Editorial

THIS issue of Polibits includes ten papers by authors from nine different countries: Brazil, Colombia, Cuba, France, Germany, India, Mexico, Portugal, and Spain. The majority of the papers included in this issue are devoted to such topics as software engineering, automated code generation, cloud computing, analysis of the web, and analysis and retrieval of images.

E. Torres Orue et al. from Cuba in their paper *Process for Unattended Execution of Test Components* describe a methodology to perform software tests. With the increasing complexity of modern software, the quality of the products crucially depends on the correctness and completeness of their testing. While much attention has been paid to totally automatic product verification techniques, in practice they are still far from being applicable, hence the importance of methodologies for manually testing products in a well-organized and systematic way. In this paper, the whole organization of the testing process is addressed, indicating the roles, phases, activities, and artifacts involved in the process. In addition, the authors present a tool that supports the process and allows unattended execution of tests.

G. Vargas-Solar et al. from France, Spain, and Brazil in their paper *Reliable Web Services Composition: An MDD Approach* present an approach for modeling and associating policies to service-based applications. This paper concerns the area of cloud computing, a relatively new software paradigm that receives in the recent years tremendous interest from the industry and user community. Within this paradigm, it is believed that soon we will not use personal computers, including laptops, tablets, etc.—instead, we will send requests to services located at powerful servers and large distributed networks, with the benefits of constant improvement, professional maintenance, reliable and redundant storage, and all this in a manner transparent for the end user. In this paper, the authors propose various extensions for the SOD-M model driven method of service-based computing that improve its reliability and interoperability.

H. Ordoñez et al. from Colombia in their paper *MultiSearchBP: Environment for Search and Clustering of Business Process Models* present an environment for search and clustering of business processes in a multimodal repository of business process models. Their three-level architecture consists of a presentation level, business level, and Storage level and supports operations typical for information retrieval settings. Internally their architecture uses a vector space model to represent the business process models stored in the repository. The authors show that their approach performs well in the experiments.

S. Schnitzer et al. from Germany in their paper *Combining Active and Ensemble Learning for Efficient Classification of Web Documents* address the problem of classifying web documents. With the enormous amount of documents available in Internet nowadays, compact presentation of thematic groups of documents to the user, as well as investigation of the structure of large collections of documents is required for effective access to the information contained in those documents and its adequate use. These processes include classification and grouping of the documents. Usual methods employed for automatic classification of texts involve large text corpora manually marked up by human experts. Development of such corpora is a very slow and expensive process. The authors show how active learning can be employed to minimize human effort by only requiring human intervention when the algorithm cannot reliably perform classification basing on the strategy learnt from its previous interaction with human annotators.

S. Pérez Lovelle et al. from Cuba in their paper *A Proposal to Incorporate More Semantics from Models into Generated Code* consider the problem of reflecting the semantics of software models the code automatically generated from them. Automatic generation of code greatly simplifies and speeds up software development process, as well as reduces the probability of software coding errors. Software development process supported by automatic generation allows the developers to concentrate on important strategic issues: what to develop and not how to code it. The developers' ideas of what is to be developed are specified in formal models, usually expressed by Unified Modeling Language (UML). However, it has been noted that some parts of the semantics of the UML models is lost when automatically converting these models into working code. The authors address the issue of how the semantics of such models is reflected in the code generated by the AndroMDA tool and indicate how it can be improved.

N. Das et al. from India and Portugal in their paper *Comparison of Different Graph Distance Metrics for Semantic Text Based Classification* continue the discussion of text classification, whose importance has just been discussed. They consider semantic approach to the task: documents that are devoted to semantically similar topics are to be clustered together. The clustering process is based on measuring similarity between texts, in this case based on their semantics. There are various commonly used semantic representations, most of them being either based on graphs or equivalent to graphs. However, comparing graphs is computationally expensive. In this paper, the authors show how to reduce the computational complexity of comparing the semantic

representations of texts by using shorter summaries of the texts instead of their full text, which makes the semantic representation simpler and the involved graphs smaller. The authors compare five different graph distance measures.

O. Rodríguez Zalapa et al. from Mexico in their paper *Distance Measurement System using Images to Determine the Position of a Sphere using the XBOX Kinect Sensor* present a method to measure the distance from an image of an object to a given reference point in the same image. This is an important task in computer vision domain. Computer vision has many practical applications, from monitoring of traffic or security to control or autonomous robots and military application. In particular, the existence of truly independent autonomous robots, such as home-helpers or automatic cars, is only possible with high-quality algorithms for visual orientation and, in particular, measuring distances between objects in the images. The authors show that their technique has better accuracy than instruments specifically designed for distance measurements.

V. M. Alonso-Rorís et al. from Spain in their paper *Information Extraction in Semantic, Highly-Structured, and Semi-Structured Web Sources* consider the task of extracting formal and structured information from open web documents, which are very heterogeneous in their nature, with wide variations from totally unstructured texts, images, sounds, and videos to highly structured databases. Better understanding of the nature and relationships of such information sources, as well as of the information they contain, leads to important applications for end users. In this paper, the authors show how the information automatically extracted from open web sources can be employed for two significantly different practical applications: in a recommender system for educational resources on the one hand, and in interactive digital TV applications on the other hand.

L. Flores-Pulido et al. from Mexico in their paper *Computing Polynomial Segmentation through Radial Surface Representation* address the problem of visual information retrieval. While traditional information retrieval operates with texts, in this case both the query and the objects to be retrieved are images. Obviously, this task requires detailed and sometimes sophisticated image-processing techniques, which are the object of the research in this paper. The authors address the problem of constructing appropriate models for representing images in such a way that facilitates their retrieval. The authors show that a modification of the General Principal Component Analysis procedure and other methods based on mathematical operations on the image data lead to high level of performance.

C. M. Zapata Jaramillo et al. from Colombia in their paper *Improving the consistency between textual and graphical syntax of the language of Semat* analize the formal language of the Software Engineering Method and Theory from the area of software engineering. This language has two different representations: as text and as graphical representation similar to block diagrams. The authors observed that the correspondence between these two forms of expression of the same information about software systems can sometimes not be direct, which leads to inconsistencies between these two forms of representation of the language. The authors propose modifications of the textual component of this formal language in order to achieve a more consistent relationship between the textual and graphical forms of the syntax of the core elements of formal language Semat.

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

Alexander Gelbukh
Editor in Chief