
Editorial: Special issue on text analysis in scientific communication for the performance measurement and metrics

We are pleased to release the following three papers as part of the special issue on text analysis in scientific communication. The call for papers attracted several research submissions, and selecting the quality papers from all the submissions was strenuous. This first part now publishes three research papers; other papers are in the queue for later issues.

Quality product descriptions in e-commerce platforms are challenging as they are time-consuming, and perfect descriptions are labor-intensive and time-consuming. Realizing this difficulty, the authors *Fouzi Harrag, Ouissem Touameur and Hamza Behilil*, in the paper, “**Hybrid NLP model for automating fashion product descriptions: integrating transformers and word embeddings**,” used a transformer model for accurate product descriptions. They used natural language data-driven techniques to validate their model. The sophisticated model they deployed used GPT-Neo, a transformer model, with the word-embedding model word2vec. This work marks a significant growth in the word-embedding systems.

Social media-based prediction uses several models to predict preferences, relations, sentiments and various assessments. The predictive model for traits uses several context features. In the paper “**Leveraging AraBERT for COVID-19 event monitoring on Arabic Twitter**,” the authors *Fouzi Harrag, Ouissem Touameur and Maroua Zermani* introduced a model based on social media, particularly tweets, to predict pandemic-like diseases. They proposed a deep learning-based approach utilizing *Arabert*, which exploited the morphological richness and context of Arabic NLP. The outcome is generating a reliable framework that captures context-rich entities and events.

Besides author-generated key terms, many databases now introduce machine-generated key terms to supplement the author’s key terms and enhance the concept representation of scientific texts. However, the databases use several models and techniques that produce different sets of terms, leading to consistency between the databases regarding text representation by key terms. In their paper, *Solanki Gupta and Vivek Kumar Singh*, “**Exploring distributional characteristics and similarities of scholarly keywords: a comparative study of Web of Science Keywords Plus and Dimensions Concepts**,” measured the similarity between the key terms used by Web of Science Keywords Plus and Dimensions Concepts. They used the rank frequency distribution of terms and goodness-of-fit measures to assess distributional properties. The similarity among ‘*highly frequent*’ terms is measured by Jaccard similarity. They found wide variations between the two databases in both distributional forms and similarities due to the implementation of different algorithms. This finding raises an issue of the reliability of the databases in content expression, and future research may work to find the most accurate model.

We thank *Frankie Wilson and Holt Zaugg*, the editors of Performance Measurement & Metrics, for accepting the special issue proposal, which resulted in a few scholarly papers being published. The authors took pains to generate valuable research, and the reviewers supported us with highly objective reviews.

We hope the readers find these papers interesting and use them for further research.

Pit Pichappan

