



Pontifícia Universidade Católica do Rio Grande do Sul
Faculdade de Informática
Programa de Pós-Graduação em Ciência da Computação



Encontro de PLN

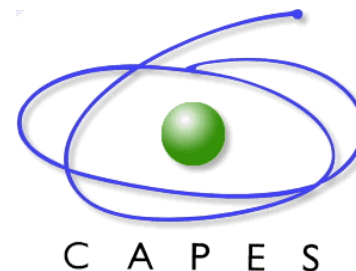
Aplicação de novas abordagens para o cálculo da similaridade intrínseca e extrínseca entre ontologias

Josiane Fontoura dos Anjos Brandolt
josiane.brandolt@pucrs.br

Vera Lúcia Strube de Lima
vera@inf.pucrs.br



Janeiro de 2008



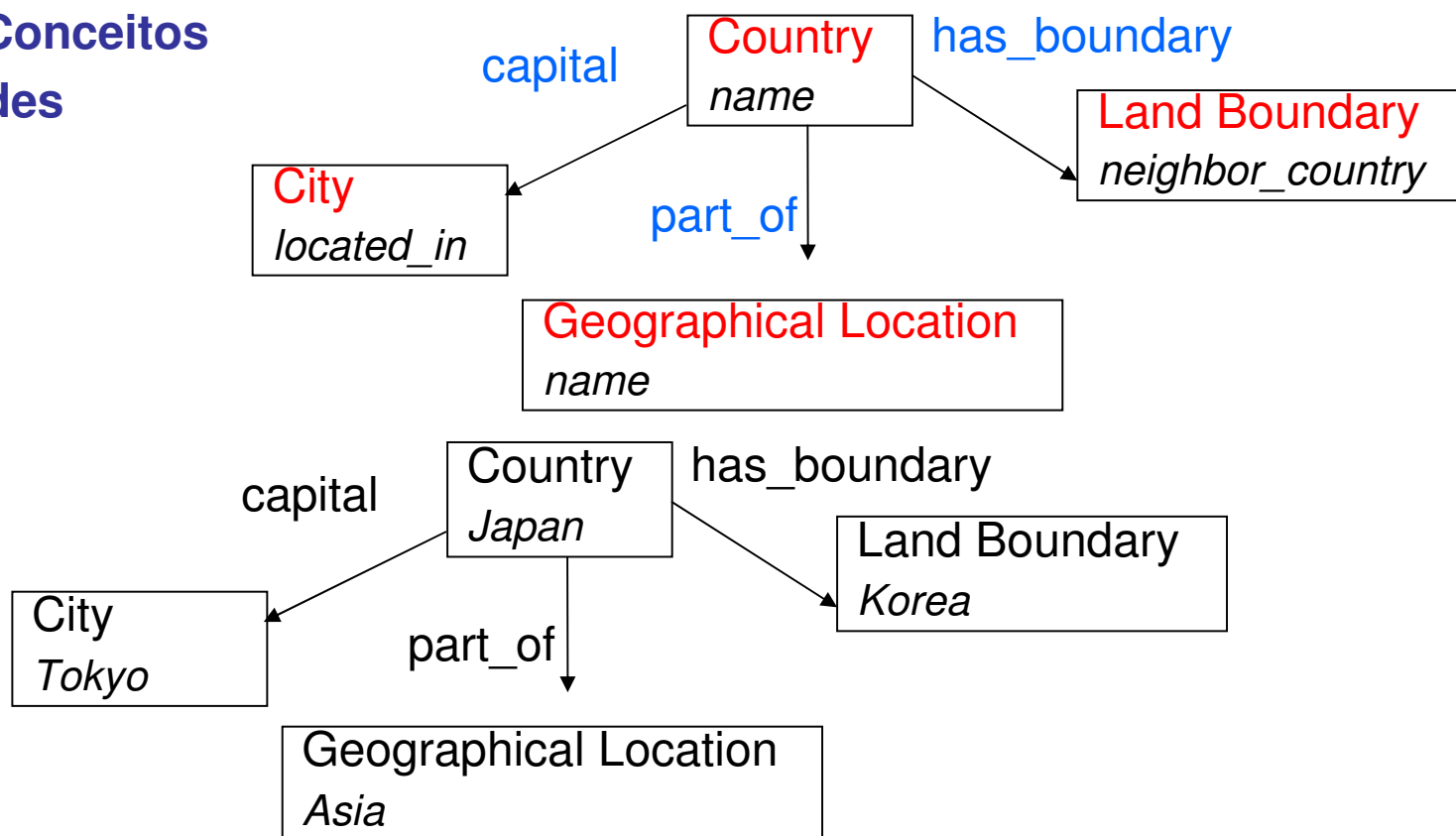
Roteiro

- ♦ **1 Introdução**
- ♦ **2 Trabalhos Correlatos**
- ♦ **3 Proposta de Trabalho**

Introdução

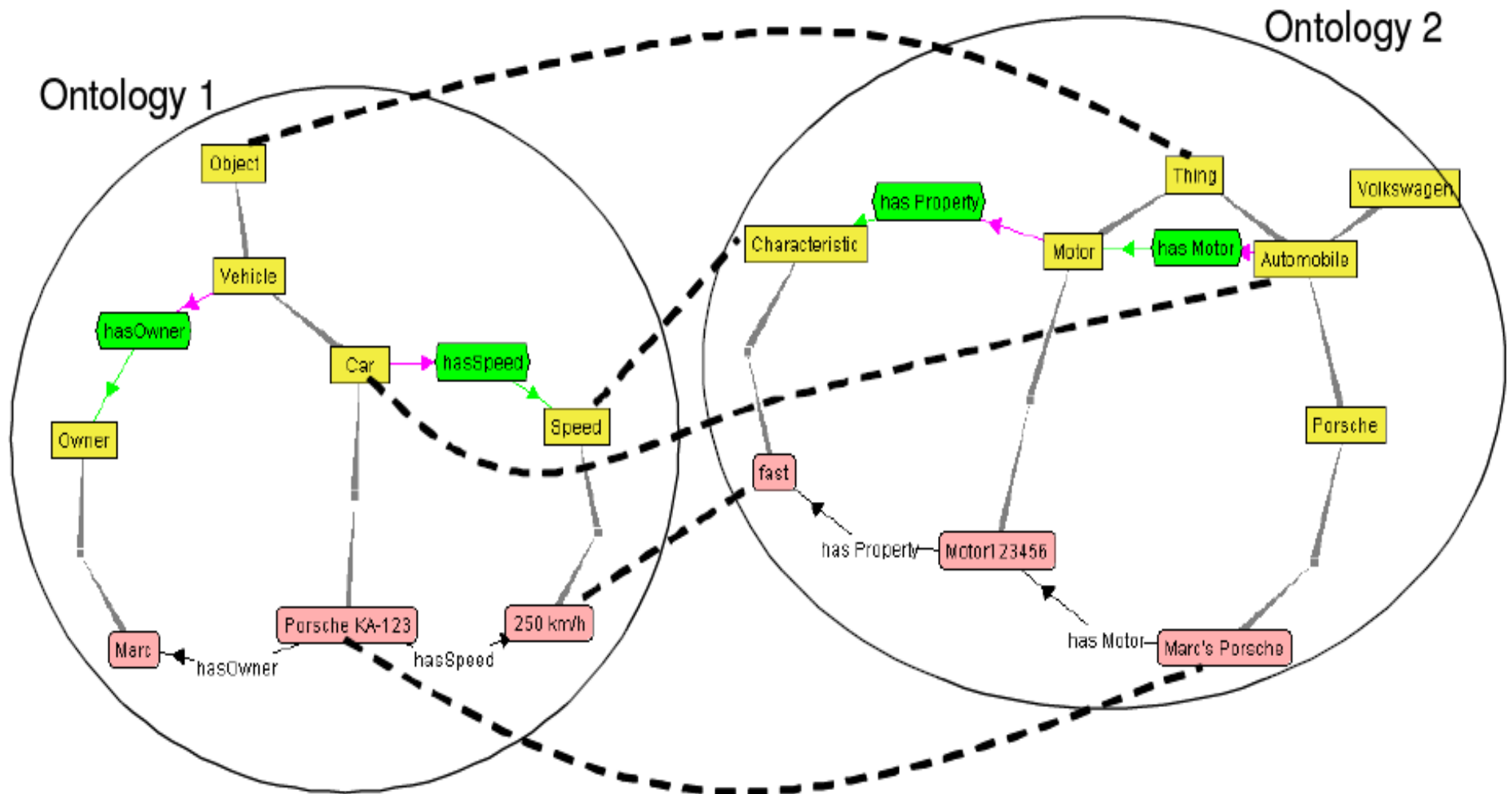
♦ Ontologias

- ♦ Classes / Conceitos
- ♦ Propriedades
- ♦ Relações
- ♦ Instâncias



Introdução

◆ Pesquisa de similaridade entre ontologias;

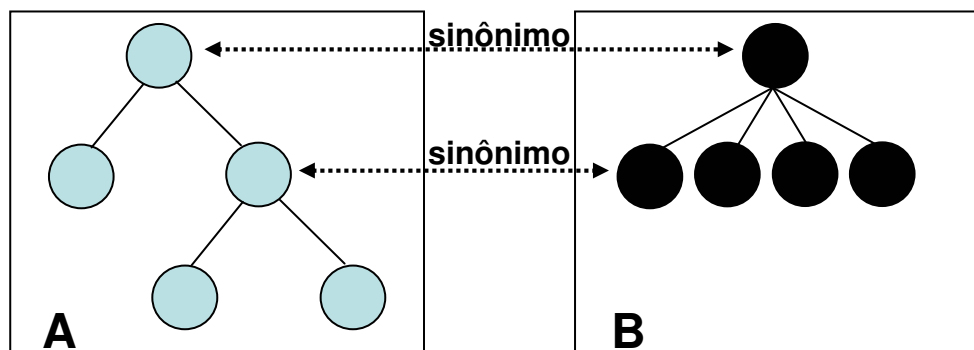


Introdução

- ♦ **Similaridade lexical e semântica;**
- ♦ **Novas visões: similaridade intrínseca e extrínseca;**
- ♦ **Medidas de similaridade sofisticadas.**

Introdução

- ♦ **Tipos de correspondências relacionadas às ontologias**
 - Alinhamento, Mapeamento e Combinação



Trabalhos Correlatos

♦ O trabalho de Freitas [FRE06]

■ Medida Sise

- Mapeamento Taxonômico de Maedche e Staab [MAE02];
- Análise de hierarquia;
- Conceitos *Semantic Cotopy* (SC) e *Common Semantic Cotopy* (CSC);
- Análise das linguagens, abstrair a sintaxe e normalizar em XML;
- Protótipo disponível ao grupo de pesquisa (usou Vocabulário Controlado Básico do Senado Federal e Vocabulário Controlado da USP);
- Avaliação da medida (*Golden Mapping*).

Trabalhos Correlatos

♦ O estudo de Tous e Delgado [TOU06]

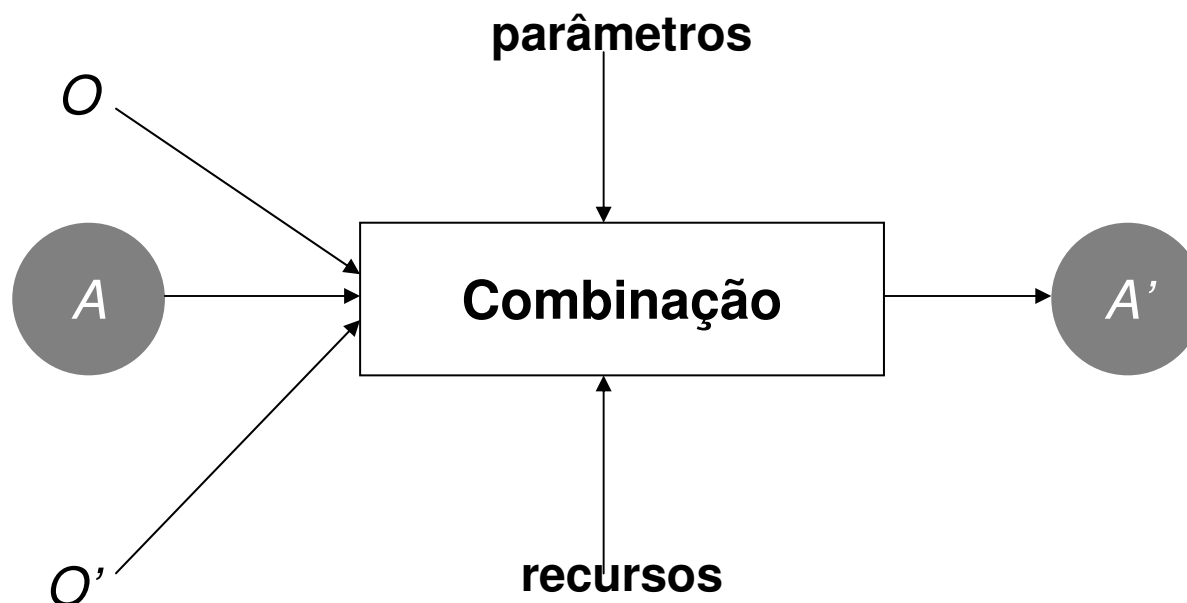
- Utiliza o Modelo de Espaço Vetorial (MEV) para alinhamento entre ontologias;
- Representação vetorial;
- Adapta o mapeamento do algoritmo de Blondel e co-autores [BLO04];
- Maioria dos algoritmos (calculam elementos próximos);
- OAEI mostrou que os melhores algoritmos combinam diferentes medidas de similaridade;
- Contribuição para a proposta.

Trabalhos Correlatos

- ♦ **As ponderações de Andreas Hess [HES06]**
 - Descreve um algoritmo iterativo;
 - Estabelece métricas de distâncias entre *strings*;
 - Métricas baseadas em uma representação vetorial entre elementos;
 - Algoritmos em conjunto;
 - Contribuição para a proposta.

Trabalhos Correlatos

- ♦ A análise de Euzenat e Shvaiko [EUZ07]
 - Processo de combinação (*matching*)



Trabalhos Correlatos

♦ A análise de Euzenat e Shvaiko [EUZ07]

- Dimensões: entrada, processo e saída;



- Utilizadas como ferramentas para encontrar os mapeamentos;
- Contribuição para a proposta.

Trabalhos Correlatos

- ♦ **A reflexão de Castano e co-autores [CAS07]**
 - Estratégias para combinação de diferentes técnicas de similaridade;
 - Duas categorias:
 - combinação de diferentes medidas: cada medida pode ser calculada independentemente ou de maneira integrada;
 - iteração do processo de combinação: cada medida de similaridade é fundada em algum tipo básico de similaridade entre os elementos a serem comparados, por exemplo, duas estruturas em grafos - leva-se em conta o seu contexto.
 - Contribuição para a proposta.

Trabalhos Correlatos

- ♦ **A aplicação de Ehrig e co-autores [EHR07b]**
 - Detecção semi-automática de sinônimos e homônimos de elementos de processos;
 - Modelos Semânticos de Processos de Negócios – redes de Petri descritas em ontologias codificadas em OWL;
 - Medidas de similaridade sintática, lingüística e estrutural;
 - Combinação de medidas;
 - Contribuição para a proposta.

Proposta de Trabalho

♦ Questão de pesquisa

- É possível compor um método que realize o cálculo de similaridade usando critérios intrínsecos e extrínsecos entre ontologias na língua portuguesa, aprimorando os resultados obtidos por Freitas ?

♦ Hipótese

- Existem, na literatura mais atual, estudos que congregam medidas de similaridade intrínseca e extrínseca utilizando representações vetoriais e combinações de algoritmos que resultam em medidas de similaridade combinadas, que podem ser aproveitadas junto à medida SiSe.

Proposta de Trabalho

♦ Objetivo geral:

- O objetivo geral da dissertação é estudar a aplicação de novas abordagens para o cálculo da similaridade intrínseca e extrínseca entre ontologias na língua portuguesa e incorporá-las na medida SiSe, proposta por Freitas.

Muito obrigada.