

UNIVERSIDADE FEDERAL DO PAMPA

Romário Laltany Goulart da Silva

**Nanopublicação Aplicada ao Domínio de  
Estudos de Análise de Redes de  
Colaboração Científica**

Alegrete  
2019



Romário Laltany Goulart da Silva

# Nanopublicação Aplicada ao Domínio de Estudos de Análise de Redes de Colaboração Científica

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal do Pampa como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr<sup>a</sup> Andréa Sabedra Bordin

Alegrete  
2019



Romário Laltany Goulart da Silva

## Nanopublicação Aplicada ao Domínio de Estudos de Análise de Redes de Colaboração Científica

Trabalho de Conclusão de Curso apresentado  
ao Curso de Graduação em Ciência da Com-  
putação da Universidade Federal do Pampa  
como requisito parcial para a obtenção do tí-  
tulo de Bacharel em Ciência da Computação.

Trabalho de Conclusão de Curso defendido e aprovado em 25 de Novembro de 2019  
Banca examinadora:

*Andréa S. Bordin*

---

Prof. Dr<sup>a</sup> Andréa Sabedra Bordin

Orientador  
UNIPAMPA

---

Prof. Dr. João Pablo Silva da Silva

UNIPAMPA

PP/ *Alexandre L. Gonçalves*

---

Prof. Dr. Alexandre Leopoldo Gonçalves  
UFSC



Este trabalho é dedicado às crianças adultas que,  
quando pequenas, sonharam em se tornar cientistas.





## AGRADECIMENTOS

A instituição UNIPAMPA pelo ambiente criativo e amigável que proporciona. À Prof<sup>a</sup>. Dr<sup>a</sup>. Andréa Sabedra Bordin pelo suporte, apoio, paciência, incentivo e empenho dedicado ao desafio proporcionado a nós diante a elaboração deste trabalho. Deixo meus agradecimentos a todos os docentes que me proporcionaram o conhecimento não apenas racional, mas a manifestação do caráter e efetividade da educação no processo de formação acadêmico e profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender. A palavra mestre nunca fará justiça aos professores dedicados aos quais sem nominar terão meus eternos agradecimentos.

A minha família por me incentivar em todos os momentos, e em especial deixo também minha eterna gratidão a minha mãe Maria Velaine Goulart da Silva Unamuzaga e irmã Luiza Vandrielli Goulart Unamuzaga, minhas heroínas que não apenas me incentivaram e apoiaram minhas escolhas, mas me deram suporte para superar todo o desânimo e cansaço que se puseram como pedras durante o caminhar desta jornada. Meus agradecimentos aos amigos que obtive durante o processo de graduação, que sempre estiveram ao meu lado não somente nos momentos de risos, mas me apoiando e não deixando abater-me. Em especial, gratidão aos meus companheiros Luana Cadaval, Paula Cordeiro, Alex Chervenski, Murilo Meglioransi, Paloma Ribeiro, Bolivar Teixeira, Rafael Fogliato e Amanda Oliveira.

Por fim, deixo registrado meus agradecimentos a todos que fizeram parte da minha formação e que vão continuar, com certeza, presentes na minha vida.



“Não vos amoldeis às estruturas deste mundo,  
mas transformai-vos pela renovação da mente,  
a fim de distinguir qual é a vontade de Deus:  
o que é bom, o que Lhe é agradável, o que é perfeito.  
(Bíblia Sagrada, Romanos 12:2)



## RESUMO

Portais de pesquisas são ferramentas muito utilizadas por pesquisadores para compartilhar e buscar resultados de estudos científicos. No entanto, com o passar do tempo, se obteve uma grande quantidade de conteúdos não estruturados disponíveis nesses repositórios, ocasionando um dos principais problemas enfrentados na recuperação de informações, a dificuldade da utilização de abordagens computacionais para a recuperação, extração e o reuso das informações contidas nos repositórios da web. Por conta disso, o uso de artigos completos talvez não seja a melhor forma de compartilhar e permitir o acesso ao conhecimento, visto que o que realmente importa são pequenos trechos de afirmações provindas de uma pesquisa, ou seja, os resultados obtidos na mesma. Para solucionar este problema foi proposto a nanopublicação, uma abordagem que utiliza tecnologias semânticas para publicar resultados de pesquisas em um nível de granularidade menor, mantendo informações sobre a sua proveniência e facilitando a interpretação computacional. Neste trabalho é proposto o uso de nanopublicações aplicadas ao domínio de Análise de Redes de Colaboração Científica. Para isso foi criada uma ontologia, através da metodologia Neon, capaz de representar os conceitos do domínio de análise de redes. A ontologia serviu para guiar a extração de informações que foram utilizadas para a criação de nanopublicações deste domínio. Criou-se também uma aplicação semântica para a recuperação de nanopublicações. A aplicação foi avaliada por uma especialista no domínio, que verificou a viabilidade da abordagem aplicada nos estudos de Análise de Rede de Colaboração Científicas. Por meio da utilização da aplicação, foi possível recuperar resultados específicos de estudos de redes de colaboração científica de forma ágil, com a garantia da credibilidade e relevância das afirmações.

**Palavras-chave:** Nanopublicações. Tecnologias Semânticas. Web Semântica.



## ABSTRACT

Research journals are efficient and powerful tools widely used by researchers to spread and share scientific studies. In recent years, due the huge amount of non-structured content available it becomes harder to use computational methods to recover, extract and reutilise all the information that can be found on these repositories. It arises the question if the use of the full text of a scientific paper still the best way to share knowledge, once the information that people seek, usually, can be found in small paragraphs within the full text. Nanopublications appears as a feasible way to tackle this problem, that uses semantics technology , in order to organise and facilitate the navigation and computational work, as the scientific results are highlighted and all the data is well structured. In this work, is proposed the use of nanopublications applied to Scientific Collaboration Network Analysis. For this, an ontology was created, through the Neon methodology, capable of representing the concepts of the network analysis domain. Ontology served to guide the extraction of information used to create nanopublications of this domain. A semantic application for nanopublishing recovery has also been created. The application was evaluated by an expert in the field, who verified the feasibility of the approach applied in the Scientific Collaboration Network Analysis studies. With the application, it was possible to quickly retrieve specific results from studies of scientific collaboration networks, ensuring the credibility and relevance of the statements.

**Key-words:** Nanopublication. Semantic Web. Semantics technologies.





## LISTA DE FIGURAS

Figura 1 – Uma possível arquitetura para a Web Semântica. . . . .	32
Figura 2 – Grafo RDF . . . . .	34
Figura 3 – Grafo representando uma tripla Resource Description Framework (RDF) . . . . .	34
Figura 4 – Código do grafo da Figura 3 . . . . .	35
Figura 5 – Aplicação do esquema RDF no código do grafo da Figura 3 . . . . .	35
Figura 6 – Cenários e fluxos de atividades . . . . .	40
Figura 7 – Representação dos gráficos das três situações . . . . .	41
Figura 8 – A estrutura do <i>Ontology Web Language</i> (OWL) . . . . .	45
Figura 9 – Codificação do Cabeçalho. . . . .	47
Figura 10 – Representação de indivíduos. . . . .	47
Figura 11 – Representação em código da frase: Todo gato é um animal. . . . .	48
Figura 12 – Exemplo de propriedade de objeto . . . . .	49
Figura 13 – Objeto . . . . .	52
Figura 14 – Consulta SPARQL . . . . .	52
Figura 15 – Resultado obtido através da consulta da figura 14 . . . . .	52
Figura 16 – Exemplo de uma unidade mínima de afirmação . . . . .	54
Figura 17 – Exemplo de proveniência da asserção . . . . .	55
Figura 18 – Exemplo de Proveniência da Nanopublicação . . . . .	55
Figura 19 – Arquitetura mínima da nanopublicação . . . . .	55
Figura 20 – Exemplo de Rede de Atores . . . . .	56
Figura 21 – Processo do mapeamento . . . . .	61
Figura 22 – Frequência de publicação anual . . . . .	64
Figura 23 – Quantidade de publicações por domínio . . . . .	64
Figura 24 – Quantidade de autores por país . . . . .	65
Figura 25 – Rede de Atores . . . . .	66
Figura 26 – Gráfico de dispersão com dados referentes ao mapeamento entre as palavras-chaves e as categorias de artigos . . . . .	68
Figura 27 – Gráfico de dispersão com dados referentes ao mapeamento entre as palavras-chaves principais e as palavras-chave de variedade de contexto . . . . .	70
Figura 28 – Processo das atividades realizadas para a concretização do trabalho . . . . .	73
Figura 29 – Cenário 1: Tarefas Básicas de criação de uma ontologia . . . . .	75
Figura 30 – Tarefas da atividade de especificação dos requisitos da ontologia. . . . .	76
Figura 31 – Tela de palavras-chave e suas frequências . . . . .	82
Figura 32 – Atividades do planejamento do projeto. . . . .	83
Figura 33 – Árvore de Decisão de modelo de ciclo de vida da Ontologia. . . . .	84
Figura 34 – Árvore de decisão. . . . .	85
Figura 35 – Modelo de ciclo de vida cascata de 5 fases. . . . .	86

Figura 36 – Cronograma do projeto . . . . .	87
Figura 37 – Escolha do tipo de reuso de recursos ontológicos. . . . .	88
Figura 38 – Processo de reuso de módulos de ontologias. . . . .	90
Figura 39 – Arquitetura da ontologia SNAMetric. . . . .	92
Figura 40 – Arquitetura da ontologia OntoSELF, representada em diagrama de classes UML. . . . .	93
Figura 41 – Arquitetura da ontologia SemSNA. . . . .	94
Figura 42 – Processo de reuso de ontologias gerais. . . . .	96
Figura 43 – Fluxo de tarefas para a conceituação do conhecimento do domínio. . . . .	100
Figura 44 – 1º nível da taxonomia de conceitos do domínio de EARCC. . . . .	101
Figura 45 – Taxonomia doos conceitos do eixo Types . . . . .	101
Figura 46 – Taxonomia dos conceitos de EdgesNumber . . . . .	102
Figura 47 – Taxonomia dos conceitos de ActorsNumber . . . . .	102
Figura 48 – Taxonomia do conceito SCNAS . . . . .	102
Figura 49 – Taxonomia dos conceitos do eixo Metrics . . . . .	102
Figura 50 – Taxonomia dos conceitos de Time . . . . .	103
Figura 51 – Taxonomia do conceito Network . . . . .	103
Figura 52 – Taxonomia do eixo Actor . . . . .	103
Figura 53 – Taxonomia doos conceitos do eixo Location . . . . .	103
Figura 54 – Diagrama de Relações binárias ad hoc . . . . .	104
Figura 55 – Modelo Conceitual do SCNAS <i>Ontology</i> . . . . .	108
Figura 56 – Processo de transcrição das CQs para SPARQL. . . . .	109
Figura 57 – Execução das consultas SPARQL na JENA . . . . .	110
Figura 58 – Processo de criação das nanopublicações . . . . .	112
Figura 59 – Arquitetura da nanopublicação adotada neste trabalho . . . . .	113
Figura 60 – Hierarquia dos níveis da nanoSCNAS . . . . .	114
Figura 61 – exemplo de cabeçalho de uma nanoSCNAS . . . . .	114
Figura 62 – Exemplo de código de geração dos axiomas das informações . . . . .	117
Figura 63 – Relacionando os Autores e DOI as proveniências de suas respectivas nanos . . . . .	117
Figura 64 – NanoSCNAS hospedado no banco SPARQL Jena Fuseki . . . . .	118
Figura 65 – Interface da aplicação semântica criada . . . . .	120
Figura 66 – Exemplo de consulta através da aplicação . . . . .	121
Figura 67 – Exemplo de consulta através da aplicação . . . . .	122
Figura 68 – Exemplo de consulta com uso de filtros através da aplicação . . . . .	123
Figura 69 – Exemplo de tooltip na coluna de SCNAS . . . . .	123

## LISTA DE TABELAS

Tabela 1 – Número de publicações retornadas em cada repositório. . . . .	63
Tabela 2 – Critérios de inclusão e exclusão dos documentos coletados . . . . .	63
Tabela 3 – Os 5 atores com maior Centralidade de Grau . . . . .	66
Tabela 4 – Tabela de Categorias adotadas para os documentos restantes . . . . .	67
Tabela 5 – Tabela de palavras-chaves . . . . .	68
Tabela 6 – Tabela de palavras-chave referentes as variedades de contexto . . . . .	69
Tabela 7 – Tabela com os trabalhos escolhidos para responderem as QPs . . . . .	70
Tabela 8 – Excerto da Primeira Lista de Questões de Competência . . . . .	79
Tabela 9 – Tabela de aplicação dos Critérios de Reúso . . . . .	94
Tabela 10 – Heurísticas usadas . . . . .	97
Tabela 11 – Tabela com os resultados obtidos na avaliação da aplicação . . . . .	124



## LISTA DE SIGLAS

**nanoSCNAS** *Nanopublication Scientific Collaboration Network Analysis*

**OWL** *Ontology Web Language*

**RDF** Resource Description Framework

**SCNAS Ontology** *Scientific Collaboration Network Analysis Ontology*

**SPARQL** SPARQL Protocol and RDF Query Language



## SUMÁRIO

1	INTRODUÇÃO . . . . .	25
1.1	Motivação . . . . .	25
1.2	Objetivo . . . . .	28
1.3	Objetivos Específicos . . . . .	28
1.4	Contribuições . . . . .	28
1.5	Organização do Trabalho . . . . .	29
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	31
2.1	Tecnologias Semânticas . . . . .	31
2.1.1	Web Semântica . . . . .	31
2.1.2	RDF . . . . .	33
2.1.3	Ontologias . . . . .	36
2.1.3.1	Metodologias de Criação . . . . .	37
2.1.4	OWL . . . . .	45
2.1.5	Noções Básicas da OWL . . . . .	46
2.1.5.1	Cabeçalho (Header) . . . . .	46
2.1.5.2	Indivíduos e Classes . . . . .	46
2.1.5.3	Camadas e Propriedades . . . . .	48
2.1.6	SPARQL . . . . .	49
2.1.6.1	Termologia . . . . .	50
2.1.6.2	Consultas SPARQL . . . . .	50
2.1.6.3	Exemplos . . . . .	51
2.2	Nanopublicações . . . . .	52
2.2.1	Modelo . . . . .	53
2.3	Análise de Redes Sociais . . . . .	56
2.3.1	Análise de Redes Sociais e a Colaboração Científica . . . . .	58
3	TRABALHOS RELACIONADOS . . . . .	61
3.1	Mapeamento Sistemático . . . . .	61
3.1.1	Definição das questões de pesquisa . . . . .	61
3.1.2	Definição das bases de dados e da <i>string</i> de busca . . . . .	62
3.1.3	Critérios de inclusão e exclusão . . . . .	63
3.1.4	Análise dos documentos . . . . .	63
4	METODOLOGIA APLICADA . . . . .	73
4.1	Coleta de Artigos . . . . .	74
4.2	Criação da Ontologia SCNAS <i>Ontology</i> . . . . .	74
4.2.1	Especificação dos requisitos da ontologia . . . . .	75
4.2.1.1	Propósito, escopo e formalidade . . . . .	77

4.2.1.2	Identificação dos Usuários . . . . .	77
4.2.1.3	Cenários de Uso . . . . .	77
4.2.1.4	Identificar os Requisitos . . . . .	78
4.2.1.5	Agrupamento dos requisitos . . . . .	79
4.2.1.6	Validação do conjunto de requisitos . . . . .	80
4.2.1.7	Priorizar os requisitos . . . . .	80
4.2.1.8	Extrair a terminologia e suas frequências . . . . .	81
4.2.2	Organização do desenvolvimento da Ontologia . . . . .	83
4.2.2.1	Selecionar o modelo de ciclo de vida da ontologia . . . . .	84
4.2.2.2	Selecionar os cenários da ontologia . . . . .	84
4.2.2.3	Atualizar o Plano Inicial . . . . .	86
4.2.2.4	Estabelecer restrições e associar recursos . . . . .	86
4.2.3	Modelo Cascata - Fase Reuso . . . . .	87
4.2.3.1	Reúso de Recursos ontológicos inteiros d domínio . . . . .	89
4.2.3.2	Reúso de Ontologias Gerais ou Comuns . . . . .	95
4.2.4	Fase de Projeto . . . . .	99
4.2.4.1	Conceituação . . . . .	99
4.2.5	Formalização . . . . .	106
4.2.6	Fase de Integração . . . . .	106
4.2.7	Fase de Implementação . . . . .	107
4.2.8	Avaliação da ontologia . . . . .	108
4.2.8.1	Identificar o tipo de Questão . . . . .	109
4.2.8.2	Determinar as respostas esperadas . . . . .	109
4.2.8.3	Extrair entidades das CQs e das respostas . . . . .	109
4.2.9	Identificar os tipos das entidades da ontologia . . . . .	110
4.2.9.1	Criar consultas SPARQL . . . . .	110
4.2.9.2	Executar as consultas SPARQL . . . . .	110
4.2.9.3	Avaliar resultados das consultas . . . . .	111
4.3	Construção das Nanopublicações . . . . .	112
4.3.1	Arquitetura da Nanopublicação . . . . .	112
4.3.2	Escolha das ontologias . . . . .	114
4.3.3	Extração das informações . . . . .	116
4.4	Armazenamento em uma Triple Store . . . . .	118
4.4.1	Aplicação Semântica . . . . .	119
4.5	Avaliação da Aplicação . . . . .	123
4.5.1	Resultados da Avaliação . . . . .	124
4.5.2	Lições Aprendidas . . . . .	125
5	CONSIDERAÇÕES FINAIS . . . . .	127



REFERÊNCIAS . . . . .	129
Índice . . . . .	135



## 1 INTRODUÇÃO

A utilização de meios eletrônicos na comunicação entre pesquisadores e acadêmicos tornou-se comum desde o surgimento da internet, pois sua estrutura proporciona benefícios para a troca de informações e comunicação, favorecendo o compartilhamento e a interatividade entre a comunidade científica (OLIVEIRA, 2008). A comunicação da produção científica é efetivada através de canais formais e/ou informais, que tomaram maior dimensão com a publicação em meio digital, quebrando as barreiras geográficas, de tempo e de acesso, atribuindo-lhe um carácter mais público (SCHENKEL, 2008). Nesse contexto, o compartilhamento da informação, entendido como conhecimento codificado, pode ser viabilizado de várias formas, através da criação de repositórios, do registro das melhores práticas, de dispositivos de memória organizacional para gerenciamento de seus ativos intelectuais, entre outros (SCHENKEL, 2008).

De acordo com Marcial e Hemminger (2010), os repositórios de dados científicos têm um papel fundamental na ciência. Neles são implementadas práticas sistemáticas de administração de dados, promovendo, assim, coleta de dados científicos adequados, curadoria, preservação, disponibilidade a longo prazo, disseminação e acesso. Esses repositórios são amplamente difundidos nas comunidades que produzem uma enorme quantidade de dados. Além disso, os repositórios de dados científicos são frequentemente propostos como instrumentos para dar suporte à publicação de dados, pois fornecem facilidades para todos os diferentes atores envolvidos nesse processo (ASSANTE et al., 2016).

### 1.1 Motivação

Com o passar do tempo, uma grande quantidade de publicações foram incluídas nos repositórios de pesquisas científicas, dificultando a recuperação e extração dos dados, assim como o reuso destes dados que podem ser relevantes para a resolução de outros problemas de pesquisa.

Embora os resultados narrativos possam ser lidos e compreendidos pelos humanos, eles não são tão facilmente digeridos pelos computadores (GOLDEN; SHAW, 2015). O problema é que as publicações eletrônicas contidas nos repositórios de publicações científicas não estão disponíveis de forma estruturada, visto que são apenas versões em PDF de artigos de pesquisas. Ou seja, são documentos com conceitos ambíguos e que não fazem uso de quaisquer técnicas de marcação e destaque de seus conteúdos, tornando essas informações ilegíveis para computadores e dificultando a pesquisa, preservação, agregação de valor e o reuso desses artigos, evitando a extração máxima do seu verdadeiro potencial (MONS; VELTEROP, 2009).

Visto os avanços que a computação tem oferecido para a comunidade científica, cabe aos pesquisadores publicarem os resultados obtidos através de pesquisas e estudos de tal maneira que possam ser facilmente analisados, coletados e comparados por programas de computadores. Pois somente assim será possível desfrutar do verdadeiro potencial da

computação (GOLDEN; SHAW, 2015).

Sabendo que os computadores trabalham de maneira mais eficiente com dados estruturados, Mons e Velterop (2009) concluíram que o uso de artigos completos talvez não seja a melhor forma de compartilhar e permitir o acesso ao conhecimento, já que o mais importante dessas publicações são os resultados obtidos durante um estudo, ou seja, pequenas unidades de afirmações. Porém, são adequados, e talvez até mesmo essenciais, para o registro detalhado dos resultados e do processo de pesquisa realizado para chegar nos mesmos (MONS; VELTEROP, 2009).

Pensando nos domínios de conhecimento cuja existência de um alto rendimento de resultados são comuns, observou-se a necessidade de criar uma estrutura complementar para os artigos de pesquisas, levando ao desenvolvimento das nanopublicações (LIPANI et al., 2014b).

Nanopublicação é uma proposta de Mons e Velterop (2009), cujo objetivo é publicar resultados de pesquisas em formato de dados estruturados, mantendo informações sobre a proveniência dos mesmos. Portanto, diferentemente das publicações tradicionais, onde seus resultados não são estruturados e poucas vezes mantém informações sobre a real origem dos dados utilizados na pesquisa, a nanopublicação através do uso de tecnologias semânticas, tais como RDF, OWL e SPARQL, etc. tem como objetivo estruturar afirmações obtidas em estudos, tais como seus resultados, mantendo seu contexto e os benefícios da abordagem tradicional.

Em outras palavras, uma nanopublicação pode ser considerada como a menor unidade de informação publicada em um documento formal: trata-se de uma extração textual estruturada atribuída a “fala do autor” (MOURA, 2009). Através da estruturação dos dados, obtém-se um melhoramento na curadoria de informações durante a realização de buscas e o reúso dos resultados científicos, de forma que a atribuição, qualidade e procedência sejam relevantes. Em contraste aos artigos tradicionais, ela se torna muito mais adequada diante à apresentação das relações entre dados de pesquisa e troca eficiente de conhecimento (GROTH; GIBSON; VELTEROP, 2010a).

Kuhn et al. (2013) exemplifica o uso de nanopublicações através de um cenário hipotético, onde um pesquisador A se questiona sobre uma determinada hipótese, porém não sabe se outros pesquisadores já pensaram sobre a mesma. Assim teria que acabar perdendo horas, ou talvez dias, fazendo buscas em repositórios para encontrar uma resposta. Em outro país, um pesquisador B pensou na mesma hipótese e conseguiu comprová-la. Porém, pergunta-se se a sua solução é algo novo, já que uma rápida pesquisa na web não trouxe respostas significativas. Com o uso das nanopublicações, os pesquisadores poderiam acessar um portal de nanopublicações, buscariam pela determinada hipótese e o sistema recuperaria todas as nanopublicações relacionadas, em particular aquelas com as mesmas sentenças, incluindo aquelas que usam diferentes expressões para expressar o mesmo significado. Logo após, o sistema compilaria e apresentaria as informações rele-

vantes: a quantidade de pesquisas existentes; se existem pesquisadores com as mesmas dúvidas, se existem questões resolvidas ou controvérsias; apoio ou se existe oposição dos pesquisadores; e referências aos artigos mais relevantes. Esse processo salvaria dias de pesquisa para pesquisador A e proporcionaria uma resposta rápida para o pesquisador B.

Desta forma, para Groth, Gibson e Velterop (2010b), enquanto artigos narrativos ainda teriam seu lugar no cenário acadêmico, pequenos fragmentos de dados formais na forma de nanopublicações deveriam assumir uma posição central na comunicação acadêmica, visto os benefícios que a abordagem traz para a interpretação computacional e humana.

Porém, apesar dos benefícios entregues pela abordagem, a nanopublicação ainda é pouco utilizada fora do seu domínio de criação, a Ciências da Vida. Paula e Moura (2015) afirmam que as nanopublicações podem ser consideradas promissoras, desde que seus aspectos teóricos e metodológicos sejam amplamente estudados, discutidos e disseminados.

A análise de redes sociais (ARS) é uma abordagem fundamentada através de modelos matemáticos e teoria dos grafos, que tem como objetivo realizar um levantamento de propriedades e conteúdos provenientes de dados de Redes Sociais (WASSERMAN; FAUST, 1999). De acordo com Newman (2010), uma rede social pode ser definido, em sua forma mais simples, como um conjunto de nós interligados através de arestas. Estes nós representam atores ou pontos, que correspondem a entidades, pessoas, empresas ou organizações, que podem ser analisados como unidades individuais ou sociais coletivas (WASSERMAN; FAUST, 1999).

O método de análise de redes sociais também é utilizado para estudar as relações de colaboração científica entre pesquisadores. De acordo com Sonnenwald (2007), a colaboração científica é definida pela interação de um conjunto de cientistas em um determinado contexto social, que proporciona o compartilhamento de atividades científicas para que se possa atingir objetivos comuns. O trabalho realizado a partir da colaboração científica proporciona economia de tempo, de recursos financeiros e materiais (MEADOWS; LEMOS, 1999). De acordo com Barabási et al. (2002), redes de co-autoria refletem as relações entre os cientistas de um domínio ou região. Com isso, todos os anos são realizados estudos de análise de redes de colaboração científica com o propósito de analisar a contribuição de produções científicas de determinados pesquisadores para um domínio. A literatura apresenta algumas métricas utilizadas nesses estudos para descrever a coesão das redes e o papel desempenhado pelos atores, como a densidade e as medidas de centralidades.

Este trabalho propõe o uso da abordagem de nanopublicação aplicado ao domínio de Estudos Análise de Redes de Colaboração Científica (EARCC). A escolha foi influenciada pela estrutura deste domínio, que é mais bem definida em comparação a outros domínios, visto que a maioria dos artigos apresentam seus resultados de acordo com as métricas de Análise de Redes Sociais. Como exemplo, a grande maioria dos artigos apresentam a

densidade da rede, assim como os valores de centralidade de grau, etc. Verificou-se que as métricas utilizadas para o estudo das redes trazem consigo resultados que podem se mostrar interessantes para outros pesquisadores, que conseqüentemente poderiam utilizar desses dados em suas pesquisas.

Visto que todos os anos são publicados artigos com novos resultados obtidos através de análises redes de colaboração científicas, a nanopublicação se mostrou uma abordagem que permite a disseminação destes resultados de forma segura, confiável e ágil.

## 1.2 Objetivo

Este trabalho possui como objetivo criar um método e um conjunto de artefatos capazes de criar e recuperar nanopublicações.

## 1.3 Objetivos Específicos

Para conseguir alcançar o objetivo deste trabalho, será necessário:

- Criar uma ontologia que represente os conceitos de Estudos de Análise de Redes de Colaboração Científica;
- Definir uma arquitetura suficiente para a criação de nanopublicações do domínio de Análise de Redes de Colaboração Científica;
- Criar uma aplicação semântica capaz de recuperar as nanopublicações criadas.

## 1.4 Contribuições

Este trabalho tem como principais contribuições científicas:

- Uma ontologia capaz de representar conceitos de EARCC;
- Uma ontologia para nanopublicações do domínio de Análise de Redes de Colaboração Científica;
- Uma aplicação capaz de recuperar nanopublicações de Análise de Redes de Colaboração Científica;
- Um processo detalhado de criação de uma ontologia através da metodologia NeOn.

A ontologia criada para representar conceitos de Estudos de Análise de Redes de Colaboração Científica, denominada Scientific Collaboration Network Analysis (SCNAS) Ontology, pode ser parcialmente utilizada em outros sub-domínios de Análise de Redes. A ontologia é capaz de apoiar atividades de recuperação, extração e representação de

conceitos do domínio de EARCC, podendo ser utilizada e reutilizada pela comunidade de pesquisadores do domínio.

A ontologia criada para as nanopublicações, denominada nanoSCNAS, tem como principal propósito permitir a identificação de asserções contidas em estudos do domínio, assim como a identificação da proveniência das asserções, da nanopublicação e dos dados utilizados nos estudos.

Neste trabalho também foi implementada uma aplicação semântica capaz de recuperar nanopublicações do domínio de Análise de Redes de Colaboração Científica. Esta aplicação teve como objetivo mostrar a viabilidade do uso das nanopublicações, no qual permitiu a recuperação mais precisa das informações buscadas, sem perder a confiabilidade das mesmas.

Este trabalho descreveu o processo detalhado de criação de uma ontologia através da metodologia NeOn, podendo servir como apoio para pesquisadores que desejam criar ontologias através desta metodologia.

Por fim, este trabalho descreve um método e um conjunto de artefatos capazes de criar e recuperar nanopublicações de um determinado domínio. No entanto, o método e os artefatos podem ser utilizados como apoio para qualquer pesquisadores que tenha como interesse criar e recuperar nanopublicações, independentemente de seu domínio.

## 1.5 Organização do Trabalho

Este trabalho está organizado da seguinte maneira:

- Capítulo 2 - é apresentada uma fundamentação teórica sobre a Web Semântica e suas tecnologias, nanopublicações e seus conceitos.
- Capítulo 3 - é apresentado os trabalhos já desenvolvidos usando a abordagem de nanopublicação.
- Capítulo 4 - é apresentado a metodologia utilizada para atingir os objetivos deste trabalho, juntamente com seus resultados .
- Capítulo 5 - Considerações finais.





## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados conceitos de tecnologias semânticas, nanopublicação e análise de redes sociais.

### 2.1 Tecnologias Semânticas

#### 2.1.1 Web Semântica

A web semântica é uma abordagem criada para acrescentar informações semânticas na web que, através de agentes inteligentes, é possível realizar buscas e tarefas de forma muito mais eficiente (BERNERS-LEE; HENDLER; LASSILA, 2001).

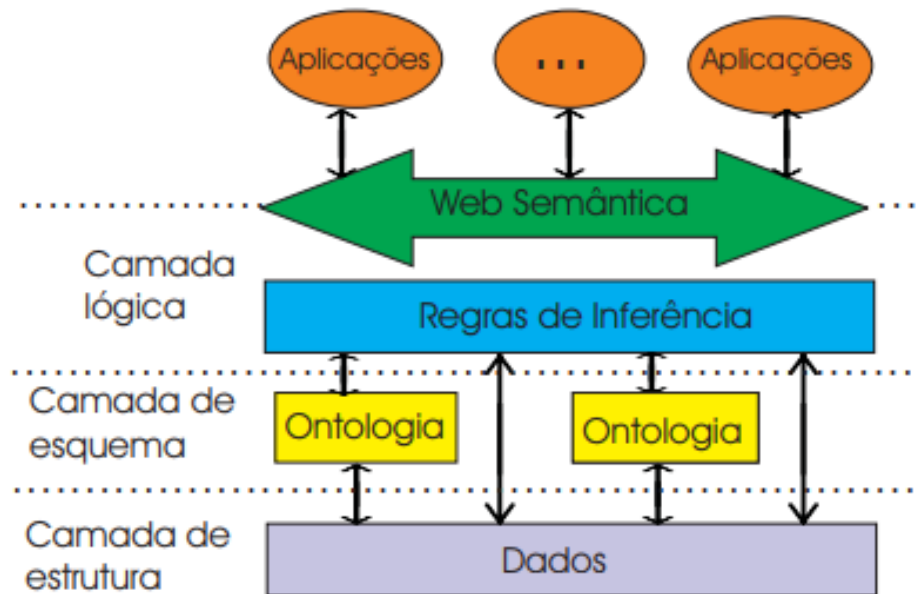
Em Berners-Lee, Hendler e Lassila (2001) é relatado que grande parte do conteúdo disponível na web são destinados apenas a leitura humana, criando uma barreira na comunicação e manipulação entre esses materiais e as máquinas. Ou seja, existem dificuldades na interpretação de conteúdos semânticos através de computadores, (BERNERS-LEE; HENDLER; LASSILA, 2001).

Com o uso da WS e suas tecnologias, é possível resolver o problema de falta de estruturação semântica dos dados contidos nos repositórios da web, tornando esses dados legíveis por máquinas. Frequentemente estas informações são disponibilizadas sem nenhum critério de organização ou padronização, dificultando o processo de busca e recuperação das mesmas por meio de mecanismos automatizados (LIMA; CARVALHO, 2005a).

"A Web Semântica representa a evolução da web atual. Enquanto a web tradicional foi desenvolvida para ser entendida apenas pelos usuários, a Web Semântica está sendo projetada para ser compreendida pelas máquinas, na forma de agentes computacionais, que são capazes de operar eficientemente sobre as informações, podendo entender seus significados. Desta maneira, elas irão auxiliar os usuários em operações na web"(DZIEKANIAK; KIRINUS, 2004).

Em Lima e Carvalho (2005a) é apresentada a representação de uma das possíveis arquiteturas que pode ser adotada pela Web Semântica, a conferir na figura 2

Figura 1 – Uma possível arquitetura para a Web Semântica.



Fonte: Lima e Carvalho (2005a)

Segundo Hendler (2001), a Web Semântica pode ser considerada como a composição de um grande número de pequenos componentes ontológicos que apontam entre si. Dessa forma, companhias, universidades, agências governamentais e grupos de interesses específicos procurarão ter seus recursos web ligados a um conteúdo ontológico, já que ferramentas poderosas serão disponibilizadas para intercambiar e processar essas informações entre aplicações web.

"A semântica não está somente relacionada ao conteúdo de um recurso, mas também à forma de como este se relaciona com os demais recursos na web. Portanto, é essencial que os recursos disponibilizados sejam expressivos o bastante para que as máquinas ou agentes sejam capazes de processar e entender o real significado do dado, intermediando as necessidades de cada usuário e as fontes de informações disponíveis" (DZIEKANIAK; KIRINUS, 2004).

A tecnologia da Web Semântica oferece novas possibilidades para que se possa realizar pesquisas mais inteligentes ao invés de apenas realizar buscas através de palavras-chave.

"A Web Semântica não é uma Web separada, mas uma extensão da atual. Nela a informação é dada com um significado bem definido, permitindo melhor interação entre os computadores e as pessoas" (BERNERS-LEE; HENDLER; LASSILA, 2001).

Para que a existência da web semântica seja possível, é necessário fazer uso de abordagens como: ontologias, um conjunto de princípios para design, grupos de trabalho

e tecnologias para a implantação desses princípios e propostas. Alguns elementos são expressos em especificações formais que incluem o modelo RDF, formatos de intercâmbio de dados (como por exemplo: RDF/XML, N3, Turtle, N-Triples) e notações, tais como RDF Schema (RDFS), Web Ontology Language (OWL), para prover uma descrição formal dos recursos e dos conceitos, termos e seus relacionamentos, num determinado domínio do conhecimento. Para a recuperação desses elementos é possível fazer uso do SPARQL Protocol and RDF Query Language (SPARQL) (CATARINO; SOUZA, 2012). Cada uma dessas técnicas tem distintas funções e padrões, que juntas compõe a web semântica. Nas próximas seções é apresentado o conceito de cada uma delas.

### 2.1.2 RDF

O *Resource Description Framework* (RDF) é uma linguagem declarativa criada e recomendada pela W3C para representar informações na *Web* (CONSORTIUM et al., 2014). Serve como um auxílio ao desenvolvimento de metadados cujo objetivo é promover a interoperabilidade entre aplicações que compartilham informações que sejam entendidas por sistemas na Web (CONSORTIUM et al., 2014). Diante da existência de várias ferramentas para lidar com dados e para trabalhar com os relacionamentos entre os mesmos, o RDF é o padrão mais fácil, seguro e poderoso criado até o momento (CONSORTIUM et al., 2014). Ele não predefine qualquer semântica nem pressupõe um domínio específico de conhecimento. Trata-se assim de um mecanismo de descrição neutro, que serve para descrever recursos de qualquer área do conhecimento (LASSILA; SWICK, 1999). O RDF recebeu a influência de várias fontes distintas, sendo que as principais vieram das comunidades de padronização da web (HTML, XML e SGML), da Biblioteconomia (metadados de catalogação), da representação do conhecimento (ontologias), da programação orientada a objetos, da linguagem de modelagem, entre outras (DZIEKANIAK; KIRINUS, 2004).

O RDF pode utilizar-se da linguagem de marcação extensível (XML), uma vez que faz parte dos objetivos do RDF realizar a possível especificação da semântica em bases XML, criando interoperabilidade entre os dados contidos nos repositórios (CONSORTIUM et al., 2014). É importante ressaltar que xml é apenas uma das sintaxes que podem ser usadas juntamente ao RDF (CONSORTIUM et al., 2014). Como exemplo de sintaxes que podem servir como complemento ao RDF.

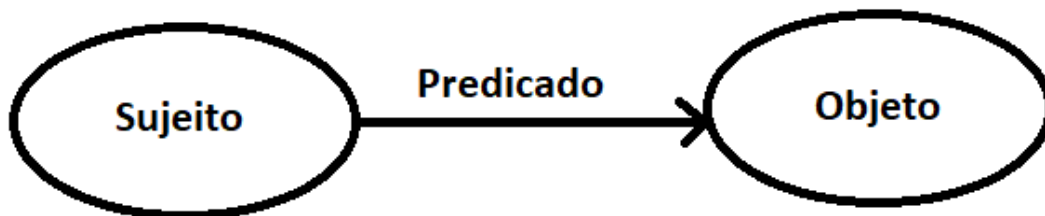
O recurso RDF tem a capacidade de definir um padrão de ontologias, que servem para descrever qualquer tipo de recurso para a Internet. Essa tecnologia estabelece um padrão de metadados que pode ser trabalhado juntamente com sintaxes distintas, como o OWL, Turtle e Xml. A complementação entre as duas tecnologias pode criar o RDF Schema, ou RDFS, que faz parte da especificação do padrão. O RDF tem como objetivo descrever os dados e os metadados por meio de um esquema de “triplas” (SOUZA; ALVARENGA, 2004).

Uma tripla é formada por um recurso, uma propriedade e um valor para a propriedade daquele recurso. Possuindo o seguinte formato: <sujeito, predicado, objeto>. Seu significado pode ser resumido como: “o recurso (sujeito) que possui a propriedade (predicado) com determinado valor (objeto)”. Um valor ou objeto pode ser tanto um outro recurso quanto um tipo primitivo definido por XML (REZENDE; BAX, 2002).

Triplas descritas em RDF podem ser representados como grafos diretos rotulados (KLEIN, 2001).

A figura 2 mostra um grafo RDF com dois nós (Sujeito e Objeto) que são ligados através de um relacionamento entre os nós (Predicado):

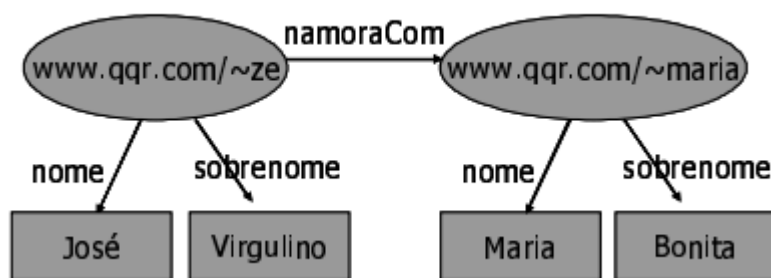
Figura 2 – Grafo RDF



Fonte: Adaptado de REZENDE e BAX (2002)

Para que se possa entender a aplicação do grafo RDF, segue um exemplo, concedido por Freitas (2003), de grafo ligando duas páginas da Web:

Figura 3 – Grafo representando uma tripla RDF



Fonte: Adaptado de Freitas (2003)

O grafo da figura 3 mostra que José Virgulino, dono da primeira página, namora com Maria bonita, proprietária da segunda página. O código do grafo da figura 3 é o equivalente ao seguinte:

Figura 4 – Código do grafo da Figura 3

```

<rdf:Description about=http://www.qqr.com/~ze>
  <nome>Jose</nome>
  <sobrenome>Virgulino</sobrenome>
  <namoraCom>
    <rdf:Description about=http://www.qqr.com/~maria>
      <nome>Maria</nome>
      <sobrenome>Bonita</sobrenome>
    </rdf:Description>
  </namoraCom>
</rdf:Description>

```

Fonte: Adaptado de Freitas (2003)

A vantagem de usar o RDF como linguagem de descrição de recursos sobre outras linguagens de marcação, está na liberdade de ignorar as imposições da estrutura do documento, o que se refere apenas a dados sobre o conteúdo. Para que houvesse um padrão a ser seguido no uso da abordagem, foram criados os esquemas RDF (RDF-Schemes ou RDFS), que fornecem tipos básicos para a criação de esquemas voltados à aplicações específicas. As definições usadas para a modelagem de novos esquemas RDF incluem classe, subclasse (herança), propriedade, sub-propriedade (para construir hierarquias de propriedades), instância e restrição (FREITAS, 2003).

Aplicando o uso desse esquema, obtem-se o seguinte código:

Figura 5 – Aplicação do esquema RDF no código do grafo da Figura 3

```

<rdf:Description ID="Pessoa">
  <rdf:type resource="http://www.w3c.org/TR/1999/PR-rdf-schema-19990303#Class">
  <rdfs:subClassOf rdf:resource="http://www.w3c.org/TR/1999/PR-rdf-schema-19990303#Resource">
</rdf:Description>

<rdf:Description ID="nome">
  <rdf:type resource="http://www.w3c.org/TR/1999/PR-rdf-schema-19990303#Property">
  <rdfs:domain rdf:resource="#Pessoa">
  <rdfs:range rdf:resource="http://www.w3c.org/TR/xmlschema-2#string">
</rdf:Description>

```

Fonte: Adaptado de Freitas (2003)

Como é possível verificar no código representado na figura 62, "Pessoa" é do tipo classe e herda características de um recurso. Já o atributo "nome" é colocado como uma propriedade da classe Pessoa, onde a propriedade é tratada como uma string. É possível visualizar que todas as definições são identificadas com uma URI própria. Como se vê, especificar código RDF pode ser um pouco complicado à primeira vista, mas, para auxiliar os usuários, existem ferramentas para edição e *parsing* sobre RDF (FREITAS, 2003).

O URI é uma forma mais geral de identificação. Qualquer pessoa e organização pode, independentemente, criá-los e usá-los para identificar objetos. Os URIs podem ser criados para referir-se a qualquer coisa que precise ser referenciada em uma declaração, não somente a recursos com endereços na Web.(LIMA; CARVALHO, 2005b)

"Por causa da sua generalidade, RDF usa URIs como base do seu mecanismo de identificação de sujeitos, predicados e objetos em declarações. Para ser mais preciso, RDF usa referências URIs. Uma referência URI é um URI juntamente com um identificador de fragmento opcional no final"(LIMA; CARVALHO, 2005b).

Resumidamente, por Lima e Carvalho (2005b), têm-se os seguintes itens como princípios, uso e aplicação do RDF:

- É proposto para situações onde as informações necessitam ser processadas por aplicações, em lugar de somente serem visualizadas por pessoas.
- Provê uma estrutura comum para expressar informações que podem ser trocadas entre diferentes aplicações sem perda de significado.
- Baseia-se no princípio de identificação de objetos usando identificadores Web, também chamados de URIs, e na descrição de recursos em termos de propriedades e valores de propriedade. Isto capacita o RDF a representar declarações simples sobre recursos como um grafo de nós e arcos, representando os recursos, suas propriedades e valores.

### 2.1.3 Ontologias

Ontologia é uma palavra de origem grega cujo significado pode ser entendido como “o estudo da existência”. Desde sua origem, a ontologia tem sido usado na filosofia para designar tanto disciplinas, quanto a sistemas de categorias, independente de linguagens, tornando-se apropriado para conceituação de teorias científicas (GUIZZARDI, 2005).

Com o passar dos anos diversas definições de ontologia foram criadas. Gruber (1995) apresenta uma definição bastante conhecida, onde: Uma ontologia é uma especificação explícita de uma conceitualização. Borst (1999) estende a definição dada por (GRUBER, 1995), apresentando a ontologia como uma especificação formal de uma conceituação compartilhada. Baseando-se nas definições anteriores, Studer, Benjamins e Fensel (1998) definem uma ontologia como: Uma especificação formal e explícita de uma conceituação compartilhada. Têm-se como formal o fato de uma ontologia ser legível por máquina. Enquanto ao explícita, refere-se aos tipos de conceitos utilizados e as restrições ao seu uso são explicitamente definidos. Conceituação se refere a um modelo abstrato que identifica os conceitos relevantes de algum acontecimento, conhecimento ou algo. Compartilhado porque a ontologia captura o conhecimento consensual aceito por um grupo.

Diferentemente de estudos científicos realizados em domínios específicos (como a física, química, biologia), que trabalham apenas com conceitos específicos de sua categoria, a ontologia permite trabalhar com  $n$  áreas de conhecimento, seus termos e relações (GUIZZARDI, 2005). Com o uso de ontologias, por exemplo, é possível definir conjuntos comuns entre termos usados para descrever e representar algum domínio, como: medicina, biblioteca, matemática (LIMA; CARVALHO, 2005a).

De forma um pouco mais detalhada, uma ontologia é usada para descrever conceitos de um domínio através de classes, propriedades, características e atributos que, junto a instâncias e indivíduos, constituem uma base de conhecimento (NOY; MCGUINNESS et al., 2001).

"As classes são o foco da maioria das ontologias. São elas que descrevem os conceitos de um determinado domínio. Por exemplo, suponhamos a existência de uma classe "vinhos", essa classe representa todos os vinhos. Vinhos específicos são exemplos dessa classe. O copo de vinho Bordeaux na sua frente, enquanto você lê este documento, é uma instância da classe de vinhos de Bordeaux. Uma classe pode ter subclasses que representam conceitos mais específicos do que a superclasse. Por exemplo, podemos dividir a classe de todos os vinhos em vermelho, branco, e vinhos rosé. Alternativamente, podemos dividir a classe vinhos em espumante e não-espumante"(NOY; MCGUINNESS et al., 2001).

Quando aplicada na computação, o uso de ontologias pode trazer benefícios como a melhora da exatidão de buscas realizadas na web, ou seja, ferramentas que usam ontologias podem ser capazes de recuperar somente páginas que sejam relevantes para o usuário. (LIMA; CARVALHO, 2005a)

A web semântica tem como característica o uso de uma grande quantidade de ontologias, que são criadas de acordo com as necessidades de uma pesquisa, ou de grupos com interesses em comuns (REZENDE; BAX, 2002). A partir da forte influência da ontologia na web semântica, foi necessário criação de tecnologias que permitissem a manipulação, especificação e representação das ontologias. Algumas dessas tecnologias são apresentadas nas próximas seções.

### 2.1.3.1 Metodologias de Criação

Metodologias têm sido desenvolvidas no intuito de sistematizar a construção e a manipulação de ontologias (López, 1999). Apesar de haver uma numerosa literatura metodológica relacionada a ontologias, não basta simplesmente eleger uma qualquer para a criação de uma nova ontologia (SUAREZ, 2010). Visto que cada metodologia segue um processo de criação diferente, e para cenários distintos, a metodologia a ser eleita deve ser adequada a ontologia a ser construída e, por esse motivo, Suarez (2010) sugere alguns requisitos para que a metodologia garanta bons resultados. São eles:

- Generalidade - Uma metodologia deve ser geral o suficiente e não deve ser conduzida para resolver casos.

- Completude - Uma metodologia deve considerar todos os casos apresentados e propor soluções para todos eles.
- Eficácia - Uma metodologia deve resolver adequadamente os casos propostos que possuem uma solução, com independência da pessoa que aplica a metodologia. Portanto, deve ser mais prescritivo do que descritivo.
- Eficiência - Uma metodologia deve ser eficiente, isto é, deve ser capaz de atingir seu objetivo ou meta. Isso significa que a metodologia deve permitir a construção de ontologias.
- Consistência - Uma metodologia deve produzir o mesmo conjunto de produtos para o mesmo problema, independentemente de quem aplica a metodologia. O conteúdo de tal conjunto de produtos será diferente dependendo do problema tratado.
- Finito - O número de elementos que compõem uma metodologia e o número de atividades devem ser finitos, ou seja, devem consumir um período razoável de tempo.
- Discernimento - Uma metodologia deve ser composta por um pequeno conjunto de componentes estruturais, funcionais e representacionais.
- Ambiente - Metodologias podem ser classificadas em científicas e tecnológicas. Nas metodologias científicas, as ideias são validadas, enquanto que, nas tecnológicas, os artefatos são construídos e avaliados. Uma metodologia tecnológica deve considerar o ciclo de vida do produto que está guiando seu desenvolvimento.
- Transparência - Uma metodologia deve ser transparente diante o decorrer do desenvolvimento, o que envolve os processos ativos ou atividades que estão sendo realizadas, quem está realizando, etc.
- Questionamentos importantes - Em cada atividade incluída na metodologia, deve-se sempre responder as seguintes perguntas: “o quê”, “quem”, “por que”, “quando”, “onde” e “como”.

Este trabalho não tem como objetivo realizar qualquer análise extensa entre os métodos ontológicos existentes, porém é importante conhecer minimamente algumas metodologias para que este conhecimento possa ser utilizado em decisões futuras. Por isso, os próximos tópicos são dedicados a apresentar, de forma sucinta, estes métodos.

### **Metodologia NeOn**

Inspirado na engenharia de software, a metodologia NeOn tem como objetivo sistematizar a criação de uma ontologia através do reúso de modelos ontológicos, ou não ontológicos, já existentes (SUAREZ, 2010). Para isso, abstrai um conjunto de 9 cenários de criação de uma nova ontologia, que define o fluxo de atividades a serem realizadas durante a fase de desenvolvimento. Esses cenários podem ser combinados para que se adaptem

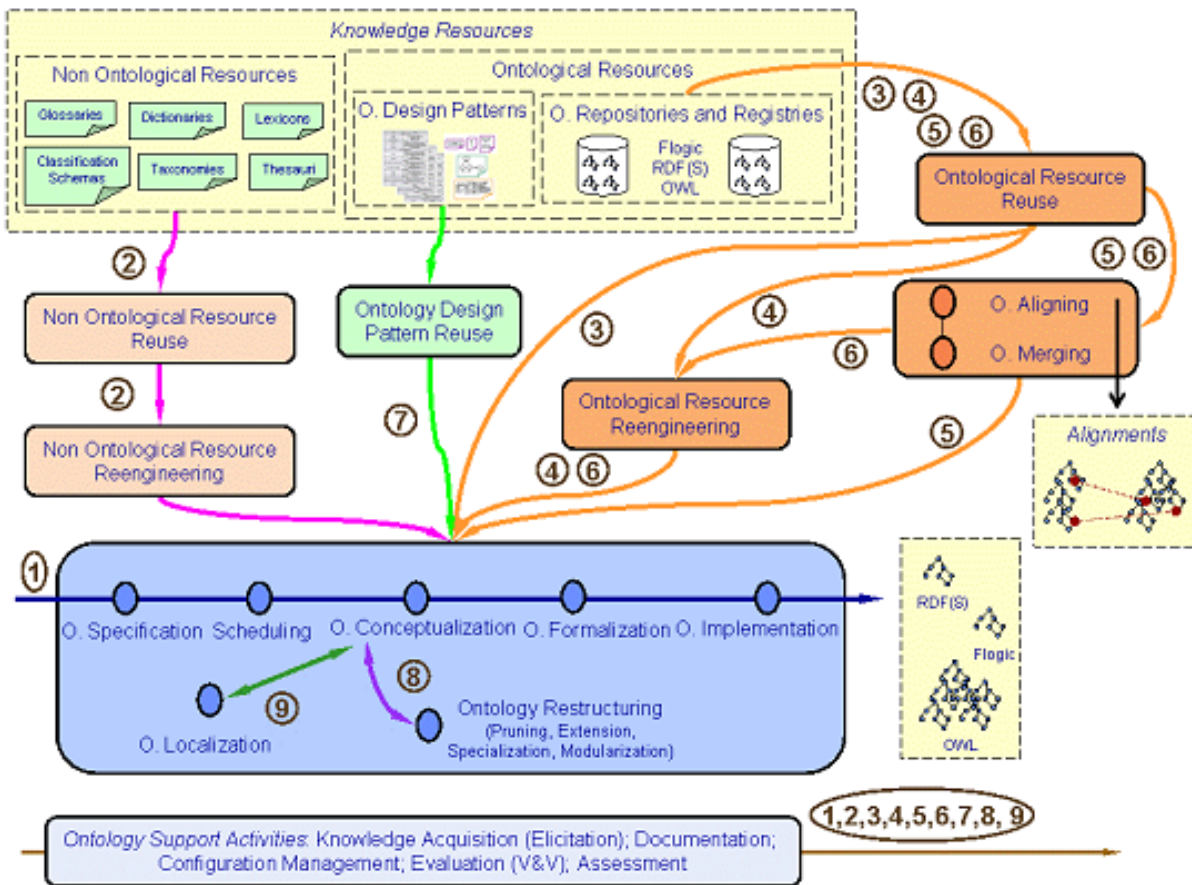


as necessidades do desenvolvimento da ontologia proposta. A seguir, é realizada uma descrição dos cenários identificados pela metodologia, seguido da figura 6 que explicita o fluxo de atividades de acordo com o cenário:

- Cenário 1 - Da especificação à implementação: A ontologia é desenvolvida desde o início, ou seja, a partir do zero. Sem reutilizar recursos de conhecimento disponíveis.
- Cenário 2 - Reutilizar e re-engenhar recursos não ontológicos: Os desenvolvedores da ontologia devem realizar o processo de reutilização de recursos não ontológicos para decidir, de acordo com os requisitos, quais NORs (recursos não ontológicos) podem ser reutilizadas para construir a ontologia.
- Cenário 3 - Reutilizando recursos ontológico: Os desenvolvedores da ontologia reutilizam recursos ontológicos inteiros, módulos ou declarações de ontologias.
- Cenário 4 - Reutilizar e re-engenhar recursos ontológicos: Desenvolvedores fazem uso de ontologias, reutilizando e reprojutando as mesmas.
- Cenário 5 - Reutilizando e fundindo recursos ontológicos: Este cenário se aplica somente em casos onde existem vários recursos ontológicos de um mesmo domínio. Os recursos selecionados para reutilização, permitindo que os desenvolvedores possam criar um novo recurso ontológico a partir de dois ou mais recursos ontológicos selecionados.
- Cenário 6 - Reutilizando, fundindo e reengenhando recursos ontológicos: Neste cenário os desenvolvedores reutilizam, fundem e reestruturam recursos ontológicos para a construção de uma nova ontologia. Este cenário é semelhante ao cenário 5, porém os desenvolvedores não fazem uso de somente um conjunto de recursos fundidos, mas fazem a reengenharia dele.
- Cenário 7 - Reutilizando Padrões de projeto de Ontologias: Os desenvolvedores reutilizam padrões de projetos de ontologias para criar uma nova ontologia.
- Cenário 8 - Reestruturação de recursos ontológicos: Os desenvolvedores reestruturam recursos ontológicos (através de módulos, reestruturação ou estendimento) para serem integrados na ontologia que está sendo construída.
- Cenário 9 - Localizando recursos ontológicos: Desenvolvedores adaptam uma ontologia a outras linguagens, obtendo uma ontologia multilíngue.

Como pode ser visto na figura 6, todos os cenários estão combinados com o fluxo de atividades do cenário 1, isso porque o mesmo contém 5 fases compostas de atividades básicas para a criação de uma ontologia. Sendo dividido em: especificação, planejamento, conceitualização, formalização e implementação.

Figura 6 – Cenários e fluxos de atividades



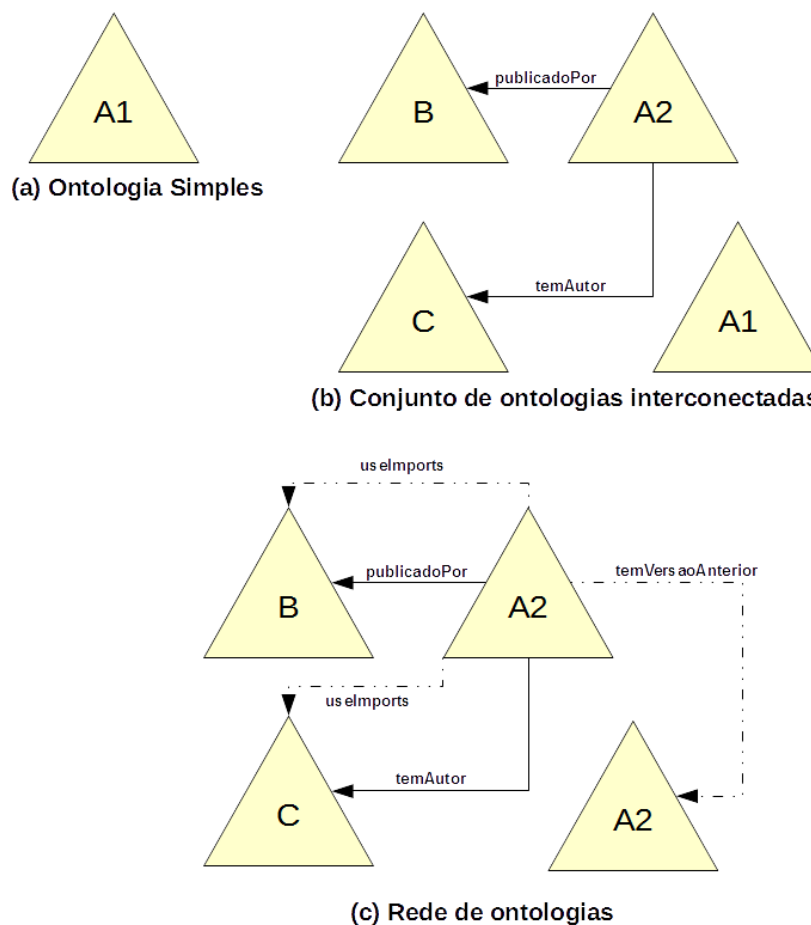
Fonte: Suarez (2010)

Suarez (2010) também realiza a identificação de três distintas situações de construção de uma ontologia, sendo elas:

- Construção de uma ontologia simples - Quando não há relacionamentos, seja dependente ou independente do domínio, com outras ontologias.
- construção de conjuntos de ontologias simples e interconectadas - Quando existe um conjunto de ontologias no qual possuem algum tipo de relacionamento dependente de domínio entre as ontologias (BORTOLATO; PAULO, 2014).
- Rede de ontologias ou ontologia em rede - Trata-se de uma coleção de ontologias interconectadas entre si por meio de uma variedade de relacionamentos, como: importação, alinhamento, modularização, versionamento e dependência (BORTOLATO; PAULO, 2014).

A figura 7 representa exemplos das três situações citadas anteriormente.

Figura 7 – Representação dos gráficos das três situações



Fonte: Suarez (2010)

Distinguir a situação de construção da ontologia é um passo importante para que se possa, durante a fase de planejamento do projeto da ontologia, identificar as atividades que serão realizadas.

A seguir será detalhado as fases que compõe o cenário 1.

### Fase 1 - Especificação

Algumas condições são necessárias para que se cumpra a especificação dos requisitos de ontologia, de forma que a atividade contenha detalhes sobre as principais atividades, entradas, saídas e atores. Apesar de não haver definições mais detalhadas sobre a especificação dos requisitos Suarez (2010) cita Gruninger e Fox (1995) quando se refere aos benefícios das questões de competência abordadas em sua metodologia, uma técnica bastante precisa, que pode ser aplicada por desenvolvedores de software e profissionais de ontologia, permitindo identificar os principais termos a serem incluídos. Suarez (2010) deixa claro que atividade de especificação de requisitos resulta em um documento de especificação que serve como um registro de informações, o qual deve conter minimamente:

- Propósito - Os objetivos da ontologia.

- Escopo - A cobertura da ontologia, juntamente com a sua granularidade.
- Nível de formalidade/Linguagem de implementação - O grau de formalidade da ontologia e a linguagem de implementação da mesma.
- Usuários esperados - A identificação dos usuários finais, a queles que irão fazer uso da ontologia.
- Usos esperados - Os principais cenários no qual a ontologia poderá ser aplicada.
- Questões de Competência (QCs) - QCs que devem ser identificadas, agrupadas, validadas e priorizadas, bem como suas respectivas respostas. Posteriormente estas QCs devem passar por um processo de extração de termos que irão compor um pré-glossário.
- Pré-Glossário de termos - Deve se manter registrado uma lista inicial de termos significativos do domínio, juntamente com suas respectivas frequências.

### **Fase 2 - Organização do desenvolvimento da Ontologia**

Esta fase é responsável por identificar os processos e atividades que serão executadas durante o desenvolver da ontologia, assim como o tempo e os recursos necessário para a conclusão da mesma. Portanto esta é uma tarefa primordial a ser executada, já que ela irá estabelecer, através de suas atividades, o tempo de vida útil da ontologia. Para que se possa constituir uma organização de atividades eficiente para o desenvolvimento da ontologia, a metodologia NeOn estabelece quatro questões importantes a serem respondidas:

- Qual modelo de ciclo de vida é o mais apropriado para a ontologia a ser desenvolvida?
- Quais os processos e atividades específicas que devem ser executadas durante o desenvolver da ontologia?
- Qual a ordem e quais as dependências existentes entre os processos e atividades a serem executados?
- Quantos recursos (humanos e tempo) são necessários para desenvolver a ontologia?

Para responder estas questões Suarez (2010) estabelece 4 atividades, são elas:

- Selecionar o modelo de ciclo de vida da ontologia
- Selecionar os cenário da ontologia
- Atualizar o plano inicial da ontologia
- Estabelecer restrições e associar recursos

A atividade de **Selecionar o modelo de ciclo de vida da ontologia** tem como objetivo definir o modelo de ciclo de vida que a ontologia proposta deve seguir. Suarez (2010) propõe dois modelos: o modelo cascata onde um estágio deve ser concluído para que o próximo tenha início, visto que os requisitos são completamente conhecidos já início do desenvolvimento da ontologia. E o modelo iterativo, baseado em um conjunto de iterações, onde cada uma é baseada em uma das cinco versões do modelo cascata propostos. O modelo Cascata inclui um conjunto de atividades de suporte que devem ser executadas em todas as fases de desenvolvimento da ontologia. Nessas atividades estão incluídas a aquisição de conhecimento necessário do domínio em que a ontologia está sendo desenvolvida, a avaliação da ontologia a partir do conhecimento adquirido do domínio e a avaliação do usuário e das necessidades das diferentes saídas de cada fase do modelo cascata. A metodologia NeOn define cinco versões distintas do modelo cascata. Estas versões foram criadas de forma incremental, sendo que cada uma contém um segmento de fases específicas para a sua versão e são adequadas a cenários diferentes, são elas:

- Modelo de 4 fases - Representa as atividades de desenvolvimento de uma ontologia, composta pela fase de iniciação, fase de projeto, fase de implementação e fase de manutenção.
- Modelo de 5 fases - Estende o modelo anterior com a reutilização de recursos ontológicos como eles são.
- Modelo de 5 fases + Merge - Segue o modelo de 5 fases, porém é adicionado uma fase de merge para que seja obtido um novo recurso ontológico através de dois ou mais recursos ontológicos selecionados na fase de reutilização.
- Modelo de 6 fases - Estende o modelo de 5 fases com a fase de reengenharia, permitindo a reengenharia de recursos do conhecimento (ontológicos e não ontológicos).
- Modelo de 6 fases + fusão - Ele estende o modelo anterior, adicionando a fase de mesclagem após a fase de reutilização.

A atividade de **Selecionar os cenários da ontologia**, tem como objetivo identificar em quais dos 9 cenários abordados pela metodologia a ontologia a ser criada se adéqua. A escolha do cenário irá identificar as atividades a serem realizadas para a concretização da ontologia.

Diante a atividade **Atualizar o Plano Inicial**, tem-se como objetivo adequar as atividades a serem realizadas, de acordo com a necessidade da ontologia, incluindo: a alteração da ordem e dependências entre os processos e atividades, inclusão ou exclusão de processos, atividades ou fases do modelo.

### **Fase 3 - Conceitualização**

Nessa fase é criado um modelo conceitual que apresenta o problema a ser resolvido,

assim como a solução para o mesmo, objetivando-se alcançar um melhor entendimento sobre o domínio a ser trabalhado. Para esta fase (SUAREZ, 2010) recomenda o uso da metodologia proposta por (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997), afim de se trabalhar com questões de competência (QC) para que se possa extrair os termos, instâncias e relações que deverão estar presentes na ontologia a ser criada. Sendo assim, são definidas as seguintes atividades a serem executadas:

- Construção de um glossário de termos
- Construção de taxonomias conceituais
- Construção de diagramas de relações binárias *ad hoc*
- Construção de um dicionário de conceitos
- Descrição de relações binárias *ad hoc*
- Descrição de atributos de instâncias
- Descrição de atributos de classe
- Descrição de constantes
- Descrição de axiomas formais
- Descrição de regras
- Descrição de instâncias

#### **Fase 4 - Formalização**

Esta fase é dedicada a transformação do modelo conceitual para o modelo formal, sendo assim, utiliza-se nesta etapa todas as saídas da fase de conceitualização. Na formalização, os conceitos são transformados em classes e logo após são organizados em hierarquia, de acordo com a taxonomia definida. As propriedades binárias e atributos de instâncias são transformados em propriedades de objetos e dados, respectivamente. Os atributos de classe são transformados em propriedades de anotação da classe que o mesmo corresponde após a transformação do conceito para classe. As constantes definidas são transformadas em propriedades de dados ou objeto, dependendo do caso.

#### **Fase 5 - Implementação**

Nesta fase ocorre a implementação da ontologia de acordo com as saídas da formalização, também leva-se em consideração os requisitos não funcionais estabelecidos na fase 2.

É importante lembrar que, ao escolher um cenário, outras fases, assim como atividades, podem ser adicionadas no desenvolvimento da ontologia.

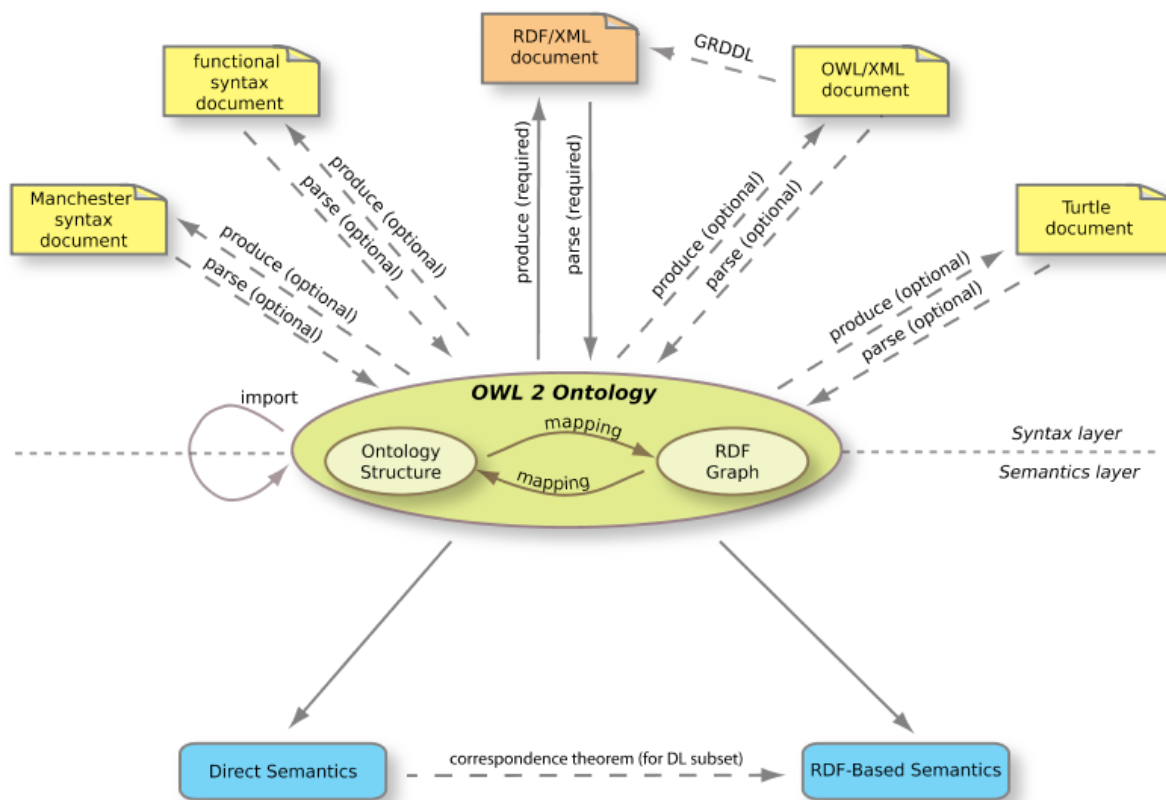
### 2.1.4 OWL

A *Ontology Web Language* (OWL) é uma linguagem de ontologias da web utilizada para desenvolver aplicações que usam como base a web semântica, usando características do RDF. A linguagem foi criada para que as aplicações possam processar o conteúdo de informações no lugar de apresentar somente as informações. Ela facilita a interpretação do conteúdo da Web pela máquina, fornecendo vocabulário adicional em conjunto com uma semântica formal. Possuindo mais facilidades para expressar significado e semântica quando comparado com XML, RDF e RDF-S. Ou seja, a linguagem OWL vai além dessas linguagens quando se trata de habilidade de representar conteúdo interpretável para a máquina (MCGUINNESS; HARMELEN et al., 2004).

Por causa da sua generalidade, RDF usa URIs como base do seu mecanismo "de identificação de sujeitos, predicados e objetos em declarações. Para ser mais preciso, RDF usa referências URIs. Uma referência URI é um URI juntamente com um identificador de fragmento opcional no final" (LIMA; CARVALHO, 2005a).

A Figura 8 apresenta os principais componentes para a criação de uma ontologia com o OWL e como eles se relacionam.

Figura 8 – A estrutura do OWL



A elipse no centro representa a noção abstrata de uma ontologia, que pode ser pensada como uma estrutura abstrata ou como um grafo RDF. No topo estão várias sintaxes concretas que podem ser usadas para serializar e trocar ontologias. Na parte inferior estão as duas especificações semânticas que definem o significado das ontologias OWL (CONSORTIUM et al., 2012).

De acordo com a W3C, a OWL fornece três sub-linguagens, onde cada um exerce um propósito distinto, atendendo diferentes usuários:

- **OWL EL** - Permite algoritmos de tempo polinomial para todas as tarefas de raciocínio padrão. É adequado para aplicações em que ontologias grandes são necessárias e onde a energia expressiva pode ser trocada por garantias de desempenho.
- **OWL QL** - Consiste na existência de consultas conjuntas que são respondidas no LogSpace (mais precisamente, AC0) usando a tecnologia de banco de dados relacional padrão. É adequado para aplicações em que ontologias leves são usadas para organizar um grande número de indivíduos e onde é útil, ou talvez necessário, acessar os dados diretamente por meio de consultas relacionais (por exemplo: SQL).
- **OWL RL** - Proporciona a implementação de algoritmos de raciocínio de tempo polinomial usando tecnologias de bancos de dados estendidos por regras, operando diretamente em triplas de RDF. É adequado para aplicações em que ontologias relativamente leves são usadas para organizar um grande número de indivíduos e onde é possível ou necessário operar diretamente em dados na forma de triplos de RDF.

## 2.1.5 Noções Básicas da OWL

### 2.1.5.1 Cabeçalho (Header)

Para criar uma ontologia em OWL é necessário definir a localização das classes primitivas, para que se possa gerar novas classes como subclasses. Sendo necessário também a determinação de uma namespace para a nova ontologia. Essas especificações ficam no cabeçalho da ontologia, também conhecido como Header (HU; LIU; WANG, 2004). A Figura 9 mostra um exemplo de codificação do mesmo.

As classes a serem definidas estarão localizadas no namespace da primeira definição. A segunda definição serve para que ontologias externas possam referenciar a ontologia sendo definida. As restantes localizam as definições primitivas de OWL, RDF, RDFS e XMLSchema.(FREITAS, 2003)

### 2.1.5.2 Indivíduos e Classes

Em OWL os indivíduos são representados pelos objetos que compõem um determinado domínio (HORRIDGE, 2011). A Figura 10 tem como objetivo representar indivíduos



Figura 9 – Codificação do Cabeçalho.

```

xml:base="http://www.semanticweb.org/x_rom/ontologies/2018/10/untitled-ontology-22"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
ontologyIRI="http://www.semanticweb.org/x_rom/ontologies/2018/10/untitled-ontology-22">
<Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#" />
<Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#" />
<Prefix name="xml" IRI="http://www.w3.org/XML/1998/namespace#" />
<Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#" />
<Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#" />

```

Fonte: Do Autor.

de um domínio.

Figura 10 – Representação de indivíduos.



Fonte: Horridge (2011)

As classes podem ser construídas de várias formas: por herança, união, intersecção, complemento, pela enumeração de instâncias ou por restrições de propriedades (FREITAS, 2003).

"As classes fornecem um mecanismo de abstração para agrupar recursos com características semelhantes. Como as classes RDF, cada classe OWL é associada a um conjunto de indivíduos, chamado de extensão de classe. Os indivíduos na extensão de classe são chamados de instâncias da classe. Uma classe tem um significado intensional (o conceito subjacente) que está relacionado, mas não é igual à sua extensão de classe. Assim, duas classes podem ter a mesma extensão de classe, mas ainda assim serem classes diferentes" (BECHHOFFER et al., 2004).

Existe uma classe geral chamada `<owl: Thing>`, que comporta todos os indivíduos e também é uma superclasse de todas as classes OWL. Assim como também existe outra classe mais específica chamada `<owl: Nothing>` que é a classe que não tem casos e uma subclasse de todas as classes OWL (JUNIOR, 2012).

Horridge (2011) exemplifica o conceito de Classe através da seguinte frase: Todo gato é um animal. Ao analisar a frase, é possível compreender que todo o "gato" é uma subclasse de "animal". A figura 11 representa o exemplo de (HORRIDGE, 2011).

Figura 11 – Representação em código da frase: Todo gato é um animal.

```

<Declaration>
  <Class IRI="#Animal"/>
</Declaration>
<Declaration>
  <Class IRI="#Gato"/>
</Declaration>
<SubClassOf>
  <Class IRI="#Gato"/>
  <Class IRI="#Animal"/>
</SubClassOf>

```

Fonte: Adaptado de Horridge (2011).

### 2.1.5.3 Camadas e Propriedades

O OWL é dividido entre 3 distintas camadas, onde cada uma tem a sua função e propriedade. São elas: (i) A camada de lógica que concede a definição de regras que irá operar nas instâncias e recursos, enquanto a (ii) camada de prova as executa, e a de (iii) confiança julga se a prova está correta ou não (KOIVUNEN; MILLER, 2001).

"Para que estas camadas entrem em operação, as camadas inferiores devem estar bem sedimentadas, o que ainda está acontecendo. Além do mais, sob o ponto de vista ontológico, não é interessante antecipar o uso de ontologias com regras, pois isto pode restringir a sua aplicabilidade. Porém regras podem ter utilidade para restringir atributos e exprimir axiomas"(FREITAS, 2003).

De acordo com Welty, McGuinness e Smith (2004) as propriedades são usadas para estabelecer relacionamentos entre entidades, ou entre indivíduos e valores de dados. Sendo assim, o OWL contém duas principais categorias de propriedade:

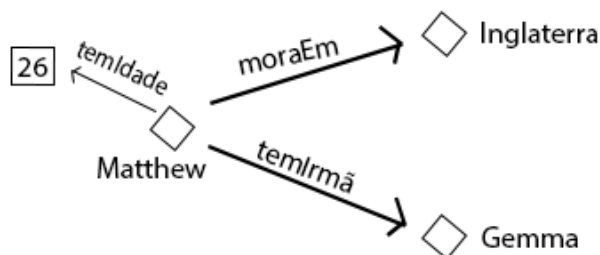
- Propriedades de dados: quando existe relação entre instâncias e valores de dados.
- Propriedades de objetos: relação entre instâncias.

Para exemplificar o princípio da propriedade, adaptou-se o exemplo de Horridge (2011) mostrando, através da figura 12, que a propriedade de objeto “temIrmã” liga o indivíduo “MAtthew” ao indivíduo “Gemma”. Assim como a propriedade de objeto “moraem” a entidade “Metthew” a entidade “Inglaterra”. Nota-se também que existe uma relação entre a entidade “Metthew” e o valor “26”, através de uma propriedade de dados.

Ainda por Welty, McGuinness e Smith (2004), existem cinco características que podem ser aplicadas nas propriedades, são elas:owl:TransitiveProperty, owl:SymmetricProperty, owl:FunctionalProperty owl:InverseFunctionalProperty e owl:inverseOf.

A OWL permite que sejam impostas restrições sobre propriedades. Uma restrição é um tipo especial de descrição de classe, isto é, descreve uma classe anônima de

Figura 12 – Exemplo de propriedade de objeto



Fonte: Adaptado de Horridge (2011).

indivíduos que satisfazem as restrições. As restrições podem ser de valores (`allValuesFrom`, `someValuesFrom` e `hasValue`) ou de cardinalidade (`maxCardinality`, `minCardinality` e `Cardinality`) (WELTY; MCGUINNESS; SMITH, 2004).

### 2.1.6 SPARQL

*SPARQL Protocol and RDF Query Language* (SPARQL), projetado pela W3C para trabalhar com a web semântica, é um conjunto de linguagens e protocolos usados para a recuperação e manipulação de dados que estão contidos em grafos RDF, permitindo que usuários possam realizar consultas em banco de dados ou em fontes que possam ser visualizadas como RDF via middleware. Por exemplo, um banco de dados relacional pode ser consultado com o SPARQL usando um software de mapeamento de bancos de dados relacional para RDF (RDB2RDF).

"Tentar usar o potencial da web semântica sem SPARQL é o mesmo que tentar utilizar um banco de dados relacional sem usar a linguagem SQL. O SPARQL possibilita a consulta de informações de bancos de dados e outras fontes diversas no mundo, em toda a web"(BERNERS-LEE; HENDLER; LASSILA, 2001, 5.3).

SPARQL não se prende a apenas um banco de dados, dado que o mesmo permite o envio de consultas e recuperação de informações através de terminais (endpoints), pois o mesmo é um protocolo de transporte que tem como base o HTTP, podendo comunicar-se com qualquer terminal SPARQL através de uma camada de transporte padronizada. Obtendo resultados em RDF que podem ser retornados em diversos formatos de intercâmbio de dados, tendo as entidades identificadas por URIs (identificador de recursos universal).

O padrão compreende as seguintes especificações: uma linguagem de consulta para RDF; uma especificação que define uma extensão do SPARQL Query Language para executar consultas distribuídas em diferentes terminais SPARQL; uma especificação que define a semântica de consultas SPARQL sob regimes de vinculação, como RDF Schema, OWL, ou RIF; um protocolo que define os meios para transmissão de consultas SPARQL arbitrárias e solicitações de atualização para um serviço SPARQL; uma especificação que define um método busca e descoberta e um vocabulário para descrever serviços SPARQL e

um conjunto de testes, para avaliação da especificação SPARQL (HARRIS; SEABORNE; PRUD'HOMMEAUX, 2013).

### 2.1.6.1 Termologia

A linguagem SPARQL também trabalha com IRIs, um subconjunto de Referências de URI RDF que omite espaços. Todas as IRIs nas consultas SPARQL são absolutas, eles podem ou não incluir um identificador, IRIs incluem URIs e URLs. A seguir, os elementos no qual o SPARQL suporta em suas cláusulas: s termos a seguir são definidos no RDF e usados no SPARQL:

- IRI (corresponde a referência URI RDF)
- literal
- forma lexical
- literal simples
- tag de idioma
- literal digitado
- nó em branco

Harris, Seaborne e Prud'hommeaux (2013) define os seguintes termos:

- Termo RDF - Inclui IRIs, nós vazios e literais
- Literal simples - Abrange literais sem tag de idioma ou tipo de dados IRI

### 2.1.6.2 Consultas SPARQL

Uma consulta SPARQL consiste em um conjunto de padrões de triplas, chamado de grafo básico. Esse padrão é parecido com o sistema de triplas do RDF, sendo composto por sujeito, predicado e objeto, a diferença é que os elementos podem ser uma variável. As soluções para as variáveis são encontradas combinando os padrões na consulta com triplas no conjunto de dados. Um padrão grafo básico é equivalente a um subgrafo dos dados RDF quando os termos RDF desse subgrafo podem ser substituídos pelas variáveis e o resultado é um grafo RDF equivalente ao subgráfico (HARRIS; SEABORNE; PRUD'HOMMEAUX, 2013).

O Protocolo SPARQL apresenta uma grande quantidade de funcionalidades, que podem ser conferidas em SPARQL:2013. Para que se possa realizar o processo de consulta é necessário compreender a estrutura e os componentes da mesma, segue, em ordem, os elementos que constituem tal estrutura:

- Declarações de prefixos, para abreviar URIs;
- Definição do conjunto de dados, informando quais grafos(s) RDF estão sendo consultados;
- A cláusula de resultado, identificando que informação deve ser retornada a partir da consulta;
- O padrão de consulta, especificando o que consultar dentro do conjunto de dados
- Modificadores de consulta, limites, ordenação, e outros que podem modificar o resultado final

De acordo Harris, Seaborne e Prud'hommeaux (2013), a linguagem especifica quatro formas distintas de consulta, onde cada uma tem sua finalidade:

- SELECT query - Usado para extrair valores brutos de um terminal SPARQL, os resultados são retornados em um formato de tabela.
- CONSTRUCT query - Usado para extrair informações do terminal SPARQL e transformar os resultados em RDFs válidos.
- ASK query - Usado para fornecer um resultado booleano (verdadeiro / falso) simples para uma consulta em um terminal SPARQL.
- DESCRIBE query - Usado para extrair um gráfico RDF do terminal SPARQL, cujo conteúdo é deixado para o terminal para decidir com base no que o mantenedor considera como informações úteis.

### 2.1.6.3 Exemplos

Abaixo um exemplo, publicado por Harris, Seaborne e Prud'hommeaux (2013), de consulta em SPARQL cujo objetivo é encontrar o título de um determinado livro através dos grafos de dados fornecidos. A consulta é composta por duas partes:

- Select - Usada para identificar as variáveis a serem exibidas nos resultados da consulta;
- WHERE - Usado para fornecer o padrão grafo básico para corresponder ao grafo de dados.

O padrão grafo básico neste exemplo consiste em um único padrão de triplas com uma única variável na posição do objeto, sendo este o título .

Figura 13 – Objeto

```
<http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title> "SPARQL Tutorial"
```

Fonte: Harris, Seaborne e Prud'hommeaux (2013)

Figura 14 – Consulta SPARQL

```
SELECT ?title
WHERE
{
  <http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title> ?title .
}
```

Fonte: Harris, Seaborne e Prud'hommeaux (2013)

Figura 15 – Resultado obtido através da consulta da figura 14

title
"SPARQL Tutorial"

Fonte: Harris, Seaborne e Prud'hommeaux (2013)

## 2.2 Nanopublicações

é uma recente arquitetura ontológica proposta por Mons e Velterop (2009) que, através de tecnologias semânticas, traz benefícios para a recuperação e reutilização de descobertas científicas. A ideia principal da criação da nanopublicação é: subdividir os resultados científicos em pequenas partes denominadas afirmações; representar essas afirmações em uma notação formal baseada em RDF; anexar informações de proveniência neste nível atômico; tratar cada uma destas pequenas entidades como uma publicação separada.

Uma vez que os clássicos processos de indexação já não satisfazem aqueles que estão em busca de informações em ambientes digitais, ao considerar o atual estado dos acervos virtuais, os usuários contemporâneos estão interessados em trechos ou segmentos de um documento de pesquisa (MOURA, 2009). Diferentemente das atuais publicações narrativas encontradas nos repositórios científicos, que não fazem uso de nenhuma técnica de marcação ou destaque do seu conteúdo, a nanopublicação evidencia trechos de afirmações, tais como resultados provindos de estudos científicos, permitindo a legibilidade de máquinas para que mais tarde estas informações possam ser recuperados facilmente e reutilizadas por pesquisadores de forma segura (GOLDEN; SHAW, 2015). Sendo assim, a nanopublicação pode ser considerada como uma pequena unidade de afirmação publicável que, assim como um documento tradicional de pesquisa, precisa ser citável, atribuível e passível de revisão. A nanopublicação também deve ser facilmente agregada e identificada em toda a Web, além de precisar ser facilmente recuperada (GROTH; GIBSON;

VELTEROP, 2010a).

Visto que a estrutura da nanopublicação permite que pesquisadores possam referenciar os autores de tais afirmações em atribuições e citações, de forma que exista uma conexão com a fonte da afirmação, usa-se dessa função como forma incentivadora para que pesquisadores possam disponibilizar suas publicações científicas em formatos padrões que impulsionam a acessibilidade e a interoperabilidade dos dados (GOLDEN; SHAW, 2015).

### 2.2.1 Modelo

Apesar de ter sido originada no campo da Bioinformática, com o objetivo de favorecer as pesquisas feitas nas áreas das Ciências da Vida, a nanopublicação foi projetada para ser extensível, de modo que os elementos possam ser modificados ou criados, conforme necessário (SCHULTES et al., 2012). Deixa-se claro que, independente do domínio de conhecimento no qual se aplicará a nanopublicação, é necessário que exista uma boa descrição do conhecimento para que se possa ter uma boa compreensão, reutilização e recuperação das informações desejadas. Para isso, a criação das nanopublicações devem ser guiadas por ontologias que descrevem e estabelecem relações entre os elementos e conceitos de um determinado domínio (SCHULTES et al., 2012). Assim, levando em conta os vários formatos que podem ser usados na hora de criar uma nanopublicação, é necessário que alguns requisitos básicos sejam atendidos:

- capacidade de identificar exclusivamente um conceito.
- capacidade de identificar exclusivamente uma declaração.
- capacidade de se referir a todos os conceitos unicamente identificados e declarações.

Considera-se também que durante a criação de uma nanopublicação é importante considerar a identificação da origem das declarações, do responsável pelas declarações e da entidade responsável pela criação das nanopublicações. Essas identificações são necessárias para dar confiabilidade a nanopublicação criada (SCHULTES et al., 2012).

De acordo com o portal nanopub.org, os 4 princípios de criação de uma nanopublicação são:

- **Princípio do Valor** - A afirmação surge de um procedimento ou observação bem documentada. Por exemplo, o Predicado estabelecendo uma associação entre o Sujeito e o Objeto poderia surgir de um modelo matemático, co-ocorrência em texto, um novo conjunto de dados experimental etc.
- **Princípio da Transparência** - A Proveniência e Condição referem-se a quem, o que, onde, quando do Predicado, permitindo que a qualidade da nanopublicação seja avaliada por outros.

- **Princípio da Evitação da Ambiguidade** - Os argumentos da nanopublicação (todos os conceitos na afirmação, na condição e na proveniência) podem ser colocados em conceitos únicos.
- **Princípio da Referência Global** - Onde, autoridade, espaço de nomes, acesso e versão de qualquer argumento de nanopublicação já foi estabelecido na Web, o Identificador único de Recursos (URI) do conceito deve ser usado. Onde não existe nenhum URI, um Identificador Universal e Exclusivo (UUID) pode ser gerado usando o ConceptWiki.

Diante as capacidades e princípios descritos, Schultes et al. (2012) propõe três níveis minimamente necessários em uma nanopublicação, sendo eles:

- **Identificação da Nanopublicação:** Cada nanopublicação é considerada única, e tem sua própria IRI.
- **Asserção (*Assertion*):** A unidade mínima de uma afirmação deve ser expressa através de uma tripla, ou seja, deve evidenciar o relacionamento entre dois conceitos (Sujeito e Objeto) usando um terceiro conceito (Predicado). Essa unidade pode ser composta por uma ou mais triplas, de acordo com a ontologia definida (nanopub.org). A figura 16, exemplifica o nível de asserção.

Figura 16 – Exemplo de uma unidade mínima de afirmação

```
:Asserção {
  ex:trastuzumabe ex:é-indicado-para ex:câncer-mama
}
```

Fonte: adaptado de nanopub.org

- **Proveniência da Asserção:** Consiste em metadados que fornecem informações sobre a origem das afirmações contidas na unidade "asserção", o que inclui informações como a identificação dos autores, instituições, registros de tempo, concessões, links para DOIs, URLs sobre a afirmação. Ou seja, quem a gerou, quando foi gerada, onde a afirmação foi obtida e qualquer outra informação similar. A Figura 17 exemplifica a proveniência da asserção.
- **Proveniência da Nanopublicação:** Essa unidade pertence tanto à asserção quanto à proveniência. Contém uma ou mais triplas RDF que são semelhantes as triplas da proveniência, porém se referem a nanopublicação como um todo. A Figura 18 exemplifica uma proveniência de nanopublicação.

Em NPs são necessárias boas descrições para que se possa ter uma boa compreensão, reutilização e recuperação das informações. Por isso é necessário uma abordagem



Figura 17 – Exemplo de proveniência da asserção

```

:proveniência {
  :asserção prov:geradonotempo "02-03-2012T14:38:00".
  :asserção prov:foiDerivadoDe :experimento.
  :asserção prov:foiAtribuidoA :experimentoCientifico.
}

```

Fonte: adaptado de nanopub.org

Figura 18 – Exemplo de Proveniência da Nanopublicação

```

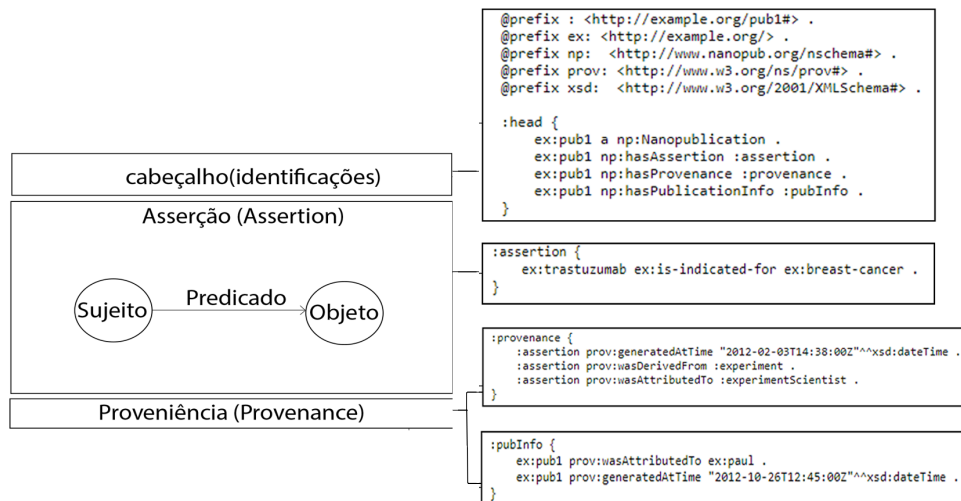
:pubInfo {
  ex:pub1 prov:foiAtribuídoA ex:paul .
  ex:pub1 prov:geradoNoTempo "26-10-2012T12:45:00".
}

```

Fonte: adaptado de nanopub.org

orientada pela comunidade para representar a estruturação dos dados (nanopub.org). A figura 19 representa a arquitetura mínima de uma nanopublicação, cujos níveis já foram detalhados.

Figura 19 – Arquitetura mínima da nanopublicação



Fonte: adaptado de Schultes et al. (2012)

A nanopublicação deve ser representada, e pode ser consultada, através de tecnologias semânticas (RDF, OWL, SPARQL). É interessante ressaltar que, através da proveniência, a técnica deve garantir a integridade dos dados e sua proveniência. Ressalta-se também que o acesso as informações, explicitadas na camada de proveniência, permite que os usuários avaliem a confiabilidade dos dados (nanopub.org).

### 2.3 Análise de Redes Sociais

A análise de redes sociais (ARS) é uma abordagem fundamentada através de modelos matemáticos e teoria dos grafos, que estuda um determinado conjunto de nós interligados através de arestas, como se fosse uma teia que ocupa um espaço em um estipulado ambiente (NEWMAN, 2010). Estes nós representam atores ou pontos, que correspondem a entidades, pessoas, empresas ou organizações, que podem ser analisados como unidades individuais ou sociais coletivas (WASSERMAN; FAUST, 1999).

A relação entre esses nós tem várias denominações, como: vínculo, ligação, arco, interação e através das ligações dos nós é possível representar diversas situações em inúmeras ciências e áreas de interesse (FARINA, 2004).

“Muito do que se fala e do que se constata nos diversos ambientes que frequentamos parece estar estruturado como uma rede. Para a Economia e o Marketing, há o interesse em saber quem vende e quem compra. Já na Biologia, há o interesse em saber quem se alimenta de quem, quando se estuda a cadeia alimentar. O cérebro faz ligações entre neurônios; as sinapses, para que a pessoa lembre ou resolva algum problema ou questão. A internet é uma rede na qual as pessoas se conectam e se comunicam. As doenças podem se propagar de uma pessoa para outras, deflagrando uma epidemia” (FARINA, 2004).

Um modelo de rede de atores é exemplificado através da Figura 20. Esta rede apresenta seis atores relacionados através de relações de confiança. Sendo assim, é possível observar que o ator 1 confia no ator 2 e no ator 3, mas não confia nos atores 6, 4 e 5, apesar de o ator 5 confiar nele. Os atores 2 e 3 também confiam no ator 1. Ninguém da rede confia no ator 6, o mesmo ator também não confia nos outros atores. Outra possibilidade de analisar essa rede é pensar no oposto, por exemplo: o ator 1 tem a confiança dos atores 2, 3 e 5, porém o ator 4 não confia nele (FARINA, 2004).

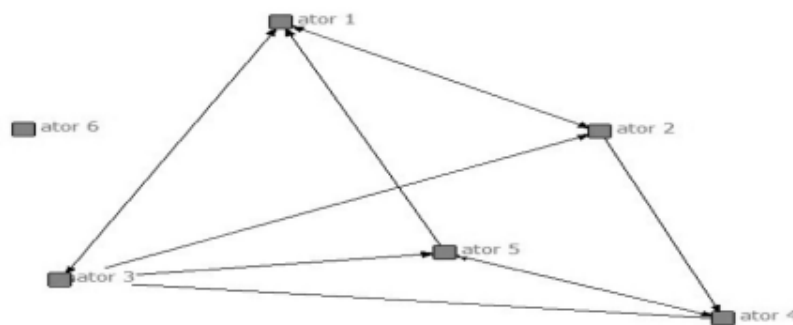


Figura 20 – Exemplo de Rede de Atores

De acordo com Souza e Alvarenga (2004), os tipos mais comuns de laços que ligam os nós encontrados em uma rede são: a avaliação individual (por exemplo, amizade ou respeito); a transação e a transferência de recursos materiais (uma transação de compra e venda entre duas empresas); a transferência de recursos não materiais (a troca de mensagens eletrônicas) ou não; a associação ou afiliação que ocorre quando os atores participam de eventos em comum (festas); a interação (sentar-se próximo a outra pessoa); a movimentação e a conexão física e social; laços entre papéis formais (laço de autoridade chefe-subordinado em uma empresa); relações biológicas (pai e filho).

Quando se trata de métricas, a análise de redes sociais conta com diversas e distintas estratégias e ferramentas, cada uma com o objetivo de se obter diferentes informações sobre as relações. A seguir, algumas métricas básicas usadas na ARS:

- **Centralidade:** de acordo com Kosorukoff (2011), a métrica de centralidade tem em vista buscar os atores mais importantes da rede, ela abrange mais três métricas que permitem identificar e quantificar tal importância, são elas:
  - **Intermediação:** Permite medir o grau de intervenção de cada ator relativamente a outros atores da rede.
  - **Proximidade:** A proximidade é um dos conceitos básicos de um espaço topológico. Intuitivamente, é possível dizer que dois conjuntos estão próximos se eles estão arbitrariamente próximos uns dos outros.
  - **Grau:** A centralidade de grau é definida pelo número de ligações que um nó possui.
- **Densidade:** Proporção de laços que um indivíduo contém.
- **Coesão:** O grau em que os atores estão conectados diretamente uns aos outros por laços coesivos. Os grupos são identificados:
  - **clique** - é considerado como um subgrupo de uma rede, pois os atores mantêm relações mais fortalecidas e estão mais densamente conectados em relação aos demais membros da mesma rede. Assim, para que se possa caracterizar a existência de um clique, deve-se existir, no mínimo, três atores que se conectem intensamente (MARTELETO; TOMAÉL, 2005).
  - **Círculos sociais** - Dois ou mais atores que interagem entre si e compartilham características semelhantes.
- **Equivalência Estrutural:** Refere-se a um mesmo conjunto comum de links que um grupo de nós compartilham entre si.
- **Alcance:** Mede a extensão do contacto que um ator estabelece com outros atores na rede.

- Transitividade: Mede o grau de flexibilidade e cooperação duma determinada rede.

### 2.3.1 Análise de Redes Sociais e a Colaboração Científica

O método de análise de redes sociais também é utilizado para estudar as relações de colaboração científica entre pesquisadores. De acordo com Sonnenwald (2007), a colaboração científica é definida pela interação de um conjunto de cientistas em um determinado contexto social, que proporciona o compartilhamento de atividades científicas para que se possa atingir objetivos comuns .

Para Meadows e Lemos (1999), observa-se um aumento no número de pesquisas científicas compartilhadas, principalmente através das atuais facilidades de comunicação oferecidas pelo intercâmbio de informações web. Além disso, o trabalho realizado a partir da colaboração científica proporciona economia de tempo, de recursos financeiros e materiais. Proporcionando estímulos por parte das agências financiadoras de pesquisas.

A maior parte do avanço da ciência ocorre devido a interação entre os cientistas Meadows e Lemos (1999). Para Digiampietri et al. (2014) é fundamental a inserção efetiva dos cientistas em redes acadêmicas para que se possa ter evolução de temas científicos, e não apenas o financiamento da ciência em centros.

"Mapear uma área do conhecimento a partir da análise de suas redes sociais passa a se tornar tarefa viável devido a quantidade de dados hoje disponíveis e que podem ser coletados de forma automática. Mapeamento múltiplo que pode fornecer dados relevantes sobre várias camadas em que essas relações ocorrem: estratégias de conectividade utilizadas, políticas de relação, ações recorrentes denotando padrões em potencial, linguagem e expressões características, objetos de pesquisa preferenciais, modos de relação entre instituições, departamentos, grupos de pesquisa, pesquisadores, professores e alunos. A rede vira objetivo de estudo, artefato de pesquisa e espaço de intervenção por onde se materializa e se atua nos fluxos de comunicação, nas estruturas e dinâmicas de conversação, nos modos de articulação e gestão"(MARTINS, 2012).

A análise de redes de colaboração científica permite compreender as estratégias e extrair informações sobre a organização dos pesquisadores em diversas áreas de conhecimento (MARTINS, 2012). Para Newman (2001), esse tipo de rede forma pequenos mundos e seus pares, escolhidos de forma aleatória, são relacionados a partir de intermediários, como dois cientistas ligados numa mesma rede de colaboração científica se tiverem, em comum, autores com os quais dividem a autoria.

De acordo com Vanz e Stump (2010), o colaborador de uma Rede Social de Colaboração Científica pode ser qualquer indivíduo que contribui diretamente para uma parte de uma pesquisa.

Como exemplos de publicações na subárea de Colaboração científica, podemos citar:

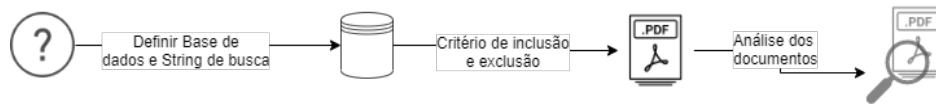
- 
- Hayashi, Hayashi e Lima (2008) que realizou um estudo que busca identificar e explicitar as conexões entre os atores do PPGEEs na produção do conhecimento científico veiculado por quatro tipos de publicações: artigos de periódicos, livros, capítulos de livros, e trabalhos completos apresentados em eventos.
  - Mendonça et al. (2018) que gerou redes de colaboração científica da Comunidade Brasileira de Interação Humano Computador, analisando os 29 autores mais prolíficos da comunidade de IHC. Onde, ao final do trabalho, os autores constatam que, através de análise de bibliométrias e técnicas de redes sociais, existe uma forte coesão da comunidade de IHC ao se identificar inúmeras colaborações entre seus autores.
  - Alvarez e Vanz (2017) que apresenta uma análise bibliométrica fundamentada no indicador de colaboração institucional dos artigos científicos brasileiros de Física de Altas Energias (FAE) indexados pelo Science Citation Index (SCI) da Web of Science (WoS) entre 1983 e 2013.



### 3 TRABALHOS RELACIONADOS

Os trabalhos relacionados foram recuperados através do protocolo de mapeamento sistemático proposto por Petersen et al. (2008).

Figura 21 – Processo do mapeamento



Fonte: Adaptado de Petersen et al. (2008)

A figura 21 apresenta o processo utilizado nesse mapeamento, primeiro foi necessário definir as questões de pesquisa. Logo após, foram definidas as bases de dados e string de busca. Os documentos retornados passaram por critérios de exclusão e inclusão, a fim de manter somente os documentos que trabalhavam com nanopublicações. Logo após, foi realizada uma análise nos documentos para que se pudesse extrair os dados necessários para responder as questões de pesquisa definidas. Durante a análise dos documentos, foi realizado a atividade de keywording, como sugere a metodologia.

#### 3.1 Mapeamento Sistemático

##### 3.1.1 Definição das questões de pesquisa

Através deste mapeamento, buscou-se a resposta para a questão central deste trabalho: Como criar e avaliar nanopublicações? Para responder este questionamento foram definidas as seguintes questões específicas de pesquisa:

- Q1 - Quais os domínios dos artigos encontrados?
- Q2 - Qual a fonte de dados da nanopublicação?
- Q3 - Quais os métodos e ferramentas de extração de nanopublicação são utilizados?
- Q4 - Quais ontologias foram usadas para extrair a Proveniência e a Asserção?
- Q5 - Quais as ferramentas computacionais estão sendo utilizadas para a criação e recuperação das nanopublicações? Qual a sintaxe de representação do RDF e banco de dados?
- Q6 - Quais os métodos de avaliação utilizados?

A partir do primeiro questionamento, pretende-se descobrir o domínio das publicações que estão presentes no cenário de nanopublicações. Através desta resposta será possível perceber o quão aplicado está sendo este modelo de publicação, se o mesmo tem sido utilizado como uma solução real ou se ainda se encontra apenas como propostas.

Através da segunda questão, tem-se como objetivo identificar quais as fontes de dados estão sendo utilizadas para a criação de possíveis nanopublicações. Com esta informação será possível descobrir se os dados utilizados já estavam em modelo de triplas RDF, em artigos PDF ou até mesmo em documentos físicos, ou seja, em documentos não estruturados ou, simplesmente, pesquisas registradas em papéis.

O terceiro questionamento traz como resposta as ferramentas utilizadas para a extração de nanopublicações. A partir desta resposta, foi possível analisar a possibilidade de fazer uso de uma dessas ferramentas para o processo de extração de nanopublicações propostas por este trabalho.

A questão de número quatro tem como objetivo identificar as ontologias usadas na criação das nanopublicações. Através desta resposta foi possível tomar decisões sobre qual a ontologia mais adequada a ser usada na criação de nanopublicações de Análise de Redes.

A quinta questão irá mostrar quais ferramentas, sintaxes, bancos de dados e bibliotecas foram utilizadas para a criação, armazenamento e recuperação das nanopublicações dos trabalhos analisados. Estas informações irão abrir um leque de ferramentas computacionais que poderão ser analisadas de forma que se possa eleger a mais adequada para este trabalho, assim como para trabalhos futuros.

Diante a literatura já estudada sobre a tecnologia, a explicitação das metodologias de avaliação sobre as nanopublicações foram poucas vezes abordadas, criando dúvidas sobre qual seria a forma mais adequada de avaliar as nanopublicações criadas neste trabalho. Por este motivo, a sexta e última questão se refere as metodologias de avaliação abordadas nos trabalhos analisados, a fim de explorá-las e talvez adotá-las neste trabalho.

### 3.1.2 Definição das bases de dados e da *string* de busca

Após definir as questões a serem respondidas, mostrou-se necessário a escolha dos repositórios científicos a serem explorados. Decidiu-se realizar as buscas pelos artigos científicos nos seguintes repositórios: ACM, IEEE xplora, Scielo, Science Direct, Scopus e Google Acadêmico.

Para esta busca, utilizou-se a seguinte string em todos os repositórios. Através da string de busca, pretende-se retornar quaisquer artigos científicos, contidos nos repositórios, que tenham citado nanopublicações em seu corpo.

Nanopublication OR Nanopublicação

A tabela 2 explicita a quantidade de publicações retornadas em cada repositório. Observa-se, portanto, que foram retornadas o total de 578 publicações.

É possível notar, através da tabela 2, que o google acadêmico, como esperado, foi o repositório que retornou mais publicações, já que ele indexa documentos de diversos repositórios. Em contraste temos o repositório Scielo, que retornou zero publicações.



Tabela 1 – Número de publicações retornadas em cada repositório.

Repositório	Quantidade
ACM	4
IEEE	9
Scielo	0
Science Direct	28
Scopus	38
Google Acadêmico	499

Fonte: Do Autor.

### 3.1.3 Critérios de inclusão e exclusão

Após coletar as publicações retornadas, aplicou-se sobre elas critérios de inclusão e exclusão, afim de realizar as análises, propostas por este mapeamento. somente em artigos que pudessem trazer informações relevantes. Os critérios estabelecidos estão explicitados na tabela 2. Após aplicar os critérios restaram 57 publicações.

Tabela 2 – Critérios de inclusão e exclusão dos documentos coletados

Inclusão	Exclusão
Estudos que estão relacionados a Nanopublicações	Estudos irrelevantes para a pesquisa, ou seja, publicações que não trabalham diretamente com nanopublicações.
Estudos que foram escritos nas linguagens Português ou Inglês.	Publicações duplicadas. Estudos sem abstract. Estudos que não abordam nanopublicações em seu Abstract ou palavras-chave. Artigos com revisões de literatura Documentos que não estejam disponíveis para a biblioteca da universidade.

Fonte: Do Autor.

### 3.1.4 Análise dos documentos

A partir de uma breve análise realizada nos artigos restantes, foi possível obter informações como a frequência anual de publicação sobre a abordagem de nanopublicação (figura 22) e a quantidade de artigos por domínio.

Como pode ser visto no gráfico representado pela figura 22, o ano com mais publicações realizadas foi 2015, com um total de 21 publicações. Seguido pelos anos de 2012 e 2014, com 9 publicações. Os anos com menos publicações foram 2009 e 2010, visto que em

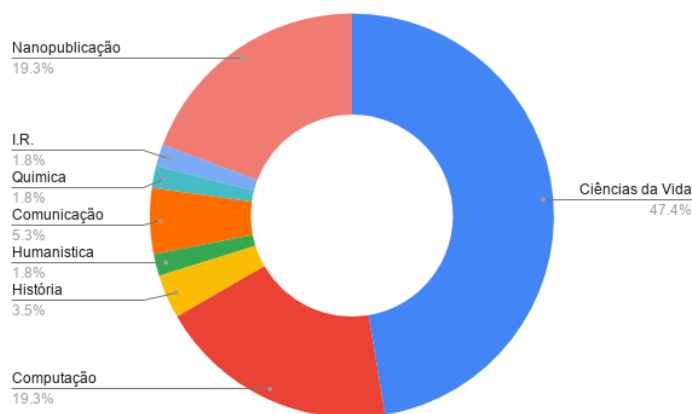
Figura 22 – Frequência de publicação anual



Fonte: Do Autor.

2009 foi o ano da proposta da nanopublicação e 2010 a primeira proposta de arquitetura da abordagem.

Figura 23 – Quantidade de publicações por domínio



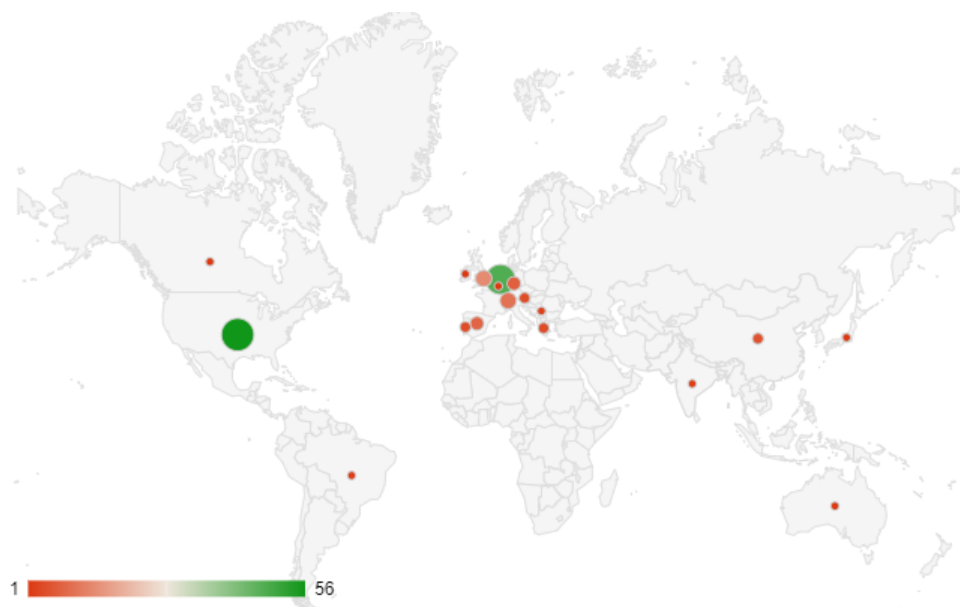
Fonte: Do Autor.

Ao analisar o gráfico representado pela figura 23 foi possível responder a Q1 e perceber que a maioria dos artigos sobre nanopublicação está concentrada no domínio de origem da abordagem, Ciências da Vida. Foram encontrados 27 artigos, que representam 47.4% das publicações coletadas. Algo interessante de destacar é a quantidade de domínios encontrados relativamente pequena, visto que 19.3% das publicações são artigos cujo foco são somente as nanopublicações.

Através da afiliação dos autores dos artigos foi possível obter a quantidade de autores por país (Figura 24). É possível perceber que o país com mais pesquisadores com

publicações sobre nanopublicações é os Estados Unidos, com 56 autores. Seguido pela Holanda, com 48. Enquanto que nos outros países o número de pesquisadores é menor que 15 para cada país marcado no mapa.

Figura 24 – Quantidade de autores por país



Fonte: Do Autor.

Também foi possível evidenciar informações sobre a colaboração científica dos autores de nanopublicações e aplicar algumas métricas de análise de redes (Figura 25). Para obter informações sobre a frequência de colaboração entre os autores e a frequência com a qual eles publicaram, foi criada uma rede social com as informações de coautoria dos artigos para posterior análise com a aplicação Gephi.

A figura 25 representa a rede criada com os autores dos artigos coletados. A rede conta com 167 autores (nós) que se relacionam através de 867 arestas. Ao aplicar métricas de rede, observa-se que a rede tem densidade de 0.063, indicando uma rede pouco densa. A rede contém 17 componentes conectados, mostrando que existem alguns grupos cujos autores estão conectados dentro do grupo, mas desconectados dos demais grupos. A tabela 3 apresenta os 5 autores com maior centralidade de grau da rede, ou seja, os que mais colaboraram.

Ao retomar o protocolo de Petersen et al. (2008), foi realizada a classificação dos artigos através de um processo de *keywording* que foi dividido em duas etapas: categorizar os documentos restantes e desenvolver um conjunto de palavras-chaves que reflitam a contribuição de cada artigo. Esta atividade permite uma classificação mais rápida dos artigos coletados, classificação que pode servir como um guia de quais artigos podem ser utilizados para a extração das informações necessárias para responder as questões de pesquisa.

Figura 25 – Rede de Atores



Fonte: Do Autor.

Tabela 3 – Os 5 atores com maior Centralidade de Grau

Autor	Centralidade de Grau
Marco Roos	48
Barend Mons	47
Mark Thompson	41
Rajaram Kaliyaperumal	41
Eelke Van Der Horst	41

Fonte: Do Autor.

Para este trabalho adotou-se as mesmas categorias de documentos científicos presentes na metodologia, representadas e descritas na tabela 4.

Para a criação das palavras-chave utilizadas para este trabalho, foram analisados todos os resumos dos trabalhos, onde cada trabalho deveria ter ao menos uma palavra-chave que representasse a contribuição do trabalho. Ou seja, se um determinado artigo tinha como objetivo criar nanopublicações de um determinado domínio, a palavra-chave atribuída a esse documento seria "Nanopublicação", visto que esse seria a principal colabo-

Tabela 4 – Tabela de Categorias adotadas para os documentos restantes

<b>Categoria</b>	<b>Descrição</b>
Artigos de Validação	Artigos cujas técnicas investigadas são novas e ainda não foram implementadas na prática. As técnicas usadas são, por exemplo, experimentos, ou seja, trabalho realizado em laboratório.
Artigos de Avaliação	Artigos onde técnicas são implementadas na prática e uma avaliação da técnica é realizada. Ou seja, é mostrado como a técnica é implementada na prática (solução implementação) e quais são as consequências da implementação em termos de benefícios e desvantagens (avaliação da implementação). Isso também inclui identificar problemas na indústria.
Proposta de Solução	Artigo onde solução para um problema é proposta, a solução pode ser nova ou significativa extensão de uma técnica existente. Os benefícios potenciais e a aplicabilidade da solução é mostrada por um pequeno exemplo ou uma boa linha de argumentação.
Artigos Filosóficos	Estes documentos esboçam uma nova maneira de ver as coisas existentes, estruturando o campo em forma de taxonomia ou estrutura conceitual.
Artigos de opinião	Estes documentos expressam a opinião pessoal de alguém, se uma determinada técnica é bom ou ruim, ou como as coisas devem ser feitas. Eles não contam com trabalhos relacionados e metodologias de pesquisa.
Artigos de experiência	Os documentos de experiência explicam o que e como algo foi feito na prática. Tem ser a experiência pessoal do autor.

Fonte: Petersen et al. (2008)

ração dele. Caso a leitura do resumo não fosse o suficiente para a identificação da palavra-chave, a leitura da conclusão era realizada. Ao obter um conjunto de palavras-chave, foram selecionadas as mais significativas. De forma que, através de um entendimento de alto nível, fosse possível identificar contribuição dos artigos. A tabela 5 representa o conjunto de palavras chaves selecionadas para este trabalho, acompanhadas de suas descrições.

Após a escolha das palavras-chave a serem usadas no mapeamento, utilizou-se uma planilha<sup>1</sup> *online* para atribuir uma palavra-chave e categorizar cada artigo, seguido de um breve comentário sobre o motivo da categoria escolhida. A figura 26 representa um gráfico de dispersão com base nos dados atribuídos para cada documento.

Como pode ser observado no gráfico, em ordem decrescente, foram obtidos 19 documentos representados pela palavra-chave Nanopublicação, sendo todos eles classificados como artigos de proposta de solução; 10 artigos representados pela palavra-chave Nanopublicação e categorizados como artigos de experiência; 5 artigos representados pela palavra-chave Ontologia e categorizados como propostas de solução; 4 artigos com palavra-chave URI e categorizados como propostas de solução. Foram encontrados 3 artigos com a palavra-chave Nanopublicação categorizados como artigos de avaliação e artigos de opi-

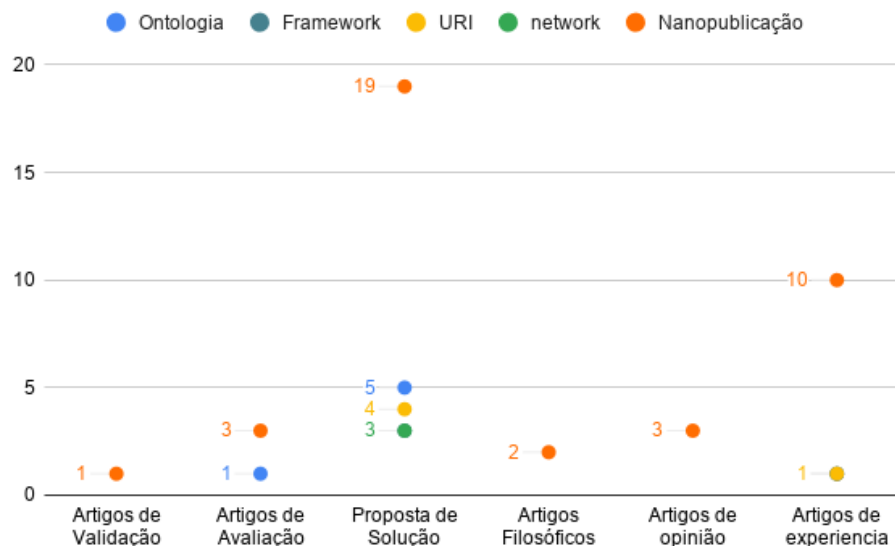
<sup>1</sup> <http://tiny.cc/k3qegz>

Tabela 5 – Tabela de palavras-chaves

Palavra-Chave	Descrição
Ontologias	Representa os artigos que relatam trabalhos ligados somente a ontologias, cujo existe a pretensão de criar nanopublicações, porém a criação das mesmas não é relatada. Ex.: O uso de uma determinada ontologia em uma determinado nível da nanopublicação. Ou a extração de informações através de uma determinada ontologia, afim de criar nanopublicações.
Framework	Representa os artigos que relatam, de alguma forma, o uso de frameworks para a manipulação de nanopublicações.
URI	Representa os estudos que, apesar de trabalharem com nanopublicações, tem como focoprincipal a URI.
Redes	Representa os estudos que tem como foco principal o domínio, ou subdomínio, de Redes de computadores aplicado a nanopublicações.
Nanopublicação	Representa os artigos que tem como foco o trabalho direto com nanopublicações. Ex.: Documento que relatam a criação de nanopublicações, ou discutem sobre asua arquitetura.

Fonte: Do Autor.

Figura 26 – Gráfico de dispersão com dados referentes ao mapeamento entre as palavras-chaves e as categorias de artigos



Fonte: Do Autor.

nião, seguidos de 2 artigos representados pela mesma palavra-chave, porém categorizados como artigos filosóficos. Foi obtido 1 artigo de palavra-chave Nanopublicação e 1 o URI, classificados como artigos de validação e experiência, respectivamente. Por fim, 1 artigo

com a palavra-chave *Ontologia* e *Framework* que, juntamente com 3 artigos representados pela palavra-chave *Framework*, não aparecem no gráfico devido a sobreposição dos dados, onde os dois primeiros são categorizados como artigos de experiência e os 3 últimos fazem parte dos artigos de proposta de solução.

Após realizar o mapeamento dos documentos, conforme as palavras-chave e as categorias dos artigos, foi criado um novo conjunto de palavras-chave. Este novo conjunto é composto por palavras que representam as variedades de contextos que podem ser atribuídos aos documentos, juntamente as palavras-chave principais. Através deste novo conjunto, foi possível obter artigos mais específicos para a extração das informações necessárias para responder as questões de competência.

Para a criação desse novo conjunto, foram analisadas as justificativas dadas para a agregação da palavra-chave principal. A cada justificativa lida era criada uma palavra chave que, ao final da tarefa, foram reunidas, analisadas e combinadas para que se pudesse criar um novo conjunto de palavras referentes aos contextos dos documentos. A tabela 6 explicita o novo conjunto de palavras criado.

Tabela 6 – Tabela de palavras-chave referentes as variedades de contexto

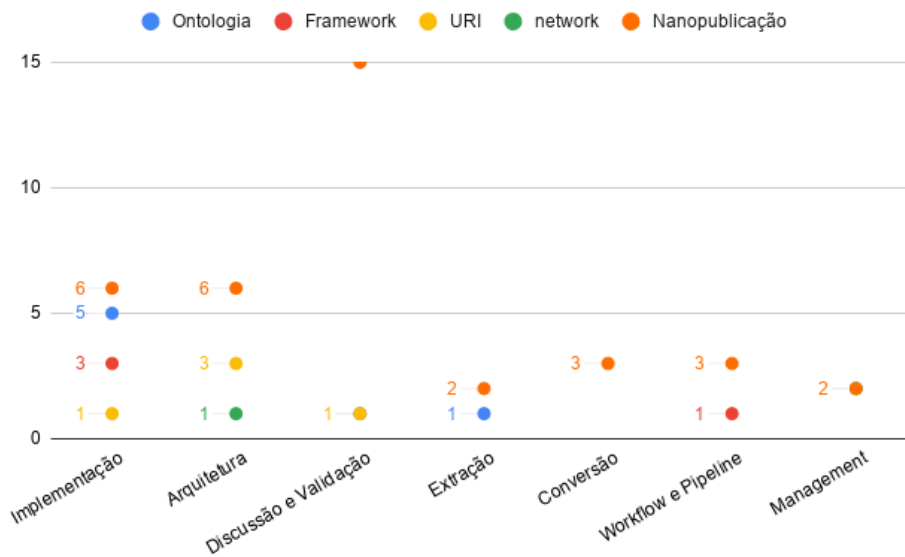
<b>Autor</b>	<b>Centralidade de Grau</b>
Implementação	Representa os artigos que tem como foco na criação de algo.
Arquitetura	Representa os artigos que focam somente em uma ou mais parte da arquitetura dananopublicação, sem trabalhar com a estrutura completa.
Discussão e Validação	Artigos que apenas discutem sobre o uso de uma abordagem, ou que relatam os benefícios, ou malefícios, obtidos através do uso da abordagem.
Extração	Artigos que focaram somente na extração de dados.
Conversão	Artigos que fizeram a conversão de uma abordagem para outra abordagem.
<i>Workflow e Pipeline</i>	Artigos que sugerem ou relatam o uso de um workflow ou pipeline para a criação de uma nano.
Gestão	Artigos que focam em gestão de dados, como o armazenamento e recuperação.

Fonte: Do Autor.

Após a definição do novo conjunto, foi feito um mapeamento que levou em consideração as palavras-chave principais e as de variedade de contexto. A figura 27 explicita um gráfico de dispersão com base no mapeamento realizado.

Como pode ser visto na figura 27, em ordem decrescente, 15 estudos referentes a palavra-chave “Nanopublicação” foram atribuídos ao contexto de “Discussão e Validação”, e 6 artigos com a mesma palavra-chave que foram atribuídos a arquitetura e implementação. Também foram atribuídos 5 artigos de ontologia para o contexto de implementação. O restante dos artigos foram atribuídos aos outros contexto em grupo que variam de 1 a 3 artigos.

Figura 27 – Gráfico de dispersão com dados referentes ao mapeamento entre as palavras-chaves principais e as palavras-chave de variedade de contexto



Fonte: Do Autor.

Após todos os artigos restantes terem passado pelo processo de classificação, decidiu-se escolher os artigos representados pelas palavras-chaves principais “Nanopublicação” e “Framework”, com contextos de implementação e “*Workflow e Pipeline*”. Os artigos também deveriam estar classificados como “proposta de solução” e “artigos de experiência”. Essa escolha se deu devido ao escopo e objetivo das questões de pesquisas criadas, cujo a estrutura destes artigos teriam uma maior capacidade de resposta. Sendo assim, a tabela 10 explicita os artigos selecionados.

Tabela 7 – Tabela com os trabalhos escolhidos para responderem as QPs

ID	Referência
1	McCusker et al. (2018)
2	Queralt-Rosinach et al. (2016)
3	González et al. (2014)
4	Maiatsky et al. (2018)
5	Schneider et al. (2015)
6	Sernadela e Oliveira (2017b)
7	Kuhn (2015)
8	Horst et al. (2015)
9	Lipani et al. (2014a)
10	Sernadela e Oliveira (2017a)
11	McCusker et al. (2014)
12	Lopes, Sernadela e Oliveira (2013)

Fonte: Do Autor.



O processo de extração de dados para se obter respostas para as QPs criadas, foi realizado através da leitura de todos os documentos onde, com o auxílio de um planilha<sup>2</sup> *online*, todas as respostas foram registradas e organizadas por colunas.

Ao analisar a coluna referente a **Q2**, observa-se que grande parte das fontes de informações mantém dados referentes ao domínio de ciências da vida, já que esta é a área de conhecimento de grande parte dos artigos. Como exemplo é possível citar a DrugBank, um banco de dados disponibilizado pela Universidade de Alberta, que armazena dados químicos e farmacológicos detalhados a respeito de medicamentos. Foca em registrar informações sobre substâncias ativas, como a seqüência, estrutura química e via metabólica. O banco de dados contém aproximadamente 4.300 registros de drogas. O drugBank proveu informações de asserção que foram utilizadas pelo trabalho 1. Como exemplo de uma fonte de dados fora do domínio de Ciências da Vida, o trabalho de número 9 faz uso de informações providas de documentos do NTCIR-10 Math Task, um desafio de recuperação matemática a partir de documentos PDF.

Diante a **Q3** é possível observar que grande parte dos trabalhos fazem a extração de dados, para a criação das nanopublicações, de forma automática. O trabalho 8 se destaca por, em vez de extrair as informações necessárias, realizar a conversão a conversão de dados RDF para já prontos para o formato de nanopublicação. Essa conversão foi realizada através de um *script* escrito em *python*, que foi disponibilizado pelos autores. Nota-se que o documento de número 11 realiza o mesmo processo, porém através da ferramenta SADI. Visto que grande parte das extrações são feitas de forma automatizada através do uso de PNL, o trabalho 9 realiza, através de uma *workflow* extrações manuais e automatizadas, para que essas extrações pudessem ser comparadas e avaliadas. No trabalho 4 a extração de dados foi realizada de forma semi-automática, onde após a extração automática dos dados os termos são revisados por um grupo de pesquisadores, afim de garantir a integridade dos dados.

Ao se referir a **Q4**, objetivou-se analisar principalmente as ontologias utilizadas para identificação das informações contidas no nível de proveniência das nanopublicações criadas. As ontologias usadas no nível de proveniência da nanopublicação, devem ser capazes de identificar a real proveniência das afirmações contidas no nível de asserção das nanopublicações. Sendo assim, notou-se que grande parte dos estudos fazem uso de ontologias com termos de proveniência específicos de um domínio. No entanto, os trabalhos 1 e 2 fizeram uso da ontologia PROV-O, como recomendado nas *guidelines* da nanopublicação. Uma ontologia capaz de descrever a proveniência de diferentes tipos. O trabalho 2 também faz uso da ontologia PAV, uma ontologia com base na PROV-O que permite descrever a autoria, curadoria e criação digital de recursos *online*. Duas ontologias que se mostraram interessantes e relevantes para a criação das nanopublicações propostas por este trabalho.

---

<sup>2</sup> <http://tiny.cc/cehggz>

Diante as ontologias utilizadas para a criação das asserções, os trabalhos, como esperados, fizeram uso de ontologias que permitissem a representação de informações específicas de cada domínio. As ontologias utilizadas nos trabalhos foram identificadas e registradas na tabela de resultados. Apesar de identificadas, não cabe a este trabalho discutir sobre tais ontologias, visto que não se mostra relevante para o objetivo do estudo.

Ao responder a **Q5** notou-se que pouco foi falado sobre o armazenamento das nanopublicações, sendo possível identificar somente a *triple store* Bigdata RDF, utilizada pelo trabalho 11. Diante a linguagem de consulta utilizada pelos trabalhos, todos descrevem o uso de consultas SPARQL. Sobre a sintaxe usada na criação das nanopublicações, as mais utilizadas foram as OWL/XML e RDF/XML.

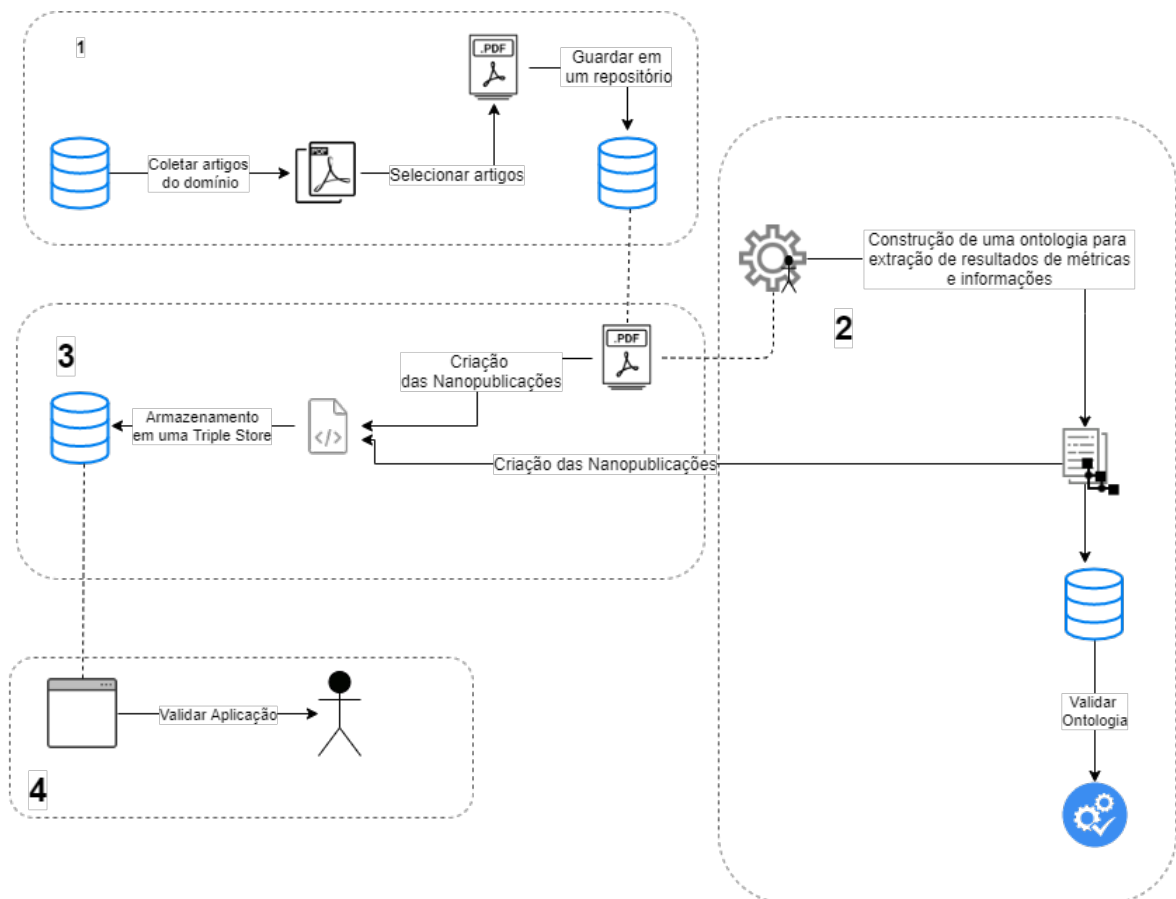
Através da **Q6** buscou-se descobrir quais metodologias de avaliação das nanopublicação estavam sendo utilizadas pelos pesquisadores. Através da busca por essa informação, foi observado que o trabalho 11 realizou a avaliação através de uma interface que permitia a recuperação das nanopublicações criadas, a aplicação e a viabilidade das nanopublicações foram validadas através de usuários e scripts. Já o trabalho 9 realizou uma análise, através de pesquisadores do domínio de I.R, nas informações contidas nas nanopublicações. Tendo como objetivo verificar se essas informações são suficientes para a reprodução de experimentos. O estudo 8 realiza cálculos de *Deletion disruption*, *Insertion disruption* e *Substitution rate* para validar as nanopublicações criadas. Através de um conjunto de consultas, o trabalho 5 tem como objetivo avaliar a recuperação das nanopublicações. Diante uma lista de questões de competência, o trabalho 2 realiza consultas em um servidor SPARQL, a fim de se obter nanopublicações que respondem as questões de forma correta. O trabalho 1 realiza a avaliação das nanopublicações criadas através do método Fisher's, uma metodologia estatística que combina resultados de diferentes testes independentes, com base numa mesma hipótese geral.

A partir da análise de todos os resultados obtidos para as QPs, foram tomadas decisões importantes para este trabalho. As metodologias adotadas nos trabalhos 11 e 2, inspiraram o método de avaliação das nanopublicações propostas por este trabalho.

#### 4 METODOLOGIA APLICADA

Este capítulo apresenta as atividades realizadas para que os objetivos deste trabalho pudessem ser concluído. A Figura 28 explicita o processo destas atividades, que estão divididas em 4 grupos.

Figura 28 – Processo das atividades realizadas para a concretização do trabalho



Fonte: Do Autor.

O primeiro grupo (Seção 4.1) é responsável pela coleta e seleção de Estudos de Análise de Redes de Colaboração Científica. Estes estudos foram coletados de um repositório científico e são essenciais para as atividades do grupo 2 e 3.

O segundo grupo (Seção 4.2) abrange a criação de uma ontologia capaz de extrair os resultados de métricas e informações importantes dos estudos selecionados, guiando as atividades de identificação/extração de afirmações e estruturação das mesmas em triplas RDF. A criação dessa ontologia foi realizada através de uma metodologia capaz de guiar todas as etapas necessárias para a construção de uma ontologia consistente.

O terceiro grupo consiste em realizar a criação de nanopublicações do domínio de Estudos de Análise de Redes de Colaboração Científica, para isso foi necessário fazer uso da ontologia construída anteriormente e de uma ontologia capaz de guiar a extração de informações referentes a proveniência das afirmações extraídas.

O quarto e último grupo consiste em um conjunto de atividades voltadas para a criação de uma aplicação semântica de recuperação de nanopublicações. Através desta aplicação, implementada com tecnologias web, como: php, css e javascript, e com o auxílio da linguagem de consulta SPARQL, é possível consultar as nanopublicações contidas em um banco de triplas. Após realizar testes, e corrigir erros encontrados, a aplicação passou por um processo de avaliação, realizada por pesquisadores de Estudos de Análise de Rede de Colaboração Científica.

As seções a seguir detalham as atividades do processo metodológico adotado para este trabalho.

#### 4.1 Coleta de Artigos

A atividade tem como objetivo realizar a coleta de publicações científicas do domínio definido. Primeiramente foi realizada buscas no repositório de publicações científicas Google Acadêmico. A busca foi realizada através da seguinte string de pesquisa:

```
network analysis ("scientific collaboration"OR "academic collaboration"OR "co-author"OR "research network")
```

Após a recuperação dos documentos, foram coletados os artigos presentes nas primeiras 5 páginas da busca. Logo após, foi necessário executar um protocolo de inclusão e exclusão para coletar apenas as publicações que poderiam fornecer as afirmações para a criação da ontologia e das nanopublicações à área de análise de redes de colaboração científica. Para a realização desse protocolo foram definidos os seguintes critérios:

- Foram considerados válidos apenas artigos que abordem o estudo de uma rede, através de um conjunto mínimo de métricas de análise de redes, tais como: número de nodos, número de relações, centralidade de grau, densidade.
- Foram válidos apenas as publicações que estivessem disponíveis em formato PDF.
- Coletou-se somente artigos que foram publicados no período de 2017-2018.

Foram coletados um total de 51 artigos, e ao aplicar os critérios sobraram 20 publicações. Estes serão usados nas próximas atividades.

#### 4.2 Criação da Ontologia SCNAS *Ontology*

A criação de uma ontologia foi necessária para servir como guia durante a extração das triplas de afirmações que estarão contidas no nível de “asserção” da arquitetura do modelo de nanopublicação adotada por este trabalho. A ontologia deve ser capaz de descrever os conceitos relacionados aos resultados de métricas e informações contidas em

estudos de análise de rede, domínio que deu origem ao nome da ontologia: *Scientific Collaboration Network Analysis Ontology* (SCNAS Ontology).

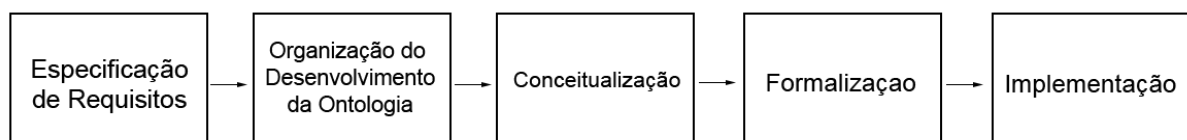
Com isso, precisou-se definir uma metodologia capaz de guiar todos os passos de desenvolvimento desta ontologia. Após analisar-se algumas metodologias, optou-se por usar a metodologia NeOn pois, além de atender os requisitos para a construção de uma ontologia consistente (citados na seção 2.1.3.1), a metodologia:

- Trabalha com diferentes cenários de criação de ontologias, explicitando, principalmente, um conjunto de práticas para a colaboração entre ontologias, como a reutilização de ontologias, por exemplo.
- As atividades propostas contém descrições bem definidas, facilitando o processos da Engenharia de Ontologia.
- É um método bastante utilizada entre engenheiros e pesquisadores de Ontologias.

Ao analisar as situações de construção de uma ontologia propostas pela metodologia (descritas na Seção 2.1.3.1), concluiu-se que a ontologia construída se encaixava na situação B (construção de conjuntos de ontologias simples e interconectadas), visto haviam pretensões de fazer reuso de ontologias secundárias.

Ao tomar conhecimento das fases básicas de construção de uma nova ontologia (representado pela figura 29), a primeira fase a ser executada é a de especificação de requisitos.

Figura 29 – Cenário 1: Tarefas Básicas de criação de uma ontologia

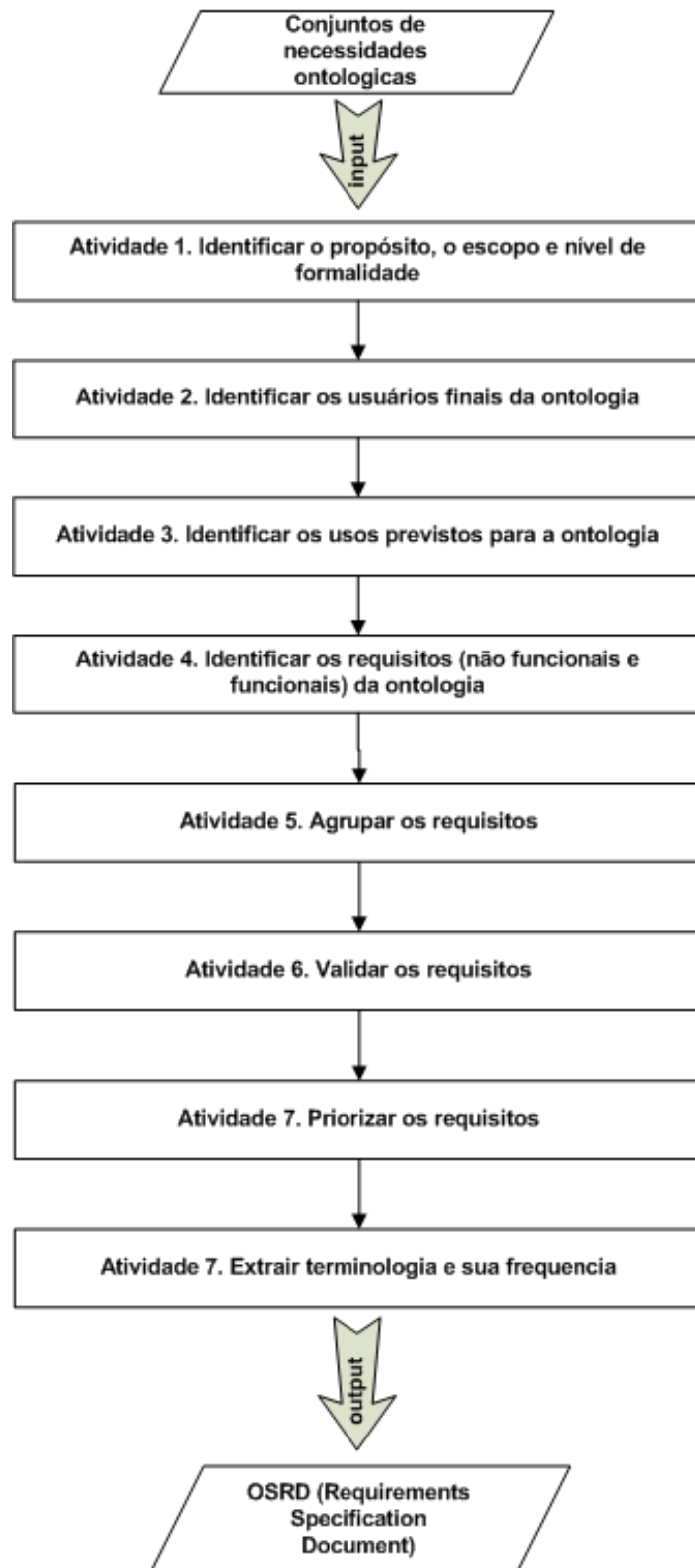


Fonte: adaptado de Suarez (2010)

#### 4.2.1 Especificação dos requisitos da ontologia

De acordo com Suarez (2010) a construção de uma ontologia passa sempre pelas atividades do cenário 1. A Figura 30 representa de forma gráfica o fluxo de atividades de especificação de requisitos da ontologia seguido de acordo com a metodologia adotada (Seção 2.1.3.1). Nas sessões seguintes as atividades serão melhor especificadas.

Figura 30 – Tarefas da atividade de especificação dos requisitos da ontologia.



Fonte: adaptado de Suarez (2010) por Bortolato e Paulo (2014)

#### 4.2.1.1 Propósito, escopo e formalidade

A criação da ontologia EARCC tem como propósito prover um modelo semântico sobre o domínio de Estudos de Análise de Redes de Colaboração Científica, auxiliando e facilitando a extração de dados de artigos científicos, os quais são usados para a criação de nanopublicações aplicadas ao domínio de EARCC.

A ontologia não deve se limitar somente a criação das NPs, podendo ser usada em outras situações, assim como também ser re-usada e/ou estendida. Visto que o domínio da ontologia são estudos de análise de Redes de Colaboração Científica, compreende-se que a mesma deve descrever conceitos relacionados a métricas de Análise de Redes de Colaboração Científica, tais como: métricas referente a rede, métricas referentes aos atores da rede e informações importantes da rede.

Em relação a formalidade da ontologia, sua implementação será realizada na linguagem OWL.

#### 4.2.1.2 Identificação dos Usuários

Tendo em vista o domínio determinado para o uso da ontologia, seus principais usuários devem ser os pesquisadores de Análise de Redes de Colaboração Científica. Não se prendendo aos seus principais usuários, a ontologia pode ser aproveitada por qualquer entidade que tenha o conhecimento mínimo e necessário para desfrutar das informações provindas pela ontologia.

#### 4.2.1.3 Cenários de Uso

Ao realizar uma análise dos cenários e do propósito da ontologia, foi possível identificar os seguintes usos da ontologia:

- Apoiar a extração de informações em estudos relacionados ao domínio de análise de redes de colaboração científica;
- Reutilizar a ontologia em projetos que trabalhem com o domínio de análise de redes de colaboração científica;
- Desenvolver aplicações semânticas que estejam dentro do domínio de Estudos de Análise de Redes de Colaboração Científica.

#### 4.2.1.4 Identificar os Requisitos

Seguindo orientações da metodologia, os requisitos serão divididos entre funcionais e não-funcionais. Referente aos requisitos funcionais, optou-se por fazer uso de uma lista de Questões de Competências (QCs), como aborda a metodologia proposta por (GRUNINGER; FOX, 1995). A lista tem como objetivo identificar e capturar os componentes necessários para a criação de uma ontologia que descreva de forma eficiente o domínio a ser trabalhado.

Para a criação das questões, foram utilizados os estudos sobre análise de rede de colaboração científica coletados na primeira etapa da metodologia adotada para este trabalho (seção 4.1). Ao total foram criadas 56 questões que, juntamente com suas respostas, foram organizadas em tabelas através da plataforma online Google Sheets. Devido ao volume de questões elaboradas, optou-se por explicitar nesta parte do trabalho, na tabela 8, um excerto da tabela de QC's. A tabela completa das questões de competência criadas nesta etapa pode ser visualizada de forma online<sup>1</sup>.

---

<sup>1</sup> <<https://bit.ly/2G0bjnN>>



Tabela 8 – Excerto da Primeira Lista de Questões de Competência

Questão	Resposta
Quais os tipos de EARCC encontrados?	Programa de Pós Graduação, Instituição de Ensino, Área de Conhecimento, Periódicos, Evento
Qual é o Nome, Densidade, Grau Médio, Número de Atores e Relações de uma determinada Instituição?	Entrada: Instituição: Universidade de Coimbra — Nome: Redes de I D da Universidade de Coimbra: análise dos projetos de IC DT financiados pela Fundação para a Ciência e Tecnologia (FCT) Densidade: 0,015 Grau Médio: 8,74 N. Atores: 553 N. Relações: 2419
Quais os nomes dos EARCC de uma determinada Área de Conhecimento e com densidade maior que X?	Entrada: Área de Conhecimento: Sustentabilidade Estratégica x = 0,005 — Saída: EARCC: A PESQUISA SOBRE SUSTENTABILIDADE ESTRATÉGICA NO BRASIL: FATOS E DADOS
Qual a localização geográfica das redes de determinado PPG?	Entrada: Programa de Pós-Graduação em Administração — Nome da Rede: Redes de Colaboração Científica: Um e studo de coutoria através da análise de redes sociais Localização: Brasil
Quais os nomes dos periódicos, nomes das redes e os periodos temporais das redes do tipo Periódico?	Revista Contabilidade, Gestão e Governança — Nome da Rede: Autores Centrais no Desenvolvimento da Rede Colaborativa: Análise da Revista Contabilidade, Gestão e Governança de 1998 a 2012. Início: 1998 Fim: 2012

Em relação aos requisitos não funcionais se definiu que a ontologia deve ser descrita em Inglês e sua linguagem de implementação deve ser OWL.

#### 4.2.1.5 Agrupamento dos requisitos

Ao analisar as questões de competência criadas, a fim de se obter um melhor entendimento do domínio, e assim poder representá-lo em formato de ontologia, decidiu-se agrupar os questionamentos elaborados em categorias:

- GQC1<sup>2</sup> -Grupo de questões de competência referentes as métricas de atores e rede
- GQC2<sup>3</sup>- Grupo de questões de competência referentes a localização geográfica

<sup>2</sup> GQC1: <<http://tiny.cc/c7bedz>>

<sup>3</sup> GQC2: <<http://tiny.cc/t8bedz>>

- GQC3<sup>4</sup>- Grupo de questões de competência referentes ao período temporal dos dados estudados

#### 4.2.1.6 Validação do conjunto de requisitos

O objetivo desta tarefa foi identificar possíveis conflitos ou contradições entre as QC's. Dentre os critérios sugeridos pela metodologia NeOn, os seguintes critérios foram utilizados neste projeto:

- **Correção** - Garante que os requisitos estão relacionados ao domínio da ontologia a ser criada.
- **Consistência** - A fim de verificar se não havia conflitos entre os requisitos levantados.
- **Compreensível** - A verificação de cada requisito, garantindo que sejam compreensível para os usuários e os especialistas do domínio.
- **Sem ambiguidades** - A realização da análise de cada requisito, assegurando se possuía apenas um significado, a fim de que eles não possam ser mal entendidos.
- **Concisão** - Analisou-se a relevância de cada requisito e se não havia duplicatas.

Essa validação foi realizada manualmente, mediante uma especialista do domínio, que analisou todas as QC's propostas. Durante 3 meses, a avaliadora realizou correções nas questões elaboradas, juntamente com o autor das mesmas. Através do *feedback* obtido da avaliadora, foram criadas novas questões de competência que passaram a abranger algumas informações importantes para um pesquisador do domínio, estas que foram passadas despercebidas em um primeiro momento pelo autor.

O processo de validação (análise, avaliação e refatoração) das QC's foi realizado manualmente mediante os critérios citados, resultando em 52 QC's. Esta tarefa dará a luz aos termos a serem descritos na futura ontologia, deixando claro que a discussão sobre as questões de competência são de extrema importância. Assim como a primeira versão da tabela de competência, o produto desta atividade pode ser visualizado de forma online<sup>5</sup>.

#### 4.2.1.7 Priorizar os requisitos

Essa atividade tem como objetivo definir prioridades aos requisitos coletados, sejam funcionais ou não funcionais. Esta atividade é recomendada pela metodologia, porém é dada como opcional, visto que é um facilitador para tomadas de decisões quando existem variados grupos de CQs, assim como muitos requisitos.

---

<sup>4</sup> GQC3: <<http://tiny.cc/kbcedz>>

<sup>5</sup> <<https://bit.ly/2kNPhNE>>

Neste trabalho a priorização dos requisitos não foi efetuada, visto que a quantidade de CQs, termos e requisitos coletados é consideravelmente pequena, não necessitando de priorização das mesmas.

#### 4.2.1.8 Extrair a terminologia e suas frequências

A partir da extração de terminologias é possível criar um pré-glossário de termos mediante suas frequências na lista de CQs e de suas respostas. Esta tarefa pode ser realizada com o uso de técnicas de extração e ferramentas que auxiliam a atividade (SUAREZ, 2010). O pré-glossário de termos é dividido em:

- termos das QCs - Através dos requisitos na forma de CQs, são extraídos terminologias (nomes, adjetivos e verbos) que serão formalmente representados na ontologia por meio de conceitos, atributos, relações ou instâncias.
- termos das respostas da QC - A partir das respostas adquiridas pelas CQs, são extraídos terminologias que podem ser representadas na ontologia como conceitos ou como instâncias.
- termos identificados como entidades nomeadas - A partir das QCS e de suas respostas, são extraídos entidades que futuramente se tornarão indivíduos.

Neste trabalho a extração de terminologias foi realizada através da aplicação da ferramenta AntConc<sup>6</sup> versão 3.5.8 (figura 31), desenvolvida por Laurence Anthony. A escolha foi realizada devido a gratuidade da ferramenta e por ser uma ferramenta muito utilizada no domínio de Linguística de *Corpus* (LC). Através da ferramenta foram geradas listas de palavras-chave, juntamente com as suas frequências de aparições, que foram analisadas por especialistas do domínio de EARCC a fim de se obter somente os termos que irão compor a versão final do ORSD.

Primeiramente objetivou-se encontrar as palavras-chave de todas as questões de competência, resultando em 64 palavras-chave<sup>7</sup>. Logo após foram excluídos todos os termos referentes a artigos definidos e indefinidos, assim como os que não iriam contribuir para a criação de um conceito, atributo, relação ou instância.

Ao analisar os termos restantes, percebeu-se que muitos eram sinônimos ou plural de termos já listados, por isso suas frequências foram somadas. Também foi analisado que alguns termos dependiam de outros para fazer sentido, por exemplo: Programa de Pós-Graduação, cujos termos listados separadamente não fazem sentido, por isso criou-se apenas uma palavra chave para representá-lo, logo após sua contagem foi realizada manualmente, já que a quantidade de questões criadas é relativamente pequena.

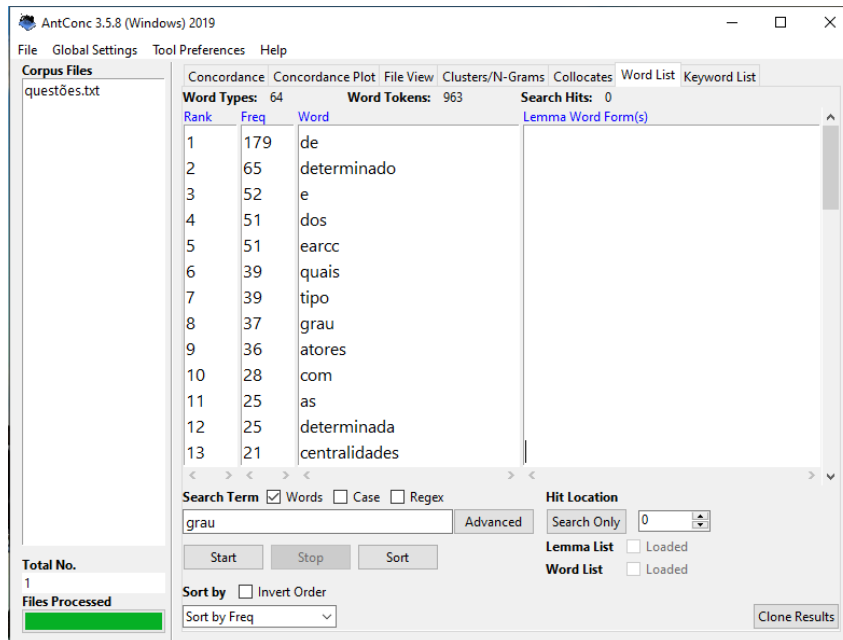
<sup>6</sup> <<https://www.laurenceanthony.net/software/antconc/>>

<sup>7</sup> <<http://tiny.cc/gysedz>>

Posteriormente, devido a estrutura das respostas, foi realizada a extração de forma manual dos termos das respostas das QCs<sup>8</sup> e dos termos identificados como nomeados<sup>9</sup>.

Após as extrações, os termos foram reunidos no documento pré-glossário<sup>10</sup>.

Figura 31 – Tela de palavras-chave e suas frequências



Fonte: Do Autor.

<sup>8</sup> <<http://tiny.cc/ugpodz>>

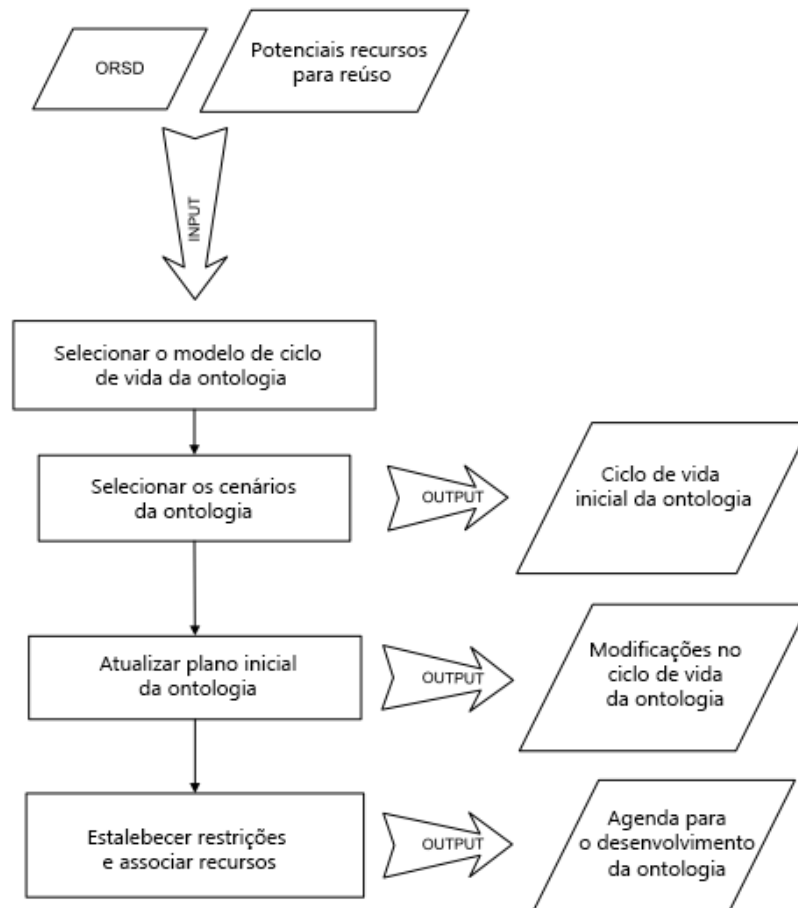
<sup>9</sup> <<http://tiny.cc/vjpodz>>

<sup>10</sup> <<http://tiny.cc/w7oodz>>

### 4.2.2 Organização do desenvolvimento da Ontologia

A metodologia NeOn propõe um fluxo de atividades necessárias para o planejamento de um projeto de desenvolvimento de uma ontologia conforme mostra a figura 32. As próximas sub-seções descrevem as atividades.

Figura 32 – Atividades do planejamento do projeto.



Fonte: Adaptado de Suarez (2010)

#### 4.2.2.1 Selecionar o modelo de ciclo de vida da ontologia

O apoio a escolha do modelo a ser seguido se dá através de uma árvore de decisão, como exibido na figura 33. Com base na resposta dada à questão, um dos modelos é escolhido.

Figura 33 – Árvore de Decisão de modelo de ciclo de vida da Ontologia.



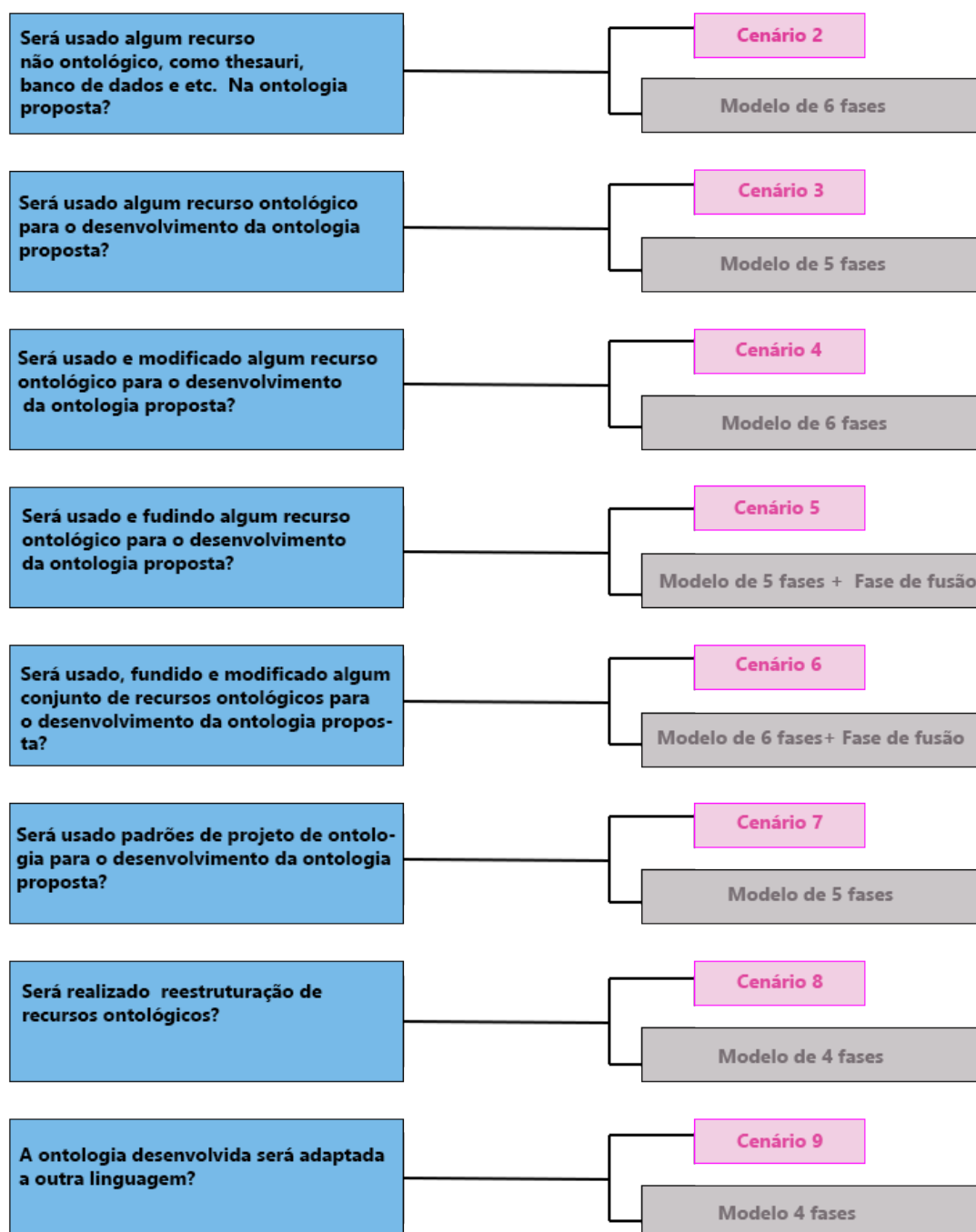
Fonte: Adaptado de Suarez (2010).

Neste trabalho o modelo cascata é o que melhor se adéqua a ontologia proposta, pois trata-se de uma ontologia com conhecimento pouco mutável, já que o conjunto de métricas aplicáveis é finito.

#### 4.2.2.2 Selecionar os cenários da ontologia

Nesta atividade utilizou-se uma árvore de decisão (Figura 34) proposta pela metodologia, cuja resposta determina a decisão do cenário e das fases do modelo de ciclo de vida cascata a ser seguido no projeto.

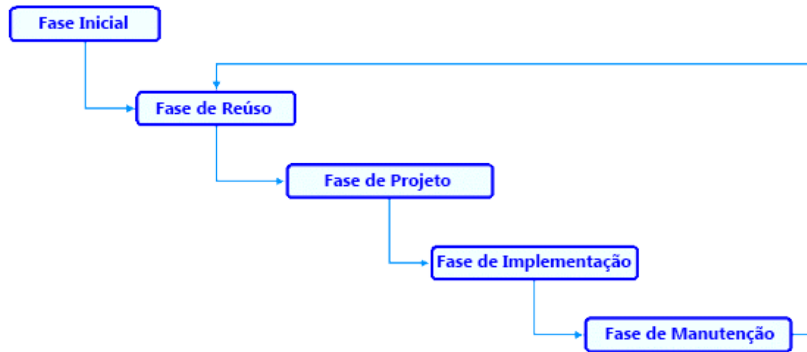
Figura 34 – Árvore de decisão.



Fonte: Suarez (2010).

Visto que foram encontradas ontologias com potencial ao reúso, o cenário eleito para a ontologia proposta foi o de número 3, estabelecendo-se o modelo de 5 fases como o ciclo de vida da ontologia EARCC. O modelo é representado pela figura 35.

Figura 35 – Modelo de ciclo de vida cascata de 5 fases.



Fonte: Adaptado de Suarez (2010).

#### 4.2.2.3 Atualizar o Plano Inicial

Diante o ciclo de vida adotado para a ontologia desenvolvida, decidiu-se seguir todas as atividades propostas para o reúso de ontologias já existentes. Porém, diante a fase de projeto, decidiu-se executar as tarefas de descrição de constantes em detalhes, descrição de axiomas formais e descrição de regras, somente nas constantes, axiomas e regras criadas exclusivamente por este trabalho. Sendo assim, as atividades citadas não foram aplicadas nas ontologias reutilizadas. Ressalta-se também que não foi necessário realizar as atividades de upgrade, tradução, modificação, update, Versioning, manutenção e sumarização, visto que as ontologias estavam de acordo para a reutilização.

#### 4.2.2.4 Estabelecer restrições e associar recursos

Após a definição das atividades e fases a serem executadas durante o ciclo de vida da ontologia, as mesmas foram organizadas e sequenciadas em um cronograma para que se pudesse planejar a execução de cada fase e atividade. Para isso, considerou-se:

- As fases e atividades dependentes de outras fases e atividades;
- As fases e atividades que poderiam ser executados em paralelo;
- As atividades excluídas do plano de desenvolvimento.

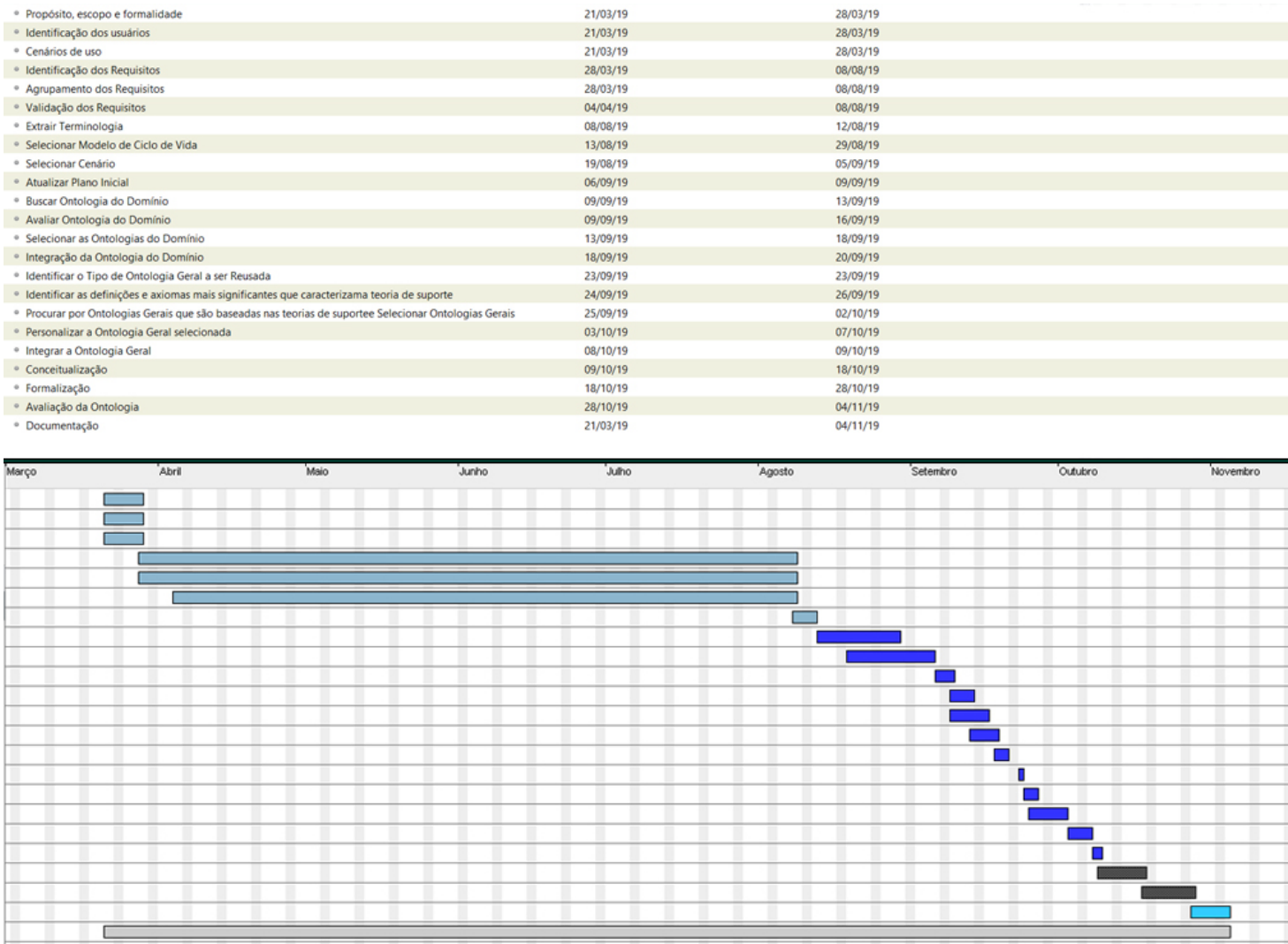
Diante a criação do cronograma de desenvolvimento da ontologia, utilizou-se a ferramenta GanttProject<sup>11</sup>, um software livre usado para gerenciar o cronograma de um

<sup>11</sup> <<https://www.ganttproject.biz/>>



projeto, ilustrando de forma gráfica os períodos de tempos e dependências determinadas para cada atividade, facilitando o entendimento e planejamento. A figura 36 explicita o cronograma determinado para a criação da ontologia proposta por este trabalho.

Figura 36 – Cronograma do projeto



Fonte: Do Autor.

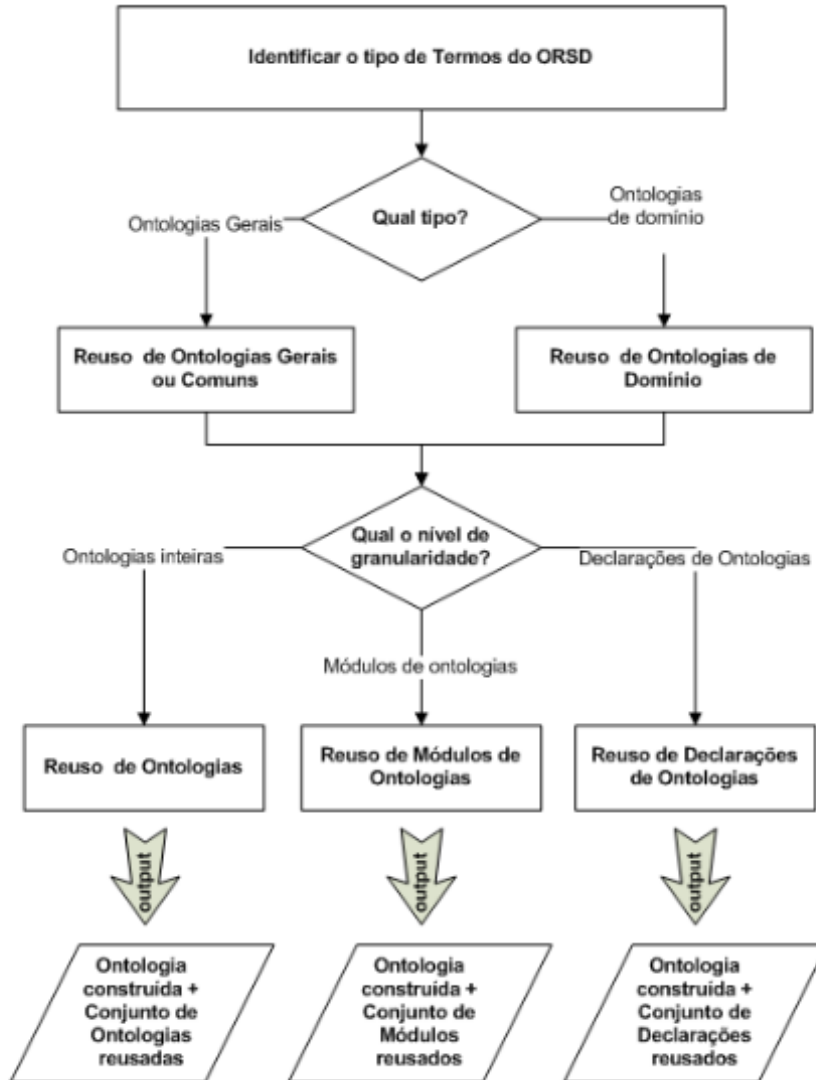
### 4.2.3 Modelo Cascata - Fase Reuso

Um dos princípios de fazer reuso de recursos ontológicos é a redução de tempo e custos associados a criação da ontologia, visto que existe uma grande quantidade de ontologias, bibliotecas de ontologias e repositórios que dispõe recursos.

Na seção 4.2.2.2 foi identificado o cenário de número 3, caracterizado pelo reuso de recursos ontológicos. Neste cenário, é realizada uma análise nos recursos ontológicos já existentes que poderão ser selecionados, ou não, para a criação da ontologia desejada. Quando se trata de reuso de recursos, Suarez (2010) destaca a importância do nível de

granularidade. A figura 37 apoia a descoberta do tipo de reuso necessário para a criação da ontologia.

Figura 37 – Escolha do tipo de reuso de recursos ontológicos.



Fonte: Adaptado de Suarez (2010).

Ao observar e analisar os termos extraídos das questões de competência, foi possível identificar termos referentes a métricas comumente usadas em Estudos de Análise de Redes de Colaboração Científica, por exemplo: Centralidade de Grau, Grau Médio, Componentes, Centralidade de intermediação, Centralidade de Proximidade, etc. Também foram encontradas informações como: localização geográfica da rede e intervalos de tempo. Verificou-se que tais termos são encontradas em ontologias de domínio e ontologias gerais já existentes. Assim, optou-se por executar os processos de reuso de ontologias do domínio e de ontologias gerais ou comuns.

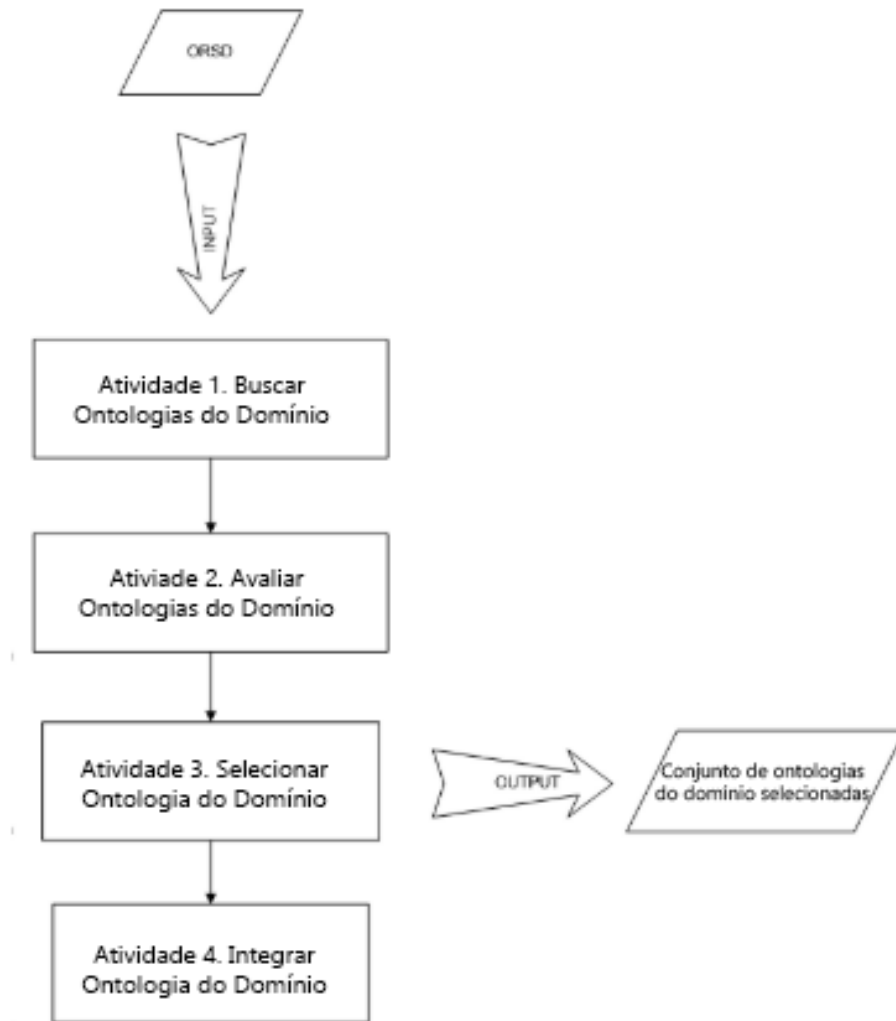
Referente a ontologias do domínio, decidiu-se adotar o processo de atividades para o reuso de ontologias inteiras, visando selecionar a ontologia a ser reusada. Esta decisão foi tomada devido a quantidade de ontologias encontradas que descrevem o domínio de Análise de Redes Sociais.

Diante do reuso de ontologias gerais ou comum, optou-se por fazer reuso de declarações de ontologias, pois trata-se de termos específicos contidos em ontologias com descrições desnecessárias para a ontologias que está sendo desenvolvida.

#### **4.2.3.1 Reuso de Recursos ontológicos inteiros d domínio**

O processo de reuso de recursos ontológicos inteiros do domínio tem como objetivo buscar, avaliar e selecionar os recursos ontológicos que serão reutilizados na ontologia que está sendo desenvolvida. Para que um recurso possa ser reusado por inteiro, o mesmo deve atender de forma integral todas as necessidades da ontologia em desenvolvimento. A figura 38 representa o processo de atividades necessárias para o reuso de ontologias inteiras.

Figura 38 – Processo de reuso de módulos de ontologias.



Fonte: Adaptado de Suarez (2010).

### Atividade 1 - Buscar Ontologias do Domínio:

Esta atividade tem como objetivo encontrar ontologias que contenham o conhecimento necessário para suprir as necessidades estabelecidas pelos termos extraídos da GQC1<sup>12</sup>. A busca por ontologias foi realizada nos repositórios ontológicos apresentados por d’Aquin e Noy (2012), e complementadas pelos seguintes repositórios: swoogle e AberOWL. Também foram utilizados repositórios científicos e acadêmicos para a busca de artigos com proposições de ontologias com métricas de Análise de Redes Sociais. Após a busca, foram selecionados os recursos mais relevantes à reutilização: SemSNA, proposto por Ereteo et al. (2009), SNAMetric, proposto por Bordin et al. (2015) e ontoSELF-v2, proposto por Yu (2008)).

### Atividade 2 - Avaliar Ontologias de Domínio:

<sup>12</sup> <<http://tiny.cc/c7bedz>>

Esta atividade tem como objetivo decidir se as ontologias selecionadas serão realmente úteis para a ontologia que está sendo desenvolvida. Assim como sugere a metodologia NeOn, para realizar a análise e avaliação foram usados os seguintes critérios:

- Escopo - Verificar se o escopo e a finalidade estabelecidos no ORSD se parecem aos da ontologia candidata;
- Requisitos Não-Funcionais - Verificar se os requisitos não-funcionais estabelecidos no ORSD, são atendidos pela ontologia candidata;
- Requisitos funcionais - Verificar se os requisitos funcionais estabelecidos no ORSD, são atendidos pela ontologia candidata de forma total, parcial ou imparcial. Para esta avaliação, foi analisado se todos os termos extraídos da ORSD (GQC1) estavam contidos na ontologia candidata.
- Questões de Competência - Verificar se a ontologia candidata tem a capacidade de responder as questões de competência contidas no ORSD, esta verificação foi realizada no GQC1.

A primeira ontologia a ser analisada foi a SNAMetric de Bordin et al. (2015), onde se observou a definição de grupos de métricas de Análise de Redes. Como pode ser visto na figura 39, a classe principal SNAMetric é usada como uma superclasse, contendo grupos de métricas que são separadas por conceitos relacionados ao nível de análise, ou seja, métricas do nível de análise do ator na rede (Actor), métricas do nível de análise da rede como um todo (Network) e métricas que analisam grupos formados por laços mais frequentes (Group). Através dos grupos de métricas que compõem a ontologia, a mesma conseguiu responder todas as questões do GQC3<sup>13</sup>.

Em relação aos requisitos não-funcionais, observou-se que a ontologia foi implementada em OWL, compatível portanto, com a linguagem da ontologia proposta. Ao analisar a hierarquia da ontologia proposta por Bordin et al. (2015), é possível observar que os conceitos são descritos com clareza e refletem seus verdadeiros significados na ontologia. Sendo assim, concluiu-se que a ontologia candidata cumpre com todos os critérios necessários para o reuso.

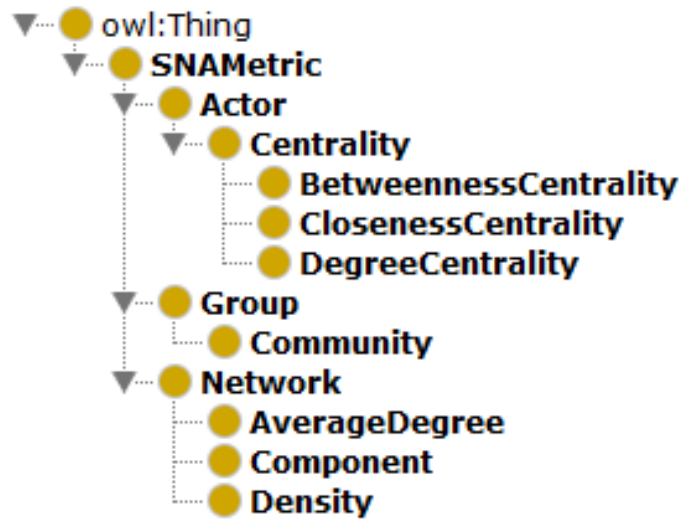
Ao analisar a ontologia ontoSELF-v2, proposta por Yu (2008), é possível observar que a mesma contém uma grande quantidade de métricas de Análise de Redes Sociais, porém, apesar das métricas representadas, em seu escopo não foi encontrado alguns termos do ORSD, impossibilitando a busca de respostas de algumas questões de competência. A figura 40 representa o diagrama de classes da OntoSELF.

Na ontologia SemSNA, proposta por Ereteo et al. (2009), é possível observar uma representação dos conceitos de Análise de Redes Sociais, cujo seu principal propósito é

---

<sup>13</sup> <<http://tiny.cc/kbcedz>>

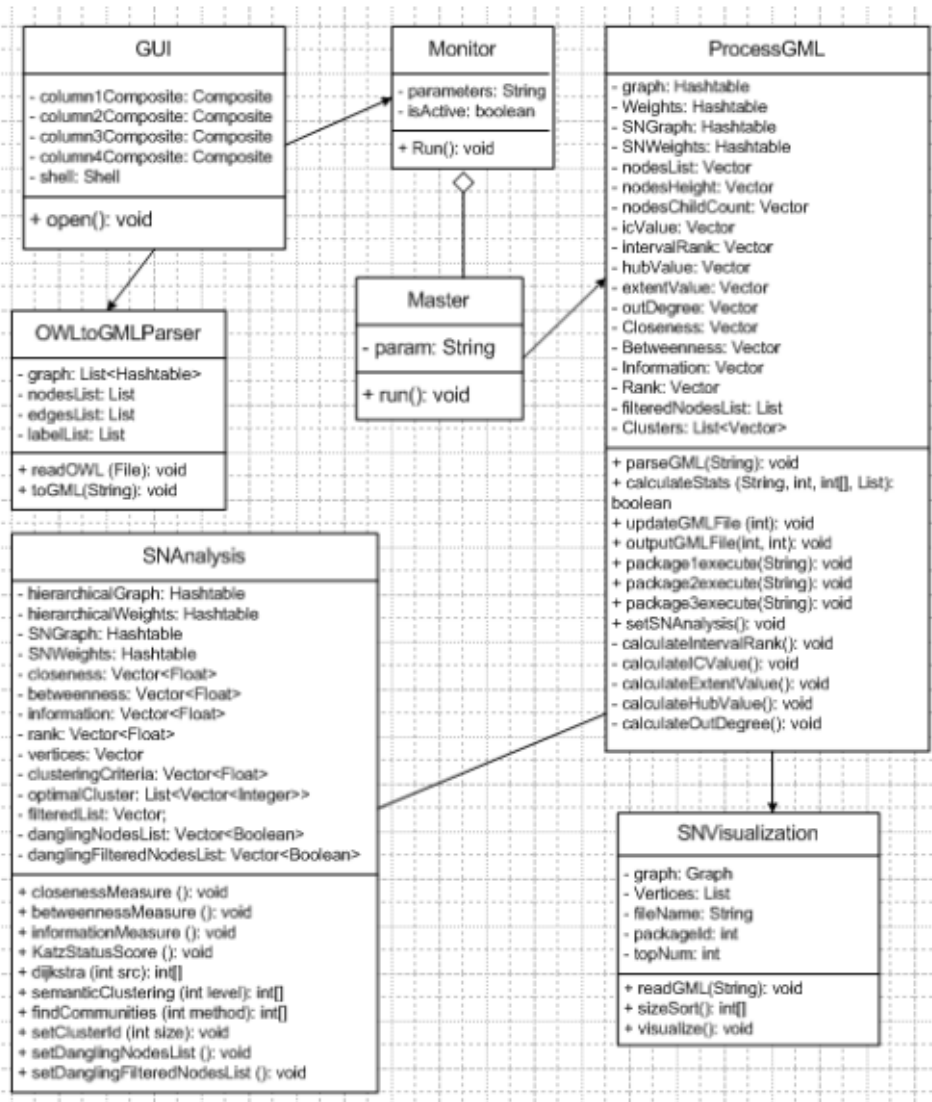
Figura 39 – Arquitetura da ontologia SNA Metric.



Fonte: Bordin et al. (2015)

calcular as principais métricas do domínio. Através de seus termos, a ontologia teve a total capacidade de responder as questões relacionadas aos Atores da rede, porém, não foi capaz de responder as questões relacionadas a rede como um todo. Sua implementação foi realizada em OWL. O esquema da ontologia SemSNA é representado na figura 41.

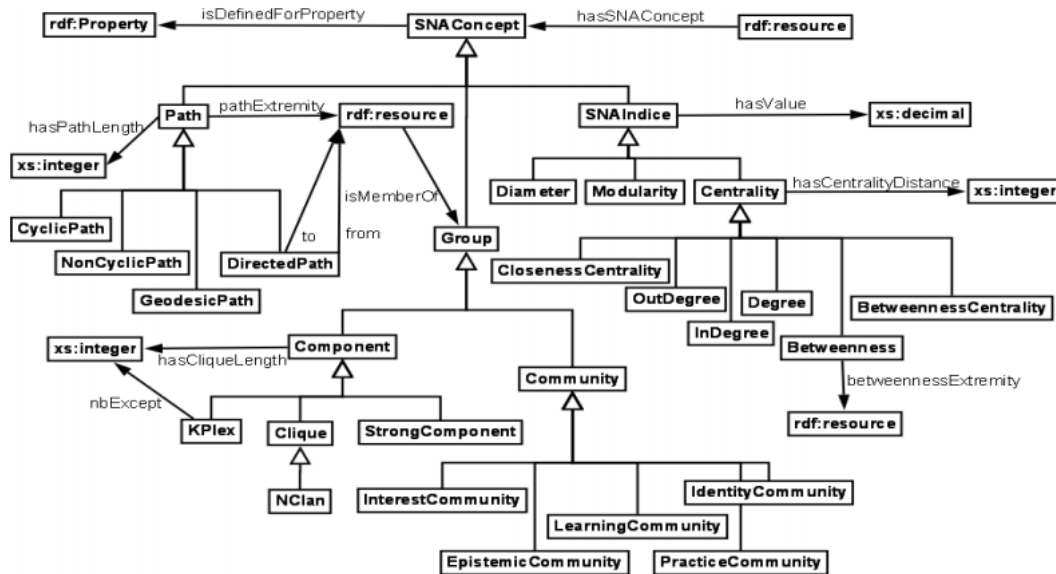
Figura 40 – Arquitetura da ontologia OntoSELF, representada em diagrama de classes UML.



Fonte: Yu (2008).

Depois de realizar a análise das ontologias candidatas, levando em consideração os critérios descritos acima, a Tabela 9 representa a aplicação da avaliação, explicitando o resultado da aplicação de cada critério de reuso das ontologias.

Figura 41 – Arquitetura da ontologia SemSNA.



Fonte: Ereteo et al. (2009).

Tabela 9 – Tabela de aplicação dos Critérios de Reúso

Critério	Avaliação	SNAMetric	SemSNA	ontoSELF-v2
Escopo	[Sim, Não, Desconhecido]	Sim	Sim	Sim
Requisitos Não-Funcionais	[Sim, Não, Desconhecido]	Sim	Não	Sim
Requisitos Funcionais	[Totalmente, Parcialmente, Imparcial]	Totalmente	Parcialmente	Parcialmente
Questões de Competência	[Totalmente, Parcialmente, Imparcial]	Totalmente	Parcialmente	Parcialmente

Fonte: Do Autor.

### Atividade 3 - Selecionar as ontologias de domínio

Esta seção tem como objetivo selecionar a ontologia de domínio a ser reusada na ontologia em desenvolvimento. Para realizar esta seleção, utilizou-se da tabela 9, que explicita os resultados da aplicação dos critérios de reúso sobre as ontologias candidatas a reúso. As ontologias que continham como resposta "não", "desconhecido" ou "imparcial", foram excluídas, pois não se mostraram adequadas às necessidades deste trabalho. As ontologias que obtiveram, em algum quadro, como resposta "Parcialmente", foram analisadas novamente, com o objetivo de dar finalidade a mesma na ontologia em construção. As ontologias que obtiveram como resposta "Sim" e "Totalmente", foram julgadas como úteis a ontologia em desenvolvimento. Com isso, foi concluído que, neste trabalho, o reúso da ontologia proposta por Bordin et al. (2015) pode ser aplicado de forma completa, visto



que os grupos de métricas contidos na ontologia suprem o conhecimento necessitado.

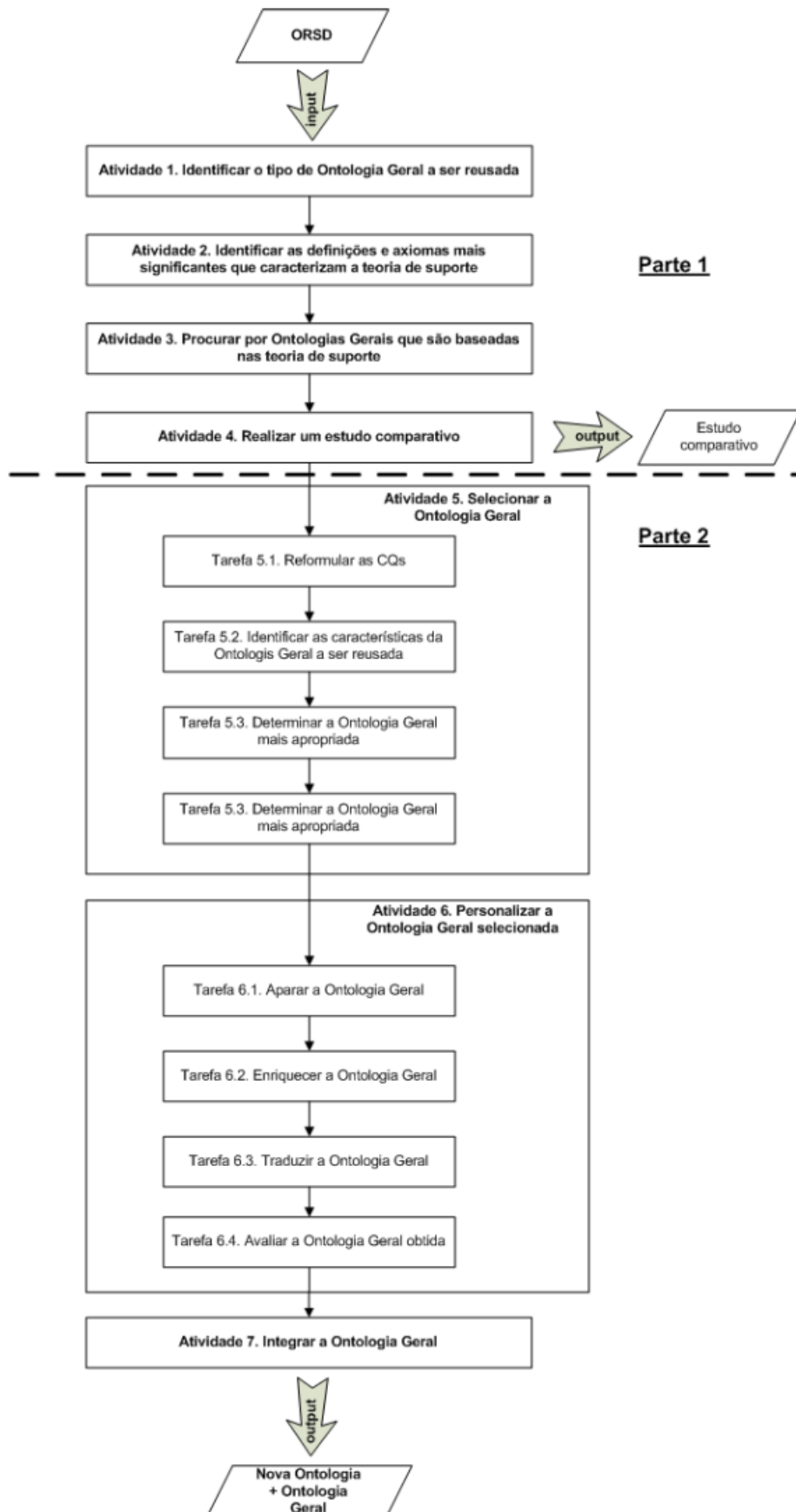
#### **Atividade 4 -Integração da Ontologia de Domínio**

A atividade de integração da ontologia de domínio é detalhada na fase de projeto, na seção integração [4.2.6].

#### **4.2.3.2 Reúso de Ontologias Gerais ou Comuns**

O objetivo dessa fase é encontrar e selecionar as ontologias gerais ou comuns - que vão além de domínios particulares - necessárias para a criação da ontologia proposta. A figura 42 representa o processo deste tipo de reúso e suas atividades são detalhadas a seguir.

Figura 42 – Processo de reuso de ontologias gerais.



### Atividade 1 - Identificar o tipo de Ontologia Geral a ser reusada

Esta atividade tem como objetivo determinar os tipos de ontologias gerais que devem ser reutilizadas. Para essa escolha, foi aplicado heurísticas baseadas nas questões de competências aplicadas no início do desenvolvimento da ontologia proposta. A tabela 10 explicita as heurísticas utilizadas.

Tabela 10 – Heurísticas usadas

Condição nas CQs da ontologia	Casos típicos	Tipo de Ontologia Geral
Referencia a Tempo	Intervalos temporais aparecem. Exemplo: Entre 2010 e 2018	Tempo
Referencia a Tempo	Referências a um ponto fixo na linha do tempo. Ex.: 2009	Tempo
Referencia a Localização	Referências a localização da rede aparecem. Ex.: Brasi	GeoSpatial

Fonte: Do Autor.

### Atividade 2 - Identificar as definições e axiomas mais significantes que caracterizam a teoria de suporte

Esta atividade tem como objetivo identificar as definições e axiomas que caracterizam a teoria de suporte das ontologias gerais candidatas a reuso. Para elaborar este levantamento, seguiu-se as teorias de (SUAREZ, 2010), considerando as seguintes definições de tempo para este trabalho:

- Pontos no tempo - É considerado pontos específicos na linha do tempo. Ex.: 2019.
- Intervalos de tempo - É considerado o tempo entre dois pontos, um intervalo de tempo. Ex.: entre 2010 e 2018.

Diante as teorias de localização, assumiu-se aquelas que são caracterizadas por uma ou mais posição no mapa. Sendo assim, um estudo pode ser realizado em diferentes cenários. Ex.: comparação entre Brasil e Holanda, ou somente no Brasil.

### Atividade 3 - Procurar por Ontologias gerais que são baseadas nas teorias de suporte e Selecionar Ontologias gerais

A busca por ontologias gerais ou comuns foram realizadas através das bibliotecas de ontologias apresentadas por d'Aquin e Noy (2012). Devido ao baixo número de ontologias encontradas, utilizou-se mais dois repositórios na realização da busca, sendo eles: o swoogle e AberOWL. E, assim como a busca de ontologias de domínio, foram utilizados repositórios científicos e acadêmicos para, através de artigos de pesquisas, encontrar proposições ontológicas. Após realizar uma extensa busca e selecionar um conjunto de

ontologias candidatas a reúso, optou-se por usar a ontologia Time-OWL para suprir o conhecimento necessário sobre o domínio de tempo, visto a sua capacidade de descrever conceitos temporais através de um vocabulário que expressa eficientemente fatos sobre relações topológicas entre instantes e intervalos. Esta decisão também foi incentivada pela linguagem de implementação da mesma (OWL), e por fazer parte das recomendações da W3C 2017. Referente a ontologia GeoSpatial, obteve-se um conjunto volumoso de ontologias candidatas porém, muitas continham termos e descrições desnecessárias para este trabalho, como coordenadas, por exemplo. Por este motivo, ao analisá-las rigorosamente, decidiu-se utilizar a ontologia Region. A ontologia Region foi implementada em OWL e proposta pelo projeto NAZOU, ela define regiões geográficas básicas, como países, estados, cidades, ruas, moedas e idiomas. O projeto NAZOU tem como objetivo projetar e verificar novas formas de processamento de informação e conhecimento em ambiente heterogêneo, em particular aquisição, representação, organização e manutenção do conhecimento atual, e desenvolver ferramentas para suportar novos modelos de conhecimento (NAZOU, ). Esta ontologia representa de forma simples o conhecimento necessário para este trabalho, facilitando seu entendimento.

### Atividade 3 - Personalizar a Ontologia Geral Selecionada

Esta atividade permite a personalização as ontologias gerais selecionadas, que deve ser realizada de acordo com as necessidades da nova ontologia e dos requisitos já abordados. Para isso, é necessário seguir os seguintes tarefas:

- Podar a ontologia geral de acordo com os recursos necessários - Através desta tarefa é feita uma poda na ontologia selecionada, levando em consideração os recursos úteis e necessários para a ontologia que está sendo criada. Por exemplo: informações sobre a moeda regional, não é necessária na ontologia de localização geográfica, por isso deve ser feita uma remover estas informações. Neste trabalho, esta tarefa foi realizada nas duas ontologias gerais eleitas, onde:
  - *Time* - Da ontologia Time foi eleito para reúso o módulo *temporal entity* composto pelas subclasses: *Time instant*, que representa um ponto em uma linha de tempo, e *Time interval* que representa um intervalo temporal.
  - *Region* - Desta ontologia foi reusado apenas um módulo e uma declaração, onde, o módulo *Administrative region* é composto por subclasses como: país, estado e cidade, afim de estabelecer uma posição geográfica. E a declaração *Geographical region* permite estabelecer o continente do país estabelecido no módulo anterior.
- Enriquecer a ontologia geral - Através desta tarefa é possível estender a ontologia selecionada com novas estruturas conceituais necessárias para a construção da nova ontologia. Esta atividade não foi realizada em nenhuma das ontologias eleitas, visto que os módulos completos atendiam as necessidades da nova ontologia.

- Traduzir a ontologia geral - Esta tarefa tem como objetivo traduzir a ontologia selecionada para a linguagem de implementação da ontologia que está sendo criada. As ontologias eleitas estavam na linguagem OWL, não necessitando tradução.
- Avaliar a Ontologia Geral obtida - Nesta tarefa é realizada uma avaliação sob o conteúdo como um todo, analisando se não existem erros e se as CQs realmente são respondidas através das ontologias gerais reusadas, Esta tarefa foi realizada manualmente por uma especialista em ontologias, validando o reuso das ontologias gerais.

### **Integrar a Ontologia Geral**

Através desta atividade foi realizada a integração das ontologias gerais selecionadas na ontologia que está em desenvolvimento. Esta atividade é detalhada na seção 4.2.6.

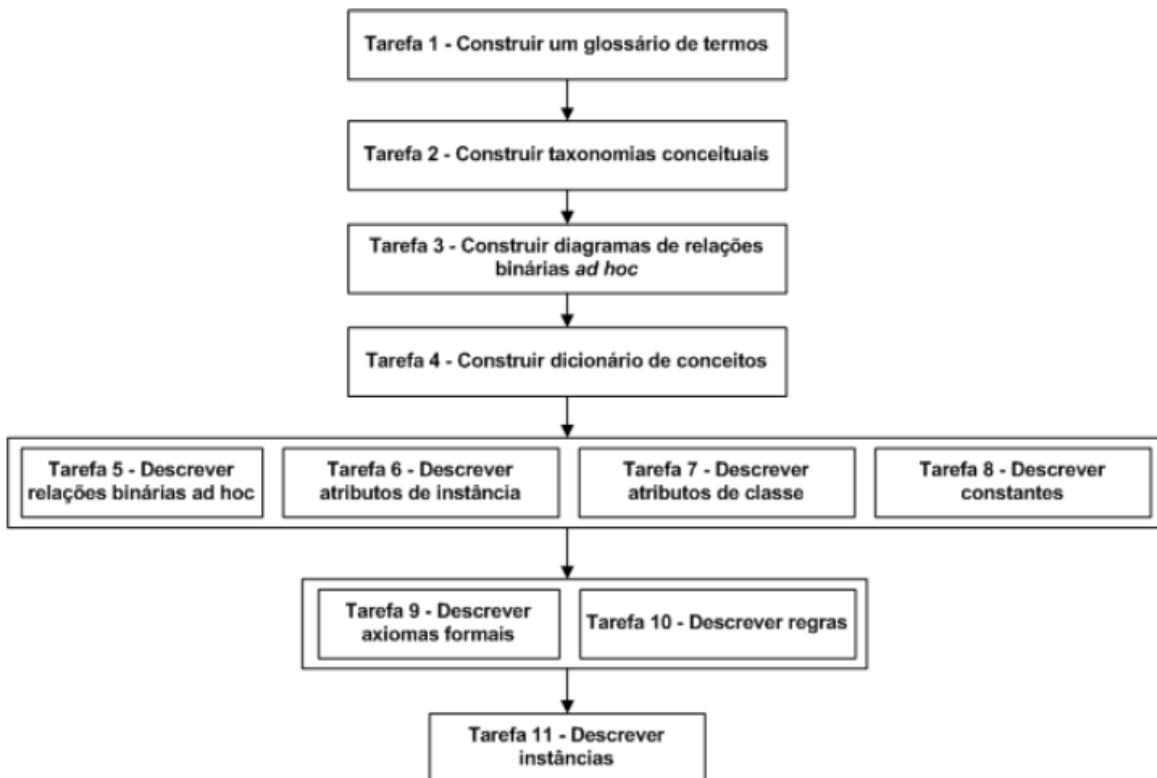
#### **4.2.4 Fase de Projeto**

Esta atividade tem como propósito a criação de um modelo conceitual que represente o problema e a solução da ontologia proposta. A partir desta tarefa, é criado um modelo formal com o uso de frames ou lógica descritiva. Estas tarefas são intituladas como conceituação e formalização, respectivamente. Os modelos criados devem cumprir com os requisitos registrados na documentação de requisitos obtido na fase de especificação. A seguir as tarefas citadas melhor especificadas..

##### **4.2.4.1 Conceituação**

Nesta atividade foi criada uma estrutura de conhecimento do domínio através de um modelo conceitual, com o objetivo de descrever o problema e a sua solução através dos termos extraídos da documentação de requisitos, obtido na fase de especificação da ontologia. A metodologia NeOn não entra em detalhes sobre esta atividade, por isso foi decidido adotar a metodologia *METHONTOLOGY* para servir como guia. A figura 43 explicita o fluxo de tarefas a serem realizadas para a conclusão da conceituação do conhecimento do domínio.

Figura 43 – Fluxo de tarefas para a conceituação do conhecimento do domínio.



Fonte: Adaptado de Fernández-López, Gómez-Pérez e Juristo (1997).

### Tarefa 1 - Construir glossário de termos

A primeira tarefa a ser executada é a de construção de um glossário de termos do domínio, que serão usados para a construção da ontologia. Para a criação deste glossário, serão utilizados os termos identificados no Pré-Glossário (Seção 4.2.1.8) e, conforme discussões e leitura de estudos do domínio, termos que foram identificados e julgados como importantes para estarem contidos na ontologia. Para a classificação dos termos utilizou-se das descrições feitas por Corcho et alli (2005), que permite a tipagem dos termos como: conceito, relação, instância, constante, atributo, axioma e regra.

Em uma planilha<sup>14</sup> online, é explicitado o glossário de termos obtidos para este trabalho, sendo dividida nas seguintes colunas: conceito (nome do termo), Sinônimo, Acrônimo, Descrição (elaboradas a partir de pesquisas sobre o termo aplicado ao domínio da ontologia) e Tipo.

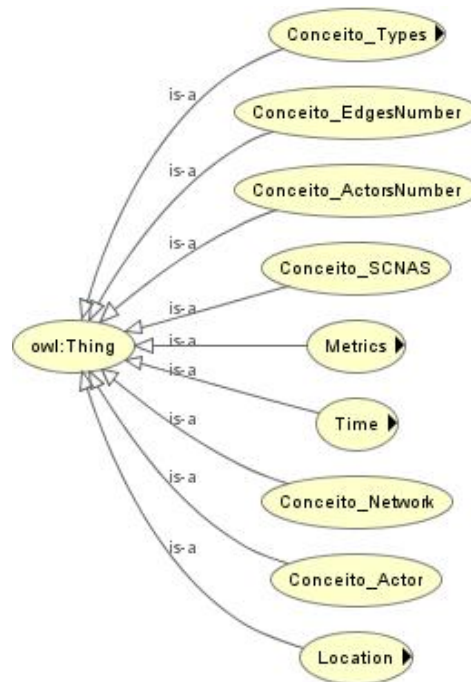
### Tarefa 2 - Construir taxonomias conceituais

Utilizou-se todos os termos tipificados como conceito no glossário para a construção da taxonomia dos mesmos, possibilitando a hierarquização de termos necessários para descrição do conhecimento do domínio trabalhado. A figura 44 representa o primeiro nível da taxonomia de conceitos criada para a ontologia proposta. A taxonomia é iniciada pelo

<sup>14</sup> <<https://bit.ly/36CdQzm>>

conceito Thing, cujo deriva os demais conceitos.

Figura 44 – 1º nível da taxonomia de conceitos do domínio de EARCC.

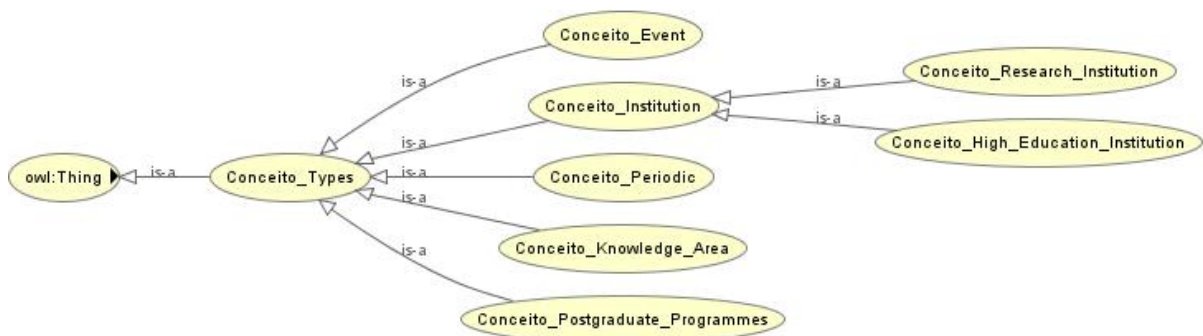


Fonte: Do Autor

Como é possível analisar na figura 44, o primeiro nível da taxonomia é composto por oito eixos. Cada um deles será melhor detalhada a seguir.

1. **Conceito\_Types:** Agrupa conceitos referentes a tipos de rede, sendo: Post-graduate Programmes, Event, Periodic, Intitution e afins.

Figura 45 – Taxonomia doos conceitos do eixo Types

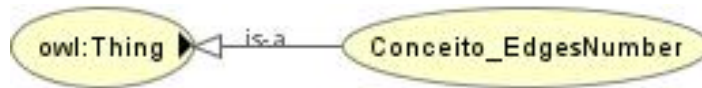


Fonte: Do Autor

2. **Conceito\_EdgesNumber:** Conceito referente ao número de relações entre os atores contidos na rede estudada.

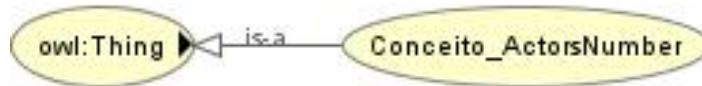
3. **Conceito\_ActorsNumber:** Conceito referente ao número de atores contidos na rede estudada.

Figura 46 – Taxonomia dos conceitos de EdgesNumber



Fonte: Do Autor

Figura 47 – Taxonomia dos conceitos de ActorsNumber



Fonte: Do Autor

4. **Conceito\_SCNAS:** Este conceito representa um estudo de análise de rede de colaboração científica.

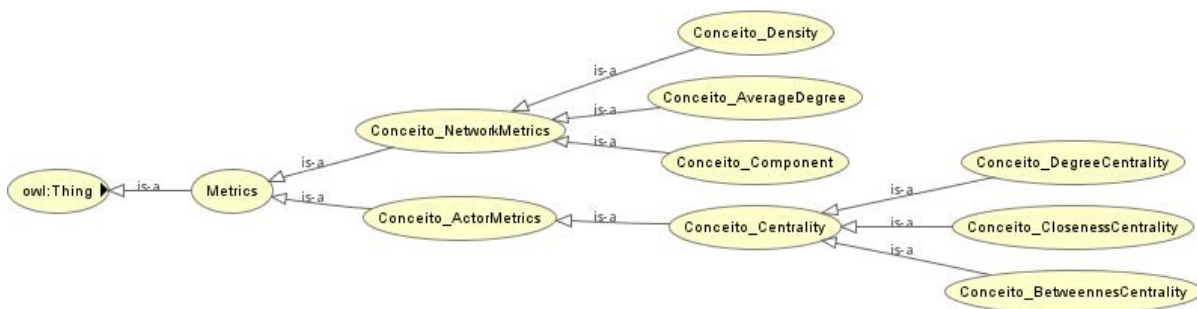
Figura 48 – Taxonomia do conceito SCNAS



Fonte: Do Autor

5. **Metrics:** Agrupa os conceitos de métricas estudadas de EARCC, seja de ator ou rede.

Figura 49 – Taxonomia dos conceitos do eixo Metrics

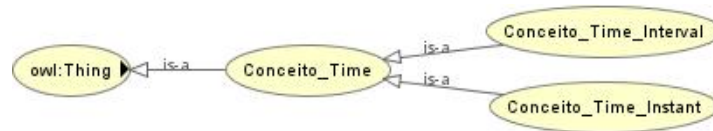


Fonte: Do Autor



6. **Time:** Agrupa todos os conceitos referentes a tempo, seja um intervalo ou um instante. Os termos desta taxonomia serão supridos pela ontologia Time, que supre as necessidades de tempo necessárias para a ontologia.

Figura 50 – Taxonomia dos conceitos de Time



Fonte: Do Autor.

7. **Conceito\_Network:** Este conceito representa uma rede estudada em um EARCC.

Figura 51 – Taxonomia do conceito Network



Fonte: Do Autor.

8. **Conceito\_Actor:** Este conceito representa um ator contido em um EARCC.

Figura 52 – Taxonomia do eixo Actor



Fonte: Do Autor.

9. **location:** Agrupa conceitos referentes a localização como país e estado. Objetiva-se que estes conceitos estejam minimamente descritos e declarados na ontologia a ser reutilizada.

Figura 53 – Taxonomia dos conceitos do eixo Location



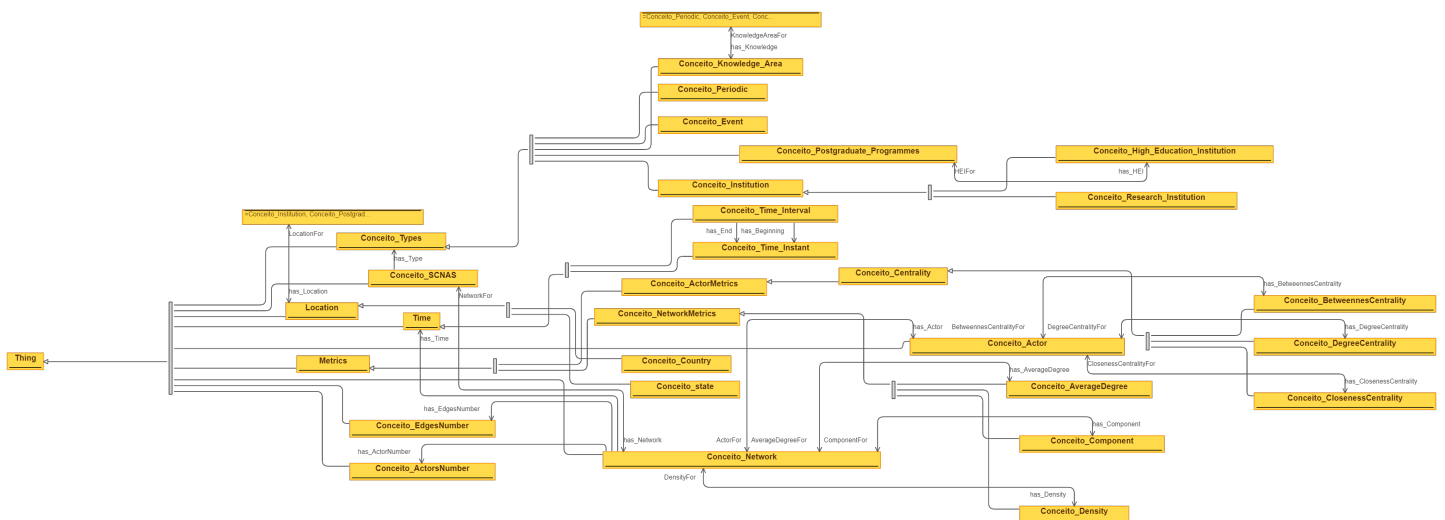
Fonte: Do Autor.

### Tarefa 3 - Construir Diagrama de relações binárias *ad hoc*

Através da leitura de estudos e discussões com uma pesquisadora do domínio, foram identificadas as relações existentes entre os conceitos das taxonomias criadas. Nesta tarefa também foram identificadas as relações inversas, por exemplo, se houver uma relação "has\_Network" entre o conceito "SCNAS" e o conceito "KnowLedge Area", também existe uma relação entre "KnowLedge Area" e "SCNAS" através do relacionamento inverso "NetworkFor". Através da relação inversa é possível recuperar todas as classes relacionadas a uma determinada classe, sendo assim, através de uma visão semântica, seria possível recuperar todos os indivíduos de "SCNAS" relacionados a um determinado indivíduo de "KnowLedge Area".

A figura 54 representa o diagrama de relações construído para a ontologia.

Figura 54 – Diagrama de Relações binárias ad hoc



Fonte: Do Autor

### Tarefa 4 - Construir dicionário de conceitos

O dicionário de conceitos especifica as instâncias, atributos, propriedades e relações binárias dos conceitos que compunham a ontologia criada. O dicionário de conceitos deste trabalho foi desenvolvido na plataforma Google Sheets, e pode ser consultado online<sup>15</sup>.

### Tarefa 5 - Descrever relações binárias *ad hoc*

Nesta tarefa foi descrita em uma planilha<sup>16</sup> online, todas as relações binárias possíveis entre as classes da ontologia criada. A tabela é composta pelas colunas: relação, que mantém o nome das propriedades de objetivos, conceito origem, que mantém o conceito

<sup>15</sup> <<http://tiny.cc/e5xkez>>

<sup>16</sup> <<https://bit.ly/2sqTOSX>>

de onde parte a relação unidirecional, a cardinalidade, que explicita as cardinalidades de cada relação, o conceito alvo, que mantém os conceitos de destino de cada relação, e a relação inversa, que mantém as propriedades de objetos referentes ao inverso de determinada relação.

#### **Tarefa 6 - Descrever atributos de instância**

Esta tarefa tem como propósito detalhar os atributos de instâncias que estão presentes no dicionário de conceitos, criando uma tabela de atributos de instância. Os atributos de instâncias são atributos (ou propriedades) que são associados a cada instâncias, ou seja, um atributo próprio para cada instância de classe. Visto que todos os atributos usados foram importados, não houve a criação de atributos de instâncias. No entanto, a descrição dos atributos de instâncias presentes na ontologia estão detalhados em uma tabela<sup>17</sup> *online*.

#### **Tarefa 7 - Descrever atributos de classe**

Nesta tarefa é criado uma tabela de atributos de classe, que detalha os atributos de classe contidos no dicionário de conceitos. Atributos de classes têm como função descrever os conceitos e assumir valores nas classes que estão vinculadas. Neste trabalho, nenhum atributo de classe foi criado.

**Tarefa 8 - Descrever constantes** Esta tarefa tem como objetivo criar uma tabela de constantes, detalhando as constantes presentes no glossário de termos. A tabela deve conter detalhes como: nome da constante, tipo de valor, valor, a unidade de medida para constantes numéricas e os atributos que podem ser inferidos ou calculados usando a constante.

As constantes utilizadas na ontologia estão descritas em uma tabela<sup>18</sup> *online*.

#### **Tarefa 9 - Descrever axiomas formais**

Através desta tarefa é realizada a identificação dos axiomas formais necessários para a ontologia que está sendo desenvolvida, as organizando em uma tabela onde deve conter detalhes sobre a mesma. Neste trabalho esta tarefa não foi realizada, visto que a ontologia é composta por reuso de módulos de ontologias já existentes.

#### **Tarefa 10 - Descrever regras**

Através desta tarefa é identificado as regras essenciais para o desenvolvimento da ontologia proposta. Estas regras são estruturadas e detalhadas em uma tabela de Regras. Neste trabalho esta tarefa não foi realizada devido ao escopo adotado para a ontologia proposta.

#### **Tarefa 11 - Descrever instâncias**

Esta atividade tem como objetivo definir as instâncias relevantes que estão presentes no dicionário de conceitos. Estas instâncias são detalhadas e organizadas em uma tabela de instâncias, onde contém as seguintes informações: nome da instância, o conceito o qual ela pertence e os valores dos seus atributos, quando conhecidos.

---

<sup>17</sup> <<http://tiny.cc/052ggz>>

<sup>18</sup> <<http://tiny.cc/fr3ggz>>

Uma planilha<sup>19</sup> online foi criada para organizar e descrever as instâncias da ontologia a ser criada.

#### 4.2.5 Formalização

A formalização tem como objetivo transformar o modelo conceitual (desenvolvido na fase de conceituação) em um modelo formal. Para realizar a formalização da ontologia proposta, foram utilizados as seguintes documentações: Glossário de Termos, Taxonomias Conceituais, Dicionário de Conceitos, Diagrama de Relações Binárias, Tabela de Relações Binárias, Tabela de Atributos de Instância, Tabela de Atributos de Classe, Tabela de Constantes e a Tabela de Instâncias. Mediante as documentações, os conceitos encontrados foram transformados em classes, que foram estruturadas em hierarquias já definidas taxonomicamente. Os atributos de instâncias foram transformados em propriedades de dados, assim como as relações binárias proporcionaram a criação das propriedades dos objetos.

A fim de facilitar o entendimento da ontologia, para aqueles que futuramente irão fazer uso da mesma, criou-se uma propriedade de anotação chamada “comment” que será associada a todas as classes, com uma breve descrição em língua inglesa.

É importante observar que grande parte do conhecimento do domínio necessário para a ontologia, foi suprido pelos módulos reusados na ontologia propostas, a integração destes módulos é detalhado na seção 4.2.6. Por tanto, necessitou-se formalizar apenas nos termos não supridos pelas ontologias reusadas. Uma planilha *online* explicita os termos formalizados neste trabalho.

#### 4.2.6 Fase de Integração

Através desta atividade, é realizada integração dos módulos eleitos para reuso na nova ontologia. Diante as candidatas a reuso, a metodologia NeOn propõe três formas de integração:

1. Realizar a integração sem quaisquer alterações, ou seja, da mesma forma que se encontra nas especificações da implementação original.
2. Realizar a integração com alterações, ou seja, houve alterações na ontologia que está sendo integrada.
3. Realizar o merge entre um conjunto de ontologias de um mesmo domínio, afim de obter uma nova ontologia.

Para este trabalho se optou por fazer uso da 1º forma de integração, visto que foram integrados módulos completos de ontologias, sem quaisquer alterações. Neste trabalho a

---

<sup>19</sup> <<http://tiny.cc/ix3ggz>>

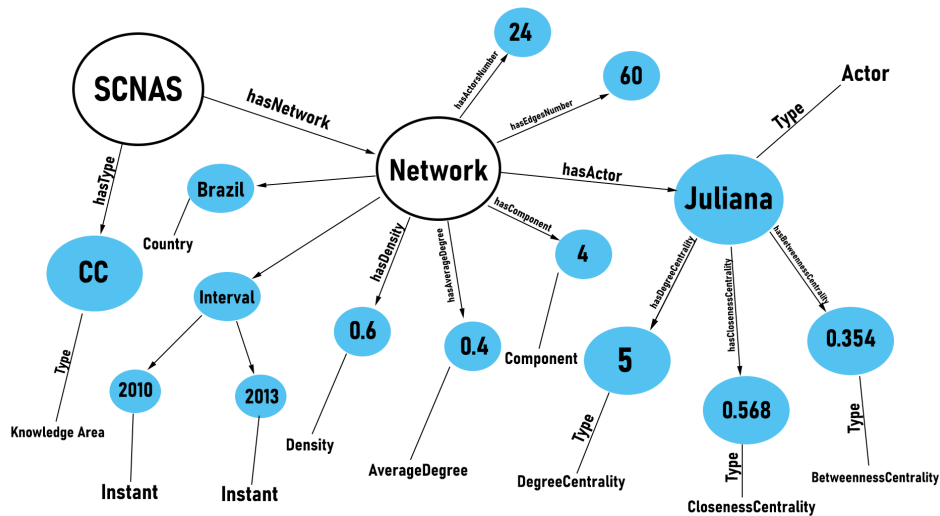
integração dos módulos foi feito através do *direct imports*, onde, primeiramente, realizou-se uma cópia dos módulos eleitos para novas ontologias vazias distintas, e posteriormente estas ontologias foram importadas para o projeto da ontologia proposta. Esta atividade foi realizada na ferramenta protégé, um *framework* de ontologias de código aberto, que, através de interface gráfica, permite a criação e manipulação de ontologias.

Considerando a documentação e o conhecimento sobre o domínio, neste trabalho a identificação das relações foi realizada de forma manual. Considera-se também que a fase de integração foi realizada de forma paralela à fase de Conceituação, sendo assim os conceitos, relações e propriedades identificados puderam ser analisados nas ontologias eleitas.

#### 4.2.7 Fase de Implementação

Nesta seção contém alguns detalhes sobre a implementação da ontologia proposta. Para este trabalho, utilizou-se a linguagem de implementação OWL para a criação da nova ontologia, seu código foi gerado de forma automática através da ferramenta Protégé mediante o modelo formal fornecido. A versão final da ontologia pode ser encontrada através do seguinte endereço: <https://github.com/Laltany/SCNAS-Ontology>.

A figura 55 mostra o um modelo conceitual da ontologia construída por neste trabalho, denominada *SCNAS Ontology*. Como pode ser observado, um indivíduo do tipo “SCNAS” se relaciona com o indivíduo “CC”, do tipo *Knowledge Area*. Através da propriedade de objeto “hasNetwork” um SCNAS pode se relacionar com, ou mais, indivíduo do tipo *Network*. Uma Rede (*Network*) pode se relacionar com indivíduos de tipo Métricas de Rede, como: *Density*, *Average Degree*, *Componenet*. Uma Rede também pode se relacionar com um indivíduo do tipo “Country”, que no diagrama é representado pelo indivíduo *Brazil*. A ontologia também permite identificar o tempo dos dados utilizados na rede, para isso uma Rede se relaciona com um indivíduo do tipo “Interval” que se relaciona com dois indivíduos do tipo “instant”. Através da propriedade de objeto “hasActor” uma rede se relaciona com um, ou mais, indivíduos do tipo “Actor”. Um *Actor* pode se relacionar com indivíduos com tipos referentes a métricas de atores, como: *Degree Centrality*, *Betweenness Centrality* e *Closeness Centrality*.

Figura 55 – Modelo Conceitual do SCNAS *Ontology*

Fonte: Do Autor.

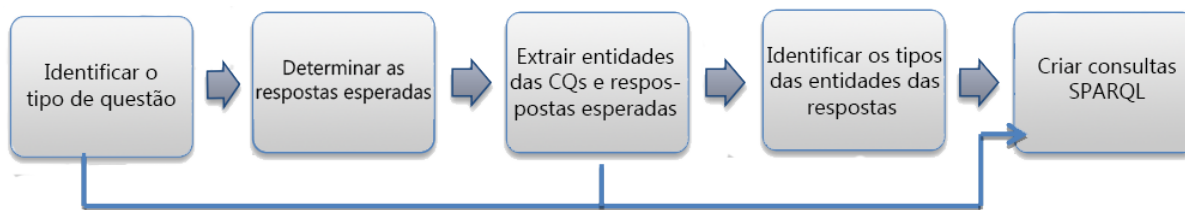
#### 4.2.8 Avaliação da ontologia

De acordo com Suarez (2010), o processo de avaliação da ontologia tem como propósito avaliar os componentes que estão presentes na ontologia criada, de forma que é determinado o que a ontologia consegue definir de forma correta ou incorreta. Assim como, também é definido as incapacidades de definições da ontologia. A avaliação proposta pela metodologia NeOn é dividida em: verificação e validação.

A verificação tem como objetivo avaliar se a ontologia está sendo construída de forma correta, considera-se neste caso os requisitos contidos na documentação de requisitos e o planejamento de construção da ontologia. Neste trabalho, a atividade de verificação acompanhou todo o processo de desenvolvimento da ontologia, visando garantir o produto estimado de cada fase.

A validação tem como objetivo comparar os significados dos conceitos e definições contidos na ontologia criada, com os conceitos e definições do domínio determinado para a ontologia. Assim como Bortolato e Paulo (2014), adotou-se a proposta de Zemmouchi-Ghomari e Ghomari (2013), que condiz em validar a ontologia através das saídas obtidas em consultas SPARQL. Para a criação destas consultas, é necessário realizar a transcrição das questões de competência (criadas nas especificações de requisitos) para a linguagem de consulta SPARQL. Esta transcrição é realizada através de um conjunto de atividades propostas por Zemmouchi-Ghomari e Ghomari (2013). A Figura 56 representa o processo de atividades propostas e a seguir elas são detalhadas.

Figura 56 – Processo de transcrição das CQs para SPARQL.



Fonte: Adaptado de Zemmouchi-Ghomari e Ghomari (2013).

#### 4.2.8.1 Identificar o tipo de Questão

A primeira atividade a ser realizada é a identificação das questões de competência que, de acordo com Zemmouchi-Ghomari e Ghomari (2013) podem ser classificadas como:

- Questões de definição - que começam com “O que é / são ”ou“ o que significa ”;
- Questões booleanas - suas respostas são “Sim/Não” ou "Verdadeiro/Falso";
- Questões Factuais - suas respostas são fatos ou informações precisas;
- Questões de lista - suas respostas são dadas através de uma lista de entidades.
- Questões complexas - suas respostas iniciam com “Como...” e “Por que...”. Onde se obter uma resposta precisa é bastante improvável.

Os tipos devido a extensão a quantidade de questões, foi criada uma tabela<sup>20</sup> *online* que explicita a identificação das questões.

#### 4.2.8.2 Determinar as respostas esperadas

Nesta atividade foram determinadas as respostas esperadas para cada questionamento da lista de questões de competência. Aproveitou-se nesta etapa as respostas já obtidas nas especificações de requisitos, cuja tabela<sup>21</sup> está disponível de forma *online*.

#### 4.2.8.3 Extrair entidades das CQs e das respostas

Através desta atividade foi realizada a extração de termos tanto das questões, quanto de suas respostas. Estes termos deveriam ser representados por entidades na ontologia criada, para que se possa realizar consultas. Caso estas entidades não estejam presentes na ontologia, a mesma deve ser revisada. O Pré-Glossário<sup>22</sup> criado na fase de Especificação da ontologia, consta todas as informações necessárias para esta atividade.

<sup>20</sup> <<http://tiny.cc/5cxggz>>

<sup>21</sup> <<http://tiny.cc/qwwggz>>

<sup>22</sup> <<http://tiny.cc/w7oodz>>

## 4.2.9 Identificar os tipos das entidades da ontologia

Nesta atividade, todas as entidades identificadas na seção anterior serão identificadas na ontologia como uma: classe, propriedade de dados, propriedade de objeto ou instância. Uma planilha<sup>23</sup> *online* explicita o tipo da entidade. Esta identificação é realizada devido a sintaxe da linguagem de consulta SPARQL.

### 4.2.9.1 Criar consultas SPARQL

Através das atividades de identificação das respostas esperadas e a localização das entidades, foi possível escrever um conjunto de consultas em SPARQL. A partir da resposta dada por estas consultas, será possível analisar se a ontologia realmente é válida. Uma planilha<sup>24</sup> *online* foi criada para apresentar as consultas realizadas para obter as respostas desejadas de acordo com suas questões.

### 4.2.9.2 Executar as consultas SPARQL

Para esta atividade, utilizou-se a ferramenta Jena para criar um banco de triplas, neste banco foi armazenado triplas de teste para que se pudesse realizar consultas SPARQL. Estas triplas são frutos de uma extração de teste realizado em alguns artigos coletados na seção 4.1. A figura 57 explicita as consultas sendo realizadas na através da ferramenta.

Figura 57 – Execução das consultas SPARQL na JENA

```

14 PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
15 PREFIX voc:<http://ns.inria.fr/semsna/2009/06/21/voc#>
16 PREFIX time:<http://www.w3.org/2006/time#>
17 PREFIX EARCC:<http://www.semanticweb.org/x_rom/ontologies/2019/4/EARCC#>
18 PREFIX re:<http://www.w3.org/2000/10/swap/reason#>
19 PREFIX ti:<http://www.ontologydesignpatterns.org/cp/owl/timeinterval.owl#>
20
21 select ?rede ?actor ?Centralidade ?value
22 where {
23   ?rede rdfs:type EARCC:knowledge_area.
24   ?rede EARCC:hasActor ?actor .
25   ?actor voc:hasNAMetricDegreeCentrality ?Centralidade .
26   ?Centralidade voc:hasValue ?value
27 }
28 LIMIT 25
  
```

QUERY RESULTS

Table Raw Response

Showing 1 to 3 of 3 entries

rede	actor	Centralidade	value
EARCC:i_rede2	EARCC:vaccaro	EARCC:7,000	"7,000"
EARCC:i_rede2	EARCC:Bitencourt	EARCC:7,000	"7,000"
EARCC:i_rede2	EARCC:Gomes	EARCC:10,000	"10,000"

Showing 1 to 3 of 3 entries

Fonte: Do Autor.

<sup>23</sup> <<http://tiny.cc/y0wggz>>

<sup>24</sup> <<http://tiny.cc/hhwggz>>



### 4.2.9.3 Avaliar resultados das consultas

Nesta atividade os resultados obtidos na seção 4.2.9.2 são avaliados, de forma que se descubra se a ontologia realmente tem capacidade de responder as questões de competências propostas, comparando os resultados obtidos com os resultados esperados. Assim como Bortolato e Paulo (2014), neste trabalho foi adotado um método de quantificação para avaliar a capacidade de representação da ontologia. Sendo assim, a cada critério cumprido pela ontologia, um valor foi atribuído. Os critérios utilizados foram:

- A ontologia contém todas as entidades necessárias para traduzir de forma correta a questão de competência onde, através da consulta SPARQL, obteve-se a resposta esperada de forma integral. O valor deste critério é igual a 3.
- A ontologia contém todas as entidades necessárias para traduzir de forma correta a questão de competência onde, através da consulta SPARQL, a resposta obtida foi parcialmente correta. O valor deste critério é igual a 2.
- A ontologia contém todas as entidades necessárias para traduzir de forma correta a questão de competência onde, através da consulta SPARQL, a resposta obtida não estava correta. O valor deste critério é igual a 1.
- A ontologia não contém entidades o suficiente para traduzir a questão de competência. O valor deste critério é igual a 0.

Uma planilha<sup>25</sup> *online* foi criada para apresentar a avaliação das consultas realizadas, através dos critérios descritos a cima.

Assim como recomenda a metodologia NeOn, os valores adquiridos através dos critérios descritos acima são utilizados em uma fórmula matemática que define a capacidade de cobertura da ontologia criada. Ou seja, através de um produto porcentual é possível definir a capacidade de resposta da ontologia, diante as questões de competência impostas. A fórmula matemática que define a cobertura é dada como:

$$Cobertura = \frac{\sum Valor}{n^{\circ}deQCs \times ValorReferência} \times 100$$

Onde:

- Cobertura é o produto que se espera buscar, ou seja, o produto porcentual que explicita a capacidade de respostas as CQs da ontologia.
- $\sum Valor$  é o somatório dos valores atribuídos as CQs.
- $n^{\circ}deQCs$  é a quantidade de questões de competência que devem ser respondidas pela ontologia.

<sup>25</sup> <<http://tiny.cc/jswggz>>

- ValorReferência é a escala máxima que por ser o atribuído a cada CQ, neste caso trabalharemos com ValorReferência = 3.

Aplicação da fórmula:

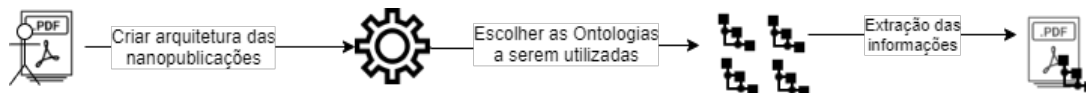
$$Cobertura = \frac{162}{54 \times 3} \times 100$$

Através da aplicação da fórmula, concluiu-se que a ontologia desenvolvida tem 100% de capacidade de responder corretamente as questões de competências impostas na ORSD.

### 4.3 Construção das Nanopublicações

Esta seção descreve a criação das nanopublicações propostas por este trabalho. O fluxo das atividades necessárias para a construção das nanos é representado pela figura 58.

Figura 58 – Processo de criação das nanopublicações



Fonte: Do Autor.

Primeiramente, definiu-se a arquitetura da nanopublicação a ser adotada para este trabalho. Logo após, foram definidas as ontologias capazes de suprir as necessidades de cada nível da arquitetura. Por fim, foi realizada através das ontologias definidas, a extração das informações que compunham a nanopublicação proposta.

As seções a seguir detalham as atividades brevemente descritas a cima.

#### 4.3.1 Arquitetura da Nanopublicação

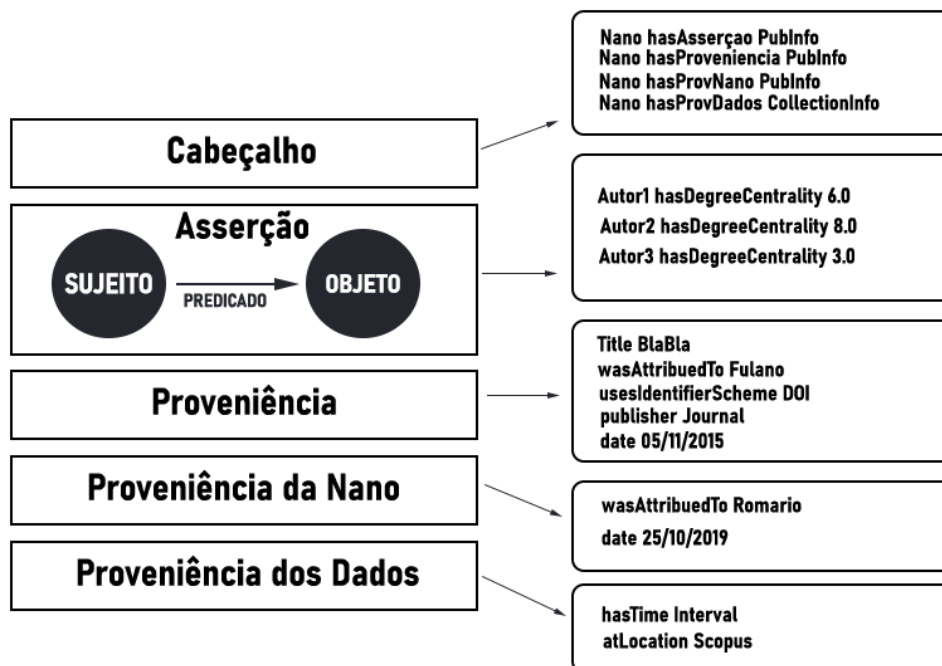
Como abordado na seção 2.2.1 a arquitetura mínima para uma nanopublicação é composta pelos níveis de asserção, proveniência da afirmação e proveniência da nanopublicação, podendo-se criar novos níveis em sua estrutura para que se possa atingir os objetivos da nanopublicação a ser desenvolvida.

Neste trabalho, a criação de nanopublicações aplicada a EARCC tem como objetivo agilizar a recuperação de pequenas afirmações contidas em estudos deste domínio. Por isso, através da análise de artigos e conversas com uma especialista do domínio, foi observada a necessidade de criar um quarto nível sob a arquitetura mínima. Dado que o nível de asserção é composto por afirmações referentes a resultados de métricas aplicados a dados, que são oriundos de uma coleta de dados, o quarto nível tem como objetivo manter triplas referentes a informações sobre a coleta dos dados estudados pelos artigos. O

tempo dos dados coletados (ex.: de 2010 a 2015) e a identificação do(s) repositório(s) que proveu os dados estudados são informações que se mostraram importantes e suficientes para o novo nível da arquitetura. Sendo assim, ao verificar o quarto nível da arquitetura, os pesquisadores podem identificar a proveniência dos dados trabalhados nos artigos do domínio.

A figura 59 explicita a arquitetura de nanopublicação adotada para este trabalho.

Figura 59 – Arquitetura da nanopublicação adotada neste trabalho



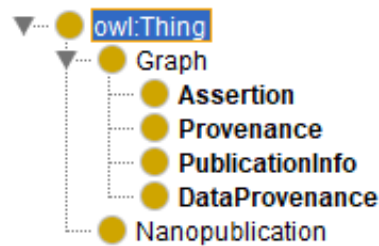
Fonte: Do Autor.

Como pode ser observado na figura 59, a arquitetura da nanopublicação adotada para este trabalho, nomeada como *Nanopublication Scientific Collaboration Network Analysis* (nanoSCNAS), contém 4 principais níveis, sendo eles:

- Asserção - Mantém as afirmações contidas nos documentos do domínio, ou seja, resultados de métricas e informações obtidas pela rede estudada.
- Proveniência - Mantém afirmações referentes a proveniência do estudo, como: título, nome dos autores, data de publicação, nome do periódico, DOI ou link do documento.
- Proveniência da Nano - Mantém afirmações referentes a criação da nanopublicação, como: quem criou a nano e a data de criação da nano.
- Proveniência dos dados - Mantém afirmações sobre a coleta dos dados que permitiram a criação da rede estudada, como: tempo dos dados e a identificação do repositório que proveram os dados.

A figura 60 representa a hierarquia de classes referentes aos níveis da nanoSCNAS. Como pode ser observado na figura 59, no topo se encontra o cabeçalho onde permanecem as identificações dos níveis da arquitetura a serem usadas na nanopublicação. Ou seja, o relacionamento entre um indivíduo de classe Nanopublicação (Nanopublication) com os indivíduos de classe Asserção (Assertion), Proveniência (Provenance), Proveniência da Nano (PublicationInfo) e Proveniência dos Dados (DataProvenance).

Figura 60 – Hierarquia dos níveis da nanoSCNAS



Fonte: Do Autor.

A figura 61 exemplifica um cabeçalho de uma nanopublicação.

Figura 61 – exemplo de cabeçalho de uma nanoSCNAS

```
<!-- http://www.nanopub.org/nschema#pub1_Nanopublication -->
<owl:NamedIndividual rdf:about="http://www.nanopub.org/nschema#pub1_Nanopublication">
  <rdf:type rdf:resource="http://www.nanopub.org/nschema#Nanopublication"/>
  <np:hasAssertion rdf:resource="http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#SCNAS1"/>
  <np:hasDataProvenance rdf:resource="http://www.nanopub.org/nschema#pub1_Collect"/>
  <np:hasProvenance rdf:resource="http://www.nanopub.org/nschema#pub1_Provenance"/>
  <np:hasPublicationInfo rdf:resource="http://www.nanopub.org/nschema#pub1_publicationInfo"/>
</owl:NamedIndividual>
```

Fonte: Do Autor.

Após a definição da arquitetura da nanoSCNAS, tornou-se necessário eleger as ontologias são capazes de identificar e descrever as informações necessárias para a criação das nanopublicações. A seção 4.3.2 relata as ontologias adotadas para cada nível da nanoSCNAS.

### 4.3.2 Escolha das ontologias

Para que fosse possível a identificação das afirmações a serem mantidas nos níveis da arquitetura da nanoSCNAS, foi necessário fazer uso de algumas ontologias. Diante o nível de asserção utilizou-se a ontologia SCNAS (detalhada na seção 4.2), já que ela tem a capacidade de identificar resultados de métricas aplicadas em redes de colaboração científica e informações importantes, como a localização da rede e o tempo da rede.

Referente ao nível de proveniência e proveniência da nano, optou-se por fazer uso da ontologia PROV-O<sup>26</sup> juntamente com a ontologia DataCite<sup>27</sup>.

Recomendada e aprovada pela W3C, a ontologia PROV-O fornece um conjunto de classes e propriedades que permitem a representação de informações de proveniência em diferentes contextos. Na nanoSCNAS a PROV-O tem o papel de identificar os autores dos estudos, assim como o autor da nanopublicação, de forma que os autores estão classificados como “*Person*” (subclasse de *Agent*) e são atribuídos a proveniência, ou proveniência da nano, através da propriedade de objeto *wasAttributedTo*. Visto que um *Agent* é uma entidade com responsabilidade por uma atividade existente e que a propriedade *wasAttributedTo* é a atribuição de uma tarefa a um *Agent*.

Para suprir a descrição das informações restantes (data de publicação, data de criação da nano, nome do periódico, *DOI* ou *link* do documento) utilizou-se a ontologia *DataCite*. A *dataCite* é uma ontologia que permite a identificação precisa e consistente de um recurso para fins de citação e recuperação. Na nanoSCNAS, ela fornece informações que não podem ser descritas pela PROV-O, através das seguintes anotações: *title*, capaz de identificar o título do estudo, e *date*, identifica a data de publicação do artigo. A ontologia também forneceu as classes "*primary resource identifier*" e "*alternate resource identifier*", que classificam os identificadores dos estudos como um DOI, ou um identificador alternativo. O identificador é atribuído a proveniência através da propriedade de objeto *usesIdentifierScheme*.

Através da DCMI Metadata Terms<sup>28</sup> foi obtido o termo de anotação "*publisher*", cujo objetivo, na camada de proveniência, é identificar o periódico vinculado ao estudo. A identificação do periódico do estudo se mostrou importante para estabelecer uma confiança a mais nos dados contidos na camada de asserção, visto que o estudo precisou passar por um processo de avaliação antes de sua publicação em determinado periódico.

Por fim, na camada de Proveniência dos dados, utilizou-se a ontologia *Time* (contida na ontologia SCNAS) para representar os instantes e intervalos de tempo dos dados estudados nos artigos, de forma que são atribuídos a camada de Proveniência dos Dados através da propriedade de objeto *hasTime* (SCNAS). Para a representação do repositório que forneceu os dados estudados, foi utilizada a classe *Location* que é atribuída ao último nível através da propriedade de objeto *atLocation*.

Após a escolha das ontologias a serem utilizadas para identificação das informações necessárias para a criação das nanopublicações, foi iniciada a atividade de extração das informações. A atividade de extração das informações está detalhada na seção 4.3.3.

<sup>26</sup> <<https://www.w3.org/TR/prov-o/#wasAttributedTo>>

<sup>27</sup> <<https://sparontologies.github.io/datacite/current/datacite.html>>

<sup>28</sup> <<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#terms-publisher>>

### 4.3.3 Extração das informações

Esta atividade consiste em realizar a extração das informações contidas nos estudos coletados na seção 4.1, através dos termos das ontologias especificadas na seção 4.3.2. Visto que a ontologia SCNAS já havia sido populada na Seção 4.2.7, e que a mesma foi importada para a estrutura criada para as nanopublicações, não houve a necessidade de identificar as afirmações do nível de asserção, pois aproveitou-se dos axiomas importados. Sendo assim, todos os indivíduos classificados como SCNAS, na ontologia SCNAS, passaram também a ser classificados como “Assertion” na ontologia da nanoSCNAS e relacionados as suas respectivas nanopublicações através da propriedade de objeto “*hasAssertion*”.

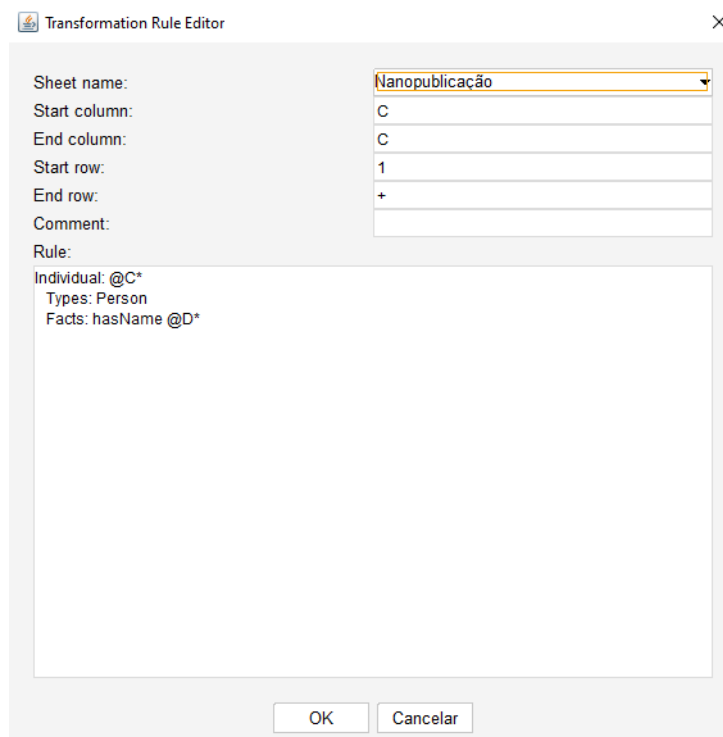
Como a quantidade de estudos a serem extraídos era pequena, decidiu-se realizar a extração dos dados necessários para os níveis de proveniência e proveniência dos dados de forma manual. Todas as informações extraídas, referentes as camadas citadas, foram armazenadas em uma planilha online através da ferramenta *Google Sheets*. No total obteve-se 140 informações que foram organizadas em diferentes abas de acordo com as suas classes, também foi criada uma coluna com o nome do individuo que iria se referir a tal informação na ontologia. Ou seja, os nomes dos autores foram colocados na aba Autores, juntamente com nome do individuo que iria representar esse autor na ontologia. É importante ressaltar que apenas informações que iriam ser classificadas na ontologia foram nomeadas, dados que proveram anotações não necessitaram de identificação. Essa abordagem necessária para que mais tarde essa planilha pudesse ser importada para o protégé, onde foram criadas todos os axiomas de forma automática através da *tools "Create axioms from excel workbook"*. Para a criação dos axiomas referentes as informações extraídas, foram identificadas as classes de cada informação, assim como suas relações e anotações, através de pequenas codificações que geraram as triplas RDF de forma automática, assim como mostrado na figura 62.

Para a criação das nanopublicações utilizou-se da mesma metodologia, em uma planilha foi criada uma coluna com o nome da nanopublicação (ex.: pub1\_Nanopublication) e logo após colunas com os nomes dos níveis das nanos (ex.: pub1\_Provenance, pu1\_DataProvenance e pu1\_PublicationInfo). Nas colunas seguintes foram colocados os nomes que identificavam os indivíduos que deveriam se relacionar com cada nível das nanos. Em seguida, através da identificação das propriedades de objeto das ontologias adotadas para cada nível da arquitetura da nanoSCNAS, todos os indivíduos foram estruturadas em seus respectivos níveis. Como exemplo dessa estruturação, a figura 63 explicita a estruturação dos níveis de proveniência.

Para o nível “*PublicationInfo*” não foi necessário fazer qualquer tipo de extração, pois o mesmo apresenta somente nome do autor das nanopublicações, sendo este o autor deste trabalho, e a data de criação das mesmas.

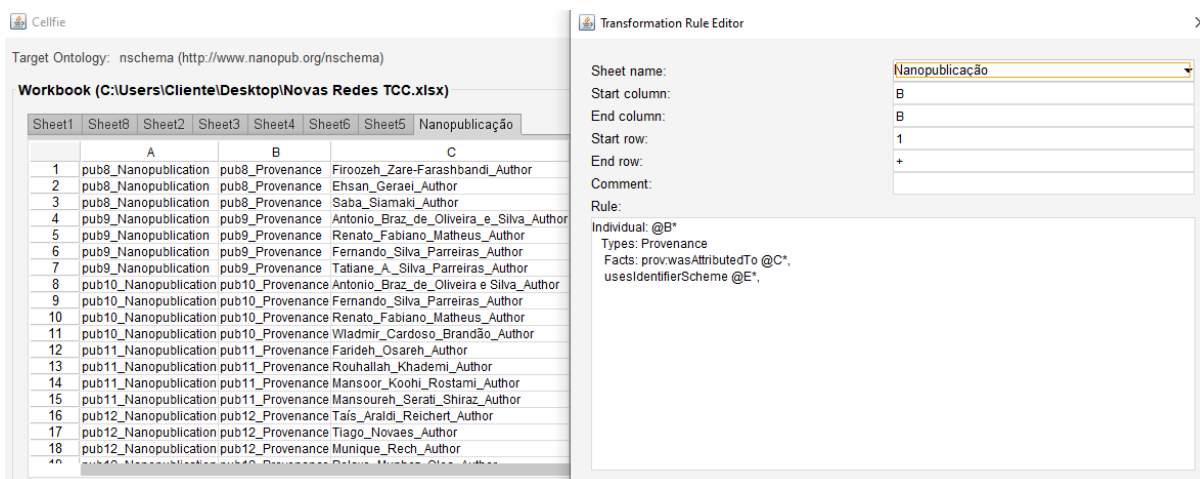
Por fim, realizou-se a estruturação das nanopublicações como um todo, onde as ca-

Figura 62 – Exemplo de código de geração dos axiomas das informações



Fonte: Do Autor.

Figura 63 – Relacionando os Autores e DOI as proveniências de suas respectivas nanos



Fonte: Do Autor.

madras de proviniência, proviniência da nano e proviniência dos dados foram relacionadas as suas respectivas nanopublicações através das propriedades de objeto *hasProvenance*, *hasPublicationInfo* e *hasDataProvenance*. As nanopublicações criadas por este trabalho estão disponíveis em um repositório<sup>29</sup> *online* e aberto para que qualquer interessado possa ter acesso.

<sup>29</sup> <<https://github.com/Laltany/nanoSCNAS>>

#### 4.4 Armazenamento em uma Triple Store

Após a criação das nanopublicações, foi necessário armazená-las em uma *triple Store* para que se pudesse fazer consultas e verificar se as nanos estavam retornando corretamente. Uma *Triple Store* nada mais é do que um servidor SPARQL, que armazena triplas RDF.

A *Triple Store* eleita para este trabalho foi o *Jena Fuseki* que, através de uma interface gráfica, permite a criação, monitoramento e administração do banco de triplas. Também fornece uma aba específica para realizar consultas no banco através da linguagem SPARQL. Neste trabalho, o *Jena Fuseki* foi instalado em uma máquina pessoal e rodou de forma local em todos os momentos. A importação da ontologia da nanoSCNA foi feita de forma fácil e rápida através da interface da ferramenta. É necessário lembrar que também foi necessário importar a ontologia SCNAS para o banco, pois os dados do nível de asserção estavam contidas nela.

Figura 64 – NanoSCNAS hospedado no banco SPARQL Jena Fuseki

The screenshot displays the Jena Fuseki SPARQL query interface. At the top, there are tabs for 'PREFIXES', 'rdf', 'rdfs', 'owl', 'xsd', and a search icon. Below this, the 'SPARQL ENDPOINT' is set to '/SCNAS/iquery', and the 'CONTENT TYPE (SELECT)' is 'JSON' and 'CONTENT TYPE (GRAPH)' is 'Turtle'. The query editor contains the following SPARQL query:

```

1
2
3 SELECT ?subject ?predicate ?object
4 WHERE {
5   ?subject ?predicate ?object
6 }
7 LIMIT 25

```

The 'QUERY RESULTS' section shows a table with 10 entries. The table has columns for 'subject', 'predicate', and 'object'. The results are as follows:

	subject	predicate	object
1	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#PB_network1_SCNAS12>	<http://ns.inria.fr/semsna/2009/06/21/SNAMetric#hasSNAMetricClosenessCentrality>	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#32_ClosenessCentrality>
2	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#PB_network1_SCNAS12>	<http://ns.inria.fr/semsna/2009/06/21/SNAMetric#hasName>	"PB"
3	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#PB_network1_SCNAS12>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/2002/07/owl#NamedIndividual>
4	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#PB_network1_SCNAS12>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#Actor>
5	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#PB_network1_SCNAS12>	<http://ns.inria.fr/semsna/2009/06/21/SNAMetric#hasSNAMetricBetweennessCentrality>	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#0_95_BetweennessCentrality>
6	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#PB_network1_SCNAS12>	<http://ns.inria.fr/semsna/2009/06/21/SNAMetric#hasSNAMetricDegreeCentrality>	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#6_DegreeCentrality>
7	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#Baisheng_Jiang_network1_SCNAS2>	<http://ns.inria.fr/semsna/2009/06/21/SNAMetric#hasSNAMetricClosenessCentrality>	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#1_4_ClosenessCentrality>
8	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#Baisheng_Jiang_network1_SCNAS2>	<http://ns.inria.fr/semsna/2009/06/21/SNAMetric#hasName>	"Baisheng Jiang"
9	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#Baisheng_Jiang_network1_SCNAS2>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/2002/07/owl#NamedIndividual>
10	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#Baisheng_Jiang_network1_SCNAS2>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.semanticweb.org/cliente/ontologies/2019/9/SCNAS#Actor>

Fonte: Do Autor.



Para verificar se as nanopublicações criadas haviam sido implementadas corretamente, utilizou-se das mesmas questões de competências elaboradas na seção 4.2.1.8 para realizar novas consultas no banco. O objetivo foi verificar se a arquitetura das nanopublicações criadas possibilitavam a recuperação das respostas esperadas de cada questão, juntamente com suas proveniências. Para esta avaliação, utilizou-se novamente uma planilha *online* da ferramenta *Google Drive*, cuja primeira coluna se refere somente as questões de competência(QC), na segunda coluna se encontram as respostas esperadas para cada QC, seguida de uma terceira coluna referente as consultas SPARQL utilizadas para responder determinada QC. Na quarta coluna se encontram os *links* para as *screenshots* das respostas obtidas através das consultas realizadas e, por último, uma coluna informando se foi possível ou não responder determinada questão, e obter a proveniência da resposta, através da consulta realizada.

Foram realizadas consultas para todas as QCs e em todas elas obteve-se resultados positivos, ou seja, as consultas retornaram todos os resultados esperados juntamente com suas proveniências. Com isso, conclui-se que as nanopublicações foram estruturadas de forma correta e que arquitetura nanoSCNAS pode ser reutilizada por pesquisadores para a criação de nanopublicações aplicada ao domínio de análise de redes de colaboração científica.

#### 4.4.1 Aplicação Semântica

Para poder demonstrar a viabilização de recuperação das nanopublicações, foi desenvolvida uma aplicação semântica capaz de recuperar as nanopublicações criadas na seção 4.3 e armazenadas numa *Triple Store* na seção 4.4. Para o desenvolvimento desta aplicação, optou-se por utilizar a linguagem python diante o back-end, devido a familiarização do autor com a linguagem. Os *frameworks Django* e *Bootstrap* foram utilizados para auxiliarem na criação do back-end e front-end da aplicação.

O *Django* é um *framework* de alto nível para desenvolvimento web, a partir da linguagem python ele permite a criação de aplicações através de uma forma limpa, ou seja, ele toma conta de toda parte pesada do desenvolvimento web (Ex.: o tratamento de requisições) para que seja possível focar no que realmente interessa.

Já o *Bootstrap* é um *framework* web com código-fonte aberto para desenvolvimento de front-end para aplicações web. Através do uso de HTML, CSS e JavaScript, baseado em modelos de design para a tipografia, visa melhorar a experiência do usuário através de uma interface amigável e responsiva. Diante o desenvolvimento da interface, o *framework* oferece em sua documentação vários componentes prontos que facilitam muito no desenvolvimento do front-end de uma aplicação.

Para conseguir realizar as requisições na Triple Store, foi utilizada a biblioteca SPARQLWrapper. Um *wrapper* escrito em linguagem *Python*, que permite executar remotamente consultas em um *endpoint*. A biblioteca converte os resultados obtidos de

uma consulta para um formato mais gerenciável, de forma que seja possível trabalhar facilmente com esses dados.

Visto que a aplicação tem como objetivo recuperar as nanopublicações criadas por este trabalho, a criação da interface e das requisições realizadas pelo sistema no banco de triplas foi em prol da recuperação de informações através de métricas estudadas em redes de colaboração científica. Como pode ser visto na figura 65, a interface da aplicação é composta por elementos encontrados na ontologia SCNAS que são apresentados para o usuário de forma mais amigável.

Figura 65 – Interface da aplicação semântica criada

**NanoSCNAS**

Knowledge Area   Periodic   Event   PGP   Geographic   Intitution ▾

<p><b>Network Metrics</b></p> <p>Density: <input type="checkbox"/></p> <p>Component: <input type="checkbox"/></p> <p>Average degree: <input type="checkbox"/></p>	<p><b>Actor Metrics</b></p> <p>Degree centrality: <input type="checkbox"/></p> <p>Closeness centrality: <input type="checkbox"/></p> <p>Betweenness centrality: <input type="checkbox"/></p>	<p><b>Other Information</b></p> <p>Region: <input type="checkbox"/></p> <p>Network time: <input type="checkbox"/></p> <p>Knowledge area: <input type="checkbox"/></p> <p>Actors number: <input type="checkbox"/></p> <p>Edges number: <input type="checkbox"/></p>
---	--	--

Fonte: Do Autor.

No topo da aplicação é possível encontrar os tipos de estudos definidos pela ontologia SCNAS, sendo eles: *Knowledge Area*, *Periodic*, *Event*, *Postgraduate (PGP)*, *Geographic* e *Institution*, que é dividido em *High Education* e *Research*. Logo abaixo são apresentadas algumas *checkboxs* referentes as métricas e informações apresentadas nos estudos. Dar *check* em uma *box* significa querer retornar a informação referente aquela *box*, por exemplo, dar *check* nas *box* referentes a *Degree Centrality* e *Closeness Centrality* informa para a aplicação que o usuário gostaria de retornar as Centralidades de Grau e Centralidades de Proximidade dos atores das redes de um determinado tipo de estudo, como mostra a figura 66.

Figura 66 – Exemplo de consulta através da aplicação

## NanoSCNAS

Knowledge Area   Periodic   Event   PGP   Geographic   Intitution ▾

**Network Metrics**

Density:

Component:

Average degree:

**Actor Metrics**

Degree centrality:

Closeness centrality:

Betweenness centrality:

**Other Information**

Region:

Network time:

Knowledge area:

Actors number:

Edges number:

Filter   Search

Quantidade de estudos retornados: 2

SCNAS	Actor	Degree Centrality	Closeness Centrality	DOI	PROV
<a href="#">Title</a>	Baisheng Jiang	6.0	1.4	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Baisheng Jiang	3.0	1.4	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Xin Xiang	4.0	2.0	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Weibing Zhang	5.0	1.0	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Gang Luo	5.0	1.64	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Lixin Yang	8.0	1.5	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Zhengfei Wang	5.0	1.0	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Fang Yang	5.0	1.7	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Bo Liu	5.0	1.91	<a href="#">link</a>	<a href="#">PROV</a>

Fonte: Do Autor.

É importante ressaltar que a aplicação trabalha com o uso da operação lógica “E” durante todas as consultas, ou seja, ao selecionar as métricas *Degree Centrality* e *Closeness Centrality*, a aplicação irá retornar somente os atores com que tiverem as duas centralidades, e não uma ou outra.

Observa-se também que o grupo de métricas “*Network Metrics*” e “*Actor Metrics*” permanecem imutáveis na interface, independente do tipo de estudo escolhido. Diferentemente do conjunto “*Other Information*” que agrupa informações específicas de cada tipo de rede, ou seja, muda de acordo com o tipo de estudo escolhido pelo usuário. Essa peculiaridade ocorre devido as regras de relacionamento adotadas para a *SCNAS Ontology*, que podem ser observadas com mais detalhes na seção 4.2.4.1.

Ao realizar uma última análise na figura 65, é possível perceber a presença dos botões “*Filter*” e “*Search*”. Clicar no botão *Search* quer dizer para a aplicação que uma

requisição a *Triple Store* devem ser enviada, ou seja, uma busca deve ser feita. Já ao clicar no botão “*Filter*” a aplicação abre um *modal* com vários *text fields*, cada uma relacionada a uma métrica ou informação. Através dessas *text fields* é possível especificar intervalos de métricas de redes e atores, ou especificar dados relacionados ao grupo de informações “*Other Information*”, como o nome da região de uma rede, por exemplo.

Figura 67 – Exemplo de consulta através da aplicação

Métricas de Rede	Métricas de Ator	Outras Informações
Density greater than: <input type="text"/>	Degree centrality greater than: <input type="text" value="5"/>	Knowledge area name: <input type="text" value="health"/>
Density less than: <input type="text"/>	Degree centrality less than: <input type="text" value="10"/>	Region name: <input type="text" value="China"/>
Component greater than: <input type="text"/>	Closeness centrality greater than: <input type="text"/>	Network time beginning: <input type="text"/>
Component less than: <input type="text"/>	Closeness centrality less than: <input type="text"/>	Network time end: <input type="text"/>
Average degree greater than: <input type="text"/>	Betweenness centrality greater than: <input type="text"/>	
Average degree less than: <input type="text"/>	Betweenness centrality less than: <input type="text"/>	

Close

Fonte: Do Autor.

A figura 67 exemplifica o uso do “*Filter*” em uma consulta, onde o usuário quer recuperar alguma informação dos estudos da área de conhecimento *health*, cujas redes são localizadas na China e que a Centralidade de Grau dos atores recuperados devem estar no intervalo entre 5 e 10. A figura 69 explicita os resultados obtidos através do uso do filtro especificado.

Ao analisar a figura 69, é possível perceber que os resultados são organizados em colunas. É importante destacar que as colunas SCNAS, DOI e PROV são fixas, independente da consulta realizada. A coluna SCNAS informa os nomes dos estudos relacionados as informações recuperadas. Na coluna DOI são encontrados os links dos DOI’s dos estudos vinculados as informações recuperadas que, ao clicar, é direcionado para o endereço do documento. A coluna PROV lista *links* que mostram para o usuário a proviniência das informações recuperadas. Observa-se que as informações destacadas em azul são *tooltips*, ou seja, para que a informação possa ser mostrada para o usuário é necessário que o mesmo sobreponha o *mouse* no *tooltip* que deseja. O uso de *tooltips* foi adotado devido a longas *strings* que deixavam a área de resultados muito poluída, como os títulos dos estudos, por exemplo.

Figura 68 – Exemplo de consulta com uso de filtros através da aplicação

Quantidade de estudos retornados: 1

SCNAS	Actor	Degree Centrality	Closeness Centrality	Location	Beginning	End	Knowledge Area	DOI	PROV
<a href="#">Title</a>	Lixin Yang	8.0	1.5	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Baisheng Jiang	6.0	1.4	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Shaohui Chen	6.0	1.0	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Pansheng Li	6.0	1.33	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Hunan	9.0	1.91	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Chinese Academy of Social Sciences	7.0	6.62	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Sun Yat-Sen University	7.0	6.66	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	China University of Political Science and Law	6.0	9.13	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Nanjing Medical University	9.0	6.55	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>
<a href="#">Title</a>	Huazhong University of Science and Technology	9.0	6.69	China	1978	2014	<a href="#">Domain</a>	<a href="#">link</a>	<a href="#">PROV</a>

Fonte: Do Autor.

Figura 69 – Exemplo de tooltip na coluna de SCNAS



Fonte: Do Autor.

Após a implantação da aplicação, foi necessário avaliá-la para descobrir se de fato o uso de nanopublicações é viável. A aplicação, em seu estado atual, se encontra disponível em um repositório<sup>30</sup> *online* aberto, para que qualquer pessoa tenha acesso. A seção 4.5 detalha o processo de avaliação da aplicação aqui criada.

#### 4.5 Avaliação da Aplicação

Com o propósito de avaliar a viabilidade da recuperação das nanopublicações através da aplicação desenvolvida, uma especialista no domínio de estudos de análise de redes de colaboração científica foi convidada para realizar os testes. Os testes consistiram em realizar consultas através da interface da aplicação, que deveriam retornar tanto as respostas das consultas, quanto as proveniências das respostas. As consultas realizadas

<sup>30</sup> <<https://github.com/Laltany/AplicacaoNanoSCNAS>>

tiveram como base as questões de competência criadas na seção 4.2.1.6, onde cada consulta deveria responder uma determinada questão e a resposta obtida deveria ser comparada com a resposta esperada. Dessa forma, foi possível julgar se a recuperação estava correta, ou não.

Inicialmente a especialista pôde fazer consultas aleatórias na aplicação para que pudesse se acostumar e entender o funcionamento do sistema, pois o objetivo desta avaliação não era avaliar a intuitividade da interface. Após entender como a aplicação funcionava, foi passado para a especialista uma lista de questões que deveriam ser respondidas através da aplicação. Ao obter as afirmações retornadas de cada consulta, a especialista conferiu a proveniência de algumas de algumas delas para conferir se estavam corretas.

Em paralelo as consultas realizadas pela especialista, era registrado em uma planilha <sup>31</sup> online se as respostas retornadas estavam corretas ou não.

Ao final dos testes a especialista fez suas considerações e passou algumas dicas de melhoramento que podem ser feitas futuramente na interface da aplicação.

#### 4.5.1 Resultados da Avaliação

Antes de iniciar a seção de resultados, é necessário lembrar que não foram pautadas as considerações relacionadas somente a interface da aplicação, visto que o principal objetivo da avaliação era obter resultados sobre as consultas realizadas no sistema. Diante disso, o sucesso ou não da realização das consultas realizadas pela avaliadora foram registradas em uma planilha *online*.

Através dos testes realizados, criou-se a tabela 11 que explicita os resultados diante a avaliação da aplicação e das nanopublicações retornadas.

Tabela 11 – Tabela com os resultados obtidos na avaliação da aplicação

Questões respondidas corretamente	100%
Proveniências corretas	100%
Relevância da Aplicação	8

Fonte: Do Autor.

Como pode ser observado na tabela 11, todas as questões foram respondidas com sucesso, de forma que todos os resultados retornados eram equivalentes aos resultados esperados. Diante as proveniências analisadas todas informações estavam corretas, possibilitando que a avaliadora pudesse analisar informações importantes sobre os estudos que proveram a informação. Caso ainda houvessem dúvidas, bastava clicar no link do endereço DOI do estudo para saná-las através da publicação completa.

<sup>31</sup> <<http://tiny.cc/a73ggz>>

Referente a relevância da aplicação, foi constatado que a mesma cumpre com o prometido. Isto é, as informações e métricas trabalhadas em estudos de análise de redes de colaboração científicas são facilmente recuperadas. Porém, ainda existe a necessidade de fazer melhorias no front-end da aplicação diante a apresentação das informações recuperadas. Também foi constatado que a aplicação pode sim ajudar e agilizar o desenvolvimento de pesquisas científicas, através da curadoria ágil que o sistema oferece diante a recuperação das informações contidas nos documentos de pesquisas.

#### 4.5.2 Lições Aprendidas

A criação de uma ontologia para a representação dos conceitos do domínio de Análise de Redes Sociais, especificamente redes de colaboração científica, mostrou que a descrição de um domínio de conhecimento não é tão fácil quanto parece, por mais que já se tenha um breve conhecimento sobre ele.

Todo o processo de criação com a metodologia Neon demandou um tempo considerável, e no entanto ainda assim destaca-se a possibilidade de não se ter coberto totalmente todas as métricas de análise de redes sociais que podem ser utilizadas em estudos como uma ameaça à validade da pesquisa. Porém, mesmo com o escopo de métricas limitado, foi possível criar uma ontologia consistente, com 100% de cobertura de resposta para as questões de competência criadas durante as especificações dos requisitos da ontologia. Durante a extração das informações necessárias para a criação das nanopublicações, foi notório que este é um dos principais problemas da abordagem, assim como afirmam Do e Mobley (2015), pois é um processo que demanda muito tempo e, em cenários onde se têm um grande volume de informações a serem extraídas, pode haver uma sobrecarga de informações incontrolável, fazendo com que as nanopublicações percam a sua confiabilidade.

A aplicação produzida por este trabalho tem como objetivo provar a viabilidade da recuperação das nanopublicações. Porém, não reflete um real cenário de utilização, pois não pode ser comparado a repositórios com centenas de documentos e informações. Para que a proposta tenha relevância prática é necessário que o volume de nanopublicações seja maior. No entanto, destaca-se que a aplicação se mostrou eficaz ao se tratar da apresentação das afirmações obtidas através dos documentos, assegurando a confiabilidade do uso das nanopublicações.





## 5 CONSIDERAÇÕES FINAIS

Um excesso de informações não estruturadas estão disponíveis em repositórios científicos, ocasionando problemas para recuperá-las, extraí-las e reutilizá-las. Para facilitar a recuperação de informações, foi proposto o uso de nanopublicação, que ao utilizar tecnologias semânticas, evidencia dados importantes, facilitando a interpretação computacional e o reuso de afirmações contidas em documentos científicos. Durante a realização do presente trabalho, foi escolhido o domínio de análise de redes de colaboração científicas para a criação e recuperação de nanopublicações. Objetivou-se, através da introdução da nanopublicação ao domínio, mostrar os benefícios que a abordagem pode trazer para os pesquisadores.

Para o cumprimento do propósito do trabalho, foi necessário a criação de uma ontologia capaz de identificar os resultados de métricas e informações comumente presentes nos documentos científicos do domínio de Análise de Redes Sociais. Para o desenvolvimento da nova ontologia, foi seguido uma série de atividades metodológicas propostas pelo grupo NeOn, o que garantiu a criação de uma ontologia consistente. Tornando-se uma ontologia adequada para ser utilizada em outros domínios de análise de rede, para além das redes de colaboração científica.

O uso da ontologia criada, juntamente a um conjunto de ontologias que descrevem a proveniência de dados, permitiram a extração de informações importantes e necessárias para a estruturação das nanopublicações do domínio, cuja arquitetura foi denominada como nanoSCNAS.

Com o objetivo de avaliar e demonstrar a viabilidade do uso das nanopublicações, criou-se uma aplicação capaz de recuperar as nanopublicações criadas por este trabalho. Onde, através de uma interface amigável, uma especialista no domínio pôde realizar testes e avaliar a aplicação e as nanopublicações geradas. Concluiu-se através desta avaliação, que o uso da aplicação traz benefícios para a curadoria de informações contidas em documentos científicos pois, através do uso de nanopublicações, mostrou-se um ganho na recuperação de informações, uma vez que a nano permite uma recuperação em um nível de granularidade menor, que os sistemas de recuperação tradicionais não permitem. Agilizando o processo de pesquisa por informações confiáveis, visto que as afirmações retornadas são acompanhadas de suas proveniências.

Por fim, observa-se que a nanopublicação ainda é um abordagem relativamente nova e pouco usada fora do seu domínio de criação. Por isso, é necessário disseminar o conhecimento sobre esta estrutura que tende a trazer muitos benefícios as pesquisas científicas, principalmente para aquelas que necessitam demandar muito tempo para realizar buscas, já que pode agilizar este processo. A nanopublicação pode ser estendida de várias maneiras, desde que se mantenha a sua estrutura mínima. Desta forma, através do uso de ontologias, é possível reestruturá-la para  $n$  domínios e de  $n$  maneiras, tornando a nanopublicação uma abordagem que ainda precisa ser muito explorada. Este trabalho teve como

objetivo criar nanopublicações do domínio de estudos de análise de redes de colaboração científica e entende-se que, através dos resultados obtidos, seja possível dar apoio a outros pesquisadores seja através da aplicação criada, ou através de sua metodologia.

Assume-se como trabalhos futuros o desenvolvimento de um extrator de dados para criação de nanopublicações no domínio de Análise de Redes Sociais. Para que se possa dar destaque a real colaboração que a nanopublicação traz ao meio científico, pretende-se fazer comparação da abordagem com outras abordagens de recuperação de informação, apontando as vantagens e desvantagens de cada uma. Têm-se também como objetivo implementar uma aplicação que permita a criação de nanopublicações de forma colaborativa, onde os próprios autores e usuários interessados poderão para criar novas nanopublicações, alimentando a base de conhecimento deste domínio.

## REFERÊNCIAS

- ALVAREZ, G. R.; VANZ, S. A. de S. Indicadores bibliométricos de colaboração científica institucional na física de altas energias brasileira (1983-2013). **Revista Científica Guillermo de Ockham**, v. 15, n. 2, 2017. Citado na página 59.
- ASSANTE, M. et al. Are scientific data repositories coping with research data publishing? **Data Science Journal**, Ubiquity Press, v. 15, 2016. Citado na página 25.
- BARABÁSI, A.-L. et al. Evolution of the social network of scientific collaborations. **Physica A: Statistical mechanics and its applications**, Elsevier, v. 311, n. 3-4, p. 590–614, 2002. Citado na página 27.
- BECHHOFFER, S. et al. Owl web ontology language reference. **W3C recommendation**, v. 10, n. 02, 2004. Citado na página 47.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific american**, JSTOR, v. 284, n. 5, p. 34–43, 2001. Citado 3 vezes nas páginas 31, 32 e 49.
- BORDIN, A. S. et al. Framework baseado em conhecimento para análise de rede de colaboração científica. 2015. Citado 4 vezes nas páginas 90, 91, 92 e 94.
- BORST, W. N. Construction of engineering ontologies for knowledge sharing and reuse. 1999. Citado na página 36.
- BORTOLATO, F.; PAULO, S. Instituto de Pesquisas Tecnológicas do Estado de São Paulo Ligando Dados Governamentais Abertos: uma ontologia do Processo Legislativo de São Paulo. 2014. Citado 4 vezes nas páginas 40, 76, 108 e 111.
- CATARINO, M. E.; SOUZA, T. Batista de. A representação descritiva no contexto da web semântica. **Transinformação**, Pontifícia Universidade Católica de Campinas, v. 24, n. 2, 2012. Citado na página 33.
- CONSORTIUM, W. W. W. et al. Owl 2 web ontology language document overview. World Wide Web Consortium, 2012. Citado 2 vezes nas páginas 45 e 46.
- CONSORTIUM, W. W. W. et al. Rdf 1.1 concepts and abstract syntax. World Wide Web Consortium, 2014. Citado na página 33.
- DIGIAMPIETRI, L. et al. Análise da rede de relacionamentos dos doutores brasileiros. In: **CSBC 2014 - BreSci**. Brasília - DF: [s.n.], 2014. Citado na página 58.
- DO, L.; MOBLEY, W. **Single Figure Publications: Towards a novel alternative format for scholarly communication**. F1000Res, 2015. Disponível em: <<https://doi.org/10.12688/f1000research.6742.1>>. Citado na página 125.
- DZIEKANIAK, G. V.; KIRINUS, J. B. Web semântica 10.5007/1518-2924.2004v9n18p20. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 9, n. 18, p. 20–39, 2004. Citado 3 vezes nas páginas 31, 32 e 33.
- D'AQUIN, M.; NOY, N. F. Where to publish and find ontologies? a survey of ontology libraries. **Journal of Web Semantics**, Elsevier, v. 11, p. 96–111, 2012. Citado 2 vezes nas páginas 90 e 97.

- ERETEO, G. et al. Analysis of a Real Online Social Network using Semantic Web Frameworks. In: **ISWC 2009**. Washington, United States: [s.n.], 2009. v. 5823/2009, p. 180–195. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00562035>>. Citado 3 vezes nas páginas 90, 91 e 94.
- FARINA, M. C. Análise de redes sociais. **Programa de Pós-Graduação em Comunicação, Universidade Municipal de São Caetano do Sul (PPGCOM-USCS)**, 2004. Citado na página 56.
- FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. American Association for Artificial Intelligence, 1997. Citado 2 vezes nas páginas 44 e 100.
- FREITAS, F. L. G. de. Ontologias e a web semântica. **Jornada de Mini-Cursos em Inteligência Artificial, SBC**, v. 8, 2003. Citado 5 vezes nas páginas 34, 35, 46, 47 e 48.
- GOLDEN, P.; SHAW, R. **Nanopublication beyond the Sciences**. [S.l.], 2015. Citado 4 vezes nas páginas 25, 26, 52 e 53.
- GONZÁLEZ, A. R. et al. Nanopublishing clinical diagnoses: tracking diagnostic knowledge base content and utilization. In: IEEE. **2014 IEEE 27th International Symposium on Computer-Based Medical Systems**. [S.l.], 2014. p. 335–340. Citado na página 70.
- GROTH, P.; GIBSON, A.; VELTEROP, J. The anatomy of a nanopublication. **Information Services & Use**, IOS Press, v. 30, n. 1-2, p. 51–56, 2010. Citado 2 vezes nas páginas 26 e 53.
- GROTH, P.; GIBSON, A.; VELTEROP, J. The anatomy of a nanopublication. **Information Services & Use**, IOS Press, v. 30, n. 1-2, p. 51–56, 2010. Citado na página 27.
- GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing? **International journal of human-computer studies**, Elsevier, v. 43, n. 5-6, p. 907–928, 1995. Citado na página 36.
- GRUNINGER, M.; FOX, M. S. Methodology for the Design and Evaluation of Ontologies. **Ijcai**, v. 95, p. 1–10, 1995. Citado 2 vezes nas páginas 41 e 78.
- GUIZZARDI, G. **Ontological Foundations for Structural Conceptual Models**. Tese (Doutorado) — Centre for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, 2005. Disponível em: <<http://www.loa.istc.cnr.it/Guizzardi/SELMAS-CR.pdf>>. Acesso em: 3.7.2011. Citado 2 vezes nas páginas 36 e 37.
- HARRIS, S.; SEABORNE, A.; PRUD’HOMMEAUX, E. Sparql 1.1 query language. **W3C recommendation**, v. 21, n. 10, 2013. Citado 3 vezes nas páginas 50, 51 e 52.
- HAYASHI, M.; HAYASHI, C. R. M.; LIMA, M. Y. d. Análise de redes de coautoria na produção científica em educação especial. **Liinc em revista**, UFRJ/IBICT; Laboratório Interdisciplinar em Informação e Conhecimento Rio de Janeiro, v. 4, n. 1, p. 84–103, 2008. Citado na página 59.

- HORRIDGE, M. **A Practical Guide to Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.3**. Retrieved March 10, 2013. 2011. Citado 4 vezes nas páginas 46, 47, 48 e 49.
- HORST, E. van der et al. Finding novel associations across domains using linked data: a case study on genetic variants disrupting transcription start sites. In: **SWAT4LS**. [S.l.: s.n.], 2015. p. 2–12. Citado na página 70.
- HU, H.; LIU, D.; WANG, S. Web ontology language owl [j]. **Computer Engineering**, v. 12, p. 000, 2004. Citado na página 46.
- JUNIOR, J. L. **WEB SEMÂNTICA, ONTOLOGIAS E LINGUAGEM OWL: UM ESTUDO DE CASO SOBRE A FERRAMENTA PROTÉGÉ**. 2012. Disponível em: <<https://cepein.femanet.com.br/BDigital/arqTccs/0911270541.pdf>>. Citado na página 47.
- KLEIN, M. Xml, rdf, and relatives. **IEEE Intelligent Systems**, IEEE, v. 16, n. 2, p. 26–28, 2001. Citado na página 34.
- KOIVUNEN, M.-R.; MILLER, E. W3c semantic web activity. **Semantic Web Kick-Off in Finland**, v. 2, p. 27–44, 2001. Citado na página 48.
- KOSORUKOFF, A. **Social network analysis: theory and applications**. [S.l.]: Passmore, D. L, 2011. Citado na página 57.
- KUHN, T. nanopub-java: A java library for nanopublications. **arXiv preprint arXiv:1508.04977**, 2015. Citado na página 70.
- KUHN, T. et al. Broadening the scope of nanopublications. In: SPRINGER. **Extended Semantic Web Conference**. [S.l.], 2013. p. 487–501. Citado na página 26.
- LASSILA, O.; SWICK, R. R. Resource description framework (rdf) model and syntax specification. 1999. Citado na página 33.
- LIMA, J. C. de; CARVALHO, C. L. de. **Ontologias-owl (web ontology language)**. [S.l.], 2005. Citado 4 vezes nas páginas 31, 32, 37 e 45.
- LIMA, J. C. de; CARVALHO, C. L. de. **Resource description framework (rdf)**. [S.l.], 2005. Citado na página 36.
- LIPANI, A. et al. Extracting nanopublications from ir papers. In: SPRINGER. **Information Retrieval Facility Conference**. [S.l.], 2014. p. 53–62. Citado na página 70.
- LIPANI, A. et al. An information retrieval ontology for information retrieval nanopublications. In: SPRINGER. **International Conference of the Cross-Language Evaluation Forum for European Languages**. [S.l.], 2014. p. 44–49. Citado na página 26.
- LOPES, R. P.; SERNADELA, P.; OLIVEIRA, J. L. Exploring nanopublishing with coeus. In: **SWAT4LS**. [S.l.: s.n.], 2013. Citado na página 70.
- MAIATSKY, M. et al. Vicoglossia: Annotatable and commentable library as a bridge between reader and scholar (a proof of concept study: Early soviet philological culture). **Umanistica Digitale**, v. 2, n. 2, 2018. Citado na página 70.

- MARCIAL, L. H.; HEMMINGER, B. M. Scientific data repositories on the web: An initial survey. **Journal of the American Society for Information Science and Technology**, Wiley Online Library, v. 61, n. 10, p. 2029–2048, 2010. Citado na página 25.
- MARTELETO, R. M.; TOMAÉL, M. I. A metodologia de análise de redes sociais (ars). **Métodos qualitativos de pesquisa em ciência da informação**. São Paulo: Polis, p. 81–100, 2005. Citado na página 57.
- MARTINS, D. L. **Análise de redes sociais de colaboração científica no ambiente de uma federação de bibliotecas digitais**. Tese (Doutorado) — Universidade de São Paulo, 2012. Citado na página 58.
- MCCUSKER, J. et al. Developing scientific knowledge graphs using whyis. In: **SemSci@ISWC**. [S.l.: s.n.], 2018. p. 52–58. Citado na página 70.
- MCCUSKER, J. et al. A nanopublication framework for biological networks using cytoscape.js. In: CITESEER. **ICBO**. [S.l.], 2014. p. 90–92. Citado na página 70.
- MCGUINNESS, D. L.; HARMELEN, F. V. et al. Owl web ontology language overview. **W3C recommendation**, v. 10, n. 10, p. 2004, 2004. Citado na página 45.
- MEADOWS, A. J.; LEMOS, A. A. B. de L. **A comunicação científica**. [S.l.]: Briquet de Lemos/livros, 1999. Citado 2 vezes nas páginas 27 e 58.
- MENDONÇA, F. C. de et al. Scientific collaboration networks of the academic brazilian community of hci. In: ACM. **Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems**. [S.l.], 2018. p. 36. Citado na página 59.
- MONS, B.; VELTEROP, J. Nano-publication in the e-science era. In: **Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)**. [S.l.: s.n.], 2009. p. 14–15. Citado 3 vezes nas páginas 25, 26 e 52.
- MOURA, M. A. Informação, ferramentas ontológicas e redes sociais: a interoperabilidade na construção de tesauros e ontologias. In: SERVICIO DE PUBLICACIONES. **Nuevas perspectivas para la difusión y organización del conocimiento: actas del congreso**. [S.l.], 2009. p. 589–604. Citado 2 vezes nas páginas 26 e 52.
- NAZOU. **NAZOU - tools for acquisition, organization and maintenance of knowledge in an environment of heterogeneous information resources**. Disponível em: <<http://nazou.fiit.stuba.sk/home/?page=about>>. Acesso em: 11/05/2019. Citado na página 98.
- NEWMAN, M. **Networks: An Introduction**. OUP Oxford, 2010. ISBN 9780199206650. Disponível em: <<https://books.google.com.br/books?id=q7HVtpYVfC0C>>. Citado na página 27.
- NOY, N. F.; MCGUINNESS, D. L. et al. **Ontology development 101: A guide to creating your first ontology**. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA, 2001. Citado na página 37.

- OLIVEIRA, É. B. P. M. Periódicos científicos eletrônicos: definições e histórico. **Informação & Sociedade**, Universidade Federal da Paraíba-Programa de Pós-Graduação em Ciência da . . . , v. 18, n. 2, 2008. Citado na página 25.
- PAULA, L. T. de; MOURA, M. A. Nanopublicações e indexação: processos e relações. **Informação & Informação**, v. 20, n. 1, p. 137–161, 2015. Citado na página 27.
- PETERSEN, K. et al. Systematic mapping studies in software engineering. In: **Ease**. [S.l.: s.n.], 2008. v. 8, p. 68–77. Citado 3 vezes nas páginas 61, 65 e 67.
- QUERALT-ROSINACH, N. et al. Publishing disgenet as nanopublications. **Semantic Web**, IOS Press, v. 7, n. 5, p. 519–528, 2016. Citado na página 70.
- REZENDE, B.; BAX, M. Projecto indexa: ferramenta de auxílio à divulgação de informações na web. **Universidade Federal de Minas Gerais**, 2002. Citado 2 vezes nas páginas 34 e 37.
- SCHENKEL, M. B. d. C. Compartilhamento do conhecimento científico em instituição estadual de ensino superior: o caso do centro de ciências humanas e da educação da udesc. **Universidade Federal de Santa Catarina**, 2008. Citado na página 25.
- SCHNEIDER, J. et al. Formalizing knowledge and evidence about potential drug-drug interactions. In: **BDM2I@ ISWC**. [S.l.: s.n.], 2015. Citado na página 70.
- SCHULTES, E. et al. The open phacts nanopublications guidelines v1. 8.1. **EU Innovative Medicines Initiative-Open PHACTS Project**, 2012. Citado 3 vezes nas páginas 53, 54 e 55.
- SERNADELA, P.; OLIVEIRA, J. L. Automated nanopublications generation from biomedical literature. In: **IEEE. 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)**. [S.l.], 2017. p. 1–4. Citado na página 70.
- SERNADELA, P.; OLIVEIRA, J. L. Coeus 2.0: automated platform to integrate and publish biomedical data as nanopublications. **IET Software**, IET, v. 12, n. 2, p. 136–141, 2017. Citado na página 70.
- SONNENWALD, D. H. Scientific collaboration. **Annual review of information science and technology**, Wiley Online Library, v. 41, n. 1, p. 643–681, 2007. Citado 2 vezes nas páginas 27 e 58.
- SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, SciELO Brasil, v. 33, n. 1, 2004. Citado 2 vezes nas páginas 33 e 57.
- STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: principles and methods. **Data & knowledge engineering**, Elsevier, v. 25, n. 1-2, p. 161–197, 1998. Citado na página 36.
- SUAREZ, M. d. C. NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse. **Landscape and Urban Planning**, v. 102, n. 4, p. 268, 2010. ISSN 01692046. Citado 20 vezes nas páginas 37, 38, 40, 41, 42, 43, 44, 75, 76, 81, 83, 84, 85, 86, 87, 88, 90, 96, 97 e 108.

WASSERMAN, S.; FAUST, K. **Social network analysis—methods and applications, revised, reprinted edn.** [S.l.]: Cambridge University Press, Cambridge, 1999. Citado na página 27.

WELTY, C.; MCGUINNESS, D. L.; SMITH, M. K. Owl web ontology language guide. **W3C recommendation, W3C (February 2004) <http://www.w3.org/TR/2004/REC-owl-guide-20040210>**, 2004. Citado 2 vezes nas páginas 48 e 49.

YU, E. **Social Network Analysis Applied to Ontology 3D Visualization.** Tese (Doutorado) — Miami University, 2008. Citado 3 vezes nas páginas 90, 91 e 93.

ZEMMOUCHI-GHOMARI, L.; GHOMARI, A. R. Translating natural language competency questions into sparqlqueries: a case study. In: **The First International Conference on Building and Exploring Web Based Environments.** [S.l.: s.n.], 2013. p. 81–86. Citado 2 vezes nas páginas 108 e 109.



**ÍNDICE**

nanoSCNAS, 113

OWL, 15, 21, 45–49, 55, 77, 79, 91, 92,  
98, 99, 107

RDF, 15, 33–36, 45–47, 49–52, 54, 55, 61,  
62, 71–73, 116, 118

SCNAS Ontology, 75

SPARQL, 33, 49–51, 110, 111, 119