

Editorial

STARTING from this issue, the Editorial Board and the owner of the journal, the Centro de Innovación y Desarrollo Tecnológico en Cómputo of the Instituto Politécnico Nacional has accepted my suggestion to change the complete title of the journal to *Polytechnic Open Library International Bulletin of Information Technology and Science*. The new title, with clear meaning in English but with the familiar acronym POLIBITS, will help positioning of our journal as an international open access publication in computer science and computer engineering, will open it to wider readership, and will help attracting better authors.

This issue of POLIBITS includes ten papers by authors from seven different countries: Chile, USA, India, Mexico, Colombia, Cuba, and Russia. The papers included in this issue are devoted to such topics as logical algorithms, data visualization, sentiment analysis, word sense disambiguation, business process modeling, ontology learning, software engineering, computer vision, recommender systems, information extraction, and question answering.

Adrián Jaramillo et al. from **Chile** in their paper “Comparing the Black Hole and the Soccer League Competition Algorithms Solving the Set Covering Problem” compare implementations of the Black Hole and Soccer League Competition algorithms in a statistical way, involving the use of non-parametric tests and supported by R statistical computing environment, considering regularity and consistency of their results. Both implementations are tested on the same benchmark sets.

Chaman Lal Sabharwal from **USA** in his paper “Exploration, Exploitation Phenomena and Regression Analysis: Propensity Metric, Anomaly Reduction, Dimensionality Reduction” applies visualization for comparison of the Ordinary linear Least Square approximation model has for the best fit regression for linear trend data with other existing methods. He has found that this technique is reliable and preferable to explain to the expert as well as non-expert. The empirical tests show the accuracy improvements over conventional methods.

Nachiappan Chockalingam from **India** in the paper “Simple and Effective Feature Based Sentiment Analysis on Product Reviews using Domain Specific Sentiment Scores” outlines a novel technique to extract features from a product’s reviews along with the corresponding sentiment expressed, using POS tagging and Dependency Parsing in conjunction. The use of these of these allows both the context and the parts of speech of a word to be employed in feature and corresponding opinion word detection. The opinion word is

given a sentiment polarity determined from a training set of positive and negative reviews. The method described in this paper is for large data sets, and requires no domain specific data for feature extraction.

Grigori Sidorov and Francisco Viveros-Jiménez from **Mexico** in their paper “One Sense per Discourse Heuristic for Improving Precision of WSD Methods based on Lexical Intersections with the Context” show how to increase precision of word sense disambiguation for some word classes of these simple methods to the level comparable with that of the most sophisticated ones. They observe that these methods usually disambiguate correctly those words that conform to the One Sense per Discourse heuristic (OSD words). They use Semcor and Wikipedia to find the OSD words and left non-OSD words without disambiguation, thus improving precision at the expense of recall. Their motivation for this situation (more precision, less recall) is, as follows. First, if one needs high-quality disambiguation and uses human evaluators, then one can reduce the cost by asking them to disambiguate only words that are really difficult for the algorithms; (2) in an automatic system, one can apply this method for disambiguation of the corresponding words, and use other more sophisticated method for disambiguation of other words, i.e., use different methods for disambiguation (meta-disambiguation). The authors experimented with the complete and simplified Lesk algorithms, the graph based algorithm, and the first sense heuristic. The precision of all algorithms increased and some algorithms reach the level of the inter annotator agreement.

Hugo Ordoñez et al. from **Colombia** in their paper “TrazasBP: A Framework for Business Process Models Discovery Based on Execution Cases” present TrazasBP, a framework for Business Process (BP) indexing and searching based on execution cases. It indexes BPs based on execution cases (traces) retrieved from log files. TrazasBP not only takes into account the textual information of BP elements, but also the causal dependence between these elements. Furthermore, due to its low computational cost, TrazasBP can be used as indexing mechanism in order to reduce the search space. Experimental evaluation shows promising values of graded precision, recall and F-measure when compared with results obtained from human search.

V. Sree Harish et al. from **India** in their paper “Unsupervised Domain Ontology Learning from Text” present an iterative focused web crawler for building corpus and an unsupervised framework for construction of Domain Ontology. The proposed framework consists of five phases: (1) Corpus Collection using Iterative Focused crawling with novel weighting measure; (2) Term Extraction using HITS algorithm;

(3) Taxonomic Relation Extraction using Hearst and Morpho-Syntactic Patterns; (4) Non Taxonomic relation extraction using association rule mining; and (5) Domain Ontology Building. Evaluation results show that proposed crawler outweighs traditional crawling techniques, domain terms showed higher precision when compared to statistical techniques and learnt ontology has rich knowledge representation.

D. Larrosa et al. from **Cuba** in their paper “GeCaP: Unit Testing Case Generation from Java Source Code” propose a tool that allows developers to automatically generate test cases for unit testing from Java source code. In this proposal, the basis path testing technique is used for the design of the test cases. The control flow graph is automatically generated from the source code being tested, in order to subsequently generate the independent paths. Finally, the combinations of test values that satisfy every linearly independent paths are generated. In the process of implementing this new tool, a case study was designed for the purpose of validation; metaheuristic algorithms were applied to generate test values and value combinations for each path. These combinations were compared against the ones obtained by other state-of-the-art algorithms. Since in this case study a 100% coverage of the independent paths is reached, the proposed tool exhibits competitive results with respect to the ones reported by tools proposed by other authors.

Antonio Alarcón-Paredes et al. from **Mexico** in their paper “Naïve Screw Nut Classifier Based on Hu's Moment Invariants and Minimum Distance” present an algorithm for classification of screw nuts by means of digital image processing. This work is part of a project where a production line was built, and is focused on the quality assessment section. The algorithm presented classifies among good and poor quality screw nuts passing by a conveyor belt, by computing Hu's moment invariants of its picture. Those moment invariants are the input of a minimum distance classifier, obtaining very competitive results compared with some other classification algorithms of the WEKA platform.

Liliya Volkova et al. from **Russia** in their paper “Recommender System for Tourist Itineraries Based on Aspects Extraction from Reviews Corpora” describe a

recommender system that takes a set of venue categories of user's interest into account to form a tourist itinerary throughout a city. The system is focused on user preferences in venue aspects. Techniques of such aspects extraction are developed in this paper, in particular from reviews corpora. User preferences are used to weigh aspects associated with particular sights and restaurants. These filtered venues along with time restrictions are subject to submit into the recommender system. A lightweight ontology is discussed which describes the domains of restaurants and sightseeing knowledge and allows venues comparative analysis to enhance the search for relevant venues. The system designed performs automated planning of tourist itineraries, flexible sights searching, and analysis of venues aspects extracted from reviews in Russian.

F. A. K. Hemant from **India** in the paper “Building an Information Extraction and Question Answering Model for Text Based on the Human Brain Process” showcase an information-extraction and question-answering model for text, which is based loosely on the human brain process. The ideology used is based on how humans perceive and interact with text, and the process of storing the text for future reference. Each word of each sentence is cross-referenced and linked with all available information and the answer is given based on matching information found. The model is basic, but the future applications and scope of improvement are also shown.

This issue of the journal will be useful to researchers, students, and practitioners working in the corresponding areas, as well as to public in general interested in advances in computer science, artificial intelligence, and computer engineering.

Alexander Gelbukh

Instituto Politécnico Nacional,
Mexico City, Mexico
Editor-in-Chief