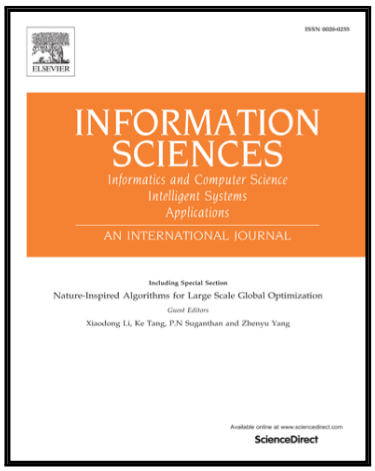

Explicitly and Implicitly Exploiting the Hierarchical Structure for Mining Website Interests on News Events

Junyu Xuan, Xiangfeng Luo, Jie Lu, Guangquan Zhang

PII: S0020-0255(17)30895-2
DOI: 10.1016/j.ins.2017.08.056
Reference: INS 13060

To appear in: *Information Sciences*

Received date: 28 January 2016
Revised date: 14 June 2017
Accepted date: 15 August 2017

Please cite this article as: Junyu Xuan, Xiangfeng Luo, Jie Lu, Guangquan Zhang, Explicitly and Implicitly Exploiting the Hierarchical Structure for Mining Website Interests on News Events, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.08.056

Explicitly and Implicitly Exploiting the Hierarchical Structure for Mining Website Interests on News Events

Junyu Xuan[a,b], Xiangfeng Luo[a], Jie Lu[b,*], Guangquan Zhang[b]

[a] *School of Computer Engineering and Science, Shanghai University, 99 Shangda Road, Shanghai, China*

[b] *Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Sydney, Australia*

Abstract

After a news event, many different websites publish coverage of that event, each expressing their own unique commentary, perspectives, and viewpoints. Websites form around a specific set of interests to cater to different audiences, and discovering these interests can help audiences C especially people and organizations that are interested in news C select the most appropriate websites to use as their sources of information. This paper presents three methods for formally defining and mining a websites interests, each of which is explicitly or implicitly based on a hierarchial structure: website-webpage-keyword. The first, and most straightforward, method explicitly uses keyword-layer network communities and the mapping relations between websites and keywords. The second method expands upon the first method with an iterative algorithm that combines both the mapping relations and the network relations from the website-webpage-keyword structure to further refine the keyword-layer network communities. In the third method, a website topic model implicitly captures the mapping relations among the websites, webpages, and keywords. The performance of three proposed methods in website interest mining is compared using a bespoke evaluation metric. The experimental results show that the iterative procedure designed in the second method is able to improve website interest mining performance, and

*Corresponding author

Email addresses: xuanjunyu@shu.edu.cn (Junyu Xuan), luoxf@shu.edu.cn (Xiangfeng Luo), Jie.Lu@uts.edu.au (Jie Lu), Guangquan.Zhang@uts.edu.au (Guangquan Zhang)