



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Modelo de Reputação e Ontologia Aplicados à Rede Social Científica do ObserveUnB

Gabriela Netto Guerra

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Orientadora
Prof.^a Dr.^a Célia Ghedini Ralha

Brasília
2012

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Mestrado em Informática

Coordenadora: Prof.^a Dr.^a Mylene Christine Queiroz de Farias

Banca examinadora composta por:

Prof.^a Dr.^a Célia Ghedini Ralha (Orientadora) — CIC/UnB

Prof.^a Dr.^a Genaina Nunes Rodrigues — CIC/UnB

Prof. Dr. Virgílio Augusto Fernandes Almeida — DCC/UFMG

CIP — Catalogação Internacional na Publicação

Guerra, Gabriela Netto.

Modelo de Reputação e Ontologia Aplicados à Rede Social Científica
do ObserveUnB / Gabriela Netto Guerra. Brasília : UnB, 2012.

136 p. : il. ; 29,5 cm.

Dissertação (Mestrado) — Universidade de Brasília, Brasília, 2012.

1. colaboração científica, 2. currículo Lattes, 3. ObserveUnB,
4. ontologia, 5. rede social, 6. reputação

CDU 004

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília-DF — Brasil

Dedicatória

Dedico este trabalho à minha família, em especial, meus pais, minha irmã, minha sogra, minha tia Joaquina, meu esposo Gsttenyo e meus filhos Gabriel e Marina.

Agradecimentos

Muitas pessoas contribuíram para o andamento deste projeto. Entre elas destacam-se meus pais e minha irmã Fabiana pelo apoio mesmo de longe, minha tia Joaquina e minha sogra pelo apoio dado sempre que recorria à sua ajuda, meus filhos Gabriel e Marina que mesmo sem entender o motivo de tanto estudo nos últimos dois anos, colaboraram aceitando a minha indisponibilidade e respeitando este momento.

Agradeço ao meu esposo Gsttenyo pelo incentivo a voltar às salas de aula e pela força nos momentos mais difíceis, pois não fosse ele não teria iniciado e nem terminado este projeto.

Agradeço também aos colegas que estudamos em conjunto durante as atividades relacionadas as disciplinas do mestrado, e também às professoras Alba Melo, Maria Emília Teles e Marisa Bräscher pelo profissionalismo, competência e pelas valiosas sugestões e críticas.

Aos professores que participaram durante a fase de validação deste projeto, respondendo às pesquisas.

Aos colegas de trabalho pela compreensão de meus diferentes horários de trabalho neste período.

Não posso deixar de agradecer a equipe do projeto Observe UnB, em especial ao Murilo, a Marcelly e ao Gedalias, pelo esforço despendido ao projeto e pela inestimável ajuda na implementação da ontologia e do modelo de reputação.

E por último, agradeço imensamente a minha orientadora Célia Ghedini Ralha. Agradeço pela proposta de trabalho, pelo apoio, sua intensa dedicação, disponibilidade e sobretudo paciência durante os últimos dois anos, e pelas suas várias e criteriosas revisões dos textos escritos nessa dissertação e nos artigos científicos. E também, agradeço ao prof. José Ralha pela inestimável ajuda com o latex ao fim da escrita desta dissertação.

Abstract

This work proposes a model of reputation and an extended ontology for use in scientific social network provided by ObserveUnB - a social network on the Web, developed with the scientific information of the University of Brasilia (UnB) academic community. The essence of ObserveUnB is based on the construction and analysis of the relationships between the researchers of the network, which was build with the aid of a reputation model and an ontology, where the ObserveUnB's big challenge is the availability of scientific knowledge, as well as to manage the integration of a large volume of data from different repositories.

The model reputation and ontology proposed associate scientific informations from a academic community of UnB, using information collected from Lattes Platform, such as the Lattes Curriculum and Directory of Research Groups and institutional systems. This model consists of an adaptation of the reputation model proposed by Cervi et al. (2011), which is based on academic researcher experience with a focus on research and scientific productivity. The proposed ontology is an extension of existing ontologies: Lattes Curriculum and Directory of Research Groups developed by the Conscientias community. These ontologies goal is to sustain the reputation model proposed by the representation of the concepts used in the reputation model, and to allow semantic queries in the domain of scientific research through the ontology structure of concepts and the use of an inference mechanism.

Thus, the reputation model and the ontology intend to contribute to the interaction of researchers and the information provision. Besides it intends to assist the formation of new research partnerships, allowing researchers to find partners for more efficiently and reliably, maximizing scientific collaboration.

Keywords: scientific collaboration, Lattes curriculum, ObserveUnB, ontology, social network, reputation

Resumo

Este trabalho propõe um modelo de reputação e uma ontologia estendida para aplicação na rede social científica disponibilizada pelo ObserveUnB - um portal *Web* de rede social científica desenvolvido com base nas informações da comunidade acadêmica da Universidade de Brasília (UnB). A essência do ObserveUnB encontra-se na construção e análise dos relacionamentos existentes entre os pesquisadores da rede, com auxílio de um modelo de reputação e uma ontologia, onde o grande desafio do ObserveUnB encontra-se na disponibilização do conhecimento científico, bem como no gerenciamento e integração de um grande volume de dados de diferentes repositórios.

O modelo de reputação e a ontologia propostos associam informações de pesquisa e produção científica da comunidade acadêmica da UnB, fazendo uso de informações coletadas da Plataforma *Lattes*, tais como, o Currículo *Lattes* e o Diretório de Grupos de Pesquisa, e de sistemas institucionais. Tal modelo consiste em uma adaptação do modelo de reputação proposto por Cervi et al. (2011), o qual baseia-se na experiência acadêmica do pesquisador com o foco na pesquisa e na produtividade científica. Enquanto a ontologia proposta é uma extensão das ontologias existentes de Currículo *Lattes* e de Diretório de Grupos de Pesquisa desenvolvidas pela comunidade Conscientias. Seu objetivo é sustentar o modelo de reputação proposto, por meio da representação dos conceitos utilizados no modelo de reputação, bem como permitir a realização de consultas semânticas dentro do domínio da pesquisa científica através da estrutura de conceitos oferecida pela ontologia e do uso de um mecanismo de inferência.

Dessa forma, o modelo de reputação aliado à ontologia visam contribuir na interação entre os pesquisadores e na disponibilização de informações. Além disso, pretendem auxiliar na formação de novas parcerias na pesquisa, permitindo que pesquisadores encontrem parceiros para suas pesquisas de forma mais eficiente e confiável, maximizando a colaboração científica.

Palavras-chave: colaboração científica, currículo Lattes, ObserveUnB, ontologia, rede social, reputação

Sumário

Lista de Figuras	x
Lista de Tabelas	xii
1 Introdução	1
1.1 Definição do Problema	2
1.2 Justificativa	3
1.3 Objetivos	4
1.4 Metodologia	4
1.5 Estrutura do Documento	5
2 Rede Social	6
2.1 Estrutura	7
2.2 Análise de Rede Social	8
2.3 Rede Social Científica	10
2.3.1 Colaboração Científica	10
2.3.2 Plataforma <i>Lattes</i>	12
3 Reputação	15
3.1 Vantagens e Aplicação	16
3.2 Componentes	17
3.3 Tipologia	18
3.4 Modelos de Cálculo	20
3.5 Modelos de Classificação	22
3.6 Problemas	23
4 Ontologia	25
4.1 Vantagens	27
4.2 Elementos Básicos	28
4.3 Classificação	28
4.4 Integração	29
4.5 Linguagens	30
4.5.1 XML	31
4.5.2 RDF	31
4.5.3 OWL	32
4.6 Metodologias	36
4.7 Ferramentas	37

4.8	Raciocinador	38
4.9	Jena	40
4.9.1	Criação de ontologias	41
4.9.2	Carga de ontologias	41
4.9.3	Manipulação de classes	42
4.9.4	Manipulação de propriedades	42
4.9.5	Manipulação de instâncias	43
4.9.6	Consultas	43
4.10	Conscientias	44
5	Trabalhos Correlatos	47
6	Proposta	55
6.1	Modelo de Reputação	55
6.1.1	Metodologia	55
6.1.2	Métrica	56
6.1.3	Componentes do Modelo de Reputação	59
6.1.4	Tipo do Sistema de Reputação	59
6.1.5	Modelo de Reputação Proposto e o Modelo de Cervi et al. (2011)	60
6.2	Ontologia	62
6.2.1	Linguagem e Ferramenta	62
6.2.2	Metodologia	63
7	Estudo de Caso	73
7.1	ObserveUnB: Visão Geral	73
7.1.1	Extração	74
7.1.2	Integração	74
7.1.3	Repositório	75
7.1.4	Interface de Visualização	76
7.2	Avaliação da Proposta	81
7.2.1	Modelo de Reputação	81
7.2.2	Ontologia	94
7.3	Análise dos Resultados	105
7.3.1	Modelo de Reputação	105
7.3.2	Ontologia	107
8	Conclusões e Trabalhos Futuros	109
A	Questionário de Avaliação do Modelo de Reputação	111
B	Consultas realizadas na Ontologia	113
	Referências	117

Lista de Figuras

1.1	Estágios do projeto.	5
2.1	Rede não direcionada.	8
2.2	Rede social científica.	12
3.1	Relacionamentos entre confiança, reputação e reciprocidade.	16
3.2	Estrutura de sistemas de reputação.	18
3.3	Tipologia de reputação.	19
3.4	Classificação baseada em perfis.	22
3.5	Classificação baseada em níveis.	22
3.6	Classificação baseada em membros.	23
3.7	Classificação baseada em pontos.	23
3.8	Classificação baseada em <i>ranking</i>	24
4.1	Amostra de classes da ontologia de currículo Lattes do grupo Conscientias.	26
4.2	Diferentes tipos de ontologia e seus relacionamentos.	29
4.3	Principais classes da ontologia de Grupos de Pesquisa.	45
4.4	Principais classes da ontologia de Currículo Lattes.	46
5.1	Rede de colaboração de dois pesquisadores no <i>ArnetMiner</i>	48
5.2	Mapa acadêmico do Brasil no <i>Microsoft Academic Search</i>	49
5.3	Rede de co-autoria de um pesquisador no portal VIVO.	50
5.4	Rede de colaboração entre pesquisadores de um INCT.	50
5.5	Mapa de Co-ocorrência de termos no Observatório da <i>Web</i>	51
6.1	Estrutura do sistema de reputação proposto.	59
6.2	Processo de desenvolvimento do Método 101.	64
6.3	Amostra de classes da estrutura acadêmica da UnB.	66
6.4	Propriedade <i>bolsista</i> acrescentada na classe <i>Autores</i>	67
6.5	Propriedade <i>temJCR</i> acrescentada na classe <i>DetalhamentoDoArtigo</i>	68
6.6	Restrição da propriedade <i>especialidade</i>	68
6.7	Restrição da propriedade <i>temArea</i>	69
6.8	Restrição da propriedade <i>temCurriculo</i>	69
6.9	Propriedade <i>nomeParaCitacao</i> sem a restrição funcionalidade.	70
6.10	Instância da classe <i>Autores</i>	71
6.11	Instância da classe <i>CurriculoVitae</i>	72
7.1	Arquitetura <i>ObserveUnB</i>	74
7.2	Amostra de classes da ontologia estendida utilizada no <i>ObserveUnB</i>	76

7.3	Modalidades de pesquisa do ObserveUnB.	77
7.4	Total de produção científica do Programa de Pós-graduação “A” da UnB. . .	77
7.5	Rede das unidades acadêmicas da UnB.	78
7.6	Pesquisa georreferenciada dos quatro <i>campi</i> da UnB.	78
7.7	Rede de relacionamentos do programa de pós-graduação “A” da UnB. . . .	79
7.8	Rede de relacionamentos de um pesquisador da UnB.	80
7.9	Visualização da estrutura de classes definida na ontologia.	80
7.10	Classificação da reputação dos docentes da Ciência da Computação da UnB.	86
7.11	Produção bibliográfica referente aos níveis de reputação.	87
7.12	Média de produção bibliográfica dos docentes classificados por nível. . . .	87
7.13	Tempo médio de carreira dos docentes classificados por nível.	88
7.14	Classificação da reputação dos docentes bolsistas PQ da UnB.	89
7.15	Classificação da reputação dos docentes bolsistas PQ1A do país.	89
7.16	Dados dos pesquisadores P1 a P18	93
7.17	Dados dos pesquisadores P19 a P33	93
7.18	Dados dos pesquisadores P34 a P49	94
7.19	Exemplo de consulta realizada no Protégé para a Questão 1.	96
7.20	Exemplo de consulta realizada no Protégé para a questão 2.	97
7.21	Exemplo de consulta realizada no Protégé para a Questão 3.	97
7.22	Exemplo de consulta realizada no Protégé para a Questão 4.	100
7.23	Exemplo de consulta realizada no Protégé para a Questão 5.	102
7.24	Exemplo de consulta realizada no Protégé para a Questão 6.	103
7.25	Consulta realizada no ObserveUnB para a questão 7.	103

Lista de Tabelas

1.1	Índices do Diretório de Grupos de Pesquisa do CNPq.	3
4.1	Projetos que fazem uso de ontologias.	28
5.1	Categorias e elementos definidos no modelo de Cervi et al. (2011).	53
5.2	Classificação da reputação definida no modelo de Cervi et al. (2011).	54
6.1	Categorias e elementos definidos no modelo de reputação.	57
6.2	Categorias e elementos.	61
7.1	Categorias, elementos e pesos.	82
7.2	Categorias, elementos e pesos revisitadas.	83
7.3	Classificação da reputação.	83
7.4	Tabela de cálculo da reputação para um pesquisador.	85
7.5	Tabela do quantitativo de entrevistas.	90
7.6	Resumo do perfil do entrevistado.	90
7.7	Resumo das respostas dos entrevistados.	92
7.8	Classes e propriedades OWL para responder à Questão 1.	95
7.9	Classes e propriedades OWL para responder à Questão 2.	98
7.10	Classes e propriedades OWL para responder à Questão 3.	99
7.11	Classes e propriedades OWL para responder à Questão 4.	99
7.12	Classes e propriedades OWL para responder à Questão 5.	100
7.13	Classes e propriedades OWL para responder à Questão 6.	101
7.14	Classes e propriedades OWL para responder à Questão 7.	104

Capítulo 1

Introdução

Comunidades na *Web* ou *sites* de Redes Sociais tem se tornado uma das mais importantes aplicações *on-line*. Isso se deve ao fato de tais ferramentas possibilitarem a interação entre os usuários, além de permitirem que usuários expressem suas opiniões sobre os mais diversos assuntos, bem como propaguem as informações, contribuindo com a disponibilização de informações na *Web*.

Uma das metas de *sites* de Redes Sociais é fornecer serviços que visam facilitar a busca e o compartilhamento de informações. No entanto, conforme cresce a quantidade de conteúdo disponível, cresce também a dificuldade de organizar e buscar as fontes certas de informação, em razão das diversas fontes de dados disponíveis.

Neste contexto, insere-se o portal ObserveUnB¹, o qual representa o primeiro portal *Web* de rede social científica desenvolvido no escopo de uma Instituição Federal de Ensino Superior no Brasil (IFES). O objetivo do ObserveUnB consiste em fomentar a colaboração científica, maximizando o potencial da produção, através da integração e disponibilização de informações e do conhecimento, que por sua vez permitem a ampliação das parcerias entre os grupos de pesquisa intra e inter-institucionais de ensino e pesquisa.

Para auxiliar na colaboração científica, o presente trabalho propõe um modelo de reputação e uma ontologia estendida da Comunidade Conscientias, base do Diretório de Grupos de Pesquisa e dos Currículos *Lattes* (Conscientias, 2002, 2006), aplicados a rede social científica disponibilizada pelo ObserveUnB. O modelo de reputação aplicado em uma rede social científica visa auxiliar na busca por potenciais parcerias de pesquisa e na recomendação de pesquisadores seniores a seus pares, bem como às entidades de fomento à pesquisa.

Com a finalidade de auxiliar na identificação e na integração das informações necessárias para o modelo de reputação, bem como adicionar aspectos semânticos às informações científicas dos pesquisadores, é proposta a utilização de uma ontologia baseada em ontologias existentes de Currículo *Lattes* e de Diretório de Grupos de Pesquisa (Conscientias, 2002, 2006). O uso da ontologia permite a organização do conhecimento, englobando o escopo da comunidade científica de forma adequada através de conceitos que integram os dados de diferentes bases, além de auxiliar na desambiguação e deduplicação desses dados. Além disso, a ontologia possibilita a recuperação de informações não explicitadas na sua definição, sendo que para isso são utilizados mecanismos de inferências disponibilizados pela ferramenta de manipulação de ontologias.

¹Disponível em: <http://www.observeunb.unb.br/>

Dessa forma, o modelo de reputação proposto e a ontologia estendida, visam auxiliar no intercâmbio de informações e disponibilizar o conhecimento produzido em uma IFES de tal forma que contribuam com a produção científica e com o esforço coordenado.

1.1 Definição do Problema

A comunidade acadêmica das IFES no Brasil é composta por um grande volume de pesquisadores, cujos dados encontram-se distribuídos de forma heterogênea em diferentes bases na *Web* e em aplicações institucionais. Repositórios que armazenam diferentes tipos de dados em múltiplos formatos estão se tornando disponíveis para uso tão rápido quanto a evolução da infraestrutura global da tecnologia da informação. Como consequência, a sobrecarga de informações resultante deste processo torna quase impossível para o usuário estar ciente dos locais, organização/estrutura e relacionamentos semânticos entre os dados dos pesquisadores distribuídos em vários repositórios (Mena et al., 2000).

Além da distribuição dos dados em diferentes repositórios, os esforços colaborativos envolvendo pesquisadores de diferentes grupos de pesquisa, áreas de conhecimento, instituições ou até mesmo de diferentes países evidenciam a necessidade de se criar veículos ou instrumentos onde a comunidade científica possa visualizar e trocar informações pertinentes à produção científica e aos projetos de pesquisa.

Como estudo de caso, podemos citar a comunidade acadêmica da UnB, que é composta por mais de 38 mil indivíduos oficialmente registrados nas bases de dados dos sistemas acadêmicos institucionais, além das bases de Currículo Lattes e de Diretório de Grupos de Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), bem como nas bases de dados dos programas de pós-graduação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Em relação à rede de colaboração científica viabilizada através do Diretório de Grupos de Pesquisa, a UnB apresenta 418 grupos de pesquisa cadastrados e certificados pela instituição, os quais são divididos em nove grandes áreas do conhecimento, onde a multidisciplinariedade tem se destacado cada vez mais na formação dos grupos. A Tabela 1.1 apresenta os quantitativos de agosto de 2011 e os referentes ao último censo de dezembro de 2010 do Diretório de Grupos de Pesquisa do CNPq, bem como os índices dos Pesquisadores, Doutores, Estudantes, Técnicos e Linhas de Pesquisa (CNPq, 2011). Note que apesar de o total de grupos ter se mantido no período, o número de estudantes aumentou de 5,6 para 8,9 estudantes por grupo (E/G), e o número de pesquisadores aumentou de 5,2 para 7,1 pesquisadores por grupo (P/G).

Dessa forma, a ontologia e o modelo de reputação propostos aplicados ao ObserveUnB enfrentam dois dos Grandes Desafios da Pesquisa em Computação no Brasil da Sociedade Brasileira de Computação (SBC). O primeiro desafio “Gestão da informação em grandes volumes de dados multimídia distribuídos” relaciona-se ao projeto visto que um grande volume de dados é extraído de bases de dados heterogêneas e integrado em um repositório de informação científica na *Web*. Outro grande desafio “Acesso participativo e universal do cidadão brasileiro ao conhecimento” é abordado no projeto desde que o ObserveUnB oportuniza a geração do conhecimento, a extensibilidade e o acesso universal. Para isso, o ObserveUnB disponibiliza informações e permite a análise dos relacionamentos da rede social, por meio do uso de um modelo de reputação e de uma ontologia.

Tabela 1.1: Dados numéricos do Diretório de Grupos de Pesquisa do CNPq (CNPq, 2011).

	Dez/2010	Ago/2011
	Total - Índice	Total - Índice
(G) Grupos Certificados pela UnB	422	418
(P) Pesquisadores	2.198 - 5,2 (P/G)	2.952 - 7,1 (P/G)
(D) Doutores	1.665 - 3,9 (D/G)	2.151 - 5,1 (D/G)
(E) Estudantes	2.358 - 5,6 (E/G)	3.741 - 8,9 (E/G)
(T) Técnicos	329 - 0,8 (T/G)	339 - 0,8 (T/G)
(L) Linhas de Pesquisa	1.701 - 4,0 (L/G)	1.715 - 4,1 (L/G)

1.2 Justificativa

A distribuição dos dados em diferentes repositórios, a carência de uma ferramenta que se adeque aos padrões nacionais da comunidade de pesquisa acadêmica e os esforços colaborativos envolvendo pesquisadores de diferentes grupos de pesquisa, áreas de conhecimento e instituições constituem a motivação para a aplicação de um modelo de reputação e de uma ontologia em uma rede social científica.

O uso de uma ontologia estendida que abrange os principais conceitos de currículo *Lattes* e de grupos de pesquisa contribui para a representação do conhecimento do domínio da pesquisa científica, facilitando a identificação das informações necessárias para o cálculo da reputação, além de proporcionar um maior refinamento nas consultas realizadas pelos usuários, obtendo maior precisão nas informações recuperadas.

Aliado aos benefícios da ontologia, o modelo de reputação aplicado aos membros de uma rede social científica auxilia na busca e na identificação de especialistas em uma determinada área de conhecimento, grupo de pesquisa ou unidade acadêmica. Além disso, a atribuição de reputação aos membros de uma rede pode servir como um apoio na distribuição de recursos para a pesquisa e na formação de novas parcerias, bem como servir como motivação para o pesquisador progredir na carreira.

Os itens a seguir resumem as motivações que impulsionam este projeto:

- facilitação e estímulo da interação entre pesquisadores na rede;
- contribuição para a integração de diversas bases de dados heterogêneas;
- possibilidade de realização de operações de inferência sobre as informações;
- facilidade de identificação de conceitos e relações do domínio de conhecimento da área de pesquisa;
- identificação de especialistas em uma área de pesquisa, em uma unidade acadêmica ou instituição;
- possibilidade de adaptação da ontologia e do modelo de reputação à diferentes aplicações.

Dessa forma, o modelo de reputação e a ontologia desempenham papéis importantes e ao mesmo tempo diversos na abordagem da colaboração científica.

1.3 Objetivos

O Objetivo geral deste trabalho consiste na disponibilização da informação e conhecimento científico na rede social acadêmica da Universidade de Brasília, através de um modelo de reputação e de uma ontologia estendida integrados no Portal do ObserveUnB.

O objetivo do uso do modelo de reputação é avaliar os membros da rede social científica do ObserveUnB, servindo como um índice de produção científica dos integrantes da rede ou indicador de especialistas em determinada área de conhecimento, podendo ser utilizado para diversos fins acadêmicos.

Enquanto que o uso da ontologia visa representar o domínio de conhecimento da pesquisa acadêmica, integrar diferentes repositórios, auxiliar na deduplicação e desambiguação dos dados, bem como permitir a realização de consultas semânticas no repositório do ObserveUnB.

1.4 Metodologia

Este projeto de pesquisa está dividido basicamente em quatro etapas:

1. Extensão da ontologia de Currículo *Lattes* e de Diretório de Grupos de Pesquisa da comunidade Conscientias e definição de um modelo de reputação:
Consiste na adaptação das ontologias e na proposta ou adaptação de um modelo de reputação ao escopo deste projeto. Estes dois processos são realizados em paralelo visto que um complementa o outro. Para definir a ontologia é preciso ter conhecimento sobre quais informações são necessárias para o modelo de reputação, por outro lado é preciso saber quais conceitos são englobados pela ontologia para definir o modelo de reputação.
2. Estudo de Caso:
Consiste na aplicação do modelo de reputação definido aos membros da rede social acadêmica do ObserveUnB e na utilização da ontologia para a realização de consultas semânticas no ObserveUnB.
3. Validação e Análise dos Resultados:
Após a implementação do modelo e da ontologia, estes são validados de forma a verificar a sua viabilidade, feito isso são analisados os resultados obtidos.
4. Escrita de artigos científicos e do documento da dissertação.

O desenvolvimento de todas as etapas baseia-se em uma pesquisa bibliográfica e documental. Bibliográfica pois para a fundamentação teórico-metodológica do trabalho foram realizadas pesquisas sobre os seguintes assuntos: ontologia, rede social e técnicas de reputação. A pesquisa também é documental porque para atingir um dos objetivos deste trabalho foi necessário investigar casos reais de aplicação do método de avaliação da rede, com a finalidade de validar o método de cálculo de reputação. O processo de trabalho deste projeto pode ser representado como na Figura 1.1.

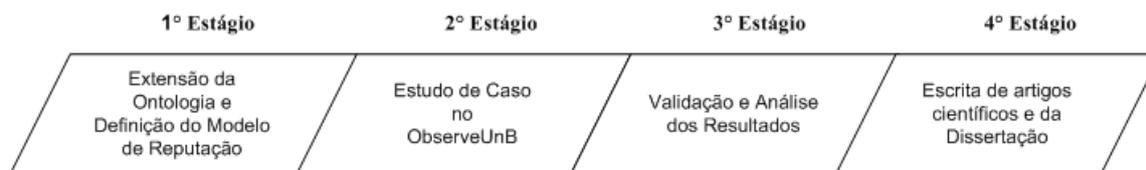


Figura 1.1: Estágios do projeto.

1.5 Estrutura do Documento

O restante deste documento está organizado da seguinte forma:

- no Capítulo 2 é realizada uma introdução dos principais conceitos de rede social relacionados a sua estrutura e análise, abordando também a Plataforma *Lattes* e as redes de colaboração científica;
- no Capítulo 3 são apresentados os conceitos relacionados a reputação, sua aplicabilidade e alguns métodos para se avaliar a reputação em um ambiente computacional;
- no Capítulo 4 são apresentados vários conceitos referentes a ontologia, bem como os elementos que fazem parte de sua construção, as vantagens e aplicações das ontologias, e as ontologias da comunidade Conscientias;
- no Capítulo 5 são descritas iniciativas de rede de colaboração científica, além de técnicas computacionais relacionadas a cálculo de reputação e a ontologia;
- no Capítulo 6 é apresentada a proposta de um modelo de reputação e de uma extensão das ontologias do grupo Conscientias para serem aplicados em uma rede social científica;
- no Capítulo 7 é realizado um estudo de caso do modelo de reputação proposto e da ontologia estendida na rede social de pesquisadores da UnB disponibilizada pelo portal ObserveUnB. Após o estudo de caso, são realizadas a validação e análise dos resultados da aplicação do modelo e da ontologia;
- no Capítulo 8 são apresentadas as principais conclusões acerca do trabalho desenvolvido, incluindo sugestões de possíveis trabalhos futuros com novas implementações e expansões.

Capítulo 2

Rede Social

Segundo Newman (2003), uma rede social é um conjunto de pessoas ou grupos de pessoas com algum padrão de contatos ou interações entre eles. Para Wasserman and Faust (1999), uma rede social é a representação de relacionamentos existentes em uma comunidade, onde a presença da informação relacional é crítica e define a característica de uma rede social.

Em uma mesma comunidade vários tipos de rede podem ser construídos, dependendo do relacionamento social: parentesco, amizade, convivência, semelhança e cooperação. Os relacionamentos podem ser estáticos (e.g., relacionamento familiar) ou mudar de acordo com o tempo (e.g., relacionamento escolar), onde a similaridade é o critério mais usado para estabelecer relacionamentos entre os componentes de uma comunidade (Tang et al., 2011b).

Uma vertente de rede social é a rede social *on-line* que permite a interação entre os membros da rede. Na *Web*, *sites* de redes sociais *on-line* estão cada vez mais atraindo a atenção de acadêmicos e pesquisadores da indústria intrigados com suas repercussões e seu poder de influência. Boyd and Ellison (2007) definem *sites* de redes sociais como serviços baseados na *Web* que permitem aos indivíduos: (i) construir um perfil público ou semi-públicos dentro de um sistema; (ii) articular uma lista de outros usuários com quem eles compartilham uma conexão e (iii) ver e percorrer sua lista de conexões e aquelas feitas por outros dentro do sistema. Para Benevenuto et al. (2011), redes sociais *on-line* são ambientes perfeitos para o estudo de vários temas da computação, incluindo sistemas distribuídos, tráfego na internet, mineração de dados, sistemas multimídia e inteligência artificial.

Segundo Tomáel et al. (2006), as redes sociais ultrapassaram o âmbito acadêmico/científico, conquistando e ganhando espaço em outras esferas. Pode-se observar esse movimento através da Internet, onde estão conquistando cada vez mais adeptos. De acordo com Tang et al. (2011b), o *Facebook*, uma das maiores redes sociais, em Janeiro de 2011 englobava mais de 600 milhões de usuários ativos.

Dessa forma, pode-se afirmar que redes sociais tornaram-se uma ponte entre a realidade e o espaço virtual na *Web*, com o intuito de aglutinar pessoas com objetivos específicos, ou apenas para desenvolver uma rede de relacionamentos, além de interagir com pessoas que já conhecem ou para conhecer novas pessoas, visto que as comunidades eletrônicas auxiliam no mapeamento de relacionamentos entre pessoas que podem não estar cientes da relação.

Isto é possibilitado por um software que, com uma interface amigável, integra recursos além dos da tecnologia da informação. O uso destes recursos gera uma rede em que os membros convidam seus amigos, conhecidos, sócios, clientes, fornecedores e outras pessoas de seus contatos para participar de sua rede, desenvolvendo uma rede de contatos profissional e pessoal, que certamente terá pontos de contatos com outras redes.

De acordo com as definições acima, atualmente existem várias redes sociais disponíveis na *Web*, que distinguem entre si de acordo com seus objetivos. O primeiro *site* de rede social foi lançado em 1997, *SixDegrees.com*. Este *site* permitia que usuários criassem perfis, lista de amigos e, no início de 1998, a navegação nesta lista (Boyd and Ellison, 2007).

Ellison et al. (2007) declaram que esses *sites* podem ser orientados a contextos, podem estar relacionados a trabalho (e.g., *LinkedIn.com*), a relacionamentos românticos (e.g., *Friendster.com*) e a outros interesses comuns, e.g., a música, a política (e.g., *MySpace.com*) e a comunidade estudantil universitária, que por sua vez constituiu a proposta original do *Facebook.com*.

2.1 Estrutura

Uma rede social é representada por meio de grafos, que são formados por um conjunto de pontos, chamados vértices, conectados por linhas, chamadas de arestas (Cormen et al., 2002). De acordo com Newman (2003), uma rede é formada pelos componentes:

- Vértice - é a unidade fundamental de uma rede, também chamado nó (na Ciência da Computação), ou ator (na Sociologia);
- Aresta - é a linha que conecta dois vértices. Também chamado aresta (na Ciência da Computação), ou um laço (na Sociologia);
- Direcionado/não-direcionado - uma aresta é direcionada se ela aponta para um lado, e se não aponta para nenhum lado ela é não direcionada. Um grafo formado por arestas direcionadas é denominado grafo direcionado e não-direcionado, em caso contrário;
- Grau - corresponde ao número de arestas conectadas ao vértice. O grau de um vértice não necessariamente é igual ao número de vértices adjacentes a ele, pois pode existir mais de uma aresta ligando dois vértices. Os vértices de um grafo direcionado possuem graus de entrada e graus de saída, que indicam os números de arestas que entram no vértice e que saem do vértice, respectivamente;
- Componente - o componente ao qual pertence um vértice é o conjunto de vértices que pode ser alcançado a partir dele por percursos ao longo das bordas do gráfico. Em um grafo direcionado um vértice tem um componente de entrada e um componente de saída, que são os conjuntos de vértices a partir do qual o vértice pode ser alcançado e que pode ser alcançado a partir dele;
- Caminho geodésico - um caminho geodésico é o caminho mais curto através da rede de um vértice a outro. Pode haver mais do que um caminho geodésico entre dois vértices;

- Diâmetro - o diâmetro de uma rede é o tamanho (em número de arestas) do mais longo caminho geodésico entre quaisquer dois vértices.

A Figura 2.1 apresenta uma pequena rede não-direcionada com oito vértices e dez arestas, onde o vértice “A” possui grau zero, o que significa que nenhuma aresta está ligada a ele.

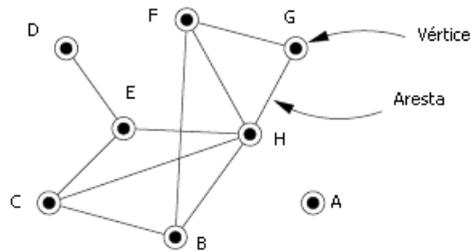


Figura 2.1: Rede não direcionada. Adaptado de Newman (2003).

Um conjunto de vértices unidos por arestas é apenas o mais simples tipo de rede, existem muitas maneiras em que as redes podem ser mais complexas do que esta. Por exemplo, podem existir diferentes tipos de vértice em uma rede, ou diferentes tipos de aresta e vértice, ou arestas podem ter uma variedade de propriedades associadas. Em uma rede social, os vértices podem representar pessoas, grupos, organizações, conceitos. No caso de uma rede social de pessoas, os vértices representam homens ou mulheres, pessoas com diferentes nacionalidades, lugares, idades, renda ou alguma outra característica. As arestas podem representar amizade, conhecidos da faculdade ou proximidade geográfica, e ainda elas podem ser distinguíveis de acordo com algum tipo de característica compartilhada pelos vértices, por exemplo, o meio como uma pessoa conhece a outra (escola, universidade, trabalho) e podem possuir pesos, representando a intensidade do relacionamento. Além disso, o grafo pode ser direcionado, podendo representar mensagens e emails, uma vez que possuem uma única direção. E ainda, grafos direcionados podem ser cíclicos, o que significa que contém circuitos fechados, ou acíclicos, ou seja não contém ciclos na rede (Newman, 2003).

Representar a rede para dezenas ou centenas de vértices, é uma questão relativamente simples, pois basta traçar um grafo da rede com pontos e linhas reais. A complexidade encontra-se na busca por respostas a questionamentos sobre a estrutura e o comportamento da rede, os quais tem sido um dos principais focos da pesquisa em redes sociais, correspondendo à Análise de Rede Social.

2.2 Análise de Rede Social

Segundo Sabater and Sierra (2002), a Análise de Rede Social (ARS) é o estudo de relações sociais existentes entre indivíduos em uma sociedade. A ARS emergiu como um conjunto de métodos para a análise da estrutura social, métodos que especificamente permitem uma investigação dos aspectos relacionais dessas estruturas. A utilização destes métodos, portanto, depende da disponibilidade de relações ao invés de dados de atributos. De

acordo com Sabater and Sierra (2002), quanto mais relações existirem entre os membros melhor é a análise da rede. No entanto, estes dados podem ser difíceis de se obter.

Atualmente, a literatura enumera significativos conceitos e metodologias de avaliação de uma rede social, dispersos em vários domínios e disciplinas do conhecimento. Marteleto (2010) define os paradigmas da ARS da seguinte forma:

- É referenciada por uma perspectiva estrutural baseada em laços interligando atores sociais;
- Está fundamentada em dados empíricos sistematizados;
- As redes são configuradas por representações gráficas;
- O estudo das redes depende do emprego de modelos matemáticos e/ou computacionais.

As análises métricas, emprego de métodos matemáticos e/ou computacionais, oferecem subsídios e instrumentação para o estudo das redes sociais na medida em que, a partir de tratamentos quantitativos, torna possível a avaliação de alguns aspectos dessas relações, através de gráficos, densidades, proximidades, similaridades, vetores, intensidades, centralidades e homogeneidades. A centralidade mede o quanto um vértice está ligado a outros e expressa a sua importância na rede (Liu et al., 2005). Assim, quanto mais centralizado é um nó na rede, maior deverá ser o seu prestígio e a sua reputação (Pujol et al., 2002).

Além do grau de centralidade, na análise da rede social pode-se avaliar o grau de proximidade de determinado vértice ou nó. Os vértices que apresentam menores distâncias de um determinado nó em relação aos demais, apresentam um grau maior de proximidade, e estando mais próximos dos demais, são mais acessíveis, o que significa uma posição favorável e vantajosa. Além da centralidade e da proximidade, na análise de rede social também é utilizado o grau de intermediação, o que significa a posição que determinado vértice ocupa em relação a um par de vértices. O nó que está posicionado entre todos os outros pares de nós da rede possui mais poder, pois outros nós dependem dele para realizar conexões (Maia and Caregnato, 2008).

Em redes que tem um grande número de membros, utiliza-se como estratégia começar a análise por um indivíduo e seus contatos mais diretos. O tamanho da rede é, portanto, importante para se definir a análise, bem como a importância dos papéis dos indivíduos nas relações e a multiplicidade, a qualidade, a intensidade, a durabilidade e a frequência das relações existentes (Acioli, 2007).

Nesse contexto, de acordo com Matheus and de Oliveira Silva (2009) a análise de redes sociais pode ser realizada de duas formas: análise individual (*ego network analysis*) e análise global (*global network analysis*). Na primeira forma, a rede de um membro é analisada, onde são identificados os indivíduos com os quais um determinado nó mantém relacionamento, a sua posição e a distância entre os nós na rede. Enquanto que na análise global, são analisadas as características da rede social como um todo, onde são analisados o diâmetro da rede, a distância média entre os nós e o número de subgrafos totalmente conectados.

Para este trabalho, será realizada a análise individual para o cálculo da reputação, pois para identificar a reputação de um nó serão analisados os indivíduos com os quais o nó está relacionado. O cálculo da reputação será abordado em detalhes no Capítulo 6.

2.3 Rede Social Científica

Nesta Seção é definido o conceito de Rede Social Científica, também conhecida como rede de colaboração científica. Para a melhor compreensão, serão apresentados os conceitos colaboração científica e ainda, a plataforma curricular do CNPq, visto que os currículos *Lattes* dos pesquisadores são a base para a construção da rede social científica do ObserveUnB.

2.3.1 Colaboração Científica

Colaboração, segundo a definição encontrada no Dicionário *on-line* da Língua Portuguesa (Michaelis, 2012) significa: trabalhar em comum com outrem, agir com outrem para a obtenção de determinado resultado e ter participação em obra coletiva, geralmente literária, cultural ou científica. Na concepção clássica de colaboração científica, dois cientistas colaboram quando compartilham dados, equipamentos e/ou idéias em um projeto, que resulta em geral, em experimentos e análises de pesquisa publicados (Katz and Martin, 1997).

Ainda segundo os autores, a colaboração científica pode ser definida como o trabalho conjunto de pesquisadores para atingir um objetivo comum que consiste em produzir novos conhecimentos científicos. O colaborador pode ser qualquer indivíduo que dá *input* para uma parte da pesquisa. Em outro extremo, também pode ser considerado colaborador aquele que contribuiu diretamente para a pesquisa durante todo o seu curso de duração. Os autores sugerem que sejam colaboradores: (i) os pesquisadores que trabalham juntos ao longo de um projeto ou durante parte considerável dele; (ii) os pesquisadores que fazem frequentes e substanciais contribuições; (iii) os pesquisadores cujos nomes aparecem no projeto de pesquisa original e (iv) os responsáveis por um ou mais elementos da pesquisa.

Em qualquer um dos casos, a lista de colaboradores deve incluir os responsáveis por qualquer passo-chave no projeto, seja a idéia original, as hipóteses ou a interpretação teórica. Entretanto são excluídos da lista os que fizeram somente uma parte pequena da pesquisa e aqueles que não são propriamente pesquisadores, como técnicos e assistentes. De acordo com Katz and Martin (1997), as fronteiras da colaboração variam consideravelmente entre instituições, campos de pesquisa, setores e países tanto quanto se modificam ao longo do tempo.

Para que ocorra a colaboração é importante o contato direto entre as pessoas. Porém, a colaboração acontece em diferentes níveis: entre pesquisadores de um mesmo departamento ou programa de pós-graduação; entre diferentes departamentos de uma mesma universidade; entre grupos de pesquisa e comunidades de uma área do conhecimento; entre diferentes instituições; entre setores da organização social, e.g., governo, universidade, empresa privada; entre regiões geográficas e países (Vanz and Stumpf, 2010).

Apesar das inúmeras possibilidades de colaboração científica, a proximidade física é um fator que encoraja a colaboração, visto que ela proporciona companhia para um trabalho que geralmente é isolado, impulsionando o pesquisador a se relacionar com outras pessoas. Assim, a colaboração frequentemente acontece entre professores e alunos, mesmo que esta não seja considerada por muitos autores como colaboração científica, dada a diferença de funções existente entre esses indivíduos.

Independente da posição dos indivíduos participantes, a transferência do conhecimento e das habilidades destaca-se como um dos maiores benefícios da colaboração (Katz and Martin, 1997). Em equipe, o conhecimento individual de cada pesquisador é socializado entre o grupo, facilitando a transferência de conhecimentos e habilidades, além de possibilitar o aprendizado de todos os participantes e estimular a criatividade e as ideias. Discutir diferentes pontos de vista pode gerar novas perspectivas, característica que se amplia quando os colaboradores são oriundos de diferentes áreas.

Além da descentralização do conhecimento, o aumento da produtividade científica pode ser considerado uma das principais motivações para a colaboração científica. A correlação entre colaboração e produtividade científica já foi comprovada em alguns estudos (Vanz, 2009) evidenciando que o princípio *publish or perish* aumenta a pressão para incluir tantos autores quanto possível em um artigo. É possível afirmar que a co-autoria de produtos gerados pela atividade científica, particularmente de publicações, é um indicador de colaboração.

De acordo com Maia and Caregnato (2008), estudos sobre co-autoria têm sido empregados para verificar características de colaboração em determinadas disciplinas e colaborações entre pesquisadores de mesma ou diferentes instituições ou países. Os resultados desses estudos revelam que as colaborações entre autores têm aumentado em todas as disciplinas. Entretanto, o grau de colaboração difere entre elas, ou seja, e.g. nas ciências naturais, não só o número de artigos em co-autoria como o número de autores por artigo é maior do que nas ciências sociais. Segundo os autores, os estudos sobre co-autoria e colaboração mostram que geralmente trabalhos teóricos geram artigos com menos autores do que aqueles de natureza experimental e demonstram que os trabalhos cooperativos internacionais têm maior impacto e visibilidade.

Associado aos trabalhos sobre co-autoria, atualmente também se observa um fortalecimento do método de análise de redes sociais nos estudos sobre colaboração científica, o que possibilita uma visão abrangente das interações entre as partes, diferentemente das análises que enfocam características individuais. As análises de redes sociais dão destaque para as relações, vínculos, influências e interações entre unidades, que podem representar tanto indivíduos quanto empresas, organizações, instituições ou nações, conforme apresentado na Seção 2.1.

A partir da literatura de diversos autores, Vanz (2009) propõe uma lista de motivações para a colaboração científica:

- aumento da popularidade científica, da visibilidade e do reconhecimento pessoal;
- aumento da produtividade;
- racionalização do uso da mão-de-obra científica e do tempo dispensado à pesquisa;
- redução da possibilidade de erro;
- obtenção ou ampliação de financiamentos, recursos, equipamentos e materiais;
- aumento da especialização na ciência;
- aumento da própria experiência através da experiência de outros cientistas;
- realização de pesquisa multidisciplinar;
- cooperação em detrimento da competição;

- divulgação de uma pesquisa e
- compartilhamento de entusiasmo, mérito e de problemas.

Seguindo os conceitos de rede social e de colaboração científica, pode-se definir rede de colaboração ou rede social científica como sendo redes de pesquisadores que colaboram entre si, ou seja, que trabalham em conjunto em prol da ciência ou ainda, que publicam em co-autoria com outros pesquisadores. Segundo Barabasi et al. (2002), redes de colaboração científica refletem as relações e as interações existentes entre os cientistas. Uma rede de co-autoria é um exemplo de um grafo não direcionado, a Figura 2.2 mostra a rede social científica de um pesquisador, onde o pesquisador Barabasi colabora com outros três pesquisadores da rede, e automaticamente esses pesquisadores colaboram com o pesquisador Barabasi.

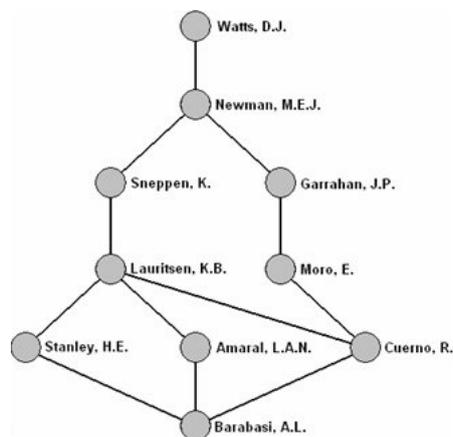


Figura 2.2: Rede social científica (Newman, 2003).

Newman (2003) afirma que, ao contrário de outros tipos de redes, as redes sociais científicas são divididas em comunidades locais onde as pessoas interagem mais do que a média encontrada em outras redes. De acordo com o autor, o processo de um cientista apresentar o colaborador a um terceiro é fundamental para o desenvolvimento das comunidades científicas. Nesse sentido, redes sociais científicas são uma importante classe de redes, sobre as quais técnicas de ARS tem sido utilizadas para explorar e aprofundar nos conceitos oferecidos por essas redes (Liu et al., 2005).

Nesse trabalho, técnicas computacionais de ARS foram aplicadas com o objetivo de identificar a importância ou reputação dos pesquisadores membros de uma rede social científica.

2.3.2 Plataforma *Lattes*

O CNPq, agência do Ministério da Ciência, Tecnologia e Inovação (MCTI), tem como principais atribuições fomentar a pesquisa científica e tecnológica e incentivar a formação de pesquisadores brasileiros. Criado em 1951, a agência desempenha papel primordial na formulação e condução das políticas de ciência, tecnologia e inovação (CNPq, 2012b).

Com o crescimento da comunidade acadêmica brasileira, o CNPq apresentou a necessidade de elaborar uma plataforma que concentrasse os currículos de seus pesquisadores. As principais preocupações eram integrar e padronizar as informações de currículo dos pesquisadores, facilitar a busca e a seleção de consultores e pesquisadores e gerar estatísticas para orientar as políticas públicas de educação. Com isso, no ano de 1999 foi criada a Plataforma *Lattes* a partir de um projeto desenvolvido pelas Universidades Federal de Santa Catarina e Federal de Pernambuco em parceria com empresas privadas (CNPq, 2012b).

A plataforma *Lattes* representa a integração de bases de dados de Currículos, de Grupos de pesquisa e de Instituições em um único sistema de informações. O Currículo *Lattes* se tornou um padrão nacional no registro da vida pregressa e atual dos docentes e discentes do país, e é adotado pela maioria das instituições de fomento, universidades e institutos de pesquisa no país (LATTES, 2012). De acordo com Mena-Chalco and Cesar-Jr. (2012), os currículos *Lattes* foram projetados para mostrar informação pública, embora, individual de cada usuário cadastrado na plataforma.

A concentração desses dados em apenas um lugar traz benefícios como a maior facilidade de controle nas alterações e a otimização nas buscas feitas por usuários, visto que a plataforma *Lattes* padroniza o formato de como são exibidas as informações. Os currículos *Lattes* são apresentados de uma única forma, porém a quantidade de dados pode variar para cada pesquisador.

O *Lattes* ainda auxilia no incentivo ao financiamento de programas universitários no Brasil, sendo considerado um recurso importante para as instituições que financiam a Ciência no país. A transparência, facilidade, a confiabilidade e a riqueza dos dados contidos nessa plataforma, tornaram-se elementos indispensáveis na análise de mérito e competência para o fomento à Ciência brasileira.

De acordo com Lane (2010), em 2010 a plataforma *Lattes* abrigava cerca de 1,6 milhões de currículos e 4 mil instituições cadastradas, sendo que 8% destes currículos são de doutores e cerca de 13% são de mestres. Segundo a autora, por criar incentivos adequados para que pesquisadores e instituições acadêmicas utilizem o banco de dados, por incentivar e avaliar de forma completa as pesquisas acadêmicas e por sua infraestrutura dinâmica e completa, o banco de dados *Lattes* é considerado um dos mais confiáveis do mundo atualmente.

Além dos Currículos, a plataforma *Lattes* abarca também o Diretório dos Grupos de Pesquisa em atividade do país. Os recursos humanos constituintes dos grupos, as linhas de pesquisa e os setores de atividade envolvidos, as especialidades do conhecimento, a produção científica, tecnológica e artística e os padrões de interação com o setor produtivo são algumas das informações contidas no Diretório (LATTES, 2012). Os grupos estão localizados em instituições de ensino superior, institutos de pesquisa e outras instituições. Segundo LATTES (2012), as informações individuais dos participantes dos grupos são extraídas dos seus Currículos *Lattes*.

E por último, o Diretório de Instituições foi englobado pela plataforma *Lattes* visando promover as organizações do Sistema Nacional de CT&I à condição de usuárias da Plataforma *Lattes*. O Diretório registra todas e quaisquer organizações ou entidades que estabelecem algum tipo de relacionamento com a agência, e.g., instituições nas quais os estudantes e pesquisadores apoiados pelo CNPq desenvolvem suas atividades, instituições onde os grupos de pesquisa estão abrigados, usuários de serviços prestados pelo CNPq

e instituições que pleiteiam participar dos programas e serviços oferecidos pela agência (LATTES, 2012).

Portanto, a disponibilização pública dos dados da Plataforma *Lattes* na *Web* permite maior transparência e confiabilidade às atividades de fomento do CNPq e das agências que a utilizam, fortalecendo o intercâmbio entre pesquisadores e instituições, sendo considerada uma fonte inesgotável de informações para estudos e pesquisas. Além disso, na medida em que suas informações são recorrentes e cumulativas, a plataforma adquire o importante papel de preservar a memória da atividade de pesquisa no país.

Neste capítulo foi apresentado uma revisão da área de Rede Social com foco na colaboração científica. No Capítulo 3 será apresentada uma breve revisão dos conceitos de reputação e seus respectivos modelos.

Capítulo 3

Reputação

O termo reputação é definido pelo dicionário Aurélio (de H. Ferreira, 1986) como sendo “o ato ou efeito de considerar, julgar, dar bom crédito a, avaliar, estimar e calcular”. Outro dicionário (Michaelis, 2012) considera a reputação como “fama, renome” ou como “conceito em que uma pessoa é tida com bom ou mau nome”.

Sabater and Sierra (2001) definem reputação como “opinião ou visão de alguém sobre algo” e apresenta três noções de reputação: individual, social e ontológica. Reputação individual refere-se a como as impressões de um único indivíduo são julgadas por outros. Reputação social refere-se a impressão sobre indivíduos baseado na reputação do grupo social a que pertencem. Reputação ontológica depende do contexto em que se encontra a reputação, onde a reputação de um indivíduo pode ter diferentes classificações em comunidades distintas.

Segundo Ferris et al. (2003) a reputação é formada dentro de um contexto específico, que inclui o conjunto de papéis do indivíduo, as normas sociais e o ambiente. As questões sobre os relacionamentos entre esses fatores e as formas pelas quais eles afetam a formação da reputação são subjetivas e complexas, pois dependem da percepção que as pessoas tem de uma dada situação.

Os sistemas de reputação procuram gerenciar as opiniões sobre o comportamento dos nós ou membros de uma rede e/ou a qualidade dos dados envolvendo estes nós. Estas opiniões são combinadas para formar a reputação, que é utilizada posteriormente para determinar a confiança do nó. A confiança é um importante aspecto do relacionamento entre duas entidades, sendo a base para a formação de coligações, ou seja, comunidades fortes são formadas por entidades que tem confiança entre si.

Para Josang et al. (2007) o conceito de reputação e confiança estão intimamente ligados. Dasgupta (2000) afirma que a confiança é baseada na reputação e que esta deve ser adquirida com base no comportamento através do tempo. Da mesma forma, a reputação também é adquirida com o tempo, porém ela deve ser pública. Para de Oliveira Albuquerque (2008), a reputação pode ser definida em um cenário em que não há informação suficiente para realizar a inferência sobre a confiança de um membro, e para descobrir se a entidade é confiável, outros membros da rede são questionados.

A reputação tem sido objeto de estudo de diversas áreas de pesquisa, tais como em Ciência da Computação e em Ciências Humanas. Apesar de a reputação ser um artefato antigo, é um conceito que vem sendo bastante utilizado, principalmente em comunidades eletrônicas, mais precisamente em redes sociais, visto que estas comunidades precisam das

noções de confiança e credibilidade a fim de fortalecer o comprometimento e a responsabilidade entre os parceiros.

No âmbito da ARS, a reputação é um valor derivado da medida de confiança dos membros de uma rede social, com base nas referências ou avaliações de uma comunidade, sendo globalmente visível para todos os membros da rede. Enquanto que, a confiança é uma análise pessoal e subjetiva que se baseia na experiência pessoal e nas referências de confiança dos demais membros (Josang et al., 2007). Adali et al. (2010) afirmam que a confiança pode servir para identificar membros ou nós influentes em uma rede, e determinar como a informação fluirá nela.

Outro conceito relacionado a reputação e confiança é a reciprocidade, que consiste em uma troca mútua de ações e realizações, tais como favores ou revanches. A reciprocidade pode ser considerada como um valor associado ao relacionamento entre dois membros da rede, quanto mais alto este valor maior a probabilidade de interação entre estes dois membros. Para Mui et al. (2002), a confiança, a reputação e a reciprocidade estão relacionadas da seguinte forma:

- um aumento na reputação de uma entidade ou membro pode melhorar a confiança que os demais membros possuem nele;
- um aumento na confiança depositada em uma entidade pode aumentar a probabilidade de esta agir em reciprocidade;
- um aumento das ações recíprocas pode melhorar a reputação de uma entidade.

A Figura 3.1 ilustra o relacionamento entre confiança, reputação e reciprocidade, onde a direção das setas indica o sentido da influência entre os conceitos.

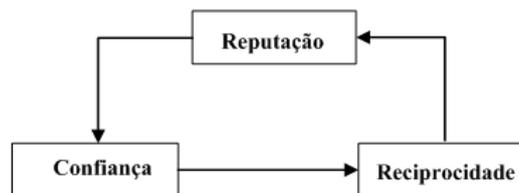


Figura 3.1: Representação dos relacionamentos entre confiança, reputação e reciprocidade. Traduzido de Mui et al. (2002).

Portanto, um sistema com modelo de reputação tem o objetivo de coletar, distribuir e agregar informações sobre o comportamento passado de seus membros, de maneira a ajudá-los a decidir em quem confiar, encorajando comportamentos leais e dissuadindo comportamentos desonestos (Resnick et al., 2000). Algumas vantagens do uso de sistemas de reputação são apresentadas na Seção 3.1.

3.1 Vantagens e Aplicação

Um dos grandes benefícios da Internet é o fato de ela permitir interações, tanto comerciais como não comerciais, entre as pessoas, organizações ou apenas entre computadores. No

entanto, o valor real de uma interação depende fortemente da confiabilidade entre as entidades envolvidas, e.g. no comércio eletrônico um comprador *on-line* pode obter melhores itens ou menor custo dos comerciantes, bem como pode ser lesado por um produto de baixa qualidade (Friedman et al., 2007).

Ainda segundo o autor, no caso de a confiabilidade ou reputação de cada entidade da rede se tornar pública ou visível, novas interações poderão ou não ocorrer em decorrência da reputação. Além disso, a expectativa de que o resultado da avaliação da entidade se torne visível no futuro pode impedir atitudes ilícitas e incentivar atitudes leais e honestas.

Sistemas de reputação foram projetados para uso em diversos ambientes e tem se tornado um componente principal no âmbito computacional. Segundo Cervi et al. (2011), a utilização de métodos de reputação pode-se dar na (i) detecção de clientes em sites de vendas *on-line*; (ii) recomendação de produtos; (iii) recomendação de especialistas para a participação de bancas; (iv) detecção de pesquisadores para a formação de rede de colaboração científica; (v) organização de comitê de programas científicos; (vi) organização de corpo editorial e também em *sites* de compartilhamento de notícias, *sites* de redes sociais e fóruns de mensagens.

Além dessas aplicações, a medida de reputação é bastante aplicada em comunidades criadas com o propósito do gerenciamento do conhecimento, uma vez que envolve o intercâmbio de informação e conhecimento entre membros da comunidade. O cálculo da reputação dá uma idéia da confiança que um membro pode ter em outro e serve para guiar a procura por especialistas (Pujol et al., 2002).

Em todos estes casos, a reputação desempenha um papel vital na garantia de qualidade das interações entre os usuários, sendo comum a sua utilização na *Web*. Na Seção 3.2 são descritos os componentes básicos de um sistema de reputação.

3.2 Componentes

Segundo Friedman et al. (2007), os sistemas de reputação podem ser divididos em diferentes componentes de acordo com aspectos do comportamento da rede social ou do usuário, entretanto, todos eles compartilham alguns componentes básicos. Da perspectiva do fluxo informacional, sistemas de reputação podem ser definidos como sistemas que usam tecnologias de rede para coletar experiências e opiniões dos usuários como “informação de reputação”, e então processar, armazenar e tornar públicas estas informações para os usuários (Liu et al., 2010). De acordo com Liu et al. (2010), um sistema de reputação é composto pelos componentes exibidos na Figura 3.2.

A Figura 3.2 apresenta os componentes básicos de um sistema de reputação, a saber:

1. Entrada - refere-se a coleta de informações utilizadas no cálculo da reputação e outros dados relacionados.
2. Processamento - o objetivo do processamento é aplicar a métrica de cálculo. A métrica define como os parâmetros referentes às transações são reunidos em um único valor indicando a reputação da entidade avaliada, assim como quais parâmetros são avaliados. Ela pode ser vista como uma função que mapeia o conjunto de parâmetros a uma imagem, cujo resultado pode ser utilizado para se comparar os nós. Na maioria dos casos, sistemas usam métricas simples, como o somatório ou média dos valores da reputação.

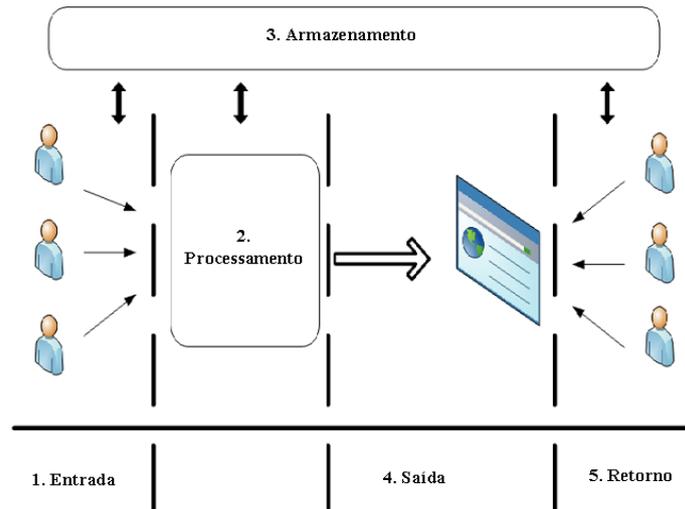


Figura 3.2: Estrutura de sistemas de reputação. Traduzido de Liu et al. (2010).

3. Armazenamento - as informações de reputação podem ser armazenadas de diferentes formas, ou seja, no nó avaliado ou no nó que fez a avaliação ou em algum repositório centralizado. Quando o armazenamento é feito em cada nó, torna-se mais difícil a identificação de uma determinada reputação, visto que é necessário uma busca na rede.
4. Saída - é o processo de se tornar público o valor da reputação.
5. Retorno - garante a acurácia do valor da reputação. Alguns sistemas de reputação permitem que os usuários opinem sobre o cálculo da reputação, e.g, a *Amazon* solicita aos usuários que avaliem sobre a reputação calculada e pede para classificá-la quanto a sua utilidade.

Conforme a identificação destes componentes, vários modelos de reputação podem ser construídos e por conseguinte, sistemas de reputação podem avaliar os membros de uma rede de diferentes maneiras. Algumas dessas métricas são abordadas nas Seções 3.3 e 3.4.

3.3 Tipologia

Baseado na revisão da literatura e em várias noções de reputação, Mui (2002) propôs uma tipologia de reputação, a qual é apresentada na Figura 3.3.

Segundo Mui (2002), os tipos de reputação podem ser:

- Individual e Grupo - a reputação pode indicar um indivíduo ou um grupo de indivíduos. Sistemas de reputação utilizados pelos sites *eBay*, *Amazon*, *Free Haven*, baseiam-se na reputação dos indivíduos. A reputação de grupo pode ser calculada como a média das reputações de seus membros. Sabater and Sierra (2001), em sua pesquisa, tem estudado este tipo de reputação que é inferida de um grupo de reputações individuais.

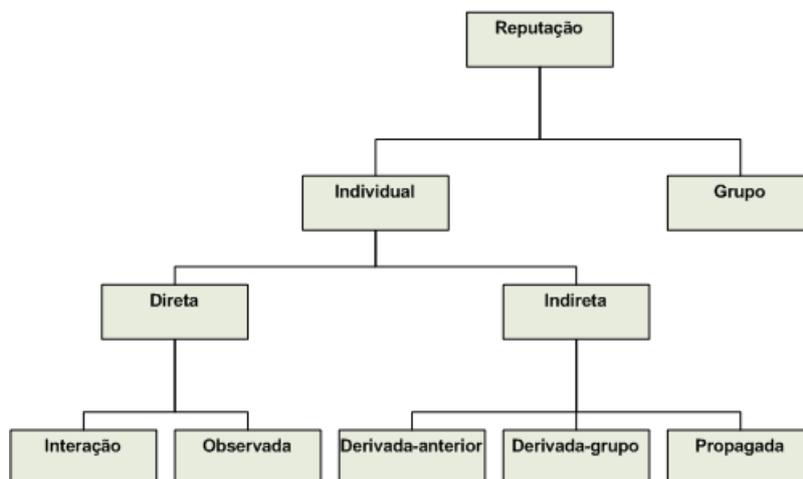


Figura 3.3: Tipologia de reputação, traduzido de Mui (2002).

- Direta e Indireta - pode-se considerar que a reputação individual pode ser obtida à partir de informações obtidas diretamente, ou à partir de inferências de informações obtidas indiretamente. A reputação direta refere-se à estimativas realizadas por um avaliador com base em experiências diretas, ou seja, experiências obtidas pelo avaliador. Reputação indireta refere-se à estimativas obtidas de terceiros, ou seja, experiências que não foram realizadas pelo avaliador.
- Observada - esta reputação corresponde a *feedbacks* de usuários sobre os outros membros, os quais eles interagem diretamente.
- Interação - reputação “observada” difere da “interação” pelo fato que esta última baseia-se em interações atuais entre o avaliador e o membro avaliado. O fator de impacto *Journal Impact Factor* determinado pela análise de citações é uma reputação “observada”. Entretanto, pesquisadores podem não concordar com o fator de impacto. Neste caso, cada pesquisador pode revisar a reputação com base no conhecimento que possui de cada publicação.
- Derivada-anterior - esta reputação considera o conhecimento prévio que um indivíduo possui sobre outro. Quando não possui conhecimento, o avaliador distribui um valor de forma uniforme a todos os avaliados, de forma que nenhum membro fique com baixa reputação por falta de conhecimento do agente avaliador.
- Derivada-grupo - este modelo de reputação provê estimativas de reputação para um indivíduo baseada na reputação do grupo o qual ele pertence. Sabater and Sierra (2001) pesquisaram diferentes mapeamentos entre a reputação individual de um membro desconhecido e o grupo do qual ele é proveniente.
- Propagada - neste tipo, a reputação de um indivíduo é calculada com base em informações transmitidas por outros indivíduos do mesmo grupo. Neste caso, a reputação pode ser transmitida de membro à membro.

Além da tipologia, os sistemas de reputação se diferenciam a partir de alguns critérios. De acordo com Casare and Sichman (2005), são descritos alguns desses critérios:

- Natureza da reputação - identifica que tipo de entidade a reputação se aplica, e.g., indivíduo, grupo, produto, serviço, organização;
- Papel - identifica quais os papéis que as entidades envolvidas na formação da reputação assumem, e.g., papel de alvo, de avaliador ou de transmissor da reputação;
- Tipo da reputação - conforme apresentado na Figura 3.3, classifica a reputação a partir da origem da informação utilizada na sua formação, podendo ser individual ou de grupo, e direta ou indireta;
- Valor da reputação - identifica o tipo do valor da reputação, podendo ser um valor consolidado ou um valor detalhado por meio de atributos;
- Fator de avaliação - identifica quais os fatores considerados na avaliação da reputação, e.g., a informação fornecida por uma entidade ou as relações existentes entre elas e,
- Manutenção da reputação - identifica o método utilizado para a propagação ou cálculo da reputação, e.g., uma entidade é responsável pela coleta e distribuição da informação ou esta pode ser realizada por qualquer nó.

Os tipos de reputação apresentados têm sido estudados e aplicados por diversos pesquisadores em sistemas atuais. Na Seção 3.4 são mostrados como estes tipos podem ser utilizados em modelos de cálculo de reputação.

3.4 Modelos de Cálculo

Diante da diversidade de aplicações e contextos, diferentes técnicas de cálculo de reputação tem sido implementadas, com resultados diversos. Segundo Pujol et al. (2002), o mecanismo usual de medida de reputação baseia-se no fornecimento de informações pelos usuários ou membros da rede. Estas informações indicam a autoridade ou reputação que um usuário pode ter, onde a cada membro da rede é associado um grau de autoridade que indica a sua importância na rede. Para o autor, esta autoridade pode ser propagada de um membro para outro, e a autoridade de um nó ou membro depende da autoridade dos nós relacionados a ele.

Este mecanismo permite que os membros avaliem a atuação dos indivíduos e identifiquem as reputações avaliadas perante a opinião da comunidade, criando assim suas redes de confiança.

A desvantagem deste mecanismo de mensuração é a necessidade do envolvimento explícito e freqüente dos usuários na rede. Isto implica que uma boa medição da reputação depende criticamente do envolvimento dos usuários e de sua contínua contribuição na mensuração.

Neste contexto, Cruz et al. (2007) descrevem três formas de calcular a reputação em serviços oferecidos na *Web*:

- Comércio eletrônico - em *sites* de comércio eletrônico (e.g., *Amazon*), os usuários avaliam produtos através de notas e comentários. Essas avaliações são utilizadas para recomendar produtos similares para os próprios usuários ou para outros clientes com preferências similares às deles. Para garantir a credibilidade das recomendações,

o sistema de reputação coleta a opinião dos usuários sobre as avaliações de produtos, onde eles indicam se uma avaliação foi útil ou não na sua decisão de compra. Com base nessas avaliações, os clientes podem adicionar avaliadores às suas redes de confiança ou bloquear avaliadores cujas opiniões eles não confiam.

- Leilão - os *sites* de leilão *Ebay* e MercadoLivre coletam informações (*feedback*) dos usuários sobre as negociações realizadas. O usuário avalia sua transação de forma positiva (+1), negativa (-1), ou neutra (0). A reputação de um comprador ou vendedor é representada pelo saldo de avaliações positivas e negativas que ele recebe de diferentes usuários com quem negociou. Se um vendedor recebe várias avaliações repetidas de um mesmo comprador, o sistema só contabiliza uma dessas avaliações. Isso garante que a reputação do vendedor seja um reflexo da opinião de diferentes compradores que negociaram com ele e não de apenas um.
- Notícias - em *sites* de notícia (e.g. *Slashdot*), os usuários postam e comentam notícias. Os comentários podem ser avaliados por todos os outros usuários (moderadores) através de notas que contam pontos positivos ou negativos para o autor do comentário. As avaliações dos moderadores são julgadas por usuários mais antigos (meta-moderadores), para minimizar a ação de moderadores injustos ou desonestos. Assim, usuários com maus comportamentos perdem pontos de reputação e privilégio de moderação.

Um método alternativo de medir a reputação baseia-se na localização de cada membro em sua comunidade da rede social. A localização de um dado membro na rede social pode ser usada para inferir alguma propriedade sobre seu grau de experiência, i.e. reputação. Membros que são bem conhecidos e respeitados pela maioria dos outros membros da comunidade, tendem a ser facilmente identificados como os nós mais conectados em um grafo de rede social da sua comunidade. Esta informação pode ser uma base para o mecanismo de reputação usado ao invés de capturar graus de autoridade distribuídas por cada usuário. Alguns algoritmos usados na classificação de páginas na *Web* baseiam-se neste método de reputação, podemos citar o *HITS* (Kleinberg, 1999), o *PageRanking* (Page et al., 1999) e o *NodeRanking* (Pujol et al., 2002).

Independente do método de reputação, segundo Resnick et al. (2000) um sistema de reputação deve possuir algumas características básicas:

- Cada membro deve ter um tempo mínimo de vida, a fim de que as interações entre eles sirvam de instrumento para expectativas futuras;
- Deve-se levar em conta a disposição dos usuários em informarem dados para o cálculo da reputação, atividade para a qual certas vezes é necessário oferecer incentivo;
- O histórico da reputação de um membro deve influenciar decisões atuais.

Baseado no que foi exposto, não existe uma fórmula única para que sejam construídos sistemas de reputação. O fato de um modelo se adaptar a uma rede não significa que possa servir a outras. Basta considerar que os propósitos de cada rede nem sempre são iguais, assim como o tipo de relação que se estabelece entre os membros de uma rede.

3.5 Modelos de Classificação

Para facilitar a identificação da reputação do usuário ou membro em uma comunidade, pode-se atribuí-la a uma hierarquia de níveis, onde a reputação pode ser classificada de diferentes maneiras. De acordo com Lima (2010) a reputação pode ser classificada como:

- Baseada em Perfil - este modelo de classificação é baseado na existência de perfis de usuário, onde cada usuário pertence a um perfil que pode ser alterado conforme a participação do mesmo na comunidade. Este modelo estimula encontros na rede e a cooperação entre os usuários devido à união de usuários com perfis semelhantes. É recomendado para modelos de reputação que dependem da interação do usuário. O *site* de vendas MercadoLivre classifica a reputação de seus usuários em perfis conforme é exibida na Figura 3.4.

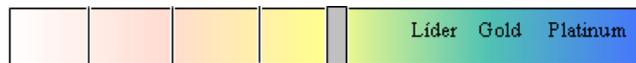


Figura 3.4: Classificação baseada em perfis.

- Baseada em níveis numéricos - este modelo é utilizado para visualizar a reputação do usuário e dar a idéia de sua evolução na comunidade. Esse tipo de classificação se adapta melhor à comunidades mais competitivas que a modalidade anterior devido ao fato de a comparação entre números ser mais intuitiva que a comparação entre nomes de categorias. A Figura 3.5 exhibe uma classificação baseada em níveis numéricos.

Nível	Pontos
0	0
1	1 até 750
2	751 até 2000
3	2001 até 7500
4	7501 até 15000
5	Mais de 15000

Figura 3.5: Classificação baseada em níveis.

- Baseada em identificação de membros - utilizado em comunidades que fazem uso de rótulos, ao invés de números ou perfis, para identificação de membros a fim de legitimar os usuários como confiáveis ou portadores de certos papéis dentro daquele contexto. Dessa forma, ao invés de um usuário pertencer a um perfil denominado “Gold” ou a um nível numérico “Nível 5”, é atribuído a ele um ou mais rótulos que claramente o distingue dos demais usuários, e.g., “Contribuidor Top” ou “Membro da elite”. Este modelo é muito utilizado para distinguir representantes de uma comunidade e usuários de um sistema conforme é apresentado na Figura 3.6.
- Baseada em pontos - este modelo geralmente é usado em conjunto com outros modelos de classificação como “Baseado em Perfil” e “Baseado em níveis numéricos”, e é

Tipos de Usuários de um Sistema
Operador
Usuário
Gerente
Administrador

Figura 3.6: Classificação baseada em membros.

indicada a comunidades bastante competitivas, e.g., jogos de esportes. A Figura 3.7 apresenta uma classificação dos clubes de futebol brasileiro baseada na pontuação obtida pelos clubes em cada campeonato brasileiro do ano de 2009.

Clubes Brasileiros	240	180	090	060	045	024	022	020	016	015	014	012	010	009	007	006	005	004	003	002	Total
1 São Paulo	03	03	06	02	00	00	02	00	01	00	00	00	00	22	00	00	00	00	00	00	2178
2 Santos	02	03	03	01	06	00	00	00	03	00	00	00	00	19	00	00	00	00	00	00	1839
3 Flamengo	01	01	06	01	02	01	00	01	01	00	00	00	00	31	00	00	00	00	00	00	1449
4 Internacional	01	02	03	01	01	00	02	00	00	00	00	00	00	00	39	00	00	00	00	00	1292
5 Grêmio	01	02	02	00	04	00	00	01	00	00	00	00	00	00	36	00	00	00	00	00	1234
6 Palmeiras	00	01	06	01	03	00	00	01	05	00	00	00	00	22	00	00	00	00	00	00	1213
7 Cruzeiro	00	02	01	02	05	01	01	00	00	00	02	00	01	00	35	00	00	00	00	00	1124
8 Corinthians	01	00	05	00	03	00	00	00	02	00	00	00	00	26	00	00	00	00	00	00	1091
9 Vasco	00	02	04	01	01	00	00	00	02	00	00	00	00	22	00	00	00	00	00	00	1055
10 Fluminense	00	00	03	00	01	00	00	00	02	00	00	00	00	30	00	00	00	00	00	00	617
11 Atlético Mineiro	00	00	01	02	00	00	00	00	00	00	00	00	00	40	00	00	00	00	00	00	490
12 Botafogo	00	00	02	01	00	00	00	00	04	00	00	00	00	19	00	00	00	00	00	00	475
13 Bahia	00	00	01	00	01	00	00	00	00	00	03	00	00	00	43	00	00	00	00	00	429
14 Sport	00	00	00	00	01	00	00	01	00	01	00	02	00	00	39	00	00	00	00	00	338
15 Coritiba	00	00	01	00	00	00	00	00	00	00	00	00	00	00	34	00	00	00	00	00	294
16 Atlético Paranaense	00	00	01	00	00	00	00	00	00	01	00	00	00	00	22	00	00	00	00	00	237
17 Piazinho	00	00	00	00	00	00	01	00	02	00	00	01	00	00	00	00	00	44	00	00	236
18 Vitória	00	00	00	00	00	00	00	00	00	00	04	00	00	00	26	00	00	00	00	00	204
19 Cianã	00	00	00	00	00	00	00	00	01	00	00	03	00	00	00	00	22	00	00	00	133
20 Criciúma	00	00	00	00	01	00	00	00	01	00	00	00	00	00	09	00	00	00	00	00	114

Figura 3.7: Classificação baseada em pontos.

- Baseada em *ranking* - é indicado a comunidades altamente competitivas, onde os membros se interessam basicamente em saber sua posição em relação aos outros membros, ou se pertencem a um grupo diferenciado de membros. Este tipo de classificação é utilizado em conjunto ao sistema baseado em pontos e diferencia-se dos demais, pois sempre destaca os primeiros classificados no *ranking*, conforme apresentado na Figura 3.8.

3.6 Problemas

Em geral, problemas em sistemas com modelo de reputação podem ocorrer devido ao mau uso do sistema. Josang et al. (2007) apresentam diversos destes problemas, dentre os quais podem ser citados:

- Falta de incentivo - em sistemas onde o membro da rede participa ativamente da avaliação de um outro nó, por poucos benefícios o usuário pode não realizar a avaliação. Essa falta de retorno ao sistema pode diminuir a precisão das informações;



Figura 3.8: Classificação baseada em *ranking*.

- Avaliações injustas - podem ocorrer no caso em que nós maliciosos podem realizar avaliações erradas intencionalmente, com o objetivo de diminuir a reputação do nó avaliado. E também no caso em que membros são mal avaliados a partir do processamento da métrica de cálculo de reputação;
- Inconsistência - a falta de informações ou a inconsistência delas pode influenciar no valor da reputação;
- Inflação da reputação - ocorre em casos em que um nó adultera as suas informações com o objetivo de aumentar a sua reputação;
- Traição - este problema ocorre quando um nó se comporta bem até obter uma reputação alta, a partir disso o nó muda de comportamento de forma a prejudicar os outros nós.

Neste capítulo foi realizada uma breve introdução sobre temas relacionados a reputação, sua tipologia e estrutura, bem como sua aplicabilidade em diferentes sistemas computacionais. No Capítulo 4 serão elucidados conceitos relacionados à ontologia e sua aplicação na área da Computação.

Capítulo 4

Ontologia

Um dos desafios encontrados no desenvolvimento de sistemas é fornecer aos usuários a capacidade de consultar as diferentes fontes de informação de forma eficiente e eficaz sem obrigá-los a lidar com os aspectos de heterogeneidade de estrutura, representação, vocabulário, além da inconsistência de dados.

A complexa questão do significado dos dados e sua descrição são apresentadas por Fonseca and Egenhofer (1999), onde são apresentados três tipos de heterogeneidade:

- Esquemática - onde um objeto do mundo real é representado por diferentes conceitos em um banco de dados;
- Sintática - onde os bancos de dados usam diferentes paradigmas e
- Semântica - onde um fato pode ter mais de uma descrição.

Desta forma, diversos estudos e técnicas têm sido desenvolvidos com o intuito de disponibilizar a informação de forma compartilhada e reutilizável para satisfazer as consultas do usuário. Uma das técnicas, denominada *Otimização Semântica de Consultas*, objetiva reformular uma consulta em outra mais eficiente, a qual é semanticamente equivalente, ou seja, as consultas fornecem a mesma resposta (Necib and Freytag, 2004).

Dentre as técnicas de *Otimização Semântica de Consultas*, destaca-se o uso de ontologias. O termo ontologia tem origem no grego *ontos*, ser, e *logos*, palavra. Originalmente vem da palavra “categoria”, definida por Aristóteles, e de uma das áreas da Filosofia, podendo ser usada para classificar coisas. O Dicionário Michaelis (2012) define ontologia como “Ciência do ser em geral”. Quando se trata de linguagem e cognição, ontologia refere-se a tudo que existe no mundo composto por objetos, mudanças e relações entre eles.

Esse termo foi também adotado pelas comunidades de Ciência da Computação e Ciência da Informação para se referir a conceitos e termos que podem ser usados para descrever alguma área do conhecimento ou construir uma representação deste. Em organização da informação, ontologia é um “catálogo de tipos de coisas” em que se supõe existir um domínio, na perspectiva de uma pessoa que usa uma determinada linguagem. Trata-se de “uma teoria que diz respeito a tipos de entidades e, especificamente, a tipos de entidades abstratas que são aceitas em um sistema com uma linguagem” (Almeida and Bax, 2003).

Entre as várias definições de ontologia existentes, uma muito utilizada é apresentada por Gruber (1996): ontologia é uma especificação formal e explícita de uma conceitualização compartilhada. Onde: (i) “especificação formal” diz respeito a algo que seja passível de

ser processado por computadores; (ii) “explícita” são os conceitos, propriedades, relações definidos de forma explícita; e (iii) “conceitualização compartilhada” pode ser entendida como a conceitualização consensual, ou seja, o conhecimento de um domínio é aceito por um grupo de pessoas.

Guarino (1998) define uma ontologia como sendo um artefato computacional, baseado em um vocabulário formal, utilizado para descrever uma conceitualização particular comumente aceita do mundo. E segundo Almeida and Bax (2003), as ontologias definem uma linguagem por meio de um conjunto de termos para ser utilizada na formulação de consultas, por meio dos quais os agentes de software terão acesso no momento da inferência dos dados.

Na prática, uma ontologia fornece um conjunto de termos e conceitos e seus relacionamentos semânticos. Estes relacionamentos são criados por especialistas, e toda esta estrutura é utilizada por usuários durante a formulação de uma consulta. Graficamente, as ontologias parecem árvores de conceitos, em que cada folha é um termo específico relacionado a um galho, que é um conceito mais geral. Sendo assim, dependendo do domínio em questão, uma mesma entidade pode ser caracterizada de diferentes maneiras (Bräscher and Carlan, 2010).

Ao analisar o conceito de manga em diferentes domínios, determinadas características são ressaltadas. Na Biologia trata-se de uma fruta tropical¹. Pode-se, por exemplo, estabelecer relação entre manga e plantação, no contexto da biologia, já na área de vestuário, entre manga e camisa. Evidentemente que o conceito completo de manga contempla essas e outras características. No entanto, cada domínio destaca aquelas que são particularmente importantes.

A Figura 4.1 apresenta uma pequena amostra das classes presentes na ontologia *Conscientias* utilizada neste trabalho (Conscientias, 2006). Note que a classe produção bibliográfica e suas subclasses foram criadas para contextualizar os indivíduos, e.g., livros publicados ou organizados (*LivrosPublicadosOuOrganizados*), trabalhos em eventos (*TrabalhosEmEventos*), detalhes (*Detalhamento*), capítulos de livros publicados (*CapitulosDeLivrosPublicados*), artigos (*Artigos*), demais tipos de produção bibliográfica (*DemaisTiposDeProducaoBibliografica*) e textos em jornais ou revistas (*TextosEmJornaisOuRevistas*).

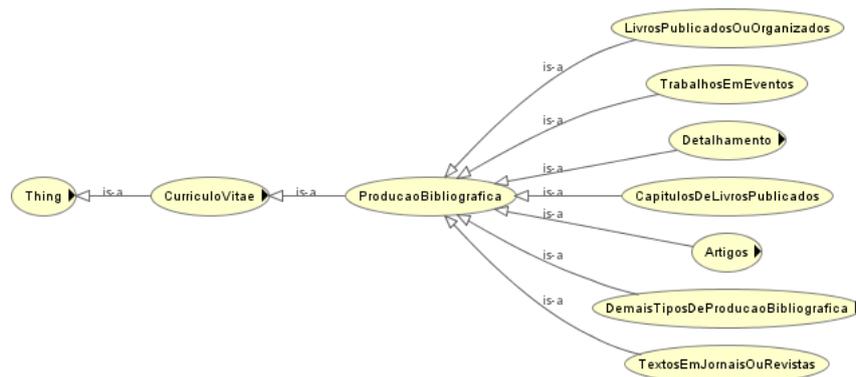


Figura 4.1: Amostra de classes da ontologia Conscientias (Conscientias, 2006).

¹url: <http://pt.wikipedia.org/wiki/Manga>

4.1 Vantagens

Segundo Guimarães (2002) existem muitas vantagens no uso de ontologias em sistemas computacionais, a saber:

- Compartilhamento do conhecimento - a ontologia de um determinado domínio de conhecimento pode ser compartilhada e utilizada por outras aplicações deste mesmo domínio, permitindo o reuso e aumentando a comunicação entre pessoas e organizações. A utilização das ontologias de Diretório de Grupos de Pesquisas e de Currículo *Lattes* do Conscientias (Conscientias, 2002, 2006) neste trabalho pode ser considerado um exemplo;
- Representação do conhecimento - conceitos são representados por vocabulários de forma a evitar interpretações ambíguas, facilitando a interoperabilidade entre os sistemas;
- Precisão nas consultas - a aplicação pode fazer inferências em uma consulta por meio dos relacionamentos definidos entre os conceitos da ontologia, retornando ao usuário as informações que realmente são do seu interesse.

A aplicação de ontologias pode ser vista em diferentes áreas do conhecimento:

- Na Inteligência Artificial - mais especificamente no processamento da linguagem natural, o conhecimento do domínio é muito importante para uma compreensão coerente do texto. Esse conhecimento do domínio pode ser obtido por meio de uma ontologia sobre o discurso do texto. Além disso, a ontologia auxilia na desambiguação de termos e funciona como um dicionário de conceitos dentro do domínio do texto. Para auxiliar na desambiguação e na deduplicação de dados, podem ser utilizadas funções de similaridade que calculam o quanto duas entidades são similares através dos relacionamentos e de termos extraídos dos conceitos da ontologia (Morris et al., 2008) ou simplesmente restrições de funcionalidade podem ser aplicadas aos conceitos da ontologia.
- Na gestão do conhecimento - a recuperação da informação de forma rápida e eficiente é de vital importância dentro de uma organização. Nesta área, o uso de ontologias se mostra bastante útil pois fornece a estrutura básica para se organizar o conhecimento de forma estruturada, facilitando o acesso à informação.
- No comércio eletrônico - o uso de uma ontologia permite o entendimento comum dos termos, facilitando as transações e a troca de informações e a integração de catálogos entre as lojas.
- Na *Web Semântica* - onde uma das dificuldades existentes na *Web* consiste em obter resultados precisos em buscas realizadas via Internet. As ferramentas de busca que existem atualmente não conseguem descobrir o significado preciso de determinada página na *Web*. A *Web Semântica* pesquisa como resolver este tipo de problema adicionando semântica às páginas da *Web*, sendo que o uso de ontologias é uma das propostas em uso.

Atualmente existem vários projetos de diferentes áreas de aplicação que fazem uso de ontologias. A Tabela 4.1 apresenta alguns desses projetos, segundo Thiago Semprebom and Mendonça (2007).

Tabela 4.1: Projetos que fazem uso de ontologias.

Projeto	Área de Utilização
PlanetOnto	Gestão de Conhecimento
MKBEEEM	Comércio Eletrônico
Oncoterm	Processamento de Linguagens Naturais
OntoSeek	Web Semântica
Smartrainer	Educação

4.2 Elementos Básicos

Segundo Gruber (1996), uma ontologia é formalizada através do uso de cinco tipos de componentes: classes, relações, axiomas, funções e instâncias. Alguns autores não consideram as instâncias como parte de uma ontologia (Maedche, 2002). De acordo com a estrutura de Gruber (1996), Gómez-Pérez (1999) descreve os componentes de uma ontologia:

- Classes - são os conceitos organizados em estrutura hierárquica formando uma taxonomia, onde uma classe agrupa elementos com características similares, e.g., um artigo científico é um subconceito de publicação bibliográfica;
- Relações - representam o tipo de interação entre os conceitos de um domínio, e.g., a relação existente entre os conceitos pesquisador e grupo de pesquisa é o relacionamento pertence-à;
- Axiomas - usados para modelar sentenças sempre verdadeiras, e.g., todo orientador tem um aluno;
- Funções - uma função é um caso especial de relacionamento em que um elemento de uma relação tem uma relação única com outro elemento, e.g., Inteligência Artificial é sub-área de Ciência da Computação;
- Instâncias - utilizadas para representar os elementos de uma classe, e.g., Márcia Falcão é uma instância da classe autores.

4.3 Classificação

São propostas mais de uma classificação para ontologia. De acordo com Maedche (2002), são apresentados os quatro tipos propostos de classificação de uma ontologia:

- Ontologias de alto nível - descrevem conceitos mais gerais como tempo, espaço, evento. Esse tipo de ontologia oferece maior capacidade de reuso por ser mais genérica.
- Ontologias de domínio - descrevem o conhecimento relacionado a um determinado escopo, e.g., veículos e imóveis.
- Ontologias de tarefa - descrevem um conhecimento relacionado a uma tarefa ou atividade, e.g., ciclo de vida de processos de um tribunal, gerenciamento de projetos.

- Ontologias de aplicação - são as mais específicas. Esta ontologia especializa conceitos das ontologias de domínio e de tarefa, e.g., a ontologia estendida do grupo Conscientias especializa ontologias de Diretório de Grupos de Pesquisa e de Currículo *Lattes* do Conscientias (Conscientias, 2002, 2006). Por serem muito específicas, estas ontologias são as que possuem menor capacidade de reuso.

A Figura 4.2 apresenta a classificação mais conhecida de ontologias segundo Guarino (1998) e mostra a dependência entre os níveis de classificação, na qual as setas representam relações de especialização.

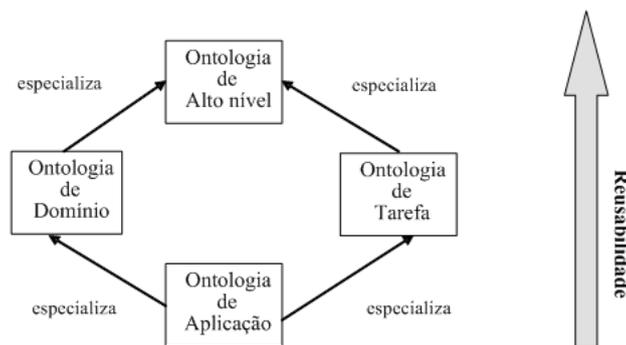


Figura 4.2: Diferentes tipos de ontologia e seus relacionamentos. Traduzido de Guarino (1998).

4.4 Integração

Um dos problemas no desenvolvimento de aplicações com repositório semântico consiste na integração, também denominado *merge*, de ontologias. A *Web* é composta de uma variedade de informações, e quando há necessidade de extrair informações de diferentes origens, a integração é necessária (Calvanese and Lenzerini, 2001).

Na integração de ontologias uma nova ontologia é criada a partir de duas ou mais ontologias com partes delas se sobrepondo, ou então as ontologias podem ser usadas de forma disjunta. Em ambos os casos as ontologias precisam ser alinhadas tornando-se consistentes e coerentes. O alinhamento de conceitos entre ontologias é especialmente difícil porque ele requer compreensão do significado dos conceitos. Assim, alinhar duas ontologias significa em mudanças em pelo menos uma delas, o que vai resultar em uma nova versão da ontologia que foi alterada. Para exemplificar, pode-se citar a integração das ontologias de Diretório de Grupos de Pesquisa e de Currículo *Lattes* da comunidade Conscientias (Conscientias, 2002, 2006) realizada neste trabalho.

Segundo Noy (2004), integrar ontologias consiste em realizar o mapeamento sintático e semântico através dos passos:

- Encontrar os termos nas ontologias que se sobrepõem;

- Relacionar conceitos que são semanticamente equivalentes;
- Checar a consistência, coerência e a não redundância de resultados.

4.5 Linguagens

Para a construção de ontologias existem diversas linguagens, cada uma fornecendo diferentes funcionalidades e fazendo uso de diferentes paradigmas de representação do conhecimento (Horridge, 2011). Segundo Su and Ilebrikke (2002), os três paradigmas de representação mais utilizados pelas linguagens de implementação de ontologias são: o cálculo de predicados de primeira ordem, as linguagens baseadas em *frames* e em orientação a objetos, e as lógicas descritivas.

O cálculo de predicados de primeira ordem utiliza fatos, objetos e relações para representar um domínio. Em cálculo de predicados, o fato que representa “a produção de Célia é alta” pode ser descrito através do predicado binário Produção (Célia, Alta), onde Célia e Alta são as constantes que representam os objetos e o predicado Produção representa a relação entre estes objetos (Su and Ilebrikke, 2002).

Ainda segundo os autores, em linguagens baseadas em *frames* e em orientação a objetos, um *frame* oferece o contexto para a modelagem de um determinado aspecto de um domínio. Neste caso, o exemplo anterior poderia ser representado por meio do *frame* Produção{alvo: Célia; avaliação: Alta}, onde Célia e Alta são os valores dos atributos alvo e avaliação.

O último paradigma, as lógicas descritivas (*Description Logics -DL*), representam um alto grau de expressividade, permitindo a representação do conhecimento bem como o raciocínio automático sobre esse conhecimento. As lógicas descritivas oferecem um conjunto de construtores que permitem a representação de conceitos e papéis. Os conceitos, que correspondem às classes, são interpretados como conjunto de objetos, enquanto os papéis, que correspondem às relações são interpretados como relações binárias entre os objetos. Neste caso, o fato que descreve o exemplo citado poderia ser representado pela classe Autor, que representa o conjunto de objetos que são pessoas, a classe Intensidade que representa o conjunto de objetos que classificam a intensidade da produção (baixa, média, alta) e pela relação “Tem Produção” entre os objetos “Célia” e “Alta” (Casare and Sichman, 2005).

De acordo com os autores, uma das principais características das lógicas descritivas é o processo de inferência realizada pelo raciocinador ou *Reasoner*. Posea and Harzallah (2004) afirmam que existem dois métodos de inferência: a inferência interna e a externa. Na inferência interna, o sistema faz a validação dos conceitos (detecção de inconsistências) e a classificação automática dos conceitos baseado nas características da ontologia, ou seja, verifica se uma classe é subconjunto de outras classes através da comparação de suas definições e da classificação de instâncias. A inferência externa é realizada explorando-se a estrutura de conceitos da ontologia, fazendo uso da similaridade semântica.

A distância semântica entre os conceitos é obtida por meio da quantidade de relacionamentos existentes entre os conceitos no grafo da ontologia, bem como pelos tipos de relacionamentos ou pela distância da próxima classe ancestral comum.

Várias linguagens, que seguem os paradigmas descritos, foram criadas para expressar formalmente as ontologias e algumas delas vem sendo estudadas de tal forma que

representem a ontologia de maneira fácil e eficiente. O W3C (*World Wide Web Consortium*) vem liderando a pesquisa das linguagens para ontologias. Inicialmente foram propostas as linguagens XML (*eXtensible Markup Language*) e RDF (*Resource Description Framework*)/RDFS (*RDF Schema*) como uma linguagem de ontologia primitiva. A linguagem OWL (*Ontology Web Language*) que herda a sintaxe do XML e a semântica do RDF, é uma linguagem descritiva recomendada pela W3C visto que fornece um rico e expressivo vocabulário (Kim and Kim, 2008).

Serão apresentadas algumas destas linguagens, sendo descrito de forma detalhada a OWL, utilizada na extensão da ontologia de Currículo Lattes e Diretório de Grupo de Pesquisa neste projeto.

4.5.1 XML

A linguagem XML foi criada para carregar e transportar dados, permitindo também o desenvolvimento de documentos em formatos próprios (Bray et al., 2008). Assim como documentos em HTML (*HyperText Markup Language*), documentos XML contém texto. Esse texto contém primeiramente o conteúdo do documento e a linguagem de marcação na forma de *tags*. Essa marcação permite um programa interpretar as diversas partes do conteúdo como “elementos”. A linguagem XML permite a seus usuários definir suas próprias linguagens de marcação (*tags* e as estruturas nas quais elas podem aparecer) adaptadas para seus próprios requisitos específicos.

Opcionalmente, um documento XML pode incluir uma declaração de tipo de documento XML para definir condições adicionais na estrutura do documento e dar suporte ao uso de unidades pré-definidas de texto com o documento. A declaração de tipo de documento (introduzido com *doctype*) contém ou aponta para declarações que definem uma gramática para o documento, que é conhecida como Definição de Tipo de Documento (DTD).

As declarações em DTD especificam algumas regras como quais elementos XML e atributos podem aparecer em documentos XML correspondentes a essa DTD, quais os relacionamentos entre esses elementos e atributos e quais destes elementos ou atributos são obrigatórios. A declaração de tipo de documento pode apontar para um conjunto de declarações localizadas fora do documento (chamado subconjunto externo, que pode ser usado para permitir que declarações comuns sejam compartilhadas através de vários documentos), pode incluir as declarações diretamente no documento (chamado de subconjunto interno), ou pode ter subconjuntos internos e externos. As ontologias de Currículo Lattes e Diretório de Grupo de Pesquisa do grupo Conscientias foram disponibilizadas na linguagem XML.

4.5.2 RDF

O RDF é uma linguagem para troca de dados na *Web* desenvolvida pelo W3C, a qual possui características que facilitam esta troca de informação e suporta a evolução e mudanças de dados (Klyne and Carroll, 2004). O RDF foi desenvolvido para situações nas quais a informação precisa ser processada por aplicações, apesar de ser visualizada apenas por pessoas. Esta linguagem provê um *framework* comum para expressar essa informação de forma que ela possa ser trocada entre aplicações sem perdas de significado. A habilidade

de trocar informações entre aplicações significa que a informação pode ser disponibilizada para aplicações diferentes daquelas para as quais a informação foi criada originalmente.

A linguagem é baseada em *frames*, na qual um domínio de conhecimento é representado em uma coleção de triplas, onde cada uma consiste em um sujeito, um predicado e um objeto. Cada tripla representa um relacionamento entre os conceitos representados na ontologia e o conjunto dessas triplas formam um grafo direcionado em RDF. Em um grafo RDF os nós representam os sujeitos e objetos, e os arcos representam o predicado, onde a direção do arco é importante, sempre aponta para o objeto.

4.5.3 OWL

A OWL é uma linguagem baseada em lógica descritiva, sendo a mais recente desenvolvida pelo W3C (Horridge, 2011). Uma ontologia em OWL contém classes, propriedades e instâncias de classes. A definição das classes e das propriedades em OWL pode envolver primitivas de RDF e de esquemas RDF/RDFS. Nestes casos, são utilizados os prefixos `rdf` e `rdfs` no nome do construtor, e.g. o termo `rdfs:subClassOf` indica que o construtor `subClassOf` faz parte do vocabulário de esquema RDF, enquanto o termo `owl:Class` indica que o construtor é primitivo de OWL.

A OWL é considerada a linguagem mais completa visto que permite descrever mais propriedades, relacionamentos e classes que outras linguagens permitem: relações inter-classes (e.g. disjunção), cardinalidade da classe (e.g. exatamente um), eqüidade, tipagem de propriedades mais rica, características de propriedades (e.g. simetria), operadores (e.g. interseção, união e negação), classes enumeradas. É possível também estabelecer restrições de propriedades herdadas das superclasses pelas subclasse, bem como estabelecer que duas classes são disjuntas, onde uma instância não pode ser membro das duas classes.

A seguir são descritos como os principais componentes de uma ontologia podem ser representados em OWL.

1. Classes em OWL

Em OWL todas as classes são subclasses da classe *Thing* e representam um conjunto de indivíduos. Uma classe é definida através de descrições de classe, que são combinadas em axiomas de classe. Conforme Horridge (2011), para definir uma classe podem ser usadas as seguintes funcionalidades e respectivos construtores²:

- Subclasse (`rdfs:subClassOf`) - usado para criar hierarquia de classes;
- Equivalência (`owl:equivalentClass`) - usado para definir que duas classes são equivalentes, ou seja, possuem as mesmas instâncias;
- Enumeração (`owl:oneOf`) - permite definir uma classe por meio da enumeração das instâncias que a formam;
- Disjunção (`owl:disjointWith`) - define que duas classes não podem ter nenhuma instância comum;

²A expressão que aparece entre parênteses após o nome da funcionalidade corresponde ao construtor associado à mesma, podendo pertencer originalmente ao RDF, aos esquemas RDF, ou ainda ser uma primitiva OWL

- União (owl:unionOf), interseção (owl:intersectionOf) e complemento de classe (owl:complementOf) - usados para definir uma classe por meio da combinação de outras classes.

2. Propriedades em OWL

As propriedades em OWL representam o relacionamento entre dois indivíduos que podem pertencer a duas categorias principais: propriedade de objeto (*object properties*) e propriedade de tipo de dado (*datatype properties*). As propriedades de objeto relacionam dois indivíduos, e.g. o relacionamento entre duas instâncias das classes *Autores* e *curriculoVitae* através da propriedade *temCurriculo*. As propriedades de tipo de dado relacionam indivíduos aos tipos de dados, que podem ser tanto literais como tipos de dados definidos de acordo com esquemas XML, e.g. string (xsd:string), data e hora (xsd:dateTime). De acordo com Horridge (2011), as funcionalidades e respectivos construtores podem ser aplicados às propriedades de uma ontologia, sendo que somente as funcionalidades *Funcional inversa* e *Range* podem ser aplicadas às propriedades de tipo de dado:

- Sub-propriedade (rdfs:subPropertyOf) - permite criar hierarquias de propriedades, e.g., a propriedade *temOrientador* pode ser sub-propriedade de *temProfessor*;
- Domínio (rdfs:domain) - limita os indivíduos aos quais a propriedade se aplica, e.g., a propriedade *especialidade* pode estar relacionada apenas aos indivíduos das classes *GrandeArea*, *Area*, *SubArea* e *Especialidade*;
- Imagem (rdfs:range) - limita os indivíduos que a propriedade tem como valor, e.g., a propriedade *sexo* pode assumir o valor “Feminino” ou “Masculino”;
- Equivalência (rdfs:equivalentProperty) - define a equivalência entre duas propriedades;
- Inversa (owl:inverseOf) - define que uma propriedade é o inverso de outra, e.g., a propriedade *temArea* é o inverso de *pertenceGrandeArea*, ou seja, se uma Grande Área tem uma Área de Conhecimento, então pode-se inferir que esta Área pertence à Grande Área de Conhecimento;
- Transitiva (owl:transitiveProperty) - define a transitividade de uma propriedade, e.g., se uma Grande Área possui uma Área e esta possui uma Especialidade, então pela transitividade, pode-se inferir que a Grande Área possui a Especialidade;
- Simétrica (owl:symmetricProperty) - especifica que uma propriedade é igual a sua inversa, e.g., se um autor X é co-autor de um autor Y em uma publicação, então pode-se inferir que o autor Y é co-autor do autor X;
- Assimétrica (owl:asymmetricProperty) - especifica se uma propriedade é assimétrica, e.g., se um indivíduo X é relacionado a outro Y pela propriedade *temOrientador* então não pode-se inferir que Y não relaciona-se a X pela propriedade *temOrientador*;
- Funcional (owl:functionalProperty) - especifica que a propriedade tem um único valor, e.g., se o indivíduo Maria Joana Oliveira tem um currículo Lattes X e

Maria Joana Oliveira Silva tem o currículo Lattes X, então pode-se inferir que os dois autores são a mesma pessoa;

- Funcional inversa (`owl:inverseFunctionalProperty`) - especifica que a propriedade inversa é funcional, e.g., dado que o currículo Lattes X pertence à Maria Joana Oliveira e à Maria Joana Oliveira Silva e esta propriedade é inversa, então pela funcionalidade da propriedade, Maria Joana Oliveira e Maria Joana Oliveira Silva são a mesma pessoa, e pela relação inversa da propriedade pode-se inferir que os autores possuem o currículo X.
- Reflexiva (`owl:reflexiveProperty`) - especifica que um indivíduo se relaciona a ele próprio através de uma propriedade, e.g., um indivíduo deve se relacionar com ele próprio através da propriedade conhece;
- Irreflexiva (`owl:irreflexiveProperty`) - indica que um indivíduo não pode relacionar com ele mesmo através de uma propriedade irreflexiva, e.g., o indivíduo X é professor do indivíduo Y, onde qualquer um dos indivíduos não pode ser professor dele próprio.

3. Restrições de Propriedade em OWL

Em OWL a restrição de propriedade (*property restriction*) permite descrever uma classe por meio de restrições que devem ser obedecidas por todos os indivíduos pertencentes à classe. Essas limitações podem ser de dois tipos: restrição de valor (*value constraint*) e restrição de cardinalidade (*cardinality constraint*). As seguintes funcionalidades e construtores podem ser usados para definir restrições de valor em uma propriedade:

- Restrição existencial (`owl:someValuesFrom`) - especifica a existência de pelo menos um relacionamento entre os indivíduos, e.g., a propriedade *temArea* da classe *GrandeArea* especifica que os indivíduos desta classe relacionam a pelo menos um indivíduo da classe *Área*;
- Restrição universal (`owl:allValuesFrom`) - especifica que através de uma propriedade todos os relacionamentos existentes entre os indivíduos de uma classe associam aos indivíduos de outra classe, e.g., todos os indivíduos da classe *Curriculo Vitae* se relacionam a classe *Autores*;
- Restrição de indivíduo (`owl:hasValue`) - especifica que através de uma propriedade pelo menos um dos relacionamentos dos indivíduos de uma classe é com um indivíduo.

Além das restrições de valor, as seguintes funcionalidades e construtores restringem a cardinalidade de uma propriedade:

- Cardinalidade mínima (`owl:minCardinality`) - define o número mínimo de relacionamentos que um indivíduo deve ter;
- Cardinalidade máxima (`owl:maxCardinality`) - define o número máximo de relacionamentos que um indivíduo deve ter;
- Cardinalidade (`owl:cardinality`) - define o número exato de relacionamentos que um indivíduo deve ter.

Além disso, o modelo lógico da ontologia permite a utilização de um mecanismo de inferência, o qual pode verificar se as declarações e as definições da ontologia são consistentes. À partir disso, o mecanismo de inferência pode ajudar a manter a hierarquia entre as classes, o que é útil quando existem casos em que uma subclasse tem mais de uma superclasse (Horridge, 2011).

Conforme McGuinness and van Harmelen (2004), a OWL é projetada para aplicações que necessitam processar informação e não apenas apresentar essa informação. Ela facilita enormemente a interpretabilidade da máquina acerca de conteúdo *Web*, muito mais do que XML e RDF são capazes de fornecer, pois fornece vocabulário adicional e uma semântica formal.

As linguagens OWL se subdividem em três tipos:

- OWL-Lite;
- OWL-DL;
- OWL-Full.

A característica principal de cada sub-linguagem é a sua expressividade: a OWL-Lite é a menos expressiva; a OWL-Full é a mais expressiva; a expressividade da OWL-DL está entre a OWL-Lite e a OWL-Full.

A OWL-Lite é a sub-linguagem sintaticamente mais simples. Destina-se a situações em que apenas são necessárias restrições e uma hierarquia de classe simples. Por exemplo, o OWL-Lite pode fornecer uma forma de migração para tesouros existentes, bem como de outras hierarquias simples.

A OWL-DL é mais expressiva que a OWL-Lite e baseia-se em lógica descritiva, um fragmento de lógica de primeira ordem, passível portanto de raciocínio automático. Através do *Reasoner* é possível computar automaticamente a hierarquia de classes e verificar inconsistências na ontologia. Além disso, a OWL-DL permite o uso de propriedades não suportadas pela OWL-Lite, aumentando a capacidade de representação semântica de um domínio de conhecimento.

A OWL-Full é a sub-linguagem OWL mais expressiva. Destina-se a situações onde alta expressividade é mais importante para garantir a decidibilidade ou completeza da linguagem. É mais flexível que as demais, e.g., permite que uma classe seja também uma instância ou uma propriedade. No entanto, essa flexibilidade não permite que sistemas de raciocínio automático possam ser usados para inferência de informações. Enquanto que a OWL-DL e OWL-Lite possuem regras de formação mais rígidas e viabilizam o uso da lógica de descrição para a realização de inferência. A OWL-Full pode ser vista como uma extensão do RDF, enquanto que OWL-Lite e OWL-DL são consideradas extensões de uma visão restrita do RDF (McGuinness and van Harmelen, 2004). Dessa forma, a OWL-DL vem sendo bastante utilizada na construção de ontologias que permitam a realização de inferência de conhecimento.

Uma ontologia OWL é também uma ontologia OWL 2. A linguagem OWL 2 é uma extensão e revisão da OWL adicionando novas funcionalidades. Ontologias em OWL 2, assim como no seu predecessor, provê classes, propriedades, instâncias e são armazenadas como documentos *Web* - arquivos ou páginas *Web* identificados por um IRI (*Internationalized Resource Identifier*).

OWL 2 oferece facilidades adicionais (e.g. união disjunta de classes) e mais expressividade ao OWL, como: chaves, cadeia de propriedades, tipos e intervalos de dados mais avançados, restrições qualificada de cardinalidade, propriedades assimétricas, reflexivas, disjuntas e mais funcionalidades relacionadas a *tags* (W3C, 2009).

A OWL 2 não segue uma sintaxe ou um formato de arquivo em particular. No entanto, na prática uma ontologia em OWL 2 precisa estar escrita em alguma sintaxe para que seja armazenada e reutilizada por outras aplicações. A W3C sugere cinco tipos de sintaxes para ontologias em OWL 2:

- RDF/XML - representa grafos RDF. É uma implementação obrigatória com foco na interoperabilidade, ou seja, softwares compatíveis com OWL 2 devem no mínimo implementar esta sintaxe;
- OWL/XML - representa uma serialização direta a XML com foco na facilidade de serialização, facilitando o processamento por ferramentas XML;
- Funcional - representa uma especificação estrutural, portanto facilita na visualização da estrutura formal da ontologia;
- Manchester - é uma sintaxe próxima da DL, portanto, facilita na leitura/escrita de ontologias em DL;
- Turtle - mapeia-se a grafos RDF e facilita a leitura de triplas RDF.

4.6 Metodologias

Assim como no desenvolvimento de software, existem diferentes metodologias para a construção de uma ontologia. No entanto, a grande maioria dos desenvolvedores utiliza seus próprios critérios no processo de desenvolvimento de uma ontologia e isso pode acarretar problemas como (Almeida and Bax, 2003):

- Inexistência de um modelo conceitual da ontologia;
- Dificuldade de reúso da ontologia;
- Dificuldade no entendimento do domínio ou escopo da ontologia;
- Dificuldade no desenvolvimento de ontologias complexas;
- A capacidade de descrição conceitual do domínio fica restrita à linguagem de implementação da ontologia.

Dessa forma, percebe-se a importância da adoção de uma metodologia para minimizar as dificuldades citadas. Algumas metodologias são apresentadas:

- TOVE (*TO*ronto *Virtual Enterprise*) - utiliza a lógica de primeira ordem para representar atividades, estados, tempo, recursos e custo em uma arquitetura de ambiente de integração (Grüninger and Fox, 1995);
- ONIONS (*ON*tologic *I*ntegration *O*f *N*aive *S*ources) - consiste em um conjunto de métodos especialmente gerados para integrar múltiplas informações de origens diferentes, com ênfase em ontologias de domínio (Gangemi et al., 1999);

- METHONTOLOGY - uma das mais conhecidas e mais maduras metodologias de construção de ontologias, sendo fortemente influenciada por metodologias de Engenharia de Software e de Engenharia do Conhecimento. Prevê um ciclo de vida baseado na prototipagem de ontologias, de acordo com a evolução do processo de desenvolvimento (especificação, conceitualização, formalização, implementação e manutenção), inspirando-se em preceitos do ciclo de vida de software (López, 1999);
- UPON (*United Process for ONtologies*) - é uma metodologia baseada na abordagem UML (*Unified Modeling Language*). Baseia-se em casos de uso, é incremental e iterativa (De Nicola et al., 2009);
- OntoClean: é fundamentada por conceitos formais, que são gerais o suficiente para serem usados em qualquer ontologia, independentemente do domínio (Guarino and Welty, 2004);
- OTK (*On-To-Knowledge*) - é uma metodologia de desenvolvimento de ontologias fruto da cooperação de várias entidades européias, tendo como intuito desenvolver ontologias para serem empregadas em Sistemas de Gestão do Conhecimento (Sure and Studer, 2002);
- Sensus - baseia-se na construção de ontologias a partir de outras ontologias, através da identificação de termos relevantes para o domínio ligando-os à ontologia mais abrangente (López, 1999);
- Método 101 - segundo Noy and McGuinness (2001), não existe um modo correto ou metodologia de desenvolvimento de ontologias. Por isso, os autores sugerem apenas um processo para tal, denominado *Ontology Development 101*. Este processo consiste em um guia de passos iterativos executados livremente no desenvolvimento de ontologia, a saber: (i) determinar o domínio e escopo da ontologia; (ii) considerar o reuso de ontologias existentes; (iii) enumerar os termos importantes da ontologia; (iv) definir classes e hierarquia de classes; (v) definir as propriedades das classes; (vi) definir os valores das propriedades e (vii) criar instâncias. Este método tem como objetivo guiar a construção de uma ontologia através de um roteiro, e destaca que não há apenas uma forma ou metodologia correta para o desenvolvimento de ontologias.

O método 101 foi adotado para a extensão da ontologia por ser um método simples e intuitivo, proporcionando maior facilidade no desenvolvimento e na integração das ontologias envolvidas neste trabalho.

4.7 Ferramentas

Diversas ferramentas foram criadas para auxiliar na edição e construção de ontologias. Estas ferramentas geralmente oferecem uma interface gráfica que permite que o usuário crie e edite as ontologias sem usar diretamente uma linguagem de implementação de ontologia. Além disso, permitem conexão com sistemas de inferência, através dos quais pode-se verificar a consistência das ontologias.

De acordo com Duineveld et al. (2000), alguns fatores tornam uma ferramenta popular: facilidades de uso, entendimento intuitivo da interface, interfaces gráficas, conexão

a repositórios, portabilidade, interoperabilidade, organização dos arquivos gerados, documentação e extensibilidade.

Bergman (2010) disponibiliza uma lista com dezenas de ferramentas para edição de ontologias, das quais pode-se citar:

- **Altova SemanticWorks** - é um editor RDF e OWL que gera RDF/XML, não existe versão *open source* (ALTOVA, 2012);
- **TopBraid Composer** - é um ambiente para desenvolver ontologias para Web Semântica e aplicações *Web*. Oferece suporte para desenvolvimento, gerenciamento e teste do modelo e base de conhecimento. Existe uma versão de livre acesso para ontologias pequenas (TopQuadrant, 2011);
- **OilEd** - é um editor de ontologias para as linguagens OIL e DAML-OIL. Permite a verificação de consistência da ontologia, a importação e exportação de ontologias, bem como a classificação automática da taxonomia (Bechhofer et al., 2001);
- **WebODE** - é um ambiente integrado para a construção de ontologias baseado na tecnologia *Web*. Permite manipular ontologias realizadas de acordo com a metodologia METHONTOLOGY, apoiando as atividades envolvidas no ciclo de desenvolvimento de ontologias acordo com esta metodologia (Óscar Corcho et al., 2002);
- **Text-to-onto** - proporciona um ambiente para a construção de ontologias a partir de textos. Os textos podem ser em linguagem natural ou formatados em HTML. O sistema é composto por um módulo de gerenciamento de textos e um extrator de informações. Os resultados são armazenados na linguagem XML (Maedche and Staab, 2000);
- **Ontolingua** - provê um ambiente colaborativo para criar e editar ontologias na linguagem de mesmo nome, pode ser usado de forma local ou remota (Farquhar et al., 1996);
- **Protégé** - consiste em uma ferramenta *open source* com arquitetura baseada em componentes escritos em Java, permite a extensão de suas funcionalidades através do desenvolvimento de *plug-ins*, o que o torna um ambiente flexível para prototipagem rápida e desenvolvimento de aplicativos. O Protégé suporta dois modos de modelar ontologias através dos editores Protégé-Frames e Protégé-OWL. As ontologias do Protégé podem ser exportadas nos formatos RDF(S), OWL, e XML Schema. Esta ferramenta oferece acesso a mecanismos de inferências, tais como *Pellet* e *Racer*, que permitem o raciocínio automático e a verificação da consistência da ontologia. O Protégé foi a ferramenta utilizada para a integração e extensão das ontologias Conscientias (Conscientias, 2002, 2006) utilizadas neste projeto (Stanford, 2011).

4.8 Raciocinador

O raciocinador ou *reasoner* é um mecanismo oferecido por algumas linguagens e ferramentas para gerar inferências ou raciocínio automático à partir da ontologia. Um *reasoner* consiste em uma entidade capaz de responder a determinados questionamentos e afirmações realizados em relação a uma ontologia. Através de mecanismos de inferência,

conduzidos por raciocinadores automatizados, informações adicionais podem ser deduzidas. Os raciocinadores são construídos com base nas regras estabelecidas pela linguagem descritiva adotada, sendo que quanto maior o poder de expressividade da linguagem, mais complexo é o trabalho do raciocinador. À seguir são listadas algumas tarefas executadas pelo raciocinador sobre a estrutura de um ontologia:

- Construção da hierarquia de classes - consiste em gerar a hierarquia entre as classes;
- Verificação de instância - consiste em verificar se um indivíduo pertence a um conceito específico;
- Verificação de consistência da base - para verificar se cada restrição de relacionamento está sendo atendida;
- Identificação conceitual - para encontrar o conceito que descreve em mais detalhe um determinado indivíduo;
- Busca de indivíduos a partir de um determinado conceito.

Conforme Foundation (2012), utilizar o *reasoner* para gerar automaticamente a hierarquia de classes é um dos maiores benefícios do uso de um raciocinador. No caso de ontologias muito grandes (mais de 200 classes) o uso de um raciocinador para processar as relações entre as classes torna-se quase vital. Sem o raciocinador é muito difícil manter ontologias muito grandes em um estado sustentável e logicamente correto. Este mecanismo de raciocínio automático ajuda a manter a ontologia correta e modular, bem como promover a sua reutilização por outras ontologias e aplicações, além de minimizar erros humanos.

Entretanto, o *reasoner* adiciona muitas informações à ontologia que podem não ser úteis para a aplicação, visto que a adição destas informações torna o modelo mais complexo ou até mesmo confuso. Como algumas ferramentas não conseguem diferenciar declarações inferidas e declarações afirmadas no modelo, muitos desenvolvedores e editores de ontologias optam por não utilizar o mecanismo de inferência (Foundation, 2012).

Em muitos outros casos, no entanto, é a adição do *reasoner* que justifica a utilização da ontologia. Para exemplificar, considere que João e Maria são autores da publicação “O uso do Jena”. A relação co-autor não é declarada, mas na ontologia sabe-se que *coautorDe* é uma sub-propriedade da propriedade *autorDe*. Neste caso, se a afirmação “João e Maria são autores da publicação O uso do Jena” é verdadeira, então a afirmação “João é co-autor de Maria” também é verdadeira.

Para fazer uso do mecanismo de inferência é preciso definir o *reasoner* que será utilizado na ferramenta. Alguns oferecem maior quantidade de recursos, porém menor velocidade, enquanto que outros apresentam maior velocidade, porém oferece menos recursos. Alguns desses raciocinadores disponíveis, podem ser citados como:

- FaCT++ - um raciocinador gratuito, portátil e eficiente desenvolvido para ontologias escritas em lógica descritiva. Implementado em C++, suporta a linguagem OWL-DL e parcialmente a OWL 2 (Tsarkov and Horrocks, 2006);
- RACER - é um raciocinador para lógica descritiva, cujas principais funcionalidades são a verificação de consistência e a classificação das classes, incorporando técnicas de otimização que permitem o manuseio de grandes ontologias. Contudo, o

raciocínio com instâncias não é completo, causando resultados inesperados quando estas são utilizadas na definição de classes, e.g. por meio do uso da propriedade *owl:hasValue* (Haarslev and Moller, 2001);

- RacerPro - é a versão comercial (pago) do raciocinador RACER (Co.KG, 2012);
- Pellet - é um raciocinador gratuito desenvolvido para ontologias escritas em OWL. Ao contrário do FaCT e do RACER, que podem ser usados com várias lógicas descritivas, o objetivo do Pellet é ser usado exclusivamente em OWL. Dessa forma, explora as funcionalidades da OWL, oferecendo raciocínio sobre tipos de dados e consultas sobre instâncias. Outra funcionalidade oferecida é a conversão de ontologias dentro da linguagem OWL, transformando uma ontologia escrita em OWL-Full, que não permite raciocínio automático, em outra ontologia em OWL-DL, possibilitando o mecanismo de inferência (Sirin et al., 2007);
- MSPASS - é uma extensão do provador de teorema de primeira ordem SPASS, que pode ser usado como um provador de teorema de lógica modal, provador de teorema para lógicas descritivas e provador de teorema para cálculo relacional. É um dos mais rápidos e mais sofisticados provadores de teoremas para lógica de primeira ordem (Hustadt and Schmidt, 2000).

4.9 Jena

Diversas APIs (*Application Programming Interface*) já foram desenvolvidas para possibilitar a manipulação de ontologias em sistemas. Foram desenvolvidas APIs em linguagem C, C++, Java e outras, sendo que as desenvolvidas em linguagem Java foram mais bem aceitas e têm recebido mais colaborações no meio, devido à popularidade do Java. Foram pesquisadas diferentes APIs para manipulação de ontologias que atendessem ao propósito deste trabalho: OWL do Protégé, *BeanGenerator*, JRDF e Jena; sendo que a API Jena foi escolhida por se mostrar a mais robusta, com melhor documentação disponível e por possuir suporte para a linguagem OWL.

De acordo com Foundation (2012), o Jena é um projeto *open source* que se originou dentro do núcleo de pesquisa em *Web Semântica* da HP (*Hewlett-Packard*) no ano 2000, com o objetivo de proporcionar um *framework* na linguagem Java para construir aplicações de *Web Semântica*. Ele fornece uma extensa biblioteca de componentes Java que suportam a manipulação de ontologias nas linguagens RDF, RDFS, OWL e SPARQL em consonância com as recomendações do W3C. Jena inclui uma máquina de inferência (*reasoner*) baseada em regras para executar o raciocínio automático em ontologias RDFS e OWL, e também uma variedade de opções de armazenamento, podendo ser em memória ou em disco, de triplas RDF.

Segundo Foundation (2012), o *framework* Jena fornece um conjunto de ferramentas e bibliotecas para ajudar no desenvolvimento de uma aplicação da *Web Semântica*:

- uma API para leitura, processamento e gravação de dados RDF em XML;
- API para a manipulação de uma ontologia em OWL e em RDFS;
- um mecanismo de inferência baseado em regras para as linguagens RDF e OWL;

- um mecanismo de consulta compatível com a mais recente especificação da linguagem SPARQL;
- permite que dados RDF possam ser publicados em outros aplicativos.

O Jena trabalha de forma transparente em relação à linguagem adotada para representar a ontologia. De acordo com a linguagem, determinados recursos são habilitados ou não, porém, a forma de manipular e criar ontologias é a mesma para qualquer linguagem de representação (Frozza and Gonçalves, 2010). Serão apresentados os principais métodos e classes de manipulação de ontologias no Jena.

4.9.1 Criação de ontologias

As ontologias em Jena são criadas e manipuladas como grafos onde os nós são os recursos ou classes e os arcos são as propriedades ou relacionamentos. Em Jena, um grafo é chamado de modelo e é representado pela *interface Model*. A criação de uma ontologia é realizada por um gerenciador denominado *ModelFactory*. Em sua forma mais simples, criar uma ontologia no Jena resume-se a:

```
OntModel ontologia = ModelFactory.createOntologyModel();
```

Desta forma cria-se uma ontologia sem associação a uma linguagem específica. Porém, em geral, criam-se as ontologias seguindo um modelo de representação, para que se possa compartilhar as mesmas posteriormente. Neste caso, basta especificar a URI da ontologia ao criar a mesma, da seguinte forma:

```
OntModel ontologia = ModelFactory.createOntologyModel(URIIdaOntologia);
```

4.9.2 Carga de ontologias

O Jena disponibiliza algumas maneiras de carregar uma ontologia: (i) a partir de um arquivo; (ii) a partir de um endereço na *Web* ou (iii) a partir de um conjunto de dados. Para carregar uma ontologia, basta instanciar um objeto *OntModel* e carregar a ontologia através do método *read*. A chamada deste método varia de acordo com a origem dos dados, o método abaixo carrega uma ontologia à partir da sua url:

```
read(String url);
```

Sobre a carga de ontologias, um aspecto que merece destaque é o tratamento de referências a outras ontologias contidas na ontologia sendo carregada. Para definir se as ontologias associadas devem ou não ser carregadas, pode-se utilizar um *DocumentManager*, que é uma classe responsável pela manipulação de documentos representando ontologias. Para evitar a carga das ontologias associadas, basta chamar o método *setProcessImports* do *DocumentManager* da ontologia em questão, passando como parâmetro uma constante *false*:

```
OntModel m = ModelFactory.createOntologyModel();
OntDocumentManager dm = m.getDocumentManager();
dm.setProcessImports(false);
```

Pode-se também evitar apenas a carga de determinadas ontologias associadas, bastando para tal passar suas URIs ao método *addIgnoreImports*.

4.9.3 Manipulação de classes

Todas as entidades de uma ontologia que representam valores têm como classe base *OntResource*. Cada recurso de uma ontologia tem um conjunto de atributos para os quais os métodos à seguir permitem a sua manipulação: *add*<atributo>, *set*<atributo>, *get*<atributo>, *has*<atributo> e *remove*<atributo>.

As classes da ontologia são representadas por objetos da classe *OntClass*, por meio da qual existem duas maneiras de recuperar uma classe:

```
String autor = "http://www.observeunb.unb.br/ontologia/autores#";
Resource r = m.getResource(autor + "Autores");
OntClass autores = (OntClass) r.as(OntClass.class);
```

Ou:

```
OntClass cAutores = m.getOntClass(autor + "Autores");
```

Para listar as classes utiliza-se o método *listSubClasses*, e para criar classes utiliza-se o método *createClass*, podendo informar a URI da classe. Caso uma URI seja informada, o Jena verifica se a classe já existe. Em caso positivo, apenas retorna a mesma, sem criar uma nova classe. Se não é informada uma URI o Jena cria uma classe anônima.

```
OntClass cAutores = m.createClass(autor + "Autores");
```

4.9.4 Manipulação de propriedades

No Jena, os atributos ou propriedades de uma ontologia são representados pela classe *OntProperty* que possui os métodos básicos *add*, *get*, *set*, *list*, *has* e *remove*, para manipular os atributos. Em uma ontologia, a propriedade denota o nome de um relacionamento entre classes ou recursos, ou entre uma classe e um valor. Com o método *listDeclaredProperties*, pode-se obter as propriedades de uma classe. Nesta listagem, pode-se incluir propriedades das superclasses ou não, conforme o parâmetro informado.

Como suporta OWL, Jena define dois tipos básicos de propriedades: *ObjectProperty* e *DatatypeProperty*. O primeiro tipo suporta objetos em um intervalo, enquanto que o segundo tipo suporta apenas literais (*strings*, números). Além destes tipos básicos de propriedades, Jena define outros tipos mais avançados de propriedade como *TransitiveProperty*, *FunctionalProperty*, *SymmetricProperty* e *InverseFunctionalProperty*, as quais podem ser acessadas como qualquer propriedade através de uma simples chamada a um método:

```
public TransitiveProperty asTransitiveProperty();
public FunctionalProperty asFunctionalProperty();
public SymmetricProperty asSymmetricProperty();
public InverseFunctionalProperty asInverseFunctionalProperty();
```

Além das restrições apresentadas, existem outros tipos de restrições associadas às propriedades no Jena: *has value*, *all values from*, *some values from*, *cardinality*, *min cardinality* e *max cardinality*.

4.9.5 Manipulação de instâncias

O Jena permite a manipulação de instâncias através dos métodos: *listInstances*, *createIndividual* e *dropIndividual*. Estes métodos listam instâncias, criam indivíduos em uma classe e removem indivíduos de uma determinada classe. Para criar instâncias existem duas maneiras, nas quais a classe à qual a instância pertence deve ser informada:

```
OntClass cAutores = m.createClass(autor + "Autores");  
Individual autor1 = m.createIndividual(autor + "autor1", cAutores);
```

ou

```
Individual autor1 = cAutores.createIndividual(autor + "autor1");
```

A diferença entre as duas abordagens é que na segunda maneira a criação do indivíduo é feita no mesmo modelo ao qual a classe está relacionada.

4.9.6 Consultas

O Jena disponibiliza consultas por meio das linguagens RDQL (*RDF Data Query Language*) (W3C, 2004) e SPARQL (*Protocol and RDF Query Language*) (W3C, 2008), sendo que esta última foi escolhida para este projeto por ser a mais atualizada. O propósito do SPARQL, que é a recomendação do W3C desde 2008, consiste em permitir que arquivos RDF sejam consultados através de uma linguagem orientada a objetos.

A vantagem do SPARQL é a sua semelhança com a linguagem SQL (*Structure Query Language*), facilitando o aprendizado. A linguagem pode ser utilizada para formular consultas que variam desde as mais simples às mais complexas. Consultas complexas podem incluir novos recursos como, união (*UNION*), como funções de agregação (e.g. *COUNT*), expressões na cláusula *SELECT* (e.g. adição, multiplicação), negação (*EXISTS* ou *NOT EXISTS*), consultas aninhadas e consultas que podem ser construídas através de novos grafos, os quais são gerados a partir do resultado de uma consulta.

Assim como no SQL, as principais cláusulas do SPARQL são: *SELECT*, *FROM*, *WHERE*, *ORDER BY* e *UNION*, que possui funcionamento diferente do seu correspondente na linguagem SQL. Outras cláusulas são específicas do SPARQL:

- **BASE** - define a URI base de um recurso;
- **FILTER** - aplica um filtro sobre as linhas recuperadas pela consulta;
- **LIMIT** - limita a quantidade de linhas recuperadas pela consulta;
- **OFFSET** - permite que seja aplicado um deslocamento sobre o conjunto de linhas recuperadas pela consulta;
- **OPTIONAL** - permite que uma linha seja recuperada mesmo que não exista o valor de uma propriedade;
- **PREFIX** - cria um apelido para a URI de um arquivo RDF/OWL.

Da mesma forma como no RDF, a linguagem SPARQL é construída sobre uma tripla a qual consiste em sujeito, objeto e predicado. No SPARQL variáveis são identificadas com os símbolos “?” ou “\$”, enquanto que literais são delimitados por aspas simples ou duplas.

4.10 Conscientias

A Comunidade para Ontologias em Ciência, Tecnologia e Informações de Aperfeiçoamento de Nível Superior (Conscientias) é uma extensão da Comunidade LMPL (Linguagem de Marcação da Plataforma Lattes), e foi estabelecida no ano 2000 como responsável pela criação e manutenção das gramáticas XML da Plataforma Lattes. Além disso, este grupo foi criado para desenvolver ontologias que visam o intercâmbio de informações entre agências de fomento e instituições ligadas ao tema Ciência, Tecnologia, Inovação e Informações de Aprimoramento de Nível Superior (Conscientias, 2012).

Caracterizam-se como responsabilidades da Comunidade Conscientias a concepção, elaboração, recomendação e manutenção das gramáticas relacionadas às ontologias submetidas pelas agências ou instituições conselheiras.

No ano 2002 foi definida a ontologia para os Grupos de Pesquisa na linguagem de definição de esquema XSD (*XML Schema Definiton*), quatro anos depois foi criada a ontologia para o Currículo *Lattes*. Inicialmente a ontologia foi construída utilizando a linguagem de definição de tipos DTD (Raggett et al., 1999). Posteriormente, com a homologação da linguagem XML Schema pelo Consórcio W3C, a comunidade Conscientias-LMPL construiu uma nova gramática para o Currículo *Lattes* utilizando a linguagem de esquemas XML.

Na comunidade Conscientias, as ontologias são representadas pela linguagem de marcação XML. Com essa linguagem, o XML *Schema*, pode-se utilizar recursos não disponíveis na linguagem antecessora - o DTD, como mecanismo de controle de tipos, a utilização de namespaces e a reutilização de código. Além disso, a linguagem XML facilita o intercâmbio de informações curriculares entre as bases de dados das instituições de ensino e pesquisa e as bases do Sistema de Currículo *Lattes*.

De acordo com Conscientias (2006), o formato de representação em XML foi definido para manter o padrão de informação da Plataforma *Lattes*. Esse padrão mantido pela Comunidade Conscientias-LMPL foi elaborado seguindo a estrutura das informações representadas no Sistema de Currículos Lattes e de Grupos de Pesquisa do CNPq. Para as instituições que utilizam esses sistemas, a adoção do padrão nacional definido pela Comunidade Conscientias-LMPL de exportação e importação de currículos garante segurança e estabilidade nas regras de tradução entre suas estruturas de dados e a estrutura do Currículo *Lattes*, bem como facilita o intercâmbio de informações entre diversas bases institucionais e a base da Plataforma *Lattes*.

As Figuras 4.3 e 4.4 exibem em forma de grafo as principais classes das ontologias de Grupos de Pesquisa e de Currículo *Lattes*, respectivamente, definidas pelo Conscientias.

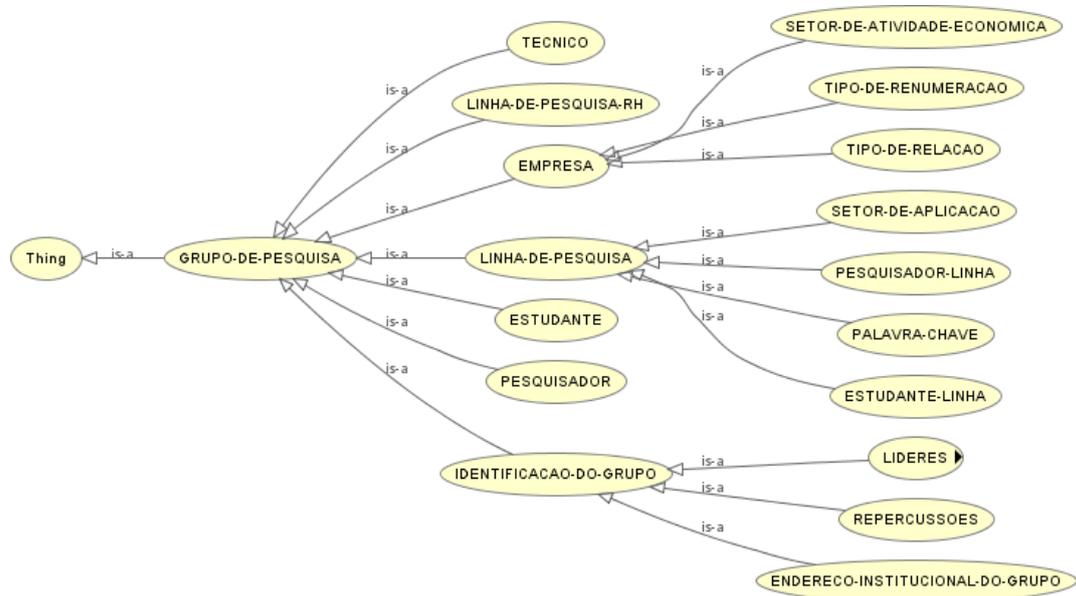


Figura 4.3: Principais classes da ontologia de Grupos de Pesquisa (Conscientias, 2002).



Figura 4.4: Principais classes da ontologia de Currículo Lattes (Conscientias, 2006).

Capítulo 5

Trabalhos Correlatos

Recentemente vários trabalhos vem apresentando soluções integrando pesquisadores na *Web* e possibilitando relacionamentos por meio de produções científicas. No entanto, são poucas as iniciativas de construção de rede social científica a qual faça uso de um sistema de reputação visando avaliar os membros da rede, bem como utiliza recursos de uma ontologia para disponibilizar informações semânticas do domínio de conhecimento da comunidade acadêmica de pesquisa.

O *Google Scholar*¹ é uma das mais poderosas ferramentas de colaboração científica. Ela permite que pesquisadores compartilhem e encontrem publicações científicas. Além de ser utilizada como um meio de medir a produção individual de pesquisadores, onde várias ferramentas baseiam-se em seus dados para calcular a produção dos pesquisadores. Para uma publicação ser indexada nesta ferramenta ela tem que ter pelo menos uma referência a um artigo já indexado no *Google Scholar* (Labbé, 2009)

A ferramenta *ArnetMiner*² desenvolvida por (Tang et al., 2008) fornece serviços como a busca de especialistas, busca associativa de pessoas, visualização em grafos e extração de informações semânticas baseada na comunidade de pesquisadores na *Web*. Para isso a ferramenta estende a ontologia FOAF (Brickley and Miller, 2000) para unificar os perfis dos pesquisadores e permite a pesquisa de especialistas por algum tópico ou palavra chave, fazendo uso do algoritmo de reputação *PageRank*. De acordo com Tang et al. (2011a), atualmente este sistema coleta informações de mais de 1 milhão de pesquisadores, 3 milhões de publicações de artigos e 8 mil conferências obtidas de páginas pessoais dos pesquisadores e de diferentes bibliotecas digitais internacionais, como DBLP (DBLP, 2011), *CiteSeer* (CiteSeer, 2011) e ACM (ACM, 2011). Este sistema está disponível na *Web* desde 2006 recebendo um considerável volume de acessos provenientes de 190 países. A Figura 5.1 exibe a rede de colaboração formada por pesquisadores fictícios P23 e P15, onde os tipos de relacionamento se distinguem pelas cores, podendo ser do tipo orientador, orientando ou co-autor. Assim como no *ArnetMiner*, no presente projeto é utilizada uma ontologia para a representação e integração da informação e também um modelo de reputação para a identificação de especialistas na rede. Entretanto, ainda não são capturados dados de bibliotecas digitais internacionais como DBLP, *CiteSeer* e ACM.

¹Disponível em: <http://scholar.google.com.br>

²Disponível em: <http://www.arnetminer.org>

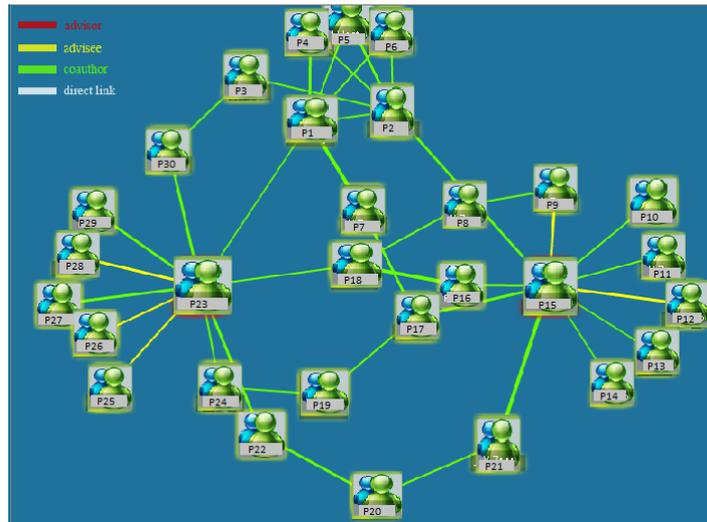


Figura 5.1: Rede de colaboração de dois pesquisadores no *ArnetMiner*.

Outro projeto interessante é o *Microsoft Academic Search*³ (MAS), uma ferramenta de livre acesso desenvolvida pela *Microsoft Research*, destinada à realização de pesquisa básica aplicada a diferentes áreas de conhecimento. O MAS permite consultas sobre publicações, autores, conferências, periódicos, instituições e palavras-chave, mediante um mecanismo de busca que, assim como *ArnetMiner*, coleta informações na *Web* de bases internacionais como DBLP, *CiteSeer*, ACM, entre outros. Estes dados podem ser visualizados em diferentes formatos, como rede de relacionamento, gráficos de publicações de pesquisadores e mapas acadêmicos georreferenciados. A Figura 5.2 apresenta um mapa de parte da América do Sul, destacando o Brasil. No mapa o tamanho da circunferência é proporcional a quantidade de publicações, onde a maior circunferência representa a Universidade de São Paulo com 119 mil e 572 publicações até o ano de 2011. Neste projeto, é possível obter a rede de relacionamentos de pesquisadores através de suas publicações científicas e de grupos de pesquisa, e a transitividade de relacionamentos por meio do uso da ontologia.

Yang et al. (2009) propõe a idéia de construção de uma rede social de pesquisadores baseada na coleta de metadados de publicações científicas e de dados de perfis pessoais na *Web*. De acordo com os dados selecionados são construídas diferentes redes de relacionamento. Dentre os dados investigados estão: palavras-chave de publicação, interesses pessoais, dados sobre conferências onde são realizadas as publicações, e co-autores de publicações. Os dados dos perfis são extraídos da *Web* e convertidos para o formato RDF (Klyne and Carroll, 2004), facilitando o cálculo de similaridade entre dois pesquisadores. Esse método calcula a semelhança entre dois pesquisadores medindo a distância entre dois conjuntos de dados que foram coletados para representá-los.

O portal VIVO⁴ é uma ferramenta que disponibiliza uma rede de cientistas baseada em dados e ontologias institucionais com o objetivo de permitir a descoberta de pesquisadores e a colaboração científica entre instituições e países, através do compartilhamento de in-

³Disponível em: <http://academic.research.microsoft.com>

⁴Disponível em: <http://vivoweb.org/>



Figura 5.2: Mapa acadêmico do Brasil no *Microsoft Academic Search*.

formações dos pesquisadores e suas atividades. O projeto foi financiado pela Universidade da Flórida, Universidade *Cornell* e pela Universidade de Indiana em *Bloomington*, juntamente com quatro outras instituições parceiras, tendo a sua última versão sido iniciada em Abril de 2010 (Krafft et al., 2010). VIVO suporta a pesquisa rápida de informações relacionadas ao perfil e à publicação do pesquisador a nível individual, institucional e internacional, bem como a pesquisa semântica. Para isso, todos os dados do VIVO são representados em RDF através da extensão das ontologias FOAF (Brickley and Miller, 2000) e BIBO (D’Arcus and Giasson, 2009), que representam o perfil do pesquisador e sua documentação bibliográfica. Apesar de exibir a rede de co-autoria do pesquisador em forma de grafo, o VIVO não faz uso de um sistema de reputação em sua rede. A Figura 5.3 mostra a rede de colaboração de um pesquisador no portal VIVO.

Na mesma linha do estudo deste trabalho, uma integração entre pesquisadores por meio de publicações, rede social e ontologia é o modelo para redes de conhecimento científico proposto por Paiva et al. (2009). No entanto, não há menção de avaliação da rede e da aplicação do modelo em uma ferramenta como estudo de caso.

A nível nacional, o Portal Brasileiro de Ciência e Tecnologia⁵ - CiênciaBrasil, desenvolvido por Laender et al. (2011), disponibiliza um mapa da produção científica e constrói a rede de co-autoria dos atuais pesquisadores vinculados aos Institutos Nacionais de Ciência e Tecnologia (INCT), bem como utiliza as informações dos autores para fornecer diferentes visualizações do perfil de publicação. Para isso, os dados individuais dos pesquisadores são coletados da plataforma Lattes do CNPq por meio de técnicas de extração, integração, desambiguação e deduplicação, e análise resultantes de pesquisas realizadas no âmbito *InWeb*, um dos 126 INCTs existentes. O portal, no entanto, não faz uso de ontologias bem como não utiliza algoritmo ou modelo de reputação nos membros da sua rede. A figura 5.4 mostra a rede de colaboração entre os pesquisadores de um mesmo INCT.

Outro projeto de iniciativa do *InWeb*, o Observatório da *Web*⁶, refere-se à descoberta de

⁵Disponível em: <http://www.pbct.inweb.org.br/pbct/>

⁶Disponível em: <http://observatorio.inweb.org.br/>

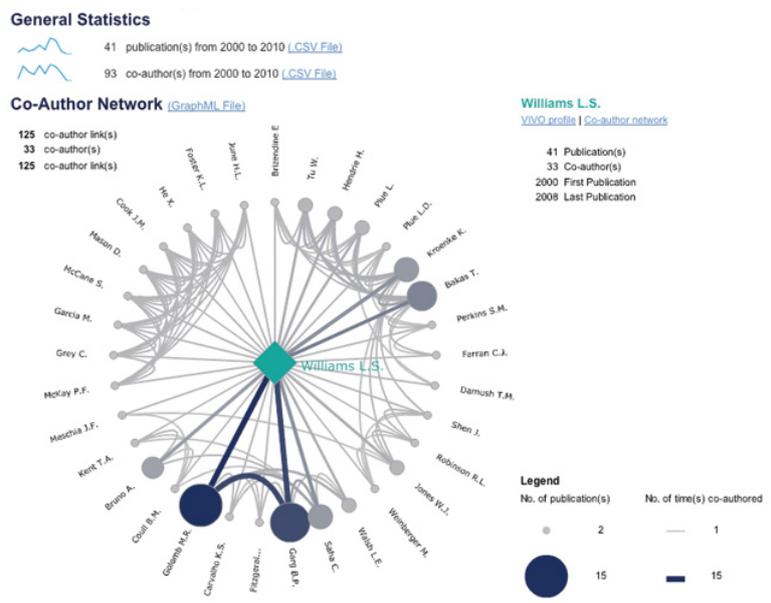


Figura 5.3: Rede de co-autoria de um pesquisador no portal VIVO.

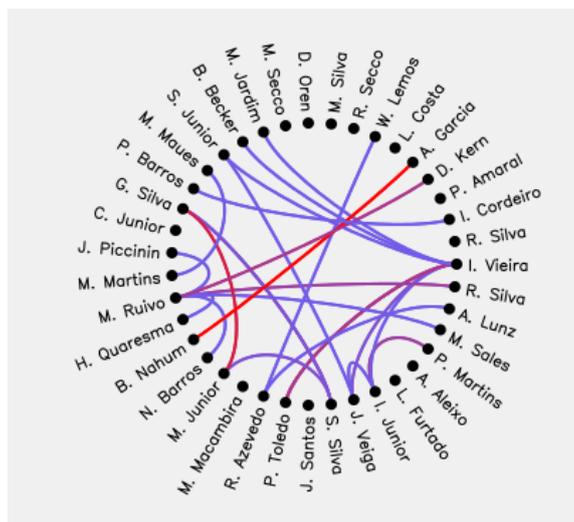


Figura 5.4: Rede de colaboração entre pesquisadores de um INCT.

informações através de redes sociais *on-line* (Meira et al., 2010). O portal foi criado para apresentar ao usuário uma síntese do que está sendo falado sobre determinados assuntos (Eleições de 2010, dengue, copa do mundo 2010) nas mais diversas fontes de conteúdo da *Web*, incluindo jornais, revistas, portais, redes sociais e o *Twitter*. A ferramenta apresenta a informação ao usuário em grafo. A Figura 5.5 mostra quais palavras relacionadas a Dengue normalmente aparecem juntas no *Twitter*, onde a espessura da linha mostra com que frequência as duas aparecem juntas. A ferramenta faz uso de alguns métodos de avaliação de rede para aferir critérios julgados relevantes para a aplicação e para o usuário, contudo não utiliza ontologia como recurso semântico de consulta.



Figura 5.5: Mapa de Co-ocorrência de termos no Observatório da *Web*.

Dentre os trabalhos de cálculo de reputação, a popularização do *PageRank* fez com que diversos trabalhos no âmbito acadêmico utilizassem deste algoritmo para a análise da reputação de pesquisadores (Page et al., 1999). O algoritmo *PageRank* tem o objetivo de estimar a importância de uma página na *Web* baseado na reputação das páginas que fazem referência a ela. Dessa forma páginas na *Web* que são referenciadas por várias páginas de alta reputação tornam-se mais importantes. Trabalhos como os de Gollapalli et al. (2011), Singh et al. (2011) e Tang et al. (2008), aplicam versões adaptadas do *PageRank* a conjuntos de dados provenientes de bibliotecas digitais a fim de ordenar autores, artigos e conferências.

Tang et al. (2008) na ferramenta *ArnetMiner* utilizam um método baseado na avaliação do candidato pelo usuário, na propagação da reputação e nos relacionamentos existentes entre os pesquisadores da rede para encontrar especialistas em um determinado assunto, onde para calcular a reputação são utilizadas informações obtidas do perfil do pesquisador. No *ArnetMiner*, se uma pessoa tem co-autoria em vários documentos ou se o nome dessa pessoa ocorre várias vezes em um determinado tópico de pesquisa, essa pessoa tem potencial para ser especialista neste tópico. Ou ainda, se uma pessoa conhece várias pessoas em um tópico de pesquisa ou se o nome dessa pessoa co-ocorre várias vezes em publicações de especialistas nesta área, então esta pessoa pode ser uma especialista.

O método definido por Singh et al. (2011) é utilizado para classificar as publicações de várias áreas de pesquisa em diversas conferências. Esta classificação é utilizada também para ordenar as conferências e autores. Singh et al. (2011) consideram que a importância da publicação é diretamente proporcional ao número de trabalhos que a referenciam, além de a qualidade da conferência depender da qualidade de suas publicações e não apenas da quantidade. Neste método, a reputação do autor depende não somente da qualidade de suas publicações mas também das conferências em que são publicadas.

Gollapalli et al. (2011) argumentam que em bibliotecas digitais diversas fontes de dados estão disponíveis, e.g. a localização da publicação e citações para outros documentos. Baseado nestas informações, os autores propõem um método que utiliza outras fontes de dados além das citações e co-autorias para a classificação de autores.

Os trabalhos de Matsuo et al. (2004) e Hogan et al. (2006) adaptam o *PageRank* para ordenar informações da *Web Semântica*, sendo que o último leva em consideração o contexto em que os dados se encontram. Outro algoritmo clássico para classificação de páginas na *Web* é o HITS (Kleinberg, 1999). HITS estima a importância das páginas através do cálculo da pontuação dos atributos *hub* e autoridade, onde a pontuação de um membro depende da pontuação do outro.

Diferentemente dos métodos anteriores, os quais consideram a rede de citação para o cálculo da reputação, o modelo proposto por Cervi et al. (2011) considera toda a trajetória acadêmica do pesquisador para aferir a sua reputação, utilizando um modelo de perfil adaptativo formado por elementos obtidos de seu currículo *Lattes*. A Tabela 5.1 apresenta as categorias definidas, seus elementos e as siglas de cada um deles entre parênteses.

Para auxiliar na identificação de tais elementos e representar o perfil do pesquisador, o modelo proposto por Cervi et al. (2011) faz uso de uma ontologia, que por sua vez é dividida em cinco categorias:

1. *Identification* - representa a titulação (*Education Degree*) do pesquisador, cujas opções são mestrado, doutorado ou pós-doutorado;
2. *Advisory* - representa as orientações do pesquisador, onde o elemento *Master Dissertation Advisor* representa as orientações de mestrado e o elemento *Phd Thesis Advisor* representa as orientações de doutorado. E as orientações de pós doutorado são representadas pelo elemento *Postdoctoral Advisor*;
3. *Examining Board* - indica a participação em bancas examinadoras pelo pesquisador, onde os elementos *Participation in Examination Boards Master Dissertation* e *Participation in Examination Boards Phd Thesis* representam a participação em bancas de mestrado e de doutorado, respectivamente;
4. *Membership* - representa as conferências onde o pesquisador foi coordenador (*Conference Committee Coordinator*) ou membro de comitê de programa (*Conference Committee Member*), bem como se foi ou ainda é revisor de periódico (*Reviewer of Journals*) ou membro de corpo editorial (*Editorial Board Member*);
5. *Production* - representa a produção científica e outros elementos relacionados a produção do pesquisador. O elemento *Articles in Scientific Journals* representa os artigos publicados em periódicos. Capítulos e livros publicados são representados pelos elementos *Book Chapter Published* e *Books Published*, respectivamente. Artigos publicados em conferências são representados pelo elemento *Complete Work*

Published in Conference Proceedings. Os elementos *H-Index* e *Network Co-authorship* dizem respeito às citações dos artigos do pesquisador por outros pesquisadores e ao número de pesquisadores que possuem trabalhos publicados em conjunto com o pesquisador avaliado. O elemento *Research Projects* representa o número de projetos de pesquisa que o pesquisador é membro. E os elementos *Software With Register* e *Software Without Register* representam os softwares com e sem registro em que o pesquisador teve participação.

Tabela 5.1: Categorias e elementos definidos no modelo de Cervi et al. (2011).

Categoria	Elemento
1. Identification (ID)	Education Degree (ED)
2. Advisory (ADV)	Postdoctoral Advisor (PA) Phd Thesis Advisor (PTA) Master Dissertation Advisor (MDA)
3. Examining Board (EB)	Participation in Examination Boards Phd Thesis (PEBPT) Participation in Examination Boards Master Dissertation (PEBMD)
4. Membership (MS)	Conference Committee Coordinator (CCC) Conference Committee Member (CCM) Editorial Board Member (EBM) Reviewer of Journals (RJ)
5. Production (PROD)	Articles in Scientific Journals (ASJ) Books Published (BP) Book Chapter Published (BCP) Complete Work Published in Conference Proceedings (CWPCP) H-Index (HI) Network Co-authorship (NC) Research Projects (RP) Software With Register (SWR) Software Without Register (SWOR)

A partir das categorias e elementos da Tabela 5.1, Cervi et al. (2011) propõe uma métrica para calcular a reputação do pesquisador. A métrica é definida por meio de um somatório entre os vários elementos que compõem o modelo, sendo que para isso são atribuídos pesos às categorias e elementos, cujo somatório dos pesos equivale a 100. Além dos pesos, a métrica faz uso do valor máximo obtido de cada elemento da Tabela 5.1 para cada pesquisador. Ao final do cálculo da reputação, o pesquisador é classificado em uma tabela de níveis de reputação que varia de 1 a 5 conforme apresenta a Tabela 5.2, os quais são utilizados na métrica proposta por Cervi et al. (2011).

Conforme exposto, as iniciativas de construção de rede social científica com modelo de reputação e aspectos semânticos para suporte à comunidade acadêmica brasileira, ainda não se apresentam de forma totalmente explorada. Nesta direção, este projeto propõe um modelo de reputação baseado no modelo de Cervi et al. (2011) e uma ontologia estendida de currículo *Lattes* e de grupos de pesquisa aplicados a rede social de pesquisadores

Tabela 5.2: Classificação da reputação definida no modelo de Cervi et al. (2011).

Nível	Reputação
1	≥ 0 e < 20
2	≥ 20 e < 40
3	≥ 40 e < 60
4	≥ 60 e < 80
5	≥ 80 e < 100

vinculados a UnB, construída à partir da integração de dados provenientes de diferentes repositórios nacionais, características ainda inexistentes em uma ferramenta no âmbito das IFES. No capítulo 6 serão descritos o modelo de reputação e a ontologia propostos neste trabalho.

Capítulo 6

Proposta

Neste capítulo é descrito o modelo de reputação com detalhamento da métrica utilizada. Também será descrita a ontologia estendida proposta neste trabalho.

6.1 Modelo de Reputação

Métricas científicas de reputação exercem um importante papel na comunidade acadêmica, visto que auxiliam no processo de medição da qualidade da produção científica e na identificação de especialistas em uma determinada área do conhecimento.

A maioria das métricas de reputação aplicadas à comunidade científica baseia-se na rede de citações de pesquisadores. No entanto, estes métodos de avaliação não consideram a trajetória acadêmica do pesquisador, onde a citação e a publicação são os principais critérios utilizados para o cálculo da reputação.

Diante desse contexto, Cervi et al. (2011) propôs um modelo de reputação mais abrangente englobando a vida acadêmica do pesquisador, que busca um equilíbrio entre diferentes elementos que fazem parte da trajetória científica de um pesquisador, ao invés de levar em consideração apenas métricas como o *H-index* ou o *G-index*, que baseiam-se apenas em citações bibliográficas. Dessa forma, o método de Cervi et al. (2011) (Capítulo 5), por considerar toda a trajetória acadêmica do pesquisador e ser um modelo flexível e adaptativo a todas as áreas de conhecimento, foi considerado mais adequado ao contexto desse projeto. A adaptabilidade do modelo envolve desde a inserção e eliminação de categorias e elementos, bem como o ajuste de seus pesos.

O objetivo do modelo de cálculo de reputação proposto neste trabalho é servir como um indicador de produção científica de pesquisadores e de especialistas em determinada área, de forma a contribuir com a interação e a colaboração entre os pesquisadores da comunidade acadêmica.

6.1.1 Metodologia

Assim como o modelo de Cervi et al. (2011), o modelo proposto considera os elementos da carreira acadêmica do pesquisador obtidos do Currículo *Lattes* e distribuídos em categorias ponderadas para o cálculo da reputação, onde os pesos são utilizados para privilegiar elementos considerados mais relevantes na área de pesquisa. A seguir são descritas as oito categorias definidas no modelo:

1. Identificação - corresponde à formação do pesquisador (graduação, mestrado, doutorado e estágio de pós-doutorado), sendo que somente a última formação do pesquisador será considerada;
2. Orientação - corresponde às orientações concluídas (iniciação científica, mestrado, doutorado, pós-doutorado);
3. Banca - corresponde às participações do pesquisador em bancas de mestrado e doutorado;
4. Membro - indica se o pesquisador participa ou já foi membro de corpo editorial de periódicos, de projeto de pesquisa ou se é revisor de periódico;
5. Bolsista - indica se o pesquisador recebe bolsa de produtividade em pesquisa do CNPq;
6. Prêmios e Títulos - corresponde à premiações ou títulos recebidos pelo pesquisador;
7. Produção - especifica a produção científica do pesquisador que corresponde a artigos publicados com e sem JCR (*Journal Citation Reports*¹), livros publicados ou organizados, capítulos de livros, trabalhos em eventos, textos de jornais ou revistas e demais tipos de produção bibliográfica, bem como a produção artística (partituras, notas musicais) e a técnica (softwares e demais tipos de produção técnica);
8. Co-autoria - corresponde aos pesquisadores vinculados ao pesquisador em análise através de publicações em co-autoria.

Cada categoria é dividida em elementos, que compõem a carreira da pesquisa conforme apresentado na Tabela 6.1.

6.1.2 Métrica

Assim como no modelo de Cervi et al. (2011), a métrica para calcular a reputação foi definida pelo somatório entre os vários elementos que compõem a carreira acadêmica do pesquisador. Visando privilegiar elementos de maior significância de acordo com os parâmetros estabelecidos pela CAPES, o modelo utiliza pesos distintos entre os diversos elementos que compõem as oito categorias definidas conforme consta na Tabela 6.1. Devido as particularidades de cada área de conhecimento, no modelo proposto os pesos podem variar de acordo com a área em que o modelo for aplicado, onde os valores dos pesos serão apresentados no Capítulo 7.

Para cada categoria é definida uma fórmula para se obter o valor da reputação correspondente, com os pesos definidos conforme aplicação da métrica. À seguir são apresentadas as fórmulas que compõem a métrica para calcular a reputação dos pesquisadores:

1. Identificação (IDE) - corresponde ao peso de apenas um dos elementos da categoria, considerando apenas a formação máxima;

$$Valor_{IDE} = Peso_{IDE}$$

¹O *Journal Citation Reports* é um recurso definido por Reuters (2012) que permite avaliar e comparar publicações científicas utilizando dados de citações extraídos de revistas acadêmicas e técnicas e o impacto destas na comunidade científica. O JCR permite verificar os periódicos mais citados em uma determinada área e a relevância da publicação para a comunidade científica por meio do Fator de Impacto.

Tabela 6.1: Categorias e elementos definidos no modelo de reputação.

Categoria	Elemento
1. Identificação (IDE)	Pós-doutorado (PD) Doutorado (DOU) Mestrado (MES) Graduação (GRA)
2. Orientação (ORI)	Pós-doutorado (PD) Doutorado (DOU) Mestrado (MES) Iniciação Científica (INI)
3. Banca (BAN)	Doutorado (DOU) Mestrado (MES)
4. Membro (MEM)	Corpo Editorial (CE) Revisor de Periódico (RP) Projeto de pesquisa (PP)
5. Bolsista (BOL)	Bolsista CNPq (BOL)
6. Prêmios e Títulos (PT)	Prêmios e Títulos (PT)
7. Produção (PROD)	Artigos publicados com JCR (ARPJ) Artigos publicados sem JCR (ARPS) Livros publicados ou organizados (LIV) Capítulos de livros (CAP) Trabalhos em eventos (EVE) Texto em jornal ou revista (TEX) Demais tipos de produção bibliográfica (DEM) Técnica (TEC) Artística (ART)
8. Co-autoria (COA)	Pesquisador (PES)

2. Orientação (ORI) - corresponde ao somatório da multiplicação de cada elemento da categoria 2 pelo respectivo peso;

$$Valor_{ORI} = Ocorrencias_{INI} * Peso_{INI} + Ocorrencias_{MES} * Peso_{MES} + Ocorrencias_{DOU} * Peso_{DOU} + Ocorrencias_{PD} * Peso_{PD}$$

3. Banca (BAN) - corresponde ao somatório da multiplicação de cada elemento da categoria 3 pelo respectivo peso;

$$Valor_{BAN} = Ocorrencias_{MES} * Peso_{MES} + Ocorrencias_{DOU} * Peso_{DOU}$$

4. Membro (MEM) - corresponde ao somatório da multiplicação de cada elemento da categoria 4 pelo respectivo peso;

$$Valor_{MEM} = Ocorrencias_{CE} * Peso_{CE} + Ocorrencias_{RP} * Peso_{RP} + Ocorrencias_{PP} * Peso_{PP}$$

5. Bolsista (BOL) - corresponde ao próprio peso do elemento da categoria Bolsista, sendo considerado apenas a bolsa vigente;

$$Valor_{BOL} = Peso_{BOL}$$

6. Prêmios e Títulos (PT) - corresponde a multiplicação da quantidade de ocorrências do elemento da categoria 6 pelo peso;

$$Valor_{PT} = Ocorrencias_{PT} * Peso_{PT}$$

7. Produção (PROD) - corresponde ao somatório da multiplicação de cada elemento da categoria 7 pelo respectivo peso;

$$Valor_{PROD} = Ocorrencias_{ARPJ} * Peso_{ARPJ} + Ocorrencias_{ARPS} * Peso_{ARPS} + Ocorrencias_{LIV} * Peso_{LIV} + Ocorrencias_{CAP} * Peso_{CAP} + Ocorrencias_{EVE} * Peso_{EVE} + Ocorrencias_{TEX} * Peso_{TEX} + Ocorrencias_{DEM} * Peso_{DEM} + Ocorrencias_{TEC} * Peso_{TEC} + Ocorrencias_{ART} * Peso_{ART}$$

8. Co-autoria (COA) - nesta categoria, cada relacionamento de co-autoria possui um peso, que é atribuído de acordo com a classificação da reputação do pesquisador que compõe o relacionamento (ver Tabela 7.3). A fórmula para calcular a pontuação de co-autoria corresponde ao somatório dos pesos dos relacionamentos multiplicado pelo peso da categoria 8, onde i indica o número de relacionamentos.

$$Valor_{COA} = \sum_{i=0}^N Peso_{relacionamento} * Peso_{PES}$$

O valor final da reputação do pesquisador ($REP_{pesquisador}$) é identificado como Índice de Produção Científica (IPC) e corresponde ao somatório dos valores referentes a cada categoria dividido pela quantidade total de elementos das categorias. Como na primeira categoria apenas um dos elementos é considerado para cálculo, também será considerado apenas um elemento no quantitativo, neste caso o total de elementos considerado é 22:

$$REP_{pesquisador} = Valor_{IDE} + Valor_{ORI} + Valor_{BAN} + Valor_{MEM} + Valor_{BOL} + Valor_{PT} + Valor_{PROD} + Valor_{COA} / Total_{elementos}$$

6.1.3 Componentes do Modelo de Reputação

De acordo com a estrutura básica de um sistema de reputação apresentada na Seção 3.2, pode-se representar o sistema de reputação proposto para a rede social científica do ObserveUnB conforme é mostrado na Figura 6.1, com os seguintes componentes:

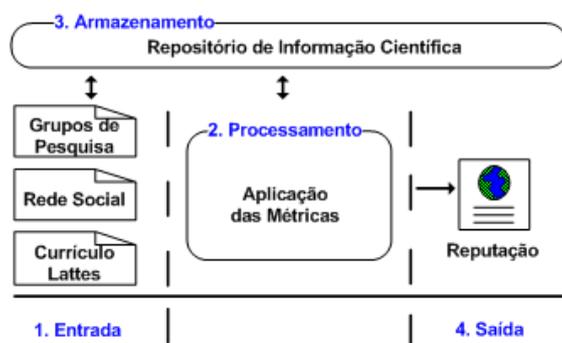


Figura 6.1: Estrutura do sistema de reputação proposto.

1. Entrada - refere-se às informações coletadas dos Currículos *Lattes*, dos Diretórios de Grupos de Pesquisa e da rede social científica utilizadas no cálculo da reputação;
2. Processamento - neste componente são implementadas as métricas apresentadas na Seção 6.1.2 que correspondem ao somatório dos valores referentes a cada categoria dividido pela quantidade de elementos definidos na Tabela 6.1;
3. Armazenamento - as informações coletadas e processadas são armazenadas no repositório de informações científicas;
4. Saída - a reputação do pesquisador pode ser classificada de acordo com uma tabela de níveis numéricos, conforme apresentada na Tabela 7.3 da Seção 7.2.1 e disponibilizada para consulta.

Nota-se que o componente *Feedback* definido no sistema de reputação da Figura 3.2 não é aplicado no modelo proposto, visto que as opiniões do usuário não são coletadas pelo sistema de reputação.

6.1.4 Tipo do Sistema de Reputação

Com base no que foi exposto e na tipologia definida por Mui (2002), apresentada na Seção 3.3, pode-se classificar o sistema de reputação no modelo proposto como:

- Individual - a reputação é calculada para cada membro da rede;
- Indireta - os elementos utilizados no cálculo são obtidos de experiências do pesquisador registradas em seu currículo *Lattes*;
- Derivada grupo - a tabela de níveis de classificação da reputação do pesquisador é definida de acordo com o resultado do cálculo da reputação do grupo avaliado (Ciência da Computação da UnB);

- Propagada - a propagação da reputação de um indivíduo é utilizada no cálculo da pontuação da categoria Co-autoria, onde o valor parcial da reputação do indivíduo relacionado ao pesquisador avaliado é considerado na métrica.

6.1.5 Modelo de Reputação Proposto e o Modelo de Cervi et al. (2011)

O modelo de reputação proposto difere do modelo de Cervi et al. (2011) com relação à métrica, às categorias e aos elementos utilizados no cálculo. Alguns elementos do modelo original foram suprimidos, em contrapartida foram adicionados outros elementos e categorias. Fazendo uma comparação com o modelo original, na Tabela 6.2 os elementos são destacados em três cores, onde: (i) a cor azul representa os elementos e as categorias que não existem no modelo de Cervi et al. (2011) e foram acrescentados no modelo proposto; (ii) a cor vermelha representa os elementos e categorias do modelo original que foram suprimidos no modelo proposto e (iii) a cor preta indica os elementos que existem no modelo de Cervi et al. (2011) mantidos no modelo proposto.

Os elementos incluídos no modelo proposto visam adequá-lo à realidade acadêmica da UnB, bem como abranger todas as áreas de conhecimento, conforme descritos a seguir:

- Foi incluído o elemento Graduação (Identificação) para abranger os docentes com esta titulação máxima, os quais correspondem ao total de 22 docentes, segundo informações do Decanato de Pesquisa e Pós-Graduação (DPP)/UnB de 2011;
- O elemento Iniciação Científica (Orientação) justifica-se devido ao fato de a UnB investir fortemente em programas de iniciação científica;
- A categoria Bolsista visa incentivar os docentes à adesão de bolsas de produtividade em pesquisa (PQ) do CNPq;
- A categoria Prêmios e Títulos incentiva a participação em concursos realizados na comunidade científica e a melhoria na pesquisa e produção científica;
- A distinção entre artigos publicados com JCR e sem JCR na categoria Produção segue recomendações da CAPES para todas as áreas, que incentiva a publicação em jornais, revistas e periódicos internacionais;
- Os elementos Texto em jornal ou revista, Demais tipos de produção bibliográfica e Técnica e Artística da categoria Produção visam englobar todos os dados referentes à produção de um pesquisador disponibilizados no ObserveUnB, abrangendo a comunidade acadêmica das diversas áreas de ciências.

Em contrapartida, alguns elementos que constam no modelo de Cervi et al. (2011) foram excluídos do modelo proposto. Em razão de os elementos *H-index*, Membro e Coordenador de Comitê de Conferência não estarem identificados de forma computável no currículo *Lattes* não foi possível contemplá-los no modelo. Os elementos Software com registro e sem registro são englobados no elemento Técnica da categoria Produção.

Outra diferença entre os modelos encontra-se na métrica definida para o cálculo. Em seu modelo, Cervi et al. (2011) considera no cálculo o valor máximo da reputação em cada categoria na rede de pesquisadores e a quantidade de níveis de classificação, tornando-se

Tabela 6.2: Categorias e elementos.

Categoria	Elemento
Identificação	Pós-doutorado Doutorado Mestrado Graduação
Orientação	Pós-doutorado Doutorado Mestrado Iniciação Científica
Banca	Doutorado Mestrado
Membro	Coordenador de Comitê de Conferência Membro de Comitê de Conferência Corpo Editorial Revisor de Periódico Projeto de pesquisa
Bolsista	Bolsista CNPq
Prêmios e Títulos	Prêmios e Títulos
Produção	Artigos publicados com JCR Artigos publicados sem JCR Livros publicados ou organizados Capítulos de livros Trabalhos em eventos H-index Texto em jornal ou revista Demais tipos de produção bibliográfica Software com registro Software sem registro Técnica Artística
Co-autoria	Pesquisador

necessário o conhecimento do valor máximo por categoria. No modelo proposto, o valor da pontuação referente às categorias de 1 a 7 da tabela 6.1 é calculado para cada co-autor relacionado ao pesquisador avaliado, visando obter o valor da pontuação da categoria co-autor. Além disso, na métrica proposta a reputação não possui um valor limite, ela cresce à medida em que aumenta o número de ocorrências dos elementos avaliados no currículo *Lattes* do pesquisador.

Para maiores esclarecimentos sobre o modelo de reputação será apresentado um estudo de caso no Capítulo 7.

6.2 Ontologia

Nos últimos anos, o aumento exponencial dos dados disponíveis tem conferido importância significativa às técnicas de organização da informação. Além disso, informações relevantes estão dispersas em diferentes recursos de informação e o meio mais utilizado para encontrá-las é através de máquinas de busca. Entretanto, as máquinas de busca tradicionais recuperam uma lista de recursos, oferecendo pouca ou nenhuma informação sobre as relações semânticas existentes entre eles. Por conseguinte, o usuário tem de despender uma quantidade substancial de tempo acessando-os e lendo-os, para então descobrir como esses recursos de informação estão relacionados.

No contexto dos sistemas acadêmicos não é diferente, e visando organizar e acrescentar a semântica aos dados, foi estendida uma ontologia à partir da integração das ontologias de Currículo *Lattes* e de Diretório de Grupos de Pesquisa da comunidade Conscientias (Conscientias, 2002, 2006), apresentadas na Seção 4.10. As Seções 6.2.1 e 6.2.2 descrevem o processo de extensão das ontologias.

6.2.1 Linguagem e Ferramenta

Dentre as diversas ferramentas disponíveis no mercado para a construção de ontologias relacionadas na Seção 4.7, a ferramenta Protégé e a linguagem OWL 2 foram utilizadas no processo de integração e extensão das ontologias. A escolha da linguagem OWL 2 deve-se às seguintes características:

- Contém as primitivas para a representação das categorias, conceitos e relacionamentos da ontologia, ou seja, as classes, subclasses e propriedades. Além disso, estas primitivas oferecem a expressividade necessária, tais como a representação da transitividade entre as classes e a definição de classes a partir da combinação de outras classes;
- Permite a inferência lógica realizada por mecanismo de raciocínio automático (FaCT++, Pellet, RACER, conforme citados na Seção 4.8);
- É amplamente utilizada no meio acadêmico;
- Permite a utilização da ontologia em ambientes *Web*;
- Pode ser gerada por meio de um editor de ontologias, o Protégé;
- Pode ser manuseada por meio da API Jena, conforme apresentado na Seção 4.9.

A principal desvantagem da linguagem OWL 2 é a sintaxe complexa que dificulta a implementação e manutenção da ontologia. Essa dificuldade foi mitigada por meio da utilização de um editor, o Protégé, e de um raciocinador automático, o Pellet. O primeiro gerou automaticamente o código em OWL 2 a partir da interface gráfica, enquanto o segundo permitiu checar a consistência da ontologia durante o processo de extensão, de forma a identificar rapidamente a inserção de definições inconsistentes (conforme apresentado na Seção 4.8).

Em relação às ferramentas apresentadas na Seção 4.7, foi escolhido o Protégé conforme suas principais funcionalidades:

- A interface gráfica que permite ao desenvolvedor concentrar-se na modelagem das entidades sem se preocupar com o formato do arquivo ou linguagem de saída;
- A flexibilidade na utilização de *plugins* que oferecem serviços como a representação gráfica da ontologia e sua exportação para outros ambientes;
- É um *software open source* e de livre acesso sendo amplamente utilizado pela comunidade acadêmica;
- Oferece vasta documentação e listas de discussão na *Web*;

Para auxiliar no processo de extensão e entendimento da ontologia, o *plugin* OWLViz foi utilizado para auxiliar na visualização (Horridge, 2004). Este *plugin* permite que todas as classes da ontologia sejam visualizadas em forma de grafo, onde é representada a hierarquia de conceitos, conforme exibido nas Figuras 4.3 e 4.4. Nos grafos, as classes são expressas como nós, os arcos expressam os relacionamentos e são capazes de expressar a herança múltipla, contudo as propriedades das classes não são expressas nos grafos. Ao selecionar um nó ou aresta, a ferramenta exibe a informação detalhada em outra interface.

A escolha do Pellet como motor de inferência deve-se à facilidade de uso junto ao Protégé e às características do raciocinador, tais como a rapidez de execução e principalmente a sua adequação à linguagem OWL, a qual outros raciocinadores não suportavam em sua completude.

6.2.2 Metodologia

Analisando as diversas metodologias conforme apresentadas na Seção 4.6, para a extensão das ontologias do Conscientias optou-se por seguir o Método 101 (Noy and McGuinness, 2001). O método 101 é considerado uma metodologia simples e comum a maioria dos métodos atualmente utilizados. Além disso, essa decisão deve-se ao fato de esta metodologia assemelhar-se aos métodos de Engenharia de Software, tornando mais fácil e intuitiva a sua utilização.

Por se tratar de um guia, o *Ontology Development 101* contribui com uma visão clara de como se dá interatividade no desenvolvimento de ontologias. Por isso, todos as fases do método foram adotadas no processo de extensão da ontologia proposta. A Figura 6.2 ilustra em (a) os sete passos sugeridos por Noy and McGuinness (2001) e em (b) um exemplo de como os passos podem ser empregados durante o desenvolvimento de uma ontologia. As Seções de 6.2.2 a 6.2.2 descrevem os passos adotados para a extensão da ontologia deste trabalho.

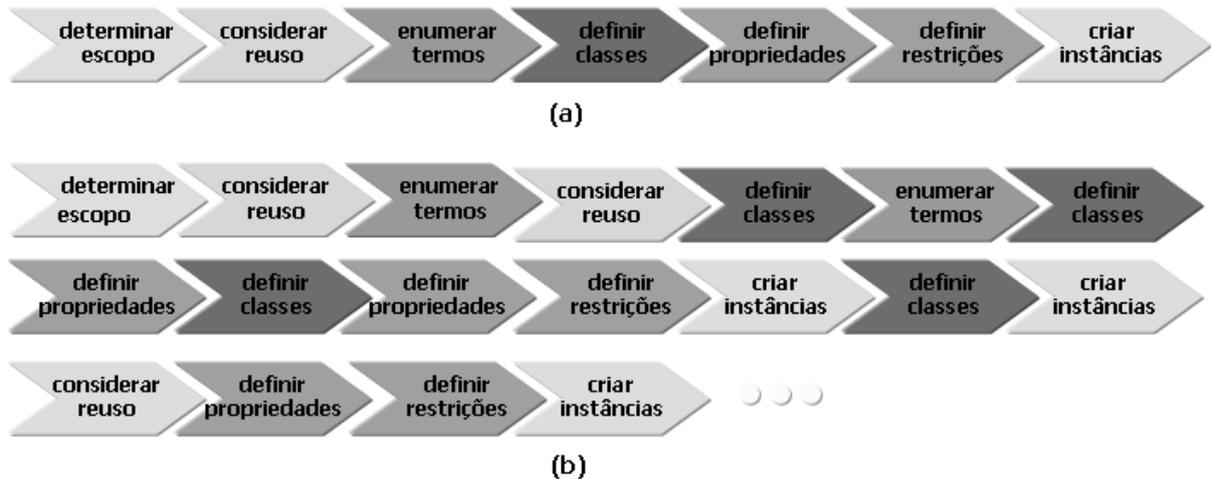


Figura 6.2: Processo de desenvolvimento do Método 101 (Noy and McGuinness, 2001).

Passo 1: Determinar o Domínio e Escopo da Ontologia

O primeiro passo consiste na identificação do domínio e do escopo da ontologia. Para isso foram identificados o propósito e os cenários de utilização da ontologia, onde o propósito consiste em auxiliar na representação do domínio de conhecimento da área de pesquisa de uma IFES, representando os elementos do modelo de reputação, bem como auxiliar na realização de consultas semânticas. O escopo consiste em representar a pesquisa e produção científica de uma IFES. Seguindo este escopo foram selecionados os seguintes conceitos para constar na ontologia:

- pesquisador;
- grupo de pesquisa;
- grande área de conhecimento, área de conhecimento, sub-área e especialidade;
- produção bibliográfica que corresponde à artigos publicados ou aceitos para publicação, livros e capítulos de livro, textos publicados em jornais ou revistas, trabalhos em eventos;
- produção técnica que corresponde à registro ou patente, software, trabalhos técnicos, processos ou técnicas e demais produções técnicas, e.g., maquete, editoração, relatório de pesquisa, organização de evento;
- produção artística e cultural;
- orientações concluídas;
- titulação que corresponde à formação acadêmica do pesquisador;
- prêmios e títulos;
- bolsistas em produtividade do CNPq;
- palavra-chave;

- instituição e unidade acadêmica.

Passo 2: Reutilizar Ontologias Existentes

Foram realizadas consultas em repositórios de ontologias *on-line* (e.g, o portal de ontologias OntoLP) buscando ontologias que se adequassem ao propósito deste trabalho. A partir desta busca foram identificadas as ontologias de Currículo *Lattes* e de Diretório de Grupo de Pesquisa do CNPq da comunidade Conscientias (Conscientias, 2002, 2006).

Como as ontologias de Diretório de Grupos de Pesquisa e de Currículo *Lattes* do Conscientias são disponibilizadas nas linguagens DTD e XML, e para este projeto foi escolhida a linguagem OWL para a extensão da ontologia, foi necessário a conversão das ontologias. Para auxiliar nesse processo, foi utilizada a ferramenta *TopBraid Composer* (TopQuadrant, 2011). Através da *TopBraid Composer* os esquemas de dados foram importados e migrados para a linguagem OWL.

O processo de conversão das ontologias para OWL gerou problemas como a falta de hierarquização das classes, e.g. as subclasses das classes *curriculoVitae* e *GrupoDePesquisa* forma hierarquizadas de forma incorreta. Após a correção manual dos erros, por meio da ferramenta Protégé foi realizado um *merge* das duas ontologias do Grupo Conscientias, ocasionando novos problemas como a duplicação de tipos de dados e de classes, e.g., as classes “Autores” e “Pesquisador” representando o mesmo conceito no domínio da aplicação. Dessa forma, manteve-se na ontologia final apenas a classe “Autores” representando o conceito de pesquisador.

Passo 3: Listar Termos Importantes

Os termos importantes foram relacionados de forma a serem englobados na ontologia. Para isto, foram listadas as questões que competem à ontologia responder:

1. Quem são os pesquisadores de determinada área de atuação?
2. Quem são os pesquisadores de determinada área de atuação que possuem produção científica?
3. Quem são os pesquisadores vinculados a determinada unidade acadêmica?
4. Quem são os pesquisadores vinculados a determinado grupo de pesquisa?
5. Quem são os co-autores de uma publicação?
6. Quem são os pesquisadores de determinada área de atuação que possuem publicação com JCR?
7. Quais são os termos relacionados a determinada área de conhecimento, sub-área ou especialidade?

As questões listadas em linguagem natural foram convertidas para as linguagens de Lógica Descritiva no Protégé e para a linguagem SPARQL de forma que possam ser executadas através do módulo de integração que utiliza a API Jena, conforme apresentados na Seção 7.2.2.

Passo 4: Definir Classes e Hierarquia

Dentre os termos selecionados, foram definidas as classes e a hierarquia entre elas. A ontologia resultante da integração entre as ontologias de currículo *Lattes* e de Diretório de Grupos de Pesquisa resultou em mais de 200 classes. No entanto, para atender ao escopo definido na subseção 6.2.2 foram necessárias 99 classes desta ontologia. Após a seleção das classes, foi incluída a hierarquia de classes para representar a estrutura acadêmica da UnB composta por 168 conceitos, visto que a instituição compõe o escopo deste trabalho. Na Figura 6.3 é apresentada a hierarquia de classes da unidade acadêmica Instituto de Ciências Exatas, composto por três departamentos - Matemática, Ciência da Computação e Estatística, incluídas na ontologia. A classe *Orgao* representa os órgãos governamentais, e.g., a UnB.

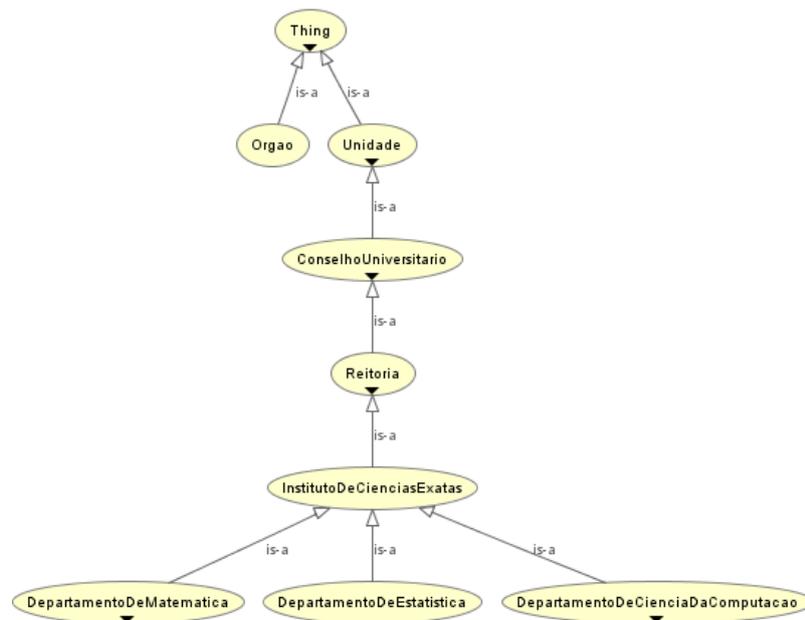


Figura 6.3: Amostra de classes da estrutura acadêmica da UnB.

Passo 5: Definir Propriedades das Classes

Definida a estrutura entre as classes, suas propriedades *datatype property* e *object property* foram selecionadas permitindo a identificação dos objetos representados e o relacionamento entre eles, sendo que a propriedade *bolsista* foi incluída na classe *Autores* para indicar se um pesquisador é bolsista de produtividade em pesquisa (PQ) do CNPq. A Figura 6.4 mostra a propriedade *data property* definida com a característica funcional,

visto que pode assumir apenas um dos seguintes valores: PQ1A, PQ1B, PQ1C, PQ1D, PQ2 e PQSr.

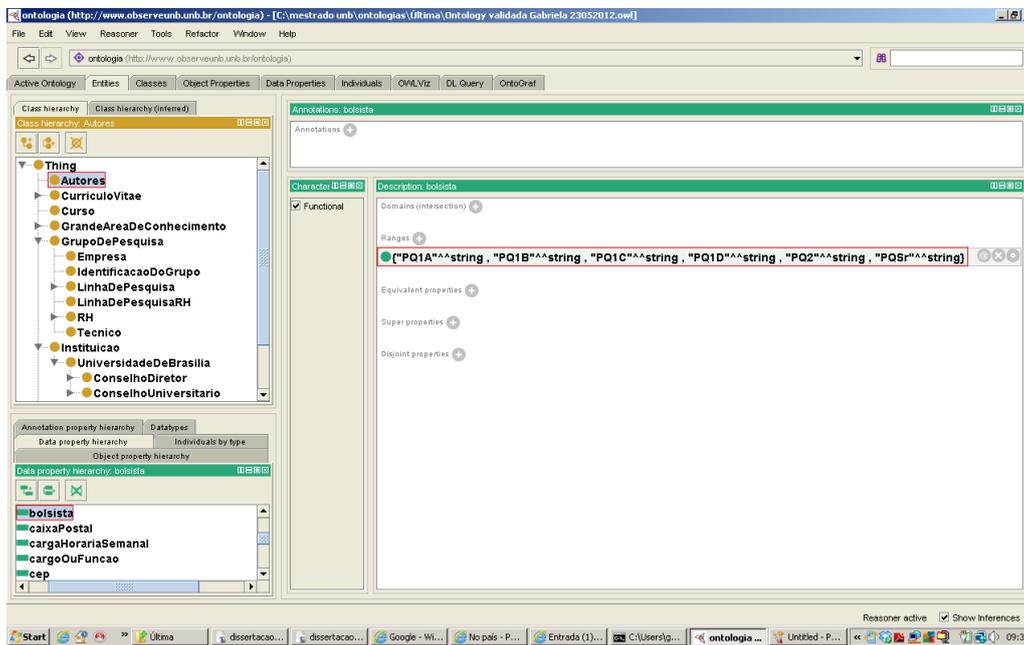


Figura 6.4: Propriedade *bolista* acrescentada na classe *Autores* da ontologia.

Outra propriedade relevante para o modelo de reputação incluída na ontologia é a *datatype property temJCR*. Essa propriedade indica se o artigo possui o fator de impacto JCR, podendo assumir os valores “Sim” ou “Não”, sendo que para isso a propriedade foi atribuída à classe *DetalhamentoDoArtigo*, conforme apresentado na Figura 6.5.

Passo 6: Definir Restrições das Propriedades

Foram definidas algumas restrições para as propriedades da ontologia. A Figura 6.6 mostra que a *object property especialidade* da classe *AreaDeAtuacao* está restrito a um domínio de classes *Domain*, onde pode estar ligado a apenas um dos indivíduos pertencentes às classes *Area*, *Especialidade* ou *SubArea*.

A Figura 6.7 exhibe a *object property temArea* que relaciona as classes *GrandeAreaDeConhecimento* e *Area*. Nota-se que a propriedade é Transitiva, ou seja, permite a transitividade no relacionamento, e.g., a grande área Ciências Exatas e da Terra possui a área Matemática, que por sua vez pertence à grande área Ciências Exatas e da Terra.

Conforme apresentado na Figura 6.8, a restrição funcional foi definida para a propriedade *object property temCurriculo* que por sua vez relaciona a classe *Autores* com a classe *CurriculoVitae*. Esta restrição de funcionalidade indica que um autor pode ter apenas um currículo *Lattes*.

Enquanto que a Figura 6.9, mostra que a restrição funcional foi retirada da propriedade *nomeParaCitacao* da classe *Autores*, visto que um autor pode ter mais de um nome para citação, solucionando o problema da ambigüidade de dados.

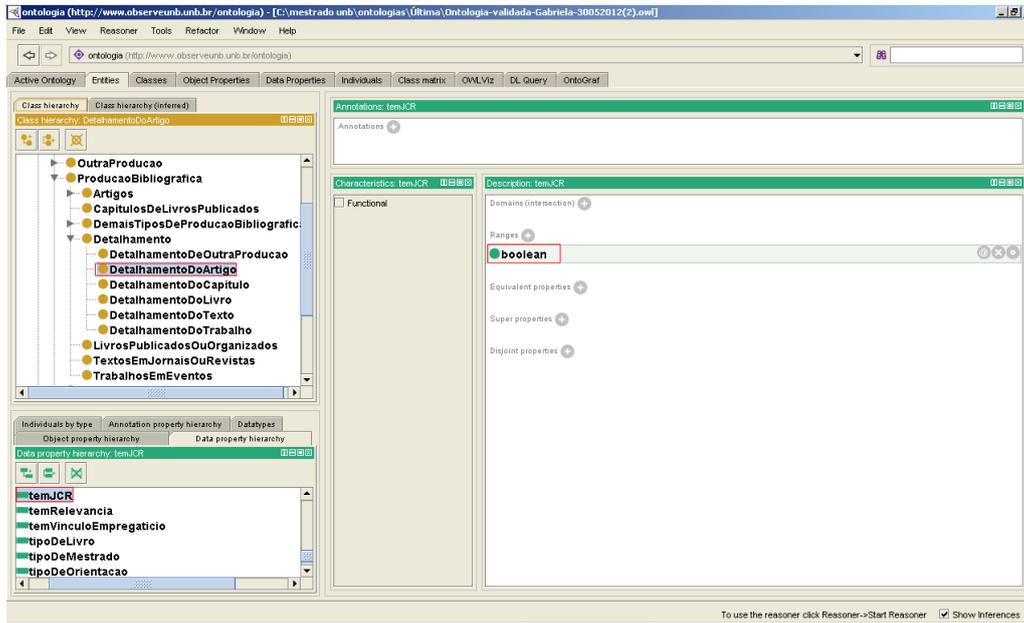


Figura 6.5: Propriedade *temJCR* acrescentada na classe *DetalhamentoDoArtigo* da ontologia.

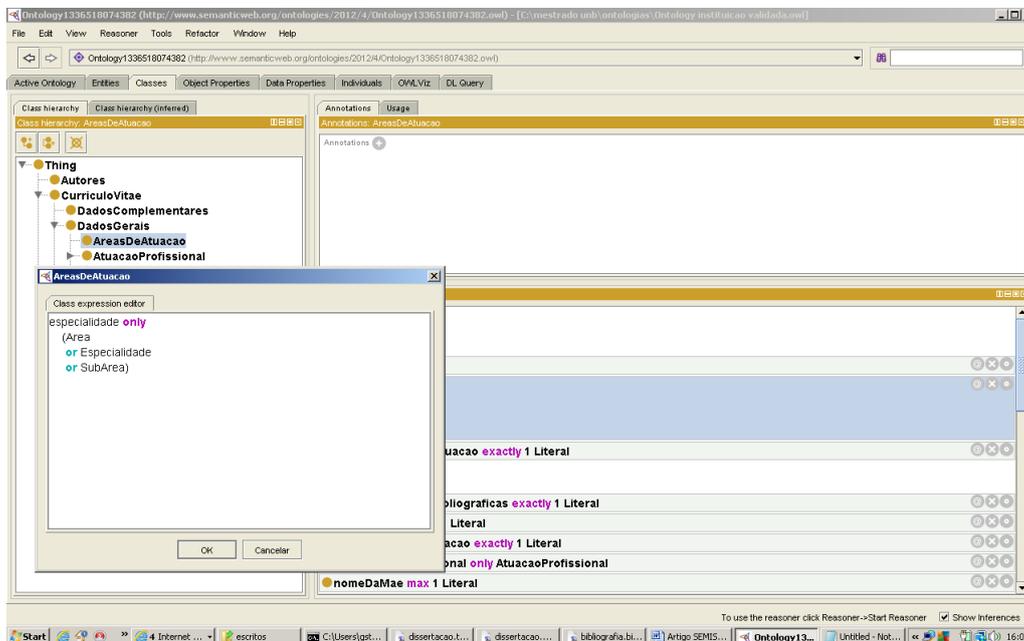


Figura 6.6: Restrição da propriedade *especialidade* da ontologia.

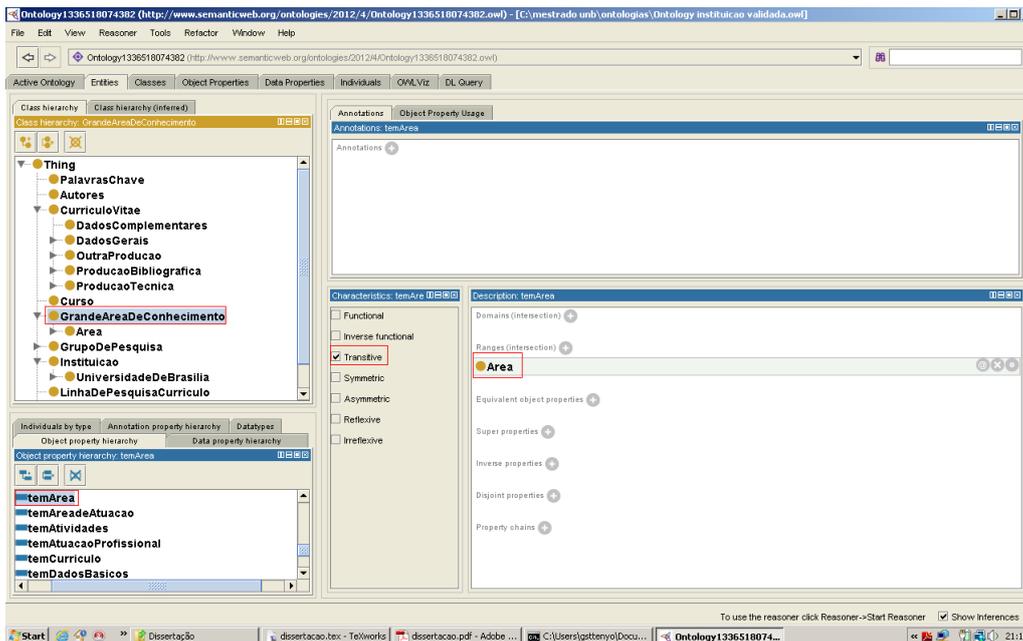


Figura 6.7: Restrição da propriedade *temArea* da ontologia.

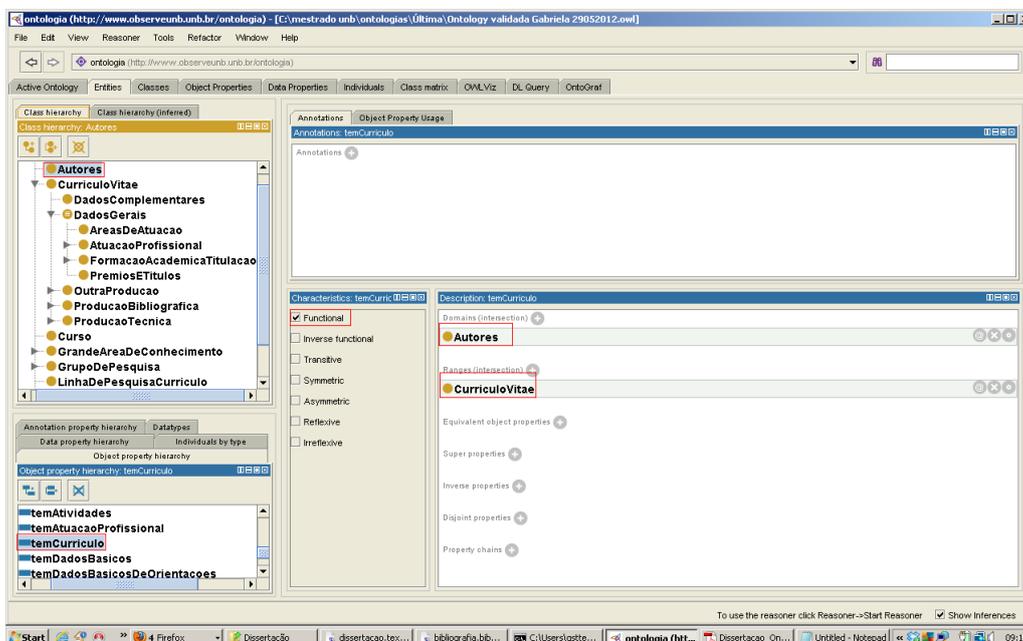


Figura 6.8: Restrição da propriedade *temCurriculo* da ontologia.

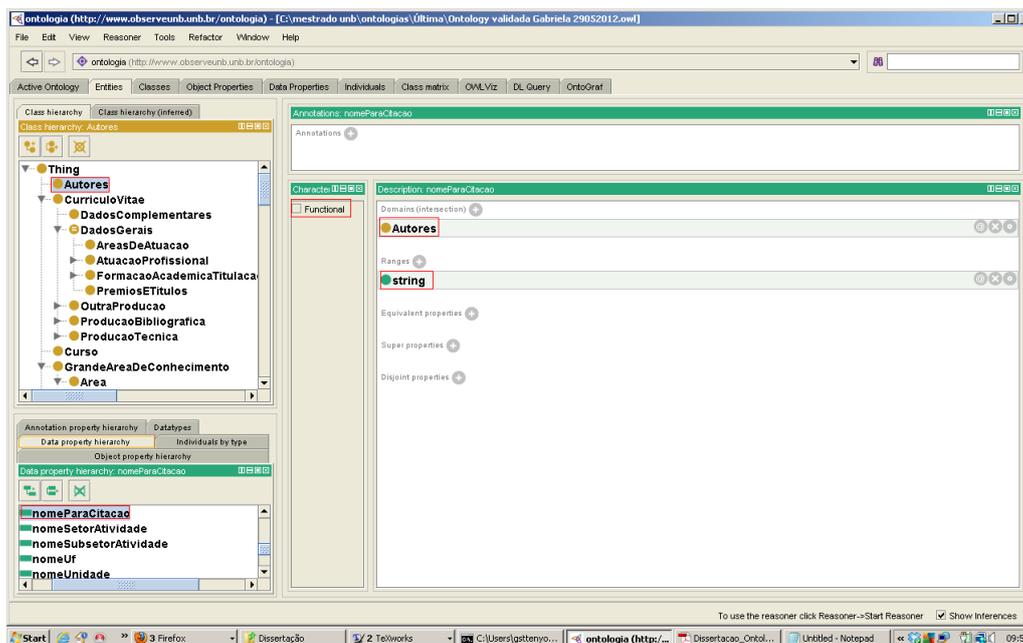


Figura 6.9: Propriedade *nomeParaCitacao* sem a restrição funcionalidade da ontologia.

Passo 7: Criar Instâncias

Em razão do tempo despendido a esta etapa, foram instanciadas uma amostra de dados nas classes julgadas mais importantes para o escopo deste trabalho as quais são listadas à seguir:

- Autores: foram instanciados os 49 autores (pesquisadores) do Departamento de Ciência da Computação da UnB com as propriedades nomeCompletoDoAutor, nomeParaCitacao, cpf, bolsista, pertenceGrupoDePesquisa e temCurriculo;
- CurriculoVitae: para cada um dos 49 autores foi instanciado um currículo com as propriedades temDadosGerais, temProducaoBibliografica, temProducaoTecnica e temOutraProducao;
- ProducaoBibliografica: para cada indivíduo da classe Autores foi instanciada uma produção bibliográfica, onde uma produção com mais de um autor foi instanciada apenas uma vez, solucionando o problema da duplicidade de dados;
- DadosGerais: para cada indivíduo da classe curriculoVitae foram instanciados os dados gerais com as propriedades temAtuacaoProfissional e temAreaDeAtuacao;
- AreasDeAtuacao: para cada indivíduo da classe DadosGerais foram instanciados os indivíduos na classe AreasDeAtuacao com a propriedade especialidade;
- AtuacaoProfissional: para cada indivíduo da classe DadosGerais foram instanciados os indivíduos na classe AtuacaoProfissional com a propriedade temAtividades;
- Atividades: para cada indivíduo da classe AtuacaoProfissional foram instanciados os indivíduos na classe Atividades com a propriedade pelaUnidade que relaciona-se a classe unidade;

- GrandeAreaDeConhecimento: foram criados indivíduos nesta classe para representar as 9 grandes áreas de conhecimento, conforme definido pela CAPES (CAPES, 2008);
- Area: foram instanciados todos os indivíduos que representam as áreas de conhecimento com as propriedades nomeDaArea, temSubArea e pertenceGrandeArea;
- SubArea: foram instanciados todos os indivíduos que representam as sub-áreas de conhecimento com a propriedade pertenceArea;
- GrupoDePesquisa: para cada indivíduo da classe Autores foi instanciado um GrupoDePesquisa para representar os grupos aos quais os pesquisadores pertencem;
- Unidade: foi instanciada a classe DepartamentoDeCienciaDaComputacao e algumas outras unidades acadêmicas da UnB, e.g. Instituto de Ciências Exatas, Departamento de Matemática, Departamento de Estatística.

Durante a instancição de algumas classes surgiram problemas relacionados a tipos, restrições e intervalos de valores de algumas propriedades. Por exemplo, a propriedade *nomeParaCitacao* da classe *Autores* originalmente foi definida para conter exatamente um indivíduo, no entanto, como diversos autores possuem 0 ou mais nomes para citação, esta propriedade foi alterada para conter no mínimo 0 indivíduos.

A Figura 6.10 apresenta a instância *Autor1* criada na classe *Autores* para representar uma pesquisadora, enquanto que a Figura 6.11 apresenta a instância *cur1* criada na classe *CurriculoVitae* para representar o currículo da instância *Autor1*, com foco nos artigos publicados em periódicos.

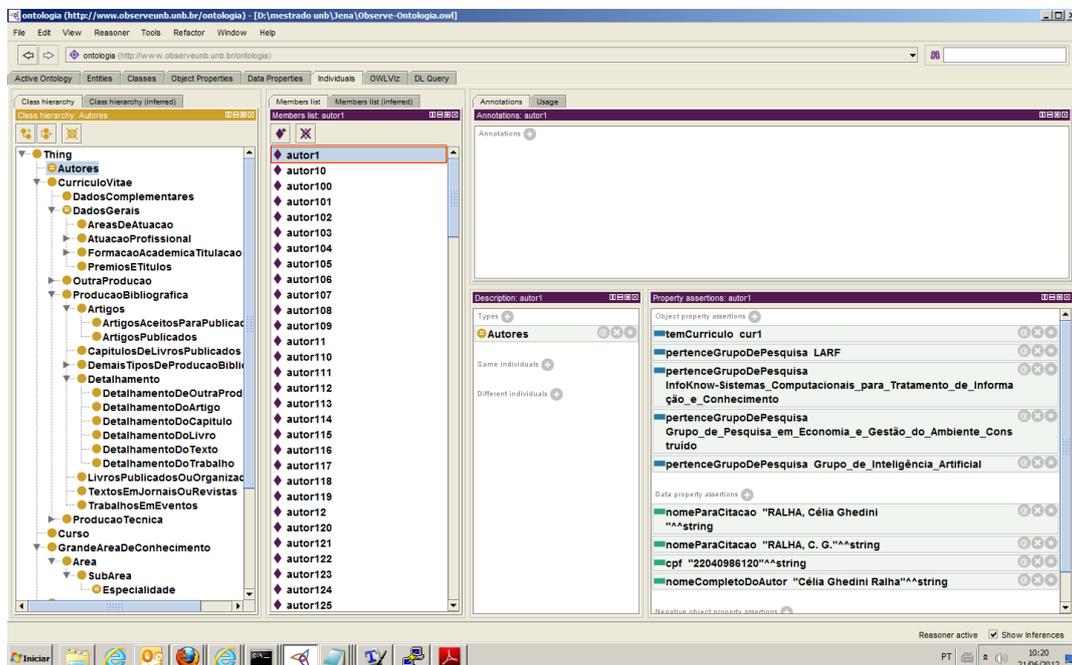


Figura 6.10: Instância da classe *Autores*.

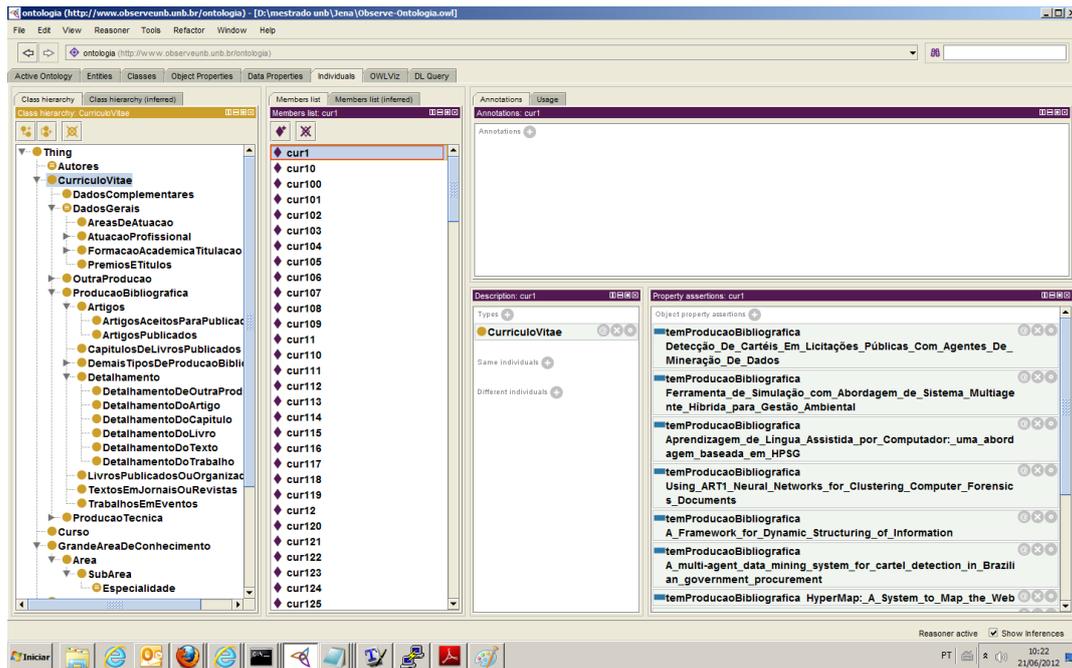


Figura 6.11: Instância da classe *CurrículoVitae*.

Neste capítulo foi apresentado o modelo de reputação definido com base no modelo de Cervi et al. (2011), além da ontologia estendida formada pela integração das ontologias de currículo *Lattes* e de diretório de grupos de pesquisa da comunidade *Conscientias*. No Capítulo 7 será apresentado um estudo de caso para ilustrar o uso do modelo proposto neste capítulo.

Capítulo 7

Estudo de Caso

Neste capítulo é apresentado o portal *Web* de rede social científica da UnB intitulado ObserveUnB. Esta ferramenta é utilizada como estudo de caso da aplicação do modelo de reputação e da ontologia estendida propostos neste projeto.

7.1 ObserveUnB: Visão Geral

O ObserveUnB é um projeto idealizado e coordenado pelo Decanato de Pesquisa e Pós-Graduação (DPP) da UnB e implementado pela equipe técnica vinculada: (i) ao Programa de Pós-Graduação de Informática (PPGInf) do Departamento de Ciência da Computação da UnB e (ii) ao Laboratório do Ambiente Construído, Inclusão e Sustentabilidade (Lacis) da Faculdade de Arquitetura e Urbanismo. Este projeto foi iniciado em março de 2011 e a primeira versão foi disponibilizada na *Web* à comunidade acadêmica em novembro de 2011.

A proposta da ferramenta consiste em servir como um observatório do conhecimento científico, com foco na pesquisa e na produção científica realizadas pelos docentes e discentes vinculados à instituição. O ObserveUnB permite que pesquisadores encontrem potenciais parcerias para pesquisa através dos relacionamentos apresentados na rede de colaboração científica. Desta forma, o ObserveUnB tem como objetivo principal responder quatro perguntas básicas relacionadas aos pesquisadores:

1. Quem são os pesquisadores?
2. O que os pesquisadores estão produzindo?
3. Onde os pesquisadores estão localizados?
4. Com quem os pesquisadores estão pesquisando?

A Figura 7.1 apresenta a arquitetura do ObserveUnB com seus quatro módulos básicos: extração, integração, repositório e interface de visualização. As Seções de 7.1.1 a 7.1.4 descrevem detalhes de implementação desses módulos. A versão disponível na *Web* utiliza os dados dos sistemas institucionais da UnB, do currículo *Lattes* dos docentes da UnB e as informações dos programas de pós-graduação da UnB disponíveis na CAPES, bem como faz uso do modelo de reputação e da ontologia propostos neste trabalho.

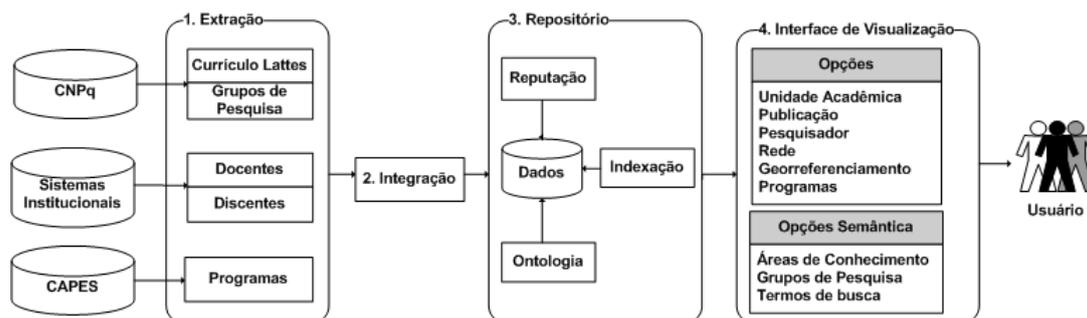


Figura 7.1: Arquitetura ObserveUnB.

7.1.1 Extração

Esse módulo abrange o processo de extração de dados dos docentes e discentes utilizados e armazenados no repositório, os quais são obtidos das bases de dados de sistemas institucionais da UnB: Sistema de Pessoal - SIPES, Sistema de Informação Acadêmica da Graduação – SIGRA, Sistema de Informação Acadêmica da Pós-Graduação - SIPPOS. Os dados dos 36.199 discentes e 2.181 docentes são fornecidos pelo Centro de Informática da UnB em arquivos texto com o formato CSV (*Comma Separated Values File*).

Os dados dos programas de pós-graduação da UnB, envolvendo 1.368 pesquisadores vinculados aos programas de pós, coletados através do aplicativo Coleta de Dados da CAPES, são extraídos utilizando o código da instituição e gerando arquivos do tipo texto convertidos para o formato CSV.

Para executar a extração dos currículos *Lattes* do CNPq foi implementado um robô na linguagem Java. O Robô-CV que utiliza os dados dos docentes e discentes da UnB obtidos dos sistemas institucionais (SIPES, SIGRA e SIPPOS), acessa a plataforma *Lattes* do CNPq através de um endereço IP pré-definido e procede à coleta dos arquivos de currículo *Lattes* na linguagem XML.

O processo de extração enfrentou vários desafios referentes ao grande volume de dados, seja pelo limite de quota para coleta diária definido pelo CNPq às instituições parceiras, ou pelo próprio processo de tratamento dos dados no repositório do ObserveUnB. Desta forma, a extração dos currículos *Lattes* está programada para ocorrer uma vez por mês, sendo que a última coleta realizada em abril de 2012 totalizou 14.369 currículos de uma população total de 38.380 indivíduos. Destes, 2.100 referem-se a docentes e 12.269 a discentes, o que representa um percentual de 96% dos docentes e 33% dos discentes que possuem currículo *Lattes*.

7.1.2 Integração

Os dados dos arquivos XML e dos arquivos texto com o formato CSV gerados durante o módulo de extração são filtrados através do módulo de integração, o qual foi desenvolvido na linguagem PHP, convertendo os dados armazenados em uma base de dados relacional MySQL, cuja estrutura foi modelada de acordo com as ontologias do Grupo Conscientias.

Durante a integração foram detectados alguns problemas como a inconsistência, a ambiguidade e a duplicação de dados: (i) pesquisadores vinculados a mais de uma unidade

acadêmica na instituição; (ii) pesquisadores sem currículo *Lattes*; (iii) pesquisadores diferentes com nomes iguais; (iv) pesquisadores com mais de um nome para citação; (v) publicação vinculada a mais de um autor e (vi) duplicação de pesquisadores na base.

No caso de dados incorretos, a correção foi realizada manualmente e na inexistência de currículo *Lattes*, apenas os dados institucionais do pesquisador foram integrados. Enquanto que no caso de duplicação e ambigüidade de dados, além do processo manual de correção, estão sendo utilizados os recursos da estrutura de classes e relacionamentos entre os conceitos fornecidos pela ontologia. Também estão sendo pesquisados algoritmos e funções de similaridade que calculam o quanto duas entidades são similares através dos relacionamentos e de termos extraídos dos conceitos da ontologia (Morris et al., 2008).

7.1.3 Repositório

Na versão atual do ObserveUnB, o repositório consiste em uma base de dados relacional composta por dados pessoais (nome, filiação, nacionalidade, cpf e vinculação acadêmica), dados dos programas de pós-graduação e currículos *Lattes* de pesquisadores, obtidos no módulo de extração, cujo tamanho total da base atual é de 970 Mb. Além desses dados, o repositório agrega a ontologia estendida do Diretório de Grupo de Pesquisa e do Currículo *Lattes* do Grupo Conscientias, além das informações da reputação dos membros da rede, conforme detalhados no Capítulo 6.

Para viabilizar a integração da ontologia ao repositório de informações científicas, foi implementado um módulo na linguagem Java utilizando a API Jena, utilizado para criar protótipos, manipular classes, instanciar e recuperar os conceitos da ontologia.

Tendo em vista a complexidade da ontologia estendida que engloba mais de 200 classes, bem como o escopo do projeto que representa a produção científica dos pesquisadores da rede da UnB, foram selecionados os seguintes conceitos para integrar a ontologia: (i) autores; (ii) grupo de pesquisa; (iii) grande área de conhecimento, área de conhecimento, sub-área e especialidade; (iv) produção bibliográfica (artigos publicados com e sem JCR, livros e capítulos de livro, textos publicados em jornais ou revistas, trabalhos em eventos); (v) produção técnica (registro ou patente, software, trabalhos técnicos, processos ou técnicas e demais produções técnicas); (vi) outra produção (produção artística e cultural, orientações concluídas); (vii) palavras-chave e (viii) órgão (universidade) e unidades (conselhos universitários, institutos, faculdades e departamentos).

A Figura 7.2 apresenta um substrato da ontologia do ObserveUnB, cujas classes originam de duas ontologias diferentes do Grupo Conscientias, sendo as classes representadas em amarelo provenientes do Currículo *Lattes* e as classes representadas em vermelho oriundas do Diretório de Grupos de Pesquisa. A classe unidade (em azul) e suas 168 subclasses representando a estrutura acadêmica da UnB foram incluídas na ontologia.

O repositório do ObserveUnB armazena também as informações de reputação dos integrantes da rede social científica. O objetivo do modelo de reputação proposto aplicado à rede do ObserveUnB é qualificar pesquisadores em suas áreas de pesquisa, servindo como um instrumento de colaboração e fomento à produção científica. Dentre os diversos métodos e algoritmos de reputação estudados, foi proposto um modelo de reputação (descrito na Seção 6.1) baseado no modelo definido por Cervi et al. (2011), uma vez que ele considera toda a trajetória acadêmica do pesquisador fazendo uso de um modelo de perfil adaptativo.

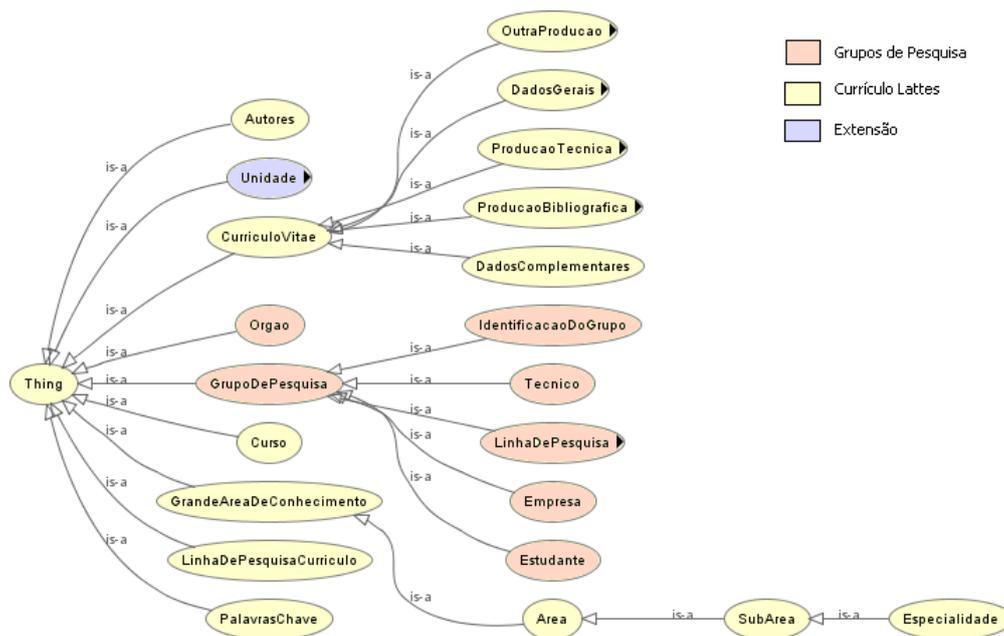


Figura 7.2: Amostra de classes da ontologia estendida utilizada no ObserveUnB.

7.1.4 Interface de Visualização

Para responder as quatro perguntas básicas de pesquisa apresentadas na Seção 7.1, o ObserveUnB disponibiliza ao usuário como opções de pesquisa as apresentadas na Figura 7.3: (i) dados da UnB; (ii) georreferenciamento; (iii) rede; (iv) programas; (v) publicação e (vi) pesquisador. Neste caso, foi informado pelo usuário a palavra-chave “Sistema de Informação” e como retorno, o sistema exibe uma lista de publicações que apresentam este termo como palavra-chave.

Ao selecionar a opção Unidade Acadêmica ou Dados da UnB o sistema exibe o total da produção científica da Universidade e o total por unidade acadêmica. Dado o nome de uma publicação, o sistema retorna detalhes da publicação. Dado o nome de um pesquisador, o sistema retorna informações de seu currículo, sua produção científica e os relacionamentos com outros pesquisadores ou redes de pesquisa. A opção de busca Programas exibe os pesquisadores vinculados aos programas de pós-graduação e o total de produção científica correspondente.

A Figura 7.4 mostra o quantitativo de produção científica dos pesquisadores vinculados ao Programa de Pós-graduação “A” (conceito 7 na CAPES).

A opção de Rede, apresenta a rede das unidades acadêmicas conforme mostra a Figura 7.5. Quando realizada a busca por uma unidade acadêmica específica o sistema apresenta os pesquisadores e a rede de colaboradores vinculada ao pesquisador.


Universidade de Brasília
 Decanato de Pesquisa e Pós-graduação

[Home](#) | [Dados da UnB](#) | [Georreferenciamento](#) | [Rede](#) | [Programas](#) | [Contato](#)



Buscar

Resultado da Pesquisa para o termo : *Sistema de Informação*

Estudo para a Implementação de um Sistema de Informação Logística do Agribusiness Brasileiro

Ano: 1998
[Mais detalhes](#)

Sistema de Informação Gerencial da Oficina Mecânica Central do Transporte Rodoviário de Carga de Nicarágua - Manuais Operacional das Áreas de Trabalho da Empresa - Sistemas Logísticos (Estoque-Produção)

Ano: 1987
[Mais detalhes](#)

Sistema de Informação Operacional da Oficina Mecânica Central do Transporte Rodoviário de Carga de Nicarágua

Ano: 1987
[Mais detalhes](#)

Figura 7.3: Modalidades de pesquisa do ObserveUnB.

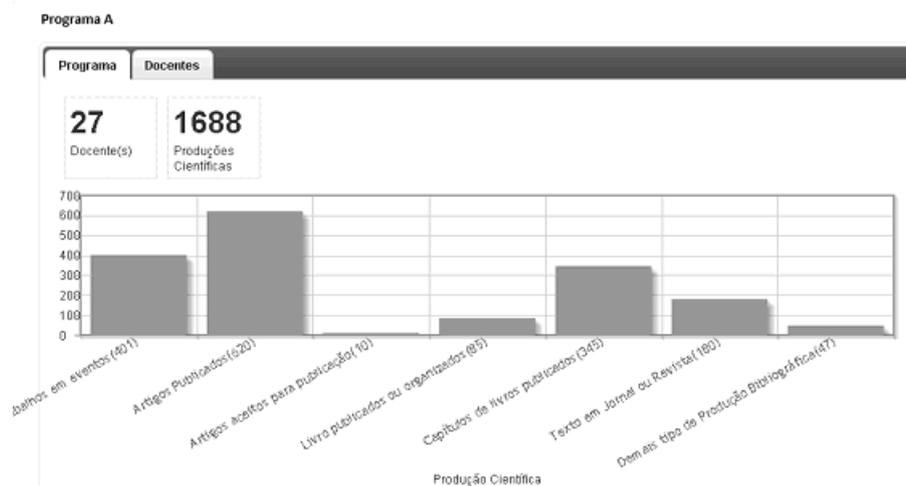


Figura 7.4: Total de produção científica do Programa de Pós-graduação “A” da UnB, nível 7 conforme a CAPES.

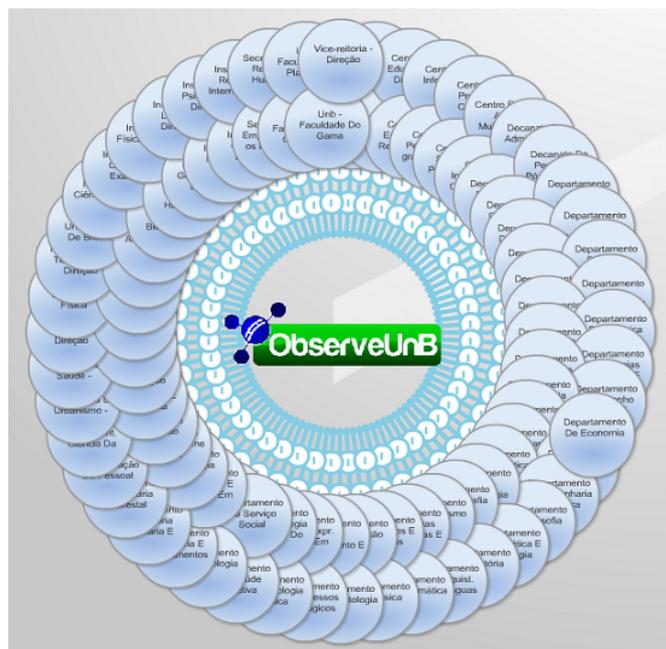


Figura 7.5: Rede das unidades acadêmicas da UnB.

Para localizar os pesquisadores dentro da Universidade, a ferramenta exibe um mapa obtido do Google Maps onde, a partir do georreferenciamento, são localizadas suas unidades acadêmicas. A Figura 7.6 exibe a tela com a localização georreferenciada dos quatro *campi* da UnB.

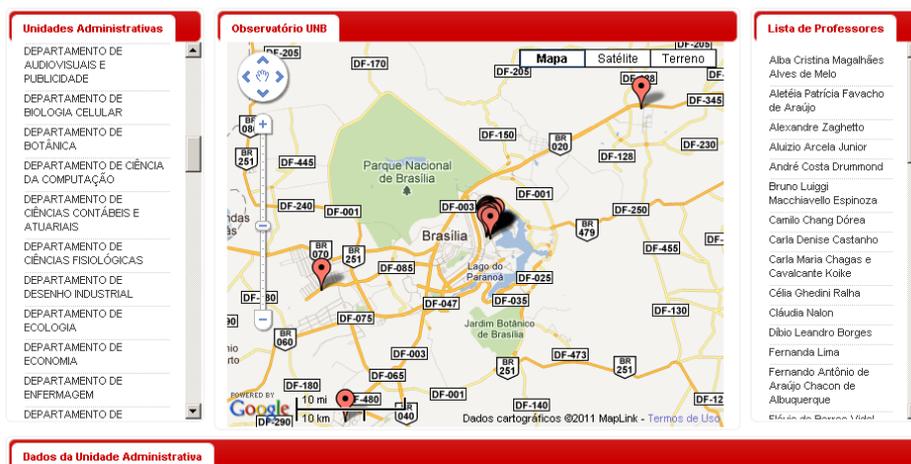


Figura 7.6: Pesquisa georreferenciada dos quatro *campi* da UnB.

Como nas demais redes sociais, o sistema permite a visualização de conexões entre pesquisadores, possibilitando ao usuário identificar as parcerias existentes. Foram definidos três tipos de relacionamentos na rede, a saber: (i) “T” – representa a integração do pesquisador a uma Unidade Acadêmica da UnB; (ii) “R” - representa a colaboração de pesquisa entre pesquisadores via produção científica e (iii) “P” – representa a participação

do pesquisador em determinado programa de pós-graduação da UnB, conforme é apresentado na Figura 7.7. Para a visualização das redes de relacionamento foi utilizada a biblioteca *Relation Browser* (Stefaner, 2012), destinada a apresentar estruturas de rede social a partir de redes de relacionamentos complexas.

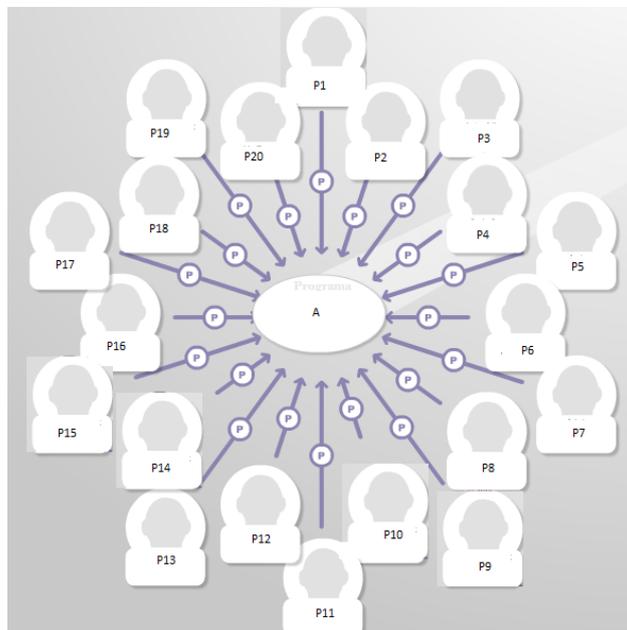


Figura 7.7: Rede de relacionamentos do programa de pós-graduação “A” da UnB.

Visando amenizar o impacto da divulgação do cálculo de reputação do pesquisador integrante da rede social científica da UnB, o valor da reputação do pesquisador está sendo apresentado no portal do ObserveUnB como Índice de Produção Científica – IPC. Este índice é apresentado para cada pesquisador membro da rede social do portal ObserveUnB como sendo o resultado da aplicação da métrica de reputação (ver Seção 6.1.2) definida a partir da Tabela 7.1 no modelo proposto. A Figura 7.8 apresenta a rede de um pesquisador “P31” e sua reputação (IPC).

Além das opções de pesquisa por pesquisador ou publicação, com o uso da ontologia é possível a realização de consultas mais elaboradas por meio de relacionamentos definidos entre os conceitos da ontologia e do mecanismo de inferência. A partir do módulo de integração desenvolvido para acessar a ontologia, o sistema monta consultas com os termos fornecidos na pesquisa pelo usuário, a partir dos quais são identificados os indivíduos que atendem ao critério de busca (conforme apresentado na Figura 7.25 da Seção 7.2.2).

No caso da busca por uma sub-área de uma área, por meio da restrição de transitividade definida na propriedade destes conceitos, será recuperada a sub-área associada à área, bem como as especialidades associadas à subárea, as quais são inferidas do relacionamento que associam os dois conceitos. A Figura 7.9 mostra um exemplo de busca da produção bibliográfica dos pesquisadores de uma área de conhecimento, onde por meio da restrição de transitividade definida no relacionamento entre esses conceitos, serão recuperados as produções e os pesquisadores que melhor atenderem ao critério de busca.

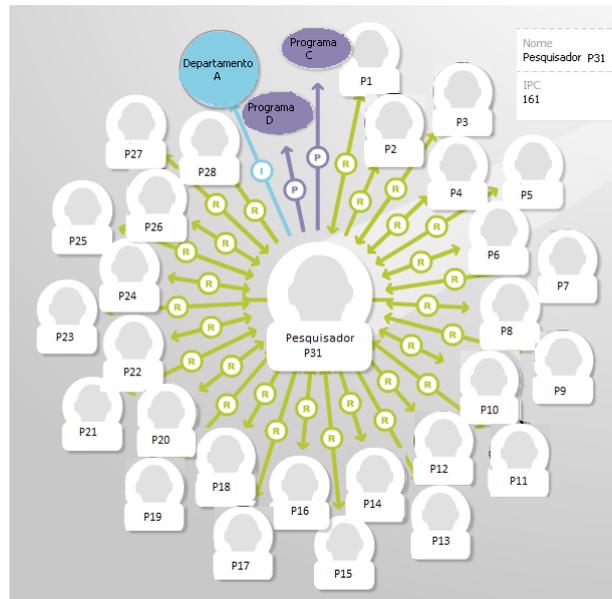


Figura 7.8: Rede de relacionamentos de um pesquisador da UnB.



Figura 7.9: Visualização da estrutura de classes definida na ontologia.

7.2 Avaliação da Proposta

O objetivo deste capítulo é apresentar o procedimento utilizado e os resultados obtidos na avaliação do modelo de reputação e da ontologia estendida. Para a avaliação do modelo de reputação foram realizadas entrevistas com uma amostra de 11 pesquisadores docentes do Departamento de Ciência da Computação da UnB. No que diz respeito à ontologia, esta foi classificada pelo motor de inferência Pellet fornecido pelo Protégé e posteriormente foram realizadas consultas à ontologia de forma que as questões delimitadas no escopo deste projeto, conforme apresentadas na Seção 6.2.2, fossem respondidas.

7.2.1 Modelo de Reputação

O modelo de reputação proposto na Tabela 7.1 foi definido a partir do modelo de (Cervi et al., 2011) em conjunto com alguns pesquisadores das áreas de Ciência da Computação, Estatística, Serviço Social, Psicologia e Veterinária. Neste modelo os pesos dos elementos foram definidos em consenso com as diferentes áreas dos pesquisadores envolvidos, onde buscou-se dar maior peso aos elementos de maior dificuldade e atribuir menor peso aos elementos de menor relevância no currículo do pesquisador. Tal modelo foi aplicado a rede social científica do ObserveUnB composta por 2.181 docentes, onde o resultado do cálculo da reputação encontra-se disponibilizado como IPC no portal do ObserveUnB.

Entretanto, com base na análise dos resultados obtidos a partir da implementação do modelo e em *feedbacks* de pesquisadores avaliados, percebeu-se a necessidade de refinamento do modelo visando atender às particularidades de cada uma das áreas.

Dessa forma, para a validação do modelo proposto foi selecionado o corpo docente da área de conhecimento Ciência da Computação da UnB composto por 49 docentes, cujos dados foram obtidos do ObserveUnB em abril de 2012.

No modelo proposto cada categoria é dividida em elementos, que por sua vez são relacionados a pesos, onde o somatório dos pesos das categorias é igual a 100. Para proceder a definição dos pesos dos elementos buscou-se embasamento na experiência de docentes da área Ciência da Computação, bem como no documento de área da Ciência da Computação (CAPES, 2009), onde por exemplo, priorizou-se a produção de artigos em periódicos com JCR (peso igual a 13) em desvalorizou-se a produção de livros (peso igual a 5) e capítulos (peso igual a 3). A Tabela 7.2 apresenta as categorias com seus elementos e respectivos pesos, e entre parênteses estão as siglas correspondentes aos elementos.

Classificação

Visando identificar a reputação do pesquisador em sua comunidade, o valor da reputação poderá ser classificado em uma tabela de níveis. Dependendo do contexto ou da área de utilização, pode-se definir quantos níveis de reputação desejar. Para exemplificar, a Tabela 7.3 sugere uma classificação em cinco níveis. Conforme descrito na Seção 3.5 este modelo de classificação foi escolhido devido ao fato de a comparação entre números ser mais intuitiva que entre nomes de categorias.

Os níveis da Tabela 7.3 foram definidos com base nos valores obtidos do cálculo de reputação na comunidade acadêmica do Departamento de Ciência da Computação da UnB. Conforme análise dos dados, os níveis 1 e 2 indicam pesquisadores em início de

Tabela 7.1: Categorias, elementos e pesos.

Categoria	Peso	Elemento	Peso
1. Identificação (IDE)	14	Pós-doutorado (PD) Doutorado (DOU) Mestrado (MES) Graduação (GRA)	14 12 8 4
2. Orientação (ORI)	12	Pós-doutorado (PD) Doutorado (DOU) Mestrado (MES) Iniciação Científica (INI)	5 4 2 1
3. Banca (BAN)	7	Doutorado (DOU) Mestrado (MES)	4 3
4. Membro (MEM)	7	Corpo Editorial (CE) Revisor de Periódico (RP) Projeto de pesquisa (PP)	3 3 1
5. Bolsista (BOL)	14	Bolsista CNPq (BOL)	14
6. Prêmios e Títulos (PT)	3	Prêmios e Títulos (PT)	3
7. Produção (PROD)	40	Artigos publicados com JCR (ARPJ) Artigos publicados sem JCR (ARPS) Livros publicados ou organizados (LIV) Capítulos de livros (CAP) Trabalhos em eventos (EVE) Texto em jornal ou revista (TEX) Demais tipos de produção bibliográfica (DEM) Técnica (TEC) Artística (ART)	10 5 8 5 3 2 1 3 3
8. Co-autoria (COA)	3	Pesquisador (PES)	3

Tabela 7.2: Categorias, elementos e pesos revisitadas.

Categoria	Peso	Elemento	Peso
1. Identificação (IDE)	14	Pós-doutorado (PD)	14
		Doutorado (DOU)	12
		Mestrado (MES)	8
		Graduação (GRA)	4
2. Orientação (ORI)	12	Pós-doutorado (PD)	5
		Doutorado (DOU)	4
		Mestrado (MES)	2
		Iniciação Científica (INI)	1
3. Banca (BAN)	7	Doutorado (DOU)	4
		Mestrado (MES)	3
4. Membro (MEM)	7	Corpo Editorial (CE)	3
		Revisor de Periódico (RP)	3
		Projeto de pesquisa (PP)	1
5. Bolsista (BOL)	14	Bolsista CNPq (BOL)	14
6. Prêmios e Títulos (PT)	3	Prêmios e Títulos (PT)	3
7. Produção (PROD)	40	Artigos publicados com JCR (ARPJ)	13
		Artigos publicados sem JCR (ARPS)	7
		Livros publicados ou organizados (LIV)	5
		Capítulos de livros (CAP)	3
		Trabalhos em eventos (EVE)	7
		Texto em jornal ou revista (TEX)	1
		Demais tipos de produção bibliográfica (DEM)	1
		Técnica (TEC)	3
8. Co-autoria (COA)	3	Pesquisador (PES)	3

Tabela 7.3: Classificação da reputação.

Nível	Reputação	Peso Relacionamento
1	≥ 0 e < 15	0,25
2	≥ 15 e < 30	0,5
3	≥ 30 e < 45	1
4	≥ 45 e < 60	1,5
5	≥ 60	2

carreira (menos de 10 anos) ou com pouca produção científica, o nível 3 representa os pesquisadores que possuem a produção científica e tempo de carreira próximos da média de produção e de tempo do Departamento em que se encontram. Enquanto que os níveis 4 e 5 indicam pesquisadores com considerável (mais de 15 anos) tempo de carreira acadêmica e com alta produção científica.

Exemplo

Para facilitar o entendimento do cálculo de reputação, é apresentado um exemplo da métrica aplicada. O exemplo da Tabela 7.4 contém o número de ocorrências de cada elemento da Tabela 6.1 para o pesquisador avaliado, que no caso está vinculado ao Departamento de Ciência da Computação da UnB.

Após a coleta dos elementos à partir do repositório do ObserveUnB, a reputação é calculada conforme a métrica definida em 6.1.2 e classificada segundo a Tabela 7.3.

1. $Valor_{IDE} = 1*14 = 14$
2. $Valor_{ORI} = 1*5 + 4*4 + 18*2 + 33*1 = 90$
3. $Valor_{BAN} = 5*4 + 8*3 = 44$
4. $Valor_{MEM} = 1*3 + 11*3 + 10*1 = 46$
5. $Valor_{BOL} = 1*14 = 14$
6. $Valor_{PT} = 12*3 = 36$
7. $Valor_{PROD} = 12*13 + 14*7 + 6*5 + 2*3 + 62*7 + 0*1 + 7*1 + 28*3 = 815$
8. $Valor_{COA} = (2 + 0,25 + 1 + 0,25 + 0,5 + 0,25) *3 = 12,75$

Conforme explicitado na Seção 6.1.2, a pontuação da categoria Co-autoria é definida pelo somatório dos pesos de cada relacionamento do pesquisador avaliado multiplicado pelo peso da categoria, onde os pesos dos relacionamentos são obtidos de acordo com a classificação da reputação do pesquisador na Tabela 7.3.

Finalmente, com a pontuação de cada categoria, obtém-se o valor final da reputação do pesquisador, que corresponde ao somatório destes valores dividido pela quantidade de elementos distintos das categorias, onde neste caso equivale a 21:

$$\begin{aligned} Reputacao_{Pesquisador} &= 14 + 90 + 44 + 46 + 14 + 36 + 815 + 12,75 \\ Reputacao_{Pesquisador} &= 1.071,75 / 21 \\ Reputacao_{Pesquisador} &= 51 \end{aligned}$$

Dessa forma, o valor obtido para a reputação do pesquisador é 51. De acordo com a Tabela 7.3 a reputação está classificada no nível 4, o que significa que o pesquisador possui uma significativa experiência acadêmica e alta produção científica.

Resultados do Modelo

Para auxiliar na avaliação do modelo, foram gerados alguns gráficos a partir da implementação do modelo de reputação no escopo da área de Ciência da Computação. Estes

Tabela 7.4: Tabela de cálculo da reputação para um pesquisador.

Categoria	Elemento	Peso	Ocorrência
1. Identificação (IDE)	Pós-doutorado (PD)	14	1
	Doutorado (DOU)	12	-
	Mestrado (MES)	8	-
	Graduação (GRA)	4	-
2. Orientação (ORI)	Pós-doutorado (PD)	5	1
	Doutorado (DOU)	4	4
	Mestrado (MES)	2	18
	Iniciação Científica (INI)	1	33
3. Banca (BAN)	Doutorado (DOU)	4	5
	Mestrado (MES)	3	8
4. Membro (MEM)	Corpo Editorial (CE)	3	1
	Revisor de Periódico (RP)	3	11
	Projeto de pesquisa (PP)	1	10
5. Bolsista (BOL)	Bolsista CNPq (BOL)	14	1
6. Prêmios e Títulos (PT)	Prêmios e Títulos (PT)	3	12
7. Produção (PROD)	Artigos publicados com JCR (ARPJ)	13	12
	Artigos publicados sem JCR (ARPS)	7	14
	Livros publicados organizados (LIV)	5	6
	Capítulos de livros (CAP)	3	2
	Trabalhos em eventos (EVE)	7	62
	Texto em jornal ou revista (TEX)	1	0
	Demais tipos de produção bibliográfica (DEM)	1	7
	Técnica (TEC)	3	28
8. Co-autoria (COA)	Pesquisador (PES)	3	6

gráficos mostram: (i) a classificação da reputação dos docentes avaliados; (ii) a produção bibliográfica destes docentes; (iii) o tempo de carreira acadêmica dos docentes e (iv) a classificação dos pesquisadores bolsistas em produtividade do CNPq.

A classificação da reputação dos docentes do Departamento Ciência da Computação da UnB é representada no gráfico da Figura 7.10. Note que o Nível 1 tem a maior concentração de pesquisadores (29). Isso se deve em parte à incompletude e inconsistência dos dados nos CV-Lattes, bem como o fato de que muitos desses pesquisadores apresentam pouco tempo de carreira acadêmica.

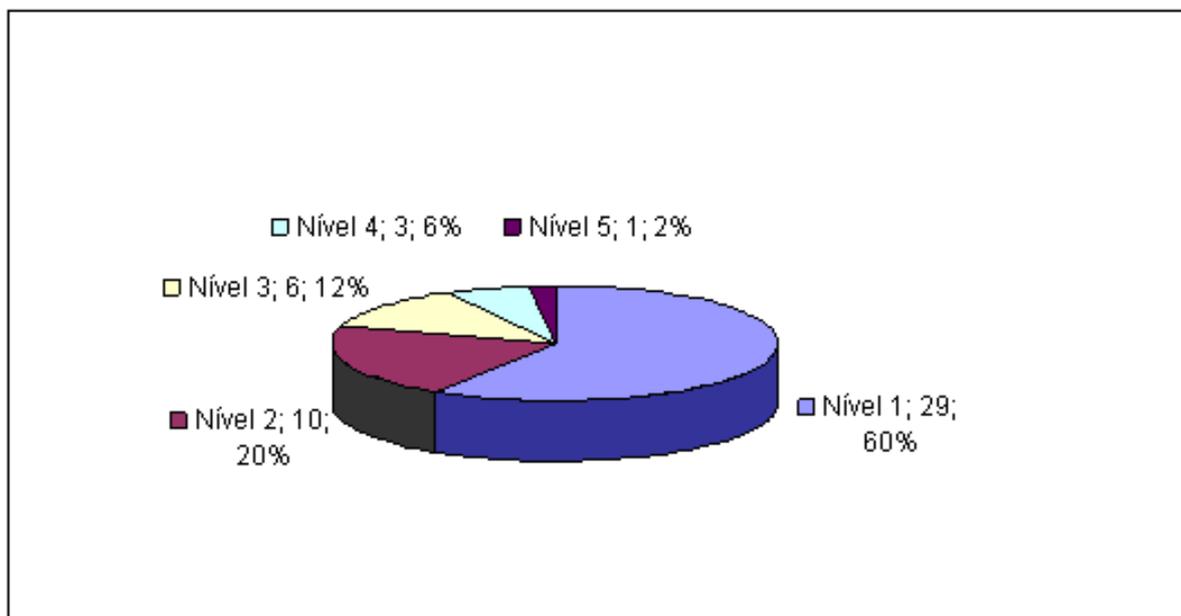


Figura 7.10: Classificação da reputação dos docentes da Ciência da Computação da UnB (nível, unidade, percentual).

A Figura 7.11 apresenta o percentual do quantitativo de produção bibliográfica referente a cada nível da Tabela 7.3. Percebe-se que apesar de o Nível 1 concentrar a maior quantidade de docentes (60% conforme a Figura 7.10), ele representa o segundo maior percentual de produção bibliográfica da Computação (24%), enquanto o Nível 2 com 10 docentes representa 30% da produção científica da unidade acadêmica.

Baseado nos gráficos das Figuras 7.10 e 7.11, pode-se obter a média de produção bibliográfica por pesquisador de cada nível, referente a sua vida acadêmica. A Figura 7.12 apresenta a média quantitativa de produção bibliográfica por nível, onde o Nível 1 detém a menor média de quantitativos de produção (21), enquanto o Nível 5 possui a maior quantidade de produção (158).

Para avaliar a influência do tempo de carreira no cálculo da reputação, foi gerado o gráfico de tempo médio de carreira acadêmica, dentro e fora da UnB, dos docentes referentes a cada nível de classificação. A partir da Figura 7.13 percebe-se que o tempo médio de carreira acadêmica cresce a cada nível de classificação. Isso reforça a tese de que quanto menor o nível de classificação menor é o tempo de carreira do pesquisador. No caso da amostra de pesquisadores avaliados, o Nível 5 representa uma exceção, visto que

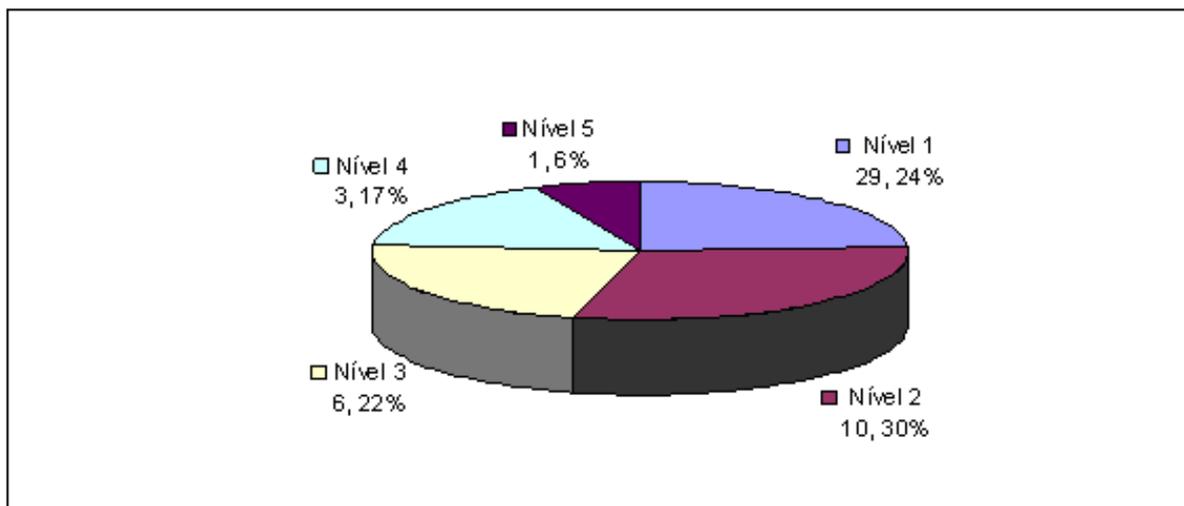


Figura 7.11: Produção bibliográfica referente aos níveis de reputação (nível, unidade, percentual).

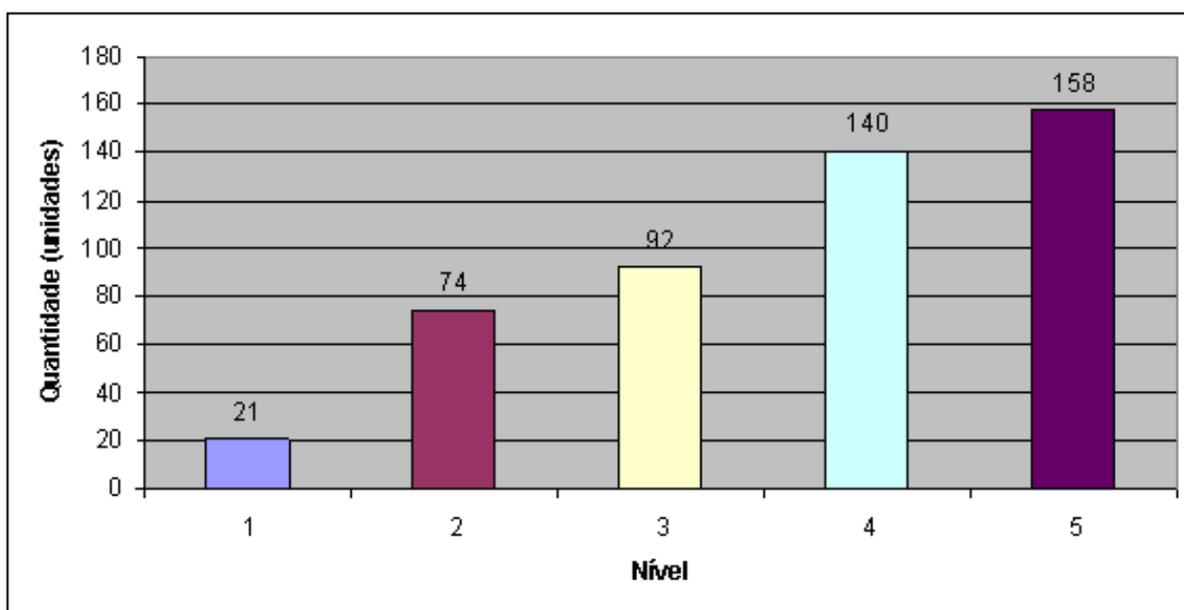


Figura 7.12: Média de produção bibliográfica dos docentes classificados por nível (nível, unidade, percentual).

representa um pesquisador com tempo de carreira (12 anos) próximo ao dos pesquisadores de Nível 1 (10 anos), apesar de já estar como titular na carreira.

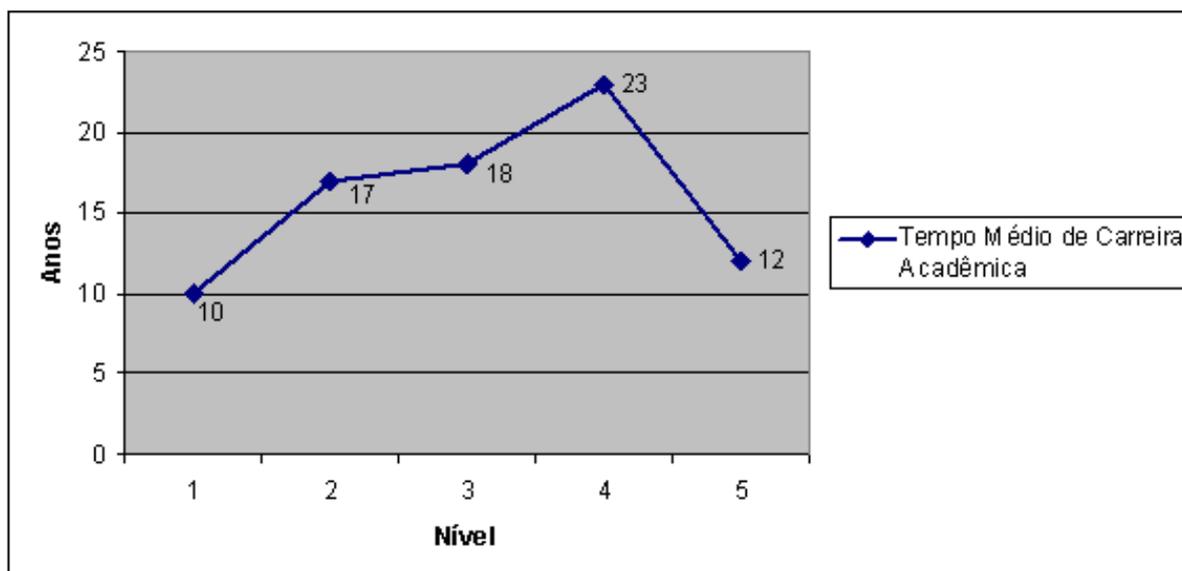


Figura 7.13: Tempo médio de carreira dos docentes classificados por nível.

Os gráficos 7.14 e 7.15 representam os bolsistas de produtividade em pesquisa (PQ) do CNPq (CNPq, 2012a). A Figura 7.14 representa os oito docentes bolsistas PQ vinculados ao Departamento de Ciência da Computação da UnB. Note que o Nível 3 concentra a maior parte dos bolsistas (4), enquanto que o Nível 2 contém 1 pesquisador, devido ao pouco tempo de carreira deste pesquisador.

O gráfico da Figura 7.15 representa os 16 pesquisadores bolsistas de produtividade e pesquisa do CNPq nível PQ1A, o maior nível de bolsa de produtividade do país (CNPq, 2012a). A partir do gráfico, verifica-se que o Nível 5 concentra a maior parte dos bolsistas (82%), enquanto o Nível 4 detém dois pesquisadores e o Nível 3 concentra um pesquisador.

Validação do Modelo

A fim de conhecer a opinião dos pesquisadores sobre o modelo de reputação, realizou-se uma entrevista semi-estruturada com alguns pesquisadores do corpo docente do Departamento de Ciência da Computação da UnB. Inicialmente, a entrevista estava sendo realizada de forma presencial, com duração estimada de 15 minutos. No entanto, devido à escassez de tempo decorrente dos cargos ocupados pelos docentes optou-se por enviar um questionário *on-line*.

O questionário enviado encontra-se no Apêndice A. Ele também foi usado como tópicoguia para as entrevistas presenciais, permitindo ao entrevistado falar livremente sobre o modelo avaliado.

Foram enviados emails contendo o formulário do questionário e a explicação do modelo a todos os docentes do Departamento de Ciência da Computação da UnB. Destes, quatro docentes responderam o formulário *on-line* e dois docentes registraram sua opinião por email, sem preencher o questionário.

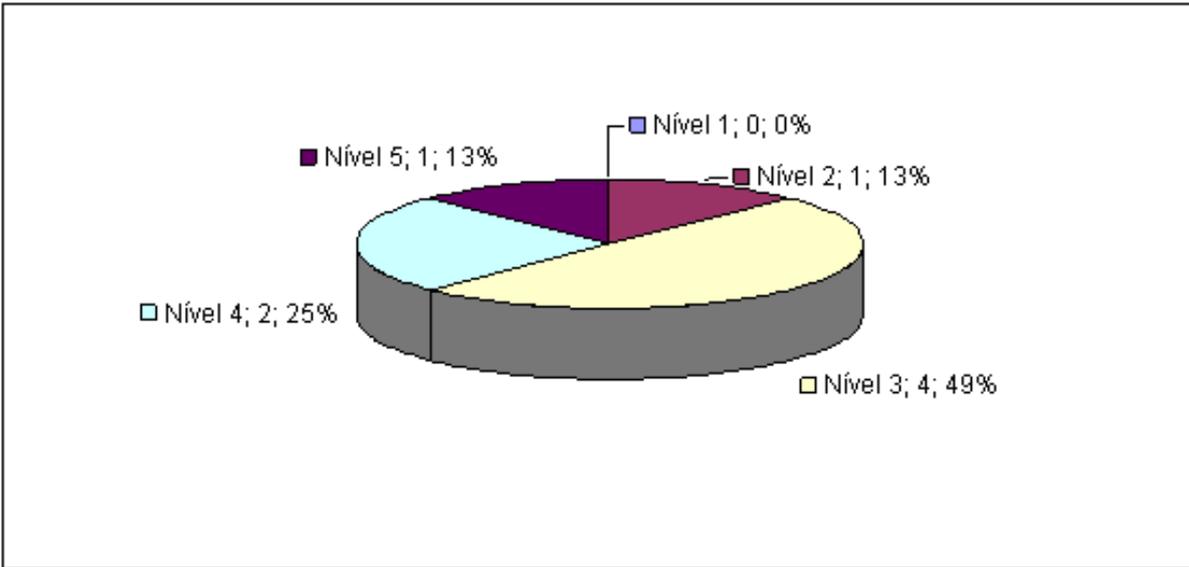


Figura 7.14: Classificação da reputação dos docentes bolsistas PQ do Dep. Ciência da Computação da UnB (nível, unidade, percentual).

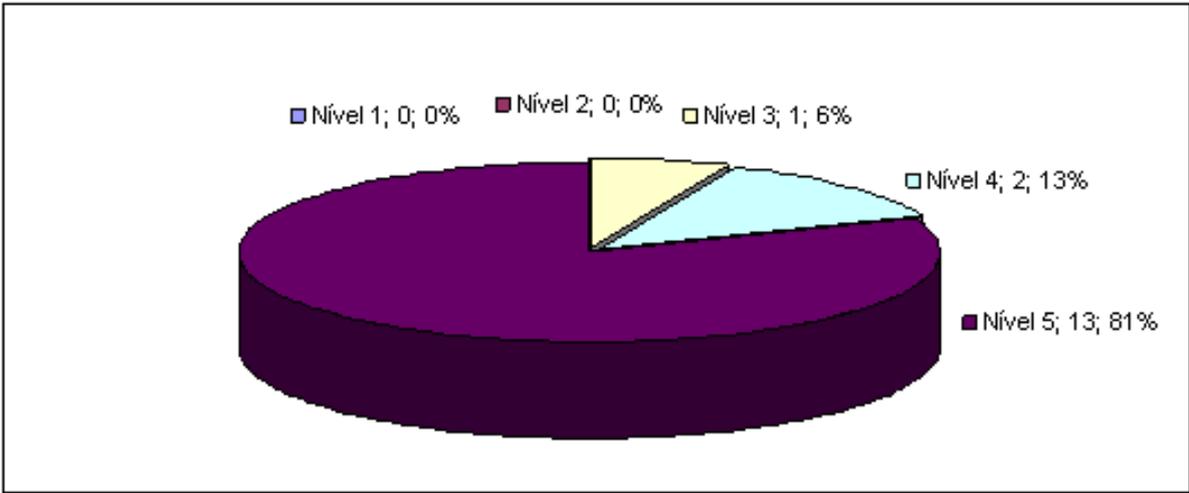


Figura 7.15: Classificação da reputação dos docentes bolsistas PQ1A da área Ciência da Computação do país (nível, unidade, percentual).

Quanto às entrevistas presenciais, tentou-se realizá-las com seis docentes da Ciência da Computação da UnB. Desses, três concederam a entrevista respondendo o questionário e dois apenas opinaram sobre o modelo proposto. A Tabela 7.5 resume o quantitativo de entrevistas.

Tabela 7.5: Tabela do quantitativo de entrevistas.

Meio utilizado	Quantitativo
Entrevista por <i>e-mail</i>	2
Entrevista por formulário <i>on-line</i>	4
Entrevista presencial com formulário	3
Entrevista presencial sem formulário	2
Total	11

A Tabela 7.6 apresenta informações sobre os entrevistados, como formação acadêmica, faixa etária, tempo de carreira na UnB e bolsista de produtividade no CNPq. As informações foram obtidas das entrevistas e do currículo *Lattes*, onde os nomes foram omitidos para preservar a identidade dos pesquisadores. O resultado das entrevistas é apresentado ao final desta Seção e a análise desses resultados é apresentada na Seção 7.3.

Tabela 7.6: Resumo do perfil do entrevistado.

Pesquisador	Faixa Etária	Formação	Tempo de Docência na UnB (em anos)	Bolsista CNPq
P1	50 a 60	Doutorado	10	Não
P2	41 a 50	Pós-doutorado	15	Sim
P3	51 a 60	Pós-doutorado	15	Sim
P4	41 a 50	Doutorado	7	Não
P5	51 a 60	Pós-doutorado	24	Sim
P6	41 a 50	Doutorado	4	Não
P7	31 a 40	Doutorado	3	Não
P8	31 a 40	Doutorado	1	Não
P9	41 a 50	Doutorado	6	Não
P10	41 a 50	Doutorado	6	Sim
P11	31 a 40	Doutorado	3	Não
Média	45	Média	9	Não

Com base na análise dos gráficos e no modelo de reputação definido para a área de Ciência da Computação (Tabela 7.2), os 11 docentes entrevistados registraram suas opiniões. Com relação a questão 2 do formulário do Apêndice A que pergunta se o pesquisador concorda com a abordagem do modelo em focar na produção científica, os pesquisadores responderam:

- P1, P2, P3, P4, P5, P6 e P7 concordam com a abordagem;
- P8 concorda em parte visto que, segundo o pesquisador, a reputação representa mais que titulações e produções científicas, onde cargos administrativos podem colaborar muito para uma boa reputação, em contrapartida podem dificultar a pesquisa;

- P11 concorda em parte, de acordo com o pesquisador há casos em que pesquisadores, cuja produção não é tão expressiva possuem índice de reputação mais alto que pesquisadores com publicações dentro e fora da instituição.

Em relação a questão 3 do formulário que pergunta se o pesquisador concorda com os elementos e categorias definidos no modelo, os pesquisadores responderam:

- P3 e P4 concordam;
- P1, P2, P5, P6, P7, P9, P10 e P11 concordam em parte;
- P1, P5 e P11 sugeriram a inclusão de outros índices internacionais ou nacionais de classificação de publicação (e.g. Scopus, H-index, Qualis). Enquanto o pesquisador P2 sugeriu a distinção entre as publicações nacionais e internacionais;
- P6 sugeriu incluir o elemento Graduação na categoria 3 (Banca) e o elemento Revisor de congresso científico na categoria 4 (Membro);
- P7 questionou a presença do elemento Bolsista CNPq, sugerindo que fosse considerado como Prêmios e Títulos;
- P8 discordou, para o pesquisador não há relação entre a produção científica e os elementos definidos, devendo ser considerado apenas a produção científica.

Considerando a questão 4 sobre a ordem de relevância (pesos) definida para os elementos, os pesquisadores responderam:

- P5 e P6 concordam;
- P1, P2, P3, P4, P7, P8, P9, P10 e P11 concordam em parte;
- P1 considerou grande a diferença de pesos entre as publicações com e sem JCR;
- P2 e P4 sugeriram aumentar a diferença entre os pesos destes elementos;
- P7 discorda da relevância atribuída a categoria Co-autoria, pois para o pesquisador, um número grande de co-autorias representa uma baixa contribuição do pesquisador avaliado, contribuindo negativamente para a reputação;
- P8 sugeriu que os pesos dos elementos da categoria Produção variassem conforme a sub-área ou especialidade, e.g. na área de redes, uma publicação em um evento específico possui maior peso que a publicação de um artigo com JCR;
- P11 sugeriu que Capítulos de livros tivessem maior peso, enquanto que artigos publicados em periódicos sem JCR deveriam ser diferenciados de artigos publicados em conferências. Sugeriu também que o elemento Doutorado da categoria Orientação tivesse maior peso;
- Alguns docentes criticaram a presença e a relevância do elemento Pós-doutorado na categoria Identificação, sugerindo que tal elemento fosse considerado como um estágio, pois segundo os entrevistados, este elemento não é reconhecido internacionalmente como uma titulação.

Com relação a questão 5 que questiona se o Gráfico 7.10 reflete a realidade acadêmica da Ciência da Computação da UnB, os pesquisadores responderam:

- P1, P4, P5, P6 e P7 disseram que sim;
- P2 e P11 responderam que o modelo reflete em parte a realidade acadêmica;
- P2 comenta que existem muitos pesquisadores recém-doutores que ingressaram recentemente na instituição, contribuindo para a baixa reputação. Para P2, seria interessante considerar o tempo de carreira do docente no cômputo da reputação;
- P11 sugeriu que o pesquisador avaliado pudesse validar cada item considerado no cálculo de sua reputação, podendo inclusive influenciar no resultado final.

Por último, alguns pesquisadores comentaram sobre o modelo:

- P2, P3, P5 e P9 acharam interessante o modelo de reputação proposto, onde P2 ressaltou a relevância da categoria co-publicação;
- P1 e P2 sugeriram que o modelo fosse temporal, onde apenas os elementos em um período de tempo fossem considerados (e.g. nos últimos 5 anos ou período trienal da avaliação da CAPES).

A Tabela 7.7 apresenta um percentual das respostas objetivas dos pesquisadores entrevistados. Analisando a tabela percebe-se que a maior parte dos pesquisadores entrevistados (64%) concordam com a abordagem do modelo de reputação proposto. Enquanto que 73% dos entrevistados concordam parcialmente com os elementos e categorias definidos no modelo, e 82% concordam em parte com os pesos atribuídos aos elementos. Por último, a maioria dos entrevistados (46%) concorda com os resultados obtidos da aplicação do modelo, onde tais resultados refletem à realidade acadêmica do Departamento de Ciência da Computação.

Tabela 7.7: Resumo das respostas dos entrevistados.

Questão	Concordam	Não Concordam	Concordam em Parte	Não Opinaram
2	64%	0%	18%	18%
3	18%	9%	73%	0%
4	18%	0%	82%	0%
5	46%	0%	18%	36%

Comparação IPC Versus Base DBLP

Além da validação do modelo de reputação por meio das entrevistas, foi realizada uma comparação dos índices de reputação obtidos com o índice DBLP dos 49 pesquisadores do Departamento de Ciência da Computação. Os gráficos 7.16, 7.17 e 7.18 apresentam os resultados dos índices IPC e DBLP para os pesquisadores identificados de P1 a P49. Os resultados desta validação são discutidos na Seção 7.3.

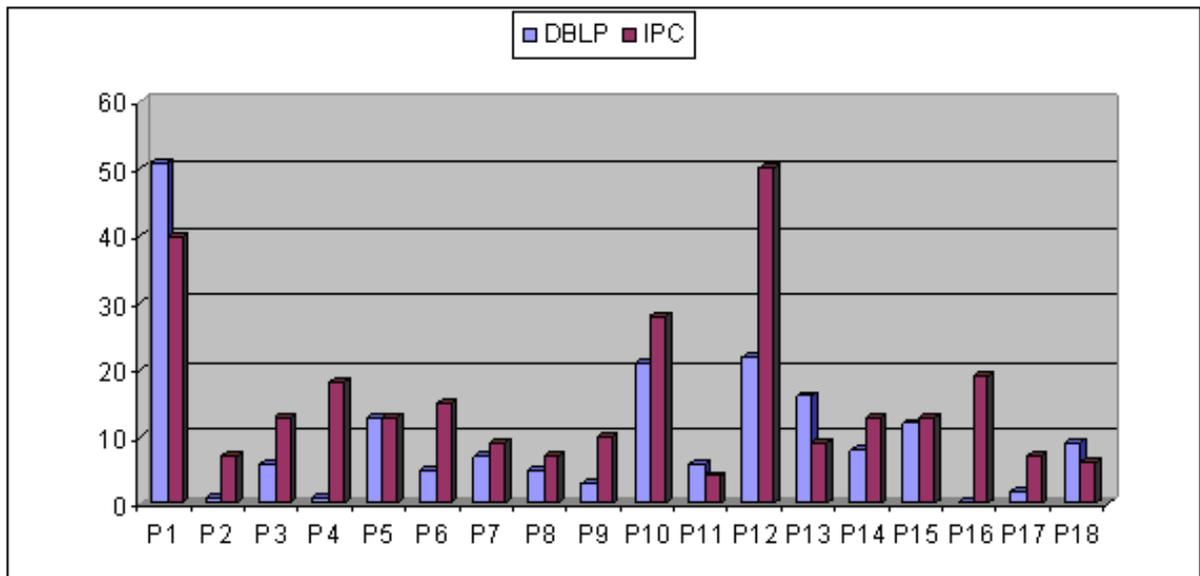


Figura 7.16: Dados dos pesquisadores P1 a P18

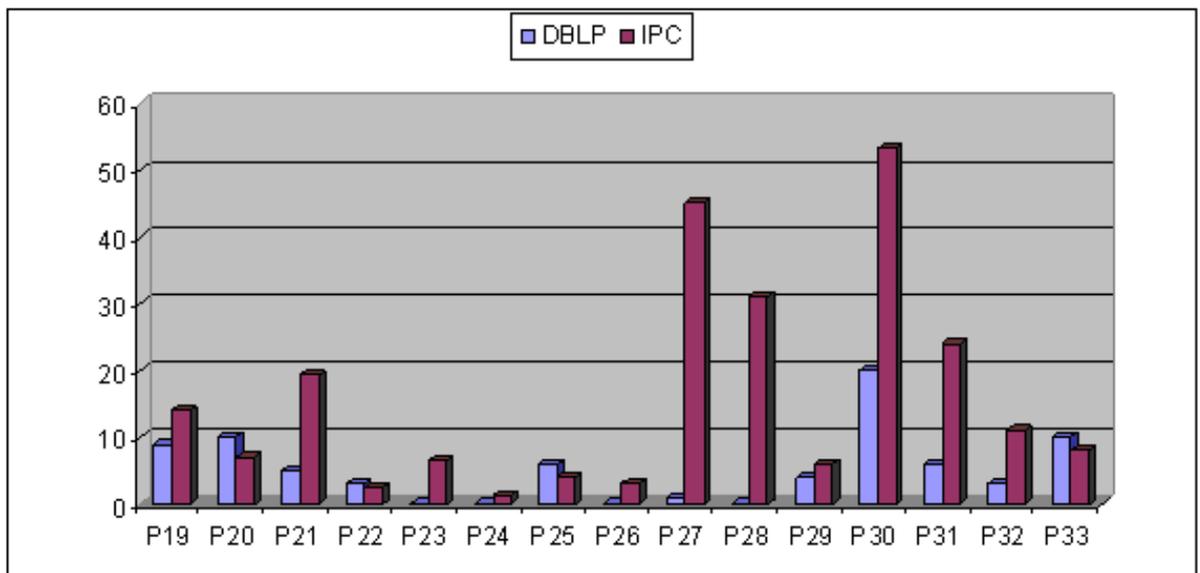


Figura 7.17: Dados dos pesquisadores P19 a P33

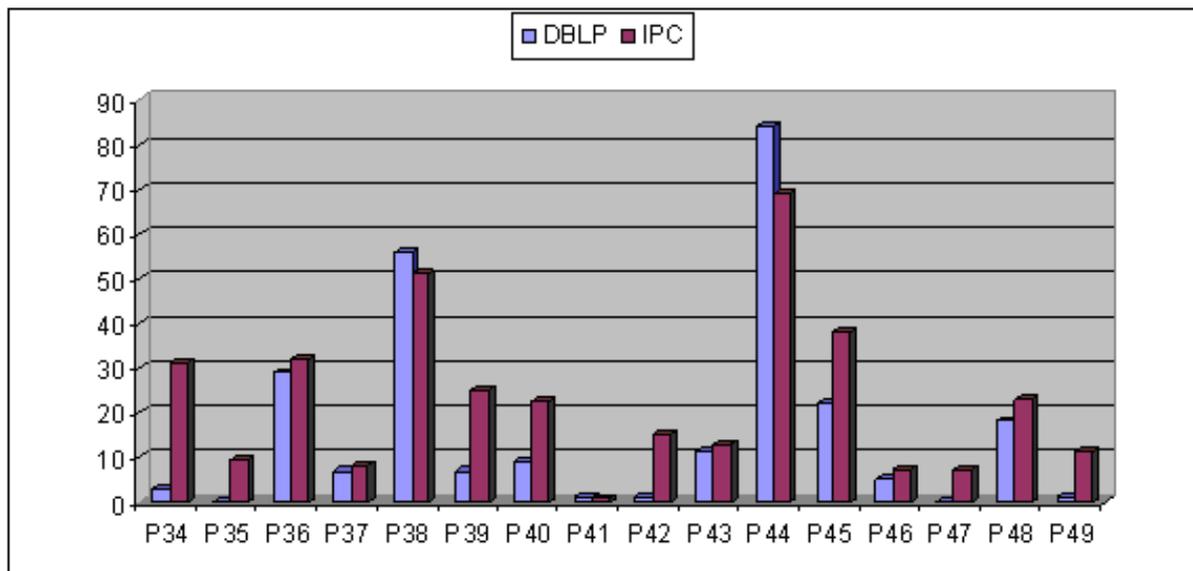


Figura 7.18: Dados dos pesquisadores P34 a P49

7.2.2 Ontologia

A avaliação da ontologia foi realizada a partir das questões de competência que delimitaram seu escopo, descritas na Seção 6.2.2. Tais questões envolvem os conceitos definidos na Seção 6.2.2 representados na ontologia através da hierarquia de classes e propriedades. Este procedimento de avaliação englobou as seguintes atividades:

- Formulação das questões de competência em linguagem formal - as questões de competência até então descritas apenas em linguagem natural (ver Seção 6.2.2), foram descritas em OWL 2. Os elementos usados na formulação destas questões, tais como classes e propriedades, foram mantidos ou incorporados à ontologia estendida;
- Produção das respostas às questões de competência por meio de mecanismo de inferência - a execução desta atividade envolveu a ontologia, os elementos usados na formulação de competência e o motor de inferência Pellet. As respostas as questões foram encontradas por meio da análise do relacionamento existente entre as classes através de suas propriedades. Assim, as classes que representam tais questões foram associadas, via inferência, a outras classes da ontologia, que representam a resposta à questão. Para tanto, foi utilizada a linguagem de consulta do Protégé *DL query*, por estar disponível na ferramenta e por ser considerada mais intuitiva. Além do *DL-Query*, foi utilizada a linguagem SPARQL no Jena, tendo em vista a utilização deste *framework* no módulo de integração da ontologia com o ObserveUnB;
- Análise do resultado do processo de inferência - a análise do resultado consistiu em avaliar e discutir a consistência e expressividade oferecidas pela ontologia, tanto para a descrição das questões de competência em seus termos, como na obtenção das respostas encontradas.

É apresentada para cada uma das questões levantadas na Seção 6.2.2, a descrição das atividades realizadas no procedimento de avaliação da ontologia. Para isso, são exibidos os

resultados de cada consulta realizada na linguagem *DL-Query* no Protégé e na linguagem SPARQL da API Jena, cujas sintaxes constam no Apêndice B.

Questão 1 : Quem são os pesquisadores de determinada área de atuação?

A Tabela 7.8 apresenta as classes e propriedades utilizadas para responder a questão, descrevendo o relacionamento existente entre os conceitos na ontologia.

Tabela 7.8: Classes e propriedades OWL para responder à Questão 1.

Classe	Propriedade	Descrição
Autores	temCurriculo	As instâncias desta classe têm no máximo uma associação através da propriedade temCurriculo com instâncias da classe CurriculoVitae.
CurriculoVitae	temDadosGerais	As instâncias desta classe podem ter somente associação através da propriedade temDadosGerais com instâncias da classe DadosGerais.
DadosGerais	temAreadeAtuacao	As instâncias desta classe têm pelo menos uma associação através da propriedade temAreadeAtuacao com instâncias da classe AreasDeAtuacao.
AreasDeAtuacao	especialidade	As instâncias desta classe têm somente associação através da propriedade especialidade com instâncias das classes Area, SubArea ou Especialidade.

Considere a consulta em linguagem natural “Quem são os pesquisadores que atuam na área Ciência da Computação?”, escrita na linguagem SPARQL e na linguagem *DL query*, cujas sintaxes são exibidas no Apêndice B. A Figura 7.19 mostra uma tela do Protégé contendo o resultado da execução da consulta. Note que 51 autores da área Ciência da Computação foram apresentados, sendo que deste total, 49 autores estão vinculados ao Departamento de Ciência da Computação.

Questão 2 : Quem são os pesquisadores de determinada área de atuação que possuem produção científica?

A Tabela 7.9 apresenta as classes e propriedades utilizadas para responder a questão, descrevendo o relacionamento existente entre os conceitos na ontologia.

Considere a consulta em linguagem natural “Quem são os pesquisadores da área Ciência da Computação que possuem produção científica?”, escrita na linguagem SPARQL e na linguagem *DL query*, cujas sintaxes são exibidas no Apêndice B.

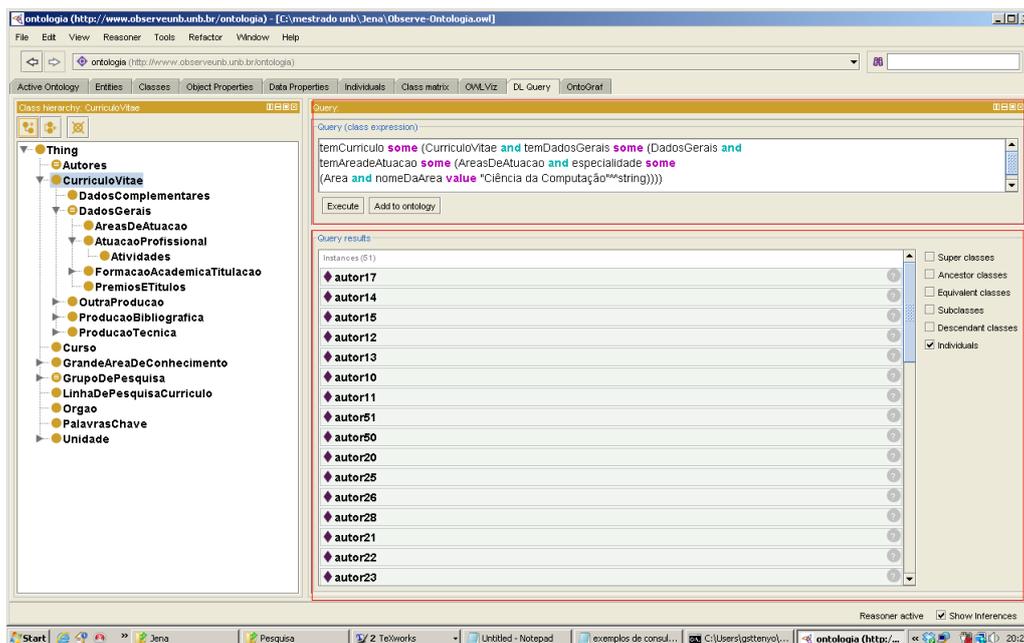


Figura 7.19: Exemplo de consulta realizada no Protégé para a Questão 1.

A Figura 7.20 mostra uma tela do Protégé contendo o resultado da execução da consulta. Note que do total de pesquisadores listados na primeira consulta (51 autores), 9 deles não possuem produção bibliográfica (artigos em periódicos).

Questão 3 : Quem são os pesquisadores vinculados a determinada unidade acadêmica?

A Tabela 7.10 apresenta as classes e propriedades utilizadas para responder a questão, descrevendo o relacionamento existente entre os conceitos na ontologia.

Considere a consulta em linguagem natural “Quem são os pesquisadores vinculados ao Departamento de Ciência da Computação da UnB?”, escrita na linguagem SPARQL e na linguagem DL *query*, cujas sintaxes são exibidas no Apêndice B. A Figura 7.21 mostra uma tela do Protégé contendo o resultado da execução da consulta. Note a semelhança com a Figura 7.19, visto que a consulta 1 abrange todos os resultados da consulta 3.

Questão 4 : Quem são os pesquisadores de determinado grupo de pesquisa?

A Tabela 7.11 apresenta as classes e propriedades utilizadas para responder a questão, descrevendo o relacionamento existente entre os conceitos na ontologia.

Considere a consulta em linguagem natural “Quem são os pesquisadores vinculados ao grupo de pesquisa Teoria da Computação?”, escrita na linguagem SPARQL e na linguagem DL *query*, cujas sintaxes são exibidas no Apêndice B. A Figura 7.22 mostra uma tela do Protégé contendo o resultado da execução da consulta, onde são apresentados os autores que integram o grupo de pesquisa Teoria da Computação.

Questão 5 : Quem são os co-autores de determinada publicação?

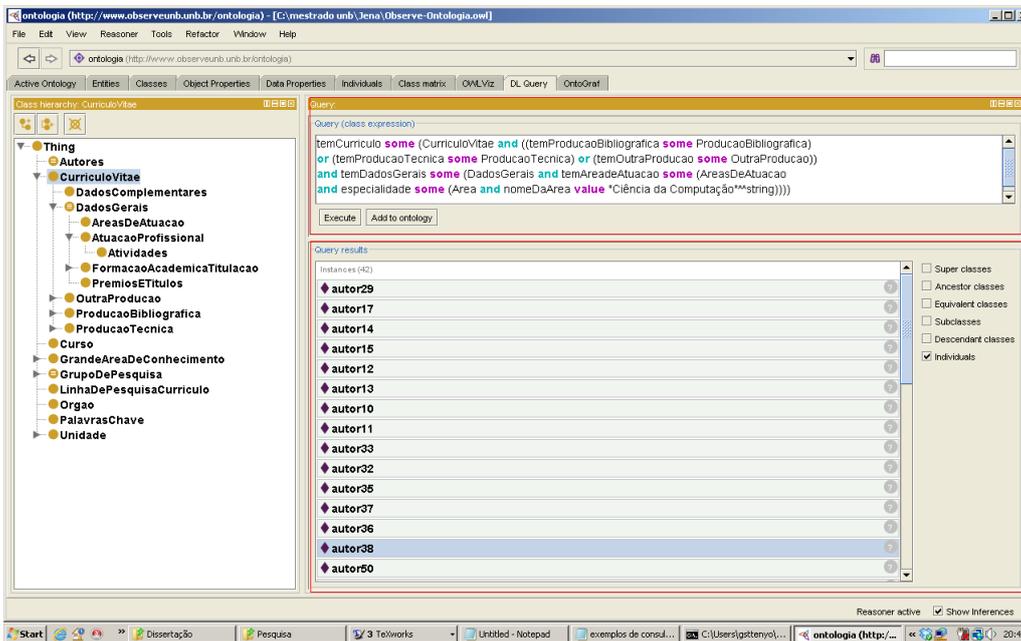


Figura 7.20: Exemplo de consulta realizada no Protégé para a questão 2.

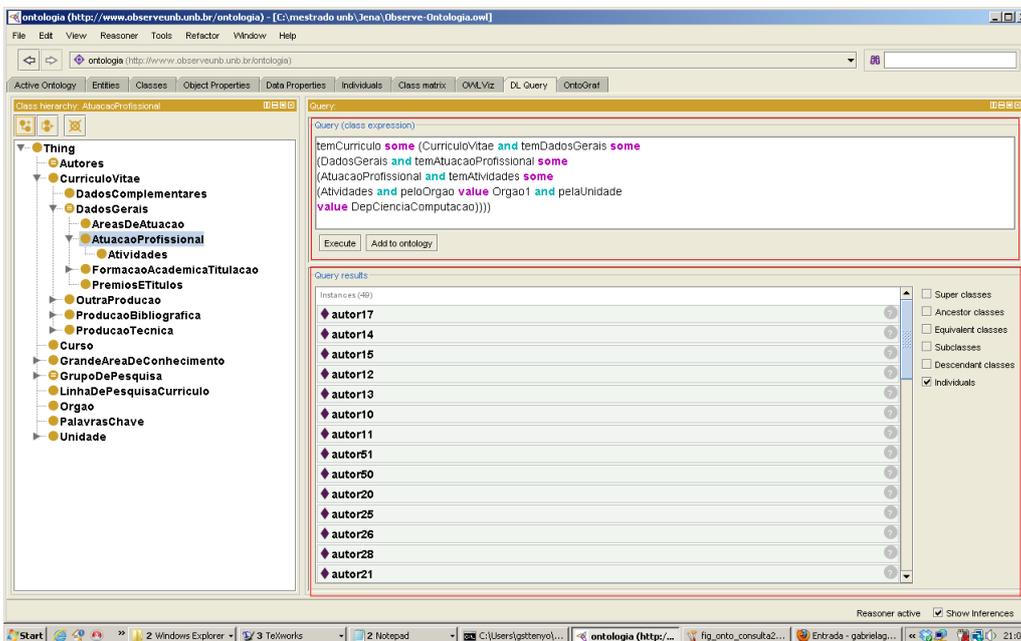


Figura 7.21: Exemplo de consulta realizada no Protégé para a Questão 3.

Tabela 7.9: Classes e propriedades OWL para responder à Questão 2.

Classe	Propriedade	Descrição
Autores	temCurriculo	As instâncias desta classe têm no máximo uma associação através da propriedade temCurriculo com instâncias da classe CurriculoVitae.
CurriculoVitae	temDadosGerais temOutraProducao temProducaoBibliografica temProducaoTecnica	As instâncias desta classe podem ter somente associação através das propriedades temDadosGerais, temOutraProducao, temProducaoBibliografica e temProducaoTecnica com instâncias das classes DadosGerais, OutraProducao, ProducaoBibliografica e ProducaoTecnica respectivamente.
DadosGerais	temAreadeAtuacao	As instâncias desta classe têm pelo menos uma associação através da propriedade temAreadeAtuacao com instâncias da classe AreasDeAtuacao.
AreasDeAtuacao	especialidade	As instâncias desta classe têm somente associação através da propriedade especialidade com instâncias das classes Area, SubArea ou Especialidade.

A Tabela 7.12 apresenta as classes e propriedades utilizadas para responder a questão, descrevendo o relacionamento existente entre os conceitos na ontologia.

Considere a consulta em linguagem natural “Quem são os co-autores da publicação bibliográfica *ObserveUnB - Um Portal Web de Rede Social Científica?*”, escrita na linguagem SPARQL e na linguagem DL *query*, cujas sintaxes são exibidas no Apêndice B. A Figura 7.23 mostra uma tela do Protégé contendo o resultado da execução da consulta. Neste caso, são apresentados três autores os quais são co-autores da publicação.

Questão 6 : Quem são os pesquisadores de determinada área de atuação que possuem produção científica com JCR?

A Tabela 7.13 apresenta as classes e propriedades utilizadas para responder a questão, descrevendo o relacionamento existente entre os conceitos na ontologia.

Tabela 7.10: Classes e propriedades OWL para responder à Questão 3.

Classe	Propriedade	Descrição
Autores	temCurriculo	As instâncias desta classe têm no máximo uma associação através da propriedade temCurriculo com instâncias da classe CurriculoVitae.
CurriculoVitae	temDadosGerais	As instâncias desta classe têm associação através da propriedade temDadosGerais somente com instâncias da classe DadosGerais.
DadosGerais	temAtuacaoProfissional	As instâncias desta classe têm associação através da propriedade temAtuacaoProfissional somente com instâncias da classe AtuacaoProfissional.
AtuacaoProfissional	temAtividades	As instâncias desta classe têm associação através da propriedade temAtividades somente com instâncias da classe Atividades.
Atividades	peloOrgao pelaUnidade	As instâncias desta classe têm associação através da propriedade peloOrgao somente com instâncias da classe Órgão e através da propriedade pelaUnidade somente com instâncias da classe Unidade.

Tabela 7.11: Classes e propriedades OWL para responder à Questão 4.

Classe	Propriedade	Descrição
Autores	pertenceGrupoPesquisa	As instâncias desta classe têm associação através da propriedade pertenceGrupoPesquisa somente com instâncias da classe GrupoDePesquisa.

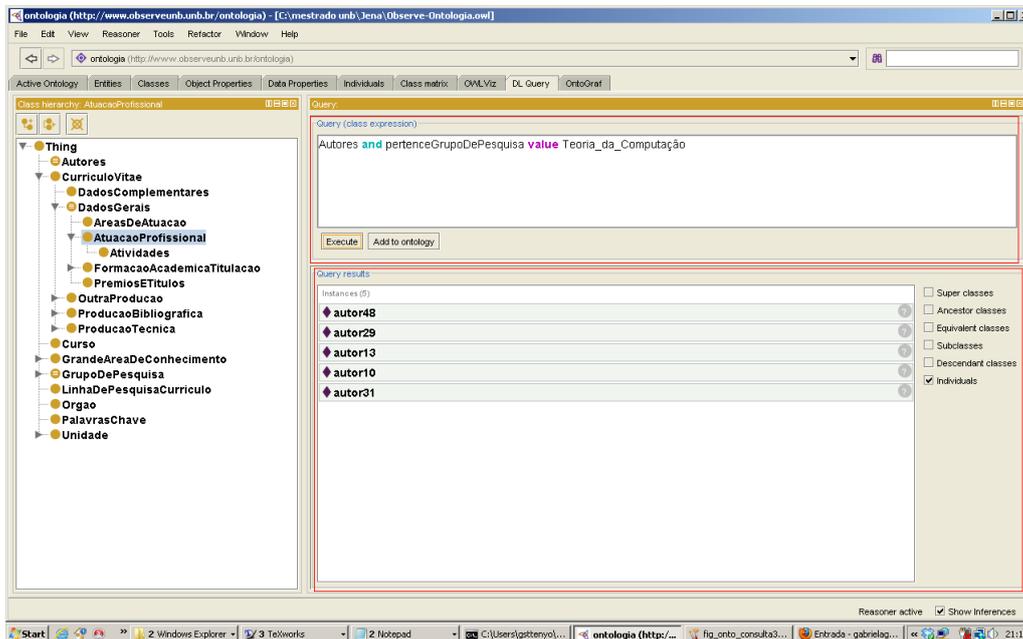


Figura 7.22: Exemplo de consulta realizada no Protégé para a Questão 4.

Tabela 7.12: Classes e propriedades OWL para responder à Questão 5.

Classe	Propriedade	Descrição
Autores	temCurriculo	As instâncias desta classe têm no máximo uma associação através da propriedade temCurriculo com instâncias da classe CurrículoVitae.
CurrículoVitae	temOutraProducao temProducaoBibliografica temProducaoTecnica	As instâncias desta classe podem ter somente associação através das propriedades temOutraProducao, temProducaoBibliografica e temProducaoTecnica com instâncias das classes OutraProducao, ProducaoBibliografica e ProducaoTecnica respectivamente.

Tabela 7.13: Classes e propriedades OWL para responder à Questão 6.

Classe	Propriedade	Descrição
Autores	temCurriculo	As instâncias desta classe têm no máximo uma associação através da propriedade temCurriculo com instâncias da classe CurriculoVitae.
CurriculoVitae	temDadosGerais temProducaoBibliografica	As instâncias desta classe podem ter associação através das propriedades temDadosGerais e temProducaoBibliografica com instâncias das classes DadosGerais e ProducaoBibliografica.
ProducaoBibliografica	temArtigo temDetalhamento	As instâncias desta classe podem ter associação através das propriedades com instâncias das classes Artigos e Detalhamento.
Detalhamento	temDetalhamentoDoArtigo	As instâncias desta classe tem associação com instâncias da classe DetalhamentoDoArtigo através da propriedade citada.
DadosGerais	temAreadeAtuacao	As instâncias desta classe têm pelo menos uma associação através da propriedade temAreadeAtuacao com instâncias da classe AreasDeAtuacao.
AreasDeAtuacao	especialidade	As instâncias desta classe têm somente associação através da propriedade especialidade com instâncias das classes Area, SubArea ou Especialidade.

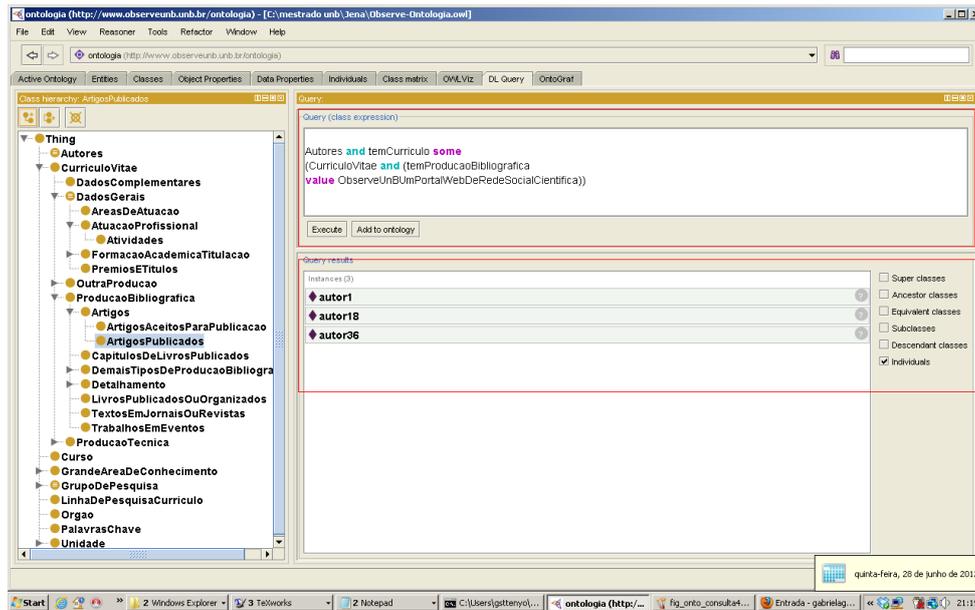


Figura 7.23: Exemplo de consulta realizada no Protégé para a Questão 5.

Considere a consulta em linguagem natural “Quem são os pesquisadores da área Ciência da Computação que possuem produção científica com JCR?”, escrita na linguagem SPARQL e na linguagem DL *query*, cujas sintaxes são exibidas no Apêndice B. A Figura 7.24 mostra uma tela do Protégé contendo o resultado da execução da consulta. Neste caso são apresentados 29 autores com publicação com JCR, visto que foi instanciado um artigo para cada autor.

Questão 7 : Quais são os termos relacionados à uma área de conhecimento, sub-área ou especialidade?

A Tabela 7.14 apresenta as classes e propriedades utilizadas para responder a questão, descrevendo o relacionamento existente entre os conceitos na ontologia.

Considere a consulta em linguagem natural “Quais são os termos relacionados a área Ciência da Computação?”, esta consulta pode ser escrita em *DL-Query*, conforme é mostrado nas sintaxes exibidas no Apêndice B. Percebe-se que a consulta 7 é mais genérica que as demais, tornando-a mais flexível. Isso ocorre em razão da vasta utilização das propriedades e relacionamentos definidos na ontologia (conforme descrito na Tabela 7.14). No caso desta consulta, foi implementado uma aplicação Java utilizando a API Jena. A Figura 7.25 mostra a tela da aplicação com o resultado da consulta 7, onde são apresentados todos os autores, os grupos de pesquisa e as produções bibliográficas relacionados a área Ciência da Computação.

Dessa forma, pode-se criar diferentes consultas utilizando diversas combinações de classes e propriedades da ontologia. Para isso, basta que os termos de busca das consultas estejam definidos na ontologia, bem como a ontologia tenha sido validada por algum raciocinador ou *reasoner*.

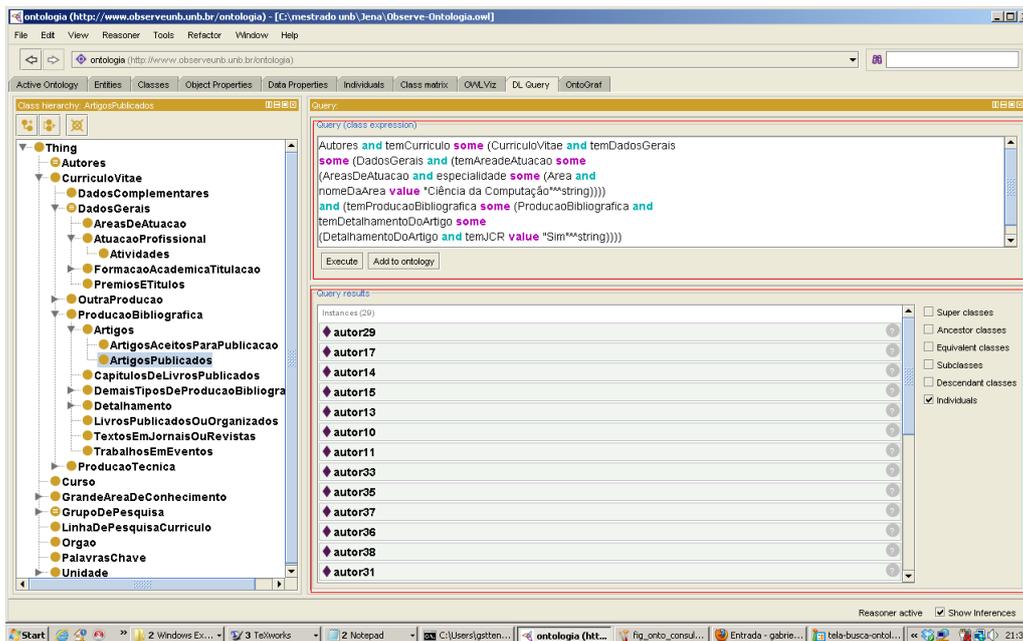


Figura 7.24: Exemplo de consulta realizada no Protégé para a Questão 6.

Buscar

Área

Buscar

Busca Utilizando Ontologia

Nome	Grupo	Produção Bibliográfica
Alba Cristina Magalhães Alves de Melo	-Sistemas de Computação -TransLab	-Exact Parallel Alignment Of Megabase Genomic Sequences With Tunable Work Distribution -Exact Parallel Alignment Of Megabase Genomic Sequences With Tunable Work Distribution
Aletéia Patrícia Favacho de Araújo		
Alexandre Zaghetto	-InfoKnow-Sistemas Computacionais para Tratamento de Informação e Conhecimento -Arquitetura da Informação -LISA	-Pre- and postprocessing for multilayer compression of scanned documents -Segmentation-Driven Compound Document Coding Based on H.264/AVC-INTRA
Aluizio Arceia Junior		
André Costa Drummond	-LARF	-Um Algoritmo Rápido, Eficiente e Justo para Agregação Dinâmica de Tráfego em Redes WDM -Path Protection WDM Networks with Impaired-Transmission -Fairness in Zone-Based Algorithms for Dynamic Traffic Grooming in WDM Mesh Networks -A Multiobjective Fuzzy Bandwidth Partitioning Model for Self-Sizing Networks -A Fixed-Parameter Tractable Approach for the Wavelength Assignment Problem in Transparent Networks
Carla Denise Castanho	-ACV Brasil -GIPSE	-A Parallel Algorithm for constructing Strongly Convex Superhulls of Points -Polynomially Fast Parallel Algorithms for Some P-Complete Problems -A Service Oriented Architecture for Ubiquitous Applications
Cíaudia Nalon	-Teoria da Computação	-Tableaux for logics of time and knowledge with interactions relating to synchrony -Causal Resolution for Normal Modal Logics
		-ObserveUnBumPortalWebDeRedeSocialCientifica

Figura 7.25: Consulta realizada no ObserveUnB para a questão 7.

Tabela 7.14: Classes e propriedades OWL para responder à Questão 7.

Classe	Propriedade	Descrição
Autores	temCurriculo pertenceGrupoDePesquisa	As instâncias desta classe têm no máximo uma associação através da propriedade temCurriculo com instâncias da classe CurriculoVitae e associação através da propriedade pertenceGrupoDePesquisa com instâncias da classe GrupoDePesquisa.
GrupoDePesquisa	area-predominante	As instâncias desta classe são associadas a propriedade listada para identificar a área do grupo.
CurriculoVitae	temDadosGerais temOutraProducao temProducaoBibliografica temProducaoTecnica	As instâncias desta classe podem ter somente associação através das propriedades listadas com instâncias das classes DadosGerais, OutraProducao, ProducaoBibliografica e ProducaoTecnica.
ProducaoBibliografica	especialidade	As instâncias desta classe têm somente associação através da propriedade especialidade com instâncias das classes Area, SubArea ou Especialidade.
DadosGerais	temAreadeAtuacao temAtuacaoProfissional	As instâncias desta classe têm pelo menos uma associação através das propriedades citadas com instâncias das classes AreasDeAtuacao e AtuacaoProfissional.
AreasDeAtuacao	especialidade	As instâncias desta classe têm associação com instâncias das classes Area, SubArea ou Especialidade.
AtuacaoProfissional	temAtividades	As instâncias desta classe têm associação com instâncias da classe Atividades.
Atividades	peloOrgao pelaUnidade	As instâncias desta classe têm associação com instâncias da classe Órgão e da classe Unidade.

7.3 Análise dos Resultados

O processo de exploração de métodos de cálculo de reputação e de ontologias proporcionou a investigação de algoritmos, ferramentas e conceitos relacionados à Web Semântica e reputação, que por sua vez implicou na pesquisa sobre redes sociais. A escolha por um ambiente científico de rede social foi apropriada, pois a ARS é um paradigma em evidência na sociedade da informação. Devido à complexidade de uma rede social, existem diversas abordagens de ARS, sendo que o cálculo da reputação dos membros da rede é uma das abordagens mais frequentemente utilizadas. Nas Seções 7.3.1 e 7.3.2 são discutidos os resultados obtidos neste trabalho.

7.3.1 Modelo de Reputação

Na direção da importância da ARS na sociedade da informação, o sistema de cálculo de reputação apresentou-se uma característica importante na rede, visto que agrega informações que visam dar suporte à composição de grupos de pesquisa e redes de colaboração, além de servir como apoio à tomada de decisões com relação a distribuição de recursos à pesquisa, entre outras iniciativas. Além dessas vantagens, a identificação da reputação pode servir como um estímulo a produção científica, desde que o pesquisador almeje alcançar um nível mais alto na tabela de classificação de reputação.

Durante o processamento do cálculo da reputação, um dos problemas encontrados concentrou-se no fato que nem todos os currículos encontram-se atualizados ou possuem dados para todos os elementos do modelo proposto (e.g., orientação concluída) o que acabou determinando um índice de reputação abaixo do esperado, principalmente para os pesquisadores classificados nos níveis 1 e 2. Outra questão que influenciou na alta concentração de pesquisadores nesses níveis, é a existência de pesquisadores que concluíram o curso de doutorado a menos de 5 anos, que estão iniciando sua carreira como orientadores de mestrado e de doutorado, possuindo pouca inserção internacional.

Além desses fatores, muitos dos pesquisadores analisados não mantêm um equilíbrio em sua trajetória de pesquisa, pois alguns elementos ficam bem pontuados e outros não. Isto faz com que a média final do cálculo do índice de reputação seja baixa. Em contrapartida, no grupo avaliado existem pesquisadores que concluíram o curso de doutorado a um bom tempo, alguns realizaram curso de pós-doutorado, possuem um número significativo de orientações, bem como possuem inserção internacional por meio de publicação de artigos e livros.

Com relação às entrevistas realizadas para avaliar o modelo de reputação proposto para a área de Ciência da Computação, algumas das sugestões dos pesquisadores avaliados são bastante válidas e relevantes:

- Inclusão de outros índices de impacto (e.g., Scopus, H-Index);
- Distinção entre os meios de publicação;
- Inclusão do tempo de carreira do pesquisador e a temporalização dos documentos;
- Validação dos itens avaliados por parte do usuário ou pesquisador avaliado.

Entretanto, devido à complexidade em se identificar essas informações no currículo *Lattes*, visto que este disponibiliza apenas algumas delas (i.e., ano e fonte da publicação,

e ano de ingresso na UnB) e no formato textual, de difícil processamento, tais melhorias poderão ser consideradas em trabalhos futuros. Com relação à validação dos itens por parte do usuário, é necessário um estudo para verificar a viabilidade da interação do usuário de forma que esta influencie de forma legítima na sua reputação.

Como o modelo de reputação foi definido baseado no currículo *Lattes*, no modelo de Cervi et al. (2011) e em recomendações da CAPES para a área, vários dos questionamentos podem ser justificados. Segundo CAPES (2009), para a avaliação do perfil do docente considera-se a titulação, o aprimoramento e a experiência. Dessa forma, valoriza-se a formação acadêmica do corpo docente, onde a renovação, a especialização e o percentual de pós-doutores do corpo docente é avaliada, bem como são valorizados os docentes com bolsas de produtividade do CNPq. Isso explica a relevância das categorias Identificação, ressaltando o elemento Pós-doutorado, e Bolsista no modelo proposto.

Em razão de o modelo focar na pesquisa e produção científica a proposta procura abranger vários itens que compõem esta abordagem na carreira. Seguindo esta abordagem, com exceção da categoria Identificação, elementos que abrangem a graduação não foram contemplados no modelo, e.g. graduação em Banca e Orientação.

Além disso, projetos institucionais e projetos de intercâmbio são valorizados na área, incentivando iniciativas bem sucedidas de transferência de conhecimento, de cooperação entre instituições nacionais, bem como a colaboração científica. Essa característica pode ser identificada na participação de projetos e na rede de co-autoria do pesquisador, onde quanto maior a rede do pesquisador, maior tende a ser a colaboração e a disponibilização de conhecimento.

Com relação a categoria Produção, os elementos que se referem a produção de livros e capítulos não foram valorizados no modelo, pois segundo o documento de área da Ciência da Computação definido pela CAPES, a área utiliza de modo pouco expressivo esta modalidade de publicação (CAPES, 2009). Nesta área, a qualidade, a quantidade e a regularidade das publicações são bastante valorizadas, onde no modelo proposto, a qualidade se justifica pela diferença existente entre as publicações com e sem JCR.

Analisando-se os gráficos 7.16, 7.17 e 7.18, percebe-se que do total de 49 pesquisadores avaliados, o IPC e o índice DBLP de 29 pesquisadores possuem valores próximos, ou seja, a diferença entre eles é menor que 50% do maior valor destes índices, e.g. um pesquisador cujos valores do IPC e do DBLP são 50 e 40, respectivamente, onde a diferença entre eles é 10, ou seja é menor que 25, corresponde a 50% do maior índice (50). Entretanto, em 20 pesquisadores, esta diferença ultrapassa o percentual de 50%, representando uma disparidade entre os valores do IPC e DBLP. A justificativa para isto encontra-se no fato que a métrica DBLP considera apenas as publicações de artigos em periódicos e em conferências. Enquanto que a métrica da reputação proposta ou IPC considera um grupo de elementos que fazem parte da trajetória científica de um pesquisador, construída ao longo de sua carreira acadêmica.

Analisando os resultados da implementação do modelo de reputação e das entrevistas realizadas com os docentes, percebe-se que o modelo proposto retrata com realidade boa parte da comunidade acadêmica do Departamento de Ciência da Computação. No entanto, nota-se que ainda está longe de um parâmetro ideal que seja geral para todas as áreas. Para que o modelo chegue próximo deste ideal, faz-se necessário uma análise mais detalhada do perfil do pesquisador e de cada área de conhecimento, bem como de suas

sub-áreas, de forma que se alcance um modelo específico para cada área ou até mesmo sub-área de conhecimento.

7.3.2 Ontologia

Visando auxiliar na identificação das informações necessárias para o cálculo da reputação, foi importante o uso das ontologias disponibilizadas pela comunidade Conscientias. No entanto, a ontologia original não contempla informações relevantes para o modelo, como Bolsista do CNPq, publicações com JCR e a estrutura acadêmica das IFES. Dessa forma, a iniciativa de primeiramente consultar os repositórios de ontologias antes de criar uma nova ontologia, fez com que esforços fossem concentrados em outras fases do projeto como a definição, o teste e a validação do modelo de reputação, bem como a extensão e a validação da ontologia utilizada.

A ontologia estendida se mostrou consistente após a integração das ontologias na linguagem OWL, a inclusão de classes e a criação das instâncias. Durante o processo de extensão deparou-se com diversos problemas, dentre os quais destacam-se: dificuldade de conversão da linguagem XSD/DTD para OWL, falhas na integração das ontologias e complexidade do código OWL e da API Jena. Essas dificuldades foram superadas através da adaptação da estrutura de conceitos, da criação de instâncias e da realização de consultas na ontologia. Tudo isso pode ser feito por meio do uso do editor de ontologias Protégé e também por meio de uma aplicação Java que faz uso da API Jena desenvolvida para este fim.

A partir da ferramenta utilizada como estudo de caso - ObserveUnB, pode-se obter dados estatísticos relacionados à produção científica dos 2.181 pesquisadores da UnB, com 115.291 unidades de produção científica, correspondendo ao total da produção acadêmica dos pesquisadores vinculados a instituição, incluindo artigos em periódicos, trabalhos em eventos, artigos aceitos para publicação, livros publicados ou organizados, capítulos de livros publicados, textos em jornais ou revistas e demais tipos de produção bibliográfica.

Com essas informações, pode-se visualizar o gráfico quantitativo de produção científica de cada unidade acadêmica e de cada programa de pós-graduação. Baseado no currículo *Lattes* é apresentado à produção científica de cada pesquisador, incluindo a área de conhecimento, palavras-chave e ano da publicação. Com a rede de pesquisa obtêm-se todos os pesquisadores com os quais um determinado docente publicou. Enquanto que, na rede de relacionamentos em forma de grafo é possível visualizar os docentes da instituição com os quais o pesquisador produziu, conforme apresentado na Figura 7.8.

A partir da rede de colaboração científica pode-se identificar o nível de reputação de cada membro e inferir respostas a questionamentos sobre pesquisadores que mais produzem isoladamente ou com quem eles produzem em parceria. Da pesquisa georreferenciada obtêm-se a localização de um pesquisador ou de uma unidade acadêmica na UnB. E os gráficos quantitativos de produção científica apresentam respostas às questões sobre quais unidades acadêmicas e programas de pós-graduação que mais produzem na instituição.

Aliado a essas informações do ObserveUnB, a ontologia oferece recursos que enriquecem as opções de pesquisa da ferramenta. Consultas que permitem a busca de um pesquisador, sua produção científica, sua vinculação com a unidade acadêmica e com projetos de pesquisa, seus relacionamentos com outros pesquisadores, suas áreas de pesquisa, são algumas das facilidades de busca oferecidas pela ontologia. O mecanismo de inferência

oferecido pelo Protégé e pelo Jena foi utilizado para validar a ontologia estendida, bem como para inferir relacionamentos entre conceitos não explicitados na ontologia.

Diante da variedade de informação disponibilizada pela ontologia cabe ao usuário ou desenvolvedor filtrá-las e relacioná-las de forma que atenda às suas necessidades científicas e acadêmicas.

Com base nos resultados obtidos por meio da aplicação e validação do modelo de reputação e da ontologia no ObserveUnB, verifica-se que a proposta deste trabalho é adequada às redes sociais científicas, uma vez que baseia-se em informações institucionais obtidas da Plataforma *Lattes* (sistema de abrangência nacional) e em recomendações de critérios de classificação de programas e publicação da CAPES, podendo ser utilizada em diferentes áreas do conhecimento e em diferentes IFES. Talvez a redefinição dos pesos e elementos do modelo de reputação seja recomendável para melhor adequação das características de área e culturais das diferentes instituições de ensino e pesquisa do país. No entanto, a utilidade do modelo é inquestionável para aplicações de avaliação científica baseada na comunidade de pesquisadores existentes na base *Lattes*.

Ainda em relação a ontologia estendida da base *Conscientias*, novos conceitos podem ser definidos para compor a ontologia, de forma a abranger aspectos importantes que não foram abordados por este trabalho, como a dedicação ao ensino de graduação no perfil dos docentes, tão importante para uma boa formação de profissionais qualificados no país.

Foram também resultados desta pesquisa a elaboração de dois artigos, sendo que um foi publicado (Ralha et al., 2012), enquanto o outro encontra-se em fase de análise (Guerra et al., 2012).

Capítulo 8

Conclusões e Trabalhos Futuros

Neste trabalho, foi apresentado um modelo de reputação e uma ontologia estendida de currículo *Lattes* e de Diretório de Grupos de Pesquisa aplicados à uma rede social científica. Como estudo de caso foi utilizado o *ObserveUnB*, um portal Web de rede social científica da UnB, o qual possui como desafio a integração de um grande volume de dados provenientes de diferentes repositórios, bem como o gerenciamento e a disponibilização do conhecimento científico.

Neste contexto, o objetivo deste trabalho que consiste na disponibilização da informação e conhecimento científico na rede social acadêmica da UnB foi alcançado. Para isso, foi utilizado um modelo de reputação para avaliar os membros da rede social do *ObserveUnB*, cujos resultados podem servir como um índice de produção científica dos integrantes da rede ou como indicador de especialistas em determinada área de conhecimento.

Além do modelo de reputação, foi utilizada uma ontologia, que auxiliou na integração de informações do domínio de conhecimento da pesquisa acadêmica, bem como serviu como meio para solucionar o problema da duplicidade e ambiguidade de dados. O uso da ontologia possibilitou a realização de consultas semânticas, através de sua estrutura de conceitos e do mecanismo de inferência disponibilizado pelo *Protégé* e pelo *Jena*. Dessa forma, o modelo de reputação e a ontologia mostraram-se úteis no grande desafio enfrentado pelo *ObserveUnB*.

No decorrer deste trabalho, foram elucidados temas relacionados ao domínio acadêmico, bem como conceitos que envolvem redes sociais, reputação e ontologias. Com isso, pode-se verificar a importância de um sistema de reputação em aplicações de comércio eletrônico até mesmo em aplicações de redes sociais, que por sua vez constitui o foco deste projeto.

No que diz respeito à ontologia, foi possível conhecer os recursos que suas linguagens oferecem, bem como suas diversas aplicações. No entanto, o uso de uma ontologia mostrou-se uma tarefa complexa, em razão de esta tecnologia estar inerente no meio acadêmico, sendo considerada ainda recente e portanto, pouco difundida em outras comunidades de tecnologia da informação.

Como trabalhos futuros, novas funcionalidades podem ser incorporadas ao *ObserveUnB* para permitir a visualização da evolução da reputação e da produção científica dos pesquisadores ao longo do tempo, como por exemplo nos triênios de avaliação dos programas de pós-graduação pela CAPES. Para viabilizar uma maior interação do cidadão brasileiro com os pesquisadores no portal, o modelo de reputação poderá ser adaptado a

uma rede social dinâmica, permitindo que alguns fatores influenciem no cálculo, e.g. a interação entre os membros da rede por meio de troca de mensagens de twitter, e-mails, ou a participação do usuário mais experiente no cálculo da reputação de outros menos experientes, possibilitado o aumento ou decréto do grau de reputação dos membros da rede.

O modelo de reputação pode ser aprimorado a partir da inclusão de novos elementos importantes para a medição da reputação, bem como a partir da realização de novas validações, e.g. comparações com outros métodos de reputação ou a sua aplicação em outras comunidades de pesquisadores.

Com relação à rede de relacionamento, pode-se apresentar visualmente diferenças no grafo, podendo diferenciar as arestas (relacionamentos) e os nós (membros) de acordo com a quantidade de relacionamento existente entre os membros e o nível de reputação do membro na rede, e.g. quanto maior a quantidade de relacionamento existente entre dois pesquisadores maior a espessura da aresta, e quanto maior o nível de reputação do pesquisador maior o diâmetro do nó. E pode-se analisar o comportamento dos usuários e membros da rede social a partir da observação da interação entre os usuários e membros da rede.

Também seria importante realizar a indexação da base de produção científica do ObserveUnB utilizando dados de repositórios internacionais. Essas bases agregariam uma avaliação qualitativa da produção científica através de outros índices de impacto além do *Journal Citation Reports*, o que associado ao cálculo de reputação retrataria a realidade acadêmica dos pesquisadores do país, conforme parâmetros estabelecidos internacionais, e requeridos pelos órgãos de fomento, como a CAPES, CNPq, Finep, entre outros.

Além das bases de dados internacionais, pode-se integrar a ontologia estendida Conscientias à ontologia FOAF visando obter maior representação dos pesquisadores brasileiros, bem como dos pesquisadores estrangeiros.

Como pode ser observado, existem diversas sugestões de trabalhos futuros, uma vez que o modelo de reputação e a ontologia aliados ao ObserveUnB oferecem uma grande riqueza de possibilidades de uso e extensão, envolvendo um trabalho de pesquisa multidisciplinar entre as áreas de Ciências Exatas, como a Ciência da Computação, Estatística e a Matemática, no que tange os modelos de cálculo e ponderação dos elementos do grafo; bem como as áreas de Ciências Humanas e Sociais como a Sociologia, a Psicologia, a Administração para auxiliar na avaliação do complexo relacionamento de elementos envolvidos na rede, o qual pode ser interpretado através de aspectos intra e inter-institucionais.

Apêndice A

Questionário de Avaliação do Modelo de Reputação

DADOS PESSOAIS

Nome:

Tempo de docência (em anos):

Tempo de docência na UnB (em anos):

Faixa etária:

- Até 30 anos
- 31 a 40 anos
- 41 a 50 anos
- 51 a 60 anos
- Não quero informar

Escolaridade:

- Pós-doutorado
- Doutorado
- Mestrado
- Especialização
- Graduação

Bolsista em Produtividade do CNPq:

- Sim
- Não

SOBRE O MODELO DE REPUTAÇÃO

1. Você já conhece o ObserveUnB?

- Sim
- Não (para conhecê-lo visite www.observeunb.unb.br)

2. **Você considera que a abordagem de definição do modelo de reputação que foca na produção científica da carreira acadêmica do pesquisador está correta?**

- Sim
- Não
- Em parte

Por favor, justifique sua resposta.

3. **Você concorda com as categorias e os elementos definidos para o modelo de reputação?**

- Sim
- Não
- Em parte

Por favor, justifique sua resposta.

4. **Você concorda com a ordem de relevância atribuída aos elementos e categorias definidos para o modelo de reputação?**

- Sim
- Não
- Em parte

Por favor, justifique sua resposta.

5. **Analisando o Gráfico 1 do documento enviado em anexo, você acha que o resultado obtido da reputação reflete a realidade acadêmica de pesquisa da sua unidade acadêmica?**

- Sim
- Não
- Em parte

Por favor, justifique sua resposta.

INFORMAÇÕES ADICIONAIS

Caso você queira acrescentar mais informações, por favor, escreva nas linhas abaixo.

Apêndice B

Consultas realizadas na Ontologia

A seguir são exibidos os códigos das 7 consultas realizadas utilizando-se a ontologia estendida do ObserveUnB. As consultas foram executadas nas linguagens *DL-Query* e SPARQL, cujos resultados são apresentados no Capítulo 7.

Questão 1 : Quem são os pesquisadores que atuam na área Ciência da Computação?

Em SPARQL:

```
query = PREFIX ontologia: <http://www.observeunb.unb.br/ontologia/#>
SELECT distinct ?autor
WHERE {?autor ontologia:temCurriculo ?curriculo .
?curriculo ontologia:temDadosGerais ?dadosgerais .
?dadosgerais ontologia:temAreadeAtuacao ?area .
?area ontologia:especialidade ?especialidade .
?especialidade ontologia:nomeDaArea
"Ciência da Computação"^^< http://www.w3.org/2000/01/rdf-schema#string>}}
```

Em *DL-Query*:

```
temCurriculo some (CurriculoVitae and temDadosGerais some (DadosGerais
and temAreadeAtuacao some (AreasDeAtuacao and especialidade some
(Area and nomeDaArea value "Ciencia da Computação"^^string))))
```

Questão 2 : Quem são os pesquisadores da área Ciência da Computação que possuem produção científica?

Em SPARQL:

```
query = PREFIX ontologia: <http://www.observeunb.unb.br/ontologia#>
SELECT distinct ?autor
WHERE {?autor ontologia:temCurriculo ?curriculo .
?curriculo ontologia:temDadosGerais ?dadosgerais .
?dadosgerais ontologia:temAreadeAtuacao ?area .
?area ontologia:especialidade ?especialidade .
?especialidade ontologia:nomeDaArea
"Ciência da Computação"^^<http://www.w3.org/2000/01/rdf-schema#string> .
?curriculo ontologia:temProducaoBibliografica ?producao}}
```

```

UNION {?curriculo ontologia:temProducaoTecnica ?tecnica}
UNION {?curriculo ontologia:temOutraProducao ?outra} .
?autores ontologia:nomeCompletoDoAutor ?autor}
ORDER BY ?autor

```

Em *DL-Query*:

```

temCurriculo some (CurriculoVitae and ((temProducaoBibliografica some
ProducaoBibliografica) or (temProducaoTecnica some ProducaoTecnica)
or (temOutraProducao some OutraProducao)) and temDadosGerais
some (DadosGerais and temAreadeAtuacao some (AreasDeAtuacao
and especialidade some (Area and nomeDaArea
value "Ciência da Computação"^^string))))

```

Questão 3 : Quem são os pesquisadores vinculados ao Departamento de Ciência da Computação da UnB?

Em SPARQL:

```

query = PREFIX ontologia: <http://www.observeunb.unb.br/ontologia#>
SELECT ?autor
WHERE {?autor ontologia:temCurriculo ?curriculo .
?curriculo ontologia:temDadosGerais ?dadosgerais .
?dadosgerais ontologia:temAtuacaoProfissional ?atuacaoprof .
?atuacaoprof ontologia:temAtividades ?atividade .
?atividade ontologia:peloOrgao ontologia:Orgao1 .
?atividade ontologia:pelaUnidade
ontologia:DepCienciaComputacao
?autores ontologia:nomeCompletoDoAutor ?autor}
ORDER BY ?autor

```

Em *DL-Query*:

```

temCurriculo some (CurriculoVitae and temDadosGerais some
(DadosGerais and temAtuacaoProfissional some
(AtuacaoProfissional and temAtividades some
(Atividades and peloOrgao value Orgao1 and pelaUnidade
value DepCienciaComputacao))))

```

Questão 4 : Quem são os pesquisadores vinculados ao grupo de pesquisa Teoria da Computação?

Em SPARQL:

```

query = PREFIX ontologia: <http://www.observeunb.unb.br/ontologia#>
SELECT ?autor
WHERE {?autor ontologia:pertenceGrupoDePesquisa
ontologia:Teoria_da_Computação .
?autores ontologia:nomeCompletoDoAutor ?autor}
ORDER BY ?autor

```

Em *DL-Query*:

Autores and pertenceGrupoDePesquisa value Teoria_da_Computação

Questão 5 : Quem são os co-autores da publicação bibliográfica *ObserveUnB - Um Portal Web de Rede Social Científica*?

Em SPARQL:

```
query = PREFIX ontologia: <http://www.observeunb.unb.br/ontologia#>
SELECT distinct ?autor
WHERE {?autor ontologia:temCurriculo ?curriculo .
?curriculo ontologia:temProducaoBibliografica
ontologia:ObserveUnBUMPortalWebDeRedeSocialCientifica
?autores ontologia:nomeCompletoDoAutor ?autor}
ORDER BY ?autor
```

Em *DL-Query*:

```
Autores and temCurriculo some
(CurriculoVitae and (temProducaoBibliografica
value ObserveUnBUMPortalWebDeRedeSocialCientifica))
```

Questão 6 : Quem são os pesquisadores da área Ciência da Computação que possuem produção científica com JCR?

Em SPARQL:

```
query = PREFIX ontologia: <http://www.observeunb.unb.br/ontologia#>
SELECT distinct ?autor
WHERE {?autores ontologia:temCurriculo ?curriculo .
?curriculo ontologia:temDadosGerais ?dadosgerais .
?dadosgerais ontologia:temAreadeAtuacao ?area .
?area ontologia:especialidade ?especialidade .
?especialidade ontologia:nomeDaArea
"Ciência da Computação"^^<http://www.w3.org/2000/01/rdf-schema#string>
. ?curriculo ontologia:temProducaoBibliografica ?producao .
?producao ontologia:temDetalhamentoDoArtigo ?detalhamento .
?detalhamento ontologia:temJCR
"Sim"^^<http://www.w3.org/2000/01/rdf-schema#string> .
?autores ontologia:nomeCompletoDoAutor ?autor}
ORDER BY ?autor}
```

Em *DL-Query*:

```
Autores and temCurriculo some (CurriculoVitae and temDadosGerais
some (DadosGerais and (temAreadeAtuacao some
(AreasDeAtuacao and especialidade some (Area and
nomeDaArea value "Ciência da Computação"^^string))))
and (temProducaoBibliografica some (ProducaoBibliografica and
temDetalhamentoDoArtigo some
(DetalhamentoDoArtigo and temJCR value "Sim"^^string))))
```

Questão 7 : Quais são os termos relacionados a área Ciência da Computação?

Em *DL-Query*:

```
temCurriculo some (CurriculoVitae and ((temDadosGerais some
(DadosGerais and temAtuacaoProfissional some (AtuacaoProfissional
and temAtividades some (Atividades and pelaUnidade value
DepCienciaComputacao)))))) or (temDadosGerais some (DadosGerais
and temAreadeAtuacao some (AreasDeAtuacao and (especialidade
some (GrandeAreaDeConhecimento and nomeDaGrandeArea
value "Ciência da Computação"^^string) or (Area and
nomeDaArea value "Ciência da Computação"^^string) or SubArea
(and nomeDaSubArea value "Ciência da Computação"^^string))))))
or (temProducaoBibliografica some (ProducaoBibliografica and
especialidade some (especialidade some (GrandeAreaDeConhecimento
and nomeDaGrandeArea value "Ciência da Computação"^^string)
or (Area and nomeDaArea value "Ciência da Computação"^^string)) or
(SubArea and nomeDaSubArea value "Ciência da Computação"^^string))))
or (GrupoDePesquisa and PesquisaAREA-PREDOMINANTE value
"Ciência da Computação"^^string) or (Area and nomeDaArea value
"Ciência da Computação"^^string) or (SubArea and pertenceaArea some
(Area and nomeDaArea value "Ciência da Computação"^^string))
```

Referências

- Acioli, S. (2007). Redes sociais e teoria social: Revendo os fundamentos do conceito. *Informação Informação*, 12(12). 9
- ACM (2011). Acm digital library. Disponível em: <http://http://dl.acm.org/>, Acesso em: Setembro, 2011. 47
- Adali, S., Escriva, R., Goldberg, M. K., Hayvanovych, M., Magdon-Ismail, M., Szymanski, B. K., Wallace, W. A., and Williams, G. T. (2010). Measuring behavioral trust in social networks. In *Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on*, pages 150–152, Vancouver, Canada. 16
- Almeida, M. B. and Bax, M. P. (2003). Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, 32:7–20. 25, 26, 36
- ALTOVA (2012). Altova semanticworks. Disponível em: http://www.altova.com/download/semanticworks/semantic_web_rdf_owl_editor.html, Acesso em: Fevereiro, 2012. 38
- Barabasi, A., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *PHYSICA A*, 311:3. 12
- Bechhofer, S., Horrocks, I., Goble, C. A., and Stevens, R. (2001). Oiled: A reason-able ontology editor for the semantic web. In *Proceedings of the Joint German/Austrian Conference on AI: Advances in Artificial Intelligence*, KI '01, pages 396–408, London, UK. Springer-Verlag. 38
- Benevenuto, F., Almeida, J. M., and Silva, A. S. (2011). *Explorando Redes Sociais Online: Da Coleta e Análise de Grandes Bases de Dados às Aplicações*, chapter 2, pages 63–94. XXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos. Sociedade Brasileira de Computação, Porto Alegre, BRA. 6
- Bergman, M. K. (2010). Listing of 185 ontology building tools. Disponível em: <http://www.mkbergman.com/904/listing-of-185-ontology-building-tools/>, Acesso em: Novembro, 2011. 38
- Boyd, D. and Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1-2). 6, 7

- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., and Yergeau, F. (2008). Extensible markup language (xml) 1.0 (fifth edition). Disponível em: <http://www.w3.org/TR/REC-xml/>, Acesso em: Novembro, 2011. 31
- Brickley, D. and Miller, L. (2000). Friend of a friend. Disponível em: <http://www.foaf-project.org/>, Acesso em: Setembro, 2011. 47, 49
- Bräscher, M. and Carlan, E. (2010). *Sistemas de organização do conhecimento: antigas e novas linguagens*, chapter 8, pages 147–176. Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento. IBICT, Brasília, BRA. 26
- Calvanese, D. and Lenzerini, M. (2001). Ontology of integration and integration of ontologies. In *Procs. of the 2001 Description Logic Workshop (DL 2001)*. 29
- CAPES (2008). Tabela de Áreas de conhecimento. Disponível em: <http://www.capes.gov.br/avaliacao/tabela-de-areas-de-conhecimento>, Acesso em: Março, 2011. 71
- CAPES (2009). Documento de Área - 2009. Disponível em: http://capes.gov.br/images/stories/download/avaliacao/COMPUTACAO_05mar10.pdf, Acesso em: Fevereiro, 2012. 81, 106
- Casare, S. and Sichman, J. (2005). Towards a functional ontology of reputation. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS '05, pages 505–511, New York, USA. ACM. 19, 30
- Cervi, C. R., Galante, R., and de Oliveira, J. P. M. (2011). Identificando a reputação de pesquisadores usando um modelo de perfil adaptativo. *Anais do XXXVIII SEMISH*. vi, vii, ix, xii, 17, 52, 53, 54, 55, 56, 60, 72, 75, 81, 106
- CiteSeer (2011). Scientific literature digital library. Disponível em: <http://citeseer.ist.psu.edu>, Acesso em: Setembro, 2011. 47
- CNPq (2011). Diretórios dos grupos de pesquisa no brasil. Disponível em: <http://dgp.cnpq.br/diretorioc/>, Acesso em: Fevereiro, 2011. 2, 3
- CNPq (2012a). Bolsas de produtividade em pesquisa e tecnologia. Disponível em: <http://efomento.cnpq.br/efomento/distribuicaoGeografica/distribuicaoGeografica.do?metodo=apresentar>, Acesso em: Maio, 2012. 88
- CNPq (2012b). Conselho nacional de desenvolvimento científico e tecnológico. Disponível em: <http://www.cnpq.br>, Acesso em: Fevereiro, 2012. 12, 13
- Co.KG, R. S. G. . (2012). Racerpro. Disponível em: <http://www.racer-systems.com/>, Acesso em: Maio, 2012. 40
- Conscientias (2012). Comunidade conscientias. Disponível em: <http://lmpl.cnpq.br/lmpl/index.jsp>, Acesso em: Maio, 2012. 44
- Conscientias, C. (2002). Padronização xml: Grupos de pesquisa. Disponível em: <http://lmpl.cnpq.br/lmpl/?go=grupo.jsp>, Acesso em: Março, 2011. 1, 27, 29, 38, 45, 62, 65

- Conscientias, C. (2006). Padronização xml: Curriculum vitae. Disponível em: <http://impl.cnpq.br/impl/?go=cv.jsp>, Acesso em: Abril, 2011. 1, 26, 27, 29, 38, 44, 46, 62, 65
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2002). *Algoritmos: Teoria e Prática*, volume 2. Editora Elsevier, Rio de Janeiro, BRA, 4 edition. 7
- Cruz, C. C. P., Motta, C. L. R., and Santoro, F. M. (2007). *ReCoP: Um Modelo para Reputação em Comunidades de Prática*, pages 92 – 102. Sociedade Brasileira de Computação, Porto Alegre, BRA. 20
- D’Arcus, B. and Giasson, F. (2009). The bibliographic ontology. Disponível em: <http://http://bibliontology.com/>, Acesso em: Maio, 2012. 49
- Dasgupta, P. (2000). *Trust: Making and Breaking Cooperative Relations*, chapter Trust as a Commodity, pages 49–72. Department of Sociology, University Oxford. This was originally published in hardcopy in 1988, PDF is 2000 electronic version. 15
- DBLP (2011). Digital bibliography and library project. Disponível em: <http://dblp.uni-trier.de>, Acesso em: Setembro, 2011. 47
- de H. Ferreira, A. B. (1986). *Novo Dicionário Aurélio da Língua Portuguesa*. Editora Nova York Fronteira, 2 edition. 15
- De Nicola, A., Missikoff, M., and Navigli, R. (2009). A software engineering approach to ontology building. *Inf. Syst.*, 34(2):258–275. 37
- de Oliveira Albuquerque, R. (2008). *Uma Proposta de um Modelo de Confiança Computacional para Grupos em Sistemas Distribuídos*. PhD thesis, Universidade de Brasília, Brasília, BRA. 15
- Duineveld, A. J., Stoter, R., Weiden, M. R., Kenepa, B., and Benjamins, V. R. (2000). Wondertools?: a comparative study of ontological engineering tools. *Int. J. Hum.-Comput. Stud.*, 52:1111–1133. 37
- Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of facebook “friends:” social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168. 7
- Farquhar, A., Fikes, R., and Rice, J. (1996). The ontolingua server: a tool for collaborative ontology construction. In *International Journal of Human-Computer Studies*. 38
- Ferris, G., Blass, F., Douglas, C., Kolodinsky, R., and Treadway, D. (2003). *Personal Reputation in Organizations*. Jerald Greenberg, New Jersey, USA. 15
- Fonseca, F. T. and Egenhofer, M. J. (1999). Ontology-driven geographic information systems. In *ACM-GIS*, pages 14–19. 25
- Foundation, T. A. S. (2012). Apache jena. Disponível em: <http://jena.apache.org/>, Acesso em: Fevereiro, 2012. 39, 40

- Friedman, E., Resnick, P., and Sami, R. (2007). *Game theory Algorithm*, chapter Manipulation-Resistant Reputation Systems. Cambridge University Press. 17
- Frozza, A. A. and Gonçalves, R. (2010). Jena – um framework web semântico em java. 41
- Gangemi, A., Pisanelli, D. M., and Steve, G. (1999). An overview of the onions project: Applying ontologies to the integration of medical terminologies. *Data and Knowledge Engineering*, 31:183–220. 36
- Gómez-Pérez, A. (1999). Ontological engineering: A state of the art. 28
- Gollapalli, S. D., Mitra, P., and Giles, C. L. (2011). Ranking authors in digital libraries. In *Proceeding of the 11th annual international ACM/IEEE joint conference on Digital libraries*, JCDL '11, pages 251–254, New York, USA. ACM. 51, 52
- Grüninger, M. and Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. In *International Joint Conference on Artificial Inteligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing*. 36
- Gruber, T. R. (1996). A translation approach to portable ontology specifications. Technical report ksl 92-71, Stanford University, Stanford, USA. 25, 28
- Guarino, N. (1998). Formal ontology and information systems. In *Proceedings of the 1st International Conference*. 26, 29
- Guarino, N. and Welty, C. (2004). *An Overview of OntoClean*, pages 151–172. The Handbook on Ontologies. Berlin:Springer-Verlag. 37
- Guerra, G. N., Ralha, C. G., Valentim, G. H. O., Sousa, M. S., Persegona, M. F. M., Godinho, M. L. S., Persegona, A. L. M., Blumenschein, R. N., Tome, M. V. D. F., and Carvalho, D. B. B. (2012). Observatório de conhecimento da rede de colaboração científica na universidade de brasília: Observe unb. *Revista Brasileira de Pós-Graduação*, pages 1–18. Sob avaliação. 108
- Guimarães, F. J. Z. (2002). Utilização de ontologias no domínio b2c. Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, BRA. 27
- Haarslev, V. and Moller, R. (2001). Racer system description. In *Proceedings of the First International Joint Conference on Automated Reasoning*, IJCAR '01, pages 701–706, London, UK. Springer-Verlag. 40
- Hogan, A., Harth, A., and Decker, S. (2006). Reconrank: A scalable ranking method for semantic web data with context. In *2nd Workshop on Scalable Semantic Web Knowledge Base Systems*. 52
- Horridge, M. (2004). Owlviz – a visualisation plugin for the protégé owl plugin. Disponível em: <http://www.co-ode.org/downloads/owlviz/OWLvizGuide.pdf>, Acesso em: Novembro, 2011. 63

- Horridge, M. (2011). A practical guide to building owl ontologies using protégé 4 and co-ode tools edition 1.3. 30, 32, 33, 35
- Hustadt, U. and Schmidt, R. A. (2000). Mspass: Modal reasoning by translation and first-order resolution. In *Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*, TABLEAUX '00, pages 67–71, London, UK. Springer-Verlag. 40
- Josang, A., Ismail, R., and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decis. Support Syst.*, 43(2):618–644. 15, 16, 23
- Katz, J. and Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1):1–18. 10, 11
- Kim, M. and Kim, M. (2008). Developing protégé plug-in: Owl ontology visualization using social network. *JIPS*, 4(2):61–66. 31
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *J. ACM*, 46:604–632. 21, 52
- Klyne, G. and Carroll, J. J. (2004). Resource description framework (rdf): Concepts and abstract syntax. Disponível em: <http://www.w3.org/RDF/>, Acesso em: Novembro, 2011. 31, 48
- Krafft, D. B., Cappadona, N. A., Caruso, B., Corson-Rikert, J., Devare, M., Lowe, B. J., and Collaboration, V. (2010). Vivo: Enabling national networking of scientists. In *WebSci10: Extending the Frontiers of Society On-Line*. 49
- Labbé, C. (2009). Ike antkare, one of the great stars in the scientific firmament. *ISSI Newsletter*, 6(2):48–52. 47
- Laender, A. H. F., Moro, M. M., da Silva, A. S., Jr., C. A. D., Gonçalves, M. A., Galante, R., Silva, A. J. C., Bigonha, C. A. S., Dalip, D. H., Barbosa, E. M., Borges, E. N., Cortez, E., Jr., P. S. P., de Alencar, R. O., Cardoso, T. N. C., and Salles, T. (2011). Ciênciabrasil - the brazilian portal of science and technology. In *SEMISH*. 49
- Lane, J. (2010). Let's make science metrics more scientific. *Nature*, 464(7288):488–489. 13
- LATTES (2012). Palatorma lattes. Disponível em: <http://lattes.cnpq.br/index.htm>, Acesso em: Maio, 2012. 13, 14
- Lima, D. O. (2010). Infraestrutura para gerenciamento de reputação de usuários e sua aplicação em um caso real. Master's thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, BRA. 22
- Liu, L., Munro, M., and Song, W. (2010). Evaluation of collecting reviews in centralized online reputation systems. In *WEBIST (2)*, pages 281–286. 17, 18
- Liu, X., Bollen, J., Nelson, M. L., and Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Inf. Process. Manage.*, 41(6):1462–1480. 9, 12

- López, M. F. (1999). Overview of methodologies for building ontologies. In *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem Solving Methods (KRR5)*, Stockholm, Sweden. 37
- Maedche, A. (2002). *Ontology Learning For The Semantic Web*. Kluwer Academic Publishers Group, Norwell, Massachusetts, USA. 28
- Maedche, E. and Staab, S. (2000). The text-to-onto ontology learning environment. In *Software Demonstration at ICCS-2000 - Eight International Conference on Conceptual Structures*. 38
- Maia, M. d. F. S. and Caregnato, S. E. (2008). Co-autoria como indicador de redes de colaboração científica. *Perspectivas em Ciência da Informação*, 13:18–31. 9, 11
- Marteleto, R. M. (2010). Redes sociais, mediação e apropriação de informações: situando campos, objetos e conceitos na pesquisa em ciência da informação. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 3(1). 9
- Matheus, R. F. and de Oliveira Silva, A. B. (2009). *Fundamentação básica para análise de redes sociais: conceitos, metodologia e modelagem matemática*, pages 239–288. Angellara Editora, São Paulo, BRA. 9
- Matsuo, Y., Tomobe, H., Hasida, K., and Ishizuka, M. (2004). Finding social network for trust calculation. In *In proc. ECAI2004*, pages 510–514. 52
- McGuinness, D. L. and van Harmelen, F. (2004). Owl web ontology language overview. Disponível em: <http://www.w3.org/TR/2004/REC-owl-features-20040210/>, Acesso em: Novembro, 2011. 35
- Meira, W., dos Santos Filho, W., Pappa, G. L., Guedes, D., Veloso, A., Almeida, V. A. F., Pereira, A. M., and Guerra, P. H. C. (2010). Observatório da web: Uma plataforma de monitoração, síntese e visualização de eventos massivos em tempo real. *ANAIS do XXX Congresso da Sociedade Brasileira de Computação*. 51
- Mena, E., Illarramendi, A., Kashyap, V., and Sheth, A. P. (2000). Observer: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. *Distrib. Parallel Databases*, 8:223–271. 2
- Mena-Chalco, J. P. and Cesar-Jr., R. M. (2012). *Bibliometria e Cientometria: reflexões teóricas e interfaces*, chapter Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. São Carlos: Pedro & João. (in press). 13
- Michaelis (2012). Dicionário moderno da língua portuguesa. Disponível em: <http://michaelis.uol.com.br/moderno/portugues>, Acesso em: Maio, 2012. 10, 15, 25
- Morris, A., Velegrakis, Y., and Bouquet, P. (2008). Entity identification on the semantic web. In *SWAP'08*, pages 1–1. 27, 75
- Mui, L. (2002). Computational models of trust and reputation: Agents, evolutionary games, and social networks. Master's thesis, Massachusetts Institute of Technology, Cambridge, USA. 18, 19, 59

- Mui, L., Mohtashemi, M., and Halberstadt, A. (2002). A computational model of trust and reputation for e-businesses. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 7 - Volume 7*, HICSS '02, page 188, Washington, USA. IEEE Computer Society. 16
- Necib, C. B. and Freytag, J. C. (2004). Using ontologies for database query reformulation. In *ADBIS (Local Proceedings)'04*, pages 1–1. 25
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM REVIEW*, 45:167–256. 6, 7, 8, 12
- Noy, N. F. (2004). Semantic integration: A survey of ontology-based approaches. *SIGMOD Record*, 33:2004. 29
- Noy, N. F. and McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. Technical report, Stanford University, Stanford, USA. Disponível em: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf, Acesso em: Abril, 2011. 37, 63, 64
- Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120. 21, 51
- Paiva, D. C., Mucheroni, M. L., and Netto, M. L. (2009). Um modelo para redes de conhecimento científico. *Workshop de Ontologias no STIL*. 49
- Posea, V. and Harzallah, M. (2004). Building a competence ontology. In *Workshop Enterprise Modeling and Ontology, International Conference on Practical Aspects of Knowledge Management (PAKM 04)*, Wien. 30
- Pujol, J. M., Sangüesa, R., and Delgado, J. (2002). Extracting reputation in multi agent systems by means of social network topology. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, AAMAS '02, pages 467–474, New York, USA. ACM. 9, 17, 20, 21
- Raggett, D., Hors, A. L., and Jacobs, I. (1999). Document type definition. Disponível em: <http://www.w3.org/TR/html4/sgml/dtd.html>, Acesso em: Novembro, 2011. 44
- Ralha, C. G., Guerra, G. N., Valentim, G. H. O., Sousa, M. S., Persegona, M. F. M., Godinho, M. L. S., Persegona, A. L. M., Blumenschein, R. N., Tome, M. V. D. F., and Carvalho, D. B. B. (2012). Observeunb - um portal web de rede social científica. *Anais do XXXIX SEMISH*, 1:1–12. 108
- Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Commun. ACM*, 43:45–48. 16, 21
- Reuters, T. (2012). Journal citation reports. Disponível em: http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/, Acesso em: Maio, 2012. 56

- Sabater, J. and Sierra, C. (2001). Social regret, a reputation model based on social relations. *SIGecom Exch.*, 3:44–56. 15, 18, 19
- Sabater, J. and Sierra, C. (2002). Reputation and social network analysis in multi-agent systems. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, AAMAS '02, pages 475–482, New York, USA. ACM. 8, 9
- Óscar Corcho, Fernández-López, M., Gómez-Pérez, A., and Óscar Vicente (2002). Webode: An integrated workbench for ontology representation, reasoning, and exchange. In *IN: PROCEEDINGS OF EKAW 2002. LNCS 2473*, pages 138–153. Springer. 38
- Singh, A. P., Shubhankar, K., and Pudi, V. (2011). An efficient algorithm for ranking research papers based on citation network. *Data Mining and Optimization (DMO), 2011 3rd Conference on*, pages 88–95. 51, 52
- Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y. (2007). Pellet: A practical owl-dl reasoner. *Web Semant.*, 5(2):51–53. 40
- Stanford (2011). Protégé. Disponível em: <http://protege.stanford.edu/>, Acesso em: Fevereiro, 2011. 38
- Stefaner, M. (2012). Relation browser. Disponível em: <http://moritz.stefaner.eu/projects/relation-browser/>, Acesso em: Abril, 2012. 79
- Su, X. and Ilebrikke, L. (2002). A comparative study of ontology languages and tools. In *Proceedings of the 14th International Conference on Advanced Information Systems Engineering*, CAiSE '02, pages 761–765, London, UK. Springer-Verlag. 30
- Sure, Y. and Studer, R. (2002). On-to-knowledge methodology. In Davies, J., Fensel, D., and van Harmelen (eds.), F., editors, *On-To-Knowledge: Semantic Web enabled Knowledge Management*, pages 33–46. Wiley. 37
- Tang, J., Zhang, J., Jin, R., Yang, Z., Cai, K., Zhang, L., and Su, Z. (2011a). Topic level expertise search over heterogeneous networks. *Mach. Learn.*, 82:211–237. 47
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. (2008). Arnetminer: extraction and mining of academic social networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 990–998, New York, USA. ACM. 47, 51
- Tang, W., Zhuang, H., and Tang, J. (2011b). Learning to infer social ties in large networks. In *Proceedings of the 2011 European conference on Machine learning and knowledge discovery in databases - Volume Part III*, ECML PKDD'11, pages 381–397, Berlin, Heidelberg. Springer-Verlag. 6
- Thiago Semprebom, M. Y. C. and Mendonça, I. (2007). Ontologias e protégé. Disponível em: [Disponível em: http://www.das.ufsc.br/~gb/pg-ia/Protege07/ontologias.pdf](http://www.das.ufsc.br/~gb/pg-ia/Protege07/ontologias.pdf), Acesso em: Maio, 2012. 27

- Tomaél, M. I., Alcará, A., and Chiara, I. D. (2006). Das redes sociais à inovação. *Ciência da Informação*, 34(2). 6
- TopQuadrant (2011). Topbraid composer. Disponível em: http://www.topquadrant.com/products/TB_Composer.html, Acesso em: Março, 2011. 38, 65
- Tsarkov, D. and Horrocks, I. (2006). Fact++ description logic reasoner: system description. In *Proceedings of the Third international joint conference on Automated Reasoning*, IJCAR'06, pages 292–297, Berlin, Heidelberg. Springer-Verlag. 39
- Vanz, S. A. (2009). As redes de colaboração científica no brasil. Master's thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, BRA. 11
- Vanz, S. A. and Stumpf, I. R. C. (2010). Colaboração científica: revisão teórico-conceitual. *Perspectivas em Ciência da Informação*, 15:42–55. 10
- W3C (2004). Rdfql - a query language for rdf. Disponível em: <http://www.w3.org/Submission/RDQL/>, Acesso em: Abril, 2012. 43
- W3C (2008). Sparql query language for rdf. Disponível em: <http://www.w3.org/TR/rdf-sparql-query/>, Acesso em: Abril, 2012. 43
- W3C (2009). Owl 2 web ontology language overview. Disponível em: <http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>, Acesso em: Abril, 2012. 36
- Wasserman, S. and Faust, K. (1999). *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York, USA. 6
- Yang, Y., Yeung, C. M. A., Weal, M. J., and Davis, H. (2009). The researcher social network: A social network based on metadata of scientific publications. In *Proceedings of WebSci'09: Society On-Line*. 48