Sink Web Pages in Web Application

Doru Anastasiu Popescu

Faculty of Mathematics and Computer Science,
University of Pitesti, Romania
dopopan@gmail.com

Abstract. In this paper, we present the notion of sink web pages in a web application. These pages allow identifying a reduced scheme of the web application, which can lead to simplifying the method of testing and verifying the entire web application. We believe that this notion can be useful in the partially supervised learning.

Keywords: Relation, Tag, HTML, Web Application.

1 Introduction

The results of this paper are related to web applications that contain web pages consisting of HTML tags and are saved in files with .html or .htm extension. The web pages can contain other elements as well, such as scripts or applets; however, these will not be used next in the paper (for example: their testing and verifying implies using specific methods). Next, we will use for these pages the name of web pages. On one hand, the number of web applications which are built using this type of web pages is a very large one; on the other hand the web applications can contain a large number of web pages. In this context, the matter of verifying and testing the web application from the point of view of the content areas ([14], [15], [10]) or from the point of view of the navigability in the application ([7], [9]). A classification of the methods and models of testing and verifying is presented in [9].

We believe that reduced scheme for a web application (presented in section 2) can be used in other areas that utilize a large number of components, as it is the case of partially supervised learning ([12], [13]).

The results presented in the following sections are related to the selection of some web pages from the set of web pages of an application, which if tested and verified assure the testing and verification of the entire application. The process of detecting these web pages is realized through a relation among the web pages of the web application (described in section 2). Different ways of defining this relation have been presented in [1], [4], [5], [6].

The method of defining the pages that will be selected (named sink web pages) will be introduced in section 3. Using sink web page notion, a reduced scheme of the web application is created, which can be used in the process of testing and verifying the web application.

F. Schwenker and E. Trentin (Eds.): PSL 2011, LNAI 7081, pp. 154–158, 2012.

2 Defining a Relation between Two Web Pages

Next, we will consider a web application having the set of web pages P={p_1, p_2, ..., p_n} and a set TG of tags.

For any web page p_i from P, we write T_i the sequence of tags from p_i, which are not a member of TG (the order in which these are encountered is important).

Definition 1. Let TG be a set of tags, p_i and p_j two web pages from P. We say that T_i=(T_{i1},T_{i2}, ..., T_{ia}) is in T_j=(T_{j1},T_{j2}, ..., T_{jb}) as a sequence, if there exists an index k in the sequence T_j, with T_{jk}=T_{i1}, T_{jk+1}=T_{i2}, ..., T_{jk+a-1}=T_{ia}.

Definition 2. Let TG be a set of tags, p_i and p_j two web pages from P. We say that p_i is in relation R with p_j and we write p_i R p_j, if:
i) T_i is in T_j as a sequence;
ii) Any tag <Tg> from T_j which appears in T_i, as well, has a closing tag <\Tg> in T_j, then <\Tg> is also in T_i.

Exemple 1. Let us consider a web application with three web pages: P={p_1, p_2, p_3}. p_1 can be found in the file Pag1.html, p_2 in the file Pag2.html, and p_3 in the file Pag3.html (table 1). Considering:

Table 1. Pag1.html, Pag2.html and Pag3.html

Pag1.html	Pag2.html
<HTML> <HEAD>	<HTML> <HEAD>
<TITLE>Web page 1</TITLE>	<TITLE>Web page 2 </TITLE>
</HEAD> <BODY>	</HEAD> <BODY>
<P> Picture 1	<B> <P> Picture 1</P>
<IMG SRC="pic.jpg">	<IMG SRC="pic.jpg">
	<P> <IMG SRC="pic.jpeg">
</BODY> </HTML>	</BODY> </HTML>
Pag3.html	
<HTML> <HEAD>	</FONT> <IMG SRC="pic.jpg">
<TITLE>Web page 3</TITLE>	<FONT SIZE=4 COLOR=red>Picture 3
</HEAD> <BODY>	</FONT> <IMG SRC="pic.jpg">
<FONT COLOR=red>Picture 3	</BODY> </HTML>

TG={<P>,</P>,<B>,</B>,<HTML>,<HEAD>,<TITLE>,</TITLE>, </HEAD>,<BODY>,</BODY>,</HTML>} we obtain:
T1=(<IMG SRC="pic.jpg">); T2=(<IMG SRC="pic.jpg">, <IMG SRC="pic.jpeg">); T3=(<FONT COLOR=red>, </FONT>, <IMG SRC="pic.jpg">, <FONT SIZE=4 COLOR=red>, <IMG SRC="pic.jpg">, </FONT>).

According to definitions 1 and 2, we obtain only two pairs of pages with are in relation R:
p_1 R p_2 and p_1 R p_3.

Observations. 1. If TG=∅, then according to definition 2, two web pages p_i and p_j will be in relation R only if these pages contain exactly the same tags after removing the tags that are members of TG. In this case the relation R is an equivalence relation. A few results using this type of relation are presented in [4], [6], [7].

2. The tags from TG have to be written correctly, in order not to negatively influence the syntactic correctitude of the web pages.

3 The Concept of Sink Web Page

Using the notations in the previous section, we define the sink web pages for a web application with the set of web pages P={p_1, p_2, ..., p_n} and a set TG of tags, as below:

Definition. The web page p_i is called *sink web page*, if the following property is fulfilled:
it does not exist p_j in P, with i≠j and p_i R p_j.

Using the relation R, we can construct for the web application an oriented graph G=(X,U) as below:
X={1, 2, ..., n} is the set of nodes. For a web page p_i, its index i is associated to it, where 1≤i≤n.
U={(i, j)| p_i R p_j, i≠j, 1≤i,j≤n} is the set of edges.
Writing $d_+(i)=|\{(i,j) \mid (i,j) \in U\}|$, we obtain:

Proposition. If i, 1≤i≤n is a node in the oriented graph G, previously defined with $d_+(i)=0$, then p_i is a sink web page.

Example. For a web application with n=15 web pages and the relations between them given as in fig. 1 we obtain the following sink web pages: p_1, p_5, p_8, p_9, p_{12}, p_{15}. The following sink web pages are being obtained:

Reduced scheme of the web pages which are part of a web application can be obtained. The scheme consists of two levels:
- the level of the sink web pages;
- the level of the web pages which are subordinated to the sink web pages through the relation R, the representation of this subordination being realized by arrows.

For the previous example, the reduced scheme of the web pages is the following:

Observations - There can exist many reduced schemes, due to the fact that a web page can be in relation R to several sink web pages, the scheme showing only one relation from all of them.
- In the process of verifying and testing, there can only be used the sink web pages, because the tags from the other pages are included (in the same order and with the same properties) in these ones.
- The errors detected in the sink web pages must be repaired in these pages, as well as in the ones that are bound to these through the relation R, symbolized by arrows in the reduced diagram.

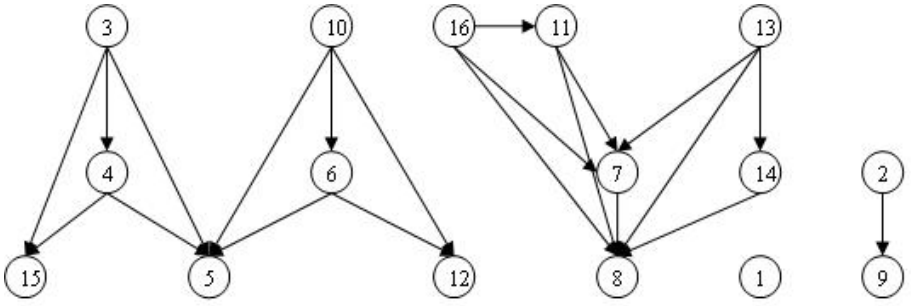

Fig. 1. The oriented graph G for a web application with 15 web pages and the relation R given by the edges

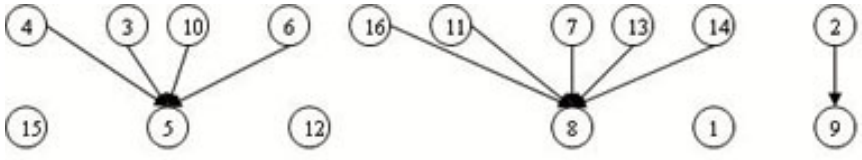

Fig. 2. A reduced scheme for the web application with 15 web pages and the relation R given in fig.1

4 Conclusion and Future Work

The concept of sink web page can be used in both processes of testing and verifying, as it was mentioned in the previous sections, but also in measuring the static complexity of the navigability ([11]) and of the content ([8], [10]) of a web application taking into consideration only these web pages. Another application of web pages sink is related to copyright, that is to simplify the comparison of two web applications, [3]. We intend to realize in the future a statistical study on categories of web applications regarding the number of sink web pages and their impact on testing, verifying, measuring, when using a complex application. In the same time we want to expand the area of applicability of the notions presented in the previous sections. The general paper from the area of partially supervised learning, like [12] and [13] show us that there is potential in this direction.

References

1. Danauta, C.M., Popescu, D.A.: Method of reduction of the web pages to be verified when validating a web site. Buletin Stiintific, Universitatea din Pitesti, Seria Matematica si Informatica (15), 19–24 (2009)
2. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to algorithms, 2nd edn. MIT Press (1990)

3. Popescu, D.A., Danauta, C.M.: Similarity measurement of web sites using sink web pages. In: 35th International Conference on Telecommunications and Signal Processing, pp. 24–24. IEEE Xplore (2011)
4. Popescu, D.A., Danauta, C.M., Szabo, Z.: A method of measuring the complexity of a web application from the point of view of cloning. In: Proceedings of the 5th International Conference on Virtual Learning, Section Models and Methodologies, October 29-October 31, pp. 186–181 (2010)
5. Popescu, D.A., Szabo, Z.: Sink web pages of web application. In: Proceedings of the 5th International Conference on Virtual Learning, Section Software Solutions, October 29-October 31, pp. 375–380 (2010)
6. Popescu, D.A.: A relation between web pages. In: CKS 2011: Proceedings of Challenges of the Knowledge Society, April 15-16, pp. 2026–2033. "Nicolae Titulescu" University and "Complutense" University, Bucharest (2011)
7. Popescu, D.A.: Reducing the navigation graph associated to a web application. Buletin Stiintifc - Universitatea din Pitesti, Seria Matematica si Informatica (16), 125–130 (2010)
8. Anastasiu, D.P.: Classification of links and components in web applications. In: The 2nd International Conference on Operational Research ICOR 2010, Constanta, Romania, September 09-12 (2010)
9. Alalfi, M.H., Cordy, J.R., Dean, T.R.: Modeling methods for web application verification and testing: State of art. JohnWiley and Sons, Ltd. (2008)
10. Mao, C.-Y., Lu, Y.-S.: A Method for Measuring the Structure Complexity of Web Application. Wuhan University Journal of Natural Sciences 11(1) (2006)
11. Sreedhar, G., Chari, A.A., Ramana, V.V.: Measuring Qualitz of Web Site Navigation. Journal of Theoretical and Applied Information Technology 14(2) (2010)
12. Zhu, X., Goldberg, A.B.: Introduction to Semi-Supervised Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool (2009)
13. Zhu, X.: Semi-Supervised Learning Literature Survey. University of Wisconsin Madison (2008)
14. Alpine HTML Doctor, `http://www.alpineinternet.com/`
15. Validome HTML/XHTML/..., `http://www.validome.org/`
16. W3C Markup Validation Service, `http://validator.w3.org`