

Selected Abstracts from *JASIST*

Authors who choose to do so prepare and submit these summaries to the editor of the *Bulletin*.

From *JASIST* v. 60 (1)

Neuhaus, C., Marx, W., & Daniel, H.D. (2009). The publication and citation impact profiles of *Angewandte Chemie* and the *Journal of the American Chemical Society* based on the sections of *Chemical Abstracts*: A case study on the limitations of the *Journal Impact Factor*. (176-183).

Study and Results: Taking *Angewandte Chemie International Edition* and the *Journal of the American Chemical Society* as examples, the study examines the publication and citation impact profiles of both journals across the sections of the bibliographic database *Chemical Abstracts*. The findings suggest that a single measure of journal citation impact such as the *Journal Impact Factor* published by Thomson Reuters is insufficient for characterizing the significance and performance of multidisciplinary and wide-scope journals.

What's New? The findings show that the information available in the *Science Citation Index* is a rather unreliable indication of the document type and is therefore inappropriate for

comparative analysis of journals. The findings further suggest that the composition of the journal in terms of contribution types, the length of the citation window and the thematic focus of the journal in terms of the sections of *Chemical Abstracts* has a significant influence on the overall journal citation impact. For the comparison of multidisciplinary and wide-scope journals more sophisticated methods such as publication and citation impact profiles across subject headings of bibliographic databases (for example, the sections of *Chemical Abstracts*) are therefore valuable.

Limitations: A high quality of indexing information is a prerequisite for the applicability of a subject classification scheme to comparative analysis.

From *JASIST* v. 60 (2)

Bose, I., & Chen, X. (2009). A method for extension of generative topographic mapping for fuzzy clustering. (363-372).

Study and Results: We developed a clustering method (GTMFCM) that combined generative topographic mapping (GTM) and fuzzy c-means algorithm (FCM). It is observed that this method performs better than FCM and Gustafson-Kessel

algorithms in terms of values of clustering validity indexes. In business applications, this new method can be used to explore segments of customers to create a complete and vivid profile of customers' behavioral patterns. Businesses can benefit from such knowledge by aligning their marketing strategies with customers' preferences.

What's New? Clustering algorithms explore the hidden behavioral patterns of customers. However the presentation and interpretation of clustering results are difficult for practitioners. Fuzzy clustering techniques are believed to be more capable of revealing information about customers' behavioral patterns because they assign data to clusters with probabilities. On the other hand, techniques such as GTM visualize the distribution of customers but cannot group them into the desired number of clusters. By combining the two techniques, we created a tool which can do visualization and clustering at the same time with acceptable performance.

Limitations: The paper is limited in that the proposed method has been validated using benchmark and simulated data sets but needs further validation using customer data collected from real business situations.

From JASIST v. 60 (2)

Pera, M. S., & Ng, Y.-K. (2009). SpamED: A spam email detection approach based on phrase similarity. (393-411).

Study and Results: We introduce a novel spam-email detection approach, denoted *SpamED*, which is designed for solving today's problem on increasing influx of spam emails that reach user's inboxes, leading to monetary loss and waste of computational resources. The premise of our investigation is to correctly identify incoming (non)-spam emails based solely on

performing *exact* and *partial* similarity matching among the phrases in an incoming email *E* and the ones in a user-identified spam email *S*, which determines how similar *E* and *S* are and subsequently establishes the likelihood of *E* being (non-)spam. Experimental results compiled by using known spam-detection corpuses on *SpamED* prove the effectiveness of *SpamED* with a 96% accuracy in correctly classifying incoming emails.

What's New? SpamED is computational inexpensive, since the word-correlation factors used for establishing the degrees of similarity

among emails are (i) pre-computed, (ii) are efficient in detecting (non)-spam emails, (iii) require little user intervention (for labelling incoming spam emails), (iv) minimize the number of misclassified legitimate emails which contain information users cannot afford to lose, (v) allow users the feasibility in expressing and modifying their preference on what constitutes spam and (vi) outperform existing spam-detection approaches in terms of accuracy.

Limitations: *SpamED* is designed for processing text-based emails. ■