

Avaliação da Ontologia Criminal Dirigida por Humanos

Jordan Ferreira Saran, Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Marília, São Paulo, Brasil, jordan.saran@unesp.br

Abstract—Ontology refers to a set of fundamental concepts and relationships, able to identify how people interpret the situation. In the field of risk management, ontology provides the development of inference models that assist in the perception and construction of better decision-making in critical situations, especially when they relate to areas that involve lives human, environmental and heritage. However, ways to evaluate the characteristics of the classes, properties, axioms, efficiency and objectivity of the ontologies are necessary to ensure a better representativeness of the information, and also, if it is fulfilling the ontological commitment. And clearly transparently the information related to the domain to which it belongs. This article aims to present the process of evaluation of the human-driven ontology Brazilian crime ontology, in order to ascertain the fulfillment of its objectives related to the criminal domain, specifically in the semantic construction of crime situations, in case of robbery and theft.

Index Terms—Ontology, Critical Domain, Ontology Evaluation

Resumo—Ontologia se refere a um conjunto de conceitos fundamentais e relações, capaz de identificar como as pessoas interpretam a situação. No domínio de gerenciamento de risco, a ontologia propicia no desenvolvimento de modelos de inferência que auxiliam na percepção e construção de melhor tomada de decisão em situações críticas, principalmente quando se relacionam em áreas que envolvem vidas humanas, meio ambiente e patrimônio. Entretanto, formas de avaliar as características das classes, propriedades, axiomas, eficiência e objetividade das ontologias são necessárias para garantir uma melhor representatividade da informação, e também, se está cumprindo com o compromisso ontológico e transparecendo de forma clara as informações relacionadas ao domínio a que pertence. Este artigo tem como objetivo apresentar o processo de avaliação da ontologia *human-driven brazilian crime ontology*, com intuito de averiguar o cumprimento de seus objetivos relacionados ao domínio criminal, especificamente na construção semântica de situações de crimes, no caso, roubo e furto.

Palavras-chave—Ontologia; Domínio crítico; Avaliação de ontologia

I. INTRODUÇÃO

Com o crescimento do uso de meios digitais para armazenamento e divulgação da informação gerou uma grande demanda por mecanismos que auxiliem o usuário a recuperar a

informação desejada. Atualmente, os pesquisadores de várias áreas da ciência, como computação, linguística e ciência da informação entre outros, tentam desenvolver e aprimorar esses mecanismos. Uma técnica interessante para intermediar essa busca é a utilização de ontologias com o propósito de guiar o usuário na navegação por entre os conceitos do domínio até a informação desejada [1].

Vários sistemas hoje em dia, se propõem a utilizar a ontologia para auxiliar em suas tomadas de decisão, principalmente quando se relacionam em áreas que envolvem vidas humanas, meio ambiente e patrimônio [2]. A ontologia *URBANITY(hUman-dRiven BrAziliaN crIme ontologY)* iniciada por [3] que é foco deste trabalho, tem o propósito de auxiliar no processo de avaliação de situação de risco, na qual a mesma tem o objetivo de representar situações de crime, advindas de boletins de ocorrência, agregadas de propriedades e relacionamentos semânticos com vistas de uma melhor representatividade.

No entanto, a ontologia necessita de uma avaliação para averiguar se a mesma realmente está cumprindo com seu objetivo. No entanto, a ontologia necessita de uma avaliação para averiguar se a mesma realmente está cumprindo com seu objetivo. Apesar da literatura envolvendo esse tema seja escassa, [4] destaca a necessidade de haver formas de avaliar uma ontologia, ao observar seus aspectos, como, características das classes, propriedades, axiomas, escopo da ontologia e entre outras métricas para obter resultados e compreender se realmente a ontologia contém compromisso ontológico, eficiência e clareza em expressas as informações relacionadas ao domínio a qual foi criada.

O objetivo desse trabalho é a aplicação de uma metodologia de avaliação de ontologia, ou seja, onde seja capaz de analisar se a mesma está cumprindo seus objetivos corretamente e consiga a partir desses resultados, prosseguir com sua expansão e evolução. Cabe ressaltar que esse trabalho não possui o foco no desenvolvimento de uma ontologia.

II. ONTOLOGIAS EM DOMÍNIOS CRÍTICOS

De acordo com [5] o termo “ontologia” teve sua origem no ramo da Filosofia, a qual estudava a natureza do “ser” e a “existência”. Para eles, os filósofos à Ontologia têm como propósito, explicar todas as coisas do mundo, onde se estabelece sistematicamente uma linha conceitual, já para a área

da computação este termo possui características que o diferem de sua área de criação. Entretanto, é possível vê-la em áreas como Web Semântica e Inteligência Artificial aplicando Ontologia como subsídios para compreender e codificar o conhecimento para solucionar problemas existentes nestas áreas, por meio de relações existentes entre os conceitos de um determinado domínio.

De acordo com [2], no domínio crítico, tomar decisões é uma tarefa complexa, pois além do gerenciamento de sistemas computacionais, o ser humano deve gerenciar uma variedade de outras variáveis de ambiente presente naquele situação, por exemplo, um piloto de aeronave, ao estar em campo necessita ao mesmo tempo analisar as informações presentes no sistema de navegação da aeronave, analisar o ambiente ao seu redor, e assim, interpretar todas as informações e tomar uma decisão. Ainda neste contexto, eles definem que o uso da computação baseada em ontologias auxilia no desenvolvimento de um modelo de inferência no qual processos computacionais conseguiriam reter a conscientização da situação.

Para [6], o uso de ontologias são para representar áreas críticas caracterizadas pela presença de incerteza. Para este fim, tal trabalho realiza um estudo sobre técnicas e ontologias capazes de representar a semântica complexa presente nessas áreas, juntamente com a incerteza que estas contêm. Como uma solução, é proposto o uso de uma ontologia probabilística (PR-OWL) que seja capaz de representar relações multientidades presentes em uma ontologia e ainda ser capaz de representar a probabilidade e incerteza da informação.

A fim de validar as técnicas acima mencionadas pelo próprio trabalho, foi realizado um estudo de caso em uma área crítica, a qual requer uma grande quantidade de informações providas de diversas fontes, e que devido ao domínio, forma de captura e outros fatores já apresentam pontos de incerteza, sendo o domínio deste trabalho a detecção de fraude nos contratos no Brasil, onde são analisadas as possíveis fontes de dados e como

elas podem ser usadas para encontrar a melhor aplicação possível de uma ontologia probabilística.

Ao observar o uso da ontologia no domínio crítico, como visto no trabalho de Carvalho et al (2013) e a necessidade de propor formas de avaliar as técnicas utilizadas no mesmo, Hicks (2017) também destaca a necessidade de avaliar ontologias após sua criação, onde o objetivo é averiguar se as mesmas estão cumprindo com seus objetivos.

III. AVALIAÇÃO DA ONTOLOGIA URBANITY

A URBANITY [3], é a ontologia abordada neste trabalho e está inserida dentro do contexto de informações criminais no Brasil. O foco da mesma é representar informações oriundas de inteligência humana, ou seja, relatos humanos sobre um crime, neste caso, Boletins de Ocorrências (BOs) de Roubo e Furto. Com base nas análises realizadas sobre os dados criminais, os BOs acessados via Lei de Acesso à Informação (LAI), os quais trazem dados descritivos e históricos. Estes contêm detalhes importantes sobre o ambiente, sendo possível identificar e caracterizar os principais elementos que compõem situações de crime, processo essencial para a construção de uma ontologia.

Contendo diversas classes, onde estas representam os principais elementos informacionais que compõem uma situação criminal, a URBANITY traz classes como Criminoso, Vítima, Denunciante, Objeto Roubado, Arma, Local respeitando questões como hierarquia e simetria entre classes e propriedades. Além das classes, tem as propriedades que relacionam as classes e criam uma dependência ou relação semântica. Dentre as propriedades existentes, alguns exemplos são: foge, fere, rouba, furta, pertence, é, tem. Por exemplo a propriedade “tem”, relaciona Vítima com Objeto Roubado, logo a vítima pode tem um objeto, outro exemplo, a propriedade “foge”, relaciona Criminoso com Local. A Figura 1 traz a representação gráfica da URBANITY, demonstrando a quantidade de classes (retângulos com círculo amarelo),

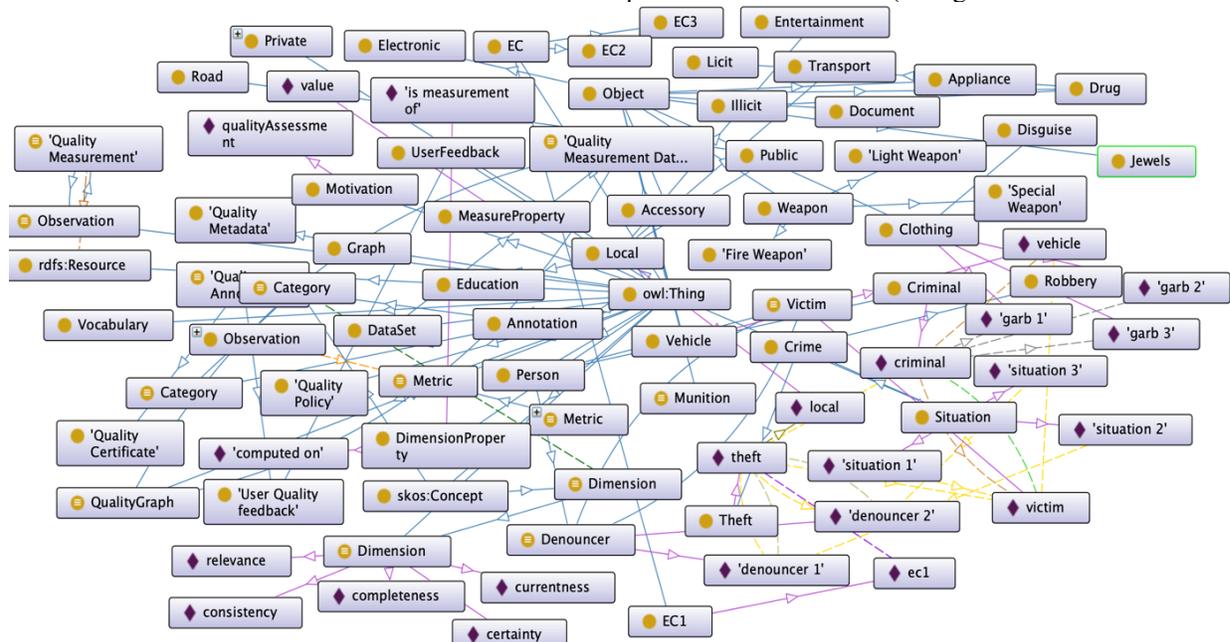


Figura 1 - Representação gráfica da URBANITY.

propriedades (setas) e algumas instâncias das classes exemplificadas (retângulos com losango roxo). Importante ressaltar que a estrutura da ontologia está em inglês devido o perfil do grupo que a desenvolveu.

Para validar a ontologia URBANITY é necessária aplicar uma metodologia de avaliação de ontologias, onde, para este trabalho é adotado a Metodologia FOCA desenvolvida por [7], na qual é uma metodologia genérica que se enquadra em diversos tipos de ontologias, e com isso, a ontologia deste artigo.

A Metodologia FOCA é baseada em três principais fundamentos, sendo eles: 1) Objetivo, Questão e Abordagem Métrica para avaliação empírica ou conhecida pela sigla GQM (*Goal, Question, Metric*) advinda do inglês; 2) os objetivos da metodologia se baseiam nos cinco princípios da Representação do Conhecimento (RC) e suas métricas, são baseadas em critérios de avaliação; 3) cada ontologia é avaliada de acordo com o tipo de ontologia.

As regras da RC, que são os objetivos da avaliação, foram consideradas como o “objetivo” de acordo com a abordagem da GQM. Os critérios de qualidade, que são as unidades de medida, foram considerados como as métricas da abordagem GQM. Por fim, as questões que se preocupam com as métricas foram consideradas como questões da abordagem GQM. A Tabela 1 demonstra as razões entre as perguntas e papéis da RC e a Tabela 2 ilustra às perguntas relacionadas aos objetivos de avaliação e as métricas a cada pergunta.

No que diz respeito à metodologia, a mesma se consiste em três etapas para realizar a avaliação, sendo elas [7]:

- **Etapa 1: verificação do tipo da ontologia:** as questões só são verificadas de acordo com o tipo de ontologia, onde o avaliador deve verificar se o tipo de ontologia pode ser: 1) Ontologia de domínio ou tarefa, onde caso seja do tipo 1, a questão 4 não deve ser verificada; 2) Ontologia de aplicação, caso seja do tipo 2, a questão 5 não deve ser verificada.
- **Etapa 2: verificação das perguntas:** o avaliador responderá às perguntas após a abordagem GQM, logo após, estabelecerá uma nota para cada pergunta.
- **Etapa 3: Verificação da qualidade:** É calculado a qualidade da ontologia utilizando o modelo de Regressão Beta, onde seu resultado final é um *score* entre 0 e 1. Este calculado pode ser realizado de duas maneiras, sendo elas: 1) Qualidade Total, onde o avaliador calcula a qualidade da ontologia considerando os 5 papéis da Representação do Conhecimento ou; 2) Qualidade Parcial, no qual o avaliador calcula a qualidade da ontologia considerando apenas alguns papéis da RC, por exemplo, considerar apenas inteligência racional. A Equação 1 demonstra o modelo de Regressão Beta.

Equação 1 - Modelo de Regressão Beta (Traduzido e adaptado de [7])

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

Tabela 1 - Regras correspondentes as perguntas junto as suas justificativas do GQM da metodologia FOCA (Traduzido e adaptado de [7])

Regras	Perguntas	Justificativa
Substituto	Q1	Esta pergunta aborda o documento que contém as questões de competência, os principais termos e os objetivos da ontologia.
	Q2	Esta questão aborda a verificação da coerência entre a documentação que contém as questões de competência, os principais termos e os objetivos da ontologia com o modelo.
	Q3	Esta pergunta aborda a reutilização de conceitos existentes para modelar o mundo real.
Compromissos ontológicos	Q4	Esta pergunta aborda a escolha da melhor representação para um domínio específico.
	Q5	Esta pergunta aborda a escolha da melhor representação para um domínio mais abstrato.
	Q6	Esta pergunta aborda a verificação da coerência com o mundo real.
Inteligência racional	Q7	Esta pergunta aborda a verificação de um raciocínio correto da ontologia.
	Q8	Esta pergunta aborda a verificação de um raciocínio correto da ontologia.
Eficiência computacional	Q9	Esta pergunta aborda um bom desempenho computacional (bem-sucedido).
	Q10	Esta pergunta aborda um bom desempenho computacional (velocidade).
Expressão Humana	Q11	Aborda a compreensão fácil da modelagem.
	Q12	Esta pergunta aborda a fácil compreensão da modelagem.
	Q13	Esta pergunta aborda a fácil compreensão da modelagem.

IV. RESULTADOS PARCIAIS

O método utilizado para a aplicação da metodologia FOCA neste trabalho, é por meio de um formulário eletrônico disponibilizado pela plataforma *Google Forms*, onde, o mesmo foi enviado via e-mail para um grupo de pessoas conhecedoras de Ontologias, com um nível de conhecimento variado. O formulário pode ser acessado por meio deste link (<https://forms.gle/WN5fqdAAgNb3xKbw9>).

O formulário foi elaborado seguindo todas as regras da metodologia FOCA, e também contém todas as informações para o acesso a própria ontologia e informações detalhadas de como utilizar as ferramentas necessárias para o acesso, entretanto, cabe ressaltar que o idioma contido no formulário é o inglês, pois ser o idioma universal.

Para os avaliadores, a ontologia desenvolvida neste trabalho, se enquadra no tipo 2, definindo como ontologia de aplicação. Se baseando neste resultado, é possível se direcionar para os cálculos necessários para o tipo 2 da ontologia. A Tabela 3 demonstra os valores obtidos pelas respostas das perguntas de um dos avaliadores, assim como a média obtida a partir da soma gerada com base no objetivo de cada grupo de perguntas.

Com as médias obtidas de cada objetivo, é aplicado seus valores na equação da “Qualidade Total”, de acordo com a terceira etapa da metodologia FOCA, e em sequência cada avaliador obtém um *score* variando entre 0 e 1. Após a obtenção do *score* de cada avaliador o procedimento final é realizar uma média aritmética entre todos os *scores* obtidos dos avaliadores, onde, neste caso o *score* final obtido até o presente momento é

Tabela 3 - GQM da Metodologia FOCA (Traduzido e adaptado de [7]).

Objetivos	Perguntas	Métricas
G1. Verifique se a ontologia está em conformidade com o substituto.	Q1. As questões de competência foram definidas?	M1. Completude
	Q2. As perguntas de competência foram respondidas?	
	Q3. A ontologia reutiliza outras ontologias?	M2. Adaptabilidade
G2. Verificar se a ontologia cumpre compromissos ontológicos.	Q4. A ontologia impôs um compromisso ontológico mínimo?	M3. Concisão
	Q5. A ontologia impôs um compromisso ontológico máximo?	
	Q6. As propriedades ontológicas são coerentes com o domínio?	M4. Consistência
G3. Verifique se a ontologia está em conformidade com o raciocínio inteligente.	Q7. Existem axiomas contraditórios?	
	Q8. Existem axiomas redundantes?	M3. Concisão
G4. Verifique se a ontologia cumpre eficiência computacional.	Q9. O motor de inferência (do inglês conhecido como <i>Reasoner</i>) trouxe erros de modelagem?	M5. Eficiência computacional
	Q10. Será que o motor de inferência executa rapidamente?	
G5. Verifique se a ontologia está em conformidade com a expressão humana.	Q11. A documentação é consistente com a modelagem?	M6. Clareza
	Q12. Os conceitos foram bem escritos?	
	Q13. Existem anotações na ontologia que mostram as definições dos conceitos?	

Tabela 2 - Exemplo de resultados obtidos por um dos avaliadores.

Grupo	Perguntas	Valores	Média	Grupo	Perguntas	Valores	Média
G1	Q1	100%	100%	G3	Q7	100%	100%
	Q2	100%			Q8	100%	
	Q3	100%		G4	Q9	75%	84,5%
G2	Q4	100%	Q10		100%		
	Q5	-	G5	Q11	75%	58,33%	
	Q6	100%		Q12	75%		
				Q13	25%		

de 0,87, como pode ser observado na Tabela 4.

Tabela 4 - Média aritmética a partir dos resultados obtidos dos avaliadores.

Avaliadores	Valores(%)
Avaliador 1	0,869281046
Avaliador 2	0,873417722
Avaliador 3	0,859812733
Avaliador 4	0,873417722
Média	0,87

V. CONCLUSÕES PARCIAIS

Com base nos resultados obtidos utilizando a metodologia FOCA para realizar a avaliação da URBANITY, pode-se concluir parcialmente que a mesma se enquadra como ontologia de aplicação, onde se relaciona claramente com o domínio criminal e está comprimindo com seus objetivos, e assim, proporcionando clareza, concisão, consistência, completude por parte das informações e relações com o domínio e por fim, adaptabilidade e eficiência computacional. Entretanto, cabe ressaltar que a avaliação ainda está em andamento.

Um ponto a ser destacado é a forma da ontologia ser avaliada,

onde, ao utilizar como meio um formulário eletrônico e solicitar ao avaliador respostas referentes a um conjunto de perguntas pré-definidas, pode acarretar uma demora na obtenção dessas respostas por parte dos avaliadores, e com isso, atrasar o desenvolvimento ou expansão da ontologia. Todavia, algumas questões surgem e demandam ações futuras no trabalho. Como o levantamento e avaliação de novas classes, propriedades e axiomas, relacionadas ao domínio criminal e que auxiliem na evolução e expansão da ontologia.

REFERENCIAS

- [1] MOREIRA, A. **Uso de ontologia em sistemas de informação computacionais.** *Perspectivas Em Ciência Da Informação*, v. 7, pp.49–60, 2002
- [2] KOKAR, M. M. MATHEUS, C. J. BACLAWSKI, K. **Ontology-based situation awareness.** *Information Fusion*, Elsevier B.V., v. 10, n. 1, p. 83–98, 2009. ISSN15662535.
- [3] BOTEAGA, L. C. et al. **Quantify: An Information Fusion Model Based on Syntactic and Semantic Analysis and Quality Assessments to Enhance Situation Awareness.** In E. Bosse & G. Rogova (Eds.), *Information Quality in Information Fusion and Decision Making* (1st ed., pp. 563–586). Springer International Publishing. <https://doi.org/10.1007/978-3-030-03643-0>, 2019.
- [4] HICKS, A. **Metrics and methods for comparative ontology evaluation.** *Ciência Da Informação*, v. 46, pp.34–42. <https://doi.org/10.18225/ci.inf.v46i1.4012>, 2017.

- [5] ISOTANI, S. BITTENCOURT, I. I. **Dados Abertos Conectados**. Novatec Editora. ISSN 24470821. ISBN 978-85-7522-449-6. 2015.
- [6] CARVALHO, R. N. et al. **Probabilistic Ontology and Knowledge Fusion for Procurement Fraud Detection in Brazil**. Uncertainty Reasoning for the Semantic Web II, p. 19–40, 2013.
- [7] BANDEIRA, J., BITTENCOURT, I. I., ESPINHEIRA, P., ISOTANI, S. **FOCA: A Methodology for Ontology Evaluation**. 2016.