

Collection Readers

Sentence Splitters

Tokenizers

Tests

...

Output  
Writer(s)

txtReader

RegEx  
(newline)

OpenNLP  
(en-sent.bin)

OpenNLP  
(en-token.bin,  
en-pos-max-ent.bin)

OffsetTokenizer  
(symbol)

OffsetTokenizer  
(whitespace)

ConceptMapper  
(symptoms)

XMI  
Writer