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

Highlights (for review)

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.

Forbidden

You don't have permission to access this resource.
