

# MISQ Archivist

## Extracting Representative Information on Intra-Organizational Blogging Platforms

*Xunhua Guo, Qiang Wei, Guoqing Chen, Jin Zhang, and Dandan Qiao*

---

### Abstract

Content generated on intra-organizational blogging platforms may help managers understand the emerging ideas, issues, and opportunities of their companies, whereas the difficulty is how to go beyond the overloaded amount of information and obtain an overall view. This paper proposes a system framework for extracting representative information from intra-organizational blogging platforms, as well as the REPSET (REPresentative SET) method, which serves as the core component of the extraction system. Drawing on a novel clustering technique, REPSET is designed to identify a small set of items that largely represent the diversified content of a huge information base. Built upon REPSET, an extraction system enables managers to locate representative articles that may serve as starting points for comprehensively understanding the hot topics, prevailing thoughts, and emerging opinions among employees. Empirical evaluations are conducted based on the massive database accumulated on an internal blogging platform at a large telecommunications company. The results from data experiments and user evaluations demonstrate that REPSET and the extraction system upon which it is based can provide outstanding performance, in comparison with benchmark methods.

**Keywords:** Intra-organizational blogging platforms, representative information extraction, enterprise social media, business intelligence, design science