

INF01006 - Projeto de Banco de Dados

Plano de ensino

Carlos A. Heuser

2010/1

1 Identificação

Nome do departamento: Informática Aplicada

Nome da atividade de ensino: INF01006 - Projeto de Banco de Dados

Curso de oferecimento: CIC – Ciência da Computação

Pré-requisito: Fundamentos de Banco de Dados

Etapa aconselhada no curso: 4^a

Corpo docente: Carlos A. Heuser

Créditos/carga horária: 4/4

2 Objetivos

A disciplina tem por objetivo permitir ao aluno aprofundar os conhecimentos de projeto de banco de dados que teve na disciplina introdutória de BD. Consta da execução de um projeto completo (especificação, projeto e construção) de uma aplicação de BD por grupos de alunos.

3 Conteúdo Programático

1. Definição da aplicação de BD. 2. Definição da metodologia. 3. Especificação de requisitos, modelagem de dados, especificação das transações. 4. Projeto e definição da plataforma de implementação.

4 Metodologia Adotada

A disciplina está montada ao redor de um projeto de porte a ser executado em grupos de no máximo dois alunos.

O semestre é dividido em duas etapas.

4.1 Etapa 1 – aulas expositivas

Na primeira parte, ocorre um conjunto de aulas expositivas tratando dos problemas que serão resolvidos no projeto (XML e similaridade). Serão propostos exercícios de fixação de conteúdo a ser resolvidos de forma individual.

4.2 Etapa 2 – confecção do projeto

Na segunda parte da disciplina, os alunos serão divididos em grupos de dois, cada grupo encarregado de um dos projetos da disciplina (ver detalhes dos projetos abaixo). A capacidade de atendimento do professor é de 20 grupos. Assim, se a turma estiver completa (40 alunos) não serão admitidos grupos de um único aluno.

Cada grupo de alunos executará um projeto. Um projeto consta de duas grandes etapas:

- Criação de um banco de dados que armazene um grande documento XML. Este banco de dados deve ser projetado de duas formas: (1) específica para a DTD do documento e (2) genérica, independente da DTD. Este banco de dados deve ser carregado e sobre ele devem ser executadas diferentes consultas XPath. O desempenho da carga e das consultas deve ser medido.
- Definição de tabelas auxiliares no banco de dados que permitam consultas por similaridade a um ou mais campos. Deve ser medida a qualidade e o desempenho destas consultas.

Mais especificamente, cada grupo seguirá o seguinte roteiro:

1. Definir a base de dados que vai utilizar entre aquelas disponibilizadas pelo professor (base contendo as obras de Shakespeare ou base de dados contendo um espelho do site DBLP).
2. Projeto de duas bases de dados relacionais diferentes, uma específica para a DTD escolhida e outra genérica. Serão disponibilizados alguns artigos mostrando estas alternativas, mas os alunos podem propor alternativas.
3. Definição de um conjunto de consultas XPath sobre o documento XML. Mapeamento destas consultas para SQL e execução das mesmas com medição de desempenho.

4. Projeto de tabelas auxiliares que permitem busca por similaridade.
5. Definição de um conjunto de consultas por similaridade, execução das mesmas e medição da qualidade obtida e do desempenho das consultas.
6. Apresentação dos experimentos para os colegas.
7. Redação de artigo curto (formato ACM, máximo de 6 páginas) relatando os resultados alcançados.

Nesta etapa, cada grupo terá um encontro semanal com o professor, em horário a ser sorteado dentro do horário da disciplina (caso tenhamos 20 grupos, este encontro terá 12 minutos – com menos alunos teremos encontros mais longos). A presença dos dois membros do grupo neste encontro é obrigatória, pois é a forma de avaliar o andamento dos alunos no projeto. O encontro serve para orientar os alunos em seu projeto.

5 Cronograma de Atividades

Aula#	Formato	Conteúdo
1	aula	Apresentação da disciplina - Aplicações típicas de XML e dados WEB
2	aula	XML - o modelo de dados e a DTD
3	aula	Entidades, APIS e namespaces
4	aula	XPath
5	aula	XQuery
6	aula	XML Schema
7	aula	Integração de esquemas e instâncias: funções de similaridade e avaliação de qualidade
8-9	orientação	Planejamento: definição da base de dados e dos artigos a usar
10-11	orientação	Projeto das bases de dados para armazenar XML
12-13	orientação	Definição da forma de implementar as consultas XPath
14-17	orientação	Implementação do armazenamento do XML
18-19	orientação	Projeto das tabelas para similaridade
20-21	orientação	Implementação das consultas por similaridade
22-25	orientação	Execução dos experimentos
26-27	orientação	Discussão dos resultados obtidos
28-29	apresentação	Grupos apresentam resultados
30	orientação	Revisão do artigo

6 Critérios de Avaliação

O conceito do aluno será obtido levando em conta os seguintes quesitos:

Peso	Quesito
10%	Resolução dos exercícios referentes às aulas expositivas.
50%	Qualidade de trabalho realizado demonstrado nas reuniões de orientação ao longo do semestre (qualidade do trabalho, prazos).
20%	Qualidade das apresentações.
20%	Qualidade do artigo apresentado no final da disciplina.

O prazo para recursos é de uma semana após a divulgação das notas.

6.1 Atividades de Recuperação

Pela natureza da disciplina, não há atividade de recuperação.

7 Bibliografia

7.1 Bibliografia referente a parte de XML

- Artigo com várias alternativas para armazenar XML. Artigo básico e referência para os demais.
Daniela Florescu - Donald Kossmann, *A Performance Evaluation of Alternative Mapping Schemes for Storing XML Data in a Relational Database*, Technical Report, INRIA, Maio 1999
- Popõe alternativas de armazenamento levando em conta a execução de consultas XPath (na realidade, de um precursor de XPath, chamado XML/QL).
Jayavel Shanmugasundaram et al. *Relational Databases for Querying XML Documents: Limitations and Opportunities* 25th VLDB Conference, Edinburgh, Scotland, 1999
<http://www.vldb.org/conf/1999/P31.pdf>
- Artigo que demonstra a tradução de XPath para SQL.
Li, Q. and Moon, B. 2001. *Indexing and Querying XML Data for Regular Path Expressions*. In Proceedings of the 27th international Conference on Very Large Data Bases (September 11 - 14, 2001).
<http://www.vldb.org/conf/2001/P361.pdf>
- Este artigo mostra um esquema para traduzir XQuery para SQL.
David DeHaan, David Toman, Mariano P. Consens, M. Tamer Özsu; A

Comprehensive XQuery to SQL Translation using Dynamic Interval Encoding. SIGMOD 2003, June 9-12, 2003, San Diego, CA.
<http://portal.acm.org/citation.cfm?id=872757.872832>

- Este artigo introduz o problema de armazenar documentos XML em que os elementos estão *ordenados*.
Igor Tatarinov; *Storing and Querying Ordered XML Using a Relational Database System*. ACM SIGMOD'2002, June 4-6, Madison, Wisconsin, USA.
<http://portal.acm.org/citation.cfm?id=564691.564715>

7.2 Bibliografia referente a parte de similaridade

- Luis Gravano, Panagiotis G. Ipeirotis, H. V. Jagadish, Nick Koudas, S. Muthukrishnan, Divesh Srivastava: *Approximate String Joins in a Database (Almost) for Free*. VLDB 2001: 491-500
<http://www.vldb.org/conf/2001/P491.pdf>
- Luis Gravano, Panagiotis G. Ipeirotis, H. V. Jagadish, Nick Koudas, S. Muthukrishnan, Lauri Pietarinen, Divesh Srivastava: *Using q-grams in a DBMS for Approximate String Processing*. IEEE Data Eng. Bull. 24(4): 28-34 (2001)
<http://sites.computer.org/debull/A01DEC-CD.pdf>