

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**INSTITUTO DE INFORMÁTICA**  
**INF01124 – Classificação e Pesquisa de Dados - Semestre 2013-I**

**Instruções para o Trabalho Final**

---

Este trabalho deve ser realizado **em grupos de duas pessoas** ou, eventualmente, de forma individual (o motivo deve ser justificado).

**O QUE DEVE SER FEITO**

O trabalho consiste em desenvolver um indexador e analisador de Currículos Lattes de pesquisadores. A ideia geral é desenvolver um programa capaz de indexar os pesquisadores e suas publicações, catalogando-os arquivo(s), permitindo sua classificação, ordenação e consulta. Este trabalho envolve, portanto, manipular arquivos-texto, arquivos binários (índices e estruturas de persistência), classificação e ordenação.

Como requisitos, o usuário deverá ser capaz de usar o software para, pelo menos:

- Listar todos os pesquisadores cadastrados;
- Listar todos os artigos de um pesquisador, organizados por:
  - tipo (em periódico ou em conferência);
  - natureza (artigo completo, estendido ou resumo);
  - quantidade de coautores;
  - QUALIS (a ser disponibilizado pelos professores em um arquivo-texto CSV).

Pelo menos outras duas funcionalidades devem ser oferecidas pelos grupos, a sua escolha. Por exemplo, o programa poderia permitir a recuperação de todos os pesquisadores que possuem artigos de um determinado tipo, natureza ou QUALIS, mostrando o resultado de maneira ordenada. Poderia ainda listar a distribuição de todos os artigos por classes.

**A abordagem** consiste em desenvolver uma aplicação que interpreta o código XML obtido a partir de CVs Lattes HTML obtidos a partir do site do CNPq a partir do endereço:

<http://buscatextual.cnpq.br/buscatextual/>

Nesse caso, você pode (a) fazer uma aplicação que funcione como um navegador, acessando o servidor e simulando o protocolo HTTP, ou (b) pode salvar o resultado de uma navegação em um arquivo local (HTML ou XML) e interpretar o resultado usando as funções tradicionais da linguagem C (ou outra preferida). Para a manipulação dos arquivos HTML ou XML, é possível utilizar bibliotecas disponíveis na rede, desde que descritas e justificadas no trabalho.

Independente da abordagem escolhida, é importante que sua aplicação seja capaz de processar e manipular o conteúdo de um conjunto de *currículos Lattes* (no mínimo 50). Com base nessas informações (e em outras que considerar relevante), seu software deve ser capaz de armazenar em um arquivo local as informações coletadas/processadas. O armazenamento não pode ser feito em um banco de dados. Você deve implementar a estrutura de arquivo mais adequada para o seu software, escolhendo entre uma das estruturas estudadas em aula (p. ex., arquivo sequencial, arquivo sequencial indexado/ISAM, arquivo indexado, arquivo direto). De preferência, use arquivos binários, mas, dependendo da situação, se bem justificada, poderá usar arquivos textuais (a evitar).

**Até o dia 14 de Maio** (até as 23:55 horas do servidor da UFRGS) cada grupo deve **encaminhar pelo Moodle** (item “Enviar definição do trabalho final”) um **documento** definindo as **funcionalidades previstas** para a aplicação, os **membros da equipe** (nomes e números UFRGS das pessoas envolvidas), os **recursos necessários** e uma **previsão de atividades e seus respectivos tempos de desenvolvimento**. Atrasos implicarão em perda de nota.

Os grupos deverão **entregar um relatório final de trabalho em formato PDF** pelo Moodle (item “Enviar relatório do trabalho final”), **até o dia 25 de Junho** (até as 23:55 horas do servidor da UFRGS). Atrasos implicarão em perda de nota.

O **relatório de trabalho** deverá conter os **itens** descritos a seguir:

1. Identificação do trabalho e dos alunos;
2. Descrição do problema e da solução proposta (algoritmos, frameworks, módulos, classes, componentes, bibliotecas, etc. utilizados). Caso utilize algum algoritmo, framework, módulo, classe, componente, biblioteca, etc., de terceiros, sua respectiva fonte/origem deve ser citada;
3. Descrição da contribuição do trabalho, ou seja, diferencial em relação ao que já existe, tanto em termos de aplicação quanto em termos de código (no caso de reuso de código/componentes, descrever em detalhes como foi feito o reuso e onde está o diferencial – sua contribuição);
4. Guia ou manual do usuário da aplicação/solução desenvolvida, descrevendo características, exemplos de uso e resultados obtidos;
5. Considerações finais descrevendo quais técnicas ou tecnologias estudadas na disciplina foram empregadas e facilitaram ou dificultaram o desenvolvimento das soluções;
6. Código-fonte dos programas implementados em arquivo adicional (não incluir no relatório) haverá outra atividade para a sua submissão.

**Apresentação do trabalho ao professor e aos colegas**, explicando a aplicação implementada. Tais apresentações ocorrerão **por sorteio, nos dias especificados no cronograma**. As apresentações/explicações podem durar até 10 minutos.

Bom trabalho.