	doc_1	doc_2	decision	id
cases	<table><tr><td>authors</td><td>• Yuki Arase • Jun'ichi Tsujii</td></tr><tr><td>title</td><td>Transfer Fine-Tuning: A BERT Case Study</td></tr><tr><td>publication_date</td><td>2019-09-02 00:00:00</td></tr><tr><td>source</td><td>SupportedSources.OPENALEX</td></tr><tr><td>journal</td><td>arXiv (Cornell University)</td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td>10.48550/arxiv.1909.00931</td></tr><tr><td>urls</td><td>• https://openalex.org/W4288122335 • https://doi.org/10.48550/arxiv.1909.00931 • http://arxiv.org/pdf/1909.00931</td></tr><tr><td>id</td><td>id-2058114509324601361</td></tr><tr><td>abstract</td><td></td></tr><tr><td>versions</td><td></td></tr></table>	<table><tr><td>authors</td><td>• Yuki Arase • Junichi Tsujii</td></tr><tr><td>title</td><td>Transfer Fine-Tuning: A BERT Case Study</td></tr><tr><td>publication_date</td><td>2019-09-03 00:00:00</td></tr><tr><td>source</td><td>SupportedSources.INTERNET_ARCHIVE</td></tr><tr><td>journal</td><td></td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td></td></tr><tr><td>urls</td><td>• https://web.archive.org/web/20200924234559/https://arxiv.org/pdf/1909.00931v1.pdf</td></tr><tr><td>id</td><td>id166437410020634166</td></tr><tr><td>abstract</td><td>A semantic equivalence assessment is defined as a task that assesses semantic equivalence in a sentence pair by binary judgment (i.e., paraphrase identification) or grading (i.e., semantic textual similarity measurement). It constitutes a set of tasks crucial for research on natural language understanding. Recently, BERT realized a breakthrough in sentence representation learning (Devlin et al., 2019), which is broadly transferable to various NLP tasks. While BERT's performance improves by increasing its model size, the required computational power is an obstacle preventing practical applications from adopting the technology. Herein, we propose to inject phrasal paraphrase relations into BERT in order to generate suitable representations for semantic equivalence assessment instead of increasing the model size. Experiments on standard natural language understanding tasks confirm that our method effectively improves a smaller BERT model while maintaining the model size. The generated model exhibits superior performance compared to a larger BERT model on semantic equivalence assessment tasks. Furthermore, it achieves larger performance gains on tasks with limited training datasets for fine-tuning, which is a property desirable for transfer learning.</td></tr><tr><td>versions</td><td></td></tr></table>	DUPLICATES	336