

УДК 004.774

Д. Л. Жусов, канд. техн. наук, сотрудник, e-mail: d.zhusov@mail.ru,
А. В. Козленко, канд. техн. наук, сотрудник, Р. В. Шульга, сотрудник,
Академия ФСО России, г. Орел

Алгоритм анализа глубины гипертекстовых переходов официальных сайтов федеральных органов исполнительной власти Российской Федерации

Предложен алгоритм анализа глубины гипертекстовых переходов официальных сайтов федеральных органов исполнительной власти России, позволяющий повысить оперативность контроля их навигационных средств.

Ключевые слова: web-сайт, гипертекстовый переход, глубина гипертекстовых переходов

Введение

В настоящее время необходимость совершенствования механизмов государственного управления и бурный рост числа пользователей сети Интернет обусловили широкое применение современных информационных технологий для организации взаимодействия государства и общества.

В целях обеспечения доступа к информации о деятельности федеральных органов исполнительной власти (ФОИВ) РФ наиболее широкое применение получили официальные web-сайты. Важность

и необходимость их качественного функционирования подчеркивается разработкой нормативных документов [1, 2], регламентирующих их работу. В целях обеспечения быстрого доступа к информации на официальном сайте ФОИВ РФ его навигационные средства должны соответствовать ряду требований, одним из которых является ограничение на максимальное число последовательных гипертекстовых переходов, начиная с главной страницы сайта. При этом число таких переходов по кратчайшей последовательности должно быть не более пяти [2].

Данный факт определяет необходимость контроля соответствия официальных сайтов ФОИВ РФ требованиям руководящих документов.

Анализ соответствия официальных web-сайтов ФОИВ руководящим документам

С января по февраль 2013 г. Фонд Свободы Информации провел ежегодный мониторинг официальных сайтов ФОИВ РФ [3]. Его результаты (рис. 1) свидетельствуют об отсутствии единого подхода к формированию структуры официальных web-сайтов ФОИВ РФ, а также о не выполнении требований к ним в полном объеме. Так, глубина вложения информации на сайтах не всегда удовлетворяет требуемому ограничению, а существующие способы их проверки достаточно трудоемки.

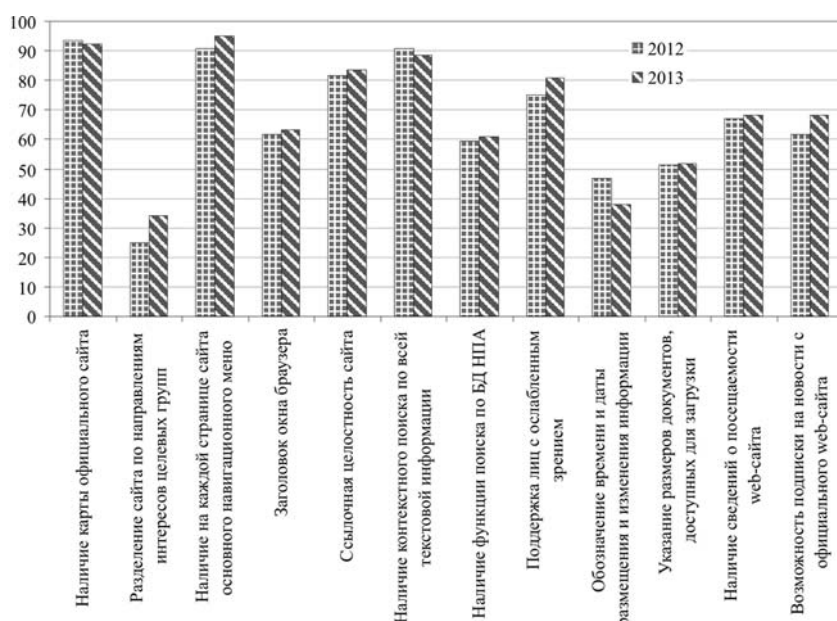


Рис. 1. Статистика соответствия официальных сайтов ФОИВ РФ требованиям руководящих документов

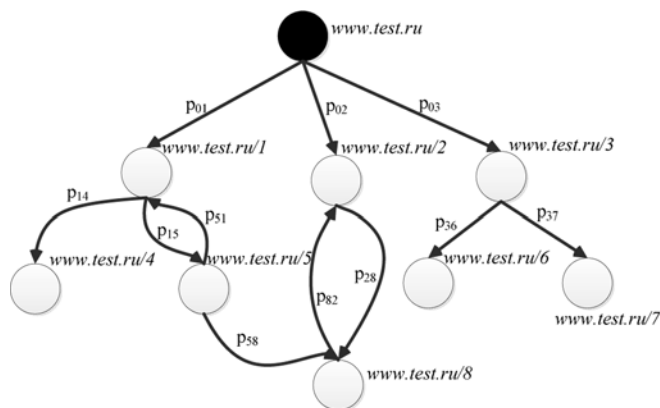


Рис. 2. Графовое представление web-сайта

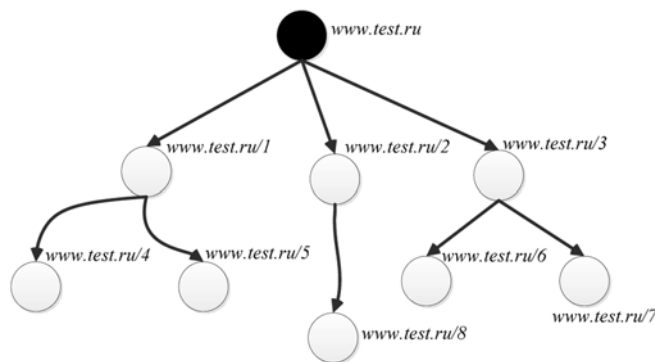


Рис. 3. Дерево web-сайта, построенное методом поиска "в ширину"

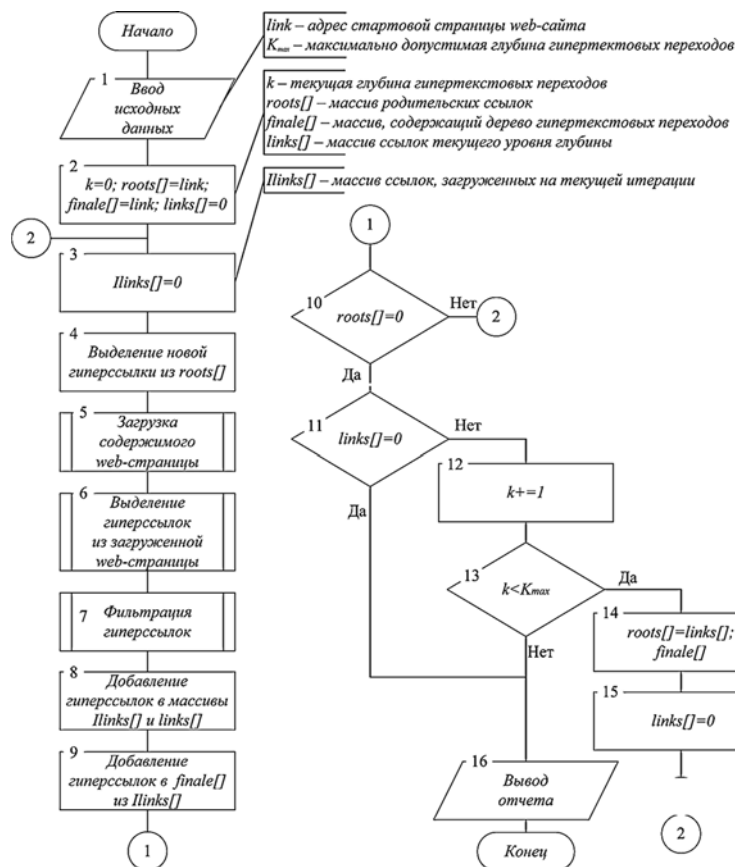


Рис. 4. Алгоритм анализа глубины гипертекстовых переходов официальных сайтов ФОИВ РФ

Следовательно, представляется необходимой разработка алгоритма анализа глубины гипертекстовых переходов web-сайтов.

Структурное представление web-сайтов

Структура каждого web-сайта может быть представлена деревом $G = (V, E)$, где V — множество вершин дерева (web-страниц), а E — множество его ребер (гипертекстовых переходов по сайту — гиперссылок) [4], а для расчета глубины гипертекстовых переходов целесообразно применять метод обхода вершин графа "поиск в ширину" [5].

Пусть имеется web-сайт, представленный в виде графа (рис. 2).

Последовательными переходами p_{01}, p_{02}, p_{03} по гиперссылкам со стартовой страницы осуществляется загрузка страниц web-сайта первого уровня: $www.test.ru/1, www.test.ru/2, www.test.ru/3$. Затем последовательно со страницы $www.test.ru/1$ переходами p_{14} и p_{15} по гиперссылкам загружаются страницы $www.test.ru/4, www.test.ru/5$; со страницы $www.test.ru/2$ переходом p_{28} загружается $www.test.ru/8$; а со страницы $www.test.ru/3$ переходами p_{36} и p_{37} соответственно $www.test.ru/6$ и $www.test.ru/7$. Загруженные страницы составляют второй уровень дерева web-сайта. Отсутствие гипертекстовых переходов со страниц второго уровня определяет остановку обхода вершин и фиксирует дерево web-сайта в виде рис. 3.

Рассмотренный метод обхода вершин дерева позволил разработать алгоритм анализа глубины гипертекстовых переходов web-сайтов (рис. 4). Стоит отметить, что объектами анализа могут являться любые web-сайты, по отношению к которым целесообразна формулировка требований по глубине гипертекстовых переходов. Возможным множеством сайтов являются официальные web-сайты ФОИВ РФ.

Основу разработанного алгоритма составляет набор рекурсивных процедур, представленный блоками 3—15. Его сложность определяется поиском в коде загруженной web-страницы тегов $\langle a \rangle$ с атрибутом *href* и числом страниц анализируемого web-сайта и вычисляется как $O(\eta) \cdot O(V + E)$, где η — число символов в коде web-страниц.

Рассмотренный алгоритм программно реализован на платформе JavaSE в виде отдельного модуля, взаимодействующего через JDBC-интерфейс с базой данных хранения результатов, функционирующей под управлением СУБД MySQL. С его помощью было проведено экспериментальное определение максимальной глубины гиперссылок как для локально расположенных сайтов, так и для размещенных в сети Интернет. Ввиду отсутствия доступа к локальной сети,

в которой размещены web-серверы с сайтами ФОИВ, для проведения фрагмента эксперимента web-сайты были предварительно полностью скачаны с помощью программного продукта Teleport Ultra 1.70. Для этих реализаций web-сайтов, размещенных на экспериментальной ПЭВМ с конфигурацией — AMD Athlon FX4000 8 ГБ ОЗУ — были получены значения времени анализа глубины гиперссылок (табл. 1).

Для реализации эксперимента на web-сайтах, размещенных в сети Интернет, использовалось подключение со скоростью 40 Мбит/с. Однако в процессе эксперимента скорость доступа к анализируемым web-страницам (их загрузки в программе) колебалась в пределах 0,1...1 Мбайт/с (табл. 2). Наиболее очевидными причинами этого можно считать предопределенная настройками web-серверов скорость обработки запросов и максимальное возможное число одновременных соединений с одного IP-адреса. Кроме того, при проведении эксперимента невозможно было ограничить нагрузку на web-серверы со стороны других пользователей сети Интернет.

Результаты свидетельствуют о том, что на время анализа глубины гипертекстовых переходов кроме объема web-сайта (число web-страниц) влияют скорость Интернет-подключения и настройки web-сервера, на котором размещен web-сайт ФОИВ (табл. 1, 2, рис. 5).

Обобщая данные таблиц 1, 2, можно представить график зависимости времени анализа от числа страниц web-сайта (см. рис. 5).

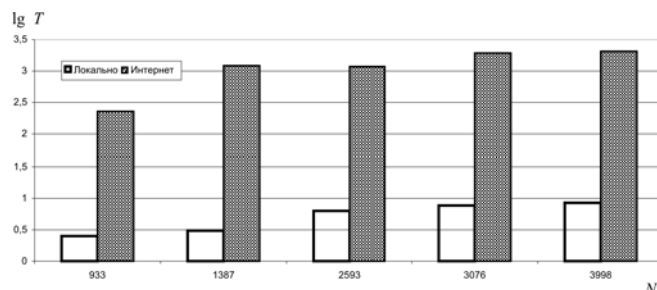


Рис. 5. Результаты экспериментальных исследований

Из рисунка видно, что время анализа глубины гиперссылок разработанным программным средством существенно меньше времени доступа к сети Интернет. Это позволяет говорить о возможности повышения оперативности контроля навигационных средств web-сайтов с помощью программной реализации представленного алгоритма при наличии высокоскоростного подключения и функционировании web-серверов в условиях неопределяемой загрузки.

Заключение

Программная реализация представленного в работе алгоритма анализа гипертекстовых переходов web-сайтов может эффективно использоваться для контроля требований нормативных документов. При этом объектами анализа могут являться не только официальные web-сайты ФОИВ РФ, но и сайты других государственных органов, организаций и учреждений.

Практическая значимость работы определяется возможностью применения представленного алгоритма для решения задачи контроля функционирования web-сайтов различного назначения.

Направлением дальнейших исследований является интеграция программной реализации алгоритма с модулем сбора и обработки данных о посещаемости web-сайтов с различных серверов сбора статистики, например, Liveinternet, Mail, Rambler и т. п.

Список литературы

1. **Федеральный закон** № 8-ФЗ "Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления" от 09 февраля 2009 г.
2. **Приказ** Министерства экономического развития РФ от 16 ноября 2009 г. № 470 "О требованиях к технологическим, программным и лингвистическим средствам обеспечения пользования официальными сайтами федеральных органов исполнительной власти".
3. **Результаты мониторинга официальных сайтов федеральных органов исполнительной власти**. 2013. URL: <http://svoboda.info/ru/node/2527>
4. **Райгородский А. М.** Модели случайных графов. М.: МЦНМО, 2011. 136 с.
5. **Кормен Т. Х., Лейзерсон Ч. И., Ривест Р. П., Штайн К.** Алгоритмы: построение и анализ. 2-е изд. М.: Вильямс, 2006. 1296 с.

Таблица 1

Время анализа глубины гиперссылок локально расположенных web-сайтов

web-сайт	Число web-страниц, N	Глубина гиперссылок, K	Время анализа, T, с	lg T
minsvyaz.ru	933	3	2,5	0,398
minfin.ru	2593	4	6,3	0,799
mintrans.ru	3998	6	8,4	0,924
rosminzdrav.ru	1387	3	3	0,477
minregion.ru	3076	5	7,6	0,88

Таблица 2

Время анализа глубины гиперссылок web-сайтов, расположенных в сети Интернет

web-сайт	Число web-страниц, N	Глубина гиперссылок, K	Время анализа, T, мин	lg T	Скорость доступа к web-сайту, Мбайт/с
minsvyaz.ru	933	3	3,87	2,366	1
minfin.ru	2593	4	19,5	3,068	1
mintrans.ru	3998	6	34,4	3,315	0,5
rosminzdrav.ru	1387	3	20,1	3,081	0,1
minregion.ru	3076	5	32,55	3,291	0,8

Hypertext Navigation Depth Analysis Algorithm for Federal Executive Authorities Official Websites in the Russian Federation

In the article the results of researches on scope of Russian Federation federal executive authorities official websites monitoring is considered. The purpose of the article was an increasing the hypertext navigation depth analysis efficiency for the websites. Achievement of the formulated purpose is realized by the solution of the hypertext navigation depth analysis algorithm development task, which considering features of websites creation. The method of graph peaks bypass "in width" and set of the recursive procedures on hyperlinks separation on each web page of the site is its cornerstone. The computing complexity assessment for developed algorithm and platform option for its program implementation are offered. The experimental results contains websites time analysis estimates depending on number of the web pages included in them and the maximum hypertext navigation depth on the shortest way. They can be an evidence of a possibility for websites monitoring efficiency increasing taking into account requirements of the guidelines. It is expedient to apply it to increase of developed algorithm application efficiency at the moments of Web-server loading limit absence in the presence of high speed connection to the Internet.

The further researches direction on number of websites visits data processing on the basis of these generally available servers of statistics is formulated.

Keywords: website, hypertext navigation, hypertext navigation depth

References

1. **Federal'nyj** zakon № 8-FZ "Ob obespechenii dostupa k informacii o dejatel'nosti gosudarstvennyh organov i organov mestnogo samoupravlenija" ot 09 fevralja 2009 g.
2. **Prikaz** Ministerstva jekonomicheskogo razvitija RF ot 16 nojabrja 2009 g. N. 470 "O trebovanijah k tehnologicheskim, programmym i lingvisticheskim sredstvam obespechenija pol'zovanija oficial'nymi sajтами federai'nyh organov ispolnitel'noj vlasti".

3. **Rezul'taty** monitoringa oficial'nyh sajtov federal'nyh organov ispolnitel'noj vlasti. 2013. URL: <http://svobodainfo.org/ru/node/2527>
4. **Rajgorodskij A. M.** Modeli sluchainyh grafov. M.: MCNMO, 2011. 136 p.
5. **Kormen T. H., Lejzerson C. I., Rivest R. P., Shtajn K.** Algoritmy: postroenie i analiz. 2-e izd. M.: Vil'jams, 2006. 1296 p.

Уважаемые авторы!

Сообщаем Вам, что в связи с подготовкой журнала "Информационные технологии" к включению в международные наукометрические базы данных редакция скорректировала требования к статьям и к сопровождающим их материалам, направляемым в адрес редакции. Просьба детально ознакомиться с новыми правилами и строго руководствоваться ими.