

UNIVERSIDADE FEDERAL DE ALAGOAS  
INSTITUTO DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL DO  
CONHECIMENTO

JUDSON MELO BANDEIRA

**FOCA: Uma Metodologia que utiliza princípios da Representação do  
Conhecimento para Avaliação de Ontologias**

Maceió

2015

JUDSON MELO BANDEIRA

**FOCA: Uma Metodologia que utiliza princípios da Representação do  
Conhecimento para Avaliação de Ontologias**

Dissertação submetida à Coordenação do Mestrado em Modelagem Computacional do Conhecimento da Universidade Federal de Alagoas como parte dos requisitos necessários para obtenção do grau de Mestre em Modelagem Computacional do Conhecimento.

Orientadores: Prof. Dr. Ig Ibert Bittencourt Santana Pinto e Prof. Dr. Patrícia Leone Espinheira Ospina

Maceió

2015

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**  
**Bibliotecária Responsável: Helena Cristina Pimentel do Vale**

B214f    Bandeira, Judson Melo.  
          FOCA: uma metodologia que utiliza princípios da representação do conhecimento para avaliação de ontologias / Judson Melo Bandeira. – 2015.  
          101 f. : il.

Orientador: Ig Ibert Bittencourt Santana Pinto.  
Dissertação (Mestrado em Modelagem Computacional do Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2015.

Bibliografia: f. 78-81.  
Apêndices: f. 82-101.

1. Usabilidade. 2. Ambientes virtuais de aprendizagem.  
3. Ontologia. I. Título.

CDU: 004.9:371.33



Membros da Comissão Julgadora da Dissertação de Mestrado de Judson Melo Bandeira, intitulada: “FOCA: Uma Metodologia que utiliza princípios da Representação do Conhecimento para Avaliação de ontologias”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas em três de agosto de dois mil e quinze, às 9h00min, no auditório do Instituto de Computação da Universidade Federal de Alagoas.

### COMISSÃO JULGADORA



---

**Prof. Dr. Ig Ibert Bittencourt Santana Pinto**

UFAL – Instituto de Computação

Orientador

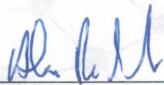


---

**Prof. Dr. Patrícia Leone Espinheira Ospina**

UFPE – Instituto de Matemática e Estatística

Orientador

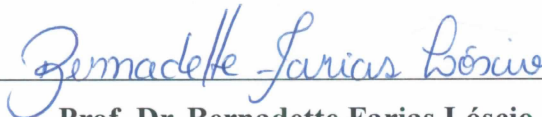


---

**Prof. Dr. Alan Pedro da Silva**

UFAL – Instituto de Computação

Examinador



---

**Prof. Dr. Bernadette Farias Lóscio**

UFPE – Centro de Informática

Examinadora

Agosto, 2015.

Dedico essa dissertação aos meus pais **José Arnaldo** e **Maricélia** e meu irmão **Udson**, por entender a importância de um curso de mestrado em minha vida e pelo apoio, tanto financeiro como emocional. Aos meus amigos de laboratório, especialmente **Williams** e **Armando**, por estarem do meu lado todos os dias, compartilhando pesquisas e distrações. Ao meu orientador, **Ig Ibert**, por me ter feito evoluir tanto pessoalmente quanto profissionalmente e a todos que direta ou indiretamente estiveram envolvidos neste processo... Sem vocês, eu não teria conseguido.

## AGRADECIMENTOS

Ao meu orientador, **Ig Ibert** por acreditar em mim desde quando entrei no laboratório de pesquisa, enxergando em mim um grande potencial, o que me fez encarar os desafios com mais coragem;

À minha mãe **Maricélia**, que mesmo com sua vida cansativa, vinha me visitar aqui em Maceió para cuidar de mim e da casa nos momentos de stress, além do enorme amor que me transmitia;

Ao meu pai **José Arnaldo**, por acreditar em mim desde o começo e pelo apoio financeiro e emocional, além das muitas visitas que me fez para aproveitar a família;

Ao meu irmão **Udson**, que me entendia nos momentos de stress e pelo carinho que me transmitia todos os dias, além do que dividia as tarefas de casa de modo que os dois pudessem estudar em perfeita harmonia;

Aos meus colegas de laboratório, especialmente **Williams** e **Armando**, por todos estes anos de projeto juntos, além de estarmos juntos no dia que mudou as nossas vidas quando vencemos o startup weekend e pela “pressãozinha” que faziam nos meus momentos de preguiça;

Aos meus outros amigos de laboratório, **Josmário, Glauber, João, Danila, Ranilson, André, Maria, Ezequiel, Denys, Thiago, Thyago, Endhe, Olavo** e os restantes membros do NEES, que formaram uma linda família dentro de um laboratório não só com objetivos de pesquisa, mas com ensinamentos para a vida.

À minha namorada **Ana Clara**, que entendeu minha rotina cansativa e me acalmou nas horas mais estressantes.

À professora **Patrícia Ospina**, que me ajudou bastante na reta final do mestrado.

A todos os outros que estiveram presentes nessa jornada.

**UM MUITO OBRIGADO!**

## RESUMO

Ontologia é um termo originado na Filosofia através dos termos gregos “ontos” (ser) e “logos” (estudo), que significa o *estudo do ser enquanto ser*. Este termo foi trazido para a Ciência da Computação como sendo um artefato computacional que representa o conhecimento de determinado universo do discurso. Uma das formas de se fazer isto é através dos cinco papéis da representação do conhecimento: substituto, compromisso ontológico, raciocínio inteligente, eficiência computacional e expressão humana. Em ciência da computação, as ontologias são aplicadas em muitas áreas como, por exemplo, em Engenharia de Requisitos, Banco de Dados, Web Semântica e Dados Conectados. A subárea responsável pelo processo de construção de ontologias é a engenharia de ontologia, que conta com aproximadamente vinte metodologias. Entretanto, conhecer bem o domínio, definir corretamente os conceitos envolvidos nele e ao mesmo tempo cumprir os papéis da representação do conhecimento tornam modelar ontologia uma tarefa bastante árdua. Desta forma, quanto maior a complexidade da ontologia, mais a sua qualidade tende a cair, reduzindo o seu potencial de reuso. Por consequência, o processo de avaliação da qualidade de uma ontologia se torna uma tarefa fundamental. A literatura da área propôs diferentes abordagens para avaliação de ontologias, porém as mesmas são baseadas em métricas de qualidade que não guiam os avaliadores na tomada de decisão sobre a qualidade da ontologia. Como resultado, o estabelecimento da qualidade da ontologia se torna uma tarefa mais dependente da experiência do avaliador e menos dependente de critérios de representação de conhecimento. Uma forma de simplificar este processo é guiar o avaliador sobre quais critérios estão relacionados com cada papel da representação do conhecimento. Este trabalho estabelece uma metodologia para avaliação de ontologias, a qual tem como base uma correspondência entre os papéis da representação do conhecimento com os principais critérios de qualidade para a avaliação de ontologia. Esta correspondência é construída através da abordagem Goal, Question, Metric. Este método consiste em três fases: Verificação do Tipo da Ontologia; Verificação das Questões e Verificação da Qualidade da ontologia. Este método foi validado através da realização de um experimento que objetiva mostrar que a metodologia conduz avaliadores com diferentes experiências em ontologia para o mesmo resultado de avaliação. Como resultado do experimento, foi possível criar uma fórmula para escores global e parcial para a qualidade da ontologia, validando-a.

**Palavras-chave: Ontologias, Avaliação de Ontologia, Representação do Conhecimento.**

## ABSTRACT

Ontology is a term originated in Philosophy through the Greek terms "ontos" (being) and "logos" (study), which means the *study of being qua being*. This term was brought to Computer Science as a computational artifact that represents knowledge of a certain universe of discourse. One way to do this is through the five roles of knowledge representation: substitute, ontological commitment, intelligent reasoning, computational efficiency and human expression. In computer science, ontologies are applied in many areas such as Requirements Engineering, Database, Semantic Web and Linked Data. The subarea responsible for ontology building process is ontology engineering, which has about twenty methodologies. However, knowing well the domain, defining the concepts involved in it correctly and at the same time comply the knowledge representation roles make ontology modeling an arduous task. Thus, how bigger is the complexity of ontology, the more its quality tends to decrease, reducing the reuse's potential. Consequently, the quality of the evaluation process of an ontology becomes a fundamental task. The literature of this area proposed different approaches for ontology evaluation, but they are based on quality metrics that do not guide the evaluators in decision making about the quality of ontology. As a result, the establishment of ontology quality becomes more dependent on the experience of the evaluator and less dependent on the knowledge representation criteria. One way to simplify this process is to guide the evaluator about what criteria are related to each role of knowledge representation. This work establishes a methodology for ontology evaluation, which is based on a match between the roles of knowledge representation with the main quality criteria for the ontology. This match is built by Goal, Question, Metric approach. This methodology consists of three phases: Ontology Type Verification; Questions Verification and ontology Quality Verification. This methodology was validated by conducting an experiment that aims to show that the methodology leads evaluators with different experiences in ontology to the same evaluation result. As a result of the experiment, was possible create a formula for global and partial scores for the quality of ontology, validating it.

**Keywords: Ontologies, Ontology Evaluation, Knowledge Representation.**



## LISTA DE FIGURAS

Figura 1. Utilização de metodologias pelos ontologistas.....	14
Figura 2. Problema da intersecção em ontologias mal modeladas.....	15
Figura 3. Definição de ontologia por Guarino.....	21
Figura 4. Estrutura da Web Semântica.....	23
Figura 5. Classificação 5-estrelas.....	25
Figura 6. Representação da abordagem GQM.....	28
Figura 7. As quatro fases do paradigma GQM.....	29
Figura 8. Metodologia proposta.....	37
Figura 9. Correspondência para o papel de Substituto.....	41
Figura 10. Correspondência para o papel de Compromisso Ontológico.....	41
Figura 11. Correspondência para o papel de Raciocínio Inteligente.....	41
Figura 12. Correspondência para o papel de Computação Eficiente.....	42
Figura 13. Correspondência para o papel de Expressão Humana.....	42
Figura 14. Histograma dos escores para qualificação de ontologias - Sem o uso da metodologia. Curvas estimadas a partir das distribuições normal e beta.....	66
Figura 15. Boxplots dos escores para qualificação de ontologias - Sem o uso da metodologia. Segundo experiência do avaliador.....	67
Figura 16. Boxplots dos escores para qualificação de ontologias - Sem o uso da metodologia. Segundo GOAL.....	68
Figura 17. Boxplots dos escores para qualificação de ontologias - Com o uso da metodologia. Segundo experiência do avaliador.....	68
Figura 18. Boxplots dos escores para qualificação de ontologias - Com o uso da metodologia. Segundo GOAL.....	69
Figura 19. Boxplots dos escores para qualificação de ontologias .....	70
Figura 20 – Gráficos de resíduo contra índices das observações (a) e gráfico normal de probabilidades com envelopes simulados (b) – Modelo de regressão beta para a construção de escores de qualificação de ontologias.....	75

## LISTA DE TABELAS

Tabela 1. As categorias de Aristóteles.....	18
Tabela 2. As categorias de Kant.....	19
Tabela 3. Metodologias e seus contextos.....	26
Tabela 4. Metodologias e suas áreas do conhecimento correspondentes.....	27
Tabela 5. Formas de avaliação propostas por Leo Obrst.....	30
Tabela 6. Critérios de qualidade definidos por Gangemi.....	31
Tabela 7. Critérios de qualidade definidos por Gruber.....	32
Tabela 8. Os critérios de qualidade de Gómez-Pérez.....	34
Tabela 9. Papeis da Representação do Conhecimento organizados em Goals.....	38
Tabela 10. Questões organizadas em Questions.....	38
Tabela 11. Critérios organizados em Metrics.....	39
Tabela 12. Níveis dos fatores.....	56
Tabela 13. Estimativas dos coeficientes do modelo da média da variável escores para qualificação de ontologias baseado na metodologia proposta.....	73
Tabela 14. Estimativas dos coeficientes do modelo da dispersão para os dados de escores para qualificação de ontologias baseado na metodologia proposta.....	74

## SUMÁRIO

<b>1</b>	<b>Introdução.....</b>	<b>12</b>
1.1	Motivação e contextualização do trabalho.....	12
1.2	Problemática.....	13
1.3	Objetivos.....	15
1.4	Escopo.....	16
1.5	Contribuições do trabalho.....	16
1.6	Organização da dissertação.....	16
<b>2</b>	<b>Fundamentação Teórica.....</b>	<b>17</b>
2.1	Ontologia.....	17
2.1.1	Perspectiva filosófica.....	17
2.1.2	Perspectiva computacional.....	20
2.1.3	Os cinco papéis da representação do conhecimento.....	21
2.2	Aplicações de ontologia.....	23
2.2.1	Web Semântica.....	23
2.2.2	Dados Conectados.....	24
2.3	Metodologias para engenharia de ontologias.....	25
2.4	A abordagem goal, question, metric (GQM).....	28
<b>3</b>	<b>Trabalhos Relacionados.....</b>	<b>30</b>
3.1	Avaliação de ontologias por Leo Obrst.....	30
3.2	Avaliação de ontologias por Aldo Gangemi.....	31
3.3	Avaliação de ontologias por Thomas Gruber.....	32
3.4	Avaliação de ontologias por Asunción Gómez-Pérez.....	34
3.5	Avaliação de ontologias por Denny Vrandečić.....	35
<b>4</b>	<b>A Proposta.....</b>	<b>37</b>
4.1	A Metodologia FOCA.....	37
4.2	Passos da metodologia.....	42
4.2.1	Passo 1: Verificação do tipo da ontologia.....	43
4.2.2	Passo 2: Verificação das questões e quantificar.....	44
4.2.3	Passo 3: Verificação da qualidade da ontologia.....	48
<b>5</b>	<b>Validação – Experimento.....</b>	<b>52</b>
5.1	Projeto de Experimento.....	52
5.1.1	Definição do problema.....	52

5.1.2	Objetivos de investigação.....	54
5.1.3	Planejamento do experimento.....	54
5.1.3.1	Questões de pesquisa e hipóteses.....	55
5.1.3.2	Fatores e variáveis de resposta.....	55
5.1.3.3	Unidades de experimento.....	56
5.1.3.4	Plano de execução.....	57
5.1.3.5	Análise de ameaças à validade.....	58
5.2	Narrativa de execução do experimento.....	58
5.3	Análise dos dados.....	60
5.3.1	Análise descritiva dos dados.....	65
5.3.2	Verificação das hipóteses.....	72
5.4	Principais conclusões.....	76
<b>6</b>	<b>Conclusões e Trabalhos Futuros.....</b>	<b>77</b>

## **Referências**

## **Apêndices**

# 1. INTRODUÇÃO

O trabalho apresentado ao longo desta dissertação trata sobre a proposição, criação e validação de uma metodologia para avaliação de ontologias que utiliza princípios da representação do conhecimento, relacionando-os com os principais critérios de qualidade de ontologias. Situa-se na linha de pesquisa de Descoberta do Conhecimento do Mestrado em Modelagem Computacional do Conhecimento, do Instituto de Computação da Universidade Federal de Alagoas.

Essa linha envolve pesquisa em representação e processamento de conhecimento em diferentes áreas, abrangendo o uso de modelos matemático-computacionais e de técnicas de inteligência artificial, numa perspectiva de apoio a processos decisórios. O trabalho desenvolvido se caracteriza como interdisciplinar, visto que a área de ontologias pode envolver diversas áreas do conhecimento, como computação, biomedicina e outras.

Esta dissertação visa auxiliar o avaliador de ontologia a tomar decisões mais precisas sobre a qualidade da ontologia que está sendo avaliada, através do estabelecimento de um metodologia para avaliação de ontologias que, contendo um passo-a-passo, facilite o processo de avaliação.

## 1.1. Motivação e contextualização do trabalho

Sistemas baseados em conhecimento (SBC) são caros de se construir, testar e manter. A complexidade da construção de SBC ocorre, por um lado, pela necessidade de representação do conhecimento de um determinado domínio e, por outro lado, pela necessidade de processamento inteligente através do conhecimento representado. Uma das formas de simplificar o ciclo de vida de desenvolvimento de SBC é através da utilização de metodologias, ou seja, é necessário o uso de metodologias baseadas em engenharia de software com especificações formais para recursos compartilhados e componentes reusáveis [Gruber, 1993]. Desta forma, é necessário um artefato que especifique um vocabulário que seja comum a tais recursos e componentes.

Considere o problema de reutilização no planejamento de sistemas baseados em conhecimento. Tal sistema possui descrições de objetos, eventos, recursos e limitações. Assim, para iniciar o planejamento, há uma dependência de uma base de conhecimento (às vezes chamado de "teoria do domínio") pois ela contém o conhecimento genérico para esta tarefa. Por exemplo, em um contexto de múltiplos agentes de software, os quais

compartilham a mesma base de conhecimento, é possível um agente trocar informações sobre o sistema com outro, e para isso é necessário o uso de um vocabulário comum para que ocorra essa interoperabilidade.

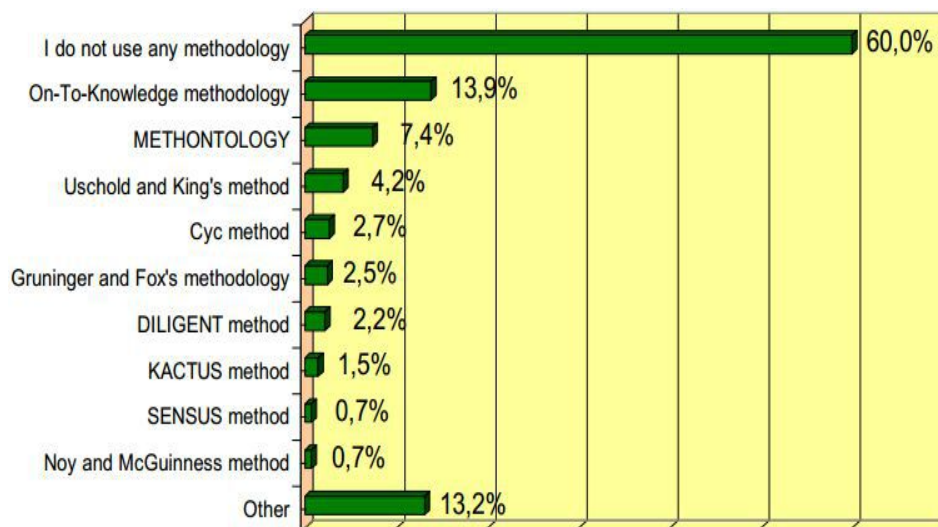
Sendo assim, um artefato computacional muito valioso que pode desempenhar esse papel é a ontologia. Uma ontologia define um vocabulário comum para um determinado domínio, de forma que qualquer desenvolvedor, que necessite conceituar este domínio para a máquina, possa reutilizá-la. Assim, é possível que sistemas de computador que utilizam a mesma ontologia para o mesmo domínio tornem-se interoperáveis, pois eles “falam a mesma língua”, sendo possível trocar conhecimento, dados e informações.

Atualmente, as ontologias estão sendo utilizadas em diferentes domínios, desde indústrias, para melhorar a comunicação entre funcionários, até a aplicação em páginas Web para melhorar o processamento e navegação de máquinas. Assim, por exemplo, para a máquina, uma página Web tradicional que contém uma imagem, sabe somente que este recurso é apenas uma imagem; quando há a utilização de uma ontologia que modela este recurso, a máquina conhece todas as propriedades daquela imagem, ou seja, toda a semântica que a envolve, desde suas propriedades até o que está sendo representada nela. Ou seja, tudo que existe numa página web tem um significado, tornando buscas muito mais poderosas e precisas.

## 1.2. Problemática

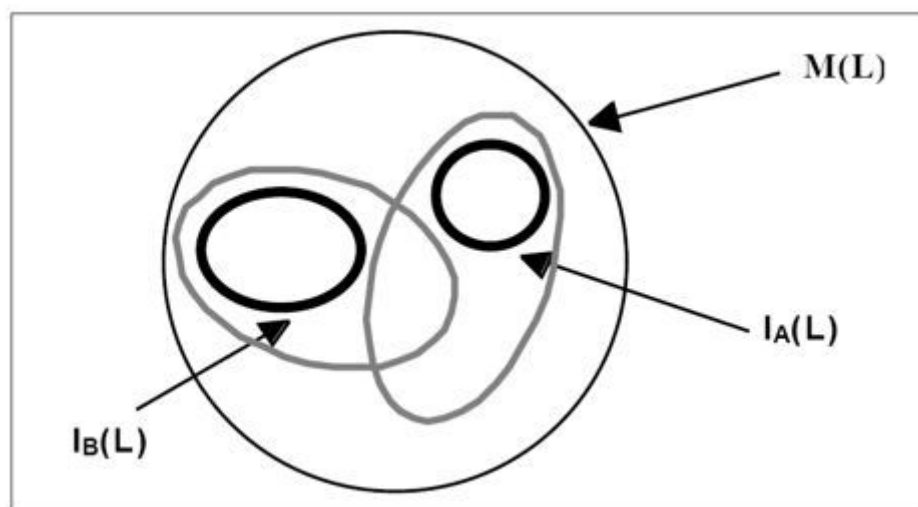
Modelar uma ontologia não é uma tarefa simples, pois o ontologista precisa conhecer profundamente o domínio que está sendo modelado, para que ele saiba capturar os principais conceitos que o envolvem, a fim de tornar a ontologia coerente com o domínio e capaz de inferir novos conceitos a respeito dele de forma eficiente. Além disso, ele precisa definir os termos de forma clara para facilitar o reuso por outros ontologistas.

Pelas dificuldades que envolvem o processo de modelagem de uma ontologia, uma pessoa que nunca modelou uma ontologia necessitará de uma metodologia para guiá-la neste processo. No entanto, uma pesquisa mostra que 60% das pessoas que modelam ontologias não utilizam nenhuma metodologia em seu processo de modelagem [Cardoso, 2007]. A Figura 1 ilustra a disparidade entre as pessoas que responderam que não utilizam qualquer metodologia e as que usam alguma metodologia.



**Figura 1. Utilização de metodologias pelos ontologistas. Fonte: [Cardoso, 2007]**

Diante de todas estas dificuldades, é possível observar que quanto mais cresce a complexidade da ontologia, mais sua qualidade tende a decrescer, diminuindo também o seu potencial de reuso. Para exemplificar este problema, considere duas ontologias, onde uma modela o domínio bicicleta e outra de motocicleta. Para modelar o domínio bicicleta, foi reutilizado o conceito pneu da ontologia de carro. Porém, sabe-se que as medidas de um pneu de bicicleta não são iguais as de um carro, o que torna uma inconsistência na modelagem do domínio de bicicleta. Se esta ontologia for reutilizada pela maioria dos ontologistas, um consenso errôneo sobre bicicleta se propagaria de uma forma absurda, o que é um grande problema. Assim, modelar uma ontologia é um processo em que se tenta definir um domínio de forma única, ou seja, sem nenhuma inconsistência. A Figura 2 ilustra esse problema. Suponha que  $M(L)$  seja o “mundo dos veículos”,  $I_a(L)$  o “entendimento de bicicleta” e  $I_b(L)$  o “entendimento de carro”. A modelagem ideal seria o que está representada pelos círculos em preto, ou seja, sem intersecção. Quanto menos restrições o ontologista insere em seu domínio, o entendimento sobre ele se torna mais geral, o que pode acontecer de conceitos (como pneu) em um domínio seja o mesmo do que de outro domínio totalmente diferente (representado pela intersecção em cinza na Figura).



**Figura 2. Problema da intersecção em ontologias mal modeladas.**

Desta forma, torna-se imprescindível não somente construir ontologias de forma sistemática, como também avaliar a qualidade de uma determinada ontologia, para que o ontologista compreenda as limitações da mesma, faça melhorias e a reuse. No entanto, muitos autores enunciam vários critérios de qualidade de ontologia (como Obrst, Gangemi, Gruber, Gómez-Perez) e, de acordo com [Vrandečić, 2009] uma boa ontologia não executa todos os critérios, pois muitos deles são contraditórios, como por exemplo “compromisso ontológico mínimo” e “completude”. Assim, dependendo do tipo de ontologia que está sendo avaliada, alguns critérios são avaliados e outros não. Além disso, as metodologias atuais apresentam os critérios, porém não estabelece um passo-a-passo de como este critério deve ser avaliado. Desta forma, é necessário um ontologista bastante experiente na área para entender do que se trata o critério para poder avaliar a ontologia, tornando a avaliação dependente da experiência de quem a usa. Além disso, [Vrandečić, 2009] diz que a falta de métodos de correspondência e avaliações experimentais travam a significatividade da avaliação da ontologia.

### 1.3. Objetivos

O principal objetivo deste trabalho é estabelecer uma metodologia para avaliação de qualidade de uma ontologia, conectando critérios de qualidade com a representação do



conhecimento. Com isso, a avaliação de uma ontologia ocorrerá independentemente da experiência do avaliador. Os objetivos específicos deste trabalho são:

- Realizar um levantamento dos principais critérios de qualidade de ontologia das metodologias para avaliação de ontologias atuais;
- Discutir e propor uma nova metodologia para avaliação de ontologias que poderá ser utilizado por ontologistas com diferentes níveis de experiências;
- Relacionar os principais critérios de qualidade de ontologias com os papeis da representação do conhecimento;
- Estabelecer métricas para cada critério avaliado;
- Avaliar a metodologia proposta e construída, apresentando suas vantagens e desvantagens, bem como os trabalhos futuros.

#### 1.4. Escopo

Este trabalho é direcionado aos ontologistas que tem o objetivo de avaliar a qualidade de ontologias OWL ou RDF-S. Além disso, diagnosticar a qualidade da ontologia em relação ao cumprimento dos cinco papeis da representação do conhecimento.

#### 1.5. Contribuições do trabalho

A metodologia do presente trabalho foi criada para contribuir com a área de Ontologia Aplicada, propondo uma abordagem sistemática para avaliação de ontologias.

#### 1.6. Organização da dissertação

Esta dissertação está organizada da seguinte forma: O Capítulo 2 apresenta os principais conceitos que envolvem a dissertação, para um melhor entendimento do trabalho; o Capítulo 3 traz os principais trabalhos relacionados e usados como base para a construção do método; o Capítulo 4 trata sobre o método em si, com seus passos em mais detalhes; o Capítulo 5 mostra como foi feita a validação do método e finalmente o Capítulo 6 traz as conclusões e trabalhos futuros.

## 2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo traz os principais conceitos que envolvem este trabalho para um melhor entendimento da área. Primeiramente, este capítulo apresenta o conceito de ontologias. Em seguida, as aplicações da ontologia em Ciência da Computação. Logo após, as metodologias para Engenharia de Ontologias. Por último, este capítulo mostra a metodologia GQM (*goal, question, metric*).

### 2.1. Ontologia

Esta seção apresenta o conceito de ontologia, sob as perspectivas filosófica e computacional e pela visão da representação do conhecimento.

#### 2.1.1. Perspectiva filosófica

Ontologia é um termo originado na Filosofia e formado através dos termos gregos *ontos* (ser) e *logos* (estudo). Ontologia como um ramo da Filosofia é a ciência do que é, dos tipos e estruturas de objetos, propriedades, eventos, processos e relações em todas as áreas da realidade. Ontologia é frequentemente usado por filósofos como sinônimo de metafísica (que pode significar como “o que vem depois da física”), um termo usado pelos primeiros alunos de Aristóteles para se referir ao que o próprio Aristóteles chamou “filosofia primeira”. Às vezes, o termo ontologia é usado em um sentido mais amplo, para se referir ao estudo do que poderia existir. [Thomasson, 2004]

Este termo surge quando se quer definir, de maneira única, qualquer ente que exista no universo. Por exemplo, ao se querer definir uma cadeira, utilizam-se conceitos primitivos que, em conjunto, formam o ente cadeira, ou seja, características de uma cadeira que a definem unicamente, porém nenhum outro ente tem estas características bem definidas. No entanto, para que uma pessoa tente definir qualquer ente que exista neste universo, é necessário que ela encontre características únicas desse ente, de forma que ele não seja confundido com nenhum outro ente, o que é uma tarefa bastante complicada de se executar.

Para facilitar a tarefa de se definir qualquer ente que existe no universo, muitos filósofos adotaram sistemas de categorias, ou seja, um modo geral para se definir algo. Por exemplo, Aristóteles enunciou um método de categorização [Thomasson, 2004], onde

a primeira categoria chamada substância primeira é um indivíduo (como homem ou cavalo) e as espécies (ser humano, animal) são substâncias secundárias. Para seguir a categorização, a pessoa faz perguntas sobre a substância, como por exemplo “O que é?” para se ter informações sobre identidade, “Como é?” para se ter informações sobre qualidade, “O que está fazendo?” para se ter informações sobre ação, e assim por diante. Desta forma, Aristóteles apresentou uma lista de categorias, mostrada na Tabela 1.

<b>Categoria</b>	<b>Exemplo</b>
Substância	Homem, Cavalo
Quantidade	Quatro-pés, Cinco-pés
Qualidade	Branco, Gramatical
Relação	Dobro, Metade
Lugar	No Liceu, No mercado
Data	Ontem, Ano passado
Postura	Está mentindo, Está sentando
Estado	Esta de sapatos, Está armado
Ação	Cortando, Queimando
Passivo	Sendo cortado, Sendo queimado

**Tabela 1. As categorias de Aristóteles. Fonte: [Thomasson, 2004]**

Kant também enunciou categorias. Enquanto Aristóteles utilizava o método da linguagem, Kant utilizou o método do pensamento (ou do julgamento), ou seja, as categorias propostas neste método dizem respeito a um objeto conhecido, porque nosso pensamento impõe uma estrutura nos objetos que ele pode conhecer. Assim, uma pessoa pode categorizar algo com base em um objeto conhecido que se assemelha ao que se quer categorizar. A Tabela 2 mostra as categorias de Kant.

<b>Categoria</b>	<b>Subcategorias</b>
Quantidade	Unidade; Pluralidade; Totalidade.
Qualidade	Realidade; Negação; Limite.
Relação	Inerência e Subsistência; Causalidade e Dependência; Comunidade.
Modalidade	Possibilidade; Existência; Necessidade.

**Tabela 2. As categorias de Kant. Fonte: [Thomasson, 2004]**

Assim como as categorias de Aristóteles, as categorias de Kant dependiam do pensamento humano, ou seja, a partir de um entendimento equivocado de uma pessoa, esta falharia na categorização de algo.

Baseados nestas dificuldades, filósofos contemporâneos (como Johansson, Crisholm, Hoffman, entre outros) aplicam o método da abstração sucessiva. Por exemplo, observando um homem em particular, por exemplo, Sócrates, pode-se dizer, “Sócrates é homem” para “Sócrates é um mamífero” para “Sócrates é um animal” até que se chegue em “Sócrates é uma substância”. Desta forma, todas as entidades do mundo teriam esta substância primária, e a partir dela, esta substância é especializada até que se chegue no que se quer definir.

### 2.1.2. Perspectiva computacional

O termo ontologia foi trazido para a Ciência da Computação com o objetivo de descrever algo, para a máquina, da mesma forma que ele existe no mundo. A primeira aparição deste termo foi em 1967, quando especialistas em processamento de dados se preocuparam em descrever um conjunto de noções sobre os dados de forma clara e de concordância comum, e isso envolvia descrever o mundo real e ideias existentes sobre o assunto na mente dos homens [Mealy, 1967]. Passados alguns anos, sistemas baseados em conhecimento que são baseados em plataformas de hardware, linguagens de programação e protocolos de rede heterogêneos necessitavam ser interoperáveis para o compartilhamento de conhecimento entre eles. Para isto, seria necessário um artefato que especificasse uma linguagem comum para que ocorresse esta comunicação.

De acordo com Gruber, ontologia é uma especificação explícita de uma conceitualização [Gruber, 1993], ou seja, todos os conceitos e relações entre estes conceitos são especificados explicitamente, em uma linguagem formal. Para Guarino, ontologia é uma explícita e parte de uma descrição de uma conceitualização [Guarino, 1998]. Guarino inicia definindo o que é conceitualização, como sendo a estrutura  $\langle D, R \rangle$ , onde  $D$  é o domínio e  $R$  é um conjunto de relações relevantes sobre  $D$ . O espaço de domínio é definido com estrutura  $\langle D, W \rangle$  onde  $D$  é o domínio e  $W$  é o conjunto de “mundos possíveis”. Por exemplo,  $D$  pode ser um conjunto de blocos em uma mesa e  $W$  seria todos os possíveis arranjos que estes blocos podem formar. Desta forma, se considerarmos uma conceitualização  $C$  como sendo  $\langle D, W, R \rangle$ , onde para cada  $w$  pertencente a  $W$ , a estrutura entendida de  $w$  de acordo com  $C$  é a estrutura onde se tem o conjunto de extensões (relativas a  $w$ ) dos elementos de  $R$  (para um mundo possível). Assim, dado uma linguagem  $L$ , com compromisso ontológico  $K$ , uma ontologia em  $L$  é um conjunto de axiomas concebidos de uma tal forma que o conjunto de seus modelos aproxima da melhor forma possível o conjunto de modelos destinados a  $L$  de acordo com  $K$ . A Figura 3 ilustra a definição dada por Guarino.

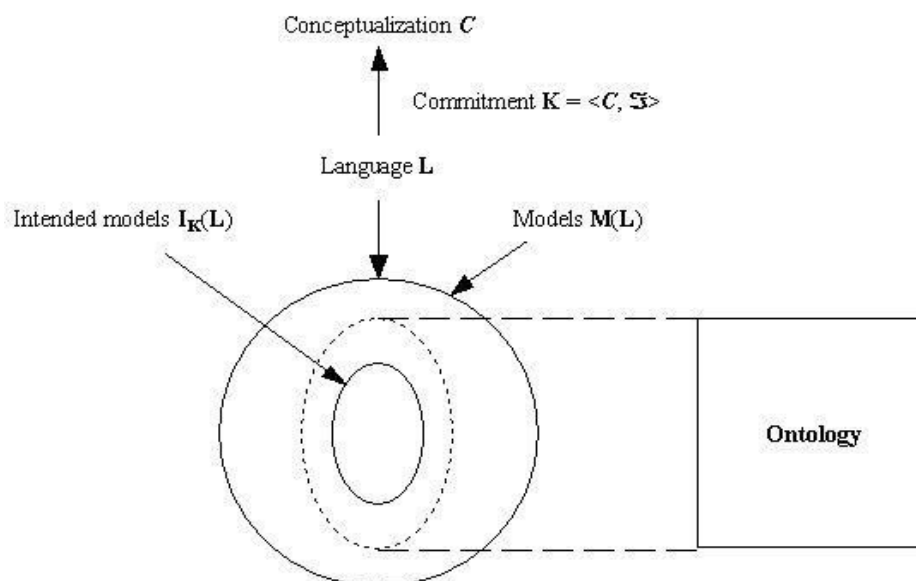


Figura 3. Definição de ontologia por Guarino. Fonte: [Guarino, 1998]

### 2.1.3. Os cinco papéis da representação do conhecimento

Uma outra forma de entender o conceito de ontologia é através dos papéis da representação do conhecimento. [Davis, 1993] enunciou os cinco papéis como sendo:

- **Substituto.** Um substituto para a coisa em si, que é usado para permitir que uma entidade determine consequências através do pensamento em vez de agir, isto é, através do raciocínio sobre o mundo em vez de agir sobre ele. Para o cumprimento deste papel é necessário pensar sobre duas questões importantes. A primeira questão reflete identidade: “O que é um substituto para isto?”, onde há a correspondência entre o substituto e a sua referente intenção no mundo. A segunda questão reflete fidelidade: “O quão perto é o substituto para a coisa real?”, ou seja “Quais atributos ele captura e torna explícitos e quais ele omite?”. Percebe-se que a princípio é impossível que um substituto seja a coisa em si pois somente a coisa em si pode ser ela mesma. A única representação que é precisa então será o ente em si, e qualquer outra torna-se imprecisa.
- **Compromissos Ontológicos.** Basicamente é a resposta para a pergunta: “Em que termos devo pensar sobre o mundo?”. Como visto no papel anterior, a coisa real pode ter várias representações, onde se quer encontrar a menos imprecisa possível, já que a coisa em si é a única representação possível. O conjunto de compromissos ontológicos é justamente a escolha da representação menos imprecisa possível, ou

seja, a mais coerente com o mundo real, aquela que além de ter os termos mais relevantes a respeito do mundo real, identifica os principais aspectos a respeito dele.

- **Raciocínio Inteligente.** O raciocínio é expresso em termos de três componentes: (1) a representação das concepções fundamentais de raciocínio inteligente, (2) o conjunto de inferências que a representação sanciona e (3) o conjunto de inferências que é recomendado. Estes componentes podem ser vistos como as respostas de representação de três questões fundamentais correspondentes: (1) O que significa raciocinar inteligentemente? (2) O que nós podemos inferir do que sabemos? e (3) O que nós devemos inferir sobre o que nós sabemos? Primeiro, raciocínio inteligente é definir propriedades essenciais para o raciocínio inteligente. Várias áreas do conhecimento, como matemática lógica, define raciocínio inteligente através de cálculos formais e deduções, na psicologia, pela resolução de problemas humanos, entre outras definições. Segundo, um conjunto de inferências sancionadas é um conjunto selecionado de inferências que são consideradas conclusões que se pode ter a partir da informação disponível. Terceiro, a partir das várias inferências que se pode fazer a partir de uma dada informação, necessita-se ter uma indicação de que inferências são adequadas para se fazer, ou seja, inteligente. Esta indicação é fornecida pelo conjunto de inferências recomendadas.
- **Computação Eficiente.** O pensamento ocorrerá em um ambiente computacional, ou seja, a máquina é a principal responsável pelo raciocínio lógico a respeito do domínio, através dos conceitos, suas relações e restrições. Este papel visa a rapidez e execução bem sucedida do raciocínio por parte da máquina.
- **Expressão Humana.** A linguagem que usamos para dizer coisas sobre o mundo, ou seja, é o papel que trata da comunicação clara dos conceitos para a máquina. Basicamente, é se perguntar, a respeito do mundo, sobre o quão é fácil falar ou pensar a respeito dele nesta linguagem, sobre que tipos de coisas são facilmente ditas na linguagem, e quais tipos de coisas são difíceis de serem entendidas. Em outras palavras, a linguagem utilizada para falar sobre o mundo deve ser entendível sem esforço algum.

## 2.2. Aplicações de ontologia

Esta seção mostra dois exemplos de aplicação de ontologias em Ciência da Computação.

### 2.2.1. Web Semântica

A Web Semântica, considerada uma extensão da web atual, visa melhorar a estrutura comum das páginas Web de forma que todos os seus recursos sejam entendíveis por máquina, através de uma linguagem comum. Tim Berners-Lee falou que a Web Semântica irá trazer significado ao conteúdo das páginas Web, criando um ambiente em que os agentes de software, percorrendo páginas, podem facilmente realizar tarefas sofisticadas para os usuários [Berners-Lee, 2001]. A ontologia tem o papel de definir semanticamente os dados que serão armazenados na aplicação, de forma que a aplicação seja capaz de realizar inferências a respeito do domínio em questão. A Figura 4 a seguir ilustra a estrutura da Web Semântica.

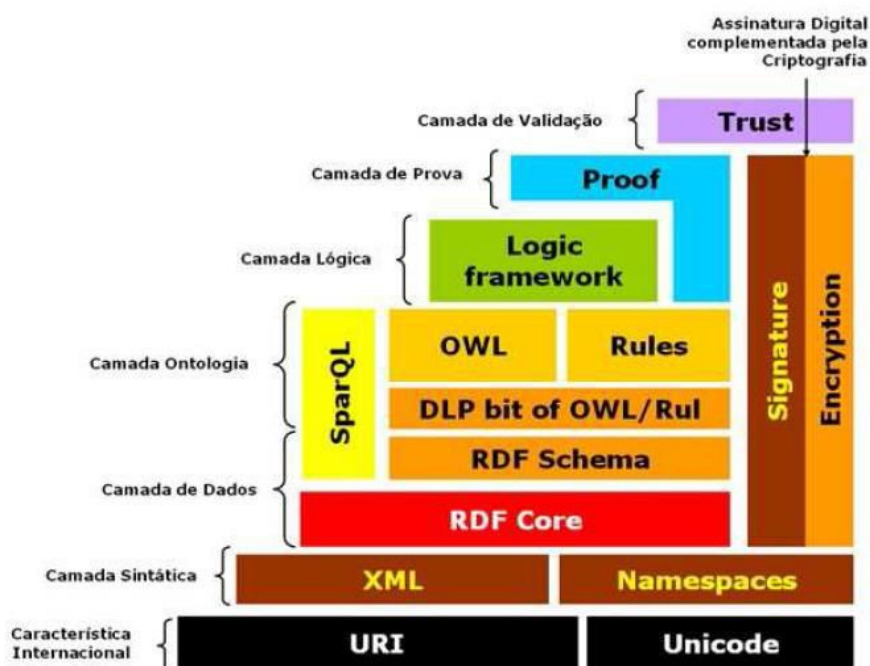


Figura 4. Estrutura da Web Semântica. Fonte: <http://www.redmondpie.com/semantic-web-wakeup-call-part-1/>



### 2.2.2. Dados Conectados

Dados Conectados, ou *Linked Data*, é fazer ligações entre os dados digitados a partir de diferentes fontes [Bizer, 2009], utilizando boas práticas para a publicação destes.

Dados Conectados possuem quatro princípios (enunciadas por Berners-Lee) para a publicação dos dados na web para que esta seja um espaço de dados global:

- **Use URI como nomes das coisas.** O uso de *Universal Resource Identifier* permite que cada recurso tenha seu identificador único, não permitindo o problema de ambiguidade de dados;
- **Use HTTP URIs para que pessoas possam procurar por estes nomes.** A partir do momento que cada dado tem seu identificador, torná-los disponíveis para busca;
- **Prover informação útil, usando padrões (RDF, SPARQL).** RDF, *Resource Description Framework*, é um modelo de representação e estruturação dos dados que permite raciocínio por parte da máquina. SPARQL é uma linguagem de consulta sobre estes tipos de dados, semelhante ao SQL. A partir do momento que o dado é descoberto na web, ele deve conter informações úteis e não apenas ser um dado comum.
- **Incluir links para outras URIs, para que eles possam descobrir mais coisas.** A partir das informações úteis que são conhecidas no dado, este contém *links* que conectam para mais informações a respeito deste ou de outras coisas.

Além dos quatro princípios, os Dados Conectados, quando abertos, também existe uma classificação 5-estrelas. Os documentos que possuem a classificação 1-estrela são quaisquer documentos encontrados na web, sob uma licença aberta, em qualquer formato, como jpg, txt, pdf, etc. Os documentos que possuem a classificação 2-estrelas são aqueles que possuem as características 1-estrela e apresentam um formato estruturado, porém é necessário um software proprietário para o manipular, como por exemplo, o formato xls. Os documentos que possuem a classificação 3-estrelas são aqueles que possuem as características 2-estrelas, porém não é mais necessário um software proprietário para o manipular, como por exemplo, o formato csv. Os documentos que possuem a classificação 4-estrelas são aqueles que possuem as características 3-estrelas e seguem os padrões do W3C (*world wide web consortium*) como RDF (utilização do framework e vocabulários existentes), SPARQL (acesso aos recursos) e que fazem a utilização de URIs

(nomear os recursos). Os documentos que possuem a classificação 5-estrelas são aqueles que possuem as características 4-estrelas e possuem também links para outras URIs de modo que as pessoas possam conhecer mais sobre o dado. A Figura 5 ilustra a classificação 5-estrelas.

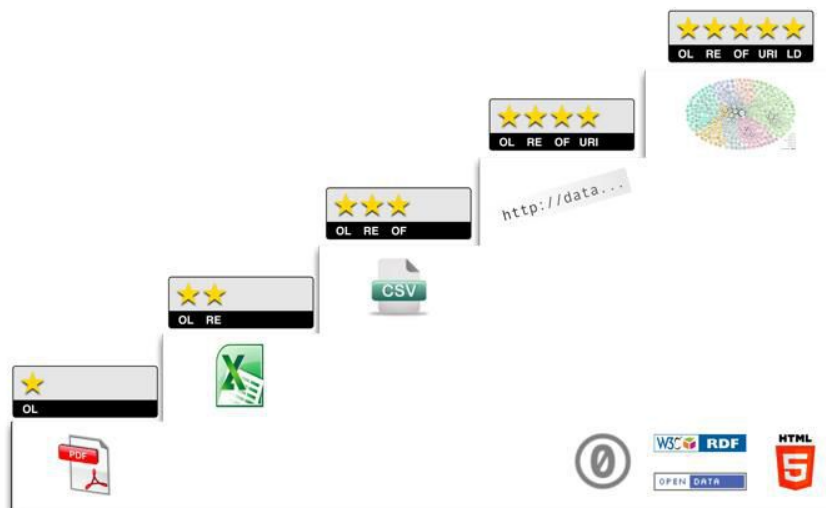


Figura 5. Classificação 5-estrelas. Fonte: <http://5stardata.info>

### 2.3. Metodologias para Engenharia de Ontologia

Esta seção apresentará um resumo sobre as principais metodologias para engenharia de ontologias.

Atualmente, existem aproximadamente 20 metodologias para engenharia de ontologias, as quais nasceram provenientes de um determinado contexto. A Tabela 3 mostra esta relação.

Metodologia	Contexto
-------------	----------

Ontology Development 101 [Noy, 2001]	Geral
Grüninger and Fox's Methodology [Grüninger 1995]	Empresa
Methontology [Fernandéz, 1997]	Química
Ontoclean [Guarino, 2002]	Geral
Unified Methodology [Uschold, 1996]	Geral
The Cyc Method [Lenat, 1985]	Sistemas Especialistas
The DILIGENT Methodology [Pinto, 2004]	Biologia
The KACTUS approach [Schreiber 1995]	Redes Elétricas
HCOME [Kotis, 2006]	Centrado em humanos
On-To Knowledge [Sure, 2004]	Gerenciamento de Conhecimento
The DOGMA approach [Jarrar, 2002]	Ontologia de Banco de Dados
ONIONS [Gangemi, 1996]	Domínio Médico
The Terminae Method [Aussenac-Gilles, 2008]	Análise de Texto
UPON [De Nicola, 2005]	Desenvolvimento de Software
IDEF5 [Menzel, 1992]	Criação e Manutenção de ontologias
Termontography [Temmerman, 2003]	Terminologias
RapidOWL [Auer, 2007]	Desenvolvimento Ágil de Software
The CommonKADS Methodology [Schreiber, 1999]	Geral
The NeOn methodology [Goméz Pérez, 2009]	Geral

**Tabela 3. Metodologias e seus contextos. Fonte: Criado pelo Autor.**

As metodologias são inspiradas em três áreas do conhecimento: Inteligência Artificial, focadas na lógica de primeira ordem na definição dos principais termos do domínio; Engenharia de Software, focadas nas fases que consistem a construção de um software e Interação Humano-Computador, focadas na interação entre os atores que desenvolvem a ontologia (engenheiros de software, engenheiros de ontologia,

engenheiros do conhecimento, etc). A Tabela 4 mostra as áreas do conhecimento e suas metodologias derivadas correspondentes:

<b>Metodologia</b>	<b>Área do Conhecimento</b>
Ontology Development 101	Engenharia de Software
Grüninger and Fox's Methodology	Inteligência Artificial
Methontology	Engenharia de Software
Ontoclean	Inteligência Artificial
Unified Methodology	Engenharia de Software
The Cyc Method	Inteligência Artificial
The DILIGENT Methodology	Interação Humano-Computador
The KACTUS approach	Inteligência Artificial
HCOME	Interação Humano-Computador
On-To Knowledge	Engenharia de Software
The DOGMA approach	Inteligência Artificial
ONIONS	Engenharia de Software
The Terminae Method	Inteligência Artificial
UPON	Engenharia de Software
IDEF5	Engenharia de Software
Termontography	Engenharia de Software
RapidOWL	Engenharia de Software
The CommonKADS Methodology	Inteligência Artificial
The NeOn methodology	Engenharia de Software

**Tabela 4. Metodologias e suas áreas do conhecimento correspondentes. Fonte: Criado pelo Autor**

#### 2.4. A abordagem *Goal, Question, Metric* (GQM)

Esta seção abordará sobre o conceito da abordagem Goal, Question, Metric, utilizada no presente trabalho para a construção da metodologia proposta neste trabalho.

De acordo com [Basili, 1992], todo processo de engenharia de software requer uma avaliação. As avaliações podem surgir de diversas perspectivas, ou seja, de desenvolvedores, gerentes, clientes ou organização. Desta forma, existem vários objetivos de avaliação por parte destes atores, com diferentes verificações e métricas, objetivando mensurar a qualidade do que foi desenvolvido.

Para uma organização mensurar o que foi desenvolvido, primeiramente deve especificar os objetivos necessários para a avaliação; traçar estes objetivos para os dados que são entendidos para definir estes objetivos operacionalmente; prover um framework para interpretar os dados relacionados aos objetivos, onde esses dados podem ser quantificados. Para estes passos, é utilizado o paradigma Goal/Question/Metric (GQM). Os objetivos são definidos sendo refinados em um conjunto de questões que são usadas para extrair as informações apropriadas dos modelos. Da mesma forma, as questões e modelos definem um conjunto de métricas e dados para coleção e provê um framework para interpretação. A Figura 6 ilustra a organização da abordagem GQM.

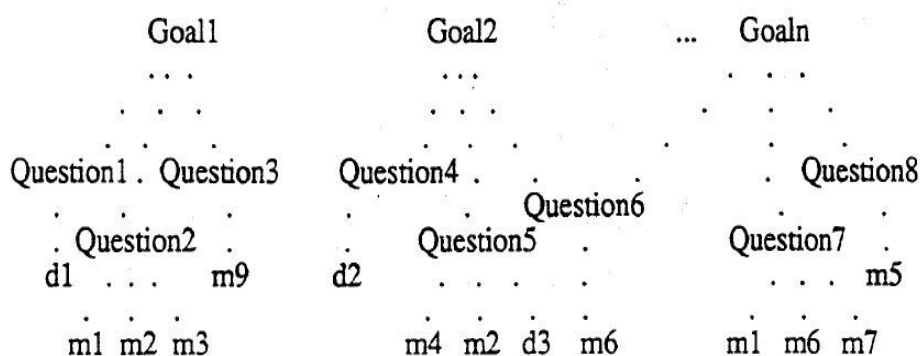


Figura 6. Representação da abordagem GQM. Fonte: [Basili, 1992]

De acordo com [Koziolek, 2008], o paradigma GQM consiste em quatro fases (também mostradas na Figura 7):

- Planejamento: Nesta fase, o time GQM é estabelecido e a área de avaliação desejada (como confiabilidade, desempenho, segurança, etc.) é identificada. Após isto, o time seleciona e caracteriza o produto ou processo a ser estudado. O resultado desta fase é o plano de projeto;

- **Definição:** Nesta fase, os objetivos de medida são definidos. Baseados nos objetivos, as questões são formuladas, perguntando por atributos específicos e tornando certos aspectos dos objetivos mais concretos. Para cada questão, uma hipótese com uma resposta esperada deve ser definida. Após isto, métricas são definidas para cada questão e checadas suas consistências e completudes. Os resultados desta fase são o plano GQM, o plano de medição e o plano de análise;
- **Coleta de dados:** Nesta fase é onde a medição ocorre de fato. A coleção dos dados pode ser feita de forma manual ou eletrônica e que pode envolver a utilização de ferramentas para a coleção de dados automatizado. A medição consiste em planilhas, ferramentas estatísticas, sistemas de banco de dados, entre outros;
- **Interpretação:** Nesta fase, os dados coletados são processados de acordo com as métricas definidas antes dos resultados de medição. As medições podem ser usadas para responder as questões e as respostas para avaliar se os objetivos iniciais terem sido atendidos.

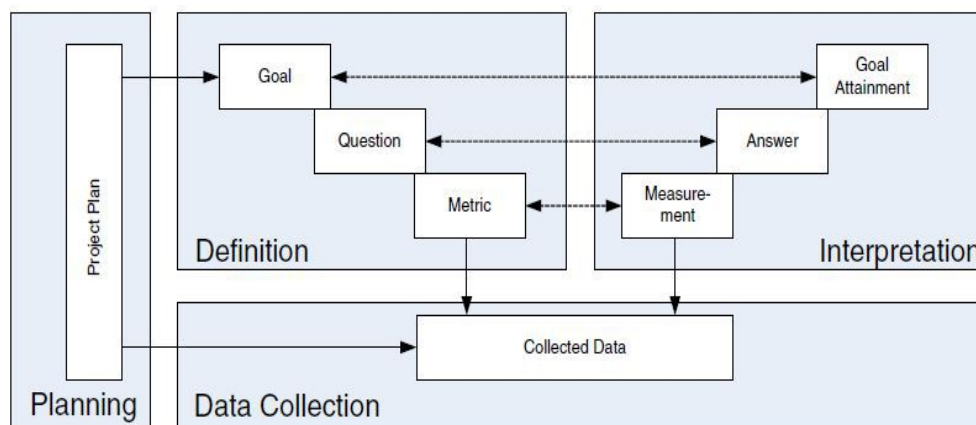


Figura 7. As quatro fases do paradigma GQM. Fonte: [Koziolk, 2008]

### 3. TRABALHOS RELACIONADOS

Este capítulo abordará os principais trabalhos relacionados com esta dissertação, os quais foram utilizados como base para a construção da metodologia proposta neste trabalho.

#### 3.1. Avaliação de ontologias por Leo Obrst

O trabalho de Leo Obrst [Obrst, 2007] propõe descrever os principais problemas em avaliação de ontologias sob diversas visões.

Antes de enunciar os problemas, Obrst mostra as diferentes formas de se avaliar uma ontologia, que são através de critérios de qualidade; de perguntas com embasamento filosófico e através da abordagem de verificação e validação. A Tabela 5 detalha as formas de avaliação propostas por Obrst.

<b>Forma de avaliação</b>	<b>Descrição</b>
Critérios de Qualidade	Avaliar através dos critérios Cobertura, Complexidade, Granularidade, Casos de Uso, Cenários, Requisitos, Aplicações, Consistência e Completude.
Questões com base filosófica	A ontologia é mapeável para algumas ontologias superiores específicas, de modo que sua avaliação será pelo menos parcialmente dependente da superior também? Qual é a teoria filosófica subjacente da ontologia sobre a realidade? (Idealista, Realista, Nominalista, Dimensionalista).
Verificação e Validação	Questões estruturais, funcionais e de usabilidade.

**Tabela 5. Formas de avaliação propostas por Leo Obrst**

O ponto forte do trabalho de Obrst é que ele mostra as diferentes visões em que uma ontologia deve ser avaliada. Além disso, mostra os principais problemas em avaliação de ontologias, sob os mais diferentes assuntos, como a Representação do

Conhecimento, Casos de Uso e Domínio de Requisitos de ontologias, Acordo Semântico e Construção do Consenso, Similaridade Semântica e Distância Semântica e Alinhamento com outras ontologias.

No entanto, a limitação do trabalho de Obrst é que são mostradas as formas de avaliar uma ontologia, porém não é mostrado como executar tais formas, ou seja, um passo-a-passo de forma a facilitar a avaliação.

### 3.2. Avaliação de ontologias por Aldo Gangemi

O trabalho de Gangemi [Gangemi, 2006] destina-se a apresentar uma abordagem abrangente para a avaliação e validação de ontologias.

Em seu trabalho, Gangemi mostra os tipos de métricas para a avaliação de ontologias, como dimensão estrutural; dimensão funcional; e finalmente a dimensão de perfil de usabilidade. Na dimensão funcional, o autor enuncia critérios para avaliação de ontologias, como: Acordo, Satisfação do Usuário, Tarefa, Tópico e Modularidade. A Tabela 6 mostra os critérios e suas descrições.

<b>Critério de Qualidade</b>	<b>Descrição</b>
Acordo	Mensurado através da proporção do acordo que os especialistas tem com relação aos elementos da ontologia, ou seja, através da medição do consenso de um grupo de especialistas.
Satisfação do Usuário	Pode ser avaliado por meio de pesquisas dedicadas ou avaliação de confiança.
Tarefa	Lida com a medição de uma ontologia de acordo com a sua aptidão para algumas metas, condições prévias, pós-condições, restrições, opções, etc.
Tópico	Trata de medir a ontologia de acordo com a sua aptidão para um repositório de conhecimento existente.



Modularidade	Mede a aptidão para um repositório de componentes reutilizáveis existente.
--------------	--

**Tabela 6. Critérios de qualidade definidos por Gangemi.**

O ponto forte do trabalho de Gangemi é que a avaliação é mostrada como três tipos de dimensões, uma que é focada na sintaxe e semântica, uso estendido da ontologia e a comunicação sobre seu contexto.

No entanto, assim como o trabalho de Obrst, todos os critérios apresentados por Gangemi não contém um passo-a-passo de cada critério que mostra como avaliar cada critério na avaliação da ontologia.

### 3.3. Avaliação de ontologias por Thomas Gruber

O trabalho de Gruber [Gruber, 1995] propôs um conjunto de critérios para guiar o desenvolvimento de ontologias.

Gruber apresenta vários critérios como Clareza, Coerência, Estendibilidade, Mínimo Viés de Codificação e Compromisso Ontológico Mínimo. Estes critérios estão organizados juntos com suas descrições na Tabela a seguir.

<b>Critério de qualidade</b>	<b>Descrição</b>
Clareza	Uma ontologia deve efetivamente comunicar o significado entendido dos termos, os quais suas definições devem ser objetivas e independentes do contexto que são modeladas e todas elas devem ser documentadas com uma linguagem natural.
	A ontologia deve sancionar inferências que são consistentes com as definições, e deve ser coerente aos conceitos que são

Coerência	definidos informalmente, tais como os descritos na documentação de linguagem natural, ou seja, se uma sentença que pode ser inferida a partir dos axiomas contradiz uma definição ou exemplo dado informalmente, em seguida, a ontologia é incoerente.
Estendibilidade	A ontologia deve ser projetada para antecipar o uso de um vocabulário compartilhado, oferecer uma base conceitual para uma série de tarefas previstas, e a representação deve ser trabalhada para que se possa ampliar e especializar a ontologia monotonamente
Mínimo Viés de Codificação	A conceituação deve ser especificada no nível de conhecimento sem depender de uma determinada codificação em nível de símbolo. Um viés de codificação resulta quando uma escolha de representação escolhida é feita exclusivamente para a conveniência de notação ou implementação. O viés de codificação deve ser minimizado, porque os agentes de compartilhamento de conhecimento podem ser implementados em representações diferentes.
Compromisso Ontológico Mínimo	Uma ontologia deve requerer o mínimo compromisso ontológico suficiente para suportar as atividades de conhecimento compartilhados entendidos, ou seja, definindo os termos que são essenciais

	para a comunicação de conhecimentos consistentes com a teoria.
--	--

**Tabela 7. Critérios de qualidade definidos por Gruber**

O ponto forte do trabalho de Gruber é que além dos critérios enunciados, ele mostra os critérios sendo aplicados em estudos de caso para o design de ontologias para engenheiros matemáticos e dados bibliográficos.

No entanto, mesmo aplicando os critérios em dois contextos, ainda há um esquecimento em estabelecer o passo-a-passo de cada critério para avaliar a ontologia

### 3.4. Avaliação de ontologias por Asunción Gómez-Pérez

O trabalho de Gómez-Pérez [Gómez-Pérez, 2001] propôs que o processo de avaliação é determinar o que define a ontologia corretamente, o que ela não faz, e o que ela faz de forma incorreta.

Além disso, assim como os demais, estabelece um critérios para avaliação de ontologias, como Completude, Concisão, Expansibilidade e Sensibilidade. A Tabela 8 mostra os critérios e suas descrições.

<b>Critério de qualidade</b>	<b>Descrição</b>
Completude	Diz respeito a provar o quão a ontologia é completa mas podemos provar a incompletude de uma definição individual, ou seja, se alguma definição está faltando na referência estabelecida
Concisão	Se a ontologia armazena qualquer definição desnecessária ou inútil.
Expansibilidade	É o que se refere ao esforço necessário para adicionar novas definições para uma ontologia

**Tabela 8. Os critérios de qualidade de Gómez-Pérez.**

O ponto forte deste trabalho é que existem critérios bem definidos, mostrados através de alguns exemplos através do *Ontolingua*.

No entanto, mesmo apresentando exemplos na apresentação de cada critério, fica difícil observar, um estabelecimento de passo-a-passo para a verificação de tais em uma ontologia qualquer, independente do uso do *Ontolingua*.

### 3.5. Avaliação de ontologias por Denny Vrandečić

O trabalho de Vrandečić [Vrandečić, 2009] organiza todos os critérios abordados pelos trabalhos vistos nas seções anteriores e insere questões a respeito dos critérios, para serem avaliados na ontologia. Dessa forma, o conjunto de critérios aborda desde o projeto da ontologia até a sua aptidão organizacional. Assim, temos:

- Precisão: Os axiomas cumprem com a especialidade de um ou mais usuários? A ontologia captura e representa corretamente os aspectos do mundo real?
- Adaptabilidade: A ontologia antecipa seu uso? Ela oferece uma base conceitual para uma série de tarefas previstas? A ontologia pode ser estendida e especializada monotonicamente, isto é, sem a necessidade de remover axiomas? Como a ontologia reage a pequenas mudanças em seus axiomas? A ontologia cumpre com procedimentos para extensão, integração e adaptação?
- Clareza: A ontologia comunica efetivamente o significado entendido dos termos? As definições são objetivas e independentes do contexto? A ontologia usa definições ou descrições parciais? As definições são documentadas? A ontologia é entendível?
- Completude: O domínio de interesse é apropriadamente coberto? As questões de competência são definidas? A ontologia pode respondê-las? A ontologia inclui todos os conceitos relevantes e suas representações léxicas?
- Eficiência Computacional: Qual fácil e bem-sucedidos os raciocinadores podem processar a ontologia? Quão rápido os serviços de raciocínio (satisfabilidade, classificação de instâncias, consultas, etc.) são aplicados a ontologia?
- Concisão: A ontologia inclui axiomas irrelevantes com o domínio a ser coberto? Ela inclui axiomas redundantes? Ela impõe um compromisso ontológico mínimo, isto é, especifica a teoria o mais fraco possível e define apenas os termos essenciais? Quão fraco são as premissas sobre a teoria filosófica subjacente da ontologia sobre a realidade?

- **Consistência:** Os axiomas levam a contradições? As descrições formais e informais da ontologia são consistentes, isto é, a documentação corresponde com a modelagem? A tradução do nível de conhecimento para a codificação mostra um mínimo viés de codificação? Todas as escolhas de representação foram feitas para a conveniência de notação ou implementação?
- **Aptidão Organizacional:** A ontologia é facilmente implantada na organização? As ferramentas baseadas em ontologia colocam restrições acima da ontologia? O processo utilizado foi adequado para a criação da ontologia? Foi certificada, se necessário? Ela cumpre requisitos legais? É de fácil acesso? Ela se alinha a outras ontologia já em uso? É bem partilhada entre os potenciais interessados?

Apesar de reunir todos os critérios dos trabalhos anteriores, ainda não há um estabelecimento de passo-a-passo para auxiliar o avaliador na tarefa de obter algum diagnóstico sobre a qualidade da ontologia que ele está verificando.

O trabalho deste autor é o principal trabalho relacionado com esta dissertação, pois, além de reunir todos critérios dos principais trabalhos sobre avaliação de ontologias, organizou-os em forma de perguntas, o que facilitou a adaptação para a metodologia GQM.

Os trabalhos que abordam qualidade de ontologias não estabelecem um padrão de critérios, ou seja, cada trabalho possui seus critérios, confundindo o avaliador, pois ele não sabe quais critérios ele deve escolher para a ontologia. Além disso, os autores tem dificuldade em explicar a execução de cada critério, e assim torna a avaliação bastante dependente de sua experiência. Indo mais além, o avaliador, mesmo sendo capaz de verificar cada critério, ele não possui uma métrica exata para qualificar a ontologia, tornando o processo muito subjetivo. Por fim, os trabalhos são antigos, e as reflexões sobre o tema tem parado.

O trabalho proposto tem como objetivo diminuir estes problemas, através de um passo-a-passo, utilizando uma abordagem conhecida pela comunidade científica, o GQM, para guiar o avaliador no processo de verificação da qualidade da ontologia, além de facilitar o estabelecimento de uma métrica, ou seja, de uma nota, para que torne a avaliação menos subjetiva.

## 4. A PROPOSTA

Este capítulo apresenta a principal contribuição desenvolvida nesta dissertação. Primeiramente, o capítulo apresentará os principais objetivos do método; logo em seguida as fases do método são apresentadas e como cada uma foi construída, com todas as justificativas.

### 4.1. A metodologia FOCA

Como foi visto nos capítulos anteriores, existem diversas barreiras que um ontologista enfrenta quando tenta modelar uma ontologia. A dificuldade de escolha de uma metodologia apropriada, assim como lidar com a complexidade que ela envolve; entender o domínio que será modelado; respeitar os papéis da representação do conhecimento; são motivos que fazem com que o ontologista, atualmente, seja motivado a avaliar a qualidade de sua ontologia.

Esta seção visa apresentar a metodologia proposta, chamada FOCA. A metodologia captura todos os critérios de qualidade de ontologias e suas respectivas questões, vistos na seção 3.5 para relacioná-las com os cinco papéis da representação do conhecimento, através da abordagem GQM. Esta abordagem foi escolhida devido a similaridade entre aspectos gerais de avaliação (como os papéis), aspectos mais específicos (como os critérios), e suas respectivas questões, além de ser uma abordagem bastante conhecida pela comunidade científica. A metodologia é composta por 3 fases: Verificação do tipo da ontologia; Verificação das questões e mensurar e Verificação da qualidade da ontologia. A Figura 8 resume a metodologia.

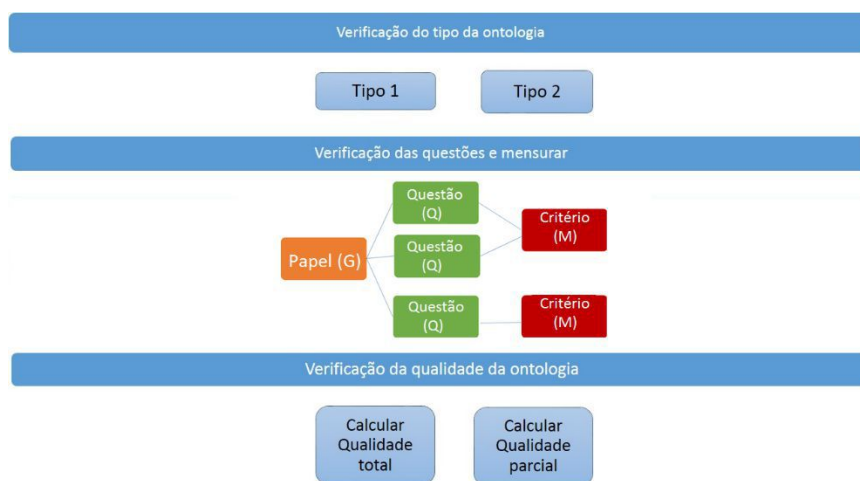


Figura 8. Metodologia proposta. Fonte: Elaborado pelo Autor.

O método foi desenvolvido tendo como base os cinco papéis da representação do conhecimento, os principais critérios de qualidade e suas respectivas questões (que sofreram algumas adaptações) de acordo com [Vrandečić, 2009], organizados de acordo a abordagem GQM. Os papéis, por serem o alvo da avaliação, foram considerados como os *GOALS* da abordagem GQM. Os critérios de qualidade, por serem as unidades de medição, foram considerados como *METRICs* da abordagem GQM. Finalmente, as questões que envolvem as métricas foram consideradas como *QUESTIONs* da abordagem GQM. Este trabalho não contém todas as questões do trabalho relacionado pois a verificação de todas tornaria o uso da metodologia extremamente exaustivo. A Tabela 9 mostra os papéis organizados em *GOALS*. A Tabela 10 mostra as questões organizadas em *QUESTIONs* e a Tabela 11 mostra os critérios organizados em *METRICs*.

<b>GOALS</b>	<b>GOALS x Papéis</b>
Goal 1	Verificar se a ontologia cumpre Substituto
Goal 2	Verificar se a ontologia cumpre Compromisso Ontológico
Goal 3	Verificar se a ontologia cumpre Raciocínio Inteligente
Goal 4	Verificar se a ontologia cumpre Computação Eficiente
Goal 5	Verificar se a ontologia cumpre Expressão Humana

**Tabela 9. Papéis da Representação do Conhecimento organizados em Goals.**

<b>QUESTIONs</b>	<b>QUESTIONs x Questões</b>
Question 1	As competências da ontologia foram definidas?
Question 2	A ontologia responde as competências definidas?
Question 3	A ontologia reusa outras ontologias?
Question 4	A ontologia impõe um compromisso ontológico mínimo?
Question 5	A ontologia impõe um compromisso ontológico máximo?
Question 6	As propriedades da ontologia são coerentes com o domínio?
Question 7	Existem axiomas contraditórios?
Question 8	Existem axiomas redundantes?
Question 9	O raciocinador traz erros de modelagem?
Question 10	O raciocinador executa de forma rápida?
Question 11	A documentação condiz com a modelagem?
Question 12	Os conceitos estão bem escritos?
Question 13	Há anotações na ontologia trazendo definições dos conceitos?

**Tabela 10. Questões organizadas em Questions.**

<b>METRICs</b>	<b>METRICs x Critérios</b>
Metric 1	Compleitude
Metric 2	Adaptabilidade
Metric 3	Concisão
Metric 4	Consistência
Metric 5	Eficiência Computacional
Metric 6	Clareza

**Tabela 11. Critérios organizados em Metrics.**

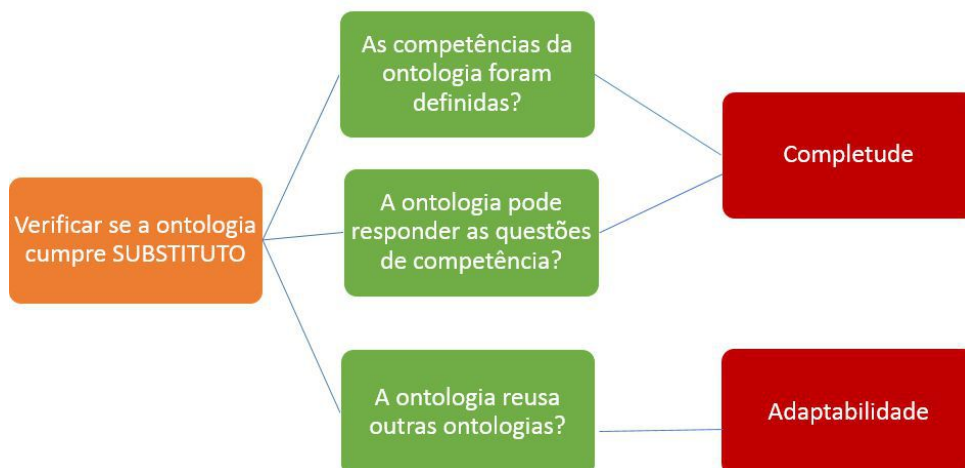
Após a organização de acordo com a abordagem GQM, foi refletida com quais GOALS as QUESTIONs estavam relacionadas. A correspondência foi feita entre a similaridade das questões com os papéis da representação do conhecimento. As justificativas das correspondências são as que seguem:

- As competências da ontologia foram definidas? **SUBSTITUTO**.
  - Justificativa: Esta questão envolve as competências da ontologia, que é a fase da engenharia de ontologias onde o ontologista vai refletir sobre o mundo real e, a partir dele, definir os principais termos e objetivos que envolvem o domínio.
- A ontologia responde as competências definidas? **SUBSTITUTO**.
  - Justificativa: Esta questão envolve a verificação dos termos trazidos do mundo real, assim como seus objetivos.
- A ontologia reusa outras ontologias? **SUBSTITUTO**.
  - Justificativa: Esta questão envolve a reutilização de termos já existentes, para a modelagem dos termos do mundo real.
- A ontologia impõe um compromisso ontológico mínimo? **COMPROMISSO ONTOLÓGICO**.
  - Justificativa: Esta questão envolve trazer a melhor representação de um domínio específico.
- A ontologia impõe um compromisso ontológico máximo? **COMPROMISSO ONTOLÓGICO**.
  - Justificativa: Esta questão envolve trazer a melhor representação de um domínio abstrato.



- As propriedades da ontologia são coerentes com o domínio? COMPROMISSO ONTOLÓGICO.
  - Justificativa: Esta questão envolve a verificação da coerência do que está sendo modelado, de acordo com o mundo real.
- Existem axiomas contraditórios? RACIOCÍNIO INTELIGENTE.
  - Justificativa: Esta questão envolve a verificação de um raciocínio correto por parte da ontologia.
- Existem axiomas redundantes? RACIOCÍNIO INTELIGENTE.
  - Justificativa: Esta questão envolve a verificação de um raciocínio correto por parte da ontologia.
- O raciocinador traz erros de modelagem? COMPUTAÇÃO EFICIENTE.
  - Justificativa: Esta questão envolve a verificação do bom desempenho computacional por parte da ontologia.
- O raciocinador executa de forma rápida? COMPUTAÇÃO EFICIENTE.
  - Justificativa: Esta questão envolve a verificação do bom desempenho computacional por parte da ontologia.
- A documentação condiz com a modelagem? EXPRESSÃO HUMANA.
  - Justificativa: Esta questão envolve a facilidade do entendimento da modelagem.
- Os conceitos estão bem escritos? EXPRESSÃO HUMANA.
  - Justificativa: Esta questão envolve a facilidade do entendimento da modelagem.
- Há anotações na ontologia trazendo definições dos conceitos? EXPRESSÃO HUMANA.
  - Justificativa: Esta questão envolve a facilidade do entendimento da modelagem.

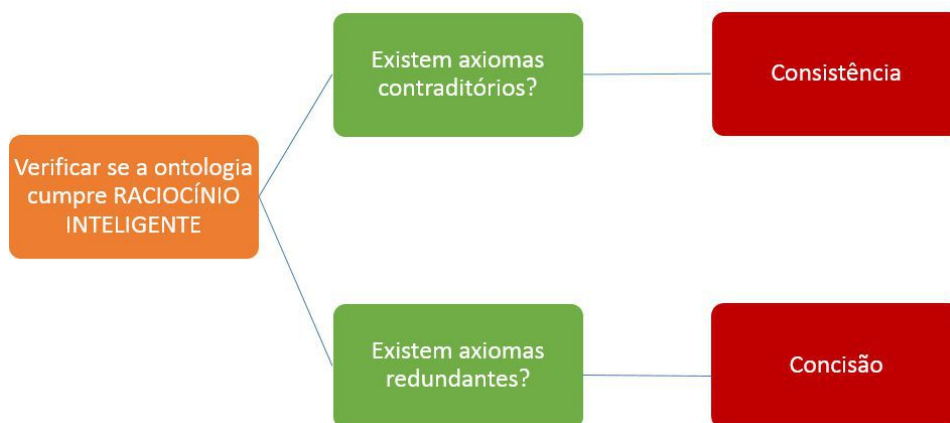
Após a reflexão e definição das correspondências entre GOALS e QUESTIONS, a construção da metodologia no formato GQM foi finalizada. As Figuras a seguir mostram toda a correspondência.



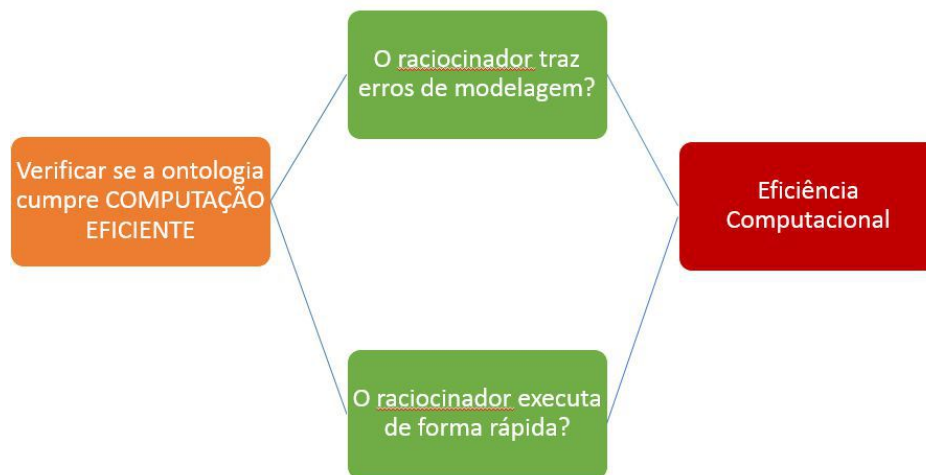
**Figura 9. Correspondência para o papel de Substituto**



**Figura 10. Correspondência para o papel de Compromisso Ontológico**



**Figura 11. Correspondência para o papel de Raciocínio Inteligente**



**Figura 12. Correspondência para o papel de Computação Eficiente**



**Figura 13. Correspondência para o papel de Expressão Humana**

As seções a seguir mostrarão a metodologia em seus três passos principais, com o objetivo de guiar o ontologista na avaliação de sua ontologia.

#### 4.2. Passos da metodologia

Esta seção abordará sobre os três principais passos da metodologia, assim como cada um deles foi construído. Os três passos são:

- **Verificação do tipo da ontologia:** É o passo em que o avaliador irá verificar qual o tipo da ontologia que será avaliada. A ontologia pode ser do tipo *task* or *domain* ou do tipo *application*;

- **Verificação das questões:** É o passo em que o avaliador irá verificar de fato a qualidade da ontologia, observando a modelagem dela e quantificando, ou seja, dando nota para a qualidade do que está sendo observado;
- **Verificação da qualidade da ontologia:** É o passo em que o avaliador irá, a partir das notas que foram dadas, calcular uma nota que representará a qualidade da ontologia.

As próximas seções apresentarão estes passos de uma forma mais detalhada, de forma que conduza melhor a execução de cada um deles.

#### 4.2.1. Passo 1: Verificação do tipo da ontologia

De acordo com [Vrandečić, 2009], uma ontologia não atende todos os critérios, pois existem critérios contraditórios, como por exemplo, nesta metodologia as questões que envolvem compromisso ontológico mínimo e compromisso ontológico máximo.

Estas duas questões só são verificadas de acordo com o tipo da ontologia. De acordo com [Guarino, 1998], uma ontologia pode ser do tipo *TOP ONTOLOGY*, *DOMAIN OR TASK ONTOLOGY* e *APPLICATION ONTOLOGY*. Desta forma, o avaliador, neste passo 1, deve verificar se a ontologia que está sendo avaliada se encaixa em:

- **Tipo 1: Ontologia de domínio ou tarefa:** são aquelas que descrevem, respectivamente, o vocabulário relacionado com um domínio genérico (como a medicina, ou automóveis) ou uma tarefa ou atividade genérica (como diagnosticar ou vender). Se a ontologia for do tipo 1, então a questão 4 não deverá ser verificada, pois ela só atende a domínios mais específicos, que não necessitam de abstração para a definição dos conceitos.
- **Tipo 2: Ontologia de aplicação:** são aquelas que descrevem conceitos, dependendo tanto de domínio específico quanto de uma tarefa, que muitas vezes são especializações dos termos de ambas as ontologias relacionadas (sistema de cadastro de alunos do curso de medicina na universidade federal de alagoas, por exemplo). Se a ontologia se encaixa neste tipo, então a questão 5 não deverá ser verificada.

É importante ressaltar que não foi considerada o tipo *TOP ONTOLOGY* pois, as questões que envolvem esta metodologia não são suficientes para a avaliação no nível de abstração que este tipo de ontologia possui.

#### 4.2.2. Passo 2: Verificação das questões

Após a verificação do tipo da ontologia, o avaliador está pronto para verificar de fato as questões seguindo o modelo GQM. Esta subseção mostra como deve ser feita a verificação de cada questão e como devem ser dadas as notas para a qualidade do que foi observado. Para a maioria das questões será necessária a utilização da ferramenta Protégé (4.0 ou superior)<sup>1</sup> para a verificação da modelagem. o avaliador terá 12 questões para responder, pois a questão 4 ou a questão 5 não será avaliada pois dependem do tipo da ontologia. A seguir, como o avaliador deve avaliar cada questão:

- Goal 1: Verificar se a ontologia cumpre Substituto.
- Question 1: As competências da ontologia foram definidas?
- Metric 1: Completude
- Como verificar: Primeiramente, verifique se existe algum documento que explique as competências da ontologia. Se não existir, a nota será automaticamente 0. Se existir, verifique:
  - Se define o objetivo da ontologia (por ex: esta ontologia modela o domínio de...)
    - Notas: 25, 50, 75, 100
  - Se define o público alvo da ontologia (por ex: esta ontologia será utilizada por...)
    - Notas: 25, 50, 75, 100
  - Se define cenários de uso desta ontologia (ou seja, em que situações ela deve ser utilizada).
    - Notas: 25, 50, 75, 100
- Goal 1: Verificar se a ontologia cumpre Substituto.
- Question 2: A ontologia responde as competências definidas?
- Metric 1: Completude
- Como verificar: Se você deu 0 na questão anterior, então as competências não foram definidas e não tem como verificar esta questão. Assim, a nota também será 0. Se as competências existem, verifique:
  - Se a ontologia responde o que foi definido nas competências.
    - Notas: 25, 50, 75, 100

---

<sup>1</sup> <http://protege.stanford.edu/>

- Goal 1: Verificar se a ontologia cumpre Substituto.
- Question 3: A ontologia reusa outras ontologias?
- Metric 2: Adaptabilidade
- Como verificar: Na ferramenta protege, na aba Active Ontology, em Ontology imports, verifique se existem importações. Se não existirem importações, a nota será 0. Se existirem importações, 100.

- Goal 2: Verificar se a ontologia cumpre Compromisso Ontológico.
- Question 4: A ontologia impõe um compromisso ontológico mínimo?
- Metric 3: Concisão
- Como verificar: Esta questão só deve ser verificada se a ontologia for do tipo 2. Se a ontologia for do tipo 1, pule para a próxima questão.

Nesta questão, verifique se a ontologia não utiliza muita abstração para definir os conceitos. Se a ontologia repleta de abstração (por exemplo: ontologia que modela o site do facebook não precisa estar definindo o que é um sistema de computador, o que é um computador, etc.) a nota será automaticamente 0. Se houver apenas algumas abstrações, dê uma nota entre estas: 25 (muita abstração), 50 (abstração razoável), 75 (quase nenhuma abstração), 100 (sem abstração).

- Goal 2: Verificar se a ontologia cumpre Compromisso Ontológico.
- Question 5: A ontologia impõe um compromisso ontológico máximo?
- Metric 3: Concisão
- Como verificar: Esta questão só deve ser verificada se a ontologia for do tipo 1. Se a ontologia for do tipo 2, pule para a próxima questão. Nesta questão, verifique se a ontologia utiliza conceitos primitivos para definir o domínio que está sendo avaliado (por exemplo, uma ontologia que modela pessoa, utilizaria os conceitos coisa -> ser vivo -> ser humano -> pessoa para definir o conceito de pessoa). Se a ontologia não utiliza nenhuma abstração, a nota será automaticamente 0. Se houver apenas algumas abstrações, dê uma nota entre estas: 25 (muito específico). 50 (algumas especificidades), 75 (algumas abstrações), 100 (muita abstração).

- Goal 2: Verificar se a ontologia cumpre Compromisso Ontológico.
  - Question 6: As propriedades da ontologia são coerentes com o domínio?
  - Metric 4: Consistência
  - Como verificar: Nesta questão, verifique se as classes e propriedades estão de acordo com o domínio modelado. Se a ontologia estiver repleta de incoerências (por exemplo, uma ontologia que modela o conceito carro tiver uma classe leão e uma propriedade quantidadeDePatatas, ou seja, não existe no domínio) a nota será automaticamente 0. Se houverem algumas incoerências, dê uma nota entre estas: 25, 50, 75. Se não houver nenhuma incoerência, 100.
- 
- Goal 3: Verificar se a ontologia cumpre Raciocínio Inteligente.
  - Question 7: Existem axiomas contraditórios?
  - Metric 4: Consistência
  - Como verificar: Nesta questão, verifique se as classes e propriedades (características funcional, transitiva, reflexiva etc) que estão coerentes com o domínio o contradizem (por exemplo: SerVivo ser subclasse de Pessoa numa ontologia que modela pessoa ou CPF não ser uma propriedade funcional, pois uma pessoa não pode ter mais de um CPF). Se a ontologia estiver repleta de contradições, a nota será automaticamente 0. Se houverem algumas contradições, dê uma nota entre estas: 25, 50, 75. Se não houver nenhuma contradição, 100.
- 
- Goal 3: Verificar se a ontologia cumpre Raciocínio Inteligente.
  - Question 8: Existem axiomas redundantes?
  - Metric 3: Concisão
  - Como verificar: Nesta questão, verifique se existem classes e propriedades que modelam a mesma coisa num mesmo sentido (por exemplo, utilizar manga para fruta e para roupa ao mesmo tempo, mouse para o animal e peça de computador ao mesmo tempo, etc). Se a ontologia estiver repleta de redundâncias, a nota será automaticamente 0. Se houverem algumas redundâncias, dê uma nota entre estas: 25, 50, 75. Se não houver nenhuma redundância, 100.
- 
- Goal 4: Verificar se a ontologia cumpre Computação Eficiente.

- Question 9: O raciocinador traz erros de modelagem?
- Metric 5: Eficiência Computacional
- Como verificar: Salve todos os seus registros até aqui. Nesta questão, verifique se, executando o raciocinador, retorna algum tipo de erro. Para isto, utilizando a ferramenta protege, clique em Reasoner, e abaixo de configure, selecione algum raciocinador (provavelmente algum deles estará selecionado, fact++, ou pellet, ou hermit) e clique em start reasoner. Se a ontologia estiver repleta de erros (ou se ocorrer algum travamento), a nota será automaticamente 0. Se houverem alguns erros, dê uma nota entre estas: 25, 50, 75. Se não houver nenhum erro, 100.
  
- Goal 4: Verificar se a ontologia cumpre Computação Eficiente.
- Question 10: O raciocinador executa de forma rápida?
- Metric 5: Eficiência Computacional
- Como verificar: Nesta questão, verifique se o raciocinador está executando de forma rápida. Se ele chega a travar algo, a nota será automaticamente 0. Se houver uma demora, dê uma nota entre estas: 25, 50, 75. Se executar de forma rápida, 100.
  
- Goal 5: Verificar se a ontologia cumpre Expressão Humana.
- Question 11: A documentação condiz com a modelagem?
- Metric 6: Clareza
- Como verificar: Nesta questão, verifique se existe a documentação da ontologia. Se a documentação não existir, a nota será automaticamente 0. Se a documentação existir, verifique:
  - Se os termos escritos na documentação estão iguais ao da modelagem.
    - Notas: 25, 50, 75, 100
  - Se a documentação explica o que é cada termo e justifica cada detalhe da modelagem.
    - Notas: 25, 50, 75, 100
  
- Goal 5: Verificar se a ontologia cumpre Expressão Humana.
- Question 12: Os conceitos estão bem escritos?
- Metric 6: Clareza
- Como verificar: Nesta questão, verifique se as classes e propriedades da ontologia estão escritas de forma entendíveis e escritas corretamente (de acordo com o português ou outro



idioma utilizado). Se a ontologia estiver muito difícil de entender lendo os conceitos, a nota será automaticamente 0. Se houverem alguns erros ou mistura de idiomas, dê uma nota entre estas: 25, 50, 75. Se a ontologia estiver bem escrita, 100.

- Goal 5: Verificar se a ontologia cumpre Expressão Humana.
- Question 13: Há anotações na ontologia trazendo definições dos conceitos?
- Metric 6: Clareza
- Como verificar: Nesta questão, verifique se existem anotações que trazem as definições dos conceitos modelados na ontologia. Para isto, utilizando a ferramenta protege, clique no conceito (classe ou propriedade) e verifique no painel do lado, chamado Annotations, se existe algo que explique o que é aquele conceito na modelagem. Se não existir nenhuma anotação, a nota será automaticamente 0. Se houverem algumas anotações, dê uma nota entre estas: 25, 50, 75. Se todos os conceitos possuem anotações, 100.

#### 4.2.3. Passo 3: Verificação da qualidade da ontologia

Após o avaliador verificar todas as questões no passo 2 e dar as notas correspondentes ao que foi observado, neste passo ele deve calcular a qualidade da ontologia. A qualidade da ontologia é calculada através da seguinte fórmula:

$$\hat{\mu}_i = \frac{\exp\{-0.44 + 0.03(Cov_S \times Sb)_i + 0.02(Cov_C \times Cr)_i + 0.01(Cov_R \times Rc)_i + 0.02(Cov_{Cp} \times Cp)_i - 0.66GExp_i - 25(0.1 \times NI)_i\}}{1 + \exp\{-0.44 + 0.03(Cov_S \times Sb)_i + 0.02(Cov_C \times Cr)_i + 0.01(Cov_R \times Rc)_i + 0.02(Cov_{Cp} \times Cp)_i - 0.66GExp_i - 25(0.1 \times NI)_i\}}$$

Através dessa fórmula, a qualidade da ontologia pode calculada de duas formas:

- **Calcular Qualidade Total:** O avaliador “i” calcula a qualidade da ontologia considerando todos os cinco papéis da representação do conhecimento. Para calcular a qualidade total:
  - $Cov_s$  é a média das notas obtidas no GOAL 1. Observe que QUESTION 1 envolve três subquestões, a nota de QUESTION 1 é a média entre elas. Finalmente,  $Cov_s$  é a média entre QUESTION 1, QUESTION 2 e QUESTION 3;

- $Cov_c$  é a média das notas obtidas no GOAL 2. Observe que a média será entre QUESTION 5 e QUESTION 6 se a ontologia for do tipo 1 e a média será entre QUESTION 4 e QUESTION 6 se a ontologia for do tipo 2.
  - $Cov_r$  é a média das notas obtidas no GOAL 3;
  - $Cov_{cp}$  é a média das notas obtidas no GOAL 4;
  - $GExp_i$  é a variável que corresponde a experiência do avaliador. Se ele se considera que possui vasta ou grande experiência em ontologias, o valor de  $GExp_i$  é 1, se não, 0;
  - $Nl$  assume o valor 1 apenas se em algum GOAL foi impossível o avaliador responder TODAS as QUESTIONS;
  - $Sb$ ,  $Cr$ ,  $Rc$ ,  $Cp$  são iguais a 1, pois para o score global estão sendo considerados todos os papeis.
- **Calcular Qualidade Parcial:** O avaliador “i” calcula a qualidade da ontologia considerando apenas alguns papeis (por exemplo: considerar apenas Substituto e Expressão Humana). Para calcular a qualidade parcial:
    - Se o avaliador estiver interessado apenas em avaliar a ontologia **em termos do papel Substituto**, ele irá considerar:
      - $Cov_s$ , a média das notas obtidas no GOAL 1;
      - $Sb = 1$ , pois o avaliador está considerando apenas o GOAL 1;
      - $Cov_c$ ,  $Cov_r$ ,  $Cov_{cp}$ ,  $Cr$ ,  $Rc$  e  $Cp$  iguais a 0, pois o avaliador não considera os GOALs 2, 3, 4 e 5;
      - Responder 1 ou 0 para  $GExp_i$  e  $Nl$ .
    - Se o avaliador estiver interessado apenas em avaliar a ontologia **em termos do papel Compromisso Ontológico**, ele irá considerar:
      - $Cov_c$ , a média das notas obtidas no GOAL 2;
      - $Cr = 1$ , pois o avaliador está considerando apenas o GOAL 2;
      - $Cov_s$ ,  $Cov_r$ ,  $Cov_{cp}$ ,  $Sb$ ,  $Rc$  e  $Cp$  iguais a 0, pois o avaliador não considera os GOALs 1, 3, 4 e 5;
      - Responder 1 ou 0 para  $GExp_i$  e  $Nl$ .
    - Se o avaliador estiver interessado apenas em avaliar a ontologia **em termos do papel Raciocínio Inteligente**, ele irá considerar:

- $Cov_r$ , a média das notas obtidas no GOAL 3;
  - $Rc = 1$ , pois o avaliador está considerando apenas o GOAL 3;
  - $Cov_s$ ,  $Cov_c$ ,  $Cov_{cp}$ ,  $Sb$ ,  $Cr$  e  $Cp$  iguais a 0, pois o avaliador não considera os GOALs 1, 2, 4 e 5;
  - Responder 1 ou 0 para  $GExp_i$  e  $Nl$ .
- Se o avaliador estiver interessado apenas em avaliar a ontologia **em termos do papel Computação Eficiente**, ele irá considerar:
- $Cov_{cp}$ , a média das notas obtidas no GOAL 4;
  - $Cp = 1$ , pois o avaliador está considerando apenas o GOAL 4;
  - $Cov_s$ ,  $Cov_c$ ,  $Cov_r$ ,  $Sb$ ,  $Cr$  e  $Rc$  iguais a 0, pois o avaliador não considera os GOALs 1, 2, 3 e 5;
  - Responder 1 ou 0 para  $GExp_i$  e  $Nl$ .
- Se o avaliador estiver interessado apenas em avaliar a ontologia **em termos do papel Expressão Humana**, ele irá considerar:
- $Cov_s$ ,  $Cov_c$ ,  $Cov_r$ ,  $Cov_{cp}$ ,  $Sb$ ,  $Cr$ ,  $Rc$ ,  $Cp$  iguais a 0, pois o avaliador não considera os GOALs 1, 2, 3 e 4;
  - Responder 1 ou 0 para  $GExp_i$  e  $Nl$ .

É possível observar que não existem variáveis para serem inseridos os valores referentes ao papel de Expressão Humana na fórmula apresentada. Existem dois motivos para isto, o motivo ontológico e o motivo matemático. O motivo ontológico é que Expressão Humana está embutido em todos os outros papeis. Se a ontologia não possui documentação, os termos são de difícil entendimento e não possui anotações, será muito difícil responder as questões referentes aos outros papeis. Por exemplo, se uma ontologia modela o conceito “Bicicleta” e em sua modelagem os conceitos são “G”, “Ro”, “A”, “hGu”, será praticamente impossível responder a pergunta: “As propriedades da ontologia são coerentes com o domínio?”, pois o avaliador sequer entende o que os termos querem dizer. Assim, Expressão Humana fica restrita a experiência do avaliador, ou seja, o avaliador ainda pode entender termos mal-modelados (escritos e representados) de acordo com sua experiência, por isso que para Expressão Humana apenas as variáveis  $GExp_i$  e  $Nl$  são consideradas. O motivo matemático é melhor explicado na seção de Análise dos dados, pois a fórmula foi obtida após a realização do experimento que

objetivou a validação da metodologia. A obtenção da fórmula, bem como a validação da metodologia é explicada no próximo capítulo.

## 5. VALIDAÇÃO – EXPERIMENTO

Assim como toda pesquisa científica, é necessária a realização de um experimento para provar a relevância do que está sendo proposto.

De acordo com [Travassos, 2002], Experimentação é o centro do processo científico. Somente experimentos verificam as teorias. Somente experimentos podem explorar os fatores críticos e dar luz ao fenômeno novo para que as teorias possam ser formuladas e corrigidas. Experimentação oferece o modo sistemático, disciplinado, computável e controlado para avaliação da atividade humana. Novos métodos, técnicas, linguagens e ferramentas não deveriam apenas ser sugeridos, publicados ou apresentados para venda sem experimentação e validação. Portanto, é preciso avaliar novas intenções e sugestões em comparação com as existentes.

As próximas seções e subseções mostrarão todo o projeto do experimento feito para a validação desta metodologia, assim como toda sua narrativa, análise e conclusões finais.

### 5.1. Projeto de experimento

Esta seção mostrará todo o projeto do experimento realizado para a validação da metodologia proposta.

A primeira etapa para a realização de um experimento é definir bem o problema que envolve o tema, explicitando uma contextualização, problema de negócio e problema técnico; logo em seguida são mostrados os objetivos de investigação do experimento. Após, será mostrado todo o planejamento do experimento, isto é, questões de pesquisa e hipóteses, fatores e variáveis de resposta, entre outros elementos. Finalmente, será mostrado o plano de execução do experimento e as análises de ameaças à validade.

#### 5.1.1. Definição do problema

Esta seção tem como objetivo apresentar o problema que envolve o experimento. Este experimento está no contexto de verificar a qualidade de uma ontologia observando questões, seguindo um passo-a-passo. A verificação da ontologia neste trabalho envolve informações como: o tipo da ontologia, seu uso, sua função de representar o mundo real, termos bem representados, a eficiência computacional, fácil entendimento da modelagem, entre outros fatores. Se uma metodologia não estabelece um passo-a-passo ou explica

como realizar a verificação de cada uma dessas especificidades, a avaliação de ontologias se torna muito difícil, custosa e imprecisa. Baseado neste problema, a metodologia proposta busca corresponder estas especificidades com algo mais genérico, que são os tipos de ontologia apresentados por Guarino (como visto anteriormente) e os cinco papéis da representação do conhecimento, para que o avaliador saiba que “função do conhecimento” está sendo representada ao verificar cada especificidade. Desta forma, a correspondência fica:

- O tipo da ontologia (*domain or task* ou *application*);
- O uso da ontologia – através das competências, principais termos e reuso em substituto;
- A sua função de representar o mundo real (compromissos máximo e mínimo em compromisso ontológico);
- Termos bem representados (contradições e redundâncias em raciocínio inteligente);
- A eficiência computacional (execução rápida e bem sucedida dos raciocinadores em computação eficiente);
- O fácil entendimento da modelagem (documentação, escrita e anotações em expressão humana).

No entanto, estabelecer um passo-a-passo, bem como corresponder estas especificidades nos levou a ter algumas dúvidas, como “A metodologia conduz bem o avaliador?” ou ainda “A metodologia depende ou independe da experiência (em ontologias) do avaliador, ou seja, a metodologia consegue conduzir um avaliador com muita experiência e um com baixa experiência para resultados parecidos?”. Desta forma, as notas dadas pelos avaliadores, assim como a experiência deles são as principais métricas para a realização do experimento.

A partir das perguntas, métricas e problema de negócio apresentado, é necessário refletir sobre os problemas técnicos que estão envolvidos. Para resolver o problema de negócio visto anteriormente, ou parcialmente, é importante validar se é adequado fazer uma correspondência entre papéis da representação de conhecimento e critérios de qualidade para a avaliação de ontologias. O problema técnico nos leva a refletir sobre algumas perguntas: “Estabelecer um passo-a-passo diminui a subjetividade na verificação

das questões?” ou “O método é adequado para pessoas com qualquer experiência em ontologias?”.

As respostas para os problemas apresentados anteriormente serão apresentadas ao longo deste capítulo. Para isto, é necessário estabelecer hipóteses e verificá-las, e verificar se a metodologia responde positivamente ou negativamente os problemas.

### 5.1.2. Objetivos de investigação

A validação foi do tipo Experimental, onde se sugere o modelo, desenvolve o método quantitativo ou qualitativo, aplica um experimento, mede e analisa, avalia o modelo e repete o processo. O experimento tem como objetivo avaliar o comportamento da metodologia para avaliação de ontologia utilizando princípios da representação do conhecimento. O experimento considerou usuários de diferentes experiências em ontologia, avaliando ontologias do mesmo domínio.

Ao final do experimento, as notas dadas foram coletadas para a construção de um modelo de regressão para que as notas dadas pelos avaliadores que não utilizaram o passo-a-passo sejam explicadas pelas notas dadas pelos mesmos avaliadores no uso da metodologia, o que será detalhado nas sessões seguintes.

Formalmente, o objetivo da investigação pode ser definido no formato GQM como **analisar** a metodologia **com a intenção de** avaliá-la a respeito de sua validade, através da criação de um modelo de regressão e possibilidade de criação de um score global e parcial **do ponto de vista** dos avaliadores de diferentes experiências avaliando ontologias **no contexto** do *sistema de currículos Lattes* e ferramenta *Protégé*.

### 5.1.3. Planejamento do experimento

Esta seção abordará todo o planejamento do experimento. Aqui, encontram-se a definição da questão de pesquisa e derivação de hipóteses, a seleção de variáveis dependentes e independentes, a identificação da unidade experimental e a seleção do modelo experimental que será utilizado, entre outros detalhes envolvidos, que serão mostrados nas subseções seguintes.

Em resumo, o experimento fará uma análise das notas dadas pelos participantes, divididos em categorias de experiências de ontologias, sem um passo-a-passo, ou seja,

apresentando apenas as questões e em seguida, com todo o passo-a-passo da metodologia. O contexto das ontologias é a modelagem do *sistema de currículos Lattes*.

#### 5.1.3.1. Questões de pesquisa e hipóteses

O objetivo deste experimento é propor um modelo de regressão em que o escore subjetivo (aquele em que os participantes não utilizam a metodologia) possa ser explicado pelo escore objetivo (aquele em que os participantes utilizam a metodologia) fornecido pela metodologia proposta para qualificação das ontologias. O modelo de regressão será útil para a construção de um escore global (aquele que considera todos os papéis da representação do conhecimento) ou parcial (aquele que considera apenas alguns papéis) para a qualidade da ontologia. Se for possível a criação destes escores, a metodologia será validada. Desta forma, as questões de pesquisa e hipóteses envolvem a validação da metodologia e a possibilidade de criação dos escores. Assim, as hipóteses são:

- A metodologia é válida?
- É possível obter um escore global para qualificar ontologias com base na metodologia proposta?
- É possível obter um escore parcial para cada GOAL com base na metodologia proposta?

#### 5.1.3.2. Fatores e variáveis de resposta

Após a definição de nossas hipóteses de pesquisa, o próximo passo do planejamento do experimento é a definição das variáveis contidas nele. Basicamente, temos dois tipos de variáveis: Primeiro, as variáveis independentes, também chamadas de fatores, são aquelas que queremos avaliar; Segundo, as variáveis dependentes, são o caso das métricas de avaliação.

Como variável independente temos:

- Metodologia de avaliação de ontologias: A metodologia de avaliação que será utilizada em determinado momento;
- Experiência dos avaliadores: O tipo da experiência dos participantes do experimento que irão influenciar na análise dos dados.

Os níveis dos fatores apresentados são definidos na Tabela 12.



<b>Fator</b>	<b>Nível</b>	<b>Descrição</b>
Metodologia	M1	Não-utilização da metodologia
Metodologia	M2	Utilização da metodologia
Experiência dos avaliadores	E1	Vasta
Experiência dos avaliadores	E2	Grande
Experiência dos avaliadores	E3	Média
Experiência dos avaliadores	E4	Pouca
Experiência dos avaliadores	E5	Pouquíssima ou Nenhuma

**Tabela 12. Níveis dos fatores. Fonte: Elaborado pelo autor.**

As variáveis dependentes são definidas a seguir:

- Notas sem a utilização de metodologia: As notas que os participantes deram observando apenas as perguntas, sem o passo-a-passo;
- Notas com a utilização de metodologia: As notas que os participantes deram utilizando toda a metodologia.

### 5.1.3.3. Unidades de experimento

O objetivo deste experimento é observar se a metodologia é capaz de explicar as notas subjetivas (sem a utilização da metodologia). Nesse projeto de experimento, após definir as questões, hipóteses e fatores e variáveis de resposta, precisamos definir qual será a unidade de experimento utilizada.

No cenário deste experimento, a unidade experimental ou, mais precisamente, as unidades experimentais serão quatro ontologias que modelam o *sistema de currículos Lattes*. As quatro ontologias foram modeladas por grupos de no máximo quatro pessoas (que não são as mesmas que participaram da avaliação sem/com a metodologia). Três das quatro ontologias foram modeladas com a utilização de metodologias para engenharia de ontologias. As metodologias utilizadas foram *HCOME*, *Grüninger and Fox methodology* e *Methontology*. Uma das ontologias foi modelada sem a utilização de metodologia.

O *currículo Lattes* é o formulário eletrônico do MCT, CNPq, FINEP e CAPES/MEC para o cadastro de dados curriculares de pesquisadores e de usuários em geral com o objetivo de criar um instrumento curricular único [Bras, 2003]. Como todos os participantes, desde a modelagem até a avaliação de ontologias, são pesquisadores, o sistema de currículos Lattes foi motivado a ser escolhido.

#### 5.1.3.4. Plano de execução

Esta seção descreve como será a execução do experimento que foi planejado e descrito ao longo das seções anteriores. A execução do experimento envolve os seguintes passos:

1. **Modelagem das ontologias:** As ontologias foram modeladas por grupos de no máximo quatro pessoas. Três das ontologias foram implementadas pelas metodologias *HCOME*, *Grüninger and Fox methodology* e *Methontology*. O domínio escolhido foi o Sistema de Currículos Lattes. As quatro ontologias podem ser melhor entendidas no Apêndice A;
2. **Seleção dos participantes:** Seis participantes foram escolhidos para a realização do experimento. Os participantes foram escolhidos através de suas experiências, que consistem no tempo em vivência em ontologias, publicações e modelagem. Dois foram considerados como vasta experiência, um como grande experiência, um como média experiência, um como pouca experiência e um como pouquíssima ou nenhuma experiência;
3. **Preparação dos documentos para avaliação:** Dois documentos foram preparados para cada participante utilizar durante a avaliação. Um deles, chamado de *Avaliação de ontologias SEM MODELO*, que continha apenas questões (sem nenhum passo-a-passo) e outro chamando *Avaliação de ontologias COM MODELO*, que continha a metodologia em sua íntegra. Estes dois documentos podem ser visualizados no Apêndice B;
4. **Execução do experimento:** Cada participante foi instruído a avaliar as quatro ontologias de acordo com o documento *Avaliação de ontologias SEM MODELO* e logo após avaliar as ontologias com o documento *Avaliação de ontologias COM MODELO*.
5. **Coleta de dados:** Os participantes foram instruídos a enviar as notas de cada documento para o experimentador. O experimentador reuniu todas as notas em uma planilha.

6. **Análise dos dados:** Foi feita uma análise estatística dos dados, com a criação de um histograma das notas (também chamadas de escores) e análise de Boxplots desses escores para responder as hipóteses.

Após a execução dos passos anteriores, os resultados poderão ser obtidos e comparados entre si. A análise dos dados bem como as principais conclusões obtidas estão apresentadas nas Seções 5.3 e 5.4, respectivamente.

#### 5.1.3.5. Análise de ameaças à validade

Embora o experimento tenha sido realizado com grande cautela para minimizar as ameaças a validade, as quais possam comprometer conclusões, existem algumas que devem ser mencionadas:

- Um único domínio modelado pelas ontologias: A utilização de apenas um domínio pode ser considerada uma ameaça a validade, pois este pode ser de conhecimento maior de parte dos avaliadores do que outros. Uma diversidade maior de domínios poderia equilibrar este caso;
- Número baixo de ontologias: Apesar serem muitas observações, quatro ontologias pode ser considerado um número baixo de ontologias. Talvez um número maior, conseqüentemente mais notas para serem coletadas, os resultados poderiam ser mais precisos;
- Avaliadores escolhidos subjetivamente: Os avaliadores foram escolhidos avulsamente observado as experiências. Nada foi medido para classificar corretamente os participantes, o que pode comprometer os resultados;
- Cansaço dos avaliadores durante o experimento: O experimento envolveu a verificação de quatro ontologias não utilizando a metodologia e, novamente as mesmas ontologias foram verificadas utilizando a metodologia, ou seja, foram feitas oito verificações, tornando o experimento um pouco exaustivo, o que pode ter influenciado em algumas das notas dos participantes.

## 5.2. Narrativa de execução do experimento

Esta seção descreve o processo de execução do experimento.

Primeiramente, as ontologias foram modeladas em grupos de no máximo quatro pessoas, distribuídas com diferentes experiências em ontologias, niveladas numa disciplina de ontologias do Programa de Mestrado em Informática da Universidade Federal de Alagoas.

O experimento se caracterizou por um estudo às cegas. Foram selecionados seis avaliadores, dos quais dois podem ser classificados reconhecidamente com vasta experiência no estudo de ontologias, um com grande experiência, um com média experiência, um com pouca e um último com pouquíssima ou nenhuma experiência. Foi solicitado a esses pesquisadores que avaliassem quatro ontologias usando critérios do seu próprio conhecimento, desde que atribuíssem um escore entre 0.001 e 0.990 (que continham em *Avaliação de ontologias SEM MODELO*). Quanto maior o escore maior a qualificação da ontologia. Uma das prerrogativas do desenho de experimento é que o avaliador desconhecesse a metodologia concreta proposta para a qualificação de ontologias. Os avaliadores analisaram e qualificaram de acordo com que havia em *Avaliação de ontologias SEM MODELO*, que eram apenas as treze questões, sem nenhum passo-a-passo. Usando critérios baseados no seu nível de conhecimento sobre ontologias, ou seja livre da metodologia proposta, os pesquisadores atribuíram os escores do tipo (0.001,0.990) para as questões. O escore final de cada GOAL foi a média entre as questions que definiam a GOAL. Desta forma temos **escores subjetivos** atribuídos pelos seis avaliadores para as cinco GOALS de quatro ontologias disponibilizadas.

Em uma segunda etapa do experimento, foi apresentado aos avaliadores a metodologia concreta para qualificar as ontologias seguindo as instruções da seção 4.3.2, em que a variável qualificadora também é um escore, que pode assumir os valores 0, 25, 50, 75 e 100 (que continham em *Avaliação de ontologias COM MODELO*). Quanto maior o valor do escore maior a qualificação da ontologia. Com base na seção 4.3.2., nota-se que, para atribuir um escore dentro de cada GOAL e para cada QUESTION é fornecido ao avaliador critérios (como um passo-a-passo) que devem ser verificados. Esses critérios recebem uma “nota” entre 0, 25, 50, 75 e 100. O escore final de cada question é a média entre os escores atribuídos a esse(s) critério(s). Finalmente, o escore final de cada GOAL é a média dos escores das questions que compõem essa GOAL. Assim, temos **escores objetivos** gerados pelo modelo preenchido pelos seis avaliadores para as cinco GOALS considerando quatro ontologias disponibilizadas.

É importante ressaltar que caso o avaliador não soubesse qualificar subjetivamente, ou seja, sem a metodologia, não era necessário estabelecer uma nota, e sim deixar a questão sem resposta. Nesse caso, para a coleta de dados, a questão que não foi respondida é excluída do experimento. No entanto, se o avaliador não soubesse responder todas as questões de uma GOAL utilizando a metodologia, a média final da GOAL assumirá o valor “0.1” e no modelo de regressão será considerada uma variável indicadora para esse caso, o que possibilita o avaliador considerar que não sabe atribuir um escore para determinada GOAL.

### 5.3. Análise dos dados

Esta seção apresentará toda a análise dos dados coletados ao final do experimento, com o objetivo de analisá-los afim de responder as hipóteses definidas. Para um melhor entendimento desta seção, será fornecida uma fundamentação teórica sobre o método estatístico utilizado. O método a ser utilizado é a modelagem linear de regressão.

#### 5.3.1. Modelos Lineares de Regressão

Estamos interessados em modelar o comportamento de uma variável através de um modelo matemático. Porém, temos que ter em mente que os modelos matemáticos são uma representação simplificada da realidade de forma que existem fatores que não podem ser controlados ou são desconhecidos, o que caracteriza a natureza aleatória da nossa variável de interesse. Outro fator a considerar no modelo é quais variáveis podem compor o modelo matemático como o objetivo de explicar a nossa variável de interesse. Estamos falando aqui de: variável de interesse ou variável resposta, variáveis que tentam explicar essa resposta ou covariadas e o componente aleatório da nossa resposta. Trata-se de um modelo de regressão estatístico. Neste contexto que se inclui o modelo de regressão linear, considere o seguinte modelo linear:

(1.1)

$$y = X\beta + \epsilon =$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \underbrace{\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}}_{(x_1 \quad x_2 \quad \dots \quad x_k)} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Podemos ver que:

- $y$  é um vetor de  $n$  observações da variável aleatória dependente, ou ainda, nossa variável resposta;
- $X$  é uma matriz  $n \times k$  formada pelas covariadas. Note que cada coluna de  $X$  é um conjunto de  $n$  observações da covariada  $x_t, t = 1, \dots, k$ . É importante ressaltar que  $X$  não é variável aleatória, ela é observada e fixa;
- Ainda temos que,  $\beta$  é um vetor de  $k$  parâmetros também fixos e desconhecidos (não são variáveis aleatórias);
- Finalmente,  $\epsilon$  é um vetor de  $n$  erros aleatórios com média zero  $(E)(\epsilon_i) = 0$  e variância constante ao longo das observações, isto é,  $var(\epsilon_i) = \sigma^2$  para todo  $i = 1, \dots, n$

O principal objetivo do modelo de regressão é a explicação da resposta de forma que o que não conseguimos explicar esteja contido no erro aleatório  $\epsilon$ . Porém, o erro deve ser bem pequeno, ou seja, esperamos que ele seja zero, ou que em média ele seja zero. Assim, surge uma das principais suposições de modelos lineares de regressão é que

(1.2)

$$E(\epsilon) = \mu_\epsilon = 0.$$

Na expressão (1.2), o termo  $E(\epsilon)$  significa valor esperado da variável aleatória  $\epsilon$ , que é o valor médio desta variável aleatória. Reafirmando que em um modelo de regressão o valor esperado do erro é supostamente zero, isto é, em média o erro é igual a zero. Com base na suposição em (1.2) e usando (1.1), temos que

$$E(y) = E(X\beta) + E(\epsilon) \Leftrightarrow E(y) = X\beta \Leftrightarrow \mu = X\beta.$$

Na prática, temos que o nosso modelo final é

$$\mu = X\beta.$$

Note que  $E(X\beta) = X\beta$  por que nem  $X$  nem  $\beta$  são variáveis aleatórias e valor esperado de uma constante é própria constante. Quanto a variância temos

$$\text{var}(y) = \underbrace{\text{var}(X\beta)}_0 + \underbrace{\text{var}(\epsilon)}_{\sigma^2} \Leftrightarrow \text{var}(y) = \sigma^2.$$

Assim,  $\text{var}(X\beta) = 0$  pois nem  $X$  nem  $\beta$  são variáveis aleatórias, pois variância de uma constante é 0. Podemos representar nosso modelo de regressão em (1.1) considerando a  $i$ -ésima observação como

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n.$$

Para conhecermos de fato o modelo acima, precisamos estimar  $\beta_1, \beta_2, \dots, \beta_k$ , que são denominados **coeficientes do modelo de regressão**. Tipicamente fazemos isto usando o método de máxima verossimilhança [Lehmann e Casella, 1998]. Assim, obtemos  $\hat{\mu}_i$  quando obtemos  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$  tal que,

$$\hat{\mu}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \dots + \hat{\beta}_k x_{ik}, \quad i = 1, \dots, n.$$

Agora, podemos obter estimativas para  $\mu_i$ . Note que em um modelo de regressão o que nos interessa é a distribuição de  $y$ , sua média, sua variância e o tipo de distribuição de probabilidades que a variável aleatória  $y$  (nossa resposta) segue.

Sabemos que a distribuição mais utilizada é a distribuição normal. No entanto, nem sempre essa é a distribuição adequada para os dados que estamos trabalhando. Se  $\mu$  é a média da variável resposta, pode assumir tanto valores positivos quanto valores negativos, então se justifica pensar na distribuição normal; se os dados são positivos, ou seja,  $\mu$  só pode assumir positivos a primeira distribuição que pensamos para construir o modelo de regressão é a gama; se  $y \in (0,1)$ , ou seja, se nossos dados assumem valores

contínuos e restritos entre zero e um, e a média da variável resposta também se encontra entre zero e um,  $\mu \in (0,1)$ , podemos pensar na distribuição beta.

Mas, se consideramos outras distribuições de probabilidades além da distribuição normal, precisamos generalizar o modelo linear. Isto é feito considerando a expressão abaixo.

$$g(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}.$$

Em que  $g(\mu_i)$  é conhecida como **função de ligação** e é utilizada para conectar a média da variável resposta e o modelo envolvendo as covariadas e os  $\beta$ 's. Mas, por que precisamos desta função de ligação? Primeiro vamos começar com o modelo normal. Se  $y_i \sim N(\mu_i, \sigma^2)$ , segue que  $g(\mu_i) = \mu_i$  é uma função de ligação adequada, ou seja,  $g(u) = u$  que chamamos de função identidade pode ser usada no modelo normal, NÃO precisamos fazer qualquer transformação na média. Isto acontece porque na distribuição normal  $\mu \in (-\infty, +\infty) = \mathbb{R}$  ( a média da população normal pode assumir qualquer valor Real). A mesma coisa deve acontecer com sua estimativa:  $\hat{\mu}_i$  Ou seja, obrigatoriamente temos que obter  $\hat{\mu}_i$  tal que,  $\hat{\mu}_i \in (-\infty, +\infty) = \mathbb{R}$ .

Então, quando trabalhamos com a distribuição normal,  $\hat{\mu}_i$  pode assumir qualquer valor Real também. Mas,

$$\hat{\mu}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \dots + \hat{\beta}_k x_{ik}, \quad i = 1, \dots, n.$$

Ou seja, o valor de  $\hat{\mu}_i$  depende dos valores  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$  pois os valores de  $x_{i2}, x_{i3}, \dots, x_{ik}$  são valores fixos conhecidos das covariadas. Só os valores de  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$  podem variar. Como  $\hat{\mu}_i$  pode assumir qualquer valor Real, não existem restrições para  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$ , eles também podem assumir qualquer valor. Assim, os valores que  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$  podem assumir estão livres, qualquer que sejam os valores de  $x_{it}, i = 1, \dots, n; t = 1, \dots, k$ .

Isto não acontece por exemplo se a variável resposta segue uma distribuição gama. Neste caso, a variável resposta só assume valores reais positivos, isto é,  $y \in (0, +\infty)$ . E ocorre o mesmo com sua média  $\mu \in (0, +\infty) = \mathbb{R}^+$  e DEVE acontecer também com  $\hat{\mu}$ , ou seja, **tem que acontecer:**  $\hat{\mu} \in (0, +\infty) = \mathbb{R}^+$ . Com esta restrição, os  $\hat{\beta}$ 's não estão livres, pois temos que garantir que  $X\hat{\beta}$  só assuma valores reais positivos. Não podemos



utilizar  $X$  (fixa e conhecida), então teríamos que realizar um processo de estimação dos  $\beta$ 's com restrição, para garantir que  $\hat{\beta}X \in (0, +\infty) = IR^+$ . Este processo pode ser bastante complicado.

A alternativa é aplicar uma função  $g$  em  $\mu_i$ , uma transformação, de forma que  $g(\mu_i) \in (-\infty, +\infty) = IR$ . Daí então os  $\hat{\beta}$ 's estarão liberados. Seja  $y$  uma variável aleatória com distribuição gama, aqui denotada por  $Y \sim G(\mu, \phi)$ , tal que

$$f_{(\mu, \phi)}(y) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi x}{\mu}\right)^\phi \exp\left(-\frac{\phi y}{\mu}\right) \frac{1}{y}, \quad y \geq 0, \mu > 0, \phi > 0, \Gamma(\phi) = \int_0^\infty t^{(\phi-1)} e^{-t} dt.$$

Se nossa variável resposta pertence aos reais positivos e decidimos usar a distribuição gama, um modelo de regressão adequado seria

$$\log(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n.$$

Isto porque, o logaritmo só pode ser calculado para valores reais positivos e o seu resultado assume valores em todos os reais, o que libera os  $\hat{\beta}$ 's. Note que  $\log(u) \in (-\infty, +\infty)$  para todo  $u \in (0, +\infty) = IR^+$ .

### 5.3.2. Modelagem de Regressão – Modelo de Regressão Beta

[Ferrari e Cribari-Neto, 2004] propõe que a média da variável resposta  $y_i$ , ou seja,  $\mu_i$  possa ser escrita como

$$g(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n.$$

No modelo de regressão beta também é necessário o uso da função de ligação, para liberar os possíveis valores que os  $\beta$ 's podem assumir. Neste caso, como  $\mu_i \in (0, 1)$  um função de ligação, ou seja, uma transformação que conduz essa média a todos os Reais é a função de ligação logito, dada por

$$\mu_i \in (0, 1) \leftrightarrow \log \left\{ \frac{\mu_i}{(1 - \mu_i)} \right\} \in (-\infty, +\infty) = IR.$$

Tipicamente, além de explicar ou modelar a média da variável resposta  $\mu_i$ , que é um parâmetro desconhecido, também modelamos a variância da variável resposta, que neste caso implica em modelar o parâmetro  $\phi$ , pois a variância depende de  $\phi_i$  e de  $\mu_i$  (que já está sendo modelada). Em geral sugerimos um modelo para  $\phi$  quando suspeitamos

que a dispersão não é constante para os dados, ou que é possível haver grupos com dispersões diferentes. Assim, [Smithson e Verkuilen, 2006] propõe um modelo de regressão beta em que

$$h(\phi_i) = \gamma_1 + \gamma_2 z_{i2} + \gamma_3 z_{i3} + \dots + \gamma_q z_{iq}, \quad i = 1, \dots, n.$$

Mais uma vez é necessário o uso de uma **função de ligação**, que neste caso denotamos por  $h(\cdot)$ . Como  $\phi > 0$ , ou seja, só assume valores positivos, uma função de ligação adequada é  $\log \phi \in \mathbb{R}$ . Precisamos estimar os  $\beta_t$ 's e os  $\gamma_j$ 's para estimarmos  $\phi_i$  e  $\mu_i$ . Isto é feito utilizando o método de máxima verossimilhança. Pesquisadores podem usar o pacote *betareg* o qual está disponível no software estatístico R.

### 5.3.3. Análise descritiva dos dados

Vamos propor um modelo de regressão em que o escore subjetivo entre (0,1) possa ser explicado pelo escore objetivo fornecido pela metodologia proposta para avaliação de ontologias. Para a construção do experimento foram consideradas outras variáveis. Assim, vamos investigar uma relação geral entre todas as variáveis envolvidas no experimento.

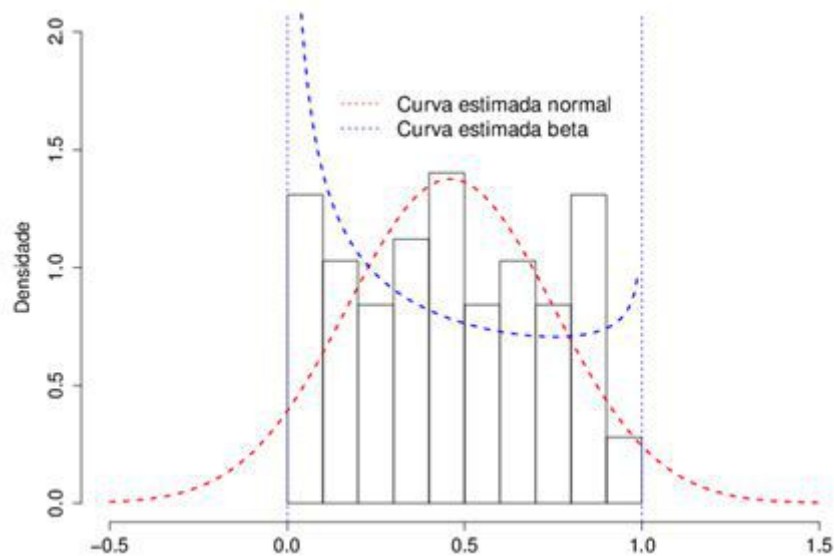
A variável resposta do modelo de regressão que investigaremos é  $y$ : o escore subjetivo obtido SEM o uso da metodologia proposta para avaliação de ontologias. Os fatores que podem interferir na nossa resposta são: o escore objetivo atribuído pelos mesmo avaliadores após o uso da metodologia proposta para avaliação ontologias: (0,100) e o nível de experiência em ontologia dos avaliadores: Vasta, grande, média, pouca e pouquíssima; GOAL: Substituto, Compromisso, Raciocínio, Computação e Expressão e as ontologias disponibilizadas: 1, 2, 3 e 4.

Inicialmente vamos investigar qual distribuição de probabilidades é mais adequada para a resposta. Temos que  $y \in (0,1)$ , logo uma distribuição adequada para modelar esses dados é a distribuição beta. A distribuição beta é comumente usada para modelar variáveis aleatórias que assumem valores no intervalo de (0,1), tais como taxas, índices, escores, proporções. A densidade beta pode assumir formas diferentes dependendo da combinação de valores de parâmetros. Seja  $y_1, \dots, y_n$  uma amostra de v.a.'s independentes tal que cada  $y_i, i = 1, \dots, n$ . Segue uma distribuição beta com densidade:

$$f(y; \mu_i, \phi_i) = \frac{\Gamma(\phi_i)}{\Gamma(\mu_i \phi_i) \Gamma((1 - \mu_i) \phi_i)} y^{\mu_i \phi_i - 1} (1 - y)^{(1 - \mu_i) \phi_i - 1}, \quad 0 < y < 1,$$

Onde  $0 < \mu_i < 1$ ;  $\mu \in (0,1)$  e  $\phi > 0$ . Aqui,  $E(y_i) = \mu_i$  e  $var(y_i) = (\mu_i(1 - \mu_i))/(1 + \phi_i)$ . Temos que  $\phi$  pode ser visto como um parâmetro de precisão, pois quanto maior  $\phi$  menor a variância de  $y_i$ , por outro lado,  $\phi^{-1}$  é um parâmetro de dispersão.

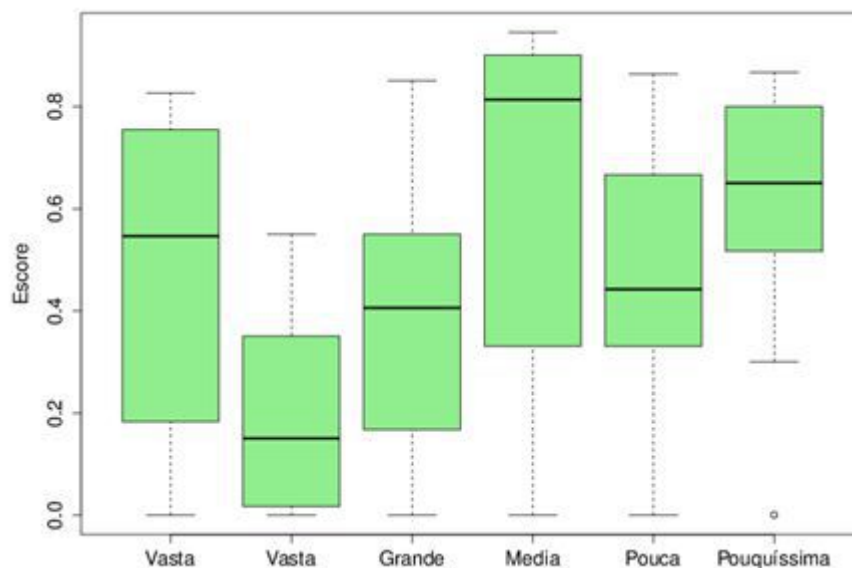
Vamos comparar duas distribuições de probabilidades competidoras, a saber: a distribuição normal e a distribuição beta. Na Figura 14 temos o histograma dos valores do escore subjetivo para a amostra estudada. Adicionalmente foram estimadas duas densidades possíveis para a variável resposta, uma densidade estimada a partir da distribuição beta (em azul) e outra densidade estimada com base na distribuição normal.



**Figura 14. Histograma dos escores para qualificação de ontologias - Sem o uso da metodologia. Curvas estimadas a partir das distribuições normal e beta.**

Podemos notar com base na Figura 14, que a densidade estimada a partir da distribuição beta é mais adequada aos dados, uma vez que todos os valores estimados abaixo da curva beta estão dentro do intervalo (0,1). Por outro lado, se fosse utilizada a distribuição normal para estimar os escores subjetivos, poderiam ocorrer valores negativos e maiores que, os quais estão fora da faixa zero e um que determina os valores possíveis da variável resposta.

Também é interessante verificar como se comporta o escore subjetivo quanto as categorias de experiência do avaliador e quanto as GOALs. Para isso forma construídos boxplots apresentados na Figuras 15 e 16.

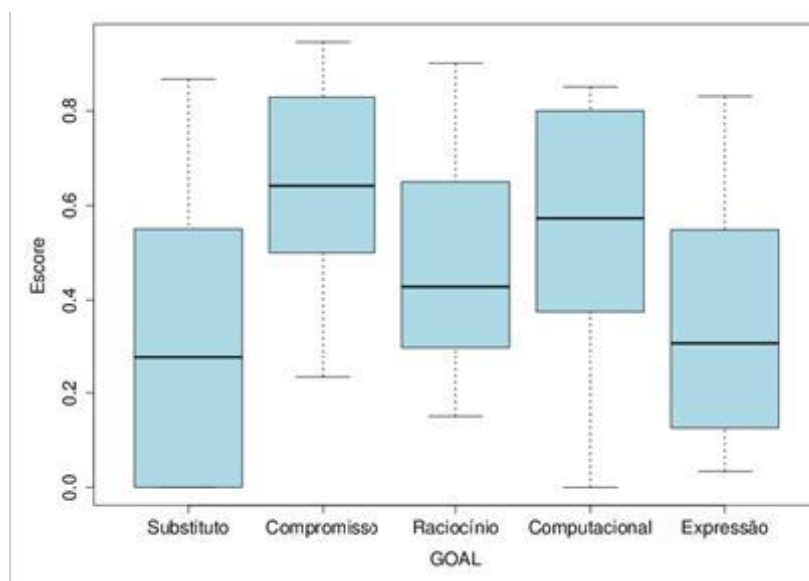


**Figura 15. Boxplots dos escores para qualificação de ontologias - Sem o uso da metodologia. Segundo experiência do avaliador.**

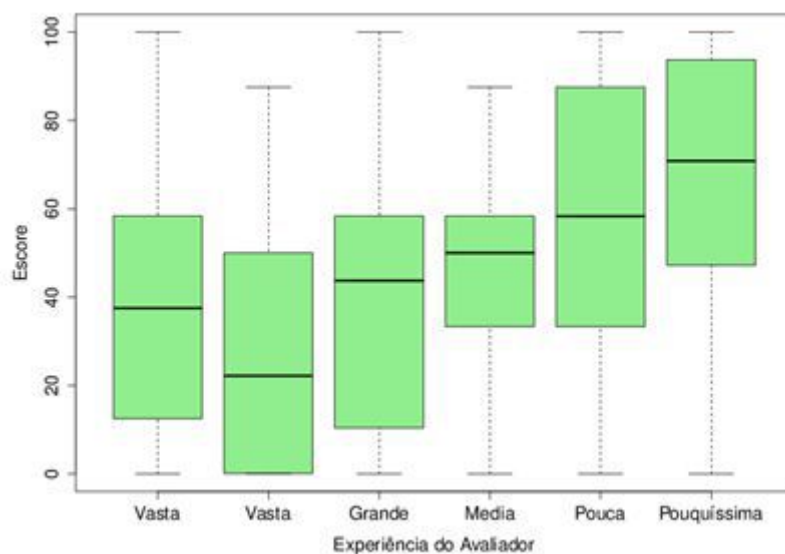
A análise da Figura 15 revela como a mediana dos escores subjetivos dos avaliadores com vasta e grande experiência é consideravelmente inferior a mediana do avaliadores menos experientes. Ou seja, ou os avaliadores com mais experiência são muito mais rigorosos em suas qualificações ou estes mesmos avaliadores, devido à experiência, identificaram que as ontologias disponibilizadas apresentavam muitas falhas e corretamente as pontuaram com escores baixos. Quando analisamos a Figura 16 percebemos que as ontologias pontuaram melhor em compromisso e computação eficiente. E o papel em que as ontologias apresentaram pior desempenho foi substituído.

A construção de boxplots se justifica quando tratamos de variáveis aleatórias. No entanto, de forma simbólica podemos construir boxplots para os escores objetivos obtidos com base na metodologia proposta. Só salientamos que NÃO vamos considerar o escore objetivo como variável aleatória e sim com um fator ou covariada que servirá para explicar a variável aleatória resposta através de um modelo de regressão. E neste caso, em um modelo de regressão tipicamente covariadas são fixas, sempre conhecidas, logo não são variáveis aleatórias. Como o escore objetivo é construído como base em uma

metodologia que segue um passo-a-passo, o caráter aleatório não se justifica. Dizemos que é uma componente sistemática e não aleatória. Mas, simbolicamente podemos construir boxplots destas medidas para avaliar seu comportamento. (Figuras 17 e 18).



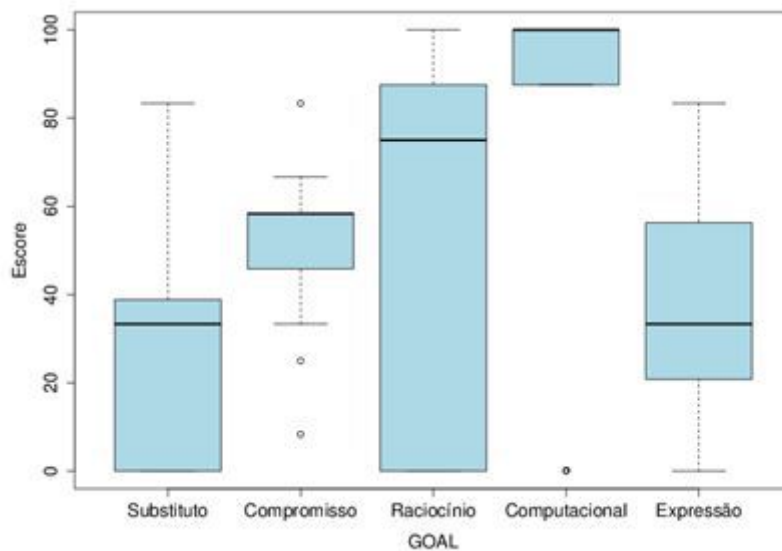
**Figura 16. Boxplots dos escores para qualificação de ontologias - Sem o uso da metodologia. Segundo GOAL.**



**Figura 17. Boxplots dos escores para qualificação de ontologias - Com o uso da metodologia. Segundo experiência do avaliador.**

A Figura 17 evidencia de forma mais organizada como de fato os avaliadores com mais experiência atribuem escores menores. De fato, quando o escore é construído com base na metodologia proposta a influência da experiência do avaliador fica mais evidente.

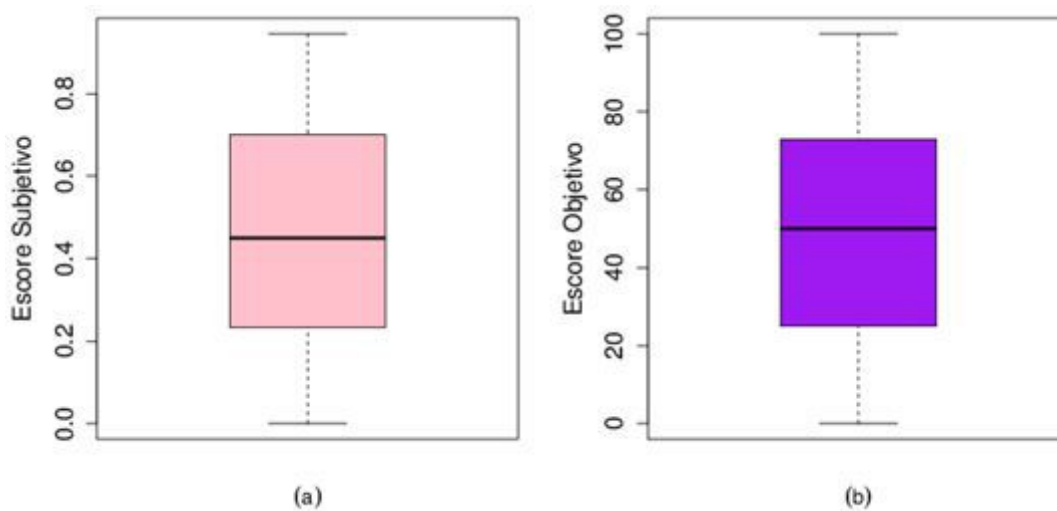
Mesmo assim o modelo consegue padronizar ou melhor sistematizar a avaliação da ontologia pois com exceção do avaliador com pouquíssima experiência, as medianas dos demais avaliadores estão próximas.



**Figura 18. Boxplots dos escores para qualificação de ontologias - Com o uso da metodologia. Segundo GOAL.**

Quanto ao desempenho das ontologias segundo as GOALs segundo a metodologia proposta, confirma-se o que a GOAL substituto é a que apresenta maiores falhas, e que a eficiência computacional é o papel com melhor desempenho.

Na Figura 19 apresentamos os boxplos dos escores obtidos sem o uso da metodologia e com o uso da metodologia. É notável a correspondência entre os comportamentos dos dois tipos de escores, evidenciando que a metologia proposta pode ser utilizada para explicar a qualidade de ontologias.



**Figura 19. Boxplots dos escores para qualificação de ontologias - Sem o uso da metodologia (a) e Com o uso da metodologia (b).**

Utilizando os conhecimentos adquiridos em 5.3.2, propomos um modelo que acreditamos que seja o mais adequado para explicar a variável resposta. Neste passo incluímos as covariadas que acreditamos que sejam importantes para a explicação da resposta. As covariadas principais que construímos foram:

- “Cov”: Escore objetivo atribuído pelos mesmo avaliadores após o uso da metodologia proposta para avaliação de ontologias: (0,100);
- “GraExp”: Se o avaliador tem média ou pouca experiência. Ou seja os avaliadores com vasta ou grande experiência foram agregados em uma única covariada. Essa covariada assume valores: 1 – se o avaliador tem média ou pouca experiência; 0 – Caso Contrário.
- “MédiaExp”: Se o avaliador tem média ou pouca experiência. Essa covariada assume valores: 1 – se o avaliador tem média ou pouca experiência; 0 – Caso Contrário.
- “Substituto”: 1 – se a GOAL representa substituto; 0 – Caso Contrário.
- “Compromisso”: 1 – se a GOAL representa substituto; 0 – Caso Contrário.
- “Raciocínio”: 1 – se a GOAL representa substituto; 0 – Caso Contrário.
- “Computação”: 1 – se a GOAL representa substituto; 0 – Caso Contrário.
- “Onto1”: 1 – trata-se da ontologia 1; 0 – Caso Contrário.

- “Onto2”: 1 – trata-se da ontologia 2; 0 – Caso Contrário.
- “Onto3”: 1 – trata-se da ontologia 3; 0 – Caso Contrário.
- “Onto4”: 1 – trata-se da ontologia 4; 0 – Caso Contrário.
- “Nula”: Uma covariada que indica a ocorrência de “Não soube responder” Apenas para o uso da metodologia – Essa covariada assume valores: 1 – se todas as questions e uma GOAL não foram avaliadas com base na metodologia, impossibilitando o cálculo da média do escore da question; 0 – Caso contrário.

Consideramos inicialmente o seguinte modelo:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \beta_{10} x_{i10} + \beta_{11} x_{i11} + \beta_{11} x_{i11} = \beta_1 + \beta_2 Cov_i + \beta_3 (Cov \times Sub)_i + \beta_4 (Cov \times Compr)_i + \beta_5 (Cov \times Rac)_i + \beta_6 (Cov \times Comp)_i + \beta_7 GraExp_i + \beta_8 MedExp_i + \beta_9 Onto1_i + \beta_{10} Onto2_i + \beta_{11} Onto3_i + \beta_{12} Nula_i, \\ 1, \dots, n.$$

e

$$h(\phi_i) = \log(\phi_i) = \gamma_1 + \gamma_2 z_{i2} + \gamma_3 z_{i3} + \gamma_4 z_{i4} + \gamma_5 z_{i5} + \gamma_6 z_{i6} + \gamma_7 z_{i7} + \gamma_8 z_{i8} + \gamma_9 z_{i9} + \gamma_{10} z_{i10} + \gamma_{11} z_{i11} + \gamma_{12} z_{i12} = \gamma_1 + \gamma_2 Cov_i + \gamma_3 (Cov \times Sub)_i + \gamma_4 (Cov \times Compr)_i + \gamma_5 (Cov \times Rac)_i + \gamma_6 (Cov \times Comp)_i + \gamma_7 GraExp_i + \gamma_8 MedExp_i + \gamma_9 Onto1_i + \gamma_{10} Onto2_i + \gamma_{11} Onto3_i + \gamma_{12} Onto3_i, \\ 1, \dots, n.$$

Em seguida estimamos esse modelo com base no conjunto de dados que temos em mãos. No nosso caso estimar o modelo implica em estimar os  $\beta_t$ 's e os  $\gamma_j$ 's para estimarmos  $\phi_i$  e  $\mu_i$ . E utilizamos o método Stepwise (StepAIC no programa **R**) para selecionar o melhor modelo. Finalmente, chegamos ao seguinte modelo de regressão:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} = \beta_1 + \beta_2 (Cov \times Sub)_i + \beta_3 (Cov \times Compr)_i + \beta_4 (Cov \times Racio)_i + \beta_5 (Cov \times Comp)_i + \beta_6 GraExp_i + \beta_7 (Cov \times Nula)_i, \\ 1, \dots, n.$$



e

$$h(\phi_i) = \log(\phi_i) = \gamma_1 + \gamma_2 z_{i2} + \gamma_3 z_{i3} + \gamma_4 z_{i4} + \gamma_5 z_{i5} = \\ \gamma_1 + \gamma_2(Cov \times Sub)_i + \gamma_3 Sub_i + \gamma_4 GraExp_i + \gamma_5 MediaExp_i, \quad 1, \dots, n.$$

Apresentamos o modelo estimado nas Tabelas 13 e 14. Na Tabela 13 encontram-se as estimativas dos  $\beta$ 's, que de fato é o modelo que nos interessa. Na Tabela 14 encontram-se as estimativas dos  $\gamma$ 's que irão definir o modelo para  $\hat{\phi}_i$  que juntamente com  $\hat{\mu}_i$  vão estimar as variâncias dos  $\hat{\beta}$ 's. Lembre-se que  $var(y_i) = (\mu_i(1 - \mu_i))/(1 + \phi_i)$ . Logo, a estimativa da **variância de y** é  $\widehat{Var}(y) = \{\hat{\mu}(1 - \hat{\mu})\}/(1 + \hat{\phi})$ . O cálculo da variância estimada de  $\hat{\beta}$  não é tão trivial, mas também dependerá de  $\hat{\mu}$  e  $\hat{\phi}$ . Então, o modelo estimado na Tabela 14 é utilizado para os cálculos das variâncias  $\hat{\beta}$ 's que irão construir as estatísticas de testes apresentadas na próxima seção.

#### 5.3.4. Verificação das hipóteses

Como discutido foi anteriormente, se conseguirmos ajustar adequadamente o modelo de regressão proposto então estaremos respondendo positivamente as hipóteses:

Trazendo de volta nossas hipóteses, temos: A metodologia é válida? É possível obter um escore global para qualificar ontologias com base na metodologia proposta? É possível obter um escore específico para cada GOAL com base na metodologia proposta? Começamos formulando as hipóteses:

$$H_0 : \beta_t = 0 \quad \text{versus} \quad H_1 : \beta_t \neq 0, \quad t = 1, \dots, 7.$$

Ou seja, as hipóteses que vamos testar são

$$H_0 : \beta_1 = 0; H_0 : \beta_2 = 0; H_0 : \beta_3 = 0; H_0 : \beta_4 = 0; H_0 : \beta_5 = 0; H_0 : \beta_6 = 0; H_0 : \beta_7 = 0.$$

Versus

$$H_1 : \beta_1 \neq 0; H_1 : \beta_2 \neq 0; H_1 : \beta_3 \neq 0; H_1 : \beta_4 \neq 0; H_1 : \beta_5 \neq 0; H_1 : \beta_6 \neq 0; H_1 : \beta_7 \neq 0.$$

As estatísticas de teste para testar as hipóteses acima são

$$\tau_1 = \frac{(\hat{\beta}_1 - \beta_1)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}}; \tau_2 = \frac{(\hat{\beta}_2 - \beta_2)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_2)}}; \tau_3 = \frac{(\hat{\beta}_3 - \beta_3)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_3)}}; \dots; \tau_7 = \frac{(\hat{\beta}_7 - \beta_7)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_7)}}.$$

Como calculamos o p-Valor considerando  $H_0$  verdadeira, então todos os  $\beta_t, t = 1, \dots, 7$ , são iguais a zero e as estatísticas passam a ser:

$$\tau_1 = \frac{\hat{\beta}_1}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}}; \tau_2 = \frac{\hat{\beta}_2}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_2)}}; \tau_3 = \frac{\hat{\beta}_3}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_3)}}; \dots; \tau_7 = \frac{\hat{\beta}_7}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_7)}}.$$

Temos que neste caso:  $\tau_t \approx N(0,1), t = 1, \dots, 7$ , ou seja, as estatísticas têm distribuição aproximadamente normal padrão e rejeitamos as hipóteses nulas se  $p\text{-Valor}_t < 0.05, t = 1, \dots, 7$  e neste caso a covariada associada ao coeficiente  $\beta_t$  é considerada importante para a explicação da resposta. Por exemplo, se  $p\text{-Valor}_3 < 0.05$  então rejeitamos  $H_0: \beta_3 = 0$ , ou seja,  $\beta_3$  é significativamente diferente de zero ao nível de 5%, e neste caso, a covariada  $Cov \times Compr$  é estatisticamente importante para a explicação do escore médio de avaliação de ontologias construído com base na metodologia proposta.

Descrição das variáveis	Parâmetros						
	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$
	Const	Cov $\times$ Sub	Cov $\times$ Compr	Cov $\times$ Racio	Cov $\times$ Comp	GraExp	Cov $\times$ Nula
estimativa	-0.44	0.03	0.02	0.01	0.02	-0.66	-25
p-Valor	0.0000	0.0001	0.0000	0.0000	0.0013	0.0549	0.0005

**Tabela 13. Estimativas dos coeficientes do modelo da média da variável escores para qualificação de ontologias baseado na metodologia proposta.**

Com base na Tabela 13 notamos que rejeitamos todas as hipóteses

$$H_0 : \beta_1 = 0; H_0 : \beta_2 = 0; H_0 : \beta_3 = 0; H_0 : \beta_4 = 0; H_0 : \beta_5 = 0; H_0 : \beta_6 = 0; H_0 : \beta_7 = 0.$$

Portanto, todas as covariadas são consideradas importantes para construir um modelo de regressão que forneça um escore em (0,1) para qualificar ontologias.

Descrição das variáveis	Parâmetros				
	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\gamma_4$	$\gamma_5$
	Const	Cov $\times$ Sub	Sub	GraExp	MediaExp
estimativa	2.55	0.08	-2.53	-0.93	-1.25
p-Valor	0.000	0.0000	0.0000	0.0000	0.000

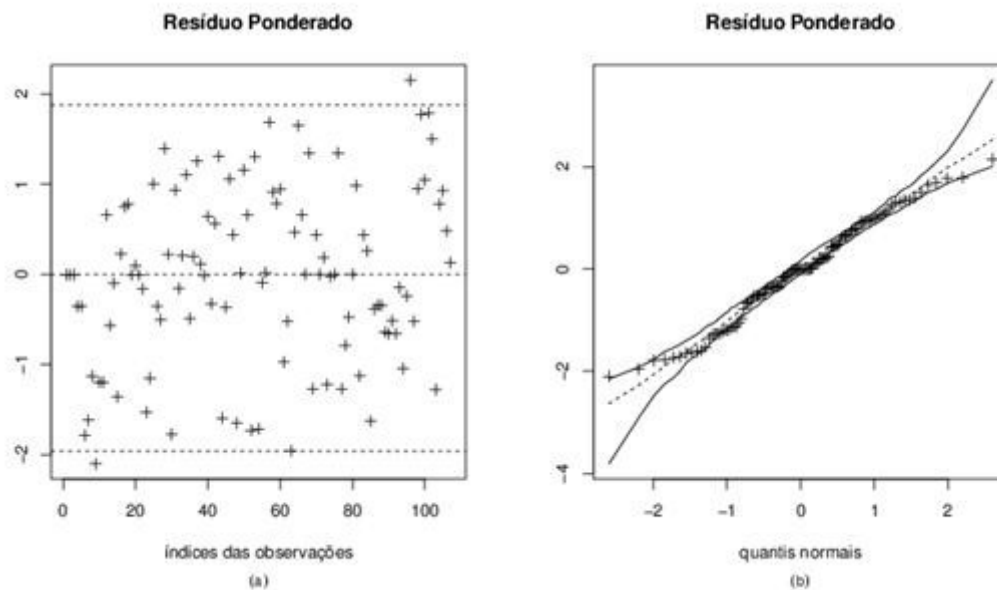
**Tabela 14. Estimativas dos coeficientes do modelo da dispersão para os dados de escores para qualificação de ontologias baseado na metodologia proposta.**

Outra Hipótese de temos que testar é: O modelo estimado com base nos dados, ou seja o modelo ajustado aos dados é de boa qualidade? Podemos fazer isso com base na análise de resíduos.

Vamos Verificar se um determinado modelo é uma representação adequada dos dados é um passo importante da análise estatística. A construção de um modelo de regressão envolve a definição da distribuição a variável de resposta, a escolha da função de ligação, a escolha das covariáveis. Tipicamente os resíduos são baseados nas diferenças entre o respostas observadas ( $y$ ) e a média estimada  $\hat{\mu}$ . Por exemplo  $r_i = y_i - \hat{\mu}_i$ , ou seja, o resíduo é uma medida de discrepância entre os dados reais e o modelo ajustado. Aqui vamos utilizar o resíduo proposto por [Espinheira et al, 2015],  $r_{p,i}^{\beta\gamma}$  denominado resíduo combinado e baseado na diferença

$$(y_i^* - \hat{\mu}_i^*), \quad \text{em que } y_i^* = \log \left\{ \frac{y_i}{(1 - y_i)} \right\} \quad \text{e } \mu_i^* = E(y_i^*).$$

Os gráficos de resíduos versus índices das observações ou versus valores preditos ( $\hat{\mu}_i$ ) são os mais básicos. Se um modelo está especificado corretamente, então estes gráficos não devem apresentar nenhuma tendência, os resíduos devem estar aleatoriamente distribuídos em torno do zero. A presença de quaisquer características sistemáticas tipicamente implica uma falha de um ou mais pressupostos do modelo. Outro gráfico de resíduos importante é o gráfico de probabilidade normal com envelope simulado, que pode ser usada mesmo quando as distribuições empíricas dos resíduos não são normais. Se o modelo está adequado aos dados, esperamos que a maioria dos resíduos estejam aleatoriamente distribuídos dentro das bandas do envelope.



**Figura 20 – Gráficos de resíduo contra índices das observações (a) e gráfico normal de probabilidades com envelopes simulados (b) – Modelo de regressão beta para a construção de escores de qualificação de ontologias.**

Com base na Figura 20(a) vemos que os resíduos estão aleatoriamente distribuídos em torno da média zero indicando uma boa qualidade de ajuste do modelo. Já na Figura 20(b) notamos que a grande maioria dos resíduos estão distribuídos dentro das bandas do envelope indicando que o modelo proposto para os dados estão adequados. Com isto confirmamos a hipótese que o modelo está bem ajustado aos dados. Finalmente, após encontrar um modelo de regressão comprovadamente adequado, envolvendo os escores subjetivos e os escores objetivos, respondemos positivamente às questões do problema. A metodologia proposta é válida, pois é possível obter um escore global para avaliar ontologias e é possível obter escores parciais por GOALS. O modelo estatístico final é dado por:

$$\hat{\mu}_i = \frac{\exp\{-0.44 + 0.03(Cov_S \times Sb)_i + 0.02(Cov_C \times Cr)_i + 0.01(Cov_R \times Rc)_i + 0.02(Cov_{Cp} \times Cp)_i - 0.66GExp_i - 25(0.1 \times NI)_i\}}{1 + \exp\{-0.44 + 0.03(Cov_S \times Sb)_i + 0.02(Cov_C \times Cr)_i + 0.01(Cov_R \times Rc)_i + 0.02(Cov_{Cp} \times Cp)_i - 0.66GExp_i - 25(0.1 \times NI)_i\}}$$

Ou seja, a fórmula que é utilizada para o terceiro passo da metodologia proposta.

#### 5.4. Principais conclusões

Como pode-se verificar, a metodologia proposta foi validada, através da criação do escore global e parcial, que podem ser utilizadas para o terceiro passo da metodologia. Com esta validação, temos:

- Um passo-a-passo para avaliação de ontologias;
- Um valor numérico o qual torna a avaliação mais precisa em relação as metodologias existentes;
- Uma equação construída através de uma validação estatística;
- Possibilidade de cálculo total ou parcial da qualidade da ontologia;
- Uma metodologia para avaliação de ontologias que independe da experiência em ontologias.

## 6. CONCLUSÕES E TRABALHOS FUTUROS

O trabalho apresentou uma nova metodologia para avaliação de ontologias utilizando princípios da representação do conhecimento, chamada. O modelo foi apresentado separado por seus três passos principais: Verificação do Tipo da Ontologia, Verificação das Questões e Mensurar e Verificação da Qualidade da Ontologia.

A necessidade da criação desta metodologia surgiu quando percebeu-se que a maioria dos trabalhos que envolve metodologias para avaliação de ontologias apenas apresentam os critérios, porém não estabelecem uma forma de guiar o avaliador em suas abordagens, o que poderia gerar imprecisão no diagnóstico final por parte do avaliador.

A metodologia proposta foi criada a partir da correspondência entre os principais critérios de qualidade de ontologias existentes na comunidade com os cinco papéis da representação do conhecimento. Após isto, foi necessário estabelecer uma dependência entre duas questões, a partir tipo da ontologia.

A metodologia proposta atingiu o objetivo proposto nesse trabalho, possibilitando aos usuários da comunidade de ontologias agora terem uma metodologia onde estabelece um passo-a-passo para avaliar a qualidade de ontologias, saberem o quão a ontologia cumpre cada papel da representação do conhecimento e saberem numericamente e precisamente a qualidade da ontologia.

A metodologia proposta, no entanto, possui algumas limitações. Uma delas é a falta de questões que torne mais distinta a avaliação da ontologia de acordo com os tipos *task or domain* e *application ontology*. Apenas duas questões fazem esta diferenciação. Outra limitação é que, avaliar duas ou mais ontologias, de acordo com três passos, treze questões e o cálculo de uma fórmula, pode tornar o uso da metodologia um pouco exaustivo.

Como trabalhos futuros, com o objetivo de eliminar as limitações, será estudada a possibilidade de inserir o tipo *top ontology* nesta avaliação, através da inserção de mais questões referentes à este tipo e será verificada a possibilidade de semi-automatizar ou automatizar por completo este processo, para diminuir a exaustão por parte do avaliador.



## REFERÊNCIAS

- Gruber, Thomas R. "A translation approach to portable ontology specifications." *Knowledge acquisition* 5.2 (1993): 199-220.
- Cardoso, Jorge. "The semantic web vision: Where are we?." *Intelligent Systems, IEEE* 22.5 (2007): 84-88.
- Vrandečić, Denny. *Ontology evaluation*. Springer Berlin Heidelberg, 2009.
- Thomasson, Amie L. "Methods of categorization." *Proceedings of the 3rd International Conference (FOIS'04)*. 2004.
- Mealy, George H. "Another look at data." *Proceedings of the November 14-16, 1967, fall joint computer conference*. ACM, 1967.
- Guarino, Nicola. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. Vol. 46. IOS press, 1998.
- Davis, Randall, Howard Shrobe, and Peter Szolovits. "What is a knowledge representation?." *AI*
- Berners-Lee, Tim, James Hendler, and Ora Lassila. "The semantic web." *Scientific american* 284.5 (2001): 28-37.
- Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked data-the story so far." *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (2009): 205-227.
- Noy, Natalya F. "Ontology Development 101: A Guide to Creating Your First Ontology: Knowledge Systems Laboratory, Stanford University." *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880* (2001).
- Gómez-Pérez, Asunción, and Mari Carmen Suárez-Figueroa. "NeOn Methodology for Building Ontology Networks: a Scenario-based Methodology." (2009).
- De Nicola, Antonio, Michele Missikoff, and Roberto Navigli. "A proposal for a unified process for ontology building: UPON." *Database and Expert Systems Applications*. Springer Berlin Heidelberg, 2005.
- Temmerman, Rita, and Koen Kerremans. "Termtography: Ontology building and the sociocognitive approach to terminology description." (2003).
- Menzel, Christopher P., Richard J. Mayer, and Michael K. Painter. *IDEF5 Ontology Description Capture Method: Concepts and Formal Foundations*. TEXAS A AND M UNIV COLLEGE STATION KNOWLEDGE BASED SYSTEMS LAB, 1992.
- Auer, Sören, and Heinrich Herre. "RapidOWL—An agile knowledge engineering



methodology." Perspectives of systems informatics. Springer Berlin Heidelberg, 2007. 424-430.

Uschold, Michael. "Building ontologies: Towards a unified methodology." TECHNICAL REPORT-UNIVERSITY OF EDINBURGH ARTIFICIAL INTELLIGENCE APPLICATIONS INSTITUTE AIAI TR (1996).

Sure, York, Steffen Staab, and Rudi Studer. "On-to-knowledge methodology (OTKM)." Handbook on ontologies. Springer Berlin Heidelberg, 2004. 117-132.

Fernández-López, Mariano, Asunción Gómez-Pérez, and Natalia Juristo. "Methontology: from ontological art towards ontological engineering." (1997).

Gangemi, Aldo, Geri Steve, and Fabrizio Giacomelli. "ONIONS: An ontological methodology for taxonomic knowledge integration." ECAI-96 Workshop on Ontological Engineering, Budapest. 1996.

Grüninger, Michael, and Mark S. Fox. "Methodology for the Design and Evaluation of Ontologies." (1995).

Lenat, Douglas B., Mayank Prakash, and Mary Shepherd. "CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks." AI magazine 6.4 (1985): 65.

Guarino, Nicola, and Christopher Welty. "Evaluating ontological decisions with OntoClean." Communications of the ACM 45.2 (2002): 61-65.

Schreiber, Guus. Knowledge engineering and management: the CommonKADS methodology. MIT press, 2000.

Jarrar, Mustafa, and Robert Meersman. "Formal ontology engineering in the dogma approach." On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE. Springer Berlin Heidelberg, 2002. 1238-1254.

Aussenac-Gilles, Nathalie, Sylvie Despres, and Sylvie Szulman. "The terminae method and platform for ontology engineering from texts." Bridging the Gap between Text and Knowledge-Selected Contributions to Ontology Learning and Population from Text (2008): 199-223.

Schreiber, Guus, Bob Wielinga, and Wouter Jansweijer. "The KACTUS view on the 'O'word." IJCAI workshop on basic ontological issues in knowledge sharing. 1995.

Pinto, H. Sofia, Steffen Staab, and Christoph Tempich. "DILIGENT: Towards a fine-grained methodology for Distributed, Loosely-controlled and evolInG." Proceedings of

the 16th European Conference on Artificial Intelligence (ECAI 2004). Vol. 110. 2004.

Kotis, Konstantinos, and George A. Vouros. "Human-centered ontology engineering: The HCOME methodology." *Knowledge and Information Systems* 10.1 (2006): 109-131.

Basili, Victor R. "Software modeling and measurement: the Goal/Question/Metric paradigm." (1992).

Koziolk, Heiko. "Goal, question, metric." *Dependability metrics*. Springer Berlin Heidelberg, 2008. 39-42.

Obrst, Leo, et al. "The evaluation of ontologies." *Semantic Web*. Springer US, 2007. 139-158.

Gangemi, Aldo, et al. *Modelling ontology evaluation and validation*. Springer Berlin Heidelberg, 2006.

Gruber, Thomas R. "Toward principles for the design of ontologies used for knowledge sharing?." *International journal of human-computer studies* 43.5 (1995): 907-928.

Gómez-Pérez, Asunción. "Evaluation of ontologies." *International Journal of intelligent systems* 16.3 (2001): 391-409.

Travassos, Guilherme Horta, Dmytro Gurov, and E. A. G. G. Amaral. *Introdução à engenharia de software experimental*. UFRJ, 2002.

Bras, Pesqui Odontol. "Organização do currículo–plataforma Lattes Curriculum vitae organization–the Lattes software platform." *Pesqui Odontol Bras* 17.Supl 1 (2003): 18-22.

Ferrari, Silvia, and Francisco Cribari-Neto. "Beta regression for modelling rates and proportions." *Journal of Applied Statistics* 31.7 (2004): 799-815.

Espinheira, Patrícia L., Luana Cecília Meireles da Silva, and Alisson de Oliveira Silva. "Prediction Measures in Beta Regression Models." *arXiv preprint arXiv:1501.04830* (2015).

Smithson, Michael, and Jay Verkuilen. "A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables." *Psychological methods* 11.1 (2006): 54.

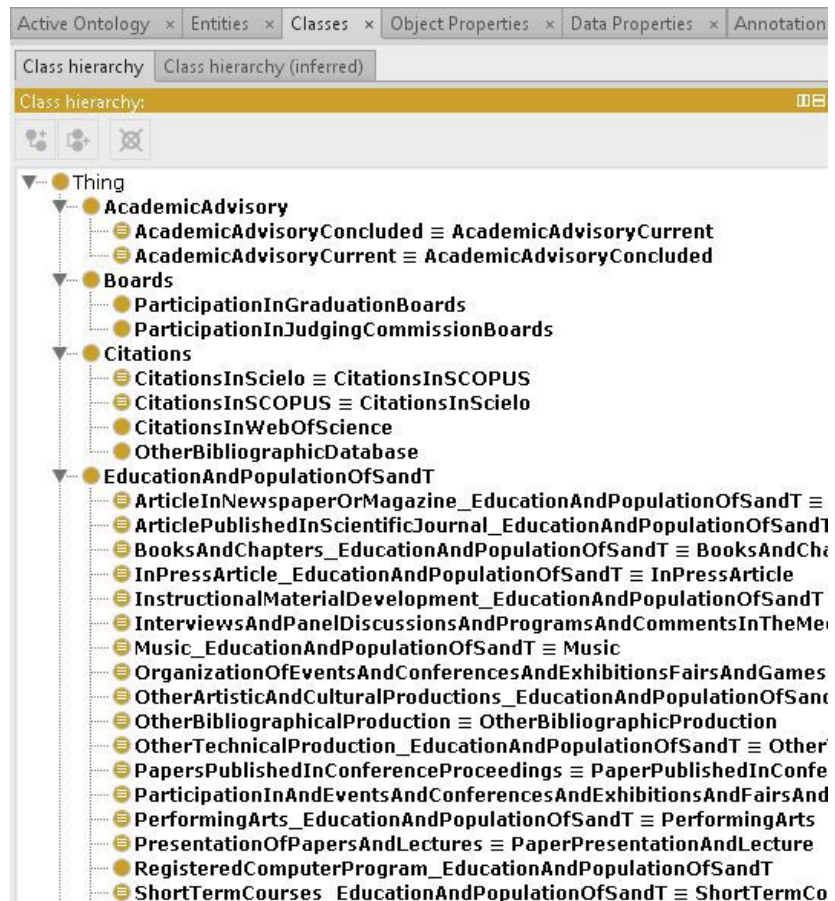
Lehmann, Erich Leo, and George Casella. *Theory of point estimation*. Vol. 31. Springer Science & Business Media, 1998.

# APÊNDICES

## APÊNDICE A – Ontologias do Lattes

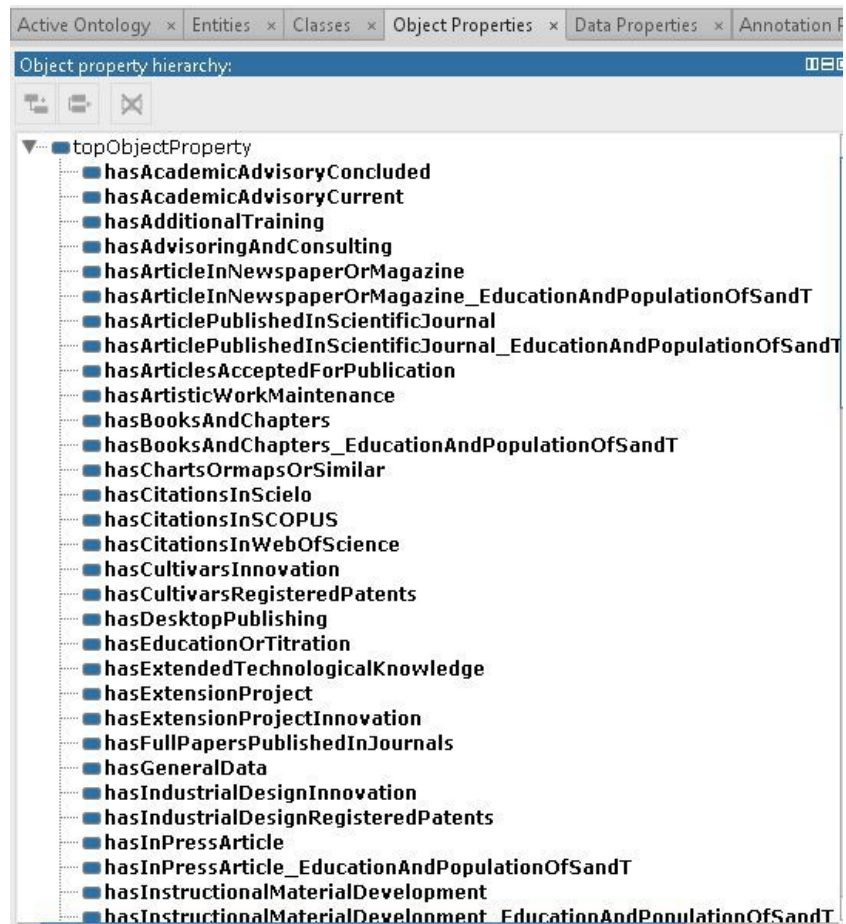
Este apêndice mostrará imagens das ontologias no editor de ontologias Protégé.

- Ontologia 1
  - Classes



Maioria das classes da ontologia do lattes 1

- Object Properties



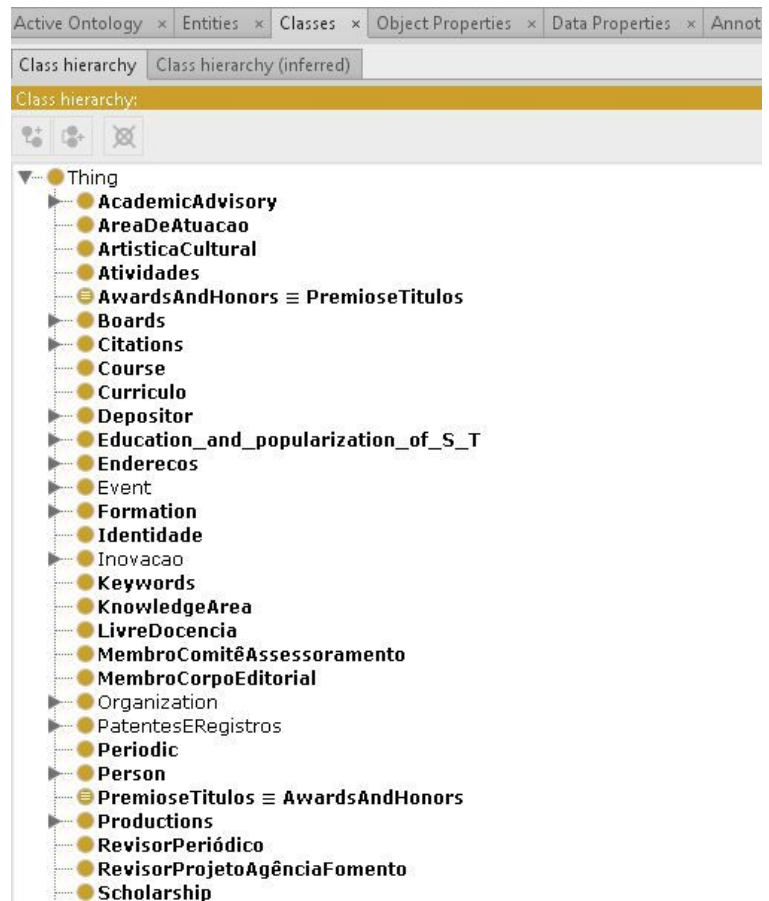
Maioria das object properties da ontologia do lattes 1

- Data Properties



Maiores das data properties da ontologia doattes 1

- Ontologia 2
  - Classes



Maioria das classes da ontologia do lattes 2

- Object Properties



Maioria das object properties da ontologia do lattes 2

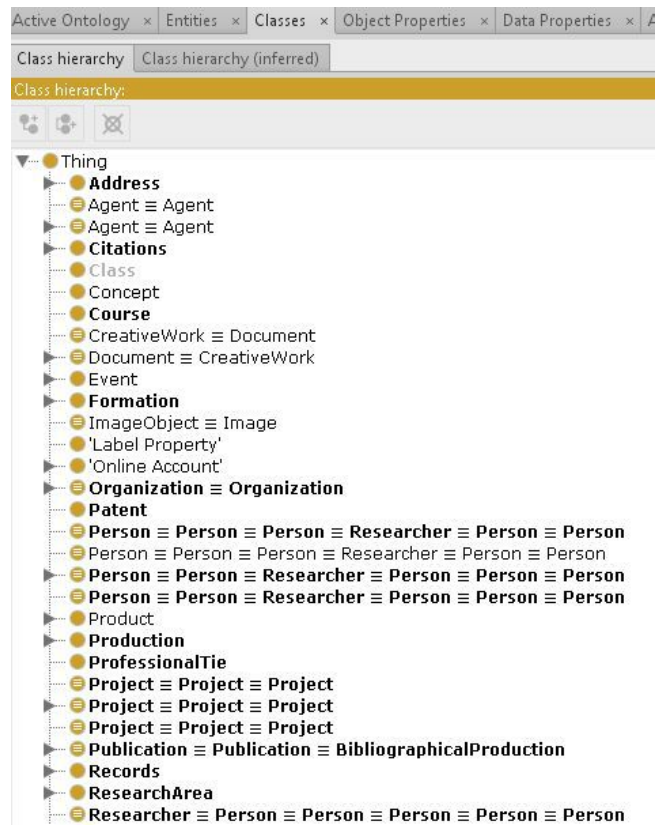
- o Data Properties





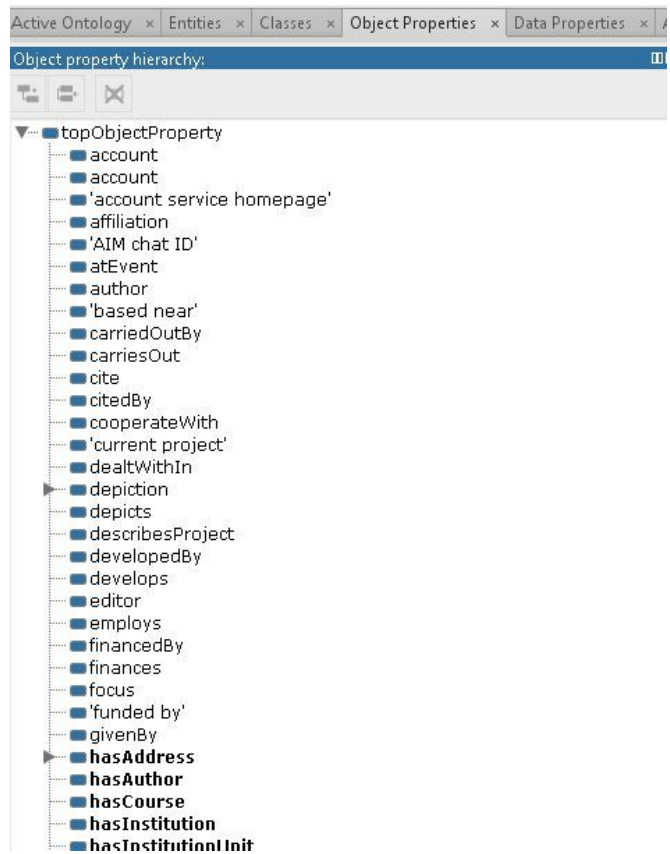
Maioria das data properties da ontologia do lattes 2

- Ontologia 3
  - Classes



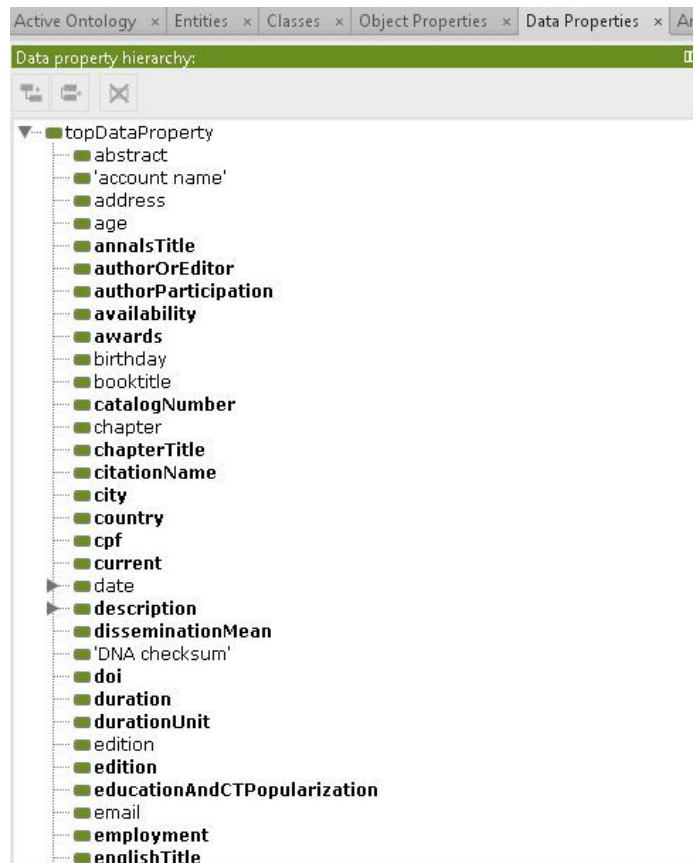
Maioria das classes da ontologia do lattes 3

- Object Properties



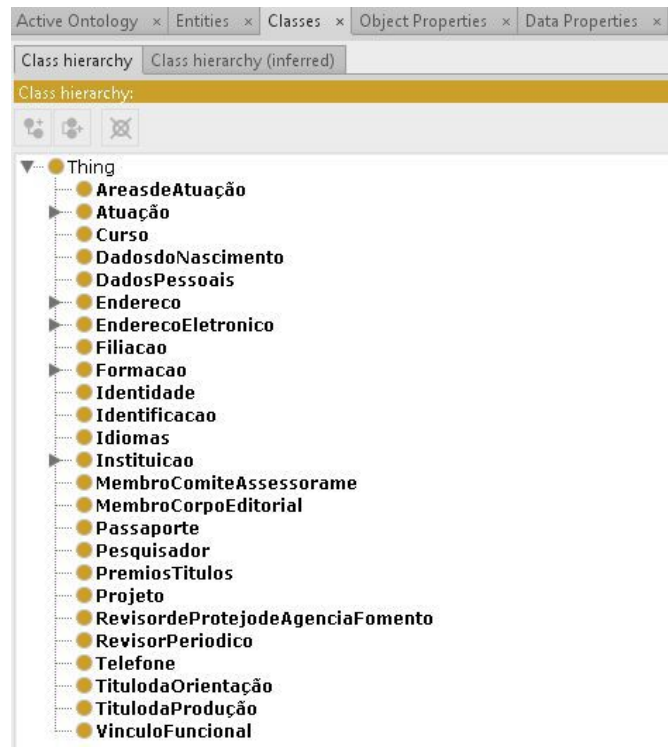
Maioria das object properties da ontologia do lattes 3

- Data Properties



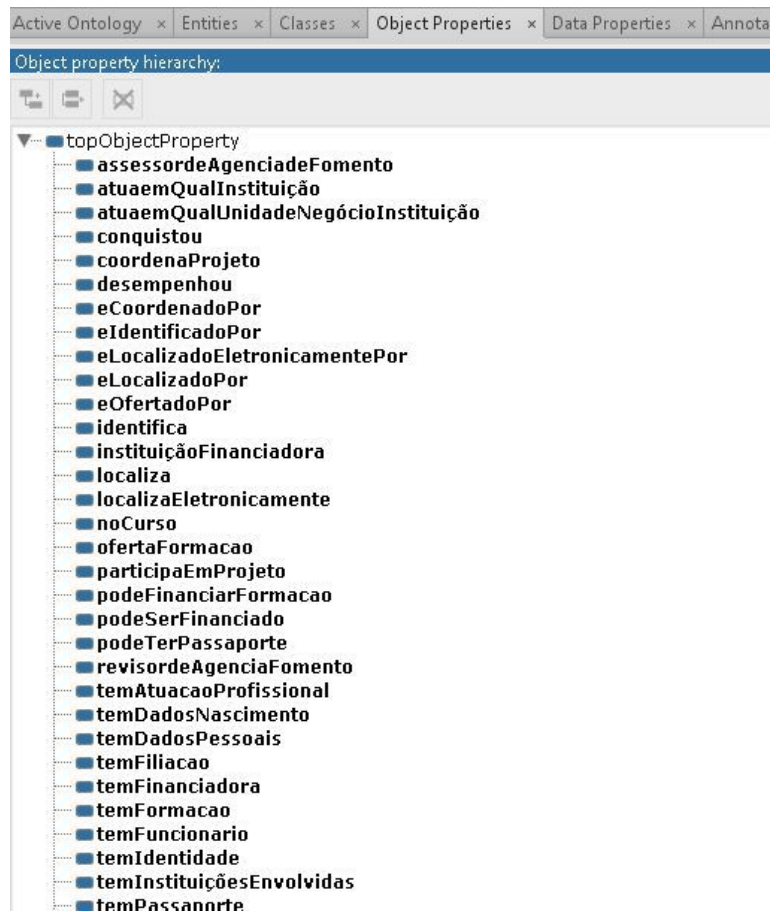
Maioria das data properties da ontologia do lattes 3

- Ontologia 4
  - Classes



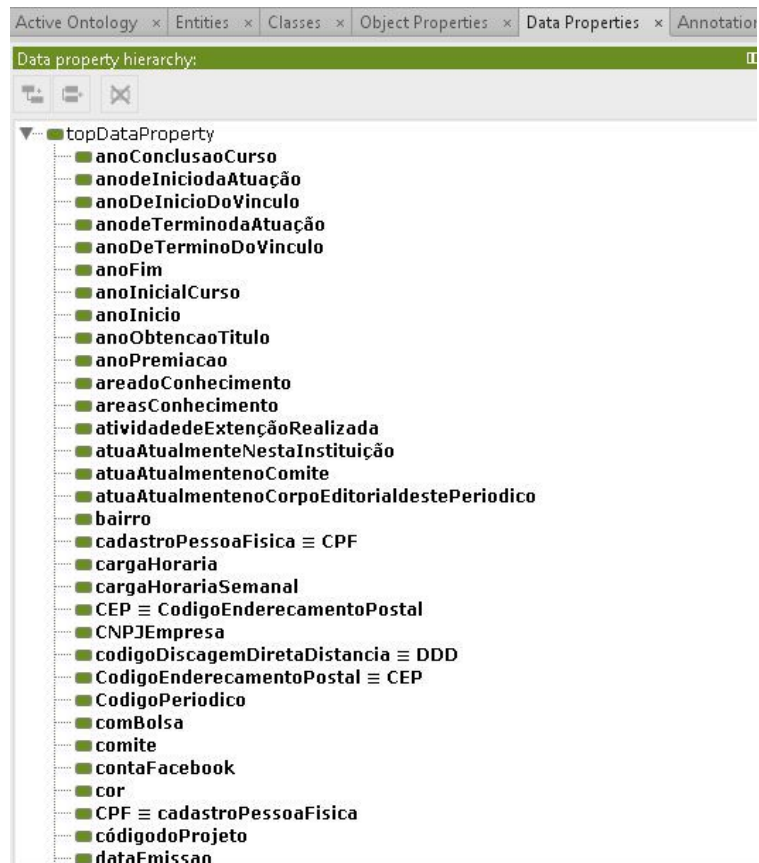
Maioria das classes da ontologia do lattes 4

- Object Properties



Maioria das object properties da ontologia do lattes 4

- Data Properties



Maioria das data properties da ontologia do lattes 4

## APÊNDICE B – Documentos do experimento

- Avaliação de ontologia SEM MODELO

### Avaliação de ontologias

Nome do avaliador:

Para cada questão, dê uma nota entre 0.001 e 0.990. Se não souber como responder, não coloque nenhuma nota (deixe em branco).

### Papel de Substituto

1. Um documento com as competências da ontologia foi criado? São definidas todas as competências da ontologia? Nota:
2. A ontologia responde as competências definidas? Nota:
3. A ontologia reusa outras ontologias? Nota:

### Papel de Compromisso Ontológico

4. A ontologia impõe um compromisso ontológico mínimo? Nota:
5. A ontologia impõe um compromisso ontológico máximo? Nota:
6. As propriedades da ontologia são coerentes com o domínio? Nota:

### Papel de Raciocínio Inteligente

7. Existem axiomas contraditórios? Nota:
8. Existem axiomas redundantes? Nota:

### Papel de Computação Eficiente

9. O raciocinador traz erros de modelagem? Nota:
10. O raciocinador executa de forma rápida? Nota:

### Papel de Expressão Humana

11. Os conceitos estão bem escritos? Nota:
12. A documentação condiz com a modelagem? Nota:
13. Há anotações na ontologia trazendo as definições dos conceitos?  
Nota:

- Avaliação de ontologia COM MODELO

## Avaliação de ontologias

Nome do avaliador:



Você está avaliando a ontologia que modela o sistema de currículos do Lattes. Para isto, siga estes três passos:

---

Primeiro Passo: Defina o tipo da ontologia que está sendo modelada

Tipo 1: A ontologia é do tipo *domain or task ontology*. Tipo de ontologia em que o domínio que a ontologia modela é mais genérico (exemplo: medicina, carro, casa, etc.)

Tipo 2: A ontologia é do tipo *application ontology*. Tipo de ontologia em que o domínio que a ontologia modela é mais específico (exemplo: curso de medicina da UFAL, um programa de computador, etc.)

---

Segundo Passo: Avalie a ontologias de acordo com as questões e dê nota.

Papel de Substituto

1. Um documento com as competências da ontologia foi criado? São definidas todas as competências da ontologia?

Primeiramente, verifique se existe algum documento que explique as competências da ontologia. Se não existir, a nota será automaticamente 0. Se existir, verifique:

- a. Se define o objetivo da ontologia (por ex: esta ontologia modela o domínio de...)
  - i. Notas: 25, 50, 75, 100
- b. Se define o público alvo da ontologia (por ex: esta ontologia será utilizada por...)
  - i. Notas para esta questão: 25, 50, 75, 100

- c. Se define cenários de uso desta ontologia (ou seja, em que situações ela deve ser utilizada).
  - i. Notas: 25, 50, 75, 100

## 2. A ontologia responde as competências definidas?

Se você deu 0 na questão anterior, então as competências não foram definidas e não tem como verificar esta questão. Assim, a nota também será 0. Se as competências existem, verifique:

- a. Se a ontologia responde o que foi definido nas competências.
  - i. Notas: 25, 50, 75, 100

## 3. A ontologia reusa outras ontologias?

Na ferramenta protege, na aba Active Ontology, em Ontology imports (lá embaixo), verifique se existem importações. Se não existirem importações, a nota será 0. Se existirem importações, 100.

## Papel de Compromisso Ontológico

### 4. A ontologia impõe um compromisso ontológico mínimo?

Você só deve verificar esta questão se a ontologia for do tipo 2. Se a ontologia for do tipo 1, pule para a próxima questão.

Nesta questão, verifique se a ontologia não utiliza muita abstração para definir os conceitos. Se a ontologia repleta de abstração (por exemplo: ontologia que modela o site do facebook não precisa estar definindo o que é um sistema de computador, o que é um computador, etc.) a nota será automaticamente 0. Se houver apenas algumas abstrações, dê uma nota entre estas: 25 (muita abstração), 50 (abstração razoável), 75 (quase nenhuma abstração), 100 (sem abstração).

### 5. A ontologia impõe um compromisso ontológico máximo?

Você só deve verificar esta questão se a ontologia for do tipo 1. Se a ontologia for do tipo 2, pule para a próxima questão.

Nesta questão, verifique se a ontologia utiliza conceitos primitivos para definir o domínio que está sendo avaliado (por exemplo, uma ontologia que modela pessoa,

utilizaria os conceitos coisa -> ser vivo -> ser humano -> pessoa para definir o conceito de pessoa). Se a ontologia não utiliza nenhuma abstração, a nota será automaticamente 0. Se houver apenas algumas abstrações, dê uma nota entre estas: 25 (muito específico). 50 (algumas especificidades), 75 (algumas abstrações), 100 (muita abstração).

6. As classes e propriedades da ontologia são coerentes com o domínio?

Nesta questão, verifique se as classes e propriedades estão de acordo com o domínio modelado. Se a ontologia estiver repleta de incoerências (por exemplo, uma ontologia que modela o conceito carro tiver uma classe leão e uma propriedade quantidadeDePatras, ou seja, não existe no domínio) a nota será automaticamente 0. Se houverem algumas incoerências, dê uma nota entre estas: 25, 50, 75. Se não houver nenhuma incoerência, 100.

### Papel de Raciocínio Inteligente

7. Existem axiomas contraditórios?

Nesta questão, verifique se as classes e propriedades (características funcional, transitiva, reflexiva etc) que estão coerentes com o domínio o contradizem (por exemplo: SerVivo ser subclasse de Pessoa numa ontologia que modela pessoa ou CPF não ser uma propriedade funcional, pois uma pessoa não pode ter mais de um CPF). Se a ontologia estiver repleta de contradições, a nota será automaticamente 0. Se houverem algumas contradições, dê uma nota entre estas: 25, 50, 75. Se não houver nenhuma contradição, 100.

8. Existem axiomas redundantes?

Nesta questão, verifique se existem classes e propriedades que modelam a mesma coisa num mesmo sentido (por exemplo, utilizar manga para fruta e para roupa ao mesmo tempo, mouse para o animal e peça de computador ao mesmo tempo, etc). Se a ontologia estiver repleta de redundâncias, a nota será automaticamente 0. Se houverem algumas redundâncias, dê uma nota entre estas: 25, 50, 75. Se não houver nenhuma redundância, 100.

### Papel de Computação Eficiente

### 9. O raciocinador traz erros de modelagem?

Salve todos os seus registros até aqui. Nesta questão, verifique se, executando o raciocinador, retorna algum tipo de erro. Para isto, utilizando a ferramenta protege, clique em Reasoner, e abaixo de configure, selecione algum raciocinador (provavelmente algum deles estará selecionado, fact++, ou pellet, ou hermit) e clique em start reasoner. Se a ontologia estiver repleta de erros (ou se ocorrer algum travamento), a nota será automaticamente 0. Se houverem alguns erros, dê uma nota entre estas: 25, 50, 75. Se não houver nenhum erro, 100.

### 10. O raciocinador executa de forma rápida?

Nesta questão, verifique se o raciocinador está executando de forma rápida. Se ele chega a travar algo, a nota será automaticamente 0. Se houver uma demora, dê uma nota entre estas: 25, 50, 75. Se executar de forma rápida, 100.

## Papel de Expressão Humana

### 11. Os conceitos estão bem escritos?

Nesta questão, verifique se as classes e propriedades da ontologia estão escritas de forma entendíveis e escritas corretamente (de acordo com o português ou outro idioma utilizado). Se a ontologia estiver muito difícil de entender lendo os conceitos, a nota será automaticamente 0. Se houverem alguns erros ou mistura de idiomas, dê uma nota entre estas: 25, 50, 75. Se a ontologia estiver bem escrita, 100.

### 12. A documentação condiz com a modelagem?

Nesta questão, verifique se existe a documentação da ontologia. Se a documentação não existir, a nota será automaticamente 0. Se a documentação existir, verifique:

- a. Se os termos escritos na documentação estão iguais ao da modelagem.
  - i. Notas: 25, 50, 75, 100
- b. Se a documentação explica o que é cada termo e justifica cada detalhe da modelagem.
  - i. Notas: 25, 50, 75, 100

13. Há anotações na ontologia trazendo as definições dos conceitos? Nota:

Nesta questão, verifique se existem anotações que trazem as definições dos conceitos modelados na ontologia. Para isto, utilizando a ferramenta protege, clique no conceito (classe ou propriedade) e verifique no painel do lado, chamado Annotations, se existe algo que explique o que é aquele conceito na modelagem. Se não existir nenhuma anotação, a nota será automaticamente 0. Se houverem algumas anotações, dê uma nota entre estas: 25, 50, 75. Se todos os conceitos possuem anotações, 100.

