

Source Documents



Text Extraction



Text Cleaning &
Normalization



Sentence Segmentation



Heuristic Sentence
Alignment



Parallel Spanish–Quechua
Corpus