TermByDocumentMatrix

Abstract

The word count matrix (document-word) approach is often referred to as latent semantic indexing and document clustering (Of course, A word frequently present in all documents will not be useful for clustering; The length of all documents is not uniform so a lengthy document will have higher word counts). This example gives parallel implementation of the Matrix-creation (In the future, the matrix sparse decomposition technique).