

Approximately Similarity Measurement of Web Sites Using Genetic Algorithms and Binary Trees

Authors : Doru Anastasiu Popescu, Dan Rădulescu

Abstract : In this paper, we determine the similarity of two HTML web applications. We are going to use a genetic algorithm in order to determine the most significant web pages of each application (we are not going to use every web page of a site). Using these significant web pages, we will find the similarity value between the two applications. The algorithm is going to be efficient because we are going to use a reduced number of web pages for comparisons but it will return an approximate value of the similarity. The binary trees are used to keep the tags from the significant pages. The algorithm was implemented in Java language.

Keywords : Tag, HTML, web page, genetic algorithm, similarity value, binary tree

Conference Title : ICICS 2016 : International Conference on Information and Communications Security

Conference Location : San Francisco, United States

Conference Dates : June 09-10, 2016