

SST(Stanford Sentiment Treebank)

언어에서 감정의 구성 효과를 완벽하게 분석할 수 있는 완전히 레이블이 지정된 구문 분석 트리가 있는 말뭉치

말뭉치는 Pang and Lee(2005)가 소개한 데이터셋을 기반으로 하며 영화 리뷰에서 추출한 11,855개의 단일 문장으로 구성됩니다.

Stanford 파서로 파싱되었으며 해당 파스 트리에서 총 215,154개의 고유한 구문을 포함하며 각각 3명의 인간 심사관이 주석을 달았다.

각 문구는 부정적, 다소 부정적, 중립적, 다소 긍정적 또는 긍정적으로 표시됩니다. 5개 레이블이 모두 포함된 말뭉치는 SST-5 또는 SST 미세 입자라고 한다.

전체 문장에 대한 이진 분류 실험(부정적이거나 다소 부정적인 대 다소 긍정적이거나 중립적인 문장은 버려짐)은 데이터 세트를 SST-2 또는 SST 이진법으로 참조.

