

Accepted Manuscript

How “Small” Reflects “Large”? — Representative Information Measurement and Extraction

Guoqing Chen, Cong Wang, Mingyue Zhang, Qiang Wei, Baojun Ma

PII: S0020-0255(17)30924-6
DOI: [10.1016/j.ins.2017.08.096](https://doi.org/10.1016/j.ins.2017.08.096)
Reference: INS 13101



To appear in: *Information Sciences*

Received date: 31 March 2017
Revised date: 1 August 2017
Accepted date: 31 August 2017

Please cite this article as: Guoqing Chen, Cong Wang, Mingyue Zhang, Qiang Wei, Baojun Ma, How “Small” Reflects “Large”? — Representative Information Measurement and Extraction, *Information Sciences* (2017), doi: [10.1016/j.ins.2017.08.096](https://doi.org/10.1016/j.ins.2017.08.096)

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.

How “Small” Reflects “Large”?

— Representative Information Measurement and Extraction[☆]

Guoqing Chen^a, Cong Wang^a, Mingyue Zhang^{a,b,*}, Qiang Wei^a, Baojun Ma^c

^a*China Retail Research Center, School of Economics and Management, Tsinghua University, Beijing 100084, China*

^b*International Business School, Beijing Foreign Studies University, Beijing 100089, China*

^c*School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China*

Abstract

While web services avail a rapid growth of data volume for use, identifying helpful information is of great value, especially when users face with an unwilling glut of information. Thus, it is deemed relevant and meaningful to provide users with a representative subset (i.e., small set) that could well reflect the original information corpus (i.e., large set). In such a large-small context, this paper addresses the issues of representativeness in light of measurement and extraction by reviewing our previous efforts. Specifically, we first discuss various metrics from different perspectives of representativeness, then present a series of related representativeness extraction methods. Finally as a supplement and extension, a recent effort is introduced, which aims to take information quality into account in deriving a ranked subset. The proposed extraction method is justified by extensive real-world data experiments, showing its superiority to others in both effectiveness and efficiency.

Keywords: representative, coverage, redundancy, consistency, compactness, extraction

1. Introduction

Increasing popularity of web services, such as search engines, product review platforms, and academic databases, has resulted in the so-called information

[☆]Partly supported by the National Natural Science Foundation of China (71490724/71372044/71402007/71110107027), and the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities of China (12JJD630001).

*Corresponding author

Email addresses: chengq@sem.tsinghua.edu.cn (Guoqing Chen), wangc3.14@sem.tsinghua.edu.cn (Cong Wang), zhangmy@bfsu.edu.cn (Mingyue Zhang), weiq@sem.tsinghua.edu.cn (Qiang Wei), mabaojun@bupt.edu.cn (Baojun Ma)

Download English Version:

<https://daneshyari.com/en/article/6856305>

Download Persian Version:

<https://daneshyari.com/article/6856305>

[Daneshyari.com](https://daneshyari.com)