

Editorial: Special Issue on Information Reuse and Integration

The increasing volumes and dimensions of information have dramatic impact on effective decision-making. To remedy this situation, Information Reuse and Integration (IRI) seeks to maximize the reuse of information by creating simple, rich, and reusable knowledge representations and consequently explores strategies for integrating this knowledge into legacy systems. IRI plays a pivotal role in the capture, representation, maintenance, integration, validation, and extrapolation of information; and applies both information and knowledge for enhancing decision-making in various application domains. This special issue includes six papers.

The six papers cover an interesting range of topics from software reuse and measurements to information and knowledge integration. The first two papers discuss issues and techniques in knowledge-based approaches to software measurements and evaluation. The next two papers investigate the development of Web-based intelligent systems for image tagging and user dialog. The last two papers present optimization issues and mappings between OO databases and XML.

The first paper by Folleco, Khoshgoftaar, Hulse and Napolitano presents their investigation into the robustness of a variety of common-used learning algorithms relative to low quality, class imbalanced measurement data. The authors identify learners from a total of 11 classification algorithms with robust performance in the presence of low quality imbalanced measurement data. Four performance metrics suitable for class imbalanced data were used to measure learner performance. Based on their results, the results demonstrate that the quality of the measurement data can impact classification performance significantly. The authors recommend using the random forest ensemble learning technique for building classification models from software measurement data, regardless of the quality and class distribution of the data.

The paper by Far, Mudigonda and Elamy focuses on the design and development of a General Purpose Software Evaluation (GPSE) system that uses statistical methods based on Multidimensional Weighted Attribute Framework (MWAF) for the evaluation of software systems. The architectural elements of MWAF are essentially survey questionnaire which gathers information from several domain experts. The GPSE system then applies principles of Analysis of Variance (ANOVA) and Tukey's pairwise comparison tests on the collected data to arrive at selection of the best suited alternative for the given problem. The authors have fully implemented the GPSE system and tested it on several projects including evaluation of multi-agent development methodologies and selection of COTS products.

The paper of Özyer proposes a collaborative infrastructure that helps users store pictures that they have seen on Web pages during surfing. The system

consists of two modules: an add-on application that will work on Firefox Mozilla browser and the Web page itself. The add-on application helps user tag and store pictures that will remind of something meaningful to him. The author has implemented a system that lets users bookmark pictures they want to keep with their tag info according to image content. Later on the users may reorganize the tags and may also perform search by using tag information rather than image content.

The paper by Shibata, Nishiguchi, and Tomiura discusses a type of open-ended dialog system that generates appropriated responses based on the vast amount of Web documents. Generated from Web documents, the candidate corpus in Japanese could maintain surface cohesion and semantic coherence. The authors have prototyped the system that is capable of meaningful conversation on various topics. They have experimented on a conversation about movies, with the result showing that the system could generate 66% appropriate responses.

In their paper, Faraz Rafi, Zaidi and Levis present two algorithms for the optimal selections of a sequential evolution of actions, in conjunction with the preconditions of their environment and their effects, depicted by Activation Timed Influence Nets, given a set of preconditions. A special case for the two algorithms is also considered where the selection of actions is further constrained by the use of dependencies among them. The algorithms are based on two different optimization criteria: one maximizes the probability of a given set of target effects, while the other maximizes the average worth of the effects' vector.

The last paper by Naser, Alhadjj, and Ridley presents a novel approach for mapping an existing object-oriented database into XML and vice versa. They first derive the so-called object graph based on characteristics of the schema to be mapped. For object-oriented schema, the object graph simply summarizes and includes all nesting and inheritance links. The inheritance is then simulated in terms of nesting to get a simulated object graph. Doing so, everything in a simulated object graph can be directly represented in XML, and thus the mapping of the actual data from the object-oriented database into corresponding XML document(s) is easily achieved. The mapping from XML into object-oriented database is similarly achieved, where everything in a simulated object graph could be directly represented in object-oriented database.

Finally, we would like to thank all the authors for their efforts in enhancing and extending their papers for this special issue. Thanks also the Informatica team for their assistance in making this special issue published in a timely fashion.

*Reda Alhadjj, University of Calgary, Canada
Kang Zhang, University of Texas at Dallas, USA*

