#### Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica

Thiago Henrique Dias Araujo

Texto apresentado
AO
Instituto de Matemática e Estatística
DA
Universidade de São Paulo
Para
O Exame de Qualificação
DE
Mestre em Ciências

Programa: Mestrado em Ciência da Computação Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Renata Wassermann

São Paulo, março de 2016

# Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica

Esta é a versão original do texto elaborado pelo candidato Thiago Henrique Dias Araujo para o exame de qualificação apresentado ao Instituto de Matemática e Estatística da Universidade de São Paulo como requisito para obtenção de título de Mestre em Ciências.

### Resumo

ARAUJO, T. H. D. Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica. 2016. 20 f. Exame de qualificação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

A comunidade científica pode ser vista como uma rede onde cada pesquisador se relaciona com outros pesquisadores através de colaborações, como na co-autoria de um artigo científico. Alguns trabalhos utilizam ontologias para modelar o domínio das redes de colaboração, e outros aplicam técnicas de aprendizado de máquina para prever ligações entre pessoas dentro dessas redes. Entretanto, algumas limitações existem nessas metodologias por utilizarem formas pouco expressivas de representação dessas relações, ou por não aproveitarem características específicas das entidades na análise desse domínio. A proposta do presente trabalho é criar uma Ontologia capaz de indicar características próprias dessas pessoas, e descrever relações entre elas, aplicando esse conhecimento em um modelo de aprendizado capaz de descobrir e prever novas relações entre esses pesquisadores.

Palavras-chave: redes de colaboração científica, ontologia, aprendizado de máquina, predição de ligações.

## Abstract

ARAUJO, T. H. D. Uma Abordagem Baseada em Ontologias para a Predição de Ligações entre Pesquisadores em Redes de Colaboração Científica. 2016. 20 f. Exame de qualificação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

Keywords: scientific collaboration networks, ontology, machine learning, link prediction.

# Sumário

Li	ista d	le Abreviaturas i	X
Lista de Figuras			xi
Li	sta d	le Tabelas xi	ii
1	Introdução		
	1.1	Considerações Preliminares	1
	1.2	Objetivos	1
	1.3	Contribuições	2
	1.4	Organização do Trabalho	2
2			3
	2.1	Fundamentos	3
3	Cor		5
	3.1	Considerações Finais	5
	3.2	Sugestões para Pesquisas Futuras	5
R	eferê	ncias Bibliográficas	7

## Lista de Abreviaturas

SPARQL SPARQL Protocol and RDF Query Language

SQL Structured Query Language SRL Statistical Relational Learning

SVM Support Vector Machine

SWRL Semantic Web Rule Language

# Lista de Figuras

# Lista de Tabelas

### Capítulo 1

## Introdução

Escrever bem é uma arte que exige muita técnica e dedicação. Há vários bons livros sobre como escrever uma boa dissertação ou tese. Um dos trabalhos pioneiros e mais conhecidos nesse sentido é o livro de Eco (2009) intitulado *Como se faz uma tese*; é uma leitura bem interessante mas, como foi escrito em 1977 e é voltado para teses de graduação na Itália, não se aplica tanto a nós.

Para a escrita de textos em Ciência da Computação, o livro de Justin Zobel, Writing for Computer Science (Zobel, 2004) é uma leitura obrigatória. O livro Metodologia de Pesquisa para Ciência da Computação de Wazlawick (2009) também merece uma boa lida. Já para a área de Matemática, dois livros recomendados são o de Nicholas Higham, Handbook of Writing for Mathematical Sciences (Higham, 1998) e o do criador do TEX, Donald Knuth, juntamente com Tracy Larrabee e Paul Roberts, Mathematical Writing (Knuth et al., 1996).

O uso desnecessário de termos em lingua estrangeira deve ser evitado. No entanto, quando isso for necessário, os termos devem aparecer *em itálico*.

```
Modos de citação:
indesejável: [AF83] introduziu o algoritmo ótimo.
indesejável: (Andrew e Foster, 1983) introduziram o algoritmo ótimo.
certo: Andrew e Foster introduziram o algoritmo ótimo [AF83].
certo: Andrew e Foster introduziram o algoritmo ótimo (Andrew e Foster, 1983).
certo: Andrew e Foster (1983) introduziram o algoritmo ótimo.
```

Uma prática recomendável na escrita de textos é descrever as legendas das figuras e tabelas em forma auto-contida: as legendas devem ser razoavelmente completas, de modo que o leitor possa entender a figura sem ler o texto onde a figura ou tabela é citada.

Apresentar os resultados de forma simples, clara e completa é uma tarefa que requer inspiração. Nesse sentido, o livro de Tufte (2001), *The Visual Display of Quantitative Information*, serve de ajuda na criação de figuras que permitam entender e interpretar dados/resultados de forma eficiente.

#### 1.1 Considerações Preliminares

Considerações preliminares<sup>1</sup>. Texto texto.

#### 1.2 Objetivos

Texto texto.

<sup>&</sup>lt;sup>1</sup>Nota de rodapé (não abuse).

2 INTRODUÇÃO 1.4

#### 1.3 Contribuições

As principais contribuições deste trabalho são as seguintes:

• Item 1. Texto texto.

• Item 2. Texto texto.

#### 1.4 Organização do Trabalho

No Capítulo 2, apresentamos os conceitos ... Finalmente, no Capítulo 3 discutimos algumas conclusões obtidas neste trabalho. Analisamos as vantagens e desvantagens do método proposto ... As sequências testadas no trabalho estão disponíveis no Apêndice

## Capítulo 2

# Conceitos

Texto texto

#### 2.1 Fundamentos

Texto texto

4 CONCEITOS 2.1

## Capítulo 3

## Conclusões

Texto texto.

#### 3.1 Considerações Finais

Texto texto.

#### 3.2 Sugestões para Pesquisas Futuras

Texto texto.

Finalmente, leia o trabalho de Alon (2009) no qual apresenta-se uma reflexão sobre a utilização da Lei de Pareto para tentar definir/escolher problemas para as diferentes fases da vida acadêmica. A direção dos novos passos para a continuidade da vida acadêmica deveriam ser discutidos com seu orientador.

<sup>&</sup>lt;sup>1</sup>Exemplo de referência para página Web: www.vision.ime.usp.br/~jmena/stuff/tese-exemplo

## Referências Bibliográficas

- Alon(2009) Uri Alon. How to choose a good scientific problem. *Molecular Cell*, 35(6):726–728. doi: 10.1016/j.molcel.2009.09.013. Citado na pág. 5
- **Eco(2009)** Umberto Eco. *Como se Faz uma Tese*. Perspectiva, 22º edição. Tradução Gilson Cesar Cardoso de Souza. Citado na pág. 1
- **Higham(1998)** Nicholas J. Higham. *Handbook of Writing for the Mathematical Sciences*. SIAM: Society for Industrial and Applied Mathematics, segunda edição. Citado na pág. 1
- Knuth et al. (1996) Donald E. Knuth, Tracy Larrabee e Paul M. Roberts. Mathematical Writing. The Mathematical Association of America. Citado na pág. 1
- Tufte(2001) Edward Tufte. The Visual Display of Quantitative Information. Graphics Pr, 2nd edição. Citado na pág. 1
- Wazlawick (2009) Raul S. Wazlawick. *Metodologia de Pesquisa em Ciencia da Computação*. Campus, primeira edição. Citado na pág. 1
- **Zobel(2004)** Justin Zobel. Writing for Computer Science: The art of effective communication. Springer, segunda edição. Citado na pág. 1