
Document Engineering Special Issue

Introduction by Steve Simske,
Associate Editor, Colorado State University

Welcome to the Journal of Web Engineering (JWE) special issue on document engineering. Thanks to River Publishers, and especially Karen Donnison for managing the issue and Nicki Dennis for encouraging the topic, for making this possible. This special issue is focused on document engineering, a topic which garners its own symposium each year (the **ACM DocEng Symposium**) and features the engineering of content for consumption in a wide variety of media and multimedia, ranging from security labels to dynamic web sites. Obviously, engineering documents for the web is an important element of this symposium, and so a partnership between this symposium and JWE was a natural choice. Authors from the past year of the conference and their colleagues were encouraged to submit research works for review by JWE editors and reviewers, and three papers were down-selected for this special issue.

The first paper, “A Comparative Analysis of Sentence Embedding Techniques for Document Ranking” by Gupta et al., provides a comparative analysis of six pre-trained sentence embedding techniques to identify the best model for document ranking in information retrieval (IR) systems. The Universal Sentence Encoder and Sentence BERT approaches outperform other techniques on all four datasets examined. Importantly, the approach can be adopted to determine the best approaches for many other types of document ranking or query reformulation systems. This paper thus addresses the document engineering of content search and retrieval.

The second paper, “A Semantic Similarity Measure for Scholarly Document based on the Study of n-gram” by Tchanchou Samen, addresses similarity measures of scholarly documents. A semantic similarity metric is introduced which takes advantage of metadata associated with the articles. Human expert ground truthing is used to validate the accuracy of the approach. This paper thus addresses the engineering of document reading

sequences; for example, suggesting the reading order of a large set of documents for curriculum development or self-training.

The third paper, “Optimal Trained Bi-Long Short Term Memory for Aspect based Sentiment Analysis with Weighted Aspect Extraction” by Archana Nagelli, introduces a novel aspect-based sentiment analysis that includes preprocessing, aspect sentiment extraction, and classification. The classification is based on the Optimized Bi-LSTM with its weights tuned by a novel Opposition Learning Cat and Mouse-Based Optimization (OLCMBO) algorithm. This approach is shown to be competitive to existing approaches using metrics such as F1-measure, specificity, accuracy, sensitivity, and precision. The important document engineering topic of determining content polarity and sentiment is nicely addressed by this paper.

Overall, these three excellent papers will give the reader an in-depth but digestible introduction to the large, complex field of document engineering. Enjoy the read!

Kind regards,

Steve Simske