cases	doc_1		doc_2		decision	id
	authors	• Yanshan Wang • N. Afzal • S. Fu • Liwei Wang • F. Shen • M. Rastegar-Mojarad • Hongfang Liu	authors	• Yanshan Wang • Naveed Afzal • Sunyang Fu • Liwei Wang • Feichen Shen • Majid Rastegar-Mojarad • Hongfang Liu	DUPLICATES	361
	title	MedSTS: a resource for clinical semantic textual similarity	title	MedSTS: A Resource for Clinical Semantic Textual Similarity		
	publication_date	2018-08-28 00:00:00	publication_date	2018-08-28 16:43:19+00:00		
	source	SupportedSources.SEMANTIC_SCHOLAR	source	SupportedSources.ARXIV		
	journal	Language Resources and Evaluation	journal	None		
	volume	54	volume			
	doi	10.1007/s10579-018-9431-1	doi			
	urls	• https://www.semanticscholar.org/paper/7304afb491930d520e194f4ca8b91a9652f0e658	urls	• http://arxiv.org/pdf/1808.09397v1 • http://arxiv.org/abs/1808.09397v1 • http://arxiv.org/pdf/1808.09397v1		
	id	id-1491436845899780656	id	id-7452776468122172646		
	abstract	None	abstract	The wide adoption of electronic health records (EHRs) has enabled a wide range of applications leveraging EHR data. However, the meaningful use of EHR data largely depends on our ability to efficiently extract and consolidate information embedded in clinical text where natural language processing (NLP) techniques are essential. Semantic textual similarity (STS) that measures the semantic similarity between text snippets plays a significant role in many NLP applications. In the general NLP domain, STS shared tasks have made available a huge collection of text snippet pairs with manual annotations in various domains. In the clinical domain, STS can enable us to detect and eliminate redundant information that may lead to a reduction in cognitive burden and an improvement in the clinical decision-making process. This paper elaborates our efforts to assemble a resource for STS in the medical domain, MedSTS. It consists of a total of 174,629 sentence pairs gathered from a clinical corpus at Mayo Clinic. A subset of MedSTS (MedSTS_ann) containing 1,068 sentence pairs was annotated by two medical experts with semantic similarity scores of 0-5 (low to high similarity). We further analyzed the medical concepts in the MedSTS corpus, and tested four STS systems on the MedSTS_ann corpus. In the future, we will organize a shared task by releasing the MedSTS_ann corpus to motivate the community to tackle the real world clinical problems.		
	versions		versions			