

Avaliações de monografias de anos anteriores

André Henrique Serafim Casimiro

4 de maio de 2011

1 ”Atualização Incremental de Data Warehouses”

Trabalho realizado em 2009 pelo aluno de graduação Pedro Paulo de Souza Bento da Silva que obteve nota 10.

1.1 Resumo

O principal objetivo deste trabalho foi criar um Data Warehouse para os dados do Hemocentro de São Paulo e estabelecer um procedimento para posteriores atualizações incrementais do mesmo. Além disso, visou-se a integração dos dados de diversos hemocentros do Brasil mas por questões de tempo o escopo teve de ser limitado.

Um Data Warehouse é um BD construído a partir de diversos outros BDs transacionais e tem como objetivo integrar fontes de dados e prover uma estrutura para realização de consultas analíticas ao longo do tempo; uma tarefa muitas vezes computacionalmente inviável nos BDs relacionais convencionais. Os DWs mantêm “fotografias” dos dados ao longo do tempo e o volume desses dados é gigantesco. Torna-se, então, indispensável a automatização do processo de atualização dos mesmos. Neste trabalho, o autor construiu o modelo multi-dimensional dos dados no DW e definiu todo o procedimento de carga de uma nova fotografia no DW com rotinas ETL (Extract Transform Load). Antes de carregar novos dados, os mesmos passam por uma verificação de consistência. Essa é uma etapa cara de ser construída pois exige muita intimidade com o modelo de dados mas foi o que garantiu a alta qualidade do DW. Ficou, então, para trabalhos posteriores a integração com os demais hemocentros brasileiros.

1.2 Avaliação da parte técnica

O texto está escrito em uma linguagem acessível a alunos de graduação, os conceitos não são somente explicados mas também contextualizados com sua aplicação. Além disso gostei também de saber as tentativas que falharam (como o uso do Pacote Pentaho).

A organização das informações está coerente e um ponto forte é a contextualização histórica do problema e das necessidades do projeto, o que conta bastante visto que o trabalho foi desenvolvido para substituir um sistema que não atendia bem as necessidades dos hemocentros e seus pesquisadores.

1.3 Avaliação da parte subjetiva

O Pedro expôs bem suas dificuldades e frustrações, e nelas eu me identifiquei um pouco. Assim como ele, tenho diversas atividades paralelas à graduação e ver que o trabalho dele obteve um resultado tão bom mesmo sob essas condições é realmente inspirador. Ele relacionou o trabalho com as matérias do curso e eu me espantei em não ver ali *MAC439 - Laboratório de Banco de Dados* que foi a matéria que me fez escolher a área de BD para o tema do TCC.

1.4 Críticas, elogios e comentários

Gostei do trabalho e pretendo definir o tema do meu TCC na mesma área de Data Warehouses. Parabéns ao autor! Não fossem os problemas que ele teve com as soluções visuais eu adoraria desenvolver algum projeto bem semelhante.

2 "Recuperação de Informações em Bancos de Dados Textuais"

Trabalho realizado em 2009 pela aluna de graduação Marcela Ortega Garcia que obteve nota 8.

2.1 Resumo

Este trabalho teve origem em uma necessidade do sistema Centro de Estudos do Genoma Humano (CEGH) que era a de criar um sistema análogo ao sistema de buscas do Google: utilizar uma única expressão de busca para encontrar informações em campos de natureza distinta, porém todos textuais. Esse é o desafio da área de Recuperação de Informação (RI): encontrar resultados relevantes em dados pouco estruturados, geralmente textuais.

Nesta difícil tarefa, alguns cuidados são tomados na construção do índice de palavras e também na interpretação da busca. São feitos alguns tratamentos nos textos, como a eliminação de palavras muito comuns (stopwords) e a identificação de palavras que possuem mesma raiz gramatical (plural, gerúndio, conjugações verbais, etc...). Para que a busca seja rápida são utilizadas estruturas de dados específicas e existem modelos de RI que permitem que a definição de relevância seja feita para cada domínio especificamente. Os conceitos estudados foram, por fim, aplicados na implementação de uma busca indexada para o sistema do CEGH e verificou-se que o ganho de desempenho é muito alto ao se comparar buscas indexadas com a convencional busca SQL.

2.2 Avaliação da parte técnica

No tempo em que o trabalho estava sendo elaborado, parece que não se sabia direito que rumo o mesmo levaria. Isso se reflete no fato de que a proposta gira, apenas, em torno de estudos e que são introduzidos alguns conceitos que nem sequer são utilizados na implementação da busca indexada no sistema do CEGH. Constitui-se, portanto, uma ótima leitura inicial para quem não conhece a área mas eu gostaria de saber mais resultados concretos sobre o tema.

2.3 Avaliação da parte subjetiva

É onde você realmente entende de onde surgiu o trabalho, de um projeto em que a própria Marcela já desenvolvia. Faltou fazer esta contextualização no começo da parte técnica. As matérias relevantes são bastante as esperadas e o futuro do projeto parece bastante interessante e tenho certeza que trabalhar com semântica é um grande desafio.

2.4 Críticas, elogios e comentários

Eu gosto bastante de estruturas de dados, o que tornou a leitura deste trabalho bem interessante para mim. Espero que, assim como a Marcela, possa fazer um TCC que me vire um bom assunto para o projeto de mestrado.