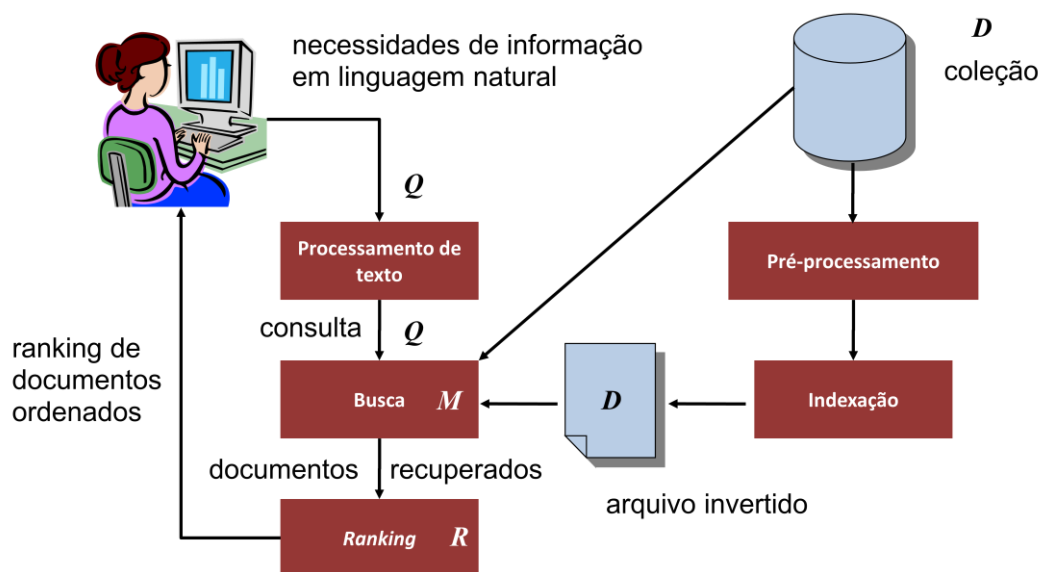


O objetivo deste trabalho é aplicar os conhecimentos adquiridos em aula para a implementação de um sistema de recuperação de informações baseado no modelo vetorial. A aplicação possui quatro objetivos principais que serão avaliados separadamente:

1. Montar um índice baseado em arquivo invertido contendo as frequências de cada termo em cada documento da coleção (50% da nota);
2. Representar cada documento no modelo vetorial (25% da nota);
3. Calcular a similaridade entre a consulta e os documentos que contém pelo menos uma das palavras (20% da nota);
4. Apresentar um ranking ordenado conforme a similaridade previamente calculada (5% da nota).

O código-fonte deve estar devidamente organizado e documentado. Cada TAD (consulta, documento, índice e ranking de documentos) e componente (pré-processamento, indexação, busca e ranking) deve ser implementado separadamente utilizando classes/módulos distintos.

Componentes de um Sistema de RI



Adaptado de Viviane Moreira – UFRGS (2007)

Fica a cargo do aluno a implementação de uma interface do usuário onde possa ser escolhida a coleção de documentos e realizadas as consultas.

O sistema pode ser desenvolvido em qualquer linguagem de programação e IDE. Podem ser utilizados quaisquer pacotes, APIs ou bibliotecas que implementem parte do problema.

O trabalho deve ser feito individualmente, duplas ou trios. Não serão aceitas quadras!

Todo o projeto, incluindo código-fonte, bibliotecas externas e executáveis devem ser entregues pelo Moodle no prazo definido e deve ser apresentado em aula pelo grupo.