

1

Introdução

1.1 Motivação

Nos últimos anos, dados dos mais variados domínios têm sido disponibilizados na Internet (Rodriguez, 2009). Crescente em tamanho e domínio, eles têm tornado cada vez mais atrativo o desenvolvimento de aplicações inovadoras que integram diferentes fontes, de forma a superar as limitações dos sistemas gerenciadores de dados tradicionais (Haase, Schmidt, & Schwarte, 2011). Tais dados também oferecem a oportunidade de desenvolver novas aplicações unindo dados proprietários com dados públicos disponibilizados na Web. Essa união é representada por ligações que precisam ser administradas de forma a se manterem consistentes, mesmo após modificações internas e externas.

De fato, o número de ligações desse tipo tende a crescer, tornando custoso e repetitivo atualizá-las sempre que a base local for modificada. Apesar disso, algumas fontes optam por essa rotina de atualização como estratégia de publicação, triplicando as entidades e suas ligações a cada modificação identificada.

Outra estratégia de publicação alternativa consiste em não realizar materializações de triplas, o que torna o processo ainda mais custoso, uma vez que as ligações precisam ser recalculadas toda vez que uma entidade for acessada.

Como terceira opção, existe uma alternativa híbrida: não materializar as entidades e materializar as ligações. Essas ligações são recalculadas somente quando ocorre alguma modificação nas entidades que estiverem envolvidas (sujeito e objeto). Alterações em alguma das entidades envolvidas em uma ligação podem resultar na necessidade de novas buscas por entidades equivalentes. Logo, tais alterações necessitam de monitoramento. Para cada tipo de alteração, tanto na base interna quanto na base externa, deve haver um plano de manutenção para as ligações envolvidas.

Visando atender essas necessidades, este trabalho apresenta um modelo e uma ferramenta para a materialização e manutenção de ligações em um Sistema

Gerenciador de Banco de Dados (SGBD). Dessa forma, a publicação de dados proprietários na Web, juntamente com suas ligações externas, é realizada apenas a partir de consultas ao banco de dados interno. Neste banco, ficam armazenadas as relações com os identificadores dos dados externos disponíveis na Web.

Como caso concreto, realizamos a materialização e manutenção de ligações entre a base de dados de um sistema interno da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) e um repositório bibliográfico on-line hospedado na Universidade de Trier na Alemanha.

1.2 Estrutura do Documento

Esta dissertação é estruturada da seguinte forma. No capítulo 2, abordamos os fundamentos relacionados ao desenvolvimento deste trabalho. No capítulo 3, apresentamos trabalhos relacionados que se preocupam com a criação das relações entre os dados publicados na Web. No capítulo 4, descrevemos a ferramenta desenvolvida, sua especificação e arquitetura, e os componentes envolvidos para a materialização e manutenção de relações, além de comentários sobre a implementação. No capítulo 5, detalhamos o estudo de caso executado e os resultados obtidos. Por fim, no capítulo 6, listamos as conclusões deste trabalho para a resolução dos problemas enumerados, além de oportunidades para trabalhos futuros.