



UNIVERSIDADE FEDERAL DE PERNAMBUCO
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA
2014.2



ANÁLISE SEMÂNTICA DE UM *MATCHER* DE ONTOLOGIAS
PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluna: Nicolle Chaves Cysneiros (ncc@cin.ufpe.br)
Orientadora: Ana Carolina Salgado (acs@cin.ufpe.br)

Recife, Outubro de 2014

1. Resumo

O volume de dados disponíveis em fontes de dados distintas e conectadas vêm crescendo, juntamente com a necessidade de extrair informações consistentes dessas fontes. O problema de integração de dados descritos por esquemas diferentes envolve a geração de um mapeamento entre conceitos do esquema, conhecida como a operação de *matching*. Ao longo dos últimos dez anos, diversas técnicas de *matching* de esquemas foram desenvolvidas, incluindo algoritmos que buscam estender as correspondências entre termos para além do nível sintático, chegando ao nível semântico do relacionamento entre conceitos. Esse Trabalho de Graduação propõe a análise das correspondências semânticas geradas pelo *matcher* semântico *SemMatcher*.

2. Contextualização

Com o crescimento da disponibilidade de informação provida pelo aumento da quantidade de computadores conectados em rede, fez-se necessária a criação de ferramentas que auxiliem o usuário a ter acesso aos dados armazenados em máquinas diferentes e organizados de acordo com esquemas diferentes. Para as aplicações que envolvem a integração dessas fontes de dados heterogêneas, o *matching* é uma importante operação de comparação de esquemas que recebe como entrada dois esquemas de dados e retorna um alinhamento identificando os elementos correspondentes [1].

Ao longo dos últimos dez anos, a operação de *matching* vem sendo estudada e novas técnicas foram desenvolvidas, incluindo algoritmos que buscam expandir a semântica das correspondências entre elementos dos esquemas para além da simples correspondência sintática dos termos [2]. Nesse contexto, o trabalho de [3] propõe o desenvolvimento de um *matcher* semântico de ontologias que recebe como parâmetros os esquemas das duas fontes de dados que se deseja integrar e um esquema, representado por uma ontologia, que descreve o conhecimento do domínio onde os dois primeiros esquemas estão inseridos.

A saída do *matcher* semântico é um conjunto de correspondências semânticas entre os elementos das ontologias de entrada. Tais relacionamentos são encontrados baseados em regras semânticas apresentadas no trabalho de [4], como equivalência, especialização, generalização, agregação (parte e todo) e disjunção. Para cada uma dessas correspondências semânticas é atribuído um peso que será utilizado no cálculo da medida de similaridade global entre os esquemas de entrada.

O *matcher* semântico apresentado em [3] foi desenvolvido e testado em diversos processos de integração de dados. Um exemplo seria o uso da medida de similaridade semântica no processo de roteamento da consulta (descobrir para quais pontos a consulta deveria ser encaminhada). Individualmente, os pesos das correspondências semânticas entre dois termos são principalmente utilizados no cálculo da medida de perda semântica da consulta durante o processo de roteamento da mesma, como apresentado no trabalho de [5].

3. Objetivos

Esse Trabalho de Graduação tem como objetivo geral a análise dos melhores pesos associados às correspondências semânticas entre dois termos geradas pelo *matcher* semântico. Já os objetivos específicos deste trabalho incluem:

- i. A extensão da ferramenta *SemMatcher* para o suporte a ontologias descritas em linguagem RDF;
- ii. A criação de uma API (*Application Programming Interface*) da ferramenta para que a mesma possa ser reutilizada em outras aplicações;
- iii. A avaliação da ferramenta dentro do contexto de roteamento de consultas.

Como a ferramenta *SemMatcher* é implementada utilizando a Linguagem Java, as extensões e melhorias da ferramenta também serão produzidas utilizando tal linguagem de programação. O projeto ainda comportará uma fase de teste onde será necessária a criação de ontologias escritas em Linguagem OWL e RDF e cenários específicos de uso para avaliar os pesos ideais das correspondências semânticas.

4. Cronograma

Nesta seção, é apresentado o cronograma de atividades previsto para o desenvolvimento desse Trabalho de Graduação (Tabela 1).

Atividade	Outubro				Novembro				Dezembro				Janeiro				Fevereiro			
Estudo da implementação atual da ferramenta			X	X	X															
Implementação das extensões propostas					X	X	X													
Testes e experimentos							X	X	X	X										
Análise dos resultados									X	X	X	X	X							
Elaboração do relatório						X	X	X	X	X	X	X	X	X	X					
Preparação e defesa														X	X	X	X			

Tabela 1: Cronograma de atividades

5. Possíveis Avaliadores

Os possíveis avaliadores para o resultado a ser obtido ao final de todas as etapas da proposta descrita neste documento são:

- Bernadette Farias Lóscio
- Patrícia Tedesco

Referências

1. MADHAVAN , J.; BERNSTEIN , P.; RAHM , E. **Generic Schema Matching with Cupid**. Microsoft Corporation. Redmond. 2001.
2. BERNSTEIN , P.; MADHAVAN , J.; RAHM , E. **Generic Schema Matching, Ten Years Later**. VLDB Endowment. [S.l.]: [s.n.]. 2011. p. 695-701.
3. PEREIRA , T. P. A. **Mapeamento Semântico de Ontologias no SPEED**. Universidade Federal de Pernambuco. Recife. 2008.
4. PIRES, C. E. **Um sistema de Gerenciamento Dados com Conectividade Baseada em Semântica**. Universidade Federal de Pernambuco. Recife. 2007.
5. FREIRE, C. A. **Uma Abordagem para Roteamento de Consultas em PDMS baseada em Aspectos Semânticos e de Qualidade**. Universidade Federal de Pernambuco. Recife. 2014.

Assinaturas

Nicolle Chaves Cysneiros
Orientando

Ana Carolina Salgado
Orientador

Recife, Outubro de 2014