

Accepted Manuscript

An optimized approach for massive web page classification using entity similarity based on semantic network

Huakang Li, Zheng Xu, Tao Li, Guozi Sun, Kim-Kwang Raymond Choo

PII: S0167-739X(17)30332-1

DOI: <http://dx.doi.org/10.1016/j.future.2017.03.003>

Reference: FUTURE 3371

To appear in: *Future Generation Computer Systems*

Received date: 16 February 2016

Revised date: 11 December 2016

Accepted date: 1 March 2017

Please cite this article as: H. Li, Z. Xu, T. Li, G. Sun, K.-K. Raymond Choo, An optimized approach for massive web page classification using entity similarity based on semantic network, *Future Generation Computer Systems* (2017), <http://dx.doi.org/10.1016/j.future.2017.03.003>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



1. A weight estimation algorithm based on the depth and breadth of Wikipedia network is used to calculate the class weight of all Wikipedia words.
2. A kinship-relation association based on content similarity was proposed to optimize the unbalance problem.
3. Bayesian classifier is used to estimate the page class probability.

ACCEPTED MANUSCRIPT

Forbidden

You don't have permission to access this resource.

Apache/2.4.25 (Debian) Server at daneshyari.com Port 443