# Chapter 68 Degree of Similarity of Web Applications

# Doru Anastasiu Popescu

University of Pitesti, Romania

# Dragos Nicolae

National College "Radu Greceanu" - Slatina, Romania

# ABSTRACT

In this chapter, the authors present a way of measuring the similarity between two Web applications. For this, they define the degree of similarity between two Web applications, taking into account only the Webpages composed of HTML tags. The authors also introduce an algorithm used to calculate this value, its implementation being made in the Java programming language.

#### INTRODUCTION

Web applications have a vast usage and a fast evolution. Consequently, various models have been created in view of web applications, especially used for verification and testing, such as those presented in (Alalfi, Cordy & Dean, 2008). The extensive development of these applications requires a mechanism for measuring their quality, these aspects having been studied in many papers, such as (Cheng-ying & Yan-sheng, 2006; Sreedhar, Chari & Ramana, 2010; Popescu & Szabo, 2010; Popescu, 2011; Popescu & Danauta, 2011). This chapter aims to determine an algorithm of measuring the similarity between two web applications. Another method of measuring the similarity between web applications has been introduced in (Popescu & Danauta, 2011) and it uses a relation between the web pages of an application, relation taken from (Popescu & Szabo, 2010; Popescu, 2011; Popescu & Danauta 2012). The formula we introduce (section 2) does not use this relation. It is based on comparing the tags of two web pages, using an algorithm for determining a common subsequence for two strings of tags. The algorithm which calculates the similarity degree is presented in section 3. The implementation and the results obtained with this algorithm are presented in section 4.

DOI: 10.4018/978-1-5225-3422-8.ch068

#### THE DEGREE OF SIMILARITY

Let WA1 and WA2 be two web applications. The application WA1 is considered to be composed of the web pages  $p_1$ ,  $p_2$ ,...,  $p_n$  and the application WA2 composed of the web pages  $q_1$ ,  $q_2$ ,...,  $q_m$ . We will also establish a set TG of tags.

For a web page  $p_i$  we build a sequence with all its tags, excluding those which are also in TG, keeping their order and removing their attributes.

### **Definition 1**

For two sequences of tags  $T_1$  and  $T_2$ , associated to the web pages  $p_i$  from WA1 and  $q_j$  from WA2, we define the degree of similarity between  $p_i$  and  $q_j$ , written  $nr_{ij}$ , as being the number equal to the maximum length of a common subsequence of tags for  $T_1$  and  $T_2$ .

#### Definition 2

For a web page p from WA1, we define de similarity degree of p with WA2 as being the number: degpage(p,WA2)=k/NT, where  $k=max\{nr_{ij} | 0 < j < m+1\}$ , NT is the number of tags from p which are not in TG and i is an index, 0 < i < n+1 for which  $p=p_i$ .

## **Definition 3**

We define the degree of similarity between WA1 and WA2 as being the number: deg(WA1,WA2)=s/n, where  $s=degpage(p_1,WA2) + degpage(p_2,WA2) + ... + degpage(p_n,WA2)$ .

**Remark 1:**  $0 < deg(WA1, WA2) \le 1$ .

**Remark 2:** If deg(WA1,WA2) = 1, then for any web page  $p_i$  from WA1, there is a web page  $q_j$  in WA2 so that  $T_1$  is a subsequence of  $T_2$ , where  $T_1$  is the sequence of tags from  $p_i$ , which are not in TG, and  $T_2$  is the sequence of tags from  $q_i$  which are not in TG.

#### Example

Let us consider the set TG={<HTML>, </HTML>, <HEAD>, </HEAD>, </HEAD>, </TITLE>, </TITLE>, </BODY>, </BODY>}, the web application WA1 composed of the web pages p1 and p2, as well as the web application WA2 composed of the web pages q1, q2 and q3. The files P1.html, P2.html for p1, p2 and Q1.html, Q2.html, Q3.html for q1, q2 and q3 are as shown in Box 1.

We obtain the following results:

The sequences of tags, which are not in TG, for each web page:

```
Tp<sub>1</sub>=(<B>, </B>, <IMG>)
Tp<sub>2</sub>=(<I>, </I>, <BR>, <BR>, <IMG>)
Tq<sub>1</sub>=(<B>, </I>, <BR>, <BR>, <IMG>)
Tq<sub>2</sub>=(<I>, </B>)
Tq<sub>3</sub>=(<I>, </I>)
Tq<sub>4</sub>=(<I>, </I>, <BR>, <BR>, <IMG>)
```

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/degree-of-similarity-of-web-applications/188272

# **Related Content**

#### A Systematic Approach for Designing Educational Recommender Systems

Patrick H. S. Brito, Ig Ibert Bittencourt, Aydano Pamponet Machado, Evandro Costa, Olavo Holanda, Rafael Ferreiraand Thiago Ribeiro (2014). *Software Design and Development: Concepts, Methodologies, Tools, and Applications (pp. 1264-1288).* 

www.irma-international.org/chapter/systematic-approach-designing-educational-recommender/77757

#### Depression Detection Model in Social Network Content: A Survey

Esraa M. Rabie, Atef F. Hashemand Fahad Kamal Alsheref (2022). International Journal of Software Innovation (pp. 1-12).

www.irma-international.org/article/depression-detection-model-in-social-network-content/309112

#### Big Data: The Path to Maturity

Stephen H. Kaisler, William H. Money, Frank Armourand J. Alberto Espinosa (2017). *International Journal of Systems and Service-Oriented Engineering (pp. 1-23).* www.irma-international.org/article/big-data/190410

#### Agile Quality Assurance Techniques for GUI-Based Applications

A. Memonand Q. Xie (2007). *Agile Software Development Quality Assurance (pp. 114-134).* www.irma-international.org/chapter/agile-quality-assurance-techniques-gui/5071

#### A Survey and Taxonomy of Intent-Based Code Search

Shailesh Kumar Shivakumar (2021). *International Journal of Software Innovation (pp. 69-110)*. www.irma-international.org/article/a-survey-and-taxonomy-of-intent-based-code-search/266283