

cases

doc_1		doc_2		decision	id
authors	<ul style="list-style-type: none"><li>Byunghyun Ban</li></ul>	authors	<ul style="list-style-type: none"><li>Ban, Byunghyun</li></ul>	DUPLICATES	228
title	A Survey on Awesome Korean NLP Datasets	title	A Survey on Awesome Korean NLP Datasets		
publication_date	2021-12-06 00:00:00	publication_date	2021-12-06 00:00:00		
source	SupportedSources.INTERNET_ARCHIVE	source	SupportedSources.CORE		
journal		journal			
volume		volume			
doi		doi	10.1109/ictc55196.2022.9952930		
urls	<ul style="list-style-type: none"><li>https://web.archive.org/web/20211208175436/https://arxiv.org/ftp/arxiv/papers/2112/2112.01624.pdf</li></ul>	urls	<ul style="list-style-type: none"><li>http://arxiv.org/abs/2112.01624</li></ul>		
id	id3750724476747850489	id	id7179234807097820350		
abstract	English based datasets are commonly available from Kaggle, GitHub, or recently published papers. Although benchmark tests with English datasets are sufficient to show off the performances of new models and methods, still a researcher need to train and validate the models on Korean based datasets to produce a technology or product, suitable for Korean processing. This paper introduces 15 popular Korean based NLP datasets with summarized details such as volume, license, repositories, and other research results inspired by the datasets. Also, I provide high-resolution instructions with sample or statistics of datasets. The main characteristics of datasets are presented on a single table to provide a rapid summarization of datasets for researchers.	abstract	English based datasets are commonly available from Kaggle, GitHub, or recently published papers. Although benchmark tests with English datasets are sufficient to show off the performances of new models and methods, still a researcher need to train and validate the models on Korean based datasets to produce a technology or product, suitable for Korean processing. This paper introduces 15 popular Korean based NLP datasets with summarized details such as volume, license, repositories, and other research results inspired by the datasets. Also, I provide high-resolution instructions with sample or statistics of datasets. The main characteristics of datasets are presented on a single table to provide a rapid summarization of datasets for researchers.Comment: 11 pages, 1 horizontal page for large tabl		
versions		versions			