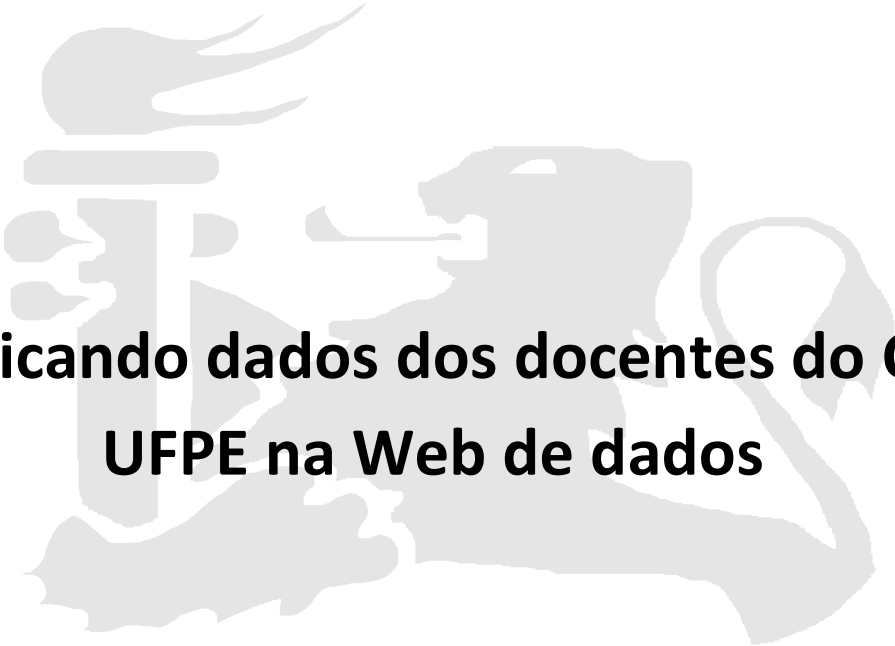


Universidade Federal de Pernambuco

GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

CENTRO DE INFORMÁTICA



Publicando dados dos docentes do CIn-UFPE na Web de dados

Aluno:	Gabriela Fernanda Silva de Vasconcelos	{gfsv@cin.ufpe.br}
Orientadora:	Dra. Bernadette Farias Lóscio	{bfl@cin.ufpe.br}
Co-orientadora:	Danusa Ribeiro Bezerra da Cunha	{drbc@cin.ufpe.br}

Agosto, 2012

Universidade Federal de Pernambuco

CENTRO DE INFORMÁTICA

Gabriela Fernanda Silva de Vasconcelos

Publicando dados do CIn-UFPE na Web de Dados

Trabalho apresentado à disciplina de Trabalho de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientadora: Profa. Dra. Bernadette Farias Lóscio

Co-orientadora: Danusa Ribeiro B. da Cunha

Recife, Agosto de 2012.

Dedico este trabalho aos meus pais e ao meu namorado.

Agradecimentos

Agradeço a paiinho, Vasconcelos, por todo seu esforço ao longo da vida.

À mainha, Socorrinho, que abdicou de sua vida em favor da educação dos seus filhos.

Aos meus oito irmãos: Viviane, Cristiane, Micheline, Patrícia, George, Fábio, Jaqueline e Gláucia que me serviram de exemplo para eu concluir a faculdade.

Ao meu namorado Papel por me fazer descobrir o amor. Tem um coração lindo e foi quem mais me ajudou e me incentivou durante todo esse tempo e ainda me proporcionou momentos maravilhosos.

Aos meus sobrinhos pelos momentos de descontração.

À Giovaninha e ao Heinho pela amizade.

Aos colegas de faculdades que tiraram as minhas dúvidas, me ajudavam nos projetos e tinha à paciência de ensinar assuntos que eu não entendia nada.

Na elaboração deste projeto, agradeço à professora Bernadette Lóscio que desde que paguei sua primeira cadeira já imaginava meu TG com ela independentemente do assunto. Sobre uma valiosa orientação, me deu a oportunidade de fazer algo para o CIn. À co-orientadora Danusa Cunha que me guiava na documentação deste projeto e me servia de incentivo quando a via estudar até tarde da noite. Ao Hélio Rodrigues, por ter me explicado e ter tirados as dúvidas do que eu deveria fazer. Ao Alberto Tavares, pessoa de extrema importância na conclusão do projeto e não mediu esforços para me ajudar.

A **Deus**, por ser a **luz** da minha vida.

...Não foi fácil, mas eu consegui.

Se chorei ou se sorri, o importante é que emoções eu vivi.

Roberto Carlos

Resumo

A Web Semântica é uma evolução da Web atual. Na Web Semântica, as informações nas páginas Web possuem um significado bem definido, isto é, os dados são estruturados em um formato padrão e estão relacionados a outros dados. No contexto da Web Semântica, vocabulários são ferramentas para organizar os dados de um domínio específico, enriquecendo esses dados adicionando significado e semântica. Isto permite que pessoas e agentes de software possam interpretá-los e trabalharem de forma mais eficaz com esses dados. Uma das tecnologias fundamentais para o funcionamento da Web Semântica é o RDF (Resource Description Framework), que trabalha com um trio de informação (sujeito, predicado e objeto) o qual expressa o significado das informações. Cada componente de uma tripla tem sua própria finalidade, e recebe uma identificação URI (Uniform Resource Identifier) única. O objetivo deste trabalho é criar um repositório com informações dos docentes do CIn, bem como suas publicações, utilizando o modelo Linked Data em um formato aberto e reutilizando algumas ontologias das mais conhecidas. E, em seguida, disponibilizar essas informações de uma forma que outros sistemas possam acessá-las.

Palavras-chave: Ontologia, *Linked Data*.

Abstract

The Semantic Web is an evolution of the current Web. In the Semantic Web, the information on the Web pages has a well-defined meaning, i.e., the data are structured in a standard format and are related to other data. In the context of Semantic Web vocabularies are tools to organize the data from a specific domain, enriching this data by adding meaning and semantics. This enables people and software agents can interpret them and to work more effectively with these data. One of the key technologies for the operation of the Semantic Web is RDF (Resource Description Framework), which works with a trio of information (subject, predicate and object) which expresses the meaning of the information. Each component of a triple has its own purpose, and receives a unique identifying URI (Uniform Resource Identifier). The objective of this work is to create a repository with information from the professors of the CIn and their publications, using the Linked Data model, in an open format ontologies and reusing some of the best known. And then make that information in a way that other systems can access them.

Keywords: Ontology, Linked Data.

Lista de Figuras

Figura 1. Arquitetura da Web Semântica.	19
Figura 2. Visão geral de conjuntos de dados publicados e seus relacionamentos em Setembro de 2011.	25
Figura 3. Interface da ferramenta Dia.	34
Figura 4. Interface do <i>Protégé</i>	35
Figura 5. Interface do SPARQL endpoint.	37
Figura 6. Esquema de Dados.	38
Figura 7. Ontologia do CInPub reutilizando outros vocabulários	41
Figura 8. Hierarquia de classes do CInPub.	44
Figura 9. Esquema da Ontologia do CInPub.	48
Figura 10. Terminal do Linux.	51
Figura 11. Linhas de comando para iniciar o Joseki.	51
Figura 12. Restrições da classe <i>cin:Area</i>	77
Figura 13. Instâncias da classe <i>cin:Area</i>	78
Figura 14. Restrições da classe <i>cin:EnderecoProfissional</i>	79
Figura 15. Instâncias da classe <i>cin:EnderecoProfissional</i>	79
Figura 16. Instâncias da classe <i>cin:LocalDaPublicacao</i>	80
Figura 17. Restrições da classe <i>cin:Colaborador</i>	81
Figura 18. Instâncias da classe <i>cin:Colaborador</i>	81
Figura 19. Restrições da classe <i>cin:Professor</i>	82
Figura 20. Instâncias da classe <i>cin:Professor</i>	82
Figura 21. Restrições da classe <i>cin:Publicacao</i>	83
Figura 22. Instâncias da classe <i>cin:Publicacao</i>	83

Lista de Tabelas

Tabela 1. Informações extraídas das fontes escolhidas.....	32
Tabela 2. Termos do vocabulário FOAF usados para a definição do CInPub	41
Tabela 3. Termos do vocabulário Dublin Core usados para a definição do CInPub	42
Tabela 4. Termos reusados do vocabulário AKT usados para a definição do CInPub...	42
Tabela 5. Termos e Relacionamentos que descrevem o domínio da Ontologia criada .	43
Tabela 6. Classes da ontologia do CInPub.	44
Tabela 7. Propriedades do tipo Datatype Property.	45
Tabela 8. Propriedades de relacionamento.....	45
Tabela 9. Propriedades inversas.	46
Tabela 10. Restrições das propriedades da classe <i>cin:Colaborador</i>	46
Tabela 11. Restrições das propriedades de relacionamento.	47
Tabela 12. Restrições das propriedades inversas.	48
Tabela 13. Prefixos utilizados nas consultas.	52
Tabela 14. Resultado da consulta 1.....	53
Tabela 15. Resultado da consulta 2.....	54
Tabela 16. Resultado da consulta 3.....	55
Tabela 17. Namespaces importados para o CInPub.....	65
Tabela 18. Ontologias importadas.	66
Tabela 19. Definição da propriedade <i>cin:temInteresse</i>	66
Tabela 20. Definição da propriedade <i>cin:ehInteresseDe</i>	67
Tabela 21. Definição da propriedade <i>cin:temEnderecoProfissional</i>	67
Tabela 22. Definição da propriedade <i>cin:ehEnderecoProfissionalDe</i>	67
Tabela 23. Definição da propriedade <i>cin:ehPublicadoEm</i>	68
Tabela 24. Definição da propriedade <i>cin:armazena</i>	68
Tabela 25. Definição da propriedade <i>cin:escreve</i>	68
Tabela 26. Definição da propriedade <i>cin:ehEscritaPor</i>	68
Tabela 27. Definição de outras propriedades da ontologia.	69
Tabela 28. Definição da classe <i>cin:LocalDaPublicacao</i>	70
Tabela 29. Definição da classe <i>cin:Area</i>	70
Tabela 30. Definição da classe <i>cin:EnderecoProfissional</i>	71
Tabela 31. Definição da classe <i>cin:Professor</i>	72
Tabela 32. Definição da classe <i>cin:Colaborador</i>	74
Tabela 33. Definição da classe <i>cin:Publicacao</i>	75
Tabela 34. Consulta 1.....	84
Tabela 35. Resultado da consulta 1.....	85
Tabela 36. Consulta 2.....	86
Tabela 37. Resultado da consulta2.	86
Tabela 38. Consulta 3.....	88
Tabela 39. Resultado da consulta 3.....	89
Tabela 40. Consulta 4.....	89

Tabela 41. Resultado da consulta 4.....	90
Tabela 42. Consulta 5.....	90
Tabela 43. Resultado da consulta5.	91
Tabela 44. Consulta 6.....	92
Tabela 45. Resultado da consulta 6.....	92
Tabela 46. Consulta 7.....	93
Tabela 47. Resultado da consulta 7.....	93
Tabela 48. Consulta 8.....	94
Tabela 49. Resultado da consulta 8.....	94
Tabela 50. Consulta 9.....	95
Tabela 51. Resultado da consulta 9.....	95
Tabela 52. Consulta 10.....	96
Tabela 53. Resultado da consulta 10.....	97

Sumário

Resumo	vi
Sumário	xi
Capítulo 1. Introdução.....	14
1.1 Motivação.....	14
1.2 Objetivos e Contribuições	16
1.3 Organização do Documento	17
Capítulo 2. Contextualização.....	18
2.1 Web Semântica.....	18
2.2. O modelo RDF (<i>Resource Description Framework</i>)	20
2.2.1. Links RDF	21
2.3 SPARQL (<i>SPARQL Protocol and RDF Query Language</i>)	22
2.4. Linked Data.....	23
2.4.1. Publicação de dados de fontes RDF como <i>Linked Data</i>	26
2.5.Vocabulários	27
2.6 Considerações Finais.....	30
Capítulo 3. Criação do CInPub.....	31
3.1 Visão geral do processo de criação do CInPub	31
3.2 Passo 1: Escolha das fontes de dados	32
3.3 Passo 2: Escolha das Ferramentas	33
3.3.1 Dia	33
3.3.2 Protégé.....	34
3.3.3 Jena TDB	36
3.3.4 Joseki	36
3.4 Passo 3: Definição do esquema do conjunto de dados	38
3.5 Passo 4: Criação da Ontologia	39
3.5.1 Domínio e Escopo da Ontologia	40
3.5.2 Reutilização de Vocabulários existentes	41
3.5.3 Definição dos Termos	42
3.5.4 Definição das Classes e Hierarquia de Classes.....	43
3.5.3 Definição das Propriedades das Classes.....	44
3.5.4 Definição das Restrições	46
3.6 Passo 5: Criação das Instâncias	49

3.7 Considerações Finais.....	49
Capítulo 4. Realizando consultas no CInPub	50
4.1 Recuperação dos dados do CInPub	50
4.1.1 Configuração do Jena TDB para Realização das Consultas.....	50
4.1.2 Realizando consultas SPARQL no CInPub.....	52
4.2 Considerações Finais.....	56
Capítulo 5. Considerações Finais	57
5.1 Contribuições	57
5.2 Dificuldades Encontradas	57
5.3 Trabalhos Futuros.....	58
6. Referências Bibliográficas.....	60
APÊNDICE A – Códigos em RDF.....	65
Namespaces e vocabulários importados	65
Definição das propriedades relacionais	66
Definição das classes.....	69
Classe <i>cin:LocalDaPublicacao</i>	70
Classe <i>cin:Area</i>	70
Classe <i>cin:EnderecoProfissional</i>	71
Classe <i>cin:Professor</i>	71
Classe <i>cin:Colaborador</i>	73
Classe <i>cin:Publicacao</i>	74
APÊNDICE B – Visão da ontologia no <i>Protégé</i>	77
Visualizações das classes com suas propriedades e restrições e exemplos de instâncias.....	77
Classe <i>cin:Area</i>	77
Classe <i>cin:EnderecoProfissional</i>	78
Classe <i>cin:LocalDaPublicacao</i>	80
Classe <i>cin:Colaborador</i>	80
Classe <i>cin:Professor</i>	81
Classe <i>cin:Publicacao</i>	83
APÊNDICE C - Consultas em SPARQL sobre o dataset CInPub.....	84
Consulta 1	84
Consulta 2	85
Consulta 3	88

Consulta 4 89

Consulta 5 90

Consulta 6 92

Consulta 7 92

Consulta 8 93

Consulta 9 94

Consulta 10 95

Capítulo 1. Introdução

Neste capítulo será feita uma breve introdução do trabalho, mostrando qual foi a motivação para o seu desenvolvimento, quais as suas contribuições e como o documento está estruturado.

1.1 Motivação

No final da década de 80, surgiu a *World Wide Web* ou Web, criada por Tim Berners-Lee. A Web se tornou um local virtual onde é armazenada e apresentada uma grande quantidade de dados que estão interligadas e que podem ser compartilhados e atualizados de uma maneira fácil por milhões de usuários espalhados em todo o mundo.

Passados mais de vinte anos após seu surgimento, uma grande quantidade de dados tem sido disponibilizada na Web, incluindo textos, áudio, imagem e vídeo. Contudo, esses dados são disseminados de forma autônoma e heterogênea. Além disso, os dados ainda são organizados para serem lidos ou compreendidos por humanos e não por agentes de software. Para que um agente de software possa entender e interpretar um dado, é necessário processar a semântica envolvida naquele dado, num determinado contexto. Semântica significa atribuir significado a elementos, dados ou expressões que precisem ser interpretadas em uma determinada situação [Souza *et al* 2008].

Em 2001, Tim Berners-Lee propôs a Web Semântica [Lee *et al* 2001], uma extensão da Web atual, onde os documentos são auto-descritíveis, permitindo a compreensão do seu conteúdo pelos agentes de software. A inclusão da semântica, além de facilitar a compreensão dos documentos pelos agentes, torna possível raciocinar e fazer inferências sobre o conteúdo de tais documentos. Dessa forma, os agentes tornam-se aptos a auxiliar as pessoas em diferentes tarefas de recuperação de informações que exijam tomada de decisões e inferência de conclusões a partir de informações não explicitamente disponíveis ou de informações contextuais [Marcondes & Campos 2008].

A ideia por trás da Web Semântica consiste em associar significado aos recursos da Web por meio da utilização de metadados que serão processados pelas máquinas. Cada recurso pode estar relacionado a outros recursos e pode ter um conjunto de metadados associados. No contexto da Web Semântica, um recurso pode ser qualquer entidade do mundo real, por exemplo, uma pessoa, uma localidade ou um evento. Entretanto, para que a Web Semântica torne-se uma realidade, é necessário que haja uma padronização de tecnologias. Neste contexto, o W3C¹ (*World Wide Web Consortium*) e a comunidade acadêmica têm unido esforços no sentido de criar, disponibilizar e padronizar linguagens e tecnologias para garantir a correta e melhor interoperabilidade possível [Filho & Lóscio 2010].

Como parte desse esforço, surgiu o conceito de *Linked Data* que pode ser definido como um conjunto de boas práticas para publicar e conectar conjuntos de dados estruturados na Web, com o intuito de criar uma Web de Dados [Bizer *et al* 2009]. Estas práticas são fundamentadas em tecnologias Web, como HTTP (*Hypertext Transfer Protocol*) e URI (*Uniform Resource Identifier*), com o objetivo de permitir a leitura dos dados conectados semanticamente, de forma automática, por agentes de software. A Web de Dados cria inúmeras oportunidades para a integração semântica de dados, motivando o desenvolvimento de novos tipos de aplicações e ferramentas, como navegadores e motores de busca.

Entretanto, para que a Web de Dados torne-se uma realidade é importante que determinados padrões sejam obedecidos a fim de garantir a correta interoperabilidade e compartilhamento de dados. Dentre estes padrões, destacam-se as ontologias e os vocabulários, os quais são definidos para domínios específicos. Uma ontologia representa uma descrição formal para o conjunto de conceitos do domínio, enquanto que o vocabulário são os conceitos propriamente ditos ou os termos usados para descrevê-los. Atualmente, existem diversas ontologias que descrevem o vocabulário de domínios específicos, como por exemplo: o FOAF² para descrever perfis pessoais e seus relacionamentos; *Dublin Core*³, que visa descrever objetos digitais; AKT⁴, que descreve a gestão de conhecimentos científicos; *BibTex*⁵, utilizados para descrever e processar

¹<http://www.w3c.br/Home/WebHome>

²<http://www.foaf-project.org/>

³<http://dublincore.org/>

⁴<http://www.aktors.org/ontology/portal#>

⁵<http://oaei.ontologymatching.org/2004/Contest/301/onto.html>

listas de referências; *Basic Geo*⁶, para informações de geolocalização, *Music Ontology*⁷, para descrição do domínio da música; *NeoGeo*⁸, ontologia dos elementos geométricos.

Atualmente, existe um grande número de conjuntos de dados disponíveis na Web que seguem os padrões de *Linked Data*. Estes conjuntos estão organizados em domínios incluindo dados sobre localizações geográficas, empresas, livros, publicações científicas, filmes, música, televisão e rádio, ensaios clínicos, comunidades online, dados estatísticos entre outros [Heath & Bizer 2011]. Para garantir o compartilhamento e futura integração de dados, vocabulários padrões já pré-definidos devem ser considerados no momento da criação dos conjuntos de dados a serem publicados na Web de dados. O uso de vocabulários comuns facilitará a identificação e busca de conceitos similares distribuídos em bases de dados distintas.

1.2 Objetivos e Contribuições

Seguindo a forte tendência de publicação de dados interligados, este trabalho tem como principal objetivo a criação de um conjunto de dados, denominado CInPub, com informações sobre os professores do Centro de Informática (CIn) da UFPE. A criação do CInPub foi norteada pelos princípios de *Linked Data* e pelo uso de vocabulários pré-definidos, como *FOAF*, *DC* e *AKT*.

Como principais contribuições do trabalho destacam-se:

- Criação de um conjunto de dados (*dataset*), seguindo os princípios *Linked Data*;
- A criação de uma ontologia baseada nos vocabulários pré-definidos para descrever formalmente o domínio de informações pessoais e bibliográficas dos professores do CIn/UFPE;
- Disponibilização de um SPARQL *Endpoint* para a definição de consultas e apresentação dos resultados sobre o conjunto de dados criados.

⁶<http://www.w3.org/2003/01/geo/>

⁷<http://musicontology.com/>

⁸<http://geovocab.org/spatial.html>

1.3 Organização do Documento

O restante deste documento está estruturado da seguinte forma:

O Capítulo 2 contém a fundamentação teórica, apresentando uma revisão sobre os tópicos relevantes para o desenvolvimento deste projeto.

O Capítulo 3 apresenta a metodologia adotada para a criação do CInPub. Dentro do contexto deste capítulo, é explanado o processo de geração e caracterização do conjunto de dados que será criado.

O Capítulo 4 descreve o conjunto de dados criado, discorrendo sobre suas principais características, aspectos de desenvolvimentos e resultados experimentais.

Finalmente, o Capítulo 5 expõe as considerações finais, além de descrever os conhecimentos adquiridos no desenvolvimento do trabalho.

Capítulo 2. Contextualização

Neste capítulo serão abordados alguns conceitos básicos e de fundamental importância para o entendimento deste trabalho. Na seção 2.1, apresentamos a Web Semântica e suas principais características. Na seção 2.2, falamos brevemente sobre o modelo RDF bem como definimos sobre os *Links* RDF essenciais para interligar os conjuntos de dados na Web de Dados. Na seção 2.3, introduzimos o conceito de *Linked Data* e os principais vocabulários e/ou ontologias utilizadas para a criação de conjuntos de dados seguindo os princípios de publicação *Linked Data*. Por fim, na seção 2.4 serão descritos alguns dos vocabulários existentes e utilizados na criação da ontologia deste projeto.

2.1 Web Semântica

Visando acrescentar significado às informações presentes na Web e proporcionar a interoperabilidade entre aplicações, Tim Berners-Lee apresentou uma nova proposta para a Web, chamada de Web Semântica [Melo & Junior 2010]. A Web Semântica é uma extensão da Web tradicional, que especifica um modelo universal para a estruturação, definição e representação da informação, acrescentando-lhe semântica. Além disso, a Web Semântica busca fornecer aos mecanismos de busca a possibilidade de usar regras de inferência com o objetivo de obter uma maior eficiência ao se recuperar informações na Web [Raymond, 2001]. Na Web Semântica os dados poderão ser relacionados, compartilhados, analisados, interpretados, processados e manipulados pelas aplicações, tornando possível a filtragem automática da informação e a execução de tarefas complexas de acordo com a interpretação das informações contidas na rede [Melo & Junior 2010].

Para permitir a interpretação das informações contidas na Web Semântica e proporcionar a interoperabilidade entre aplicações, o W3C está trabalhando na especificação de um conjunto de tecnologias, ferramentas e padrões para a identificação de recursos, representação sintática, estrutural, semântica e lógica de informações referentes a esses recursos [Melo & Junior 2010, Freitas 2011]. A arquitetura da Web

Semântica já passou por várias modificações e a sua configuração atual é ilustrada na Figura 1

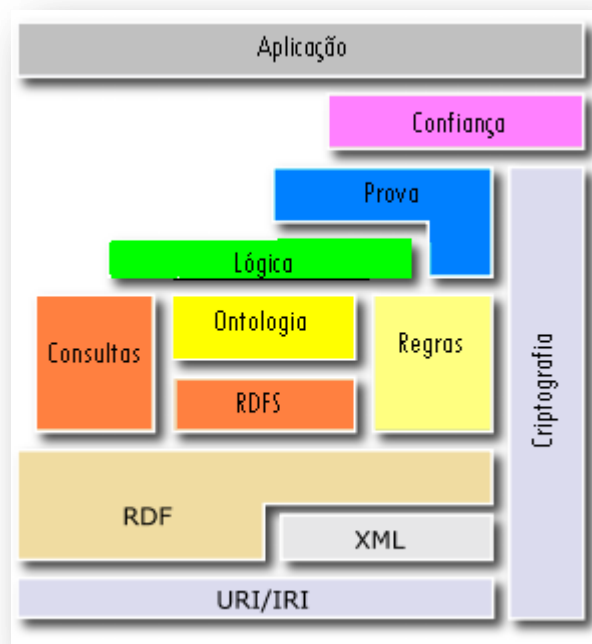


Figura 1. Arquitetura da Web Semântica.

A arquitetura mencionada é separada em camadas, onde cada uma delas possui uma determinada função. Outro fator interessante é a alta interoperabilidade entre as mesmas. Abaixo, uma breve explicação das funcionalidades de cada uma destas camadas [Filho & Lóscio 2010]:

- ✓ **Camada *URI/IRI*:** Permite identificar unicamente cada recurso (ex: figuras e páginas HTML) disponível na Web por meio de um identificador único (*Universal Resource Identifier*);
- ✓ **Camada *XML*:** Permite a criação de marcações para descrição de informações. A sintaxe das camadas superiores é baseada em XML;
- ✓ **Camada *RDF*:** Provê um modelo de descrição lógica de dados permitindo descrever assertivas e informações sobre um determinado recurso;
- ✓ **Camada *RDFS*:** Permite a criação de um vocabulário para a camada RDF. Com o RDFS, é possível criar hierarquia de classes e propriedades;

- ✓ **Camada de Ontologia:** Estende a camada RDFS, provendo um maior nível de expressividade para a definição da semântica das informações;
- ✓ **Camada de Consulta:** Responsável por prover meios para a realização de consultas sobre o modelo de dados RDFS/Ontologia;
- ✓ **Camada de Regras:** Provê um mecanismo de criação de relações entre recursos que não podem ser descritas diretamente na ontologia;
- ✓ **Camada Lógica:** É responsável pelo raciocínio e execução de inferências lógicas a partir da semântica previamente descrita;
- ✓ **Camada de Prova e Confiança:** Camada que provê um mecanismo para avaliar o nível de confiabilidade das fontes de recursos e informações;
- ✓ **Camada de Aplicação:** Camada que permite a interação entre o usuário e a aplicação Web Semântica.

2.2. O modelo RDF (*Resource Description Framework*)

A utilização de um modelo padrão para representação de dados, como o RDF, torna possível a implementação de aplicações genéricas capazes de operar sobre o espaço de dados global [Heath & Bizer 2011]. O modelo RDF é baseado no conceito de grafo, e caracteriza-se por possuir um caráter extensível e um alto nível de expressividade, facilitando, dessa forma, a interligação entre dados de diferentes fontes.

Em RDF, um recurso pode estar relacionado com dados ou com outros recursos por meio de sentenças, as quais são estruturadas no formato sujeito + predicado + objeto, onde:

- ✓ **Sujeito:** Tem como valor o recurso sobre o qual se quer escrever uma sentença. Todo recurso deve ser capaz de ser identificado unicamente.
- ✓ **Predicado:** Especifica um relacionamento entre um sujeito e um objeto. O predicado é especificado por meio de propriedades, que são relações binárias geralmente nomeadas por um verbo e permitem relacionar um recurso a dados ou a outros recursos. Uma propriedade também é um recurso e, portanto, deve ter um identificador único.

- ✓ **Objeto:** Denomina o recurso ou dado que se relaciona com o sujeito. O valor de um objeto pode ser um recurso ou um literal, que pode ser um valor numérico ou uma cadeia de caracteres.

Cada tripla faz parte da Web de Dados e pode ser usada como ponto de partida para explorar esse espaço de dados. Triplas de diferentes fontes podem ser facilmente combinadas para formar um único grafo. Além disso, é possível usar termos de diferentes vocabulários para representar os dados. O modelo RDF ainda permite a representação de dados em diferentes níveis de estruturação, sendo possível representar desde dados semi-estruturados a dados altamente estruturados.

2.2.1. Links RDF

No contexto de *Linked Data* os *links* RDF descrevem relacionamentos entre dois recursos [Heath & Bizer 2011]. Um *link* RDF consiste de três referências URI. As URIs referentes ao sujeito e objeto identificam os recursos relacionados, enquanto que a URI referente ao predicado define o tipo de relacionamento entre os recursos. Uma distinção útil que pode ser feita é com relação aos *links* internos e externos. *Links* RDF internos conectam recursos dentro de uma única fonte de dados *Linked Data*. *Links* externos conectam recursos que são servidos por diferentes fontes de dados *Linked Data*. No caso de *links* externos, as URIs referentes ao sujeito e ao predicado do *link* pertencem a diferentes *namespaces*. *Links* externos são cruciais para a Web dos Dados, visto que eles permitem interligar as fontes de dados dispersas em um espaço global de dados.

O armazenamento de dados no modelo RDF pode ser realizado por meio de grafos em memória, arquivo texto ou banco de dados específico para armazenamento de triplas RDF, chamado de *RDF Triple Store*. O armazenamento de triplas em arquivo texto usa algum formato de serialização de RDF *Notation3* ou *N3*⁹, *N-Triple*¹⁰, *Turtle*¹¹, *RDF/XML*¹² e *RDF/JSON*¹³, onde cada uma atende a um propósito específico. Vale

⁹<http://www.w3.org/TeamSubmission/n3/>

¹⁰<http://www.w3.org/2001/sw/RDFCore/ntriples/>

¹¹<http://www.w3.org/TeamSubmission/turtle/>

¹²<http://www.w3.org/TR/REC-rdf-syntax/>

¹³<http://www.w3.org/wiki/JSON+RDF>

ressaltar que há formas de conversão entre notações. A notação *N-Triple* torna as triplas mais explícitas, entretanto, esta notação possui como desvantagem a repetição de URIs. As notações *N3* e *Turtle* são muito semelhantes e possuem foco na abreviação do código. A notação *RDF/XML* descreve as triplas em XML e *RDF/JSON* é abreviada e mais fácil de ser interpretada por sistemas Web que a *Turtle* e *N3* [Covelinhas 2008].

2.3 SPARQL (*SPARQL Protocol and RDF Query Language*)

É a linguagem declarativa recomendada pelo W3C para recuperação de informações em documentos RDF/RDFS. Em suma permite consultar dados de grafos RDF com base em correspondência de padrões de tripla, que é o bloco de construção básica [Pinheiros 2011].

Alguns comandos disponíveis em SPARQL e citados em [Almeida 2009] são:

- ✓ **SELECT:** Usada para especificar a forma final em que os resultados são retornados para o usuário [Salgado & Lóscio 2012]. As variáveis começam sempre por \$ ou ?, não sendo, no entanto, parte das mesmas.
- ✓ **FROM:** É opcional e define uma URI para o conjunto de dados a serem interrogados, ou seja, especifica as fontes ou dataset a serem consultados.
- ✓ **WHERE:** Fornece um padrão de grafo de base, para fazer a correspondência com os dados do grafo. Um padrão de grafo é composto por um conjunto de padrões de triplas.
- ✓ **PREFIX:** Permite a associação de um label a uma URI, de forma a simplificar a escrita da consulta, usando-se o label definido em vez da URI, sempre que necessário;

- ✓ **CONSTRUCT:** Permite a construção de grafos em RDF e a criação de mapeamentos entre os esquemas de diferentes ontologias. No entanto a sua expressividade ainda é limitada, como no que diz respeito à concatenação (“Concat”) e divisão (“Split”) de conceitos entre ontologias.
- ✓ **FILTER:** Utilizado para restringir as soluções encontradas, segundo os valores dos atributos das triplas, geralmente o sujeito ou o objeto, em uma consulta SPARQL..

A linguagem SPARQL inclui um protocolo de acesso para RDF, em que consultas SPARQL podem ser executadas com chamadas SOAP (Simple Object Access Protocol) ou HTTP. O objetivo do protocolo SPARQL é promover a interoperabilidade, em que os clientes podem interagir com os dados SPARQL de forma consistente. Isso inclui a serialização dos resultados por meio de streams HTTP, desse modo, possibilitando uniformizar todo acesso às fontes de dados independentemente da tecnologia utilizada, pois só é preciso executar os métodos HTTP GET ou POST [Pinheiros 2011].

2.4. Linked Data

Linked Data refere-se ao conjunto de melhores práticas para a publicação de dados estruturados na Web. Essas práticas foram introduzidas por Tim Berners-Lee em [Lee *et al* 2006] e resumem-se em quatro princípios básicos:

1. Usar URIs como nome para recursos. Identificar não apenas documentos e conteúdo digital, mas também objetos do mundo real e conceitos abstratos.
2. Usar URIs HTTP para que as pessoas possam encontrar esses nomes.
3. Quando alguém procurar por uma URI, garantir que informações úteis possam ser obtidas por meio dessa URI, as quais devem estar representadas no formato RDF.
4. Incluir *links* para outras URIs de forma que outros recursos internos ou externos possam ser descobertos.

O primeiro princípio defende o uso de referências URI para identificar, não apenas documentos Web e conteúdos digitais, mas também objetos do mundo real e conceitos abstratos [Cunha *et al* 2011].

O segundo princípio defende o uso de *URIs* HTTP para identificar os objetos e os conceitos abstratos definidos pelo Princípio 1, possibilitando que essas URIs sejam dereferenciáveis sobre um protocolo HTTP. Neste contexto, dereferenciar é o processo de recuperar uma representação de um recurso identificado por uma URI, onde um recurso pode ter várias representações como documentos HTML, RDF, XML entre outros.

A fim de permitir que uma ampla gama de aplicações diferentes possa processar dados disponíveis na Web, é importante que exista um acordo sobre o formato padrão para disponibilização dos dados. O terceiro princípio *Linked Data* defende o uso de RDF como modelo para a publicação de dados estruturados na Web [Klyne *et al* 2004]. Com o RDF, é possível descrever significado sobre recursos, habilitando agentes de software a explorar os dados de forma automática, muitas vezes, agregando, interpretando ou mesclando dados.

O quarto princípio diz respeito ao uso de hiperlinks para conectar não apenas os documentos da Web, mas qualquer tipo de recurso. Por exemplo, uma ligação pode ser fixada entre uma pessoa e um lugar, ou entre um local e uma empresa. Em contraste com a Web clássica onde os hiperlinks são em grande parte não tipados, hiperlinks que conectam os recursos em um contexto de *Linked Data* são capazes de descrever a relação entre eles. Hiperlinks no contexto de *Linked Data* são chamados de *links* RDF, a fim de distingui-los dos hiperlinks existentes na Web convencional [Heath & Bizer 2011].

O exemplo mais visível da adoção e aplicação dos princípios *Linked Data* tem sido o projeto *Linking Open Data*¹⁴ fundado em janeiro de 2007 e apoiado pelo *W3C Semantic Web Education and Outreach Group*¹⁵. O objetivo principal deste projeto é identificar conjuntos de dados disponíveis sob licenças abertas e convertê-los para RDF de acordo com os princípios *Linked Data*. Na Figura 2 podem-se observar os diversos conjuntos de dados publicados no contexto do projeto *Linking Open Data*.

¹⁴<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

¹⁵<http://www.w3.org/2001/sw/sweo/>

No grafo da Figura 2, cada nó representa um conjunto de dados publicado seguindo os princípios *Linked Data*, os quais estão interligados com outros conjuntos de dados na nuvem. O tamanho de cada nó corresponde ao número de triplas RDF do conjunto de dados. As setas indicam a existência de pelo menos 50 ligações entre dois conjuntos, podendo ser unidirecionais, indicando que um certo conjunto contém triplas RDF de um outro conjunto, ou bidirecionais, indicando que ambos os conjuntos contêm triplas RDF um do outro [Cunha *et al* 2011].

O conteúdo da nuvem possui natureza diversa, incluindo dados sobre localizações geográficas, empresas, livros, publicações científicas, filmes, música, televisão e rádio, ensaios clínicos, comunidades online, dados estatísticos entre outros [Heath & Bizer 2011].

2.4.1. Publicação de dados de fontes RDF como *Linked Data*

As fontes de dados que adotam o modelo RDF são normalmente armazenadas em arquivos ou em um repositório de dados RDF (*RDF Store*). A seleção da forma de armazenamento mais adequada vai depender de fatores como o volume de dados utilizado e a frequência de atualização desses dados. Nesta subseção veremos como publicar arquivos RDF ou *RDF Stores* como *Linked Data*.

2.4.1.1. Publicação de arquivos RDF como *Linked Data*

Criar arquivos RDF estáticos e disponibilizá-los através de um servidor Web é a forma mais simples de publicar dados *Linked Data*. Essa estratégia somente é viável quando o arquivo RDF é relativamente pequeno. Embora existam vários formatos para serialização de um arquivo RDF (*RDF/XML*, *Turtle*, *Notation3*, *NTriples*), o formato mais antigo e também mais usado para publicação de *Linked Data* é o *RDF/XML*. Assim, se o arquivo for escrito manualmente em um formato mais simples, compacto e legível para o ser humano como *Turtle*, deverá ser posteriormente convertido para *RDF/XML*. A conversão de *Turtle* para *RDF/XML* pode ser realizada através de algum conversor online como *Babel*¹⁶ ou *RDF Converter*¹⁷.

Dados RDF também podem vir embutidos em documentos Web. *RDFa* [Adida & Birbeck 2008] permite a inclusão de dados RDF em um documento XHTML, aproximando a Web de Documentos da Web de Dados. Desse modo, textos e links legíveis ao ser humano também podem coexistir com dados relacionados que poderão ser facilmente processados pelas máquinas.

Até aqui falamos de arquivos RDF com conteúdo estático. No entanto, os arquivos RDF ou *RDFa* podem ter seus conteúdos gerados e preenchidos dinamicamente através de uma aplicação Web.

¹⁶<http://simile.mit.edu/babel>

¹⁷<http://www.mindswap.org/2002/rdfconvert>

2.4.1.2. Publicação de dados de RDF Store como *Linked Data*

A publicação de dados de uma RDF Store como *Linked Data* tipicamente envolve a disponibilização de uma interface *Linked Data* e de um SPARQL Endpoint para acesso aos dados. O ideal seria que toda RDF Store já tivesse esses recursos integrados para que somente fosse necessário especificar que dados seriam publicados como *Linked Data*. Entretanto em muitos casos há necessidade de se instalar serviços adicionais para prover tais recursos. Se a RDF Store não fornecer um SPARQL Endpoint embutido, este serviço pode ser instalado para permitir a execução de consultas SPARQL sobre a RDF Store. Um SPARQL Endpoint é um serviço de consulta baseado no protocolo HTTP que adere ao protocolo SPARQL, em que consultas a uma base de conhecimento podem ser enviadas ao SPARQL Endpoint e respondidas apropriadamente, além de poder enviar instruções SPARQL de atualização como *insert*, *update* e *delete* [Pinheiros 2011]. Um servidor *Joseki* ou *Fuseki* pode ser usado como SPARQL Endpoint de uma RDF Store. O *Joseki* provê suporte nativo para acesso às RDF Stores Jena TDB¹⁸ e Jena SDB¹⁹. O *Fuseki* já inclui a RDF Store Jena TDB.

2.5. Vocabulários

Na criação de conjuntos de dados seguindo os princípios *Linked Data* é considerada uma boa prática o reuso de termos de vocabulários amplamente conhecidos para facilitar tanto o processamento quanto a integração dos dados. É importante levar em consideração que novos termos devem ser definidos se não forem encontrados em vocabulários já existentes.

A reutilização de vocabulários pode reduzir o retrabalho, pois uma boa parte de termos, significados, propriedades e relações já foram previamente modelados, além de contribuir para a interoperabilidade entre as aplicações. No decorrer desta seção são apresentados alguns dos principais vocabulários utilizados na publicação de dados interligados e que estão relacionados com a proposta deste trabalho. Dentre eles

¹⁸<http://openjena.org/TDB/>

¹⁹<http://openjena.org/SDB/>

destacam-se: *Friend-of-a-Friend* (FOAF), *Semantically-Interlinked Online Communities*²⁰ (SIOC), *Simple Knowledge Organization System*²¹ (SKOS), *Description of a Project*²² (DOAP), *Creative Commons*²³ (CC) e *Dublin Core*²⁴ (DC).

- ✓ **FOAF:** O FOAF é um vocabulário usado para a construção de redes de relacionamento, facilitando a criação de páginas Web que contenham informações pessoais e sua rede de relacionamentos on-line [Bulcao 2006]. Aplicações podem utilizar os termos deste vocabulário para expressar informações sobre o perfil de contato de pessoas, bem como para fazer inferências sobre suas relações sociais com outras pessoas, com base na sua proximidade/vizinhança [Vanni 2009].

Os documentos FOAF são escritos nas linguagens OWL e RDF, podendo ser combinado com outros documentos de mesma natureza para criar uma base de dados de informação unificada [Brickley & Miller 2010].

- ✓ **Dublin Core:** A pesquisa de informação na Internet é dificultada pelas diferenças entre padrões utilizados. O *Dublin Core* pode simplificar esse processo, pois é um padrão com iniciativa global que promove a fácil acessibilidade e visibilidade de recursos [Dublin Core 2012]. De acordo com [Yamaoka 2007], as principais características do Dublin Core são:

- Simplicidade na criação e manutenção dos termos, permitindo o seu uso por não especialistas;
- Uma semântica com entendimento universal que facilita a interpretação de usuários com diferentes formações;
- Extensibilidade que permite a adição de elementos para atender as especificidades de diferentes comunidades.

Devido a uma crescente necessidade de estabelecimento de padrões de descrição de conteúdo de recursos de informação e a geração de metadados [Alves &

²⁰<http://sioc-project.org/>

²¹<http://www.w3.org/2009/08/skos-reference/skos.html>

²²<http://trac.usefulinc.com/doap>

²³<http://creativecommons.org/>

²⁴<http://dublincore.org/>

Souza 2007], *Dublin Core* vem sendo adotado internacionalmente por diversas instituições envolvidas na geração, recuperação e uso de recursos eletrônicos. Os documentos no padrão DC são considerados como objetos e integram todos os possíveis tipos de materiais e seu tratamento técnico, sejam bases de dados, imagens digitais, banco de imagens, textos eletrônicos, vídeos e filmes em formato digital e não-digital, objeto multimídia, gravações sonoras em formato digital e não-digital [Alves & Souza 2007], porém, não cobre as necessidades de descrição de todos os tipos de recursos na Web.

- ✓ **AKT:** O *AKT Reference Ontology* (AKT) é um vocabulário para publicações científicas que foi desenvolvido durante um período de seis meses por vários parceiros do projeto AKT. Esse projeto é uma colaboração de pesquisas interdisciplinares (IRCs) financiadas pela *Engineering and Physical Sciences Research Council* (EPSRC) para ajudar a identificar e resolver os problemas de TI do futuro. É um vocabulário que permite modelar o domínio de dados bibliográficos coecientíficos contendo representações para pessoas, conferências, projetos, organizações, publicações, etc. [Alani & Harith 2005].

- ✓ **SKOS:** *Simple Knowledge Organization System* (SKOS) é um modelo simples de dados que fornece um padrão de migração de bases de conhecimento para a Web Semântica, com a ajuda de uma linguagem própria para a definição de conteúdos de modelos conceituais diversos de forma compreensível pelas máquinas [W3C 2009].

Este vocabulário permite expressar a estrutura básica e o conteúdo de esquemas conceituais, tais como tesouros, esquema de classificação, lista de cabeçalhos de assunto, taxonomia e outros tipos de vocabulários controlados e esquemas conceituais embutidos em glossários e terminologia [Miles & Brickley 2005].

O SKOS ainda está em fase de desenvolvimento e compreende um conjunto de propriedades descritas em RDF, que podem ser utilizadas para expressar o conteúdo de um documento RDF, possibilitando descrever formalmente os termos e relacionamentos existentes em um tesouro [Ramalho *et al* 2007].

2.6 Considerações Finais

Neste capítulo vimos os principais conceitos e ferramentas que dão suporte a Web Sêmantica: Linked Data, RDF, Links RDF e vocabulários para descrever formalmente os conceitos de um domínio para os dados presentes na Web estejam estruturadas de modo padronizado de forma que possam ser interpretados também por agentes de software. No próximo capítulo iremos falar sobre a criação do CInPub e a metodologia utilizada para a criação da ontologia do projeto.

Capítulo 3. Criação do CInPub

Este capítulo apresenta o processo de construção da ontologia para descrição dos termos utilizados na representação das informações sobre os docentes do CIn e suas respectivas publicações. A partir desta ontologia foi criado um repositório de dados, denominado de CInPub, o qual permitirá recuperar informações sobre os docentes do CIn e estará acessível a partir de um SPARQL *endpoint*.

3.1 Visão geral do processo de criação do CInPub

Nesta seção apresentamos os passos para a criação do CInPub, são eles:

- ✓ **Passo 1: Escolha das fontes de dados** – definimos quais os locais poderiam fornecer informações relevantes sobre os professores e suas publicações. A partir dessas informações foram escolhidos os conceitos que deveriam fazer parte da ontologia.
- ✓ **Passo 2: Escolha das ferramentas** – no decorrer do projeto foi necessário utilizar algumas ferramentas para ajudar na criação do repositório. Nesta etapa foram escolhidas tais ferramentas.
- ✓ **Passo 3: Definição do esquema do conjunto de dados** – antes de criar a ontologia optamos por criar um modelo conceitual e registrá-lo na forma de um esquema para ser a base para a construção da ontologia.
- ✓ **Passo 4: Criação da ontologia** – com os passos definidos anteriormente, foi criada a ontologia utilizando como base uma metodologia simples e interativa chamada de Metodologia 101 [Noy & McGuinness 2001].
- ✓ **Passo 5: Criação das instâncias** – com o esquema da ontologia concluído, dá-se início ao processo da criação das instâncias, preenchendo os valores dos atributos de cada classe da ontologia, povoando o *dataset* do CInPub.

3.2 Passo 1: Escolha das fontes de dados

Como fontes de dados para a criação do CInPub foram utilizados o site do Centro de Informática (CIn)²⁵ e o currículo lattes do CNPq²⁶. As informações extraídas de cada uma dessas fontes estão apresentadas na Tabela 1.

Tabela 1. Informações extraídas das fontes escolhidas.

Informações extraídas do site do CIn	Informações extraídas do CV Lattes
✓ Nome; ✓ Cargo; ✓ Áreas de interesse; ✓ Email; ✓ Telefone; ✓ Fax; ✓ Pagina pessoal; ✓ Foto; ✓ Sala de trabalho na instituição; ✓ Endereço eletrônico do Currículo Lattes.	✓ Nome em citação bibliográfica; ✓ Sexo; ✓ Endereço profissional; ✓ Referências Bibliográficas das publicações; ✓ Local da publicação; ✓ Ano de publicação;

Além destas fontes, as páginas pessoais de alguns dos professores também serviram de fonte para coleta de dados, bem como os sites que disponibilizam dados bibliográficos sobre eles. A partir destas fontes, foi possível extrair outros dados como por exemplo, nome dos co-autores e a URL de artigos. Vale ressaltar que as informações coletadas destas fontes estão no formato HTML e foram extraídas manualmente.

²⁵<http://www2.cin.ufpe.br/site/index.php>

²⁶<http://lattes.cnpq.br/>

3.3 Passo 2: Escolha das Ferramentas

Nesta seção serão apresentadas as ferramentas necessárias para realização deste projeto.

3.3.1 Dia

Foi utilizado o software Dia²⁷ versão 0.97.2 para a elaboração do diagrama de classes. Este programa auxilia na elaboração de fluxogramas, organogramas e diagramas, contendo também objetos para a modelagem UML. Ele permite a criação de diagramas que podem ser exportados para os seguintes formatos: EPS, SVG, XFIG, WMF, PNG e XML. Um programa simples e fácil de trabalhar, e possibilita bons resultados nas diagramações. Este programa auxilia os analistas e desenvolvedores de sistemas na criação e integração dos diagramas de dados da UML com a lógica. Com ele é possível especificar recursos, transações, trocas de comunicação, plano de custos com tempo, esforço, entre outras. Além disso, além de modelagem de negócios voltada para informática, também é possível montar diversos tipos de diagramas. A interface, desta ferramenta é ilustrada na Figura 3.

²⁷<http://dia-installer.de/>

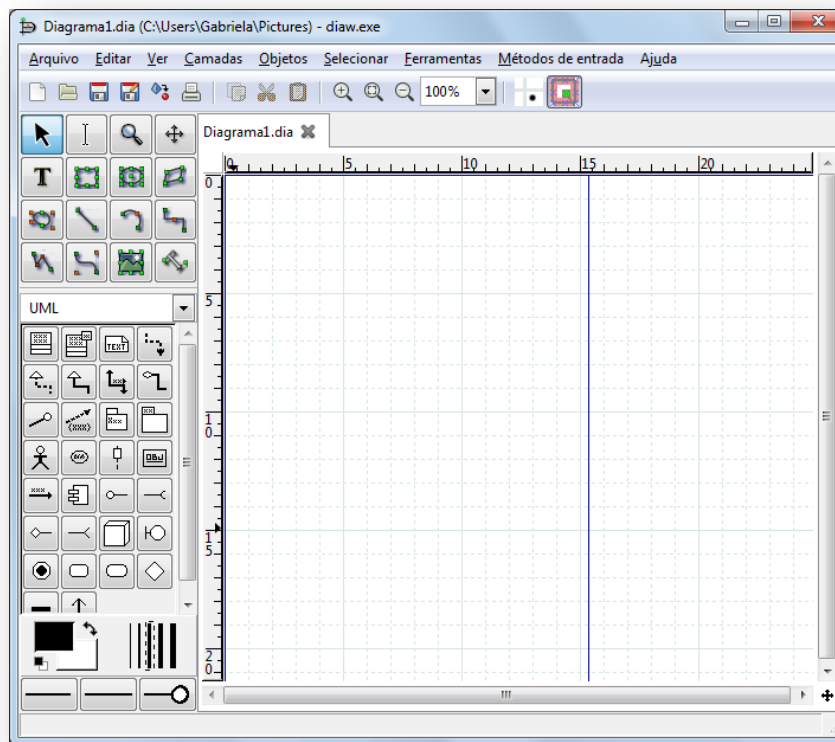


Figura 3. Interface da ferramenta Dia.

3.3.2 Protégé

A ferramenta utilizada para a construção da ontologia foi o *Protégé*²⁸ versão 3.4-8 release com suporte para OWL 1.0 e RDFS compatível com o Windows 7.

O *Protégé* é uma ferramenta que permite construir ontologias, personalizar formulários de entrada de dados, inserir e editar dados, possibilitando, então, a criação de bases de conhecimento guiadas por uma ontologia. Sua interface gráfica provê acesso a barra de menus e barra de ferramentas, além de apresentar cinco áreas de visualização (*views*) que funcionam como módulos de navegação e edição de classes, atributos, formulários, instâncias e pesquisas na base de conhecimento, propiciando a entrada de dados e a recuperação das informações [Semprebom *et al*, 2007].

O editor de modelagem do *Protégé*, *Editor Protégé-OWL*, gera o código OWL da ontologia automaticamente à medida que a ontologia está sendo criada e todas as propriedades e construtores OWL, RDF e XML podem ser utilizados durante a criação.

²⁸<http://protege.stanford.edu/>

O *Protégé* apresenta alguns menus para a criação de ontologias como, por exemplo, [Vanni 2009]:

- ✓ **OWL Classes:** permite modelar a hierarquia das classes da ontologia.
- ✓ **Properties:** permite a criação das propriedades da ontologia.
- ✓ **Forms:** mostra as propriedades criadas para as classes da ontologia
- ✓ **Individuals:** permite criar instâncias para as classes da ontologia
- ✓ **Metadata:** mostra todos os *namespaces* das ontologias e permite importar outras ontologias.

A Figura 4 apresenta a interface da ferramenta *Protégé* com os menus descritos anteriormente.

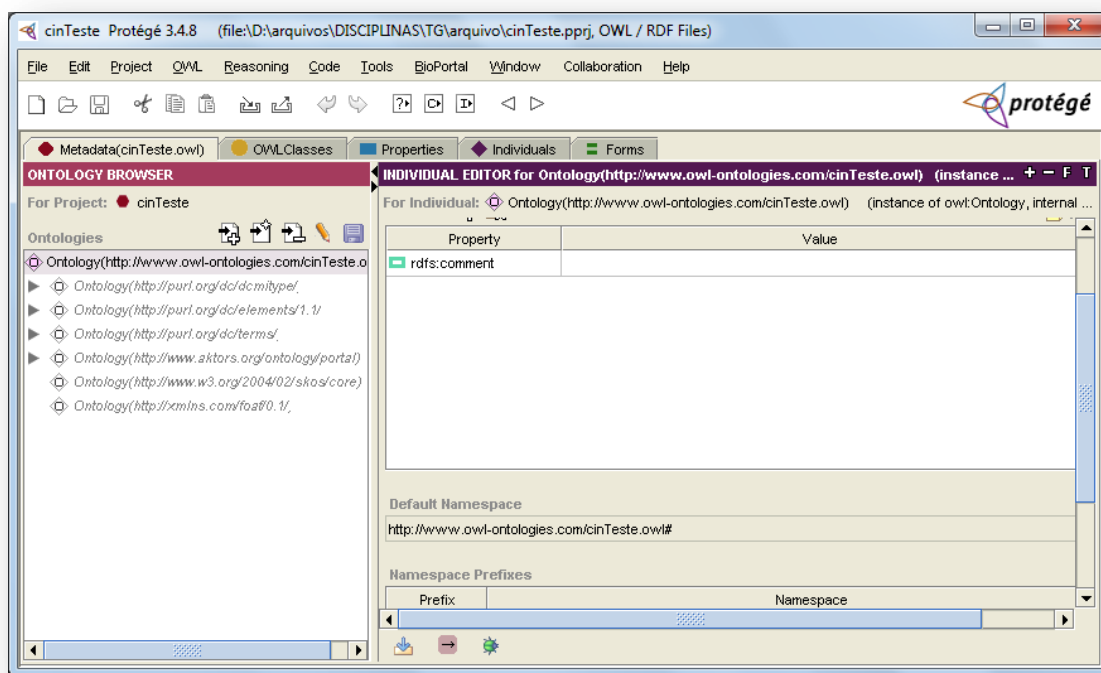


Figura 4. Interface do *Protégé*.

3.3.3 Jena TDB

Jena²⁹ é um framework Java usado para a construção de aplicações na área de Web Semântica. Este framework fornece um ambiente de programação para RDF, RDFS, OWL, SPARQL e ainda inclui um motor de inferências baseado em regras. Para o suporte a manipulação de ontologias, Jena possui um pacote chamado *API Jena 2 Ontology*. Nesse pacote existem classes para a manipulação de ontologias em RDFS, DAML+OIL e OWL. Jena³⁰ é um framework Java usado para a construção de aplicações na área de Web Semântica.

Jena trabalha de forma transparente em relação à linguagem adotada para representar a ontologia. De acordo com a linguagem, determinados recursos são habilitados, porém a forma de manipular e criar ontologias são a mesma para qualquer linguagem de representação [Frozza & Gonçalves 2010].

3.3.4 Joseki

O *Joseki*³¹ é o servidor do Jena TDB que pode ser executado sobre uma máquina virtual Java e provê um *endpoint* para consultas SPARQL sobre um conjunto de dados locais representados como conjuntos de dados RDF. Também disponibiliza um serviço HTTP que funciona como motor de busca sobre o grafo RDF armazenando na memória, obtido a partir do processamento dos conjuntos de dados pelo servidor. O servidor RDF *Joseki* possibilita [Battassini 2011]:

- ✓ Definir conjuntos de dados RDF pré-estabelecidos a partir de arquivos locais ou remotos;
- ✓ Carregar os conjuntos de dados para a memória na forma de um grafo RDF;
- ✓ Consultas o grafo do conjunto de dados atualmente carregado em linguagem SPARQL;
- ✓ Atualizar o grafo com novas triplas RDF.

²⁹<http://jena.apache.org/>

³⁰<http://jena.apache.org/>

³¹<http://joseki.sourceforge.net/>

O servidor, ao ser inicializado, processa o conjunto de dados presente no arquivo de configuração, obtém o grafo RDF que representa as triplas desse conjunto de dados e, em seguida, armazena esse grafo na memória.

A partir de um navegador, conforme visto na Figura 5, e utilizando o protocolo HTTP, a consulta SPARQL é enviada via POST ou via GET para o servidor e o mesmo retorna o resultado da consulta no formato XML. Porém, para a exibição direta na tela é solicitado ao servidor que converta o estilo de exibição de XML para HTML [Battassini 2011]. O *Joseki* ainda permite a exportação do resultado em planilhas no formato JSON, N-Triplas RDF-XML ou CSV.

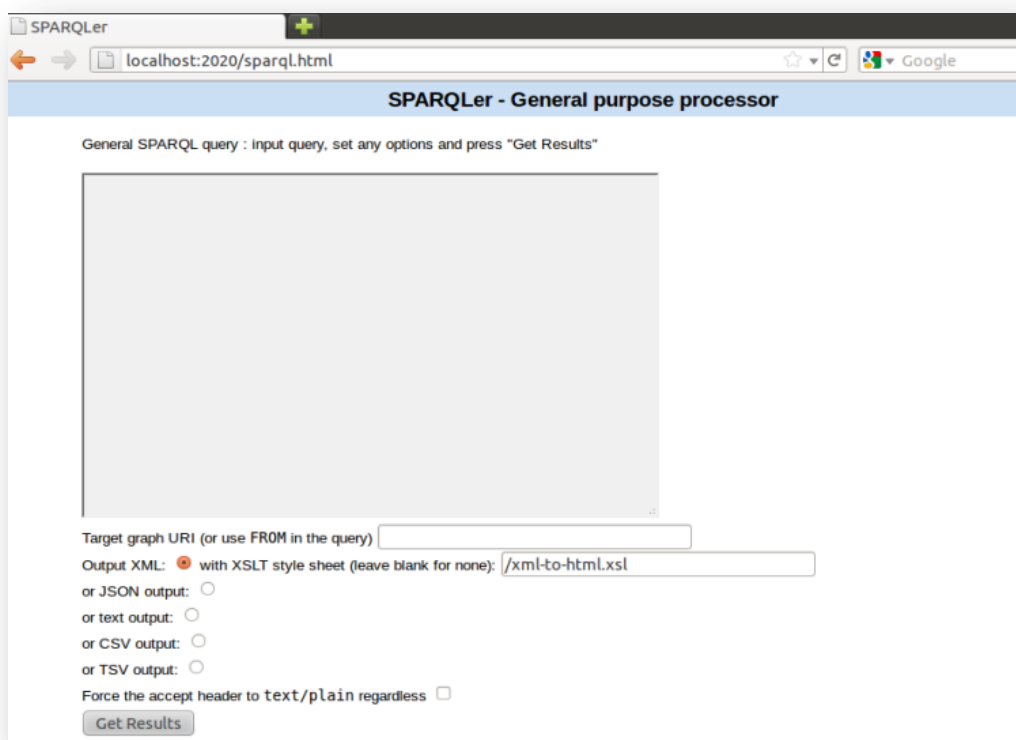


Figura 5. Interface do SPARQL endpoint.

3.4 Passo 3: Definição do esquema do conjunto de dados

Com ajuda da ferramenta DIA, foi gerado um esquema o qual foi utilizado como base para a criação da ontologia, conforme apresentado na Figura 6. Vale salientar que o esquema abaixo não foi o esquema definitivo do nosso trabalho, pois quando a ontologia estava sendo criadas algumas alterações tiveram de ser realizadas.

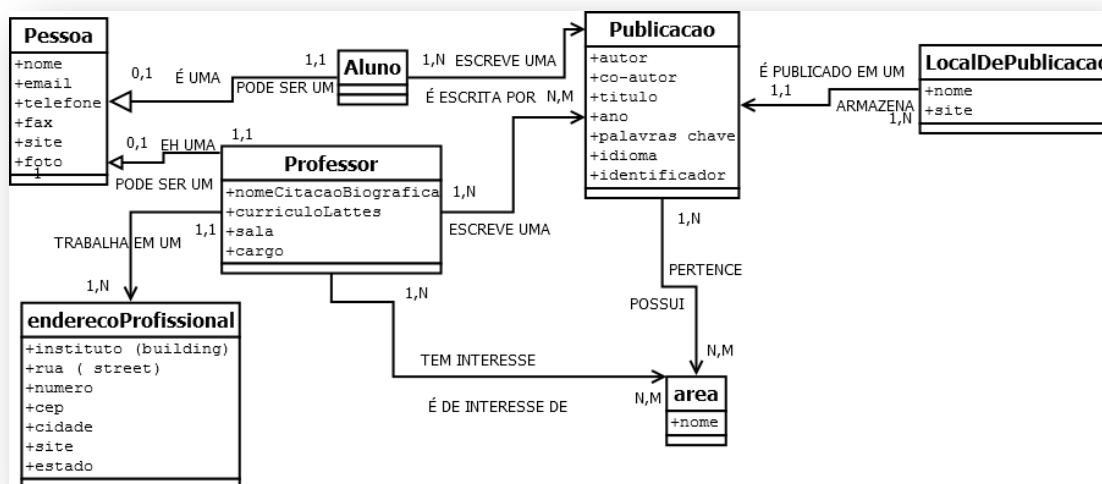


Figura 6. Esquema de Dados.

A partir do esquema da Figura 6, percebe-se que a ontologia terá informações pessoais dos professores, como nome, email, site e, também, informações profissionais como cargo, áreas de interesse e endereço profissional.

O professor, com a colaboração de alunos, escreve artigos, os quais são identificados pelo termo publicação. A ontologia possuirá termos que descrevem algumas informações relevantes de uma publicação como, por exemplo, autores, título, palavras-chave, *abstract* e a URL de acesso. Cada publicação faz parte de algum evento, congresso, seminário ou periódico, representados no esquema por *LocalDePublicacao*, que possui o nome e o site do evento.

3.5 Passo 4: Criação da Ontologia

A Metodologia 101 [Noy & McGuinness 2001] foi a metodologia escolhida para a criação da ontologia do projeto por ser simples e possuir um processo iterativo, ou seja, ao longo do processo de criação a ontologia sofre várias alterações, como criação de novos termos, mudanças na hierarquia de classes, troca de termos reusados. A Metodologia 101 define sete passos que devem ser considerados na elaboração de uma ontologia. Esses passos foram citados em [Melo & Junior 2010] como:

- ✓ **Determinar o domínio e o escopo da ontologia:** determinação do domínio a ser tratado e o escopo de abrangência do mesmo;
- ✓ **Considerar o reuso de ontologias existentes:** reutilização de termos de ontologias já previamente definidas e difundidas tornam as aplicações mais confiáveis. O reuso de ontologia é importante na medida em que o novo sistema necessite interagir com outros sistemas que se utilizam das mesmas ontologias ou vocabulários reutilizados. Dessa forma é possível refinar e estender ontologias existentes para o domínio tratado.
- ✓ **Enumerar os termos importantes:** visa listar os termos importantes que serão utilizados na ontologia sem se preocupar com o tipo das relações entre eles e restrições que eles possam vir a ter.
- ✓ **Definir classes e hierarquia:** definição das classes necessárias para a ontologia e sua organização em uma taxonomia hierárquica. A abordagem escolhida foi a *top-down*. Essa abordagem parte de uma visão geral sobre o domínio para os casos mais específicos, possibilitando uma melhor compreensão de todo o processo e do domínio como um todo, ou seja, começa definir os conceitos mais gerais e depois os mais específicos.
- ✓ **Definir as propriedades das classes:** definição da estrutura interna dos conceitos e das classes às quais ele corresponde.

- ✓ **Delimitar os valores das propriedades:** definição das restrições das propriedades, ou seja, tipo do valor, valor permitido, valor mínimo, valor máximo entre outros.

As próximas sessões detalhará cada um dos passos descritos acima.

3.5.1 Domínio e Escopo da Ontologia

A ontologia está relacionada com informações pessoais e profissionais dos docentes dos CIn, além de suas publicações. Com os dados presentes na ontologia deverá ser possível responder a algumas perguntas como, por exemplo:

- ✓ Quais foram as publicações de cada professor no ano de 2010?
- ✓ Quais as publicações que tem como palavra-chave ontologias e onde podem ser encontradas?
- ✓ Em quais salas do CIn se encontram os professores? Quais são seus endereços profissionais e Currículo Lattes?
- ✓ Quais são as áreas de interesse de cada professor do CIn?
- ✓ Quais alunos do CIn e com quais professores esses alunos publicaram algum artigo?
- ✓ Quantas publicações foram realizadas pelos professores do CIn em 2010?
- ✓ Em que eventos estão as publicações da professora Bernadette Farias Lóscio?
- ✓ Quais são as publicações que Bernadette Lóscio e Ana Carolina escreveram juntas?
- ✓ Como citar a referência bibliográfica das publicações do professor Fernando Fonseca em um documento?
- ✓ Quais são as informações referentes às publicações do Professor Aluísio do ano de 2011?

3.5.2 Reutilização de Vocabulários existentes

Um dos pilares da Web Semântica é reutilizar vocabulários já existentes. Para facilitar a criação da ontologia deste projeto, foram reutilizados termos, classes e propriedades dos vocabulários apresentados na Figura 7:

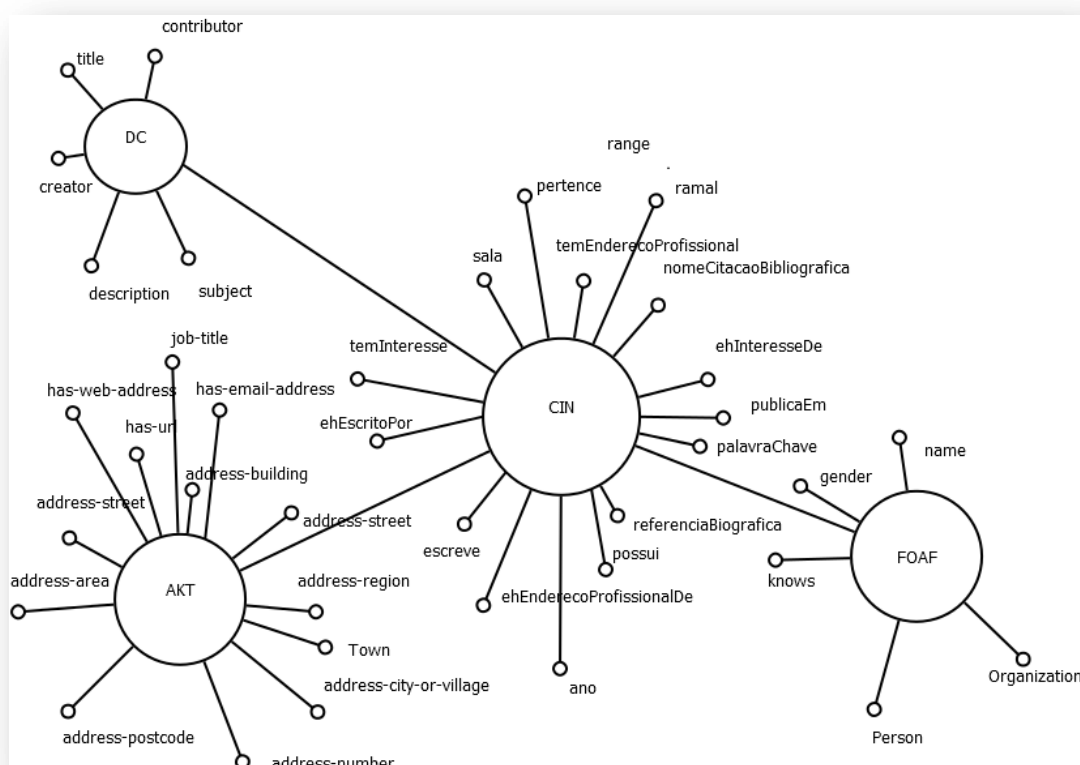


Figura 7. Ontologia do CInPub reutilizando outros vocabulários

Os termos do vocabulário FOAF utilizados para a ontologia do CInPub são mostrados na Tabela 2.

Tabela 2. Termos do vocabulário FOAF usados para a definição do CInPub

Termos do vocabulário FOAF
<i>foaf:Person</i>
<i>foaf:Organization</i>
<i>foaf:name</i>
<i>foaf:gender</i>
<i>foaf:knows</i>
<i>foaf:Interest</i>

Os termos reutilizados do vocabulário *Dublin Core* para a criação da nossa ontologia foram destacadas na Tabela 3.

Tabela 3. Termos do vocabulário Dublin Core usados para a definição do CInPub

Termo do vocabulário Dublin Core
<i>dc:creator</i>
<i>dc:contributor</i>
<i>dc:title</i>
<i>dc:description</i>

Os elementos necessários do AKT para a ontologia deste projeto são apresentados na Tabela 4.

Tabela 4. Termos reusados do vocabulário AKT usados para a definição do CInPub

Termo do vocabulário AKT
<i>akt:address-building</i>
<i>akt:address-street</i>
<i>akt:address-number</i>
<i>akt:address-area</i>
<i>akt:address-postcode</i>
<i>akt: address-city-or-village</i>
<i>akt:address-region</i>
<i>akt:has-telephone-number</i>
<i>akt:has-fax-number</i>
<i>akt:job-title</i>
<i>akt:has-web-adress</i>
<i>akt:has-email-adress</i>
<i>akt:has-URL</i>

3.5.3 Definição dos Termos

O segundo passo da Metodologia 101 define quais são os termos relevantes que descrevem o domínio em questão e que farão parte da ontologia. Segundo essa metodologia, não há

a preocupação de definir exatamente quais e de que forma estes termos irão para a ontologia.

Cada termo descrito na Tabela 5, será acompanhado de uma breve justificativa do seu uso.

Tabela 5. Termos e Relacionamentos que descrevem o domínio da Ontologia criada

Termo ou Relacionamento	Definição
Pessoa	Termo para identificar as pessoas da ontologia.
Professor	Pessoa central da ontologia e que escreve uma publicação.
Nome; Nome em citação biográfica; Cargo; Email; Sala; Ramal; Site pessoal; Currículo Lattes	Conjunto de termos para descrever um professor.
Colaborador	É uma pessoa que colabora com o professor na criação de um artigo.
Endereço Profissional	Informar o local de trabalho do professor.
Instituto; Sala; Rua; Numero; CEP; Bairro; Cidade; Site do Departamento; Telefone; Fax.	Conjunto de termos que descreve o endereço profissional do professor.
Área	Área acadêmica de interesse do professor.
Publicação	Identifica as publicações do professor.
Autor; Coautor; Título; Descrição da publicação; Ano da publicação; Referência Biográfica; URL de acesso.	Conjunto de termos que descrevem uma publicação.
Local da Publicação	Evento ou local onde foi realizado a publicação do professor.
Nome do local da publicação; Site.	Conjunto de termos que descrevem o local da publicação.
Tem endereço profissional	Relaciona um professor ao endereço profissional.
Tem interesse	Relaciona um professor com as áreas de interesse.
Conhece	Informa quais são as pessoas (professor ou colaborador) que um professor conhece.
Escreve	Identifica a relação entre o professor e o colaborador com uma publicação.
É publicado em	Relação de uma publicação e o local onde foi publicada.

3.5.4 Definição das Classes e Hierarquia de Classes

A partir dos termos identificados no passo anterior, foram definidas as classes que irão compor a ontologia do CInPub. A Tabela 6 mostra quais são os termos que se tornaram classe e a nomenclatura utilizada na ontologia.

Tabela 6. Classes da ontologia do CInPub.

Classes em linguagem natural	Classe no CInPub
Área	<i>cin:Area</i>
Colaborador	<i>cin:Colaborador</i>
Endereço Profissional	<i>cin:EnderecoProfissional</i>
Local da Publicação	<i>cin:LocalDaPublicacao</i>
Professor	<i>cin:Professor</i>
Publicação	<i>cin:Publicacao</i>

A Figura 8 mostra a hierarquia de classes da ontologia do CInPub.

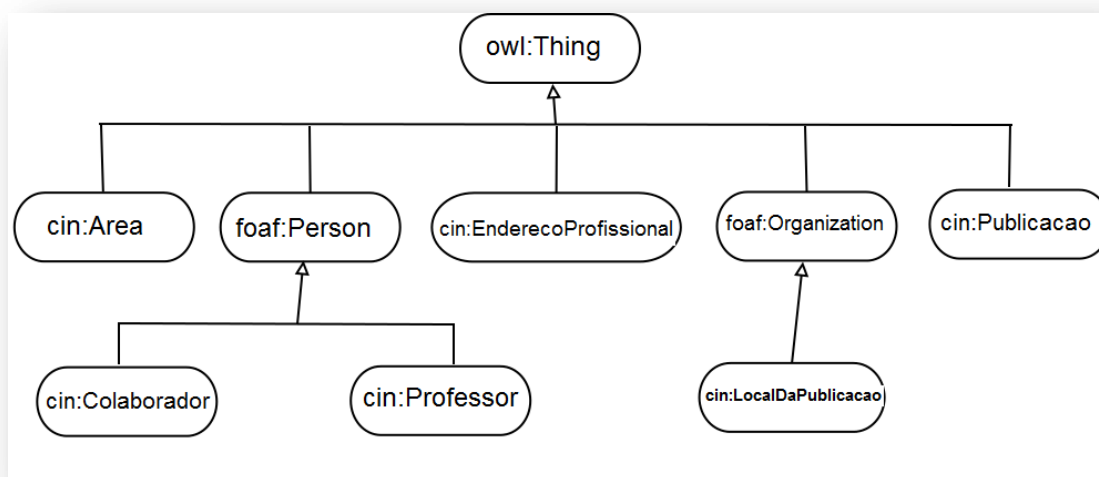


Figura 8. Hierarquia de classes do CInPub.

3.5.3 Definição das Propriedades das Classes

A partir dos termos definidos, os termos que não se tornaram classes, foram transformados em propriedades. Essas propriedades foram divididas no *Protégé* em *Datatype Property* quando se refere a um atributo da classe e *Object Property* quando a propriedade é uma relação entre duas classes.

A Tabela 7 mostra os termos que foram transformados em *datatype property*. A primeira coluna indica a classe da qual a propriedade faz parte e a terceira coluna mostra o nome da propriedade precedido do prefixo do vocabulário ao qual pertence.

Tabela 7. Propriedades do tipo Datatype Property.

Classe	Termo	Propriedade
cin:Colaborador	Nome	<i>foaf:name</i>
cin:Area	nome	<i>cin:nome</i>
cin:Professor	Nome	<i>foaf:name</i>
cin:Professor	nome em citação bibliografica	<i>cin:nomeCitacaoBibliografica</i>
cin:Professor	Cargo	<i>cin:cargo</i>
cin:Professor	Email	<i>akt:has-email-adress</i>
cin:Professor	Ramal	<i>cin:ramal</i>
cin:Professor	site pessoal	<i>akt:has-web-adress</i>
cin:Professor	endereço do Currículo Lattes	<i>akt:has-URL</i>
cin:Professor	Sala	<i>cin:sala</i>
cin:EnderecoProfissional	instituto	<i>akt:adress-building</i>
cin:EnderecoProfissional	Rua	<i>akt:adress-street</i>
cin:EnderecoProfissional	numero	<i>akt:number</i>
cin:EnderecoProfissional	CEP	<i>akt:postcode</i>
cin:EnderecoProfissional	Bairro	<i>akt:adress-area</i>
cin:EnderecoProfissional	Cidade	<i>akt:adress-city-or-village</i>
cin:EnderecoProfissional	telefone	<i>akt:has-telephone-number</i>
cin:EnderecoProfissional	Fax	<i>akt:has-fax-number</i>
cin:Publicacao	Autor	<i>dc:creator</i>
cin:Publicacao	coautor	<i>dc:contributor</i>
cin:Publicacao	Titulo	<i>dc:title</i>
cin:Publicacao	ano	<i>cin:ano</i>
cin:Publicacao	abstract	<i>dc:description</i>
cin:Publicacao	endereço de acesso	<i>akt:has-URL</i>
cin:Publicacao	referência biográfica	<i>cin:referenciaBiografica</i>
cin:Publicacao	palavra-chave	<i>cin:palavraChave</i>
cin:LocalDaPublicacao	Nome	<i>foaf:name</i>
cin:LocalDaPublicacao	Site	<i>akt: has-URL</i>

Os termos que foram definidos como *Object Property* são listados na Tabela 8 informando a nomenclatura utilizada na ontologia e quais classes estão envolvidas no relacionamento.

Tabela 8. Propriedades de relacionamento.

Propriedades	Termo	Relacionamento
tem interesse	<i>cin:temInteresse</i>	<i>cin:Professor</i> com <i>cin:Area</i>
tem endereco profissional	<i>cin:temEnderecoProfissional</i>	<i>cin:Professor</i> com <i>cin.EnderecoProfissional.</i>
escreve	<i>cin:escreve</i>	<i>cin:Professor</i> com <i>cin:Publicacao</i>
é publicado em	<i>cin:ehPublicadoEm</i>	<i>cin:Publicacao</i> com <i>cin:LocalDaPublicacao</i>

Os termos da Tabela 9 também são *object property* e são as propriedades inversas das propriedades descritas acima.

Tabela 9. Propriedades inversas.

Propriedades	Termo	Relacionamento
é interesse de	<i>cin:ehInteresseDe</i>	<i>cin:Areacom cin:Professor</i>
é endereço profissional de	<i>cin:ehEnderecoProfissionalDe</i>	<i>cin.EnderecoProfissional com cin:Professor</i>
é escrito por	<i>cin:ehEscritoPor</i>	<i>cin:Publicacaocom cin:Professor</i>
armazena	<i>cin:armazena</i>	<i>cin:LocalDaPublicacaocom cin:Publicacao</i>

3.5.4 Definição das Restrições

Neste passo, foram definidas algumas restrições com relação ao domínio:

- ✓ **Domain:** a qual classe a propriedade pertence;
- ✓ **Range:** que tipo de valor a propriedade recebe;
- ✓ **Cardinalidade** das propriedades da ontologia do projeto.

As próximas tabelas descrevem essas restrições.

Tabela 10. Restrições das propriedades da classe *cin:Colaborador*.

Domain	Propriedades	Termo Utilizado	Range	Cardinalidade
<i>cin:Colaborador</i>	Nome	<i>foaf:name</i>	String	1
<i>cin:Area</i>	Nome	<i>cin:nome</i>	String	1
<i>cin:Professor</i>	Nome	<i>foaf:name</i>	String	1
<i>cin:Professor</i>	nome em citação biográfica	<i>cin:nomeCitacaoBiografica</i>	String	≥ 1
<i>cin:Professor</i>	Cargo	<i>akt:job-title</i>	String	1
<i>cin:Professor</i>	Email	<i>akt:email-adress</i>	String	1
<i>cin:Professor</i>	Ramal	<i>cin:ramal</i>	String	1
<i>cin:Professor</i>	Site	<i>akt:has-web-adress</i>	String	1
<i>cin:Professor</i>	currículo lattes	<i>akt:has-URL</i>	String	1
<i>cin:Professor</i>	Sala	<i>cin:sala</i>	String	1
<i>cin:EnderecoProfissional</i>	Instituto	<i>akt:adress-building</i>	String	1

cin:EnderecoProfissional	Rua	akt:adress-street	String	1
cin:EnderecoProfissional	Numero	akt:adress-number	String	0 ou 1
cin:EnderecoProfissional	CEP	akt:adress-postcode	String	1
cin:EnderecoProfissional	Bairro	akt:adress-area	String	1
cin:EnderecoProfissional	Cidade	akt:adress-city-or-village	String	1
cin:EnderecoProfissional	Estado	akt:adress-region	String	1
cin:EnderecoProfissional	Site	akt:has-web-adress	String	1
cin:EnderecoProfissional	Telefone	akt:has-telephone-number	String	1
cin:EnderecoProfissional	Fax	akt:has-fax-number	String	1
cin:Publicacao	Autor	professor:nome	String	≥ 1
cin:Publicacao	Coautor	colaborador:nome	String	≥ 0
cin:Publicacao	Citulo	dc:title	String	1
cin:Publicacao	Ano	cin:data	String	1
cin:Publicacao	Abstract	dc:description	String	≥ 1
cin:Publicacao	referência bibliográfica	cin:referenciaBiografica	String	1
cin:Publicacao	URL de acesso	akt:has-URL	String	1
cin:Publicacao	palavra-chave	cin:palavraChave	String	≥ 1
cin:LocalDaPublicacao	Nome	foaf:name	String	1
cin:LocalDaPublicacao	Site	akt:has-web-adress	String	≥ 0

Tabela 11. Restrições das propriedades de relacionamento.

Propriedades	Domain	Range	Cardinal idade
cin :temInteresse	cin:Professor	cin:Area	≥ 1
cin :temEnderecoProfissional	cin:Professor	cin:EnderecoProfissional	1
cin :escreve	cin:Professor, cin:Colaborador	cin:Publicacao	≥ 1
cin :ehPublicadoEm	cin:Publicacao	cin:LocalDaPublicado	1

Tabela 12. Restrições das propriedades inversas.

Propriedades	Domain	Range	Cardinalidade
<i>cin:ehInteresseDe</i>	<i>cin:Area</i>	<i>cin:Professor</i>	≥ 1
<i>cin:ehEnderecoProfissionalDe</i>	<i>cin:EnderecoProfissional</i>	<i>cin:Professor</i>	≥ 1
<i>cin:escritoPor</i>	<i>cin:Publicacao</i>	<i>cin:Professor</i> , <i>cin:Colaborador</i>	≥ 1
<i>cin:armazena</i>	<i>cin:LocalDaPublicado</i>	<i>cin:Publicacao</i>	≥ 1

Após a definição das classes, propriedades e restrições, foi obtido o esquema apresentado na Figura 9.

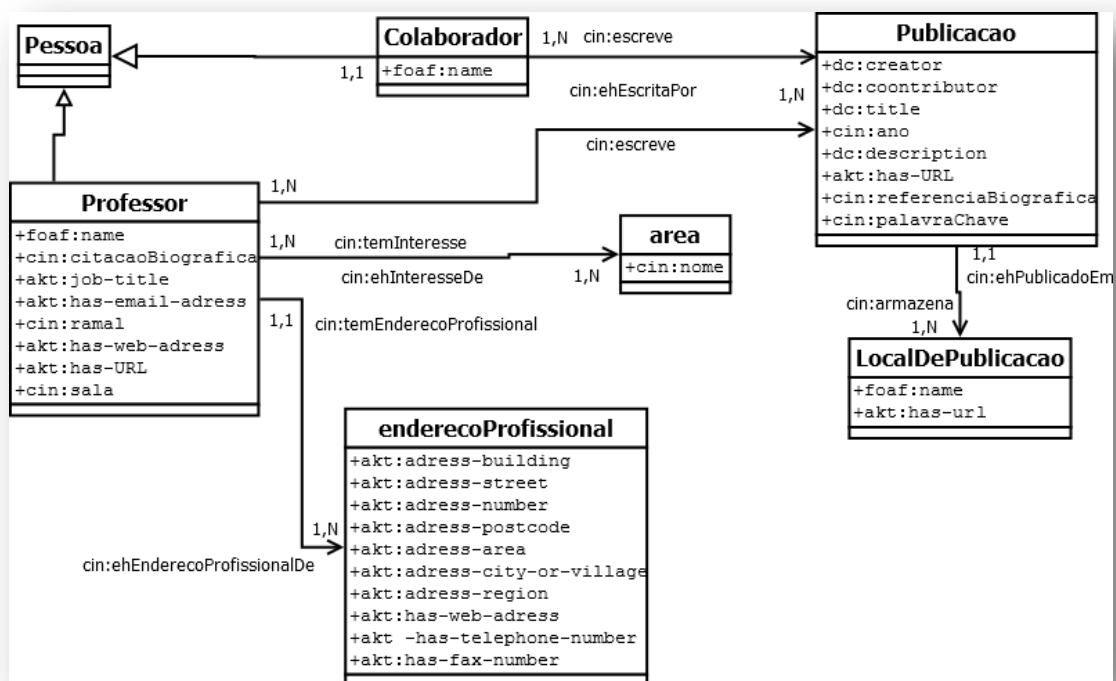


Figura 9. Esquema da Ontologia do CInPub.

3.6 Passo 5: Criação das Instâncias

Este passo se caracteriza pelo povoamento da ontologia. As instâncias foram criadas com as informações recolhidas do site do CIn, do Currículo Lattes e das páginas pessoais dos professores. Em alguns casos, foram acessadas as publicações dos docentes para extrair informações referentes a autores, *abstract* e palavras-chaves.

Como demonstração, foram inseridos no *dataset* um endereço profissional, já que o projeto é para os docentes do CIn, quarenta e oito áreas de interesses, vinte e três publicações, trinta colaboradores, oito professores e trinta locais de publicações, totalizando mil cento e sessenta e cinco triplas.

3.7 Considerações Finais

Neste capítulo vimos como se deu a criação da base de conhecimento CInPub, descrevendo passo a passo desde o processo de escolha das fontes, passando pela criação da ontologia até o povoamento do *dataset*. No próximo capítulo vamos das das configurações precisas para permitir a realização de consultas e demonstrar as consultas em cima do *dataset* criado.

Capítulo 4. Realizando consultas no CInPub

Neste capítulo será ilustrado o processo de recuperação de dados do CInPub com o uso da linguagem de consulta SPARQL.

4.1 Recuperação dos dados do CInPub

Esta sessão traz as informações de configurações para executar as ferramentas e permitir a realização das consulta no SPARQL *endpoint*.

4.1.1 Configuração do Jena TDB para Realização das Consultas

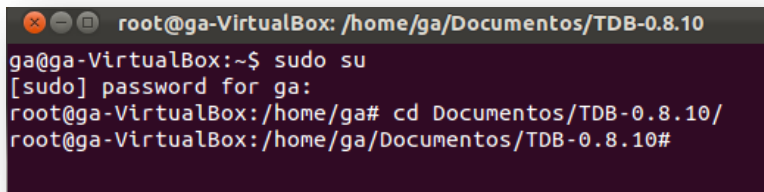
Para que seja possível recuperar os dados do *dataset* é preciso que o mesmo esteja armazenado em um servidor. Neste projeto, o servidor utilizado foi o Joseki-3.4.4 para Linux Ubutu 11³².

Após exportar a ontologia no formato OWL no *Protégé*, deve-se criar uma pasta dentro da pasta do TDB³³, componente do Jena, usado para armazenar e consultar as triplas importadas da ontologia. Feito isso, deve-se configurar o SPARQL *endpoint* informando o caminho dessa pasta para permitir a realização das consultas sobre o *dataset*.

A pasta criada neste projeto foi nomeada *MyDataset* e está localizada dentro da pasta do TDB. Para a importação do *dataset* CInPub deve-se abrir o terminal do Linux, fazer login como administrador, entrar na pasta onde se encontra o TDB conforme a Figura 10:

³²<http://www.ubuntu-br.org/>

³³<http://jena.apache.org/documentation/tdb/index.html>



```

root@ga-VirtualBox: /home/ga/Documentos/TDB-0.8.10
ga@ga-VirtualBox:~$ sudo su
[sudo] password for ga:
root@ga-VirtualBox:/home/ga# cd Documentos/TDB-0.8.10/
root@ga-VirtualBox:/home/ga/Documentos/TDB-0.8.10#

```

Figura 10. Terminal do Linux.

É preciso fazer login como super-usuário para realizar todo o processamento com o TDB.

Com o terminal direcionando para a pasta do TDB digita-se então o comando de importação do *dataset* informando o caminho da pasta *MyDataset* e o caminho onde se encontra arquivado o *dataset*:

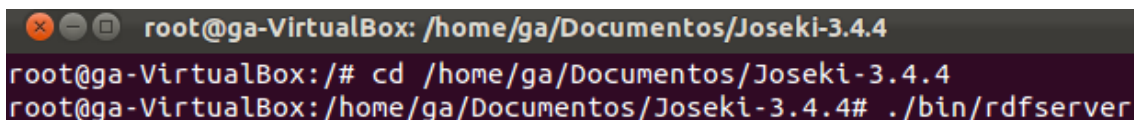
```

/home/ga/Documentos/TDB-0.8.10# ./bin/tdbloader -loc=/home/ga/Documentos/TDB-0.8.10/MyDataset/
/home/ga/Downloads/cinPub.owl

```

Após a importação, se a mesma foi realizada com sucesso, o terminal informa a quantidade de triplas armazenadas.

Para iniciar o serviço de consultas sobre os dados importados através do SPARQL *endpoint*, deve-se executar as linhas de comando para ligar o servidor RDF. Primeiro entra na pasta onde se encontra o *Joseki* e, em seguida, digita-se o comando como mostra a Figura 11.



```

root@ga-VirtualBox: /home/ga/Documentos/Joseki-3.4.4
root@ga-VirtualBox:/# cd /home/ga/Documentos/Joseki-3.4.4
root@ga-VirtualBox:/home/ga/Documentos/Joseki-3.4.4# ./bin/rdfserver

```

Figura 11. Linhas de comando para iniciar o Joseki.

Após o passo acima, para poder realizar as consultas em um SPARQL *endpoint*, abre-se o navegador e digita na URL o endereço: “localhost/2020/sparql.html”.

4.1.2 Realizando consultas SPARQL no CInPub

As consultas em SPARQL, que serão mostradas nessa seção, foram realizadas no grafo gerado a partir das triplas inseridas no repositório CInPub e respondem às perguntas feitas na Seção 3.5.1. Nestas consultas, foram utilizados os prefixos descritos na Tabela 13 para abreviar a URIS dos vocabulários necessários a fim de diminuir o tamanho das consultas e facilitar no entendimento.

Tabela 13. Prefixos utilizados nas consultas.

Prefixo	Vocabulário
Cin	">http://www.owl-ontologies.com/cin.owl#>
rdf	">http://www.w3.org/1999/02/22-rdf-syntax-ns#>
Foaf	http://xmlns.com/foaf/0.1/
Dc	http://purl.org/dc/terms/
akt	http://www.aktors.org/ontology/portal#

Consulta 1: Quais foram as publicações de cada professor realizadas no ano de 2010?

```

1 PREFIX cin: <http://www.owl-ontologies.com/cin.owl#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
4 PREFIX dc: <http://purl.org/dc/terms/>
5
6 SELECT ?professorNome ?publicacaoTitulo WHERE {
7   ?professor rdf:type cin:Professor .
8   ?professor cin:escreve ?publicacao .
9   ?professor foaf:name ?professorNome .
10  ?publicacao cin:ano 2010 ^^<http://www.w3.org/2001/XMLSchema#string> .
11  ?publicacao dc:title ?publicacaoTitulo .
12}
13 ORDER BY ASC(?professorNome)

```

Como resposta para essa consulta, duas colunas foram criadas, uma com o nome do professor e outra com o título da publicação e nas condições do *WHERE*, na linha 8, foi feita a relação dessas duas colunas a partir da propriedade *cin:escreve*, pois um professor escreve pelo menos uma publicação. As linhas 9 e 11 atribuem os valores corretos para as colunas criadas ordenado pelo nome do professor visto na linha 10. A Tabela 14 abaixo mostra o resultado da consulta.

Tabela 14. Resultado da consulta 1.

Resultado da Consulta 1	
professorNome	publicacaoNome
Abel Guilhermino da Silva Filho	MOTION DETECTION OF VEHICLES BASED ON FPGA
Abel Guilhermino da Silva Filho	NSGAI Applied to Unified Second Level Cache Memory Hierarchy Tuning Aiming Energy and Performance Optimization
Aluizio Fausto Ribeiro Araujo	Growing Self-Reconstruction Maps
Aluizio Fausto Ribeiro Araujo	Growing self-organizing surface map: learning a surface topology from a point cloud
Ana Carolina Brandao Salgado	Modelling and querying geographical data warehouses?
Ana Carolina Brandao Salgado	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels
Bernadette Farias Loscio	MeMO: A clustering based approach for multiple ontology merging
Bernadette Farias Loscio	Support Tool in the Diagnosis of Sales Price of Dental Plans
Fernando da Fonseca de Souza	Amadeus-TV: Portal Educacional na TV Digital Integrado a um Sistema de Gestão de Aprendizado
Fernando da Fonseca de Souza	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels
Patricia Cabral de Azevedo Restelli Tedesco	The Gravitational Strategy for the Timed Patrolling
Patrícia Cabral de Azevedo Restelli Tedesco	Using a Metamodel to Design Structural and Behavioral Aspects in Context-Sensitive Groupware
Robson do Nascimento Fidalgo	Modelling and querying geographical data warehouses?
Robson do Nascimento Fidalgo	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels
Valéria Cesário Times	Modelling and querying geographical data warehouses?
Valéria Cesário Times	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels

Consulta 2: Quais as publicações que têm como palavra-chave *ontologies* e onde podem ser encontradas?

PREFIX cin: <http://www.owl-ontologies.com/cin.owl#>
 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
 PREFIX dc: <http://purl.org/dc/terms/>
 PREFIX akt: <http://www.aktors.org/ontology/portal#>

SELECT ?professorNome ?publicacaoTitulo ?enderecoUrl WHERE{

?professor rdf:type cin:Professor .

```

?professor cin:escreve ?publicacao .
?professor foaf:name ?professorNome .
?publicacao dc:title ?publicacaoTitulo .
?publicacao akt:has-URL ?enderecoUrl .
?publicacao cin:palavraChave ?publicacaoPalavraChave .
FILTER(?publicacaoPalavraChave = Ontologies ^^<http://www.w3.org/2001/XMLSchema#string> ||
?publicacaoPalavraChave = ontologies ^^<http://www.w3.org/2001/XMLSchema#string>)

} ORDER BY ASC(?professorNome)

```

Tabela 15. Resultado da consulta 2.

Resultado da Consulta 2		
ProfessorNome	PublicacaoTitulo	EnderecoUrl
Ana Carolina Brandao Salgado	The Data Integration Research Group at UFPE	http://seer.lcc.ufmg.br/index.php/jidm/article/view/121/74
Ana Carolina Brandao Salgado	Retrieving Specific Domain Information from the Web through Ontologies	http://www.irma-international.org/viewtitle/2405/
Ana Carolina Brandao Salgado	Ontology-Based Clustering in a Peer Data Management System	http://www.igi-global.com/viewtitlesample.aspx?id=66054
Bernadette Farias Loscio	An incremental and user feedback-based ontology matching approach	http://delivery.acm.org/10.1145/2100000/2095607/p371-wagner.pdf?ip=150.161.2.200&acc=ACTIVE%20SERVICE&CFID=95685376&CFTOKEN=91487570&__acm__=1342275142_c5b713812e0cae f2f6bc85088cc4e76c
Bernadette Farias Loscio	Towards Automatic Generation of Application Ontologies	http://webscience.org.br/wiki/images/d/d1/Casanova_2010_ICEIS.pdf
Bernadette Farias Loscio	The Data Integration Research Group at UFPE	http://seer.lcc.ufmg.br/index.php/jidm/article/view/121/74
Fernando da Fonseca de Souza	Retrieving Specific Domain Information from the Web through Ontologies	http://www.irma-international.org/viewtitle/2405/

De forma parecida com a consulta 1, como resposta a essa consulta, foi gerada uma coluna para informar o endereço de acesso à publicação atribuindo a ele a propriedade correta da classe Publicação (*akt:has-URL*) e criado a condição da palavra-chave através de um filtro para que a busca seja feita em cima das palavras *ontologies* ou *Ontologies*. O resultado da consulta é mostrado na Tabela 15.

Consulta 3: Em qual sala do CIn se encontram os professores? Quais são seu endereços pessoal e Currículo Lattes ?

```

1 PREFIX cin: <http://www.owl-ontologies.com/cin.owl#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
4 PREFIX akt: <http://www.aktors.org/ontology/portal#>
5
6 SELECT ?professorNome ?professorSala ?professorPagina ?professorCurriculo WHERE {
7 ?professor rdf:type cin:Professor .
8 ?professor foaf:name ?professorNome .
9 ?professor cin:sala ?professorSala .
10 ?professor akt:has-web-address ?professorPagina .
11 ?professor akt:has-URL ?professorCurriculo .
12}
13 ORDER BY ?professorNome

```

O resultado da consulta é mostrada na Tabela 16 com quatro colunas: nome do professor, a sala, a página pessoal e o endereço do Currículo Lattes .

Tabela 16. Resultado da consulta 3.

Resultado da Consulta 3			
professorNome	professorSala	ProfessorPagina	professorCurriculo
Abel Guilhermino da Silva Filho	B005	www.cin.ufpe.br/~agsf	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4792886T0
Aluizio Fausto Ribeiro Araújo	C128	http://www.cin.ufpe.br/~aluizioa/aluizio-homepage.htm	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4787817Y8
Ana Carolina Brandao Salgado	B120	www.cin.ufpe.br/~acs	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4787836A0
Bernadette Farias Loscio	B001	www.cin.ufpe.br/~bfl	http://lattes.cnpq.br/2512064355660153
Patricia Cabral de Azevedo Restelli Tedesco	B110	www.cin.ufpe.br/~pcart	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4728355A3
Robson do Nascimento Fidalgo	C024	www.cin.ufpe.br/~rdnf	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4705039A6
Valéria Cesário Times	C103	www.cin.ufpe.br/~vct	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4791983P6

As demais consultas realizadas encontram-se no ANEXO C.

4.2 Considerações Finais

Neste capítulo vimos as configurações que devem ser feitas no Jena TDB para permitir as consultas nas triplas armazenadas no *dataset*. O capítulo 5 finaliza este documento com as considerações finais: dificuldades encontradas e trabalhos futuros.

Capítulo 5. Considerações Finais

Este capítulo relata as considerações finais relativas ao trabalho desenvolvido ao longo deste projeto. Nas seções a seguir, são apresentadas as contribuições, bem como as dificuldades e restrições encontradas e possibilidades de expansão do trabalho efetuado em projetos futuros.

5.1 Contribuições

Este trabalho apresentou uma metodologia seguindo os princípios de *Linked Data* para a criação de um *dataset*, reutilizando termos de outras ontologias para descrever o domínio relacionado aos docentes do CIn juntamente com suas publicações, armazenando as informações relevantes em um *dataset*, disponibilizando um *sparql endpoint* para realização de consultas para permitir que outras aplicações possam tirar proveito dessas informações.

5.2 Dificuldades Encontradas

A primeira dificuldade encontrada no decorrer deste projeto foi a leitura dos artigos em inglês. Por eu não saber inglês acabei optando pelos artigos em português, limitando ainda mais a pesquisa. A quantidade de artigos encontradas na Web não era muito grande e, ainda sim, muitas informações eram repetitivas

A segunda foi no entendimento do que é Web Semântica, *Linked Data*, RDF, diferença entre vocabulário e ontologia, outros assuntos relacionados e o que realmente seria o projeto.

Outra dificuldade foi a utilização do *Protégé*. A versão atual apresentava alguns problemas de importação de ontologias além dos tutoriais disponibilizados na Web serem baseados em versões anteriores que apresentam uma interface bem diferente. A versão anterior do *Protégé* também apresentou problemas de travamentos constantes,

erros de inicialização, falhas ao salvar a ontologia e nos cadastros de alguns valores na criação de instâncias.

A reutilização de alguns termos de vocabulários como o Dublin Core também acarretou em mudanças de última hora na ontologia para que ela estivesse funcionando até o momento da apresentação deste projeto, pois o *Protégé* não permitia adicionar o *domain* e o *range* de suas propriedades. Esses termos foram: *dc:date*, *dc:language*, *dc:subject*, *dc:identifier* e *dc:Publisher*. Apesar de existirem alguns termos mais apropriados na ontologia FOAF para representar termos da ontologia do projeto, não foi possível reutilizá-los devido a alguns problemas que apareceram na utilização do *Protégé*. Esses termos podem ser inseridos em versões futuras da ontologia. São eles: *foaf:homepage*, *foaf:phone*, *foaf:mbox*, *foaf:workInfoHomepage* e *foaf:img*.

Na etapa de criação das entidades o cadastro manual das entidades tornou esse processo bastante lento. Era preciso fazer uma pesquisa procurando o local onde estava a publicação para saber quem eram seus autores e palavras-chave, pois o Currículo Lattes só disponibilizava as iniciais e o último sobrenome encontrados na forma de referência biográfica. Além disso, o CIn tem uma grande quantidade de professores e, estes, uma grande quantidade de publicação para serem registrados no *dataset*.

5.3 Trabalhos Futuros

Este projeto tornará mais acessível se seu código estiver armazenado em um servidor do CIn e disponibilizando uma área de administração on line, onde os professores, logados no sistema, possam cadastrar suas novas publicações e alterar outras informações, como por exemplo, área de interesse.

O uso de um *script* para extrair as informações do site do CIn, outro para extrair as informações do Currículo Lattes diminuiria bastante o trabalho da criação das entidades.

Novas informações podem ser adicionadas no *dataset* para tornar mais completo o perfil de cada professor. Além disso, os termos a serem utilizados na ontologia do projeto

podem ser reutilizados de outras ontologias não só das ontologias apresentadas neste trabalho.

Uma boa oportunidade é disponibilizar essas informações não apenas a partir de um *endpoint*, mas como uma página em HTML com o uso de RDFa, onde poderiam ser listados todos os professores com suas publicações, onde cada publicação teria as informações de autores, título, abstract, um *link* para o *download* do artigo, além de outras informações. Hoje, é preciso clicar no nome do professor no site do CIn e depois ir na sua página pessoal ou no Currículo Lattes para visualizar suas publicações e ainda assim de forma incompleta, pois de um modo geral só é mostrada a referência biográfica.

Um ponto importante é a nomenclatura das URIs da ontologia do CInPub. Seria interessante ter um padrão para nomear as URIs.

Além disso, a ideia é estender o CInPub com as informações dos discentes e obter um *dataset* completo das publicações que envolvam professores e alunos que estudaram no CIn..

6. Referências Bibliográficas

- [Adida & Birbeck 2008] Adida, B.; Birbeck, M. (2008). ***RDFa Primer — Bridging the Hu-man and Data Webs***. <http://www.w3.org/TR/xhtml-rdfa-primer/>.
- [Alani & Harith 2005] Alani, H.; Harris, S.; O’Neil, Ben (2005). **Ontology Winnowing: A Case Study on the AKT Reference Ontology**. In: International Conference on Computational Intelligence for Modelling, 28-30 November 2005, Vienna, Austria.<http://oro.open.ac.uk/20041/1/iawtic05-Alani.pdf>
- [Almeida 2009] Almeida, R. G. S. F. de (2009). **Integração de Informação por Migração em Sistemas Distribuídos e Heterogêneos**. Dissertação de Mestrado. Universidade de Trás-os-Monte e Alto Douro. Vila Real 2009. http://repositorio.utad.pt/bitstream/10348/311/1/msc_rgsfalmeida.pdf
- [Alves & Souza 2007] Alves, M. D. R; Souza, M. I. F. (2007). **Estudo de Correspondência de Elementos Metadados: Dublin Cores e Marc 21**. Revista Digital de Biblioteconomia e Ciência da Informação. Campinas 2007. 24. http://www.sbu.unicamp.br/seer/ojs/index.php/sbu_rci/article/viewFile/358/237
- [Battassini 2011] Battassini, R. (2011). **Uma ferramenta para busca temporal na DBPedia a partir de uma ontologia**. Trabalho de conclusão de graduação Universidade Federal do Rio Grande do Sul. Porto Alegre. 2011. 19-26 <http://www.lume.ufrgs.br/bitstream/handle/10183/36926/000819168.pdf?sequence=1>
- [Bizer *et al* 2009] Bizer, C.; Heath, T.; Lee, T. B. 2009. **Linked data - the story so far. Preprint to the special issue on linked data**. International Journal on Semantic Web and Information Systems (IJSWIS), 2009. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.
- [Brickley & Miller 2010] Brickley, D.; Libby, M. (2010). **FOAF Vocabulary Specification 0.98**. <http://xmlns.com/foaf/spec/>

- [Bulcao 2006] Bulcao, R. F. N. (2006). **Um processo de software e um modelo ontológico para apoio ao desenvolvimento de aplicações sensíveis a contexto**. Tese de doutorado. Instituto de Ciências Matemática e de Computação – ICMC-USP, 2006. 34. <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-01032007-163908/pt-br.php>
- [Covelinhas 2008] Covelinhas, E. M. M. (2008). **O processo de Bolonha na Web Semântica**. Tese de mestrado. Instituto Superior de Engenharia de Lisboa. Lisboa, Portugal, 2008 <http://www.slideshare.net/covelinhas/o-processo-de-bolonha-na-web-semantic>
- [Cunha *et al* 2012] Cunha, D. R. ; Souza, D. ; Lóscio, B. F. Linked Data: **da Web de Documentos para a Web de Dados**. In: III Escola Regional de Computação - Ceará, Maranhão e Piauí - ERCEMAPI 2011. : SBC, 2011.
- [Dublin Core 2012] Dublin Core Metadata Initiative 2012. [http:// www.dublincore.org](http://www.dublincore.org).
- [Filho & Lóscio 2010] Filho, F.W.B.H; Lóscio, B. F (2010) **Web Semântica: Conceitos e Tecnologias**. Minicurso apresentado Escola Regional de Computação Ceará - Maranhão – Piauí 2010.
- [Freitas, 2011] Freitas, G.C.de (2011). **Perspectiva e Análise de Viabilidade Sobre a Web Semântica**. Trabalho de conclusão de graduação. Universidade São Francisco. Campinas - São Paulo 2011. <http://www.scribd.com/doc/75824460/37/Logica-Prova-e-Confianca>
- [Frozza & Gonçalves 2010] Frozza, A. A.; Gonçalves, R. (2010). **Jena – Um Framework Web Semântico em Java**. Universidade de Santa Catarina. Florianópolis – Santa Catarina. 2010. <http://www.fgoncalves.trix.net/papers/goncalvesFrozzaJena.pdf>
- [Giasson & Raymond 2007] GIASSON, F.; RAYMOND, Y. 2007. **Music Ontology Specification**. 2007. <http://musicontology.com>

- [Gruber 1995] Gruber, T. R. (1995). *Toward principles for the design of ontologies used for knowledge sharing*. 1995. International Journal Human-Computer Studies. Vol. 43, Issues 5-6, November 1995, p.907-928.
- [Heath & Bizer 2011] Heath, T.; Bizer, C. (2011) *Linked Data: Evolving the Web into a Global Data Space (1st edition)*. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool, 2011.
- [Lee et al 2011] Lee, T. B.; Hendler, J.; Lassila, O. 2001. **The Semantic Web**. Scientific American, May 2001, p. 29-37.
- [Marcondes & Campos 2008] Marcondes, C. H.; Campo, M. L. de A. (2008). **Ontologia e Web Semântica: O Espaço da Pesquisa Em Ciências da Informação**. Ponto de Acesso. Salvador. v.2, n.1, 2008 107-136. <http://www.portalseer.ufba.br/index.php/revistaici/article/view/2669/1885>
- [Melo & Junior 2010] Melo, I. S. M.; Junior, J. A. S. V. (2010). **Recuperação da Informação de Serviços Odontológicos Baseada em Ontologias de Tarefa/Aplicação**. Trabalho de conclusão de graduação. Universidade Federal do Pará. Belém - Pará 2010. http://docs.google.com/fileview?id=0By0q-rYyZyETNzA2NjgyODctOTJiOS00NjA2LTlhYzYtMGY2ZTcwNzI4Y2Ri&hl=pt_BR&authkey=CLOE3eMK
- [Miles & Brickley 2005] Miles, A.; Brickley, D. (2005). **SKOS Core Vocabulary Specification. W3C Working Draft**. <<http://www.w3.org/TR/2005/WD-swbpskos-core-spec-20051102/>>
- [Miranda et al 2011] Miranda, D. S.; Azevedo, L. L. de S.; Magalhães, R. P. (2011). **Consumindo Linked Data na Web**. Encontro Unificado de Computação em Parnaíba 2011. Piauí.<http://www.enucomp.com.br/conteudos/artigos/linkedData.pdf>

- [Noy & McGuinness 2001] Noy, N. F.; McGuinness, D. (2001). **Ontology Development 101: A Guide to creating your first Ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, Stanford, 2001.
- [Pinheiros 2011] Pinheiros, J. C. (2011). **Processamento de Consulta em um Framework Baseado em Mediador para Integração de Dados no Padrão de Linked Data**. Tese de doutorado. Fortaleza – Ceará 2011.
http://www.mdcc.ufc.br/teses/doc_download/165-008-joao-carlos-pinheiro.
- [Ramalho *et al* 2007] Ramalho, R. A. S.; Vidotti, S. A. B. G.; Fujita, M. S. L. (2007). **Web Semântica: uma investigação sob o olhar da Ciências da Computação**. DataGramaZero – Revista de Ciências da Informação. – v-8 n6 2007.
http://www.dgz.org.br/dez07/Art_04.htm
- [Salgado & Lócio 2012] Salgado, A. C.; Lócio, B. F. **Web Semântica**. 2012.
- [Semprebom *et al* 2007] Tiago; Camada, M.; Medonça, I. (2007). **Protégé e Ontologias**. Universidade Federal de Santa Catarina. Santa Catarina 2007. 16.
http://www.das.ufsc.br/~gb/pg-ia/Protege07/ontologia_protege.pdf.
- [Souza *et al* 2008] Jairo, Sean Siqueira, Leonardo Azevedo, Fernanda Baião, Mauro Lopes, Flavia Santoro, Claudia Cappelli, Vanessa Nunes, Andréa Magdaleno. **Gestao de Ontologias**.
<http://www.seer.unirio.br/index.php/monografiasppgi/article/viewFile/225/222>
- [Vanni 2009] Vanni, R. M. P. (2009). **Integração de serviços em ambientes heterogêneos: uso de semântica para comunicação entre entidades em mudanças de contexto**. Biblioteca Digital – USP. 2009, 136-137-143-144.
<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-25012010-115143/pt-br.php>

[Xavier 2011] Xavier, O. C. (2011). **Serviços Web Semânticos Baseados em RESTul. Um Estudo de Caso em Redes Sociais Online.** INF - Instituto de Informática – UFG. Goiana – Goiás (2011). 29-31.
<http://www.inf.ufg.br/mestrado/sites/www.inf.ufg.br.mestrado/files/uploads/Dissertacoes/OtavioCala%C3%A7a.pdf>

[YAMAOKA 2007] Yamaoka, E. J. (2007). **Metadados e Dublin Core.** Seminário de Gestão da Informação Jurídica em Espaços Digitais. Brasília 2007.
<http://www.stf.jus.br/arquivo/sijed/05.pdf>

[W3C 2009] **Simple Knowledge Organization System (SKOS)** (2009). 13, 89.
<http://www.w3.org/2001/sw/wiki/SKOS>, 2009k.

APÊNDICE A – Códigos em RDF

Ao concluir a criação da ontologia, o protege gera um código na sintaxe RDF/XML automaticamente. Serão mostrados nesse anexo:

- ✓ *Namespaces* e vocabulários importados;
- ✓ As propriedades relacionais;
- ✓ As propriedades relacionais inversas;
- ✓ Classes que formam a ontologia do CInPub.

Namespaces e vocabulários importados

No cabeçalho do código RDF estão os *namespaces* dos vocabulários presentes na ontologia a fim de tornar o código mais enxuto e mais legível. A Tabela 17 mostra esses *namespaces*.

Tabela 17. Namespaces importados para o CInPub.

Namespaces
<pre> <?xml version=1.0?> <rdf:RDF xmlns:dcterms=http://www.domain2.com# xmlns:foaf=http://xmlns.com/foaf/0.1/ xmlns:swrl=http://www.w3.org/2003/11/swrl# xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema# xmlns:owl=http://www.w3.org/2002/07/owl# xmlns:akt=http://www.aktors.org/ontology/portal# xmlns:xsp=http://www.owl-ontologies.com/2005/08/07/xsp.owl# xmlns:swrlb=http://www.w3.org/2003/11/swrlb# xmlns:dcmitype=http://www.domain3.com# xmlns:j.0=http://purl.org/dc/terms/ xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns# xmlns:protege=http://protege.stanford.edu/plugins/owl/protege# xmlns:xsd=http://www.w3.org/2001/XMLSchema# xmlns:cin=http://www.owl-ontologies.com/cin.owl# xmlns:skos=http://www.w3.org/2004/02/skos/core# xmlns:p1=http://www.domain4.com# xml:base=http://www.owl-ontologies.com/cin.owl> </pre>

Os vocabulários que foram reusados precisaram ser importados para que fosse permitida a utilização de seus termos. Vale lembrar que alguns vocabulários foram importados automaticamente pelo *Protégé*. A seguir o trecho que faz essa importação.

Tabela 18. Ontologias importadas.

Importação dos vocabulários
<pre> <owl:Ontology rdf:about=> <owl:imports rdf:resource=http://xmlns.com/foaf/0.1/> <owl:imports rdf:resource=http://www.w3.org/2004/02/skos/core/> <owl:imports rdf:resource=http://www.aktors.org/ontology/portal/> <owl:imports rdf:resource=http://purl.org/dc/terms/> <owl:imports rdf:resource=http://purl.org/dc/elements/1.1/> <owl:imports rdf:resource=http://purl.org/dc/dcmitype/> </owl:Ontology> </pre>

Definição das propriedades relacionais

As propriedades relacionais são identificadas como *Object Property*, possuindo uma propriedade *rdf:domain*, que especifica para qual classe ela pertence, uma propriedade *rdfs:range*, que especifica o tipo do valor que essa propriedade recebe, e uma propriedade *owl:inverseOf* que informa que a propriedade em questão é a inversa de uma propriedade específica. As próximas tabelas trazem o código RDF dessas propriedades.

Tabela 19. Definição da propriedade *cin:temInteresse*.

Propriedade: <i>cin:temInteresse</i>
<pre> <owl:ObjectProperty rdf:about=#temInteresse> <owl:inverseOf rdf:resource=#ehInteresseDe/> <rdfs:range rdf:resource=#Area/> <rdfs:domain rdf:resource=#Professor/> </owl:ObjectProperty> </pre>

Tabela 20. Definição da propriedade *cin:ehInteresseDe*.

Propriedade inversa: <i>cin:ehInteresseDe</i>
<pre> <owl:ObjectProperty rdf:about=#ehInteresseDe> <owl:inverseOf rdf:resource=#temInteresse/> <rdfs:domain rdf:resource=#Area/> <rdfs:range rdf:resource=#Professor/> </owl:ObjectProperty> </pre>

Tabela 21. Definição da propriedade *cin:temEnderecoProfissional*.

Propriedade: <i>cin:temEnderecoProfissional</i>
<pre> <owl:ObjectProperty rdf:about=#temEnderecoProfissional> <owl:inverseOf> <owl:ObjectProperty rdf:ID=ehEnderecoProfissionalDe/> </owl:inverseOf> <rdf:type rdf:resource=http://www.w3.org/2002/07/owl#FunctionalProperty/> <rdfs:domain rdf:resource=#Professor/> <rdfs:range rdf:resource=#EnderecoProfissional/> </owl:ObjectProperty> </pre>

A propriedade *cin:temEnderecoProfissional* é do tipo funcional. Isso significa que todo professor deve ter um endereço profissional.

Tabela 22. Definição da propriedade *cin:ehEnderecoProfissionalDe*.

Propriedade inversa: <i>cin:ehEnderecoProfissionalDe</i>
<pre> <owl:ObjectProperty rdf:about=#ehEnderecoProfissionalDe> <rdfs:range rdf:resource=#Professor/> <rdf:type rdf:resource=http://www.w3.org/2002/07/owl#InverseFunctionalProperty/> <rdfs:domain rdf:resource=#EnderecoProfissional/> <owl:inverseOf rdf:resource=#temEnderecoProfissional/> </owl:ObjectProperty> </pre>

Tabela 23. Definição da propriedade *cin:ehPublicadoEm*.

Propriedade: <i>cin:ehPublicadaEm</i>
<pre> <owl:ObjectProperty rdf:about=#ehPublicadoEm> <owl:inverseOf rdf:resource=#armazena/> <rdfs:domain rdf:resource=#Publicacao/> <rdfs:range rdf:resource=#LocalDaPublicacao/> </owl:ObjectProperty> </pre>

Tabela 24. Definição da propriedade *cin:armazena*.

Propriedade inversa: <i>cin:armazena</i>
<pre> <owl:ObjectProperty rdf:ID=armazena> <rdfs:domain rdf:resource=#LocalDaPublicacao/> <owl:inverseOf> <owl:ObjectProperty rdf:about=#ehPublicadoEm/> </owl:inverseOf> <rdfs:range rdf:resource=#Publicacao/> </owl:ObjectProperty> </pre>

Tabela 25. Definição da propriedade *cin:escreve*.

Propriedade: <i>cin:escreve</i>
<pre> <owl:ObjectProperty rdf:about=#escreve> <owl:inverseOf rdf:resource=#ehEscritaPor/> <rdfs:domain rdf:resource=http://xmlns.com/foaf/0.1/Person/> <rdfs:range rdf:resource=#Publicacao/> </owl:ObjectProperty> </pre>

Tabela 26. Definição da propriedade *cin:ehEscritaPor*.

Propriedade inversa: <i>cin:ehEscritaPor</i>
<pre> <owl:ObjectProperty rdf:about=#ehEscritaPor> <owl:inverseOf> <owl:ObjectProperty rdf:about=#escreve/> </owl:inverseOf> <rdfs:range rdf:resource=http://xmlns.com/foaf/0.1/Person/> <rdfs:domain rdf:resource=#Publicacao/> </owl:ObjectProperty> </pre>

Propriedades do tipo *Datatype Property* também tiveram seus *domain* e *range* especificados. A Tabela 27 mostra o código RDF de algumas dessas propriedades.

Tabela 27. Definição de outras propriedades da ontologia.

Outras propriedades do CInPub
<pre> <rdf:Property rdf:about=http://purl.org/dc/terms/title> <rdfs:domain rdf:resource=#Publicacao/> </rdf:Property> <rdf:Property rdf:about=http://purl.org/dc/terms/creator> <rdfs:range rdf:resource=#Professor/> <rdfs:domain rdf:resource=#Publicacao/> </rdf:Property> <rdf:Property rdf:about=http://purl.org/dc/terms/contributor> <rdfs:range rdf:resource=#Aluno/> <rdfs:domain rdf:resource=#Publicacao/> </rdf:Property> <rdf:Property rdf:about=http://purl.org/dc/terms/description> <rdfs:domain rdf:resource=#Publicacao/> </rdf:Property> <owl:DatatypeProperty rdf:ID=nome> <rdfs:domain rdf:resource=#Area/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID=referenciaBiografica> <rdfs:domain rdf:resource=#Publicacao/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID=sala> <rdfs:domain rdf:resource=#Professor/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID=nomeCitacaoBiografica> <rdfs:domain rdf:resource=#Professor/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID=ano> <rdfs:domain rdf:resource=#Publicacao/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID=ramal> <rdfs:domain rdf:resource=#Professor/> </owl:DatatypeProperty> </pre>

Definição das classes

Cada uma das classes criadas na ontologia foi definida e atribuída algumas das propriedades RDFS e OWL. Nesta sessão serão mostrados os códigos RDF dessas classes.

Classe *cin:LocalDaPublicacao*

Já foi falado que essa classe é uma subclasse de *foaf:Organization*. Essa relação é mostrada no trecho da tabela Tabela 28:

Tabela 28. Definição da classe *cin:LocalDaPublicacao*.

Classe: <i>cin:LocalDaPublicacao</i>
<pre><rdfs:Class rdf:ID=LocalDaPublicacao> <rdfs:subClassOf rdf:resource=http://xmlns.com/foaf/0.1/Organization/> </rdfs:Class></pre>

Classe *cin:Area*

A classe *cin:Area* possui um propriedade *cin:ehInteresseDe* que pode receber como valor pelo menos um pro professor. Logo uma área precisa ser de interesse de pelo menos um professor. Essas restrições são encontradas no código RDF da classe mostrada na Tabela 29 através das propriedades *owl:minCardinality*.

Tabela 29. Definição da classe *cin:Area*.

Classe: <i>cin:Area</i>
<pre><owl:Class rdf:about=#Area> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:ID=ehInteresseDe/> </owl:onProperty> <owl:minCardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int >1</owl:minCardinality> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about=#ehInteresseDe/> </owl:onProperty> <owl:someValuesFrom rdf:resource=#Professor/> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf rdf:resource=http://www.w3.org/2002/07/owl#Thing/> </owl:Class></pre>

Classe *cin:EnderecoProfissional*

A classe *cin:EnderecoProfissional* possui a propriedade *cin:ehEnderecoProfissional*. Para essa propriedade é necessário que ao menos um professor possua como endereço uma instância da classe *cin:EnderecoProfissional*. No código RDF da classe as restrições são representadas pelas propriedades *owl:someValuesFrom* e *owl:minCardinality* e podem ser observadas na Tabela 30.

Tabela 30. Definição da classe *cin:EnderecoProfissional*.

Classe: <i>cin:EnderecoProfissional</i>
<pre> <owl:Class rdf:about=#EnderecoProfissional> <rdfs:subClassOf> <owl:Restriction> <owl:minCardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int >1</owl:minCardinality> <owl:onProperty> <owl:ObjectProperty rdf:ID=ehEnderecoProfissionalDe/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:someValuesFrom rdf:resource=#Professor/> <owl:onProperty> <owl:ObjectProperty rdf:about=#ehEnderecoProfissionalDe/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf rdf:resource=http://www.w3.org/2002/07/owl#Thing/> </owl:Class> </pre>

Classe *cin:Professor*

Essa classe também é uma subclasse da classe *foaf:Person*. Possui a propriedade *cin:temInteresse*, onde todos os valores que essa propriedade recebe é uma entidade do tipo *cin:Area* e o professor deve ter interesse em pelo menos uma área de pesquisa. Essas restrições é vista no código abaixo pelo uso das propriedades *owl:allValuesFrom* e *owl:someValuesFrom* e *owl:minCardinality*.

Outra propriedade que ela possui é a *cin:temEnderecoProfissional*. Esquema afirma que o professor só deve ter apenas um endereço profissional. Essa limitação é expressa no código pela propriedade *owl:cardinality* e tem como valor o numero 1 .

A propriedade *cin:escreve* pode ser descrita de forma semelhante à *cin:temInteresse*, aplicando as mesmas restrições para informar que todos os valores aceitos pela propriedade *cin:escreve* deve ser instancia de *cin:Publicacao* e que o professor escreve pelo menos uma publicação. A representação da classe *cin:Professor* é mostrada na Tabela 31.

Tabela 31. Definição da classe *cin:Professor*.

Classe: <i>cin:Professor</i>
<pre> <owl:Class rdf:ID=Professor> <rdfs:subClassOf> <owl:Restriction> <owl:someValuesFrom> <owl:Class rdf:ID=Area/> </owl:someValuesFrom> <owl:onProperty> <owl:ObjectProperty rdf:ID=temInteresse/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:FunctionalProperty rdf:ID=temEnderecoProfissional/> </owl:onProperty> <owl:cardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int >1</owl:cardinality> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:someValuesFrom> <owl:Class rdf:ID=Publicacao/> </owl:someValuesFrom> <owl:onProperty> <owl:ObjectProperty rdf:ID=escreve/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about=#escreve/> </owl:onProperty> <owl:allValuesFrom> <owl:Class rdf:about=#Publicacao/> </owl:allValuesFrom> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf rdf:resource=http://xmlns.com/foaf/0.1/Person/> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> </pre>

```

<owl:ObjectProperty rdf:about=#escreve/>
</owl:onProperty>
<owl:minCardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int
>1</owl:minCardinality>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Restriction>
<owl:allValuesFrom>
<owl:Class rdf:ID=EnderecoProfissional/>
</owl:allValuesFrom>
<owl:onProperty>
<owl:FunctionalProperty rdf:about=#temEnderecoProfissional/>
</owl:onProperty>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Restriction>
<owl:minCardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int
>1</owl:minCardinality>
<owl:onProperty>
<owl:ObjectProperty rdf:about=#temInteresse/>
</owl:onProperty>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty>
<owl:FunctionalProperty rdf:about=#temEnderecoProfissional/>
</owl:onProperty>
<owl:someValuesFrom rdf:resource=#EnderecoProfissional/>
</owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty>
<owl:ObjectProperty rdf:about=#temInteresse/>
</owl:onProperty>
<owl:allValuesFrom rdf:resource=#Area/>
</owl:Restriction>
</rdfs:subClassOf>
</owl:Class>

```

Classe *cin:Colaborador*

Essa classe é subclasse da classe *foaf:Person* e sua propriedade *cin:escreve* segue a mesma definição dessa propriedade em *cin:Professor*.

Tabela 32. Definição da classe *cin:Colaborador*.

Classe: <i>cin:Colaborador</i>
<pre> <owl:Class rdf:ID=Colaborador> <rdfs:subClassOf rdf:resource=http://xmlns.com/foaf/0.1/Person/> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about=#escreve/> </owl:onProperty> <owl:minCardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int >1</owl:minCardinality> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about=#escreve/> </owl:onProperty> <owl:allValuesFrom> <owl:Class rdf:about=#Publicacao/> </owl:allValuesFrom> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:someValuesFrom> <owl:Class rdf:about=#Publicacao/> </owl:someValuesFrom> <owl:onProperty> <owl:ObjectProperty rdf:about=#escreve/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> </owl:Class> </pre>

Classe *cin:Publicacao*

Essa classe apresenta a propriedade inversa *cin:ehEscritaPor* cujo o mínimo de pessoas que escreveu a publicação é um, ou seja, pelo menos uma pessoa escreveu cada publicação. Essa limitação pode ser vista no código abaixo a partir da propriedade *owl:minCardinality* que recebe um como valor. Além disso, a propriedade *cin:ehEscritaPor* só poderá receber como valor ou uma entidade professor ou uma entidade colaborador.

Outra propriedade que possui é *cin:ehPublicadoEm*, onde uma publicação é publicada em pelo menos um *cin:LocalDaPublicacao*. Essas restrições é mostrada no código

RDF abaixo através das propriedades *owl:minProperty*, *cin:someValuesFrom*, *cin:allValuesFrom*.

Tabela 33. Definição da classe *cin:Publicacao*.

Classe: <i>cin:Publicacao</i>
<pre> <owl:Class rdf:about=#Publicacao> <rdfs:subClassOf> <owl:Restriction> <owl:minCardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int >1</owl:minCardinality> <owl:onProperty> <owl:ObjectProperty rdf:ID=ehEscritaPor/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about=#ehEscritaPor/> </owl:onProperty> <owl:someValuesFrom rdf:resource=#Professor/> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:cardinality rdf:datatype=http://www.w3.org/2001/XMLSchema#int >1</owl:cardinality> <owl:onProperty> <owl:ObjectProperty rdf:ID=ehPublicadoEm/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:someValuesFrom rdf:resource=#Colaborador/> <owl:onProperty> <owl:ObjectProperty rdf:about=#ehEscritaPor/> </owl:onProperty> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about=#ehPublicadoEm/> </owl:onProperty> <owl:someValuesFrom rdf:resource=#LocalDaPublicacao/> </owl:Restriction> </rdfs:subClassOf> <rdfs:subClassOf> <owl:Restriction> <owl:onProperty> <owl:ObjectProperty rdf:about=#ehPublicadoEm/> </owl:onProperty> <owl:allValuesFrom rdf:resource=#LocalDaPublicacao/> </owl:Restriction> </pre>

```
</rdfs:subClassOf>  
<rdfs:subClassOf rdf:resource=http://www.w3.org/2002/07/owl#Thing/>  
</owl:Class>
```

APÊNDICE B – Visão da ontologia no *Protégé*

Visualizações das classes com suas propriedades e restrições e exemplos de instâncias.

Este anexo mostrará algumas telas com as informações das classes, de suas propriedades, das restrições dessas propriedades e exemplos de instâncias de cada classe.

Para as restrições, os símbolos, que são reconhecidos facilmente, representam o tipo da restrição utilizada para a propriedade.

Classe *cin:Area*

A Figura 12 mostra a restrição feita na propriedade *cin:ehIntereseDe* que diz que a área é de interesse de professor e que existe pelo menos um professor interessado nessa área.



Figura 12. Restrições da classe *cin:Area*.

As instâncias da classe foram criadas e são mostradas a seguir, exemplificando a instância Banco de Dados, pode-se perceber que essa é uma área de interesse dos professores: Ana Carolina, Bernadette Lóscio, Fernando Fonseca e Valéria Times.

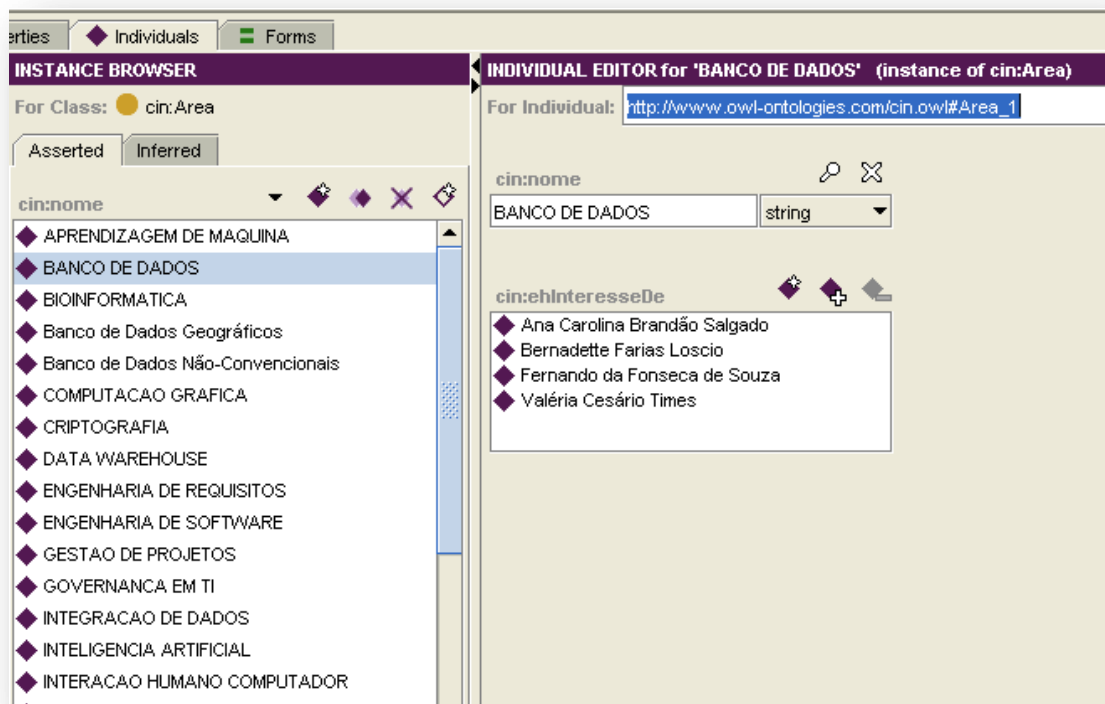


Figura 13. Instâncias da classe *cin:Area*.

Classe *cin:EnderecoProfissional*

A Figura 14 abaixo mostra a restrição feita na propriedade *cin:ehEnderecoProfissionalDe* para que possa receber como valor pelo menos uma instância da classe *cin:Professor*.

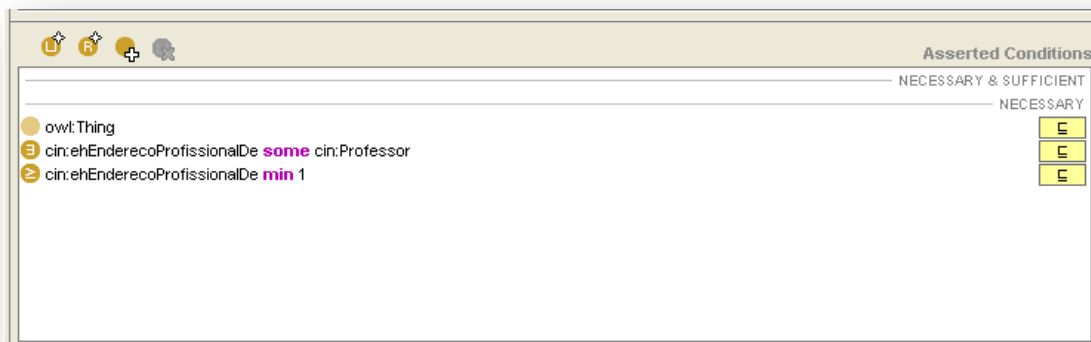


Figura 14. Restrições da classe *cin:EnderecoProfissional*.

Para o projeto, apenas um endereço profissional foi criado já que a priori todos os professores ensinam no mesmo lugar. A instância desse lugar é mostrada na Figura 15 com o nome do departamento, rua, número, telefone e outras informações além de mostrar de quais professores esse lugar é o endereço profissional.

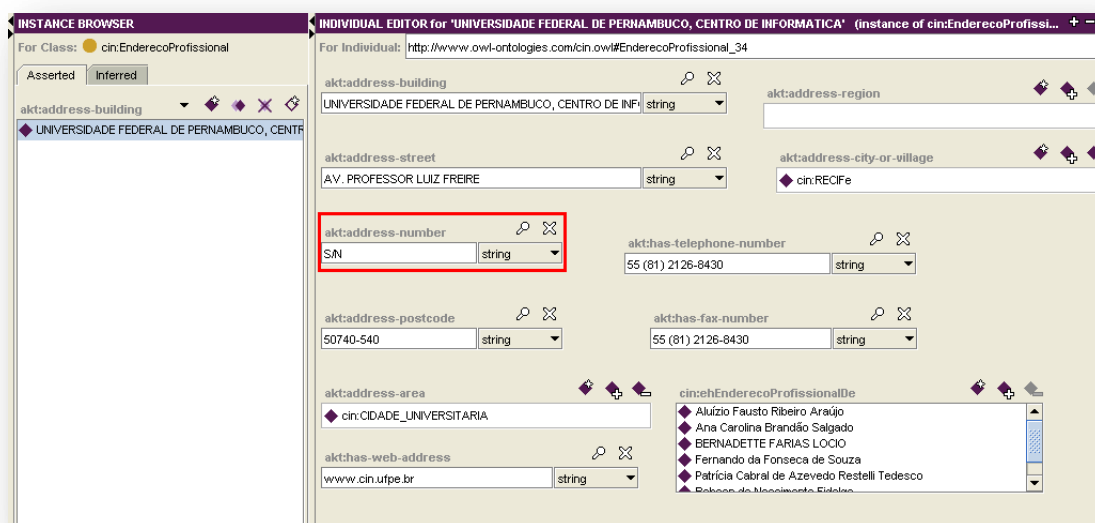


Figura 15. Instâncias da classe *cin:EnderecoProfissional*.

Classe *cin:LocalDaPublicacao*

As instâncias dessa classe representam um evento, jornal ou revista em que um artigo foi publicado. A Figura 16 mostra um evento que publicou o artigo *Support Tool in the Diagnosis of Sales Price of Dental Plainse* o site desse evento.

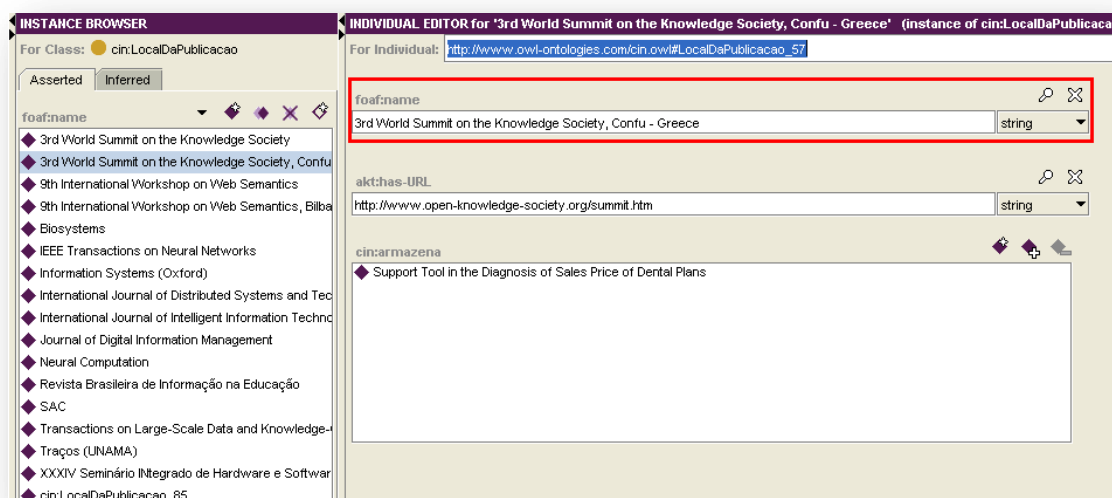


Figura 16. Instâncias da classe *cin:LocalDaPublicacao*.

Classe *cin:Colaborador*

A classe *cin:Colaborador* possui a propriedade *cin:Escreve* que restringe seu valor para publicação, ou seja, essa propriedade não pode receber outro valor que não seja uma publicação. Além dessa restrição, um colaborador escreve pelo menos uma publicação. Pode-se observar essas restrições na Figura 17.



Figura 17. Restrições da classe *cin:Colaborador*.

As instâncias dessa classe são mostradas a seguir na Figura 18, exemplificando a colaboradora Fabiana Freire que escreveu dois artigos com professores do CIn.

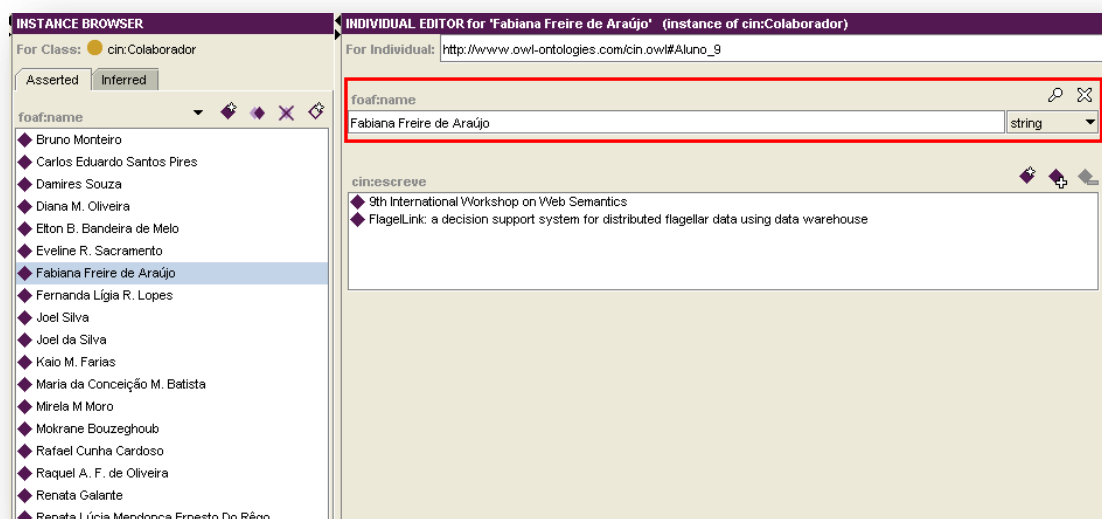


Figura 18. Instâncias da classe *cin:Colaborador*.

Classe *cin:Professor*

As restrições mostradas na Figura 19 foram feitas nas propriedades *cin:escreve*, que recebe como valor pelo menos uma publicação, *cin:temEnderecoProfissional*, que diz que um professor só pode ter um endereço profissional e *cin:temInteresse*, que diz que o professor pode se interessar por mais de uma área.

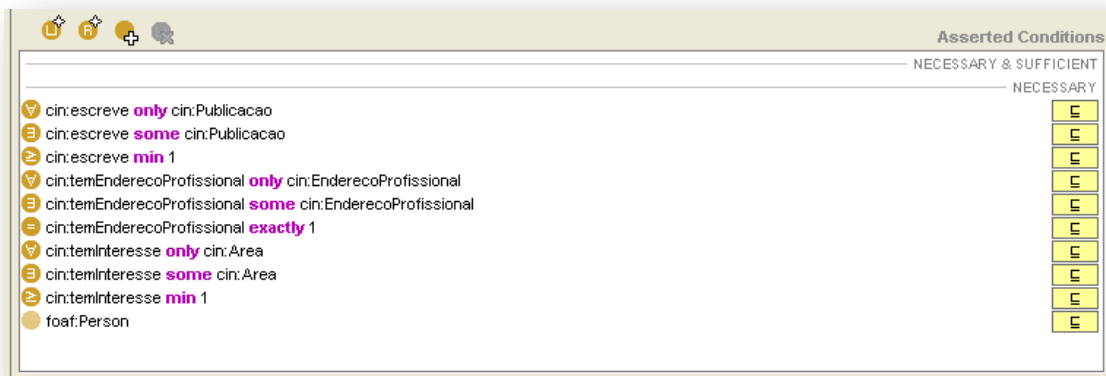


Figura 19. Restrições da classe *cin:Professor*.

Algumas instâncias de professor foram criadas para o CInPub e a Figura 20 mostra a professora Ana Carolina com suas informações pessoais, áreas de interesse, as publicações realizadas e os professores ou colaboradores que ela conhece.

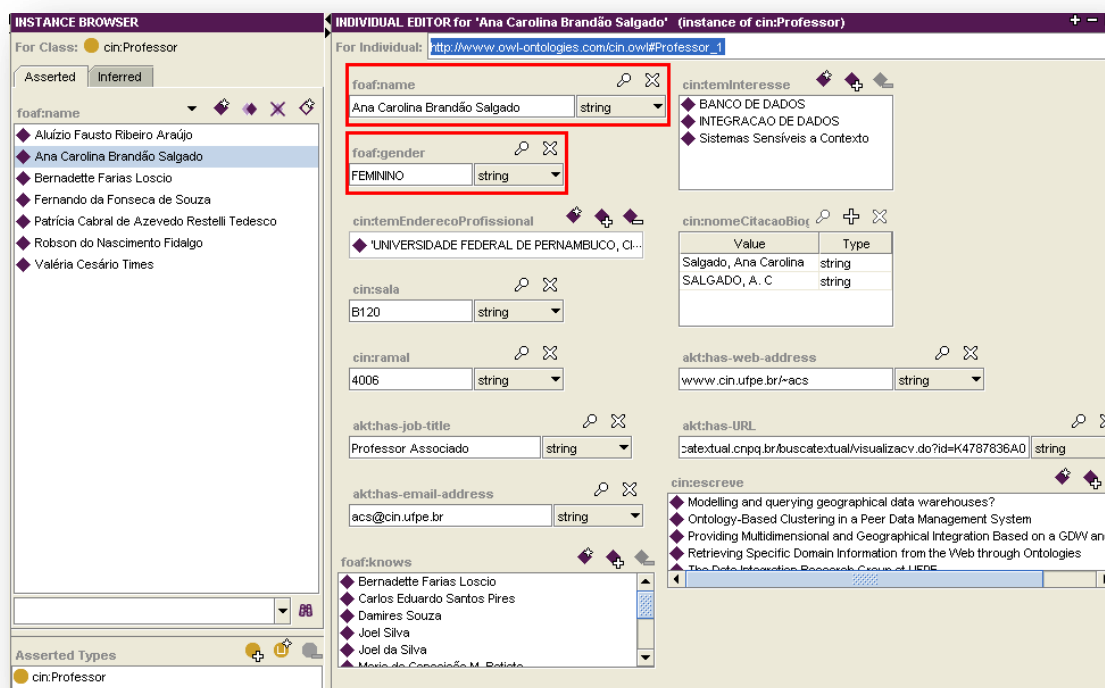


Figura 20. Instâncias da classe *cin:Professor*.

Classe *cin:Publicacao*

A classe *cin:Publicacao* possui a propriedade *cin:ehEscritoPor* que recebe como valor pelo menos uma instância de professor. Outra propriedade com restrição é *cin:PublicadoEm* que só pode receber apenas um valor e esse valor é uma instância de *cin:LocalPublicacao*. A Figura 21 representa essas restrições.



Figura 21. Restrições da classe *cin:Publicacao*.

As instâncias da classe são mostradas na Figura 22 exemplificando o artigo Amadeus-TV com informações do tipo título, referência biográfica, url de acesso e os professor e seus colaboradores.

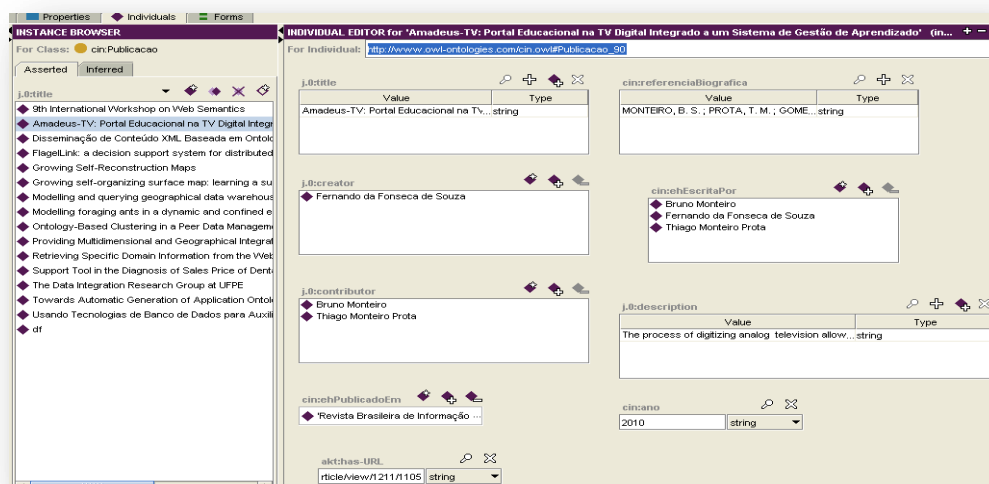


Figura 22. Instâncias da classe *cin:Publicacao*.

APÊNDICE C - Consultas em SPARQL sobre o dataset CInPub

As consultas em SPARQL sobre o dataset CInPub para responder as perguntas da seção 3.5.1.

Consulta 1

Pergunta: Quais foram as publicações de cada professor realizadas no ano de 2010?

Tabela 34. Consulta 1.

Consulta 1
<pre> PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc: <http://purl.org/dc/terms/> SELECT ?professorNome ?publicacaoTitulo WHERE { ?professor rdf:type cin:Professor . ?professor cin:escreve ?publicacao . ?professor foaf:name ?professorNome . ?publicacao cin:ano 2010 ^^<http://www.w3.org/2001/XMLSchema#string> . ?publicacao dc:title ?publicacaoTitulo . } ORDER BY ASC(?professorNome) </pre>

Tabela 35. Resultado da consulta 1.

Resultado da Consulta 1	
professorNome	publicacaoNome
Abel Guilhermino da Silva Filho	MOTION DETECTION OF VEHICLES BASED ON FPGA
Abel Guilhermino da Silva Filho	NSGAI Applied to Unified Second Level Cache Memory Hierarchy Tuning Aiming Energy and Performance Optimization
Aluizio Fausto Ribeiro Araujo	Growing Self-Reconstruction Maps
Aluizio Fausto Ribeiro Araujo	Growing self-organizing surface map: learning a surface topology from a point cloud
Ana Carolina Brandao Salgado	Modelling and querying geographical data warehouses?
Ana Carolina Brandao Salgado	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels
Bernadette Farias Loscio	MeMO: A clustering based approach for multiple ontology merging
Bernadette Farias Loscio	Support Tool in the Diagnosis of Sales Price of Dental Plans
Fernando da Fonseca de Souza	Amadeus-TV: Portal Educacional na TV Digital Integrado a um Sistema de Gestão de Aprendizado
Fernando da Fonseca de Souza	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels
Patricia Cabral de Azevedo Restelli Tedesco	The Gravitational Strategy for the Timed Patrolling
Patrícia Cabral de Azevedo Restelli Tedesco	Using a Metamodel to Design Structural and Behavioral Aspects in Context-Sensitive Groupware
Robson do Nascimento Fidalgo	Modelling and querying geographical data warehouses?
Robson do Nascimento Fidalgo	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels
Valéria Cesário Times	Modelling and querying geographical data warehouses?
Valéria Cesário Times	Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels

Consulta 2

Pergunta: Quais as publicações que tem como palavra-chave *ontologies* e onde podem ser encontradas?

Tabela 36. Consulta 2.

Consulta 2
<pre> PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc: <http://purl.org/dc/terms/> PREFIX akt: <http://www.aktors.org/ontology/portal#> SELECT ?professorNome ?publicacaoTitulo ?enderecoUrl WHERE{ ?professor rdf:type cin:Professor . ?professor cin:escreve ?publicacao . ?professor foaf:name ?professorNome . ?publicacao dc:title ?publicacaoTitulo . ?publicacao akt:has-URL ?enderecoUrl . ?publicacao cin:palavraChave ?publicacaoPalavraChave . FILTER(?publicacaoPalavraChave = Ontologies ^^<http://www.w3.org/2001/XMLSchema#string> ?publicacaoPalavraChave = ontologies ^^<http://www.w3.org/2001/XMLSchema#string>) } ORDER BY ASC(?professorNome) </pre>

Tabela 37. Resultado da consulta2.

Resultado da Consulta 2

ProfessorNome	PublicacaoTitulo	enderecoUrl
Ana Carolina Brandao Salgado	The Data Integration Research Group at UFPE	http://seer.lcc.ufmg.br/index.php/jidm/article/view/121/74
Ana Carolina Brandao Salgado	Retrieving Specific Domain Information from the Web through Ontologies	http://www.irma-international.org/viewtitle/2405/
Ana Carolina Brandao Salgado	Ontology-Based Clustering in a Peer Data Management System	http://www.igi-global.com/viewtitlesample.aspx?id=66054
Bernadette Farias Loscio	An incremental and user feedback-based ontology matching approach	http://delivery.acm.org/10.1145/2100000/2095607/p371-wagner.pdf?ip=150.161.2.200&acc=ACTIVE%20SERVICE&CFID=95685376&CFTOKEN=91487570&__acm__=1342275142_c5b713812e0cae f2f6bc85088cc4e76c
Bernadette Farias Loscio	Towards Automatic Generation of Application Ontologies	http://webscience.org.br/wiki/images/d/d1/Casanova_2010_ICEIS.pdf
Bernadette Farias Loscio	The Data Integration Research Group at UFPE	http://seer.lcc.ufmg.br/index.php/jidm/article/view/121/74
Fernando da Fonseca de Souza	Retrieving Specific Domain Information from the Web through Ontologies	http://www.irma-international.org/viewtitle/2405/

Consulta 3

Pergunta: Qual sala do CIn se encontram os professores? Qual é seu endereço pessoal e Currículo Lattes ?

Tabela 38. Consulta 3.

Consulta 3
<pre> PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX akt: <http://www.aktors.org/ontology/portal#> SELECT ?professorNome ?professorSala ?professorPagina ?professorCurriculo WHERE { ?professor rdf:type cin:Professor . ?professor foaf:name ?professorNome . ?professor cin:sala ?professorSala . ?professor akt:has-web-address ?professorPagina . ?professor akt:has-URL ?professorCurriculo . } ORDER BY ?professorNome </pre>

Tabela 39. Resultado da consulta 3.

Resultado da Consulta 3			
professorNome	Profess orSala	ProfessorPagina	professorCurriculo
Abel Guilhermino da Silva Filho	B005	www.cin.ufpe.br/~agsf	http://buscatextual.cnpq.br/ buscatextual/visualizacv.do ?id=K4792886T0
Aluizio Fausto Ribeiro Araújo	C128	http://www.cin.ufpe.br/~al uizioa/aluizio- homepage.htm	http://buscatextual.cnpq.br/ buscatextual/visualizacv.do ?id=K4787817Y8
Ana Carolina Brandao Salgado	B120	www.cin.ufpe.br/~acs	http://buscatextual.cnpq.br/ buscatextual/visualizacv.do ?id=K4787836A0
Bernadette Farias Loscio	B001	www.cin.ufpe.br/~bfl	http://lattes.cnpq.br/251206 4355660153
Patricia Cabral de Azevedo Restelli Tedesco	B110	www.cin.ufpe.br/~pcart	http://buscatextual.cnpq.br/ buscatextual/visualizacv.do ?id=K4728355A3
Robson do Nascimento Fidalgo	C024	www.cin.ufpe.br/~rdnf	http://buscatextual.cnpq.br/ buscatextual/visualizacv.do ?id=K4705039A6
Valéria Cesário Times	C103	www.cin.ufpe.br/~vct	http://buscatextual.cnpq.br/ buscatextual/visualizacv.do ?id=K4791983P6

Consulta 4

Pergunta: Quais são as áreas de interesse de cada professor do CIn?

Tabela 40. Consulta 4.

Consulta 4
PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc: <http://purl.org/dc/terms/> SELECT ?professorNome ?areaInteresseNome WHERE { ?professor rdf:type cin:Professor . ?professor foaf:name ?professorNome . ?professor cin:temInteresse ?areaInteresse . ?areaInteresse cin:nome ?areaInteresseNome . } ORDER BY ?professorNome

Tabela 41. Resultado da consulta 4.

Resultado da Consulta 4	
ProfessorNome	areaInteresseNome
Abel Guilhermino da Silva Filho	Algoritmos Inteligentes para Exploração de Arquitetura
Abel Guilhermino da Silva Filho	SISTEMAS EMBARCADOS
Abel Guilhermino da Silva Filho	Sistemas Reconfiguráveis
Aluizio Fausto Ribeiro Araújo	REDES NEURAIAS
Aluizio Fausto Ribeiro Araújo	Sistemas dinâmicos
Aluizio Fausto Ribeiro Araújo	robotica
Ana Carolina Brandao Salgado	BANCO DE DADOS
Ana Carolina Brandao Salgado	INTEGRACAO DE DADOS
Ana Carolina Brandao Salgado	Sistemas Sensíveis a Contexto
Bernadette Farias Loscio	BANCO DE DADOS
Bernadette Farias Loscio	INTEGRACAO DE DADOS
Bernadette Farias Loscio	WEB SEMANTICA
Bernadette Farias Loscio	XML
Fernando da Fonseca de Souza	BANCO DE DADOS
Fernando da Fonseca de Souza	Interação Homem-Máquina
Fernando da Fonseca de Souza	educacao a distancia
Patrícia Cabral de Azevedo Restelli Tedesco	INTELIGENCIA ARTIFICIAL
Patrícia Cabral de Azevedo Restelli Tedesco	Sistemas Multi-Agentes
Patrícia Cabral de Azevedo Restelli Tedesco	educacao a distancia
Patrícia Cabral de Azevedo Restelli Tedesco	informatica na educacao
Robson do Nascimento Fidalgo	DATA WAREHOUSE
Robson do Nascimento Fidalgo	Integração de Sistemas
Robson do Nascimento Fidalgo	BANCO DE DADOS
Valéria Cesário Times	Banco de Dados Não-Convencionais
Valéria Cesário Times	MINERACAO DE DADOS
Valéria Cesário Times	Sistemas de Suporte à Decisão

Consulta 5

Pergunta: Quais alunos do CIn e com quais professores esses alunos publicaram algum artigo?

Tabela 42. Consulta 5.

Consulta 5
PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc: <http://purl.org/dc/terms/> SELECT DISTINCT ?alunoNome ?professorNome WHERE {

```

?aluno rdf:type cin:Colaborador .
?professor rdf:type cin:Professor .
?aluno cin:escreve ?publicacao .
?professor cin:escreve ?publicacao .
?aluno foaf:name ?alunoNome .
?professor foaf:name ?professorNome .
}
ORDER BY ?alunoNome

```

Tabela 43. Resultado da consulta5.

Resultado da Consulta 5	
AlunoNome	professorNome
Bruno Monteiro	Fernando da Fonseca de Souza
Carlos Eduardo Santos Pires	Ana Carolina Brandao Salgado
Damires Souza	Ana Carolina Brandao Salgado
Diana M. Oliveira	Bernadette Farias Loscio
Elton B. Bandeira de Melo	Aluizio Fausto Ribeiro Araújo
Eveline R. Sacramento	Bernadette Farias Loscio
Fabiana Freire de Araújo	Bernadette Farias Loscio
Fernanda Lígia R. Lopes	Bernadette Farias Loscio
Fernando Wagner	Bernadette Farias Loscio
Filipe Rolim Cordeiro	Abel Guilhermino da Silva Filho
Joel da Silva	Ana Carolina Brandao Salgado
Joel da Silva	Robson do Nascimento Fidalgo
Joel da Silva	Valéria Cesário Times
Jose A. F. Macedo	Bernadette Farias Loscio
Kaio M. Farias	Bernadette Farias Loscio
Maria da Conceição M. Batista	Ana Carolina Brandao Salgado
Maria da Conceição M. Batista	Bernadette Farias Loscio
Mirela M Moro	Bernadette Farias Loscio
Mokrane Bouzeghoub	Ana Carolina Brandao Salgado
Rafael Cunha Cardoso	Ana Carolina Brandao Salgado
Rafael Cunha Cardoso	Fernando da Fonseca de Souza
Raquel A. F. de Oliveira	Bernadette Farias Loscio
Renata Galante	Bernadette Farias Loscio
Renata Lúcia Mendonça Ernesto Do Rêgo	Aluizio Fausto Ribeiro Araújo
Rocir Marcos Leite Santiago	Ana Carolina Brandao Salgado
Rogério Plácido Pinheiros	Bernadette Farias Loscio
Thiago Costa	Ana Carolina Brandao Salgado
Thiago Costa	Bernadette Farias Loscio
Thiago Monteiro Prota	Fernando da Fonseca de Souza
Vilson Luiz Dalle Mole	Aluizio Fausto Ribeiro Araújo
Vânia M. P. Vidal	Bernadette Farias Loscio
Vânia M. P. Vidal	Ana Carolina Brandao Salgado
Zoubida Kedad	Ana Carolina Brandao Salgado
Ângela M. A. Pinheiro	Bernadette Farias Loscio

Consulta 6

Pergunta: Quantas publicações foram realizadas pelos professores do CIn em 2010?

Tabela 44. Consulta 6.

Consulta 6
<pre> PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc: <http://purl.org/dc/terms/> SELECT COUNT(?publicacao) WHERE{ ?publicacao rdf:type cin:Publicacao . ?publicacao cin:ano 2010 ^^<http://www.w3.org/2001/XMLSchema#string> . }</pre>

Tabela 45. Resultado da consulta 6.

Resultado da Consulta 6
<pre> 11 ^^ http://www.w3.org/2001/xmlSchema#integer</pre>

Consulta 7

Pergunta: Em que eventos estão as publicações da professora Bernadette Farias Lóscio?

Tabela 46. Consulta 7.

Consulta 7
<pre> PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX akt: <http://www.aktors.org/ontology/portal#> SELECT DISTINCT ?publicacaoURL WHERE{ ?professor foaf:name Bernadette Farias Loscio ^^<http://www.w3.org/2001/XMLSchema#string> . ?publicacao rdf:type cin:Publicacao . ?professor cin:escreve ?publicacao . ?publicacao akt:has-URL ?publicacaoURL . }</pre>

Tabela 47. Resultado da consulta 7.

Resultado da Consulta 7
localPublicacaoNome
3rd World Summit on the Knowledge Society, Confu – Greece
9th International Workshop on Web Semantics, Bilbao – Espanha
International Conference on Information Integration and Web-based Applications and Services, 2011, Ho Chi Minh City
Journal of Digital Information Management
Journal of the Brazilian Computer Society, Porto Alegre
SAC
The 21st Annual ACM Symposium on Applied Computing, 2006, Dijon
XXXIV Seminário INtegrado de Hardware e Software

Consulta 8

Pergunta: Quais são as publicações que Bernadette Loscio e Ana Carolina escreveram juntas?

Tabela 48. Consulta 8.

Consulta 8
<pre> PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc: <http://purl.org/dc/terms/> SELECT ?publicacaoTitulo WHERE{ ?publicacao rdf:type cin:Publicacao . ?professor1 cin:escreve ?publicacao . ?professor2 cin:escreve ?publicacao . ?publicacao dc:title ?publicacaoTitulo . FILTER (?professor1=cin:Professor_47 && ?professor2 = cin:Professor_1) } </pre>

Nesta consulta foi utilizado a URI das professoras ao invés dos seus nome. O identificador da professora Ana Carolina é *cin:Professor_1* e o da professora Bernadette é *cin:Professor_47*.

Tabela 49. Resultado da consulta 8.

Resultado
publicacaoTitulo
Using Agents for Generation and Maintenance of Mediators
Query reformulation for an XML-based Data Integration System
The Data Integration Research Group at UFPE

Consulta 9

Pergunta: Como citar a referência bibliográfica das publicações do professor Fernando Fonseca em um documento?

Tabela 50. Consulta 9.

Consulta 9
PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dc: <http://purl.org/dc/terms/> SELECT ?publicacaoTitulo ?publicacaoReferencia WHERE { ?publicacao rdf:type cin:Publicacao . ?professor rdf:type cin:Professor . ?publicacao cin:referenciaBiografica ?publicacaoReferencia . ?professor cin:escreve ?publicacao . ?publicacao dc:title ?publicacaoTitulo . FILTER (?professor = cin:Professor_43) }

Tabela 51. Resultado da consulta 9.

Resultado da Consulta 9	
PublicacaoTitulo	publicacaoReferencia
Amadeus-TV: Portal Educacional na TV Digital Integrado a um Sistema de Gestão de Aprendizado	MONTEIRO, B. S. ; PROTA, T. M. ; GOMES, A. S. ; SOUZA, F. F. . Amadeus-TV: Portal Educacional na TV Digital Integrado a um Sistema de Gestão de Aprendizado. Revista Brasileira de Informação na Educação, v. 18, p. 5-16, 2010.
Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels	FIDALGO, R. N. ; TIMES, V. C. ; SILVA, J. ; SOUZA, F. F. ; SALGADO, A. C. . Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels. Journal of Information and Data Management, v. 1, p. 93-106,
Usando Tecnologias de Banco de Dados para Auxiliar Tomada de Decisão.	FIDALGO, R. N. ; ROBIN, J. L. P. ; SOUZA, F. F. . Usando Tecnologias de Banco de Dados para Auxiliar Tomada de Decisão. Traços (UNAMA), v. 2, p. 78-88, 1999.
Retrieving Specific Domain Information from the Web through Ontologies	CARDOSO, R. C. ; SOUZA, F. F. ; SALGADO, A. C. . Retrieving Specific Domain Information from the Web through Ontologies. International Journal of Intelligent Information Technologies, 2006.

Consulta 10

Pergunta: As informações referentes às publicações do Professor Aluísio realizadas em 2011?

Tabela 52. Consulta 10.

Consulta 10
<pre>PREFIX cin: <http://www.owl-ontologies.com/cin.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> SELECT ?predicadoPublicacao ?objetoPublicacao WHERE{ ?publicacao rdf:type cin:Publicacao . ?publicacao ?predicadoPublicacao ?objetoPublicacao . ?professor rdf:type cin:Professor . ?professor cin:escreve ?publicacao . ?publicacao cin:ano 2011^^<http://www.w3.org/2001/XMLSchema#string> . FILTER (?professor = cin:Professor_41) }</pre>

Tabela 53. Resultado da consulta 10.

Resultado da Consulta 10	
predicadoPublicacao	objetoPublicacao
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.owl-ontologies.com/cin.owl#Publicacao>
<http://www.owl-ontologies.com/cin.owl#ehPublicadoEm>	<http://www.owl-ontologies.com/cin.owl#LocalDaPublicacao_98>
<http://www.owl-ontologies.com/cin.owl#ehEscritaPor>	<http://www.owl-ontologies.com/cin.owl#Professor_41>
<http://www.owl-ontologies.com/cin.owl#ehEscritaPor>	<http://www.owl-ontologies.com/cin.owl#Aluno_99>
<http://purl.org/dc/terms/title>	Modelling foraging ants in a dynamic and confined environment
<http://purl.org/dc/terms/creator>	<http://www.owl-ontologies.com/cin.owl#Professor_41>
<http://purl.org/dc/terms/contributor>	<http://www.owl-ontologies.com/cin.owl#Aluno_99>
<http://purl.org/dc/terms/description>	In social insects, the superposition of simple individual behavioral rules leads to the emergence of complex collective patterns and helps solve difficult problems inherent to surviving in hostile habitats. Modelling ant colony foraging reveals strategies arising from the insects' self-organization and helps develop of new computational strategies in order to solve complex problems. This paper presents advances in modelling ants' behavior when foraging in a confined and dynamic environment, based on experiments with the Argentine ant <i>Linepithema humile</i> in a relatively complex artificial network. We propose a model which overcomes the problem of stagnation observed in earlier models by taking into account additional biological aspects, by using non-linear functions for the deposit, perception and evaporation of pheromone, and by introducing new mechanisms to represent randomness and the exploratory behavior of the ants.
<http://www.owl-ontologies.com/cin.owl#referenciaBiografica>	Bandeira de Melo, Elton B. ; Araújo, Aluizio F.R. . Modelling foraging ants in a dynamic and confined environment. Biosystems (Amsterdam. Print), p. 23-31, 2011.
<http://www.owl-ontologies.com/cin.owl#palavraChave>	Stigmergy
<http://www.owl-ontologies.com/cin.owl#palavraChave>	Ant algorithms
<http://www.owl-ontologies.com/cin.owl#palavraChave>	Swarm intelligence
<http://www.owl-ontologies.com/cin.owl#ano>	2011
<http://www.aktors.org/ontology/portal#has-URL>	http://www.sciencedirect.com/science/article/pii/S030326471000239X