

vocab size 1000; 1000 documents

