



- \*\*32개 언어\*\* (7개 언어군) × \*\*100,000+ 인스턴스\*\*
- \*\*저자원 언어 포함:\*\* 하우사어(Hausa), 오로모어(Oromo), 요루바어(Yorùbá), 스와힐리어(Swahili) 등
- \*\*기준 한계 해결:\*\* 영어 중심/번역 의존에서 벗어나, 현지 커뮤니티와 협력하여 원본 데이터 수집
- \*\*출처 다양화:\*\* 소셜미디어 (Twitter, Reddit, YouTube), 연설, 뉴스, 문학 텍스트

7개 언어군(Language Families)의 전 세계적 분포

