

1 Introdução

1.1. Motivação

Grande parte das empresas precisa guardar seus dados de maneira estruturada e organizada para que seja possível efetuar uma rápida consulta quando necessitar. Para tal, muitas delas utilizam um banco de dados relacional. Um banco de dados relacional permite que dados sejam armazenados e facilmente recuperados através de consultas. Seja para o controle de seus funcionários ou para armazenar dados estratégicos de seus produtos, vendas ou previsões, tudo normalmente fica armazenado em um banco de dados.

Para que um banco de dados seja útil para um negócio, além do correto armazenamento, é preciso existir recursos de extração dos dados inseridos auxiliando assim a tomada de decisão. Sem uma maneira eficiente para extrair os dados a coleta pode se inútil.

Normalmente a solução encontrada para disponibilizar os dados contidos nos bancos de dados parte do desenvolvimento de aplicações específicas que apresentam campos de consulta, organizados em formulários. A partir destes campos o usuário fornece entradas para a execução de consultas pré-determinadas com o objetivo de recuperar dados de uma tarefa específica.

Apesar de se uma solução viável, que realiza a extração de maneira correta para o usuário final, existe uma enorme dependência no desenvolvimento de aplicativos para a recuperação de um determinado conjunto de dados. Aplicativos devem ser desenvolvidos na medida em que outros tipos de dados (novos parâmetros) são coletados podendo tornar o processo lento e muito dependente.

Para o desenvolvimento destes aplicativos é preciso um profissional que entenda de banco de dados e, além disto, conheça a estrutura de dados da empresa. Se houver diferentes tipos de dados disponíveis nos bancos de dados, podem ser necessários muitos formulários para recuperá-los, dificultando ainda o usuário achar o que realmente precisa.

Por exemplo, em uma empresa de petróleo onde são registrados dados de produção de óleo, água, vapor, etc., dados de poços (nome, sigla, localização, composição, equipamentos e etc.), sísmica, previsão de produções, funcionários, empresas contratadas, dentre outros, são necessários inúmeros formulários para que seja possível recuperar os dados existentes.

Muitas vezes, apesar de existir um formulário para recuperar um determinado dado, o usuário necessita de uma visão diferente do que foi recuperado, causando uma nova demanda para a equipe de desenvolvimento. Imagine que foi desenvolvido um formulário para recuperar as produções dos últimos anos de um determinado poço, porém o usuário necessita saber apenas as produções que são maiores que um valor qualquer ou apenas de certo período ou outra informação que não é possível extrair através do formulário criado.

Paralelamente à construção de formulários, existem outras maneiras de recuperar dados que podem contornar os problemas levantados. Por exemplo, podemos nos basear na solução adotada por mecanismos de busca na Internet que normalmente disponibilizam apenas um campo de consulta para o usuário inserir palavras-chave que está relacionado ao resultado que espera alcançar para fazer consultas em bancos de dados relacionais.

Vale lembrar que existe uma grande diferença em buscar dados na Internet ou em um banco de dados relacional. Na Internet existem documentos com todo o seu conteúdo indexado que o motor de busca, ao encontrar uma palavra-chave que o usuário inseriu, retorna a página como sugestão de resultado. Por outro lado, em bancos de dados, existem relacionamentos entre os dados que precisam ser respeitados para que o resultado seja correto.

Desta discussão preliminar é possível compreender que estudar técnicas para simplificar a formulação de consulta a bancos de dados torna-se de extrema importância. Não basta armazenar: são precisos recursos simples que facilitem a disponibilização dos dados.

1.2. Abordagem proposta

O objetivo principal desta dissertação consiste em desenvolver um sistema que, a partir de palavras-chave, gere consultas para serem executadas em um banco de dados relacional, retornando dados relevantes em função das palavras-chave de entrada.

Em linhas gerais, o sistema funcionará da seguinte maneira: através de uma interface Web, utilizando o navegador, o usuário seleciona uma aplicação que deseja consultar e fornece palavras-chave. O sistema se encarrega de gerar dinamicamente as consultas, retornando possíveis resultados ao usuário.

As consultas serão geradas utilizando apenas as palavras-chave fornecidas pelo usuário, porém é possível selecionar quais associações entre as palavras e os elementos do banco de dados (nome das colunas, descrição ou conteúdo) estão de acordo com o que deseja procurar, melhorando o seu resultado.

Existirá ainda uma área administrativa onde será possível criar aplicações (utilizadas para fazer consulta pelo usuário). Nesta área, um usuário administrador publicará os objetos (tabelas ou visões) que poderão ser consultados.

Para que o sistema consiga gerar as consultas, será necessário associar as palavras-chave de entrada com as informações do banco de dados, para tal, é essencial que, na parte administrativa, ao incluir um novo objeto, o administrador acrescente comentários descrevendo melhor cada coluna e objeto do banco de dados.

1.3. Organização da dissertação

O capítulo 2 descreve alguns trabalhos relacionados que serviram como referência para a dissertação de mestrado. O capítulo 3 apresenta uma visão geral da arquitetura proposta neste trabalho, descrevendo cada etapa, além de exemplificar as tarefas seguindo tal arquitetura. Os capítulos 4 e 5 apresentam a implementação da arquitetura proposta no capítulo 3, mostrando maiores detalhes da administração no capítulo 4 e efetivamente a busca no capítulo 5. O capítulo 6 apresenta exemplos de aplicações reais funcionando utilizando a solução proposta. Finalmente no capítulo 7 apresento as contribuições do trabalho e sugestões de trabalhos futuros. No anexo encontram-se alguns scripts utilizados no decorrer deste trabalho.