| cases | doc_1 | | doc_2 | | decision | id |
|---|---|---|---|---|---|---|
| | **authors** | • Malte Ostendorff<br>• Terry Ruas<br>• M. Schubotz<br>• Georg Rehm<br>• Bela Gipp | **authors** | • Malte Ostendorff<br>• Terry Ruas<br>• Moritz Schubotz<br>• Georg Rehm<br>• Bela Gipp | DUPLICATES | 298 |
| | **title** | Pairwise Multi-Class Document Classification for Semantic Relations between Wikipedia Articles | **title** | Pairwise Multi-Class Document Classification for Semantic Relations between Wikipedia Articles | | |
| | **publication_date** | 2020-03-22 00:00:00 | **publication_date** | 2020-03-22 12:52:56+00:00 | | |
| | **source** | SupportedSources.SEMANTIC_SCHOLAR | **source** | SupportedSources.ARXIV | | |
| | **journal** | | **journal** | None | | |
| | **volume** | | **volume** | | | |
| | **doi** | 10.1145/3383583.3398525 | **doi** | | | |
| | **urls** | • https://www.semanticscholar.org/paper/8e74106415626a21bb909ff489ef310625c769e1 | **urls** | • http://arxiv.org/pdf/2003.09881v1<br>• http://arxiv.org/abs/2003.09881v1<br>• http://arxiv.org/pdf/2003.09881v1 | | |
| | **id** | id-959959451690142782 | **id** | id768308826673893866 | | |
| | **abstract** | Many digital libraries recommend literature to their users considering the similarity between a query document and their repository. However, they often fail to distinguish what is the relationship that makes two documents alike. In this paper, we model the problem of finding the relationship between two documents as a pairwise document classification task. To find the semantic relation between documents, we apply a series of techniques, such as GloVe, Paragraph Vectors, BERT, and XLNet under different configurations (e.g., sequence length, vector concatenation scheme), including a Siamese architecture for the Transformer-based systems. We perform our experiments on a newly proposed dataset of 32,168 Wikipedia article pairs and Wikidata properties that define the semantic document relations. Our results show vanilla BERT as the best performing system with an F1-score of 0.93, which we manually examine to better understand its applicability to other domains. Our findings suggest that classifying semantic relations between documents is a solvable task and motivates the development of a recommender system based on the evaluated techniques. The discussions in this paper serve as first steps in the exploration of documents through SPARQL-like queries such that one could find documents that are similar in one aspect but dissimilar in another. | **abstract** | Many digital libraries recommend literature to their users considering the similarity between a query document and their repository. However, they often fail to distinguish what is the relationship that makes two documents alike. In this paper, we model the problem of finding the relationship between two documents as a pairwise document classification task. To find the semantic relation between documents, we apply a series of techniques, such as GloVe, Paragraph-Vectors, BERT, and XLNet under different configurations (e.g., sequence length, vector concatenation scheme), including a Siamese architecture for the Transformer-based systems. We perform our experiments on a newly proposed dataset of 32,168 Wikipedia article pairs and Wikidata properties that define the semantic document relations. Our results show vanilla BERT as the best performing system with an F1-score of 0.93, which we manually examine to better understand its applicability to other domains. Our findings suggest that classifying semantic relations between documents is a solvable task and motivates the development of recommender systems based on the evaluated techniques. The discussions in this paper serve as first steps in the exploration of documents through SPARQL-like queries such that one could find documents that are similar in one aspect but dissimilar in another. | | |
| | **versions** | | **versions** | | | |