MC336 - PROJETO 2
Descobrir palavras comuns em vários textos

O arquivo de entrada (pelo stdin) contém:

- uma linha com palavras separadas por virgulas. Essas palavras são conhecidas como *stop words* que são palavras comuns que não servem para distinguir nenhum texto. Em portugues, as stop words seriam os pronomes, preposições, verbos de ligação, etc.
- uma linha em branco
- várias linhas de um texto. O texto pode conter pontuação e parentesis, que devem ser removidos. O texto tambem pode ter letras maiusculas e minusculas, que devem ser convertidas para minusculas. O texto não contém letras acentuadas.
- uma linha em branco
- várias linhas do segundo texto
- e assim por diante.

Escreva um programa em python que imprime as tres palavras (tirando as stop words) que aparecem em um maior número dos textos lidos. Se duas ou mais palavras aparecem no mesmo número de textos, imprima-as em ordem reversa a alfabética (isso torna o código mais simples!). Imprima uma palavra por linha, um branco, e o número de textos onde ela aparece. Finalmente, imprima `sim` se a união das tres palavras aparecem em todos os textos, e `nao`, se não for o caso.

Veja que eu não estou interessado na palavra mais frequente de cada texto, apenas nas palavras que aparecem num maior número de textos.

A motivação deste projeto é descobrir uma query que vai selecionar (pelo menos) todos os textos lidos. A última pergunta do projeto responde se um OU das tres palavras que aparecem no maior número de textos é suficiente para selecionar todos os textos. O problema real é um pouco mais complicado. Voce precisa de um conjunto de textos que voce não quer que sejam selecionados pela sua query e trabalha-se não só com palavras simples mas também com sequencias de 2 palavras.