

---

## Preface

---

### Subhash Bhalla

Graduate School of Computer Science and Engineering,  
University of Aizu,  
Aizu-wakamatsu, Fukkushima, 965-8580, Japan  
Email: bhalla.subhash@gmail.com

**Biographical notes:** Subhash Bhalla is a Professor at the Graduate School of Computer Science and Engineering, University of Aizu. He obtained his PhD in Computer Science in 1984. His research interests include distributed information systems, real-time operating systems, new query languages for web data resources and information retrieval in healthcare.

---

Large-scale information systems in transportation, banking and public-utility services depend on computing infrastructure. Many research efforts are being made in related areas, such as – web-based computing, information accesses by web users, wireless computing and sensor networks. Government agencies in many countries plan to launch facilities in education, healthcare and information support as a part of e-government initiative. In this context, information interchange management has become an active research field. A number of new opportunities have evolved in design and modelling based on new computing needs of the users. Database systems play a central role in supporting networked information systems for access and storage management aspects.

In the same context, the Seventh International Workshop on Databases in Networked Information Systems (DNIS) 2011 was held on October 12–14, 2011 at University of Aizu in Japan. The workshop program included research contributions, and invited contributions. A view of research activity in related areas was provided by special session on the topics. These included, geospatial decision making, web data management systems, networked information systems-infrastructure and web query and web mining systems. Some invited contributions have been contributed by Professor M. Kitsuregawa, Professor H.V. Jagadish, and Professor Jie Li. Revised versions of these and other selected contributed manuscripts have been included in this issue.

The articles in this issue begin by introducing energy consumption of storage at data centres, from an energy-saving perspective. Energy consumption is increasing rapidly. Large storage facilities have various RAID configurations incorporating different RAID levels, numbers of drives, and media types. The manuscript investigates different ways, for RAID configuration's effect on application performance and power consumption at storage installations.

Under the network information systems-infrastructure, the paper by Li, Ye, Xu and Liu proposes a coding scheme called separate coding (SC) for collecting all data continuously in wireless sensor networks with a mobile base station (mBS). By separately encoding a certain number of

data segments in a combined segment, and storing the combined segments in the corresponding buffers of each sensor node, the scheme provides an efficient storage method. By randomly querying a small subset of sensor nodes, the mBS can reconstruct all the original data segments with high probability. The success ratio of SC-based data collection is very close to 100%. Similarly, in the performance improvement section, the paper by Saito, Kuboyama and Yasuda considers user behaviour modelling. This paper proposes a framework for modelling user behaviour from low-level computer usage logs aiming to find working patterns and behaviours of employees at work.

The subsequent article by Hashimoto, Chakraborty and Shirota considers the efficiency of extracting messages from social media sites. The authors examine how individual users post their opinions and gradually build their consensus. It is recognised as a process of pervasive collaboration. Tracking topic transitions over time on social media gains a rich insight into exploring its social context. However, topic transition detection is difficult, because messages in social media are written in a colloquial style using unknown/compound words, and topics dynamically come and go. The paper proposes to use a modularity measure that shows the quality of a division of a network into modules or communities, for topic transition detection.

The article by Jagadish, Qian and Nandi considers a problem in database design. Databases today are carefully engineered. There is an expensive and deliberate design process, after which a database schema is defined; during this design process, various possible instance examples and use cases are hypothesised and carefully analysed; finally, the schema is ready and then can be populated with data. Such an effort is a major barrier to database adoption. This paper explores the possibility of organic database creation instead of the traditional engineered approach. It lets the user start storing data in a database with a schema that is just enough to cover the instances at hand. Then it supports an efficient schema evolution as new data instances arrive. By designing the database to evolve, it can sidestep the expensive front-end cost of carefully engineering the design of the database. Indeed, the deliberate design model

complicates not only database creation, but also database transformation (i.e., schema mapping), because traditional schema mapping tasks are carefully engineered with declarative specification hidden beneath complex user interface. The paper studies the issue of organic database transformation, which automatically induces schema mappings from sample target database instances.

Similarly, the article by Madaan and Chu considers the problem of queries over web data resources. The World Wide Web has become a large source of health information. The paper-based medical resources are becoming available on the web. Hence, web-based information retrieval, automatic page-adaptation and in-depth querying are gaining importance, in the healthcare domain. To address the problems of these ever-expanding information systems over the internet, traditional information retrieval techniques are insufficient. This study aims to help users in the healthcare domain for in-depth querying of web-based healthcare information resources. It compares the existing approaches for in-depth querying for segment-level searches rather than page-level searches. It proposes a web document segmentation-based query-by-segment tag (QBT) query-interface. It uses the semantic and structural relationships among the various content groups.

The article by Kikuchi considers implementation aspects in a cloud computing setting. Large-scale public utility services are supported by cloud computing. Various applications including data quality management (DQM) aim to reconsider their structural aspects for large-scale deployment. In enterprise computing, huge efforts to apply service-oriented architecture and business process integration have been adopted. The paper considers operational issues of master data management. It is a sub-fields of DQM.

Other than introducing the large-scale computing applications, the papers focus on presenting a complete spectrum of essential components. These are complex query, infra-structural support and upcoming application in large-scale computing. These present designs and algorithmic aspects of supporting large-scale computations.