
Editorial

Mathew Palakal

School of Informatics,
Indiana University Purdue University Indianapolis,
535 West Michigan St., IT 475,
Indianapolis, Indiana, USA
Fax: 317-278-4140
E-mail: mpalakal@iupui.edu

Biographical notes: Mathew Palakal is a Professor of Informatics and Computer Science at Indiana University Purdue University at Indianapolis. He received his PhD in the Department of Computer Science at Concordia University, Montreal in 1987. His research interests include biomedical text mining, bioinformatics, and knowledge discovery.

As research in the biomedical informatics field continues to make significant advances, massive heterogeneous data are also being generated rather exponentially. The main focus of the Bioinformatics Track in the *2007 ACM Symposium on Applied Computing* was to address computational challenges that are involved in all aspects of biological data integration and analyses. This special edition of the *International Journal of Data Mining and Bioinformatics* contains the top three selected papers from the ACM Symposium.

In their paper on 'Finding new Core Promoter Elements using backward-looking strategies', Huang, Jhan and Liou suggest that core promoter elements are key players in transcription initiation and identifying these *cis*-acting elements within the core promoter region is crucial to deciphering the mechanisms of gene expression control. A novel framework for finding new core promoter elements is proposed in this paper, based on their experimental evidence. The paper then concludes that the proposed system is feasible and able to reliably identify core promoter elements, regardless of species.

A generalised framework for microarray data integration is the theme of the paper, 'A cube framework for incorporating inter-gene information into biological data mining', by Lin, Kang, Shin and Lee. Large volumes of microarray data generated from various experiments are being routinely added to public repositories. Owing to the inherent heterogeneity of these microarray experiments, the data generated from different experiments have not been directly integrated, and hence, these resources have not been fully utilised. In this paper, the authors developed and implemented a new framework that generalises two existing approaches to address the microarray data integration problem.

The final paper, 'An on demand data integration model for biological databases', by Palakal and Naidu, presents a user-centric biological query system for information integration and knowledge acquisition from distributed, semantically heterogeneous, data sources. The proposed system, BioXBase, extracts user requested information over the internet from multiple biological data sources and organises the resulting variety

of information into a homogeneous unified view to the user after data is cleaned, processed and integrated in real time, on-the-fly.

Collectively, these three papers in this special edition of the *International Journal of Data Mining and Bioinformatics* present three different, but important, computational aspects surrounding data and information management and knowledge discovery from heterogeneous biological data sources.