An empirical evaluation of document embeddings and similarity metrics for scientific articles

Gomez, J.

The comparison of documents such as articles or patents search, bibliography recommendations systems, visualization of document collections, etc., has a wide range of applications in several fields. One of the key tasks that such problems have in common is the evaluation of a similarity metric. Many such metrics have been proposed in the literature. Lately, deep learning techniques have gained a lot of popularity. However, it is difficult to analyze how those metrics perform against each other. In this paper, we present a systematic empirical evaluation of several of the most popular similarity metrics when applied to research articles. We analyze the results of those metrics in two ways, with a synthetic test that uses scientific papers and Ph.D. theses, and in a real-world scenario where we evaluate their ability to cluster papers from different areas of research.

http://dx.doi.org/https://doi.org/10.3390/app12115664