

Faça o download dos arquivos Trabl.0.zip

Cada arquivo neste diretório possui a seguinte estrutura:

journals/cl/SantoNR90::Michele Di Santo::Libero Nigro::Wilma Russo::Programmer-Defined Control Abstractions in Modula-2.

Cada uma das linhas pode ser entendida como:

- Informação bibliográfica;
- autores separados por '::';
- Título da obra

paper-id::author1::author2::....::authorN::title

- 1) A tarefa é fazer o cálculo de quantas vezes cada termo (no título da obra) acontece por autor;
- 2) Por exemplo para o autor 'Alberto Pettorossi' os seguintes termos acontecem (considerando todos os documentos): program:3, transformation:2, transforming:2, using:2, programs:2, and logic:2.
- 3) Observe os seguintes itens:

O separador de campos é “::” e separador de autores é “::”;

Cada autor pode ter escrito múltiplas obras, que por sua vez podem estar em vários arquivos;

Existe uma lista de palavras que não serão consideradas. Utilize o arquivo stopword.py do exercício prático: exclua todas estas palavras para os autores;

Se possível faça a exclusão de pontuações também, de forma que a palavra logic e logic. sejam tratadas como se fossem uma única palavra;

- 4) Faça uma implementação de acordo com o material discutido nas aulas
- 5) Objetivo: Responder quais são as duas palavras que mais acontecem para os seguintes autores:

a- “Grzegorz Rozenberg”

b- “Philip S. Yu”