
Editorial

Lorna Uden

Faculty of Computing Engineering and Sciences (FCES),
School of Computing,
Staffordshire University,
College Road, Stoke-on-Trent,
Staffordshire ST4 2DE, UK
Email: L.uden@staffs.ac.uk

Welcome to V12N2 of *IJWET*. There are four papers to this issue. The first paper is, 'Estimating similarity of rich internet pages using visual information' by Zhen Xu and James Miller. According to these authors, traditional text-based web page similarity measures fail to handle rich-information-embedded modern web pages. Current approaches regard web pages as either DOM trees or images. However, the former only focuses on the web page structure, while the latter ignores the inner connections among different web page features. Therefore, they are not suitable for modern web pages. In this paper, the authors investigate web page similarity from the visual perspective by using Gestalt laws of grouping. They introduce a block tree which contains both structural and visual information of web pages to represent a web page visually. This is done by retrieving visual information from the web page and interpreting the Gestalt laws of grouping to merge related content. A normalised Hausdorff distance is introduced to evaluate proximities; the CIE-Lab colour space and its colour difference are used to find the colour similarities; and the normalised compression distance is used to calculate image similarities.

A page similarity classification model is then built based on the block tree edit distance (B-TED). When calculating B-TED, each block tree node is labelled with all its visual features, and compares with the corresponding nodes. A visual similarity metric is proposed as the edit distance between two block trees. An experiment is preformed utilising a test set built from randomly crawling popular websites. To evaluate the correctness of B-TED as a metric for visual similarity, a ten-fold cross-validation is conducted. The overall precision, recall, and accuracy in the experiment are 90.27%, 100%, and 94.43%, respectively. This implies that B-TED is a promising metric for web page similarity evaluation, and provides satisfactory identification results. In spite of the contributions, limitations still exist for the proposed methodology. That is, the hierarchy of the block tree does not precisely reflect the visual layout when foreground text and background colours/images are separated. In the future work, a proper solution for this limitation will be useful.

The second paper is, 'A new project management approach for R&D software projects in the automotive industry – continuous V-model' by Sergiu Stefan Nicolaescu, Horatiu Constantin Palade, Danut Dumitru Dumitrascu and Claudiu Vasile Kifor. The authors of this paper describe project management and development methods that help managers in the automotive industry. The new approach called *continuous V-model – CVM*, is based on agile concepts. According to these authors, the model has been applied

on a real automotive R&D project and the resulted KPIs have been analysed and compared to the standard approach on a similar project. Further work must be carried out to validate the model.

The third paper is ‘Finding influential sources and breaking news in news media using graph analysis techniques’ by Iraklis Varlamis and Dimitrios Fasarakis Hilliard. In this paper, the authors have developed a solution that monitors news media, which is able to detect the underlying influence among news media companies and consequently provides useful knowledge concerning influential news sources, breaking news and online news media that frequently influence each other. Among the contributions of this work are: a new methodology for identifying and quantifying the implicit influence between news media, based on content similarity and a new method for the early detection of breaking news, defined as news of high potential interest to the mass media.

The methodology has been evaluated on a real dataset comprising articles collected from Greek news sites and blogs during two months period (September and October 2015). More empirical studies are needed to verify the model.

The final paper is ‘Analysing the effect of offline media on online conversion actions’ by Didier Nibbering, Flavius Frasincar and Damir Vandic. In this paper, the authors investigate how offline advertising by means of TV and radio influences the search engine advertisement that leads to users visiting a company marketing website (a conversion action). Their research is based on the search engine-driven conversion actions of the 2012 marketing campaign ‘Do Us A Flavour’ of the chips manufacturer lays, for which they have experimented with several prediction models: linear regression (linear model), support vector regression (nonlinear model), and six distributed lag models (linear autoregressive models). The results show that offline commercials positively influence the online marketing campaign. To verify the claim, more empirical data are needed.