

Resumo

Introdução

Capítulo 1

Ontologias

A palavra *Ontologia* veio do grego, assim como vários outros termos que se referem à áreas de estudo. Seu significado, no entanto, é muito mais abstrato. Diferentemente de *Biologia*, que é "o estudo da vida", a palavra cujo plural dá nome a esse capítulo quer dizer "o estudo do ser enquanto ser". O dicionário Merriam-Webster [7] estende essa definição como: "um ramo da metafísica preocupado com a natureza e as relações do ser".

Importado da Filosofia, esse conceito começa a ser trabalhado muito antes da época dos computadores. Aristóteles já estudava Ontologia em suas Categorias [1]. No entanto, apenas em 1606, com o livro *Ogdoas Scholastica*, de Jacob Lorhard, foi que a palavra em si realmente surgiu. Esse termo ficou popular em 1729 com o livro *Philosophia Prima sive Ontologia*, de Christian Wolff, com a definição "*Ontology or First Philosophy is the science of Being in general or as Being*".[8]

Para a Ciência da Computação, a definição é um pouco diferente, embora possua muita semelhança com o conceito já explícito. É estudado na área de Inteligência Artificial, que está preocupada com a automação do comportamento inteligente. Guarino [6] definiu ontologia como "um artefato de engenharia, constituído por um vocabulário específico usado para descrever uma certa realidade, mais uma série de pressupostos explícitos acerca do significado que se atribui a esse vocabulário". Logo, uma ontologia seria uma reunião de sentenças lógicas que exibem alguma informação sobre o mundo.

Fazendo um paralelo entre ambas as disciplinas, pode-se observar que, enquanto a primeira existe um estudo sistemático da existência, na segunda existe um foco maior em o que pode ser representado.

Na prática, elas funcionam como um sistema "*tell and ask*". Algumas coisas são contadas para os agentes inteligentes (uma entidade autônoma com comportamento que simula inteligência), e então, perguntas podem ser feitas para eles, embora eles não precisem saber todas as respostas.

Surgem de um contexto onde os cientistas modelam e representar o mundo para as máquinas, e isso ocorre desde a origem dos computadores. Como cada pessoa possui uma visão de mundo, cada modelagem será diferente de algum jeito, por isso, a construção de ontologias é um tópico que merece estudo. Atualmente, a área de ontologias é bastante pesquisada por sua aplicação com Web Semântica.

As ontologias denotam uma "especificação explícita de uma conceitualização"[5], e, uma vez construídas, permitem comunicação, compartilhamento e reúso de conhecimentos. Elas interessam várias áreas do conhecimento, tais como a ciência da computação, filosofia, engenharia de dados, data science, lógica e linguística.

1.1 Conceitualização

No entanto, existe um certo debate sobre o que é uma conceitualização. Gruber [5], que definiu o que seria uma ontologia, propôs que uma conceitualização “é uma visão abstrata e simplificada do mundo que se quer representar para algum propósito”. Essa definição parece boa, mas deixa algumas pontas soltas, por exemplo, em o que seria uma “visão”.

Para Guarino e Giaretta [4], uma conceitualização pode ser entendida como “uma estrutura semântica intensional que codifica as regras implícitas que determinam a estrutura de uma porção da realidade”. Essa definição é um pouco mais concreta, e já é possível pensar sobre ela computacionalmente.

Pode-se inferir que uma conceitualização é uma modelagem de parte de algum domínio do conhecimento, que nada mais seria do que uma área de estudo, como a Geografia, Música, entre outros. Tal modelagem é feita a partir de alguma linguagem formal de representação (em geral, Lógicas) e deve levar em conta uma generalidade que se aplica ao domínio escolhido.

Vale lembrar que embora amplamente utilizadas, as linguagens naturais não são consideradas linguagens formais. Isso tem uma explicação simples. Uma ontologia em Português e outra em Espanhol terão vocábulos diferentes, mas podem ter a mesma conceitualização.

Debruyne [2] fez um estudo mais sistemático sobre o que seria uma conceitualização, baseado na pesquisa de Guarino [6]. Para facilitar o entendimento das definições, é necessário entender dois conceitos vindos da filosofia:

- Contexto Extensional: Se refere à extensão de um termo, que é o conjunto de objetos que ele denota. Um contexto é extensional se alguma expressão *e* aparece e pode ser substituída por uma expressão com a mesma extensão e manter o seu valor verdadeiro. No exemplo a seguir, temos que *instrumento musical* é uma extensão de *Violão* e de *Piano*:
 - Violão é um instrumento musical.
 - Piano é um instrumento musical.
- Contexto Intensional: é o oposto do contexto acima. Ocorre quando um termo não pode ser substituído sem que a verdade sobre a sentença seja ameaçada. Por exemplo, embora saibamos que Clark Kent é o Super-Homem e também é um jornalista, a sentença a seguir é falsa, pois o Super-Homem não é um jornalista:
 - Lois Lane acredita que o Super-Homem irá investigar uma notícia com ela.

Feita este esclarecimento, seguem as definições de Guarino.

1. Noção extensional de Conceitualização: Um par $\langle D, R \rangle$, onde D é um domínio e R são as relações relevantes que existem nesse domínio.
2. Espaço domínio: Um par $\langle D, W \rangle$, com D representando um domínio e W sendo os mundos possíveis dentro dele, ou seja, os conjuntos máximos de estados das coisas desse domínio.
3. Relação conceitual: Dado um espaço domínio $\langle D, W \rangle$, uma relação conceitual ρ^n de aridade n em $\langle D, W \rangle$ é definida como uma função total $\rho^n : W \rightarrow 2^{D^n}$ de W para o conjunto de todas as relações n -árias em D .

4. Noção intensional de Conceitualização: Uma conceitualização para D é definida como uma tripla ordenada $C = \langle D, W, R \rangle$, onde R é um conjunto de relações conceituais do espaço domínio $\langle D, R \rangle$

Definindo uma linguagem lógica L com um vocabulário V , teremos:

5. Interpretação extensional de uma Linguagem: Um modelo para L é definido como uma estrutura $\langle S, I \rangle$, onde $S = \langle D, R \rangle$ é uma estrutura de mundo e $I : V \rightarrow D \cup R$ é uma função de interpretação atribuindo elementos de D para símbolos constantes de V e elementos de R para símbolos predicados de V .
6. Interpretação intensional de uma Linguagem: Por meio de uma estrutura $\langle C, I \rangle$, onde C é uma conceitualização como definido no item 3, e I é uma função de interpretação como definido no item acima, teremos que uma interpretação intensional é um comprometimento ontológico K para L . Se K é um comprometimento ontológico para L , dizemos que L se compromete com C por meio de K .
7. O conjunto de modelos pretendidos de uma linguagem de acordo com um comprometimento: O conjunto $I_K(L)$ de todos os modelos pretendidos de uma linguagem L que são compatíveis com um comprometimento $K = \langle C, I \rangle$ são todos os modelos de L que são compatíveis com K . O comprometimento garante consistência.

1.2 Definições gerais (talvez diluir essa parte no capítulo)

Smith [10] definiu alguns conceitos que nos auxiliam a compreender o que é uma ontologia.

- Representação: seria uma ideia, uma imagem, um registro, ou uma descrição que se refere (ou é sobre), ou pretende referir a alguma entidade externa à representação.
- Representação composta: é uma representação constituída de sub-representações como suas partes, de uma maneira em que os parágrafos são feitos de sentenças e as sentenças de palavras. A menor sub-representação se chama unidade representacional, que é um ícone, um símbolo, um nome, entre outros.
- Representação cognitiva: é uma representação cujas unidades representacionais são ideias, pensamentos ou crenças na mente de algum elemento cognitivo.
- Artefato representacional: é uma representação fixa em algum meio de maneira que pode servir para fazer representações cognitivas que existem nas mentes de sujeitos separados publicamente acessíveis. Exemplos: um texto, um diagrama, um mapa, uma lista, entre outros.
- Porção de Realidade: Termo usado que compreende universais únicos e particulares e suas combinações mais ou menos complexas.
- Domínio: É uma porção da realidade que forma o assunto de estudo de uma ciência ou tecnologia. Ex: Radiologia.
- Taxonomia: É um artefato representacional em um grafo na forma de árvore com os nós representando universais ou classes e as arestas sendo relações *is-a* ou *part-of*, que serão tratadas mais adiante.

- **Ontologia:** É um artefato representacional que compreende uma taxonomia própria, cujas unidades representacionais são usadas para designar uma combinação de universais, classes definidas e relações entre eles. Uma ontologia baseada em realismo é feita apenas de termos universais, para que assim ela corresponda a parte de uma teoria científica capturada pelos seus termos e interrelações.
- **Terminologia:** É um artefato representacional que possui unidades representacionais que são os termos gerais de uma linguagem para se referir a entidades em um certo domínio.
- **Inventário:** É um artefato representacional construído de termos únicos tal como nomes próprios ou identificadores alfanuméricos.

1.3 Construindo uma ontologia

Para construir uma ontologia, é necessário escolher um domínio e o nível de generalidade que é necessário que ela atinja. Também deve-se ter em mente quem vai usá-la. Para que ela atinja o máximo de utilidade, é necessário que as perguntas que se deseja que ela responda sejam feitas antes de sua construção.

Geralmente, grandes ontologias são projetadas por equipes interdisciplinares, para que haja ela seja o mais correta e abrangente quanto possível.

Quando se confecciona uma ontologia, é necessário que sejam feitas algumas decisões de projeto.

Gruber [5] fez uma proposta de critérios de design para ontologias com o objetivo de tornar o compartilhamento de conhecimento e interoperabilidade com programas baseados em conhecimento mais fácil. Os critérios são os seguintes:

1. **Clareza:** Uma ontologia deve ter uma linguagem clara e efetiva na definição de seus termos, que devem ser objetivos. Embora a definição de um conceito possa vir de situações sociais ou requisitos computacionais, ela deve ser independente destes contextos. O formalismo, no caso, o uso de uma linguagem lógica, é um meio para este fim, ou seja, quando for possível fazer uma definição usando axiomas lógicos, isso deve ser feito. Onde possível, uma definição completa (com condições necessárias e suficientes) é preferível a uma definição parcial (com condições necessárias ou suficientes). Todas devem ser documentadas usando linguagem natural.
2. **Coerência:** Uma ontologia deve permitir inferências consistentes com suas definições. No mínimo, os axiomas usados nas definições devem ser logicamente consistentes. A coerência também deve ser aplicada aos conceitos informais, definidos na linguagem natural da documentação e nos exemplos. Se uma sentença que pode ser inferida contradiz uma definição ou exemplo dado informalmente, a ontologia é incoerente.
3. **Estendibilidade:** Uma ontologia deve ser projetada para ser capaz de antecipar o uso de conhecimento compartilhado. Deve oferecer uma fundação conceitual de modo que o novo conhecimento possa ser apoiado nela e que a ontologia possa ser estendida e especializada, ou seja, novos termos podem ser definidos usando o vocabulário existente, de modo que a revisão do conhecimento anterior possa ser evitada ao máximo.
4. **Viés mínimo de codificação:** A conceitualização deve ser especificada num nível de conhecimento que não dependa de uma codificação que utiliza um nível de símbolos

particulares. Um viés de codificação ocorre quando as escolhas de representação são feitas por pura conveniência de notação ou de implementação. Isso deve ser minimizado, já que agentes de conhecimento podem ser implementados em diferentes sistemas e estilos de representação.

5. Comprometimento ontológico mínimo: Isso faz com que ela suporte as atividades de compartilhamento de conhecimento desejadas. Deve ser feito o mínimo possível de reivindicações do mundo modelado permitindo que as partes comprometidas com a ontologia sejam livres para especializar e instanciar a ontologia o quanto quiserem. Já que isso é baseado no uso consistente do vocabulário, ele pode ser minimizado usando uma teoria fraca (genérica) e definindo apenas os conceitos necessários para a comunicação do conhecimento consistente com esta teoria.

É possível notar que todos esses critérios não poderão ser atendidos ao mesmo tempo, portanto, alguns *trade-offs* deverão ser feitos. Podemos ter um conflito, por exemplo, entre os critérios 1 e 3, já que o máximo de clareza implica que as definições terão a sua interpretação restringida.

1.3.1 Componentes de uma ontologia

Computacionalmente falando, as ontologias possuem cinco partes. Para ilustrá-las melhor, vamos fazer uma ontologia sobre Música.

1. Classes: Descrevem os conceitos de um certo domínio do discurso. São o foco das ontologias. São uma coleção de todos os particulares aos quais é possível aplicar um termo geral. Por exemplo: a classe Cantor.
2. Propriedades: São atributos que descrevem características do conceito a que uma classe se refere. Em nosso exemplo, a classe Cantor possui as propriedades Nome, RitmoPredominante e Idade.
3. Relações: Ontologias são constituídas de relações hierárquicas. Por exemplo Cantor → Pessoa. A hierarquia de classes representa uma relação “is-a”[3]. Tais relações são transitivas. Existem também as relações “part-of”, que já estão dentro da Teoria dos Conjuntos. Uma classe é parte da outra se a compõe.
4. Restrições: São os tipos das propriedades, por exemplo: enquanto para Nome e RitmoPredominante uma *string* seja suficiente, para a Idade, um inteiro já está de bom tamanho.
5. Instância: é um indivíduo (por exemplo, MariahCarey), de um conjunto universal (cantores, músicos, pessoas). Vale lembrar que uma instância não é uma subclasse.

Como já vimos acima, uma classe pode ter nomes diferentes, pois não é ele que define uma classe. Sinônimos e palavras em línguas distintas não representam classes diferentes.

Várias classes subordinadas a uma superclasse são consideradas irmãs. Elas devem ter o mesmo nível de generalidade. Por exemplo, seja uma classe Musica. Suas subclasses podem ser Cancao e MusicaAmericana, e não Cancao, MusicaAmericana e MusicaBrasileira (a última é sub-classe da penúltima).

Os conceitos definidos até aqui mostram um jeito de caracterizar ontologias, que é pelo número de termos que ela possui. Uma outra métrica é o número de axiomas lógicos que ela possui.

Existem três processos para definir as classes e a hierarquia, a seguir:

- *Top-down*: vai das classes mais genéricas para as mais específicas.
- *Bottom-up*: vai das mais específicas indo para as mais genéricas, agrupando as específicas já criadas.
- *Vai-e-Vem*: define conceitos simples para generalizá-los e especificá-los. Seria um misto dos dois primeiros itens.

1.3.2 Montando a ontologia

Montar uma ontologia é um processo que segue os seguintes passos:

- Definir as classes da ontologia.
- Colocá-las em uma hierarquia taxonômica.
- Determinar propriedades e restrições para elas.
- Criar uma base de conhecimento para essas classes e propriedades, ou seja, preenchendo a ontologia com as instâncias.
- Colocar os valores das propriedades para as instâncias.

Embora pareça ser direto, esse processo é iterativo, como afirmam Noy e McGuinness [9]. Uma vez feito, deve ser repetido para que haja uma adequação das classes com as instâncias colocadas, pois, por exemplo, se uma classe acabar com apenas uma sub-classe, a modelagem pode ter um problema, ou a ontologia não está completa. E ainda, se uma classe possui mais de uma dúzia de sub-classes, novas categorias (classes ou subclasses) podem ser necessárias.

Introduzir uma classe é legal quando se tem muitas propriedades específicas em muitas instâncias, o que as faz serem relacionadas de algum jeito. Se os diferentes valores nas propriedades de uma classe tornam-se muito restritos, então, uma nova classe deve ser criada para que a distinção aconteça.

O processo é iterativo, e teoricamente, ele nunca acaba. Na prática, o processo é encerrado quando a ontologia fornece respostas suficientemente boas para a maior parte das consultas realizadas, ou seja, tal critério é subjetivo.

A criação de uma ontologia é feita por uma equipe interdisciplinar, geralmente composta por *experts* da área que se deseja cobrir e técnicos para a confecção de ontologias (vindos da área de Computação). Isso não limita a equipe de possuir profissionais de mais áreas.

Em relação às classes, algumas observações podem ser feitas. A primeira é em relação à herança múltipla. Ela acontece quando uma classe é sub-classe de várias outras classes, por exemplo, a classe *PopBrasileiro*, pode pertencer à classe *Pop* e à classe *MusicaBrasileira*. Isso é aceitável, pois no mundo real, acontece várias vezes e de diversas maneiras.

Outro aspecto interessante é o de classe abstrata. Uma classe desse tipo não pode ter instâncias. Em nossa ontologia, podemos ter locais (no caso, países onde as músicas são criadas). A

classe Pais pode ter varias sub-classes, tais como Argentina, Brasil, Portugal, entre outras. Note que essas classes são abstratas, pois no mundo real, não existe mais de um país que possa pertencer ao conceito Brasil.

As ontologias possuem suporte a classes disjuntas. Classes disjuntas são aquelas que não possuem uma intersecção. Enquanto em nossa ontologia, Cantor e Compositor não sejam disjuntas. No entanto, Cancao e HinoNacional são classes disjuntas.

Uma ontologia não precisa ter toda a informação existente sobre o domínio. Não é necessário especializar ou generalizar mais do que seja necessário para a aplicação. Além disso, as classes não precisam ter todas as propriedades possíveis e nem carregarem todas as distinções que estão no mundo. Isso significa que a ontologia deve ser o modelo mais simples para o problema que se deseja resolver.

Algumas relações podem ter uma inversa, assim como ocorre com funções matemáticas. Uma relação que possui uma inversa pode ser:

- Cantor canta Musica
- Musica cantadaPor Cantor

Para os nomes das classes, não há uma convenção específica, só há um consenso em manter um padrão de nomenclatura, por exemplo, usando snake_case, camelCase ou usando espaços, e também prestando atenção se no nome de uma classe possui o nome de sua superclasse ou não.

1.4 Usabilidade de uma ontologia

É possível achar os usuários de uma ontologia, mas nem sempre o seu autor. O razoável é assumir que elas foram evoluindo com o passar do tempo.

Ontologias feitas sobre áreas de estudo (por exemplo, a Aviação) serão muito mais úteis do que aquelas feitas sobre acontecimentos (queda do voo TAM 3054), cuja existência não faz sentido, já que o problema relacionado ao acontecimento pode ser resolvido usando a área de estudo relacionada.

1.5 Problemas relacionados

Existem alguns problemas relacionados a ontologias. Os mais comuns são os problemas de modelagem e construção. Um problema muito comum é a existência, na linguagem natural, de Homônimos e Sinônimos. Deve existir um cuidado especial com eles.

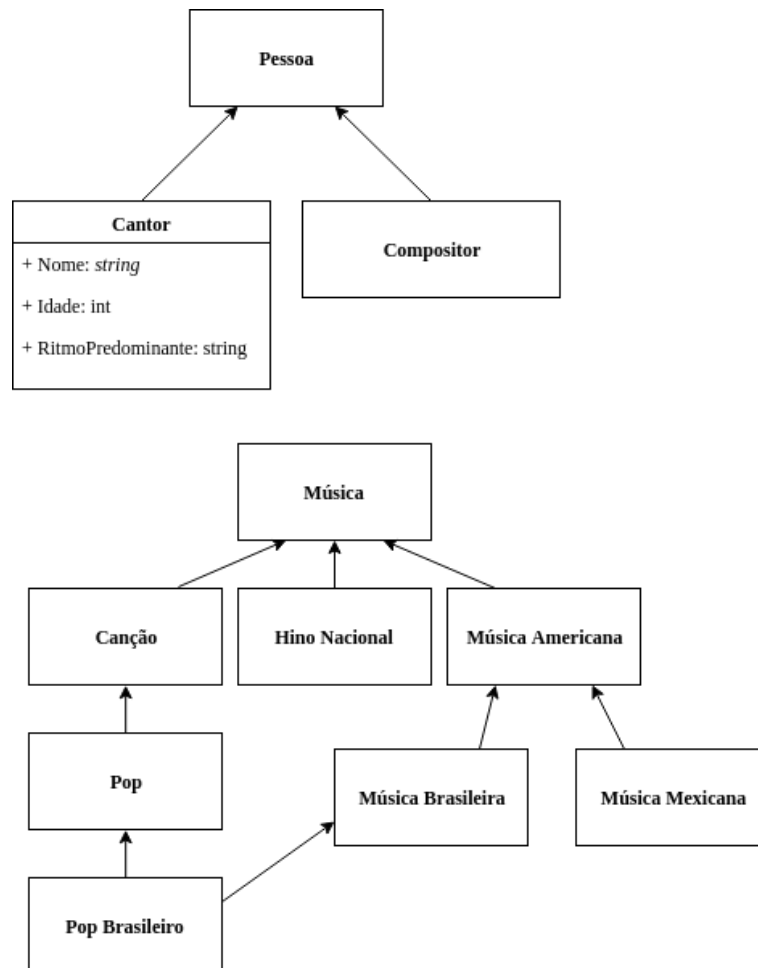
Em relação à modelagem, o problema de não utilizar uma equipe interdisciplinar especializada pode levar a uma cobertura incompatível de conceitos, ou seja, as classes ficam muito distantes da realidade. Além disso, usar fontes não confiáveis na ontologia pode fazer com que os problemas que elas estavam sendo feitas para resolver, não sejam resolvidos corretamente.

Tais problemas tornam-se muito maiores quando duas ontologias são integradas. A possível integração entre ontologias é um dos motivos para elas existirem, afinal, isso pode acelerar o seu desenvolvimento. Em nosso exemplo, se já existir uma ontologia sobre Música Brasileira, poderemos absorvê-la, mas o cuidado terá de ser redobrado em relação às questões acima.

Em relação à implementação, os pontos que surgem são em geral a respeito da linguagem utilizada, mas isso será tratado no capítulo.

Além dos já citados, pode acontecer de chegar um novo conhecimento e a ontologia ficar inconsistente. Neste caso, terão de ser usadas algumas operações de Revisão de Crenças, foco do capítulo X.

A ontologia construída nesse capítulo fica assim:



Capítulo 2

Lógicas de Descrição

Capítulo 3

Revisão de Crenças

Capítulo 4

Ferramentas computacionais

Capítulo 5

Implementação do *Plug-In*

Capítulo 6

Análise de desempenhos

Agradecimentos

Referências Bibliográficas

- [1] I. DAHLBERG, *Teoria do conceito*, Ciência da informação, 7 (1978).
- [2] C. DEBRUYNE, *The relation between a framework for collaborative ontology engineering and nicola guarino's terminology and ideas in "formal ontology and information systems"*, in VaSCo@ WebSci, 2013, pp. 34–44.
- [3] P. C. FRANÇA, *Conceitos, classes e/ou universais: com o que é que se constrói uma ontologia?*, Linguamática, 1 (2009), pp. 105–121.
- [4] P. GIARETTA AND N. GUARINO, *Ontologies and knowledge bases towards a terminological clarification*, Towards very large knowledge bases: knowledge building & knowledge sharing, 25 (1995), pp. 307–317.
- [5] T. R. GRUBER, *Toward principles for the design of ontologies used for knowledge sharing*, International journal of human-computer studies, 43 (1995), pp. 907–928.
- [6] N. GUARINO, *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*, vol. 46, IOS press, 1998.
- [7] MERRIAM-WEBSTER ONLINE, *Merriam-Webster Online Dictionary*.
- [8] M. NICKLES, A. PEASE, A. C. SCHALLEY, AND D. ZAEFFERER, *Ontologies across disciplines*, Ontolinguistics-How Ontological Status Shapes the Linguistic Coding of Concepts, (2007), pp. 23–67.
- [9] N. F. NOY, D. L. MCGUINNESS, ET AL., *Ontology development 101: A guide to creating your first ontology*, 2001.
- [10] B. SMITH, W. KUSNIERCZYK, D. SCHOBER, AND W. CEUSTERS, *Towards a reference terminology for ontology research and development in the biomedical domain*, in KR-MED, vol. 2006, 2006, pp. 57–66.