация ECMAScript 5.1 содержит синтаксические схемы [13, Приложение А], проведенное Крокфордом Д. исследование позволило сформировать исполь-

зуемые веб-обозревателями синтаксические схемы языка JavaScript [2], где основными функционально значимыми конструкциями являются «Объявление переменных», «Литерал функции», «Операторы» и «Пробел».

В результате проведенных выше заданных требований синтаксическая схема для проведения анализа интернет-страницы может быть сформирована следующим образом (рисунок).



Синтаксическая схема разбора интернет-страницы информационного ресурса

Таким образом, разработанная схема позволяет проводить анализ интернет-страниц информационного ресурса, выявляя функционально-некорректные структуры, нарушающие его полноценное безопасное функционирование.

Список литературы

1. Быстрицкий Н.Д., Макаров-Землянский Н.В. Функционирование анализатора web-ресурсов // Естественные и технические науки. – 2013. – № 6(68). – С. 295–296. ISSN 1684-2626.
2. Крокфорд Д. JavaScript: Сильные стороны. – СПб: Питер, 2012. – 176 с. :ил. ISBN 978-5-459-01263-7.
3. Постановление Правительства РФ от 14.09.2012 N 928 (ред. от 21.07.2014) «О базовых государственных информационных ресурсах». «Требования к порядку формирования, актуализации и использования базовых государственных информационных ресурсов». «Правила формирования, актуализации и использования реестра базовых государственных информационных ресурсов».
4. Постановление Правительства РФ от 24.10.2011 N 861 «О федеральных государственных информационных системах, обеспечивающих предоставление в электронной форме государственных и муниципальных услуг (осуществление функций)».
5. Приказ ФСБ РФ, ФСТЭК от 31.08.2010 N 416/489 г. Москва «Об утверждении Требований о защите информации, содержащейся в информационных системах общего пользования».
6. Приказ Минэкономразвития России N 470 от 16.11.2009 г. «О Требованиях к технологическим, программным и лингвистическим средствам обеспечения пользования официальными сайтами федеральных органов исполнительной власти».
7. Флэнган Д. JavaScript. Подробное руководство. – СПб: Символ-Плюс, 2013. – 1080 с.:ил. ISBN 978-5-93286-215-5, 978-0-596-80552-4.
8. Document Object Model (DOM) Technical Reports [Электронный ресурс] // Консорциум W3C [Официальный сайт]. Режим доступа: <http://www.w3.org/DOM/DOMTR>, свободный.
9. Standard W3C: HTML5 A vocabulary and associated APIs for HTML and XHTML. W3C Recommendation, 28 October 2014 [Электронный ресурс] // Консорциум W3C [Официальный сайт]. Режим доступа: <http://www.w3.org/TR/html5/>, свободный.
10. Standard W3C: Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification. Recommendation, 07 June 2011 [Электронный ресурс] // Консорциум W3C [Официальный сайт]. Режим доступа: <http://www.w3.org/TR/CSS21/>, свободный.
11. Standard ECMA-262: ECMAScript Language Specification 5.1 edition (June 2011) [Электронный ресурс] // ECMA International [Официальный сайт]. Режим доступа: <http://www.ecma-international.org/ecma-262/5.1/ECMA-262.pdf>, свободный.
12. Tim Berners-Lee: WorldWideWeb, the first Web client [Электронный источник] // Консорциум W3C [Официальный сайт]. Режим доступа: <http://www.w3.org/People/Berners-Lee/WorldWideWeb.html>, свободный.
13. W3C Validator Suite [Официальный сайт]. Режим доступа: <https://validator-suite.w3.org/>, свободный.

References

1. Bystrickij N.D., Makarov-Zemljanskij N.V. Funkcionirovanie analizatora web-resursov // Estestvennye i tehniicheskie nauki. 2013. no. 6(68). pp. 295–296. ISSN 1684-2626.
2. Krokford D. JavaScript: Silnye storony. SPb: Piter, 2012. 176 p. :il. ISBN 978-5-459-01263-7.
3. Postanovlenie Pravitelstva RF ot 14.09.2012 N 928 (red. ot 21.07.2014) «O bazovyh gosudarstvennyh informacionnyh resursah». «Trebovanija k porjadku formirovanija, aktualizacii i ispolzovanija bazovyh gosudarstvennyh informacionnyh resursov». «Pravila formirovanija, aktualizacii i ispolzovanija reestra bazovyh gosudarstvennyh informacionnyh resursov».
4. Postanovlenie Pravitelstva RF ot 24.10.2011 N 861 «O federalnyh gosudarstvennyh informacionnyh sistemah, obespechivajushih predostavlenie v jelektronnoj forme gosudarstvennyh i municipalnyh uslug (osushhestvlenie funkcij)».
5. Prikaz FSB RF, FSTJeK ot 31.08.2010 N 416/489 g. Moskva «Ob utverzhenii Trebovanij o zashhite informacii, sodержashhejsja v informacionnyh sistemah obshhego polzovanija».
6. Prikaz Minjekonomrazvitija Rossii N 470 ot 16.11.2009 g. «O Trebovanijah k tehnologicheskim, programmnyh i lingvisticheskim sredstvam obespechenija polzovanija oficialnymi sajтами federalnyh organov ispolnitelnoj vlasti».
7. Fljenagan D. JavaScript. Podrobnoe rukovodstvo. SPb: Simvol-Pljus, 2013. 1080 p.:il. ISBN 978-5-93286-215-5, 978-0-596-80552-4.
8. Document Object Model (DOM) Technical Reports [Jelektronnyj resurs] // Konsorcium W3C [Ofic. sajt]. Rezhim dostupa: <http://www.w3.org/DOM/DOMTR>, svobodnyj.
9. Standard W3C: HTML5 A vocabulary and associated APIs for HTML and XHTML. W3C Recommendation, 28 October 2014 [Jelektronnyj resurs] // Konsorcium W3C [Ofic. sajt]. Rezhim dostupa: <http://www.w3.org/TR/html5/>, svobodnyj.
10. Standard W3C: Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification. Recommendation, 07 June 2011 [Jelektronnyj resurs] // Konsorcium W3C [Ofic. sajt]. Rezhim dostupa: <http://www.w3.org/TR/CSS21/>, svobodnyj.
11. Standard ECMA-262: ECMAScript Language Specification 5.1 edition (June 2011) [Jelektronnyj resurs] // ECMA International [Ofic. sajt]. Rezhim dostupa: <http://www.ecma-international.org/ecma-262/5.1/ECMA-262.pdf>, svobodnyj.
12. Tim Berners-Lee: WorldWideWeb, the first Web client [Jelektronnyj istochnik] // Konsorcium W3C [Ofic. sajt]. Rezhim dostupa: <http://www.w3.org/People/Berners-Lee/WorldWideWeb.html>, svobodnyj.
13. W3C Validator Suite [Ofic. sajt]. Rezhim dostupa: <https://validator-suite.w3.org/>, svobodnyj.

Рецензенты:

Станкевич Г.А., д.э.н., профессор, ведущий научный сотрудник South IT-Uni, ИВС, г. Москва;

Бихтер И.И., д.т.н., профессор, старший научный сотрудник сектора инновационных разработок АНОО ДПФО «Национальный центр прикладных исследований», г. Саратов.