

Mahout LDA

1. mahout seqdirectory -i /user/tmp2user/news_articles_v3 -o /user/tmp2user/news_articles_seq_files
2. mahout seq2sparse -nv -i /user/tmp2user/news_articles_seq_files -o /user/tmp2user/news_articles_vectors
3. mahout rowid -i /user/tmp2user/news_articles_vectors/tfidf-vectors -o /user/tmp2user/news_articles_matrix
4. mahout cvb0_local -i /user/tmp2user/news_articles_matrix/matrix -d /user/tmp2user/news_articles_vectors/dictionary.file-0 -a 0.5 -top 3 -do /user/tmp2usr/news_articles_lda -to /user/tmp2user/news_articles_lda_topics
5. mahout vectordump -i /user/tmp2user/news_articles_lda_topics --dictionary /user/tmp2user/news_articles_vectors/dictionary.file-0 --dictionaryType sequencefile --vectorSize 10 -sort /user/tmp2user/news_articles_lda_topics