

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

Instruções para o Trabalho Final

Este trabalho deve ser realizado **em grupos de duas pessoas**. Apesar de não ser estimulado, em casos extremos por necessidade formalmente justificada, o trabalho poderá ser feito de forma individual (consulte o professor, se este for o seu caso).

O QUE DEVE SER FEITO:

O trabalho consiste em desenvolver um catalogador de Currículos Lattes de pesquisadores. O catálogo é um programa que permite indexar, classificar e consultar informações sobre os pesquisadores indexados. A ideia geral é desenvolver um programa capaz de indexar os pesquisadores, dados pessoais e publicações, catalogando-as em arquivos, permitindo a classificação, ordenação e consulta. Este trabalho envolve, portanto, manipular arquivos-texto, arquivos binários (índices e estruturas de persistência), classificação e ordenação. Além disso, a interface de usuário deverá ser de fácil utilização.

O Currículo Lattes de um pesquisador será armazenado em um arquivo XML (detalhes sobre este tipo de arquivo e como ele está organizado serão disponibilizados no Moodle da disciplina). O arquivo contém diversos dados sobre o pesquisador. O professor disponibilizará, no Moodle, um arquivo .zip contendo 50 currículos de professores do Instituto de Informática da UFRGS. Pelo menos esses deverão ser catalogados. Outros currículos podem ser obtidos no site do Currículo Lattes (<http://lattes.cnpq.br>).

Como requisitos de dados, seu programa deverá indexar e armazenar pelo menos os seguintes dados dos pesquisadores:

- Nome completo;
- Local e instituição de trabalho;
- Áreas ou Linhas de pesquisa;
- Lista de publicações em eventos;
- Lista de publicações em periódicos.

Como requisitos de consulta, seu programa deverá realizar pelo menos as seguintes operações:

- Listar todos os pesquisadores catalogados;
- Listar todos os pesquisadores catalogados cujo nome inicie-se por algum prefixo (p. ex., ana*);
- Listar todos os pesquisadores catalogados que pertençam a determinada área ou linha de pesquisa (indicada pelo usuário).

Independente do tipo de consulta realizada, os resultados deverão apresentar, no mínimo, os seguintes dados de cada pesquisador listado:

- o seu nome;
- a sua quantidade total de publicações;
- a sua quantidade de publicações em periódicos;
- a sua quantidade de publicações em eventos.

Além disso, o programa deverá permitir a ordenação dos resultados pelos seguintes critérios:

- ordem alfabética crescente de nome;
- ordem alfabética decrescente de nome;
- quantidade crescente ou decrescente de publicações (total de publicações)
- quantidade crescente ou decrescente de publicações em periódicos;
- quantidade crescente ou decrescente de publicações em eventos;

Pelo menos outras duas funcionalidades a sua escolha devem ser oferecidas. Por exemplo, o programa poderia permitir a recuperação de todos os pesquisadores que possuem artigos de um determinado tipo, natureza ou QUALIS. Poderia, ainda, listar quantidade de coautores.

Ainda em termos de requisitos, seu programa deverá evitar ao máximo o uso de memória para armazenar os dados. Toda estrutura de dados utilizada deve ser justificada em seu relatório final. Exige-se que os alunos utilizem pelo menos uma das seguintes estruturas para o armazenamento e a manipulação dos dados em arquivo (ou uma combinação delas):

- Arquivo serial indexado;
- Arquivo sequencial indexado;
- Árvore B+;
- Árvore TRIE;
- Arquivo Hash.

➤ **O armazenamento não pode ser feito em um sistema de gerenciamento de banco de dados!** De preferência, use arquivos binários, mas, dependendo da situação, se bem justificada, poderá usar arquivos textuais.

Em termos de linguagem de programação, você é livre para utilizar a linguagem que lhe for mais conveniente. Sugere-se utilizar alguma biblioteca de processamento de arquivos XML (p. ex., DOM, SAX) e que suporte o processamento de consultas XML (p. ex., XPath, XQuery). Toda biblioteca utilizada deve ser referenciada, citada e justificada em seu relatório final.

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).

ATÉ QUANDO FAZER:

Até o dia 06 de OUTUBRO (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 de matrícula das pessoas envolvidas), os **recursos necessários** e uma **previsão de atividades e seus respectivos tempos de desenvolvimento**. Atrasos implicarão em perda de nota (20% no 1º dia e 10% a cada dia subsequente).

Até o dia 20 de NOVEMBRO (até as 23:55 horas do servidor da UFRGS) os grupos deverão **entregar um relatório final de trabalho em formato PDF**. Atrasos implicarão em perda de nota (20% no 1º dia e 10% a cada dia subsequente). Esse relatório deve conter os itens descritos a seguir (ver “Sobre o relatório do trabalho, mais abaixo”).

Além do relatório, todos os códigos-fontes elaborados/utilizados devem ser encaminhados (em seu formato original – não colocá-los dentro do PDF). Para tanto junte o relatório e os arquivos-fonte em um único arquivo compactado (.zip ou .rar) e encaminhe esse arquivo resultante pelo Moodle (item “Enviar relatório do trabalho final”).

SOBRE O RELATÓRIO DO TRABALHO:

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;

Não esquecer de agregar os códigos-fonte (fora do PDF, em seus arquivos originais, conforme descrito anteriormente).

APRESENTAÇÃO DO TRABALHO:

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 20 minutos.

Conforme instruções presentes no plano de ensino da disciplina, todas as etapas do trabalho devem ser cumpridas para que a sua nota de trabalho seja contabilizada!

Bom trabalho!