

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none">Michael ConoverMatthew HayesScott BlackburnPete SkomorochSam Shah	authors	<ul style="list-style-type: none">Michael D. ConoverMatthew HayesScott BlackburnPete SkomorochSam Shah	DUPLICATES	363
	title	Pangloss: Fast Entity Linking in Noisy Text Environments	title	Pangloss: Fast Entity Linking in Noisy Text Environments		
	publication_date	2018-07-16 18:04:08+00:00	publication_date	2018-07-16 00:00:00		
	source	SupportedSources.ARXIV	source	SupportedSources.SEMANTIC_SCHOLAR		
	journal	None	journal			
	volume		volume			
	doi		doi	10.1145/3219819.3219899		
	urls	<ul style="list-style-type: none">http://arxiv.org/pdf/1807.06036v1http://arxiv.org/abs/1807.06036v1http://arxiv.org/pdf/1807.06036v1	urls	<ul style="list-style-type: none">https://www.semanticscholar.org/paper/4909eab8e8a41754d6804a11175a2152c7330329		
	id	id1654326692639457280	id	id-2837365866805975624		
	abstract	Entity linking is the task of mapping potentially ambiguous terms in text to their constituent entities in a knowledge base like Wikipedia. This is useful for organizing content, extracting structured data from textual documents, and in machine learning relevance applications like semantic search, knowledge graph construction, and question answering. Traditionally, this work has focused on text that has been well-formed, like news articles, but in common real world datasets such as messaging, resumes, or short-form social media, non-grammatical, loosely-structured text adds a new dimension to this problem. This paper presents Pangloss, a production system for entity disambiguation on noisy text. Pangloss combines a probabilistic linear-time key phrase identification algorithm with a semantic similarity engine based on context-dependent document embeddings to achieve better than state-of-the-art results (>5% in F1) compared to other research or commercially available systems. In addition, Pangloss leverages a local embedded database with a tiered architecture to house its statistics and metadata, which allows rapid disambiguation in streaming contexts and on-device disambiguation in low-memory environments such as mobile phones.	abstract	Entity linking is the task of mapping potentially ambiguous terms in text to their constituent entities in a knowledge base like Wikipedia. This is useful for organizing content, extracting structured data from textual documents, and in machine learning relevance applications like semantic search, knowledge graph construction, and question answering. Traditionally, this work has focused on text that has been well-formed, like news articles, but in common real world datasets such as messaging, resumes, or short-form social media, non-grammatical, loosely-structured text adds a new dimension to this problem. This paper presents Pangloss, a production system for entity disambiguation on noisy text. Pangloss combines a probabilistic linear-time key phrase identification algorithm with a semantic similarity engine based on context-dependent document embeddings to achieve better than state-of-the-art results (>5% in F1) compared to other research or commercially available systems. In addition, Pangloss leverages a local embedded database with a tiered architecture to house its statistics and metadata, which allows rapid disambiguation in streaming contexts and on-device disambiguation in low-memory environments such as mobile phones.		
	versions		versions			