Preprocessing Pipeline

Corpus → **Documents** → Sentencify + Tokenize → **Sentences** → Country, Number Marker

Only Sentences Having Both a Country and a Number

NER, POS Dependency Parse Chunking → **Load** → Apache Derby DB