Notas:

O algoritmo produz várias saídas na pasta files, dentre elas um arquivo de probabilidade para os termos de cada classe e um arquivo contendo o resultado da classificação dos testes (classification.text), onde, ao final, está o erro global da classificação.

* O erro ainda está alto quando usadas todas as classes (~0,53). É preciso verificar se é possível reduzir esse erro ou se ele é esperado dada a proximidade de algumas classes.
* Antes de remover as 100 palavras mais freqüentes, a classificação errava muito mais (~0,83).
* O erro com as classes inclusas (rec.autos, talk.politics.mideast e talk.politics.guns), no momento, está em torno de 0,22.

Próximo passo: implementar algoritmo de seleção de tuplas para treino e para teste, conforme especificação do trabalho.