MEASURING THE QUALITY OF THE NAVIGATION IN WEB SITES USING THE CLONING RELATION

ANASTASIU POPESCU, Doru
University of Piteşti, Romania
Faculty of Mathematics and Computer Science
dopopan@gmail.com

Abstract

In this paper, a new method of calculating the WSC number is being presented, used for measuring the quality of navigation in a web site. The method of calculating this number is based on using a relation between the web pages of the web site and constructing a reduced navigation graph.

Keywords: *Navigation, Relation, Tag, HTML, Web Site*

AMS Classification: Primary 68U35, 68W01; Secondary 68Q60, 68N30, 68M11, 68N03

1. Introduction

The results of this paper refer to web sites which contain web pages consisting of HTML tags, saved in files with the extensions .html and .htm. Next, we name these web pages. On one hand, the number of web application built using this kind of web pages is very large; on the other hand, the web applications can contain a very large number of web pages. Navigating in these web sites is an especially important mechanism. There exist different methods of measuring the navigation in these sites. In the papers [10] and [11] such methods are being described, from which we will next use the WSC number (Web Site Complexity). Calculating the WSC number involves using the entire navigation graph (which most of the times has a very large number of nodes). In the following sections, we will introduce a method of calculating it, which uses an equivalence relation between the web pages of the web site, introduced in [2], [4], [6] and a reduced navigation graph, presented in [7].

2. Defining a relation between two web pages

Next, we will consider a web application with 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, $1 \leq i \leq n$, we write T_i the sequence of tags from p_i, which are not in TG (the order in which these are encountered is important).

Definition. Let TG be a set of tags, p_i and p_j two web pages P, $1 \leq i,j \leq n$. We say that p_i and p_j are *cloned* and we write p_i C p_j, if T_i and T_j coincide.

Example 1. We will next consider a web application with three web pages: P={p_1, p_2, p_3}. p_1 is found in the file pag1.html, p_2 is found in the file pag2.html, and p_3 is found in the file pag3.html.

```
pag1.html
<HTML>
<HEAD>
<TITLE>Web page 1</TITLE>
</HEAD>
<BODY>
<p> Link 1
<a href="pag3.html">Page 3</a>
</BODY>
</HTML>

pag2.html
<HTML>
<HEAD>
<TITLE>Web page 2</TITLE>
</HEAD>
<BODY>
<B> <p>Link 2 </p> </B>
<a href="pag3.htm">Page 3</a>
<p>
Relationnnnn
</BODY>
</HTML>

pag3.html
<HTML>
<HEAD>
<TITLE>Web page 3</TITLE>
</HEAD>
<BODY>
<FONT COLOR=red>Picture 3 </FONT>
<IMG SRC="office.jpg">
<FONT SIZE=4 COLOR=red>Picture 3 </FONT>
<IMG SRC="pic.jpg">
</BODY>
</HTML>
```

Considering:

```
TG={<p>, </p>, <B>, </B>, <HTML>, <HEAD>, <TITLE>,
</TITLE>, </HEAD>, <BODY>, </BODY>, </HTML>}
```

we obtain:

```
T1=(<a href="pag3.html">, </a>);
T2=(<a href="pag3.html">, </a>);
T3=(<FONT COLOR=red>, </FONT>, <IMG SRC="pic.jpg">,
<FONT SIZE=4 COLOR=red>, <IMG SRC="pic.jpg">, </FONT>).
```

According to the previous definition we obtain that only the pair of web pages p_1 and p_2 are cloned (p_1 C p_2), because T1=T2. Between p_1 and p_3, respectively p_2 and p_3 does not exist a cloning relation, because T1≠T3.

For the relation C defined on the set of web pages P={p_1, p_2, ..., p_n} from a web site, we define the nonoriented graph CG=(X,U), where:

- X={1,2,...,n}, represents the set of nodes, i is associated to the web page p_i, 1≤i≤n.
- U={[i,j]| p_i C p_j, 1≤i,j≤n, i≠j}, the set of edges.

Notes

1. The relation C, previously defined, is an equivalence relation.
2. The equivalence classes defined on the set P realise a partition of P. Any two web pages from the same equivalence class behave the same, because they contain the same tags (except the ones from TG) and in the same order. This way, the two cloned web pages contain the same links (supposing that TG does not contain those tags that define links).
3. For the previous example, there exist two equivalence classes C_1={p_1, p_2}, C_1={p_3}.

3. The navigation graph of a web site

We will define the navigation graph of a web site with the set of web pages P={p_1, p_2, ..., p_n}, written NG=(X,U), as:

- X={1,2,...,n}, represents the set of nodes; i is associated to the page p_i, 1≤i≤n.
- U={(i,j)| there exists a link from p_i to p_j, 1≤i,j≤n, i≠j}, the set of edges.

Example 2. We will consider a web site with the pages p_1, p_2, ... , p_{12} and a set TG of tags. The relation of *cloning* C is represented in fig. 1 (the CG graph), and the navigation in a web site is represented in fig. 2 (the NG graph). The obtained equivalence classes are the following:
C_1={1,6}; C_2={2,3,5,7,9}; C_3={4,8,10}, C_4={11}, C_5={11,12}.

4. The reduced navigation graph in a web site

We will consider a web site with the set of web pages P={p_1, p_2, ..., p_n}, the navigation graph NG and a set of tags given by TG, in order to be able

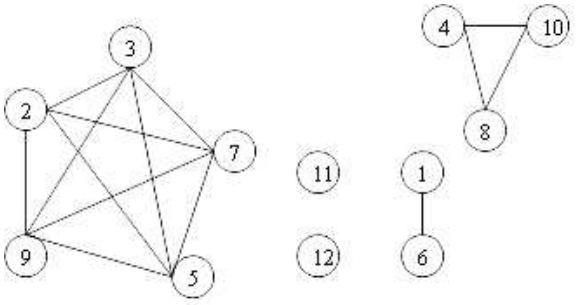

Figure 1. *The nonoriented CG graph associated to the C relation*

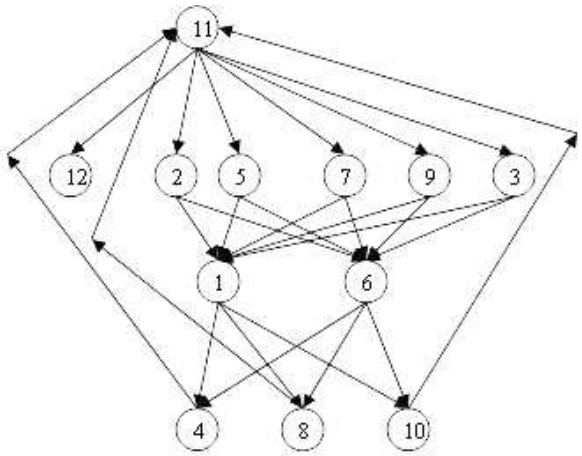

Figure 2. *Navigation Graph NG*

to use the *cloning* relation C. We write C_1, C_2, ..., C_k the equivalence classes obtained with the C relation and we obtain the CG graph.

The reduced navigation graph is a weighted oriented graph, which we will write as RNG=(X,U,c) and can be obtained as below:

- X={1,2,...,k}, represents the set of nodes; i is associated to the class C_i, $1 \leq i \leq k$.
- U={(i,j)| there exists in NG an edge (a,b) with $a \in C_i$ and $b \in C_j$, $1 \leq i,j \leq k$, $i \neq j$}, the set of edges.
- $c=(c_{ij})_{i,j=1,k}$, c_{ij} is the weight of the i-th edge, degined as the number of edges from NG that leave the nodes of the C_i class towards the nodes of the C_j class. If there does not exist an edge between i and j, then $c_{ij}=0$.

For the example in the previous section, the reduced graph is drawn in fig. 3; its weight is written on each edge.

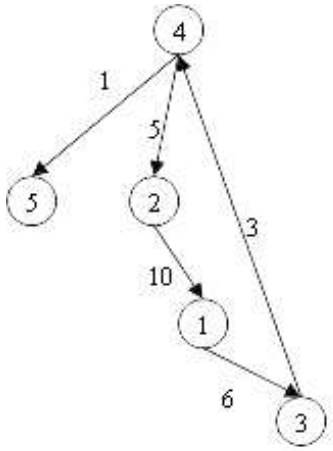

Figure 3. *The RNG graph*

5. **Measuring the quality of the navigation in web sites using the RNG graph**

In [10] and [11], several methods of measuring the navigation quality in a web site are being described. One of the methods uses the WSC number, defined using the navigation tree (defined in [10], [11], [9] or [8]). Another method of determining this number is the one that uses the RNG graph, defined in the previous section. More precisely:

$WSC=(s_1\text{-}n+1)/(n+s_2)$.

where:
s_1 is the sum of the edges' weights in RNG;
s_1 is the sum of the *return* edges' weights in RNG;
n is the number of web pages in the web site.

The *return* edges are obtained when the RNG graph is being searched (breadth or depth first, [3]), starting from the node associated to the start web page from the web site. Such an edge unites a current node with an already visited one. For the example in the previous sections, we have:
$s_1=1+5+10+6+3=25$.
$s_2=3$, for the only return edge in RNG, (3,4), the start page being in the fourth node.
$n=12$.
$WSC=(25\text{-}12+1)/(12+3)=14/15=0.9333$.

The method of calculating the WSC number that we propose implies using a graph with much less nodes and edges than the one used in [10] and [11].

6. Conclusion and future work

Algorithms of verifying the cloning relation, determining the equivalence classes and constructing the graphs used in the previous sections are presented in [2], [4], [5] and [6]. In a more general context, regarding the testing and verifying the web sites (using specific methods, [1]) where relations between the web pages can be used, the calculation of measurement numbers for the navigation quality becomes easy. I consider that a complex application, which should contain aspects introduced in this article as well would be very useful.

References

1. Alalfi, M.H., Cordy, J.R., and Dean, T.R., *Modeling methods for web application verification and testing: State of art*, John Wiley and Sons, Ltd, 2008.
2. Danauta, Catriniel Maria, Anastasiu Popescu, D., *Method of reduction of the web pages to be verified when validating a web site*, "Buletin Ştiinţific, Universitatea din Piteşti, Seria Matematică şi Informatică", 15, 19-24, 2009.
3. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C., *Introduction to algorithms*, second edition, MIT Press, 1990.
4. Anastasiu Popescu, D., Danauta, Catriniel Maria, and Szabo, Z., *A method of measuring the complexity of a web application from the point of view of cloning*, The 5th International Conference on Virtual Learning, Section Models and Methodologies, October 29 - October 31, Proceedings, 186-181, 2010.
5. Anastasiu Popescu, D., Szabo, Z., *Sink web pages of web application*, The 5th International Conference on Virtual Learning, Section Software Solutions, October 29 - October 31, Proceedings, 375-380, 2010.
6. Anastasiu Popescu, D., *A relation between web pages*, CKS 2011 Challenges of the Knowledge Society "Nicolae Titulescu" University and "Complutense" University, Bucharest, April, 15-16, CKS Proceedings, 2026-2033, 2011.
7. Anastasiu Popescu, D., *Reducing the navigation graph associated to a web application*, "Buletin Ştiinţific, Universitatea din Piteşti, Seria Matematică şi Informatică", 16, 125-130, 2010.
8. Popescu Anastasiu, D., *Classification of links and components in web applications*, The 2nd International Conference on Operational Research ICOR 2010, 09 to 12 September, Constanţa, România, 2010.
9. Cheng-Ying, M., Yan-Sheng, L., *A Method for Measuring the Structure Complexity of Web Application*, Wuhan University Journal of Natural Sciences, 11, 1, 2006.
10. Sreedhar, G., Chari, A.A., and Ramana, V.V., *Measuring Quality of Web Site Navigation*, Journal of Theoretical and Applied Information Technology, 14, 2, 2010.
11. Zhang, Y., Zhu, H., Greenwood, S., *Website Complexity Metrics for Measuring Navigability*, QSIC, 172-179, Quality Software, Fourth International Conference on QSIC'04, 2004.