

Configuration for Topic Model IR Anthology

Pre-Processing

-raw data:	19267 abstracts
-base form reduction:	Lemmatisation
-lowercasing:	yes
-stopword removal:	yes
-number removal:	yes
-punctuation removal:	yes
-Pruning:	minimum Term Frequency = 25
-POS Tag Restriction:	NOUN

→ Vocabulary Size: 2202 (*available as vocab_task34.RDS*)

Topic Model Setting:

used Library:	topicmodels::LDA
number of topics:	25
alpha:	0.05
beta:	0.1
iterations:	300
method:	Gibbs-Sampling
seed:	1234