

Project Gutenberg raw textfiles



MapReduce



CSV table with word / document as
cols / rows, tf-idf as entries



R / Rattle



Results

Project Gutenberg catalog

