

Editorial

THIS ISSUE of the Polytechnic Open Library International Bulletin of Information Technology and Science (POLIBITS) includes ten papers by authors from six different countries: France, Germany, Ireland, Mexico, Spain, and USA. The papers included in this issue are devoted to such topics as machine translation, natural language processing, information retrieval, image processing, and classification algorithms for sensor applications in the context of the Internet of Things.

Christopher G. Harris from **USA** in his paper “A Framework to Build Quality into Non-Expert Translations” proposes a framework for obtaining high-quality translation from non-expert crowdworkers, by incorporating intermediate mechanisms such as ranking and editing in addition to translation. He demonstrates the effectiveness of the proposed framework in terms of translation quality, time, and cost. The proposed framework is robust against spammers, verifiable at different steps, and consistent with little variance in quality. The proposed framework is most appropriate when it is necessary to maximize translation quality and minimize cost.

César Marrades Cortés et al. from **Ireland** in their paper “Sports Video Anonymisation and Accuracy Prediction” suggests methods for preserving anonymity of persons in video recordings such as recording of school sport examinations. For the task, the authors use machine learning predictive models to investigate the factors that affect the anonymisation program’s performance on sports videos. On one hundred video inputs, they achieve an accuracy of 94% and a specificity of 95.2%.

Angélica Hernández Rayas et al. from **Mexico** in their paper “Marching Cubes Algorithm for Transforming Images” presents a modification of the Marching Cubes algorithm to make a three-dimensional representation from surface image discretized on volumetric pixels. For this, they connect each component continuously regardless of the values of the image surface. The modification aims to achieve this effect optimizing computational resources. The obtained results show high performance of the proposed modification.

Yuming Zhai et al. from **France** in their paper “Towards Recognizing Phrase Translation Processes: Experiments on English-French” study the processes that human translators, consciously or not, resort to in their translation activity apart from the literal translation. Such processes can be idiom equivalence, generalization, particularization, semantic modulation, etc. Several typologies exist to characterize such translation processes. The authors automatically classify these fine-grained translation processes. Their results show that they can distinguish non-literal translation from literal translation

with an accuracy of 87.09%, and 55.20% for classifying among five non-literal translation processes. This demonstrates that it is possible to classify automatically translation processes even with a small amount of annotated examples.

David Muñoz et al. from **Ireland** and **Mexico** in their paper “Named Entity Recognition Based on a Graph Structure” propose a graph structure for storage and enrichment of named entities that makes use of synonyms and domain-specific ontologies in the area of computing. They experimentally measure the performance of the proposed structure is measured and compared it with other NER classifiers. This work is useful for various areas of natural language processing, such as text mining and information retrieval.

Mónica Villaverde et al. from **Spain** in their paper “A Comparison of Adaptive and Non-Adaptive Ensemble Methods for Classification Systems” analyze work four different approaches to aggregation the estimations given by the ensemble of sensors in order to obtain the final classification. A number of classifiers are analyzed: ANN, plurality majority, basic weighted majority and stochastic weighted majority. They compare the majority-voting algorithm and the artificial neural network against two proposed adaptive algorithms based on weighted majority and without previous training. In this comparison, the authors not only take into account the accuracy of each algorithm but also adaptation. The experiments show that the artificial neural network is the most accurate proposal, whereas the most innovate proposed stochastic weighted voting is the most adaptive one.

Olga Kolesnikova and **Alexander Gelbukh** from **Mexico** in their paper “Dictionary and Corpus-Based Study of Lexical Functions in Spanish” study semantic and contextual characteristics of four types of verb-noun collocations in Spanish, corresponding to a different so-called lexical function. They explain how the typology of lexical functions can be viewed as a consistent way to classify collocations by their semantic and syntactic patterns. They examine how different lexical functions as well as free word combinations can be identified automatically by supervised machine learning methods. The authors show that contextual characteristics are not powerful enough to discriminate among subtle semantic differences between lexical functions.

Sannikumar Patel et al. from **Ireland** in their paper “Word Embeddings and Length Normalization for Document Ranking” address the problem of the document length bias in information retrieval techniques based on distributed word

representations. They argue for that document length normalization is useful to address the length bias problem while using embedding-based ranking. They present experimental evidence for their claim. The proposed method of length normalization significantly improves the Mean Average Precision, by up to 47% over a baseline that uses a simple embeddings-based algorithm.

Rafael Gallardo García et al. from **Mexico** in their paper “Facial Recognition using Convolutional Neural Networks and Supervised Few-Shot Learning” present a feature-based face recognition method, consisting of two separated processes. First, a pre-trained CNN-based face detector extracts the locations and features of the faces found in the input images. Then the output of this pre-trained face detector is used to train the models for the classifiers that are used to find unknown faces in new images. The authors show that their method achieves high accuracy on datasets with several individuals but few training samples.

Björn Buchhold and **Jörg Dallmeyer** from **Germany** in their paper “Zero-Shot Learning for Topic Detection in News Articles” present a method to detect topics in news articles and represent them by a descriptive document. The authors use a

neural network that operates on two kinds of inputs: the full texts of the descriptive documents passed through the same recurrent encoder network, and a proprietary NLP pipeline and knowledge base used to recognize named entities and significant keywords. The authors evaluate and compare several model configurations on two datasets, a large one automatically created from Wikipedia and a smaller one created manually.

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,
Oscar Camacho
Instituto Politécnico Nacional,
Mexico City, Mexico
Editors-in-Chief