# Editor's Introduction

In this JIM issue, we are delighted to present four research papers. The summaries of these papers are as follows.

Ming-Dar Hwang and Yi-Chung Chu in their paper "Optimizing Software Product Development for Small Company" propose the SPDG (Software Product Develop Guideline) for a small software company which allows the execution of tasks taking into account the existing management system. Small software companies in the development process of software products must subject to tight schedule and resources to withstand a great test. The purpose of their study is to analyze a case study for over 6 years its practical experience in software product development process against Microsoft Solutions Framework (MSF, Microsoft Solution Framework) in order to identify the execution methods for increasing market compliance, reducing development cost, and shortening the development time. That allows us to develop a practice guide for software product development in small software companies. The proposed SPDG for small software companies follows the basic principles, logics, and the team model of MSFv4 and serve as the foundation for software product management. In the guidelines, their study identifies 15 key execution methods based on practical experience in the case study. Following the SPDG, small software companies can save development time by 16.1% and improve product quality and work efficiency. The SPDG is proven to be an effective software project management guide for small software companies. The paper adopted case study research method to analyze only one commercial software company. Future research may study more companies for cross-company analysis and use quantitative data to develop other insights. It confirms that the development of a set of proven, practical guidelines for gradual improvement is feasible. A small software company can use the proposed SPDG to implement and manage its new software development project.

Shin-Jer Yang and Shin-Jer Yang in their paper "Design Adaptive Resource Allocation Scheme to Improve the Processing Performance in the Cloud Computing" propose the ARAS (Adaptive Resource Allocation Scheme) that is a more effective and efficient resource-management method than the VMWare DRS; it can on-line monitor and tune the computing resources on the VM. Also, the ARAS scheme can obtain better processing time and higher computing-resource utilization than DRS to enhance the processing performance of cloud applications. Cloud computing is a collection of

virtualized computing resources such as CPU, memory, and other related resources, there is a need to support resources load balancing management. When cloud computing is running on the VMware, it utilizes DRS to create a pool for setting computing resources to each VM that then activate these VMs while creating a resource pool reserved for which child resource may be larger than their needs. Hence, the paper proposes ARAS scheme to dynamically monitor and adjust computing resources under heavy loads on the cloud. They perform simulations using Hyper pi to examine that the task scheduling in cloud environments is to be compared between ARAS and DRS via four KPIs. Hence, simulation results via Hyper pi show that the ARAS can reduce processing time about 5%~26%, enhance CPU utilization about 7.2%~10%, and increase memory Read/Write and caching ability about 2%~20%. Except using Hyper pi tool, they utilize a cloud performance simulation tool: New Relic and integrate it into cloud applications to simulate and compare the performance between ARAS and DRS. The simulation results via New Relic indicate that the ARAS can obtain better CPU, memory and caching utilizations, a higher throughput about 14.7% than DRS. Their simulation environment is only performed on the two physical machines: Server1 and Server2. Also, further research is recommended to evaluate more than two physical machines and also support live migration in virtualized cloud computing. The virtualization can encapsulate the infrastructure into VMs and create quantitative resources under cloud computing. Also, according to different resource needs of the actual environment, the ARAS method is more suitable for resource management to solve the loading of virtualized computing resources.

Jen-Yuan Yeh   Wei-Pang Yang   Hao-Ren Ke and Pei-Cheng Cheng in their paper "Extraction-based News Summarization Using Sentence Centrality in the Sentence Similarity Network" employ graph-based text analysis to model documents and investigate measures of graph-based centrality as sentence salience in summarization. One widely-adopted summarization paradigm, sentence extraction, aims at extracting important sentences and composing them into a summary. The foundation towards sentence extraction is to assess importance of sentences in the summary so as to rank sentences for extraction. The paper models documents on the same (or related) topic as a sentence similarity network, in which a sentence is regarded as a node and relationship between sentences only exists if they are semantically related. Several methods for evaluating the importance of a node (i.e., a sentence) in the network are then proposed, namely: (1) Degree Centrality; (2) Normalized Similarity-based Degree Centrality; (3) HITS Centrality; (4) PageRank Centrality; and (5) iSpreadRank Centrality. All are

designed on the basis of the idea that the importance of a node is determined not only by the number of nodes to which it connects, but also by the importance of its connected nodes. As to summary generation, CSIS (Cross-Sentence Information Sub-sumption) is employed for anti-redundancy while extracting sentences according to the sentence ranking produced based on the centrality of sentences. The proposed summarization method was evaluated using the ROUGE evaluation suite on the DUC 2004 news stories collection. Experimental results show that, while considering the ROUGE-1 metric, the performance ranking is: iSpreadRank > Normalized Similarity-base Degree > PageRank > Degree > HITS. Another experiment, conducted to combine sentence centrality with surface-level features, also presents competitive results, compared with the best participant in the DUC 2004 evaluation. The proposed summarization method is in an unsupervised manner; thus no training dataset is required. Since no domain-specific knowledge or deep linguistic analysis is exploited, the method is domain- and language-independent. However, it might lead to poor understanding of the input texts and would probably produces poor summaries, due to neither deep analysis of natural language processing performed, discourse structure considered, nor domain-specific knowledge involved in the process of summarization. The contributions of this work are threefold. First, the paper offers a sentence similarity network to model topic-related documents. Second, novel graph-based sentence ranking methods are explored to rank the importance of sentences for extraction. Finally, the proposed method had been proven successful in a case study with the DUC 2004 benchmark dataset. Directions for future research would be: (1) instead of symbolic-level analysis, to take into account semantics, such as synonymy, polysemy, and term dependency, while determining if two sentences are semantically related; (2) to investigate graph-based centrality developed in social network analysis for evaluating sentence salience in summarization; (3) to improve the cohesion and coherence of summaries using natural language processing techniques, such as sentence planning and generation.

Ya-Han Hu, Tony Cheng-Kui Huang, and Cheng-Han Yang in their paper "A Genetic Algorithm Based Approach for Text Categorization" investigate the influence of the parameters used in GA for the feature selection in TC and advance the literature in choosing GA parameters and classification techniques for optimizing the TC performance. Digital data has been accumulated rapidly resulting in the significant increase in the cost of searching information from the data source. How to effectively manage documents (i.e., text categorization, TC) has become an important research issue. However, in TC, huge amount of index terms are selected for representing document

vectors, resulting in poor prediction outcomes. The study proposes a genetic algorithm based feature selection (GAFS) method to optimize the selection of index terms. Before training classifiers, GAFS selects a reduced set of index terms that can optimize the prediction accuracy of classifiers. In experimental study, the WebKB dataset was used to evaluate the performance of GAFS. A total of six well-known classification techniques were considered, including naïve Bayesian classifier (NB), decision tree (DT), classification and regression tree (CART), random forest (RF), support vector machine (SVM) and k-nearest neighbor (kNN). The baseline model, denoted as TOTAL, is to consider complete set of index terms in all experiments. The results show that the proposed GAFS method outperforms the TOTAL method. The performance of kNN and RF classifiers deteriorates as the number of features increases. Under different number of features, the SVM, NB, and DT classifiers perform stably but the CART classifier has relatively unstable performance. The study only considers the WebKB dataset. Future research is recommended to include other well-known datasets in the TC domain. Other feature selection methods can be also considered in the experimental evaluation. Two practical implications are provided. First, this study reveals that different parameter settings in genetic algorithm (GA) can significantly affect the performance of feature selection in TC. Second, the proposed GAFS method allows users to systematically construct a robust classifier for TC.

Finally, on behalf of the editorial team, I would like to thank all the authors and reviewers for their collaborative efforts to make this issue possible. It is our sincere wish that this journal become a bilingual knowledge exchange platform among information systems researchers around the world. Finally, to our loyal readers, we hope you find the contents of the papers useful to your work or research.

Sincerely yours,

*Eldon Y. Li*

Eldon Y. Li, Ph.D.
Editor-in-Chief and University Chair Professor
Department of Management Information Systems
College of Commerce
National Chengchi University
Taipei, Taiwan
July 2014