| | | doc_1 | | doc_2 | | decision | id |
|---|---|---|---|---|---|---|---|
| cases | authors | <ul><li>Bekamiri, Hamid</li><li>Hain, Daniel S.</li><li>Jurowetzki, Roman</li></ul> | authors | <ul><li>Hamid Bekamiri</li><li>Daniel S. Hain</li><li>Roman Jurowetzki</li></ul> | | DUPLICATES | 115 |
| | | | title | A Survey on Sentence Embedding Models Performance for Patent Analysis | | | |
| | title | A Survey on Sentence Embedding Models Performance for Patent Analysis | publication_date | 2022-08-05 00:00:00 | | | |
| | publication_date | 2022-04-28 00:00:00 | source | SupportedSources.INTERNET_ARCHIVE | | | |
| | source | SupportedSources.CORE | journal | | | | |
| | journal | | volume | | | | |
| | volume | | doi | | | | |
| | doi | None | urls | <ul><li>https://web.archive.org/web/20220810000148/https://arxiv.org/ftp/arxiv/papers/2206/2206.02690.pdf</li></ul> | | | |
| | urls | <ul><li>https://core.ac.uk/download/534432190.pdf</li></ul> | id | id-3480835608685434461 | | | |
| | id | id1452323190587015960 | abstract | Patent data is an important source of knowledge for innovation research, while the technological similarity between pairs of patents is a key enabling indicator for patent analysis. Recently researchers have been using patent vector space models based on different NLP embeddings models to calculate the technological similarity between pairs of patents to help better understand innovations, patent landscaping, technology mapping, and patent quality evaluation. More often than not, Text Embedding is a vital precursor to patent analysis tasks. A pertinent question then arises: How should we measure and evaluate the accuracy of these embeddings? To the best of our knowledge, there is no comprehensive survey that builds a clear delineation of embedding models' performance for calculating patent similarity indicators. Therefore, in this study, we provide an overview of the accuracy of these algorithms based on patent classification performance and propose a standard library and dataset for assessing the accuracy of embeddings models based on PatentSBERTa approach. In a detailed discussion, we report the performance of the top 3 algorithms at section, class, and subclass levels. The results based on the first claim of patents show that PatentSBERTa, Bert-for-patents, and TF-IDF Weighted Word Embeddings have the best accuracy for computing sentence embeddings at the subclass level. According to the first results, the performance of the models in different classes varies, which shows researchers in patent analysis can utilize the results of this study to choose the best proper model based on the specific section of patent data they used. | | | |
| | abstract | None | | | | | |
| | versions | | versions | | | | |