

## Provocações acadêmicas: ontologias, tesouros, documentos, conteúdo de documentos, e... Unicórnios

*AN ACADEMIC GIVE-AND-TAKE: ONTOLOGIES, THESAURI, DOCUMENTS, DOCUMENT CONTENTS AND ... UNICORNS*

Maurício B. Almeida (ECI/UFMG)

*mba@eci.ufmg.br*

**Resumo:** Uma provocação nietzschiana sobre os valores da vida questiona: “qual o valor nos fez escolher os valores que seguimos?”. Na vida acadêmica, tal provocação deveria ser regra: questionar e questionar o que fazemos e o que nos levou a fazer o que fazemos, esse é o papel do pesquisador. É nesse espírito de provocação saudável que se apresenta esse pequeno ensaio no contexto da Organização do Conhecimento. São apresentadas questões que ainda surgem a toda hora no campo de pesquisa. Como provocação assumida que é, não tem pretensão de estabelecer a verdade, mas fomentar o debate acadêmico tão importante para o progresso da pesquisa.

**Palavras-chave:** Ontologia Aplicada. Organização do Conhecimento.

**Abstract:** A Nietzschean provocation about the values of life argues: "which value made us choose the values we follow?" In the academic life, this provocation should be a rule: to question and question again what we do and what led us to do what we do, this should be the role of a researcher. It is within this spirit of a healthy give-and-take that we present this little essay in the context of Knowledge Organization field. We present issues that still arise several times in our research field. As this essay is an assumed provocation, we do not have in mind to establish the truth, but only to foster the academic debate as important tool for the research progress.

**Keywords:** Applied Ontology. Knowledge Organization.

### 1 INTRODUÇÃO

Tradicionalmente fundamentada em técnicas e teorias milenares de classificação, a Organização do Conhecimento (OC) é um campo vibrante e dinâmico da Ciência da Informação, essencial para a ciência e mesmo para a vida cotidiana.

Esse protagonismo da OC se consubstancia nas habilidades dos profissionais ali formados em representar, organizar e recuperar o conhecimento produzido nas mais diversas áreas do pensamento.

Recentemente, passaram a compor a pesquisa em OC, técnicas provenientes da disciplina da Ontologia Aplicada, a qual traz novos *insights* filosóficos visando a manipulação automática do conhecimento registrado em meio digital. Nesse ensaio, baseado em indagações sobre a relação entre ontologias e conceitos seminais da OC já apresentadas em artigos anteriores, discutem-se questões simples, ainda que interessantes e provocativas.

Por limitações de espaço, não se apresenta o *background* necessário sobre ontologias e OC, considerando que o leitor é minimamente iniciado nesses assuntos. O restante do presente ensaio está organizado em cinco seções: as três primeiras são encabeçadas por perguntas provocativas, a quarta mostrar erros de classificação, e a última traz considerações finais.

## 2 ONTOLOGIAS E LINGUAGENS DOCUMENTÁRIAS

Duas questões relacionadas compõem a pergunta dessa seção, à saber: (a) ontologias são linguagens documentárias? (b) se não são podem ser usadas como tal?

Uma resposta à primeira parte da pergunta, à questão (a), a resposta mais rigorosa é “não”, ontologias não são linguagens documentárias (LDs). Ontologias são criadas como teorias científicas formais, e não como tipos linguagens, apesar de fazerem uso de alguma. Um renomado filósofo explica na primeira frase de seu livro: “*Esse livro é um livro sobre o mundo. Estou preocupado com ontologia, não com linguagem*” (JOHANSSON, 2004, p. 1).

O aviso é claro: existe uma diferença entre as entidades do mundo, por si, e os meios que as pessoas usam para descrever essas mesmas entidades. A ontologia se refere às entidades do mundo, não à convenção usada para se referir as entidades, ou seja, à linguagem.

A resposta para a segunda parte da pergunta, a questão (b), é um “sim”, ou seja, mesmo que ontologias não sejam linguagens documentárias, podem ser usadas como tal. Isso é verdadeiro caso esteja-se referindo ao uso de linguagens de representação adotadas para especificar ontologias, como OWL ou RDFS. O resultado nesse caso será uma LD implementada em um sistema de recuperação da informação da Web, não uma ontologia.

## 3 ONTOLOGIAS E TESAUROS

Assim como na seção 1, duas questões relacionadas compõem a pergunta dessa seção, à saber: (a) ontologias e tesauros exibem a mesma capacidade de representação? (b) se não exibem, onde residem as principais diferenças?

Já é amplamente sabido que os dois instrumentos mencionados nas questões não são diretamente comparáveis, uma vez que servem a objetivos distintos. Ao criar tesauros, empregam-se relações *broader-than* e *narrower-than* para organizar taxonomicamente o vocabulário. Essas relações surgem, por exemplo, em vocabulários controlados como o *Medical Subject Headings* (MeSH)<sup>9</sup>, criado para indexação e catalogação de informação médica. São exemplos do MeSH: i) *FetalBlood narrower-than Blood*; e ii) *Plasma narrower-than Blood*.

Do ponto de vista dos tesauros, essas relações atendem as necessidades de recuperação de documentos: uma consulta usando o termo *blood* retornará tanto artigos sobre *fetal\_blood*, quanto sobre *blood\_plasma*. Entretanto, do ponto de vista ontológico, as duas relações representam tipos

<sup>9</sup> <https://www.nlm.nih.gov/mesh/>

diferentes: enquanto na primeira relação *fetal\_blood* é um tipo de *blood*, na segunda *plasma* é parte de *blood*. Essas diferenças são importantes porque enquanto tesouros são criados para uso por pessoas, ontologias são criadas para consumo por máquinas.

#### 4 ONTOLOGIAS, DOCUMENTOS E SEU CONTEÚDO

Nessa seção, a pergunta que norteia a discussão é única: se ontologias são uma descrição de mundo, em que local do mundo se localizam documentos e seu conteúdo?

A parte da pergunta sobre documentos como entidades físicas, não apresenta grandes desafios e, portanto, será omitida aqui por limitações de espaço. A parte da pergunta sobre onde se localiza o conteúdo de documento em uma ontologia não é trivial e nem ontologias como a *Basic Formal Ontology* (BFO) (GRENON; SMITH, 2004) ou a *Information Artifact Ontology* (IAO) (IAO, 2013) se arriscam a definir o conteúdo de documentos.

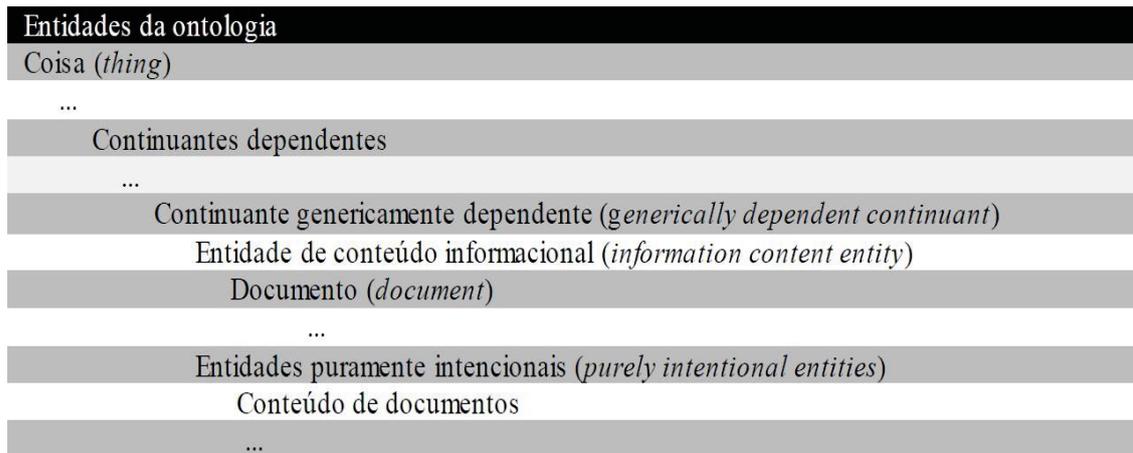
A questão reside no fato de que, no âmbito da IAO e BFO, uma *entidade de conteúdo informacional* (nível 6 da FIG. 1) sempre é sobre (*is\_about*) alguma entidade do mundo, considerado o mundo espaço-temporal. Entretanto, é fácil imaginar livros cujo conteúdo versa sobre unicórnios, coelhos da páscoa, dentre outras ficções, entidades que não existem no mundo espaço-temporal. Para lidar com essa situação, Brochhausen et al. (2013), propõem uma entidade chamada *continuante genericamente dependente*<sup>10</sup>, uma irmã da *entidade de conteúdo informacional*<sup>11</sup>, denominada *entidade puramente intencional*. Abaixo dessa entidade, é possível abrigar a entidade *conteúdo de documentos* que se refere tanto a coisas do mundo espaço-temporal, quanto coisas imaginadas sobre o mundo.

---

<sup>10</sup> Na BFO, continuantes genericamente dependentes são propriedades de outra entidade, mas tal entidade pode variar.

<sup>11</sup> Na IAO, entidades de conteúdo informacional dependem genericamente de outras e mantêm relações de *aboutness*.

FIGURA 1 – Fragmento de hierarquia com entidades “Documento” e “Conteúdo”



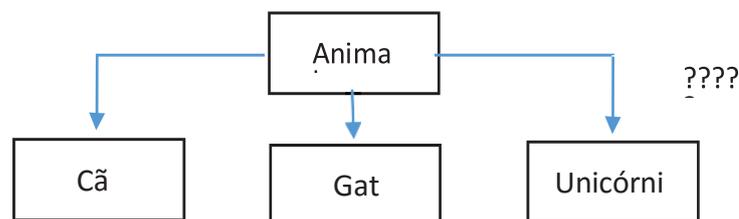
Fonte: ALMEIDA; MENDONÇA; AGANETTE, 2013, p. 15.

## 5 CLASSIFICAÇÃO DE UNICÓRNIOS, CIRURGIAS, PESSOAS, ...

Apresentou-se a localização de entidades ficcionais na ontologia, as quais poderiam ser classificadas abaixo de *entidades puramente intencionais* caso se deseje representar um conteúdo de um livro que tem como assunto unicórnios, por exemplo. Entretanto, ainda cabe questionar onde classificar tais entidades em um sistema de classificação. *Será um unicórnio um animal? Um coelho da páscoa é um coelho? Mudando um pouco o foco dos exemplos: uma cirurgia cancelada é uma cirurgia? Um diretor é uma pessoa?*

Não é difícil mostrar que a classificação desses casos em geral leva a erros básicos e podem levar à problemas em sistemas ontológicos capazes de inferência. A FIG. 2 apresenta uma hierarquia para “*unicórnio é-um Animal*”.

FIGURA 2 – Classificação de animais (setas são relação é-um)



Fonte: Elaborada pelo autor.

A ontologia contém entidades – os *continuantes* – as quais existem totalmente durante todo tempo de sua existência, persistem através do tempo mantendo identidade. Exemplos são: um coração, um animal, a cor do tomate, uma orquestra, a disposição da banana em apodrecer, etc. De fato, animais como cães e gatos são entidades chamadas nas ontologias de *continuantes independentes*: entidades portadoras de qualidades, das quais outra entidade pode depender ou ser

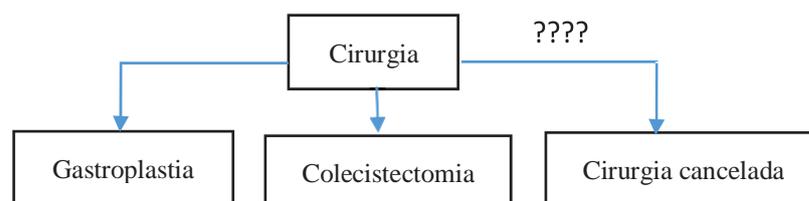
inerente a (a cor vermelha é inerente ao tomate), e que não podem depender de mais nada. Em suma, são entidades independentem da mente humana para existir.

Por outro lado, os unicórnios são entidades nunca vistas, das quais a ciência nada tem a dizer. Dinossauros também são entidades nunca vistas, mas, ao contrário dos unicórnios, a ciência comprova sua existência. O que ocorre é que unicórnios existem apenas na mente das pessoas, não são *continuanes independentes*, são, de fato, *continuanes dependentes*. Um *continuanes dependente* é uma entidade que mantém inerência (ou dependência) de outras entidades, por exemplo: o peso do José não existe sem José; a brancura do queijo depende do queijo; o pensamento de Einstein depende da existência de Einstein. Da mesma forma, unicórnios dependem da mente de alguém que os imagina, e, portanto, não são objetos nem animais, são “propriedades”. Nas ontologias de orientação aristotélica, as propriedades são chamadas *qualidades*. Sendo unicórnios qualidades, a classificação da FIG. 2 é errônea.

O caso do “*coelho da Páscoa é um coelho*” segue um raciocínio similar ao do Unicórnio. O exemplo apenas mostra outra entidade fictícia que pode carecer de classificação, sem que esteja embutida em um assunto ou conteúdo de um livro.

O caso da *Cirurgia Cancelada é um Cirurgia* é diferente, mas ilustra um erro de classificação em que não há preocupação em definir adequadamente a entidade antes localizá-la no sistema de classificação.

FIGURA 3 – Classificação de cirurgias (setas são relação é-um)



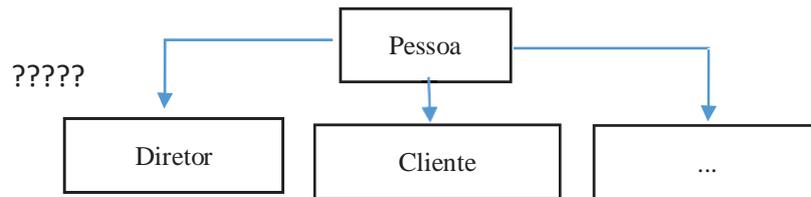
Fonte: Elaborada pelo autor.

Em ontologia existem entidades que “ocorrem” no tempo, e, portanto, denominadas *ocorrentes*. Ocorrentes se desdobram ao longo do tempo e os exemplos prototípicos dessas entidades são os processos: o processo de respiração, o processo de matrícula, etc. Uma cirurgia é um processo e, além de se desenrolar ao longo do tempo, exibe uma característica ontológica básica dos processos: a existência de um processo, requer a existência de um *continuanes independente* que participe do processo. De fato, não existe processo de matrícula sem o aluno; não existe digestão sem a comida; e não existe cirurgia sem paciente. Como uma cirurgia cancelada não possui paciente, ela não é processo e, portanto, uma cirurgia cancelada não é uma

cirurgia. A cirurgia cancelada é, na verdade, um “plano”. Planos, em ontologias, são *entidades de conteúdo informacional*, as quais dependem genericamente de outra entidade, além de manter uma relação de *aboutness* com outras. A classificação apresentada na FIG. 3 é errônea.

O caso “*diretor é-um pessoa*” é um erro comum cometido na modelagem de sistemas de informação e bancos de dados, por profissionais provenientes da informática.

FIGURA 4 – Classificação de pessoas (setas são relação é-um)



Fonte: Elaborada pelo autor.

A FIG. 4 exibe um erro comum: um diretor não é uma pessoa, um diretor é um *papel* que uma pessoa assume durante um período. Papéis são definidos em ontologias como *continuantes dependentes*, os quais não são essenciais para a identidade do continuante.

A consequência desse tipo de erro para sistemas baseados em ontologias é que podem-se obter inferências automáticas totalmente sem sentido. Considere-se, como exemplo, o papel de ser diretor, que é um papel; e considere-se também o papel de um coração artificial em bombear sangue, o qual também é um papel. Sendo esses dois tipos de papéis considerados, e na existência de uma classificação errônea a da FIG. 4, um motor de inferência poderia, hipoteticamente, gerar o fato de que o papel de bombear sangue é papel de uma pessoa. Isso certamente não faz sentido. Esse é um exemplo simples, dentre diversos outros possíveis, que não são apresentados aqui por limitações de espaço.

## 6 CONSIDERAÇÕES FINAIS

O que se pretendeu mostrar ao longo desse ensaio-provocação é a necessidade e entender a origem das entidades sob classificação para classificá-las, independentemente do tipo de sistema de organização do conhecimento. A teoria da Ontologia Aplicada pode ajudar em muito nesse aspecto.

Ainda, cabe destacar que não apenas por se usar um editor de ontologias para construir um sistema de organização de conhecimento, está se garantido que tal sistema seja de fato uma ontologia. Assim como carros, pessoas, árvores, etc., todas as entidades do mundo têm características que as determinam como tal; também as ontologias têm características que precisam

estar presentes para a estrutura seja de fato uma ontologia. É também comum ouvir que inferências não são necessárias, mas parece contraditório se dispender tanto esforço na construção de ontologias e desprezar sua principal utilidade.

Muitos dos enganos cometidos, como os exemplos apresentados aqui, se devem ao desconhecimento ou falta de crença na necessidade de uso de regras de classificação formais. Fica a consideração de que, se as regras de classificação não precisam ser seguidas, então o resultado pode não ser considerado um sistema de classificação.

## REFERÊNCIAS

ALMEIDA, M. B. Revisiting ontologies: a necessary clarification. **Journal of the American Society of Information Science and Technology**, v. 64, n. 8, p. 1682-93, 2013.

ALMEIDA, M. B.; MENDONÇA, F. M.; AGANETTE, E. C. Interfaces entre ontologias e conceitos seminais da Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: ANCIB, 2013.

BROCHHAUSEN, M. et al. Developing a semantically rich ontology for the biobank administration domain. **Journal of Biomedical Semantics**, v. 4, n. 23. 2013.

GRENON, P.; SMITH, B. **SNAP and SPAN**. 2004. Disponível em: <<http://ontology.buffalo.edu/smith/articles/>>. Acesso em: 10 de jun. de 2013.

IAO: Information Artifact Ontology. 2013. Disponível em: <<http://code.google.com/p/information-artifact-ontology/>>. Acesso em: 16 de jun. 2013.

JOHANSSON, I. **Ontological investigations**. Frankfurt: Ontos Verlag, 2004.