

Prefácio por Barry Smith

**MAURICIO BARCELLOS ALMEIDA**

# ONTOLOGIA EM CIÊNCIA DA INFORMAÇÃO

TEORIA E MÉTODO

**COLEÇÃO** Representação do Conhecimento  
em Ciência da Informação - **VOLUME 01**





Mauricio Barcellos Almeida

# ONTOLOGIA EM CIÊNCIA DA INFORMAÇÃO: Teoria e Método

Coleção Representação do Conhecimento  
em Ciência da Informação  
Volume 1

Editora CRV  
Curitiba – Brasil  
2020

Copyright © da Editora CRV Ltda.  
**Editor-chefe:** Railson Moura  
**Diagramação e Capa:** Designers da Editora CRV  
**Imagem da Capa:** pikisuperstar.com/freepik.com  
**Revisão:** Analista de Escrita e Artes Editora CRV

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
CATALOGAÇÃO NA FONTE

Bibliotecária responsável: Luzenira Alves dos Santos CRB9/1506

---

A1444

Almeida, Mauricio Barcellos.

Ontologia em Ciência da Informação: teoria e método / Mauricio Barcellos Almeida –  
Curitiba : CRV, 2020.  
374 p. (Coleção Representação do Conhecimento em Ciência da Informação – volume 1).

Bibliografia

ISBN Coleção digital 978-65-5578-688-0

ISBN Coleção físico 978-65-5578-689-7

ISBN Volume digital 978-65-5578-678-1

ISBN Volume físico 978-65-5578-679-8

DOI 10.24824/978655578679.8

1. Ciência da Informação 2. Representação do Conhecimento 3. Ontologia 4. Tecnologia da  
Informação I. Título II. Série.

CDU 004.1

CDD 004.6

---

Índice para catálogo sistemático  
1. Tecnologia da informação 004.4

ESTA OBRA TAMBÉM ENCONTRA-SE DISPONÍVEL  
EM FORMATO DIGITAL.  
CONHEÇA E BAIXE NOSSO APLICATIVO!



2020

Foi feito o depósito legal conf. Lei 10.994 de 14/12/2004

Proibida a reprodução parcial ou total desta obra sem autorização da Editora CRV

Todos os direitos desta edição reservados pela: Editora CRV

Tel.: (41) 3039-6418 – E-mail: [sac@editoracrv.com.br](mailto:sac@editoracrv.com.br)

Conheça os nossos lançamentos: [www.editoracrv.com.br](http://www.editoracrv.com.br)

## **Conselho Editorial: Comitê Científico:**

Aldira Guimarães Duarte Domínguez (UNB)	Ana Lucia de Alcântara Oshiro (UAM)
Andréia da Silva Quintanilha Sousa (UNIR/UFRN)	Arnaldo Oliveira Souza Júnior (UFPI)
Anselmo Alencar Colares (UFOPA)	Fernando do Nascimento Gonçalves (UERJ)
Antônio Pereira Gaio Júnior (UFRRJ)	Flávia Affonso Mayer (UFMG)
Carlos Alberto Vilar Estêvão (UMINHO – PT)	Flávio Lins Rodrigues (UFJF)
Carlos Federico Dominguez Avila (Unieuro)	Helena Santiago Vígata (UnB)
Carmen Tereza Velanga (UNIR)	Júlio César Machado Pinto (UNA)
Celso Conti (UFSCar)	Kelly Cristina de Souza Prudencio (UFPR)
Cesar Gerónimo Tello (Univer .Nacional	Liv Rebecca Sovik (UFRJ)
Três de Febrero – Argentina)	Lívia Fernanda Nery da Silva (UFPI)
Eduardo Fernandes Barbosa (UFMG)	Márcio Souza Gonçalves (UERJ)
Elíone Maria Nogueira Diogenes (UFAL)	Marcos Rizolli (MACKENZIE)
Elizeu Clementino de Souza (UNEB)	Maria Schirley Luft (UFRR)
Élsio José Corá (UFS)	Mohammed ElHajji (UFRJ)
Fernando Antônio Gonçalves Alcoforado (IPB)	Otacílio Amaral Filho (UFPA)
Francisco Carlos Duarte (PUC-PR)	Ricardo Ferreira Freitas (UERJ)
Gloria Fariñas León (Universidade	Sergio Augusto Soares Mattos (UFRB)
de La Havana – Cuba)	Suyanne Tolentino de Souza (PUC-PR)
Guillermo Arias Beatón (Universidade	
de La Havana – Cuba)	
Helmuth Krüger (UCP)	
Jailson Alves dos Santos (UFRJ)	
João Adalberto Campato Junior (UNESP)	
Josania Portela (UFPI)	
Leonel Severo Rocha (UNISINOS)	
Lidia de Oliveira Xavier (UNIEURO)	
Lourdes Helena da Silva (UFV)	
Marcelo Paixão (UFRJ e UTexas – US)	
Maria Cristina dos Santos Bezerra (UFSCar)	
Maria de Lourdes Pinto de Almeida (UNOESC)	
Maria Lília Imbiriba Sousa Colares (UFOPA)	
Paulo Romualdo Hernandes (UNIFAL-MG)	
Renato Francisco dos Santos Paula (UFG)	
Rodrigo Pratte-Santos (UFES)	
Sérgio Nunes de Jesus (IFRO)	
Simone Rodrigues Pinto (UNB)	
Solange Helena Ximenes-Rocha (UFOPA)	
Sydione Santos (UEPG)	
Tadeu Oliver Gonçalves (UFPA)	
Tania Suely Azevedo Brasileiro (UFOPA)	

Este livro passou por avaliação e aprovação às cegas de dois ou mais pareceristas *ad hoc*.



Em favor da honestidade intelectual:

*[...] it would be foolish to suppose that one's person solution will necessarily convince everyone else. I have presented my own proposed solutions [...] not so much in order to persuade readers that these solutions are correct, as to try to motivate them to pursue their own solutions.*

[...] seria tolice supor que a solução de uma pessoa necessariamente convença todos as outras. Eu apresento aqui minhas próprias propostas de soluções [...] não tanto para persuadir leitores que essas soluções estão corretas, mas como forma de tentar motivá-los a buscar suas próprias soluções.

*Prof. Edward J. Lowe (1950-2014).*





# SUMÁRIO

FOREWORD .....	11
<i>Barry Smith</i>	
APRESENTAÇÃO .....	15
INTRODUÇÃO .....	21
Por que ontologias? .....	21
Considerações iniciais .....	21
Ontologia em Ciência da Informação .....	23
Ontologia e tecnologia da informação .....	26
Uma breve nota: ontologia e Aristóteles .....	33
Começando a jornada .....	34
1. NOÇÕES PRELIMINARES .....	37
1.1 Ontologia: da antiguidade a contemporaneidade .....	37
1.2 Representações, artefatos e constructos .....	44
1.3 Uso corrente da ontologia .....	55
1.4 Sumário e importância .....	62
2. NOÇÕES DE ONTOLOGIA METAFÍSICA .....	67
2.1 Sistemas de categorias clássicos .....	67
2.2 Ontologia e questões metafísicas .....	81
2.3 Teorias da ontologia .....	92
2.4 Sumário e importância .....	98
3. REPRESENTAÇÃO EM ONTOLOGIA .....	103
3.1 Noções básicas de representação .....	104
3.2 Estados mentais, representação e intencionalidade .....	107
3.3 A natureza representacional da mente .....	109
3.4 A intencionalidade em Brentano e além .....	111
3.5 O computador e a representação .....	124
3.6 Sumário e importância .....	132
4. LÓGICA, SEMÂNTICA E ONTOLOGIAS .....	139
4.1 Lógica e ontologias .....	139
4.2 Semântica e ontologias .....	155
4.3 Sumário e importância .....	165

5. CATEGORIZAÇÃO EM ONTOLOGIAS .....	171
5.1 Categorização, classificação e ontologia .....	172
5.2 Por que categorias? .....	176
5.3 Impacto dos princípios metafísicos na classificação .....	179
5.4 Teorias seminais em Ciência da Informação .....	189
5.5 Sumário e importância .....	195
6. A MOBÍLIA DO MUNDO .....	201
6.1 Ontologia formal de alto nível .....	202
6.2 Os quatro eixos de entidades .....	205
6.3 A grande divisão .....	214
6.4 Doutrinas complementares .....	230
6.5 Sumário e importância .....	232
7. RELACIONAMENTOS ONTOLÓGICOS .....	237
7.1 O que são relacionamentos? .....	238
7.2 Relacionamentos linguísticos e relacionamentos lógico-formais .....	240
7.3 Relacionamento ontológicos .....	259
7.4 Sumário e importância .....	273
8. REPRESENTANDO O SOCIAL .....	279
8.1 Revisitando intencionalidade e atos da fala .....	280
8.2 Ontologia do social aplicada às instituições .....	295
8.3 Como representar a Ontologia do Social? .....	305
8.4 Sumário e importância .....	312
9. O PROJETO DE ONTOLOGIAS .....	317
9.1 Princípios e boas práticas de projeto .....	317
9.2 Métodos para desenvolvimento .....	328
9.3 Estudos de caso: setor elétrico .....	331
9.4 Estudo de caso: Biomedicina .....	341
9.5 Sumário e importância .....	353
10. CONSIDERAÇÕES FINAIS .....	359
ÍNDICE REMISSIVO .....	363

# FOREWORD

*Dr. Barry Smith*

---

## **Introduction**

By the time this book appears, the first international standard in ontology will have been published. This is the multipart standard ISO/IEC 21838, approved jointly by the International Standards Organization and the International Electrotechnical Commission, and consisting initially of two parts, on “Top-Level Ontology” and “Basic Formal Ontology” (BFO), respectively. I mention this (in my view) important development in the history of ontology since it was through working with Mauricio Almeida on top-level ontology in general and on BFO in particular that I first came in contact with Mauricio’s remarkable series of contributions to the field of Applied Ontology. Mauricio is himself one of the authors of BFO, and has played an important role in establishing BFO as the top-level ontology of choice in a wide range of application areas.

## **Mauricio Almeida’s Contributions to General Ontology**

Already in 2003 published an influential paper entitled “An overview of ontologies: research on definitions, types, applications, methods of assessment and construction”, which has been cited some 250 times. He has continued ever since to devote considerable energies to writing the sorts of comparative studies of ontology development methodology, of ontology evaluation, and of the role of ontologies in knowledge and information management, which are essential if ontology is to become established as a scientific discipline rather than as the hodge-podge collection of ad hoc proposals which it has too often seemed to be in the eyes of potential users of ontologies for application purposes.

An example of methodological work of this sort is his 2011 paper, with Renato Souza and Fred Fonseca on “Semantics in the Semantic Web: a critical evaluation”, published in the journal *Knowledge Organization*. As this paper makes clear, the term “semantics” has been widely used, not least in the context of the so-called “Semantic Web”, but unfortunately it has been used with a variety of different meanings, deriving from logic and linguistics. Some of these relate to the genuine advances in first-order model theory that underlie OWL and other Description Logics. Some relate to general theories

of meaning proposed by philosophers and linguists. And some uses of the term seem to be associated with no clear meaning at all. Almeida, Souza and Fonseca accordingly set out to provide a coherent critical account of the kind (or better: kinds) of semantics that we find in the Semantic Web in particular and in contemporary information technology in general.

## **Mauricio Almeida's Contributions to Applied Ontology**

It is however in his contributions to *Applied* Ontology that Almeida's has been most conspicuous. Here his work has encompassed areas as wide ranging as plant ontology, information security ontology, blood ontology (including a series of specific studies on leukemia and hematology), muscle development and body dysmorphia, electronic health records, and Building Information Modeling (BIM).

Areas in which he and I have most closely interacted include work on the body fluids ontology (BFLO), on obstetrics and neonatology, and on a broad range of issues connected with documents, organizations, and the legal realm. Mauricio has played a vital role above all in the creation of d-Acts, the Document Acts Ontology, and in its application especially to legal and medical (and legal-medical) documents. It is documents which allow the sorts of commitments entered into, for example through acts of promising. It is documents which, in many cases, allow organizations to be formed and sustained over time. With his most recent paper, "Toward a Document-Centered Ontological Theory for Information Architecture in Corporations", Mauricio has extended the d-Act approach into the industrial realm, thereby contributing to the work of the Industrial Ontologies Foundry (IOF), which attempts to apply BFO-based ontologies in the world of digital manufacturing.

One ontological problem relating to organizations arises where two organizations have identically the same members. The two organizations are distinguished, for example, by the fact that the same people are performing different speech acts and different document acts in their guise as members of the two different organizations. But how are these differences to account ontologically for the non-identity of the organizations themselves? As Mauricio has pointed out, it is the normative dimension that make an organization different from a swarm of bees. Members of an organization have commitments, responsibilities, stand in authority to other members. One solution to the problem therefore, consistent with the new first-order logic axiomatization of BFO, is to view organizations as object-aggregates, but to allow that object-aggregates of two sorts, which we might call extensional and intensional, where only the former behave in such a way that aggregates with the

same members are identical. Intensional object aggregates are brought into existence by articles of incorporation. They may gain and lose members and yet remain identically one and the same.

## **Ontologies in Information Science**

We remarked above that to become established as a scientific discipline ontology requires work not only on general principles and on the nature of ontologies in general, but also work on showing how these general principles are applied in specific domains. I am delighted to say that this work contributes so successfully to achieving both of these ends, and I believe that it will make an important contribution to the future of ontology, both in Brazil and in the wider world.

*New York, USA, June, 2020.*

---

## REFERENCES

AI, J.; ALMEIDA, M. B.; ANDRADE, A. Q.; RUTTENBERG, A.; WONG, D. T. W.; SMITH, B. *Towards a body fluids ontology: A unified application ontology for basic and translational science*. Proceedings of the International Conference of Biomedical Ontologies. US, 2011.

ALMEIDA, M. B.; FELIPE, E. R.; BARCELOS, R. *Toward a Document-Centered Ontological Theory for Information Architecture in Corporations*. *Journal of the Association for Information Science and Technology*, 2020, DOI: 10.1002/asi.24337.

ALMEIDA, M. B.; SOUZA, R. R.; FONSECA, F. F. *Semantics in the Semantic Web: a critical evaluation*, *Knowledge Organization*, v. 38, n. 3, p. 187-203, 2011.

ALMEIDA, M.; PROIETTI, A. B.; AI, J.; SMITH, B. *The Blood Ontology: an ontology in the domain of hematology*. Proceedings of the International Conference of Biomedical Ontologies. USA, 2011.

ALMEIDA, M. B.; SLAUGHTER, L.; BROCHHAUSEN, M. *Towards an Ontology of Document Acts: Introducing a Document Act Template for Healthcare*. OTM 2012 Workshops (LNCS 7567). DOI: 10.1007/978-3-642-33618-8\_57

FARINELLI, F.; ALMEIDA, M. B.; ELKIN, P. L.; SMITH, B. *OntONeo: The Obstetric and Neonatal Ontology*. International Conference on Biomedical Ontology, USA, 2016.

ISO/IEC 21838 1:2020, Information technology – Top-level ontologies (TLO) – Part 1: Requirements. Available in: <https://www.iso.org/standard/71954.html>

ISO/IEC 21838 2:2020. Information technology – Top-level ontologies (TLO) – Part 2: Basic Formal Ontology. Available in: <https://www.iso.org/standard/74572.html>.

RUTTENBERG, A. *et al. Basic Formal Ontology 2020*. Available in: <https://standards.iso.org/iso-iec/21838/-2/ed-1/en/>.

SMITH, B.; ALMEIDA, M. B.; BONA, J. *et al. Basic Formal Ontology 2.0: Specification and user's guide*. National Center for Ontological Research: Buffalo, USA.

# APRESENTAÇÃO



Alice e Humpty Dumpty<sup>1</sup>

Em “Alice no País da Maravilhas”, *Lewis Carroll*<sup>2</sup> narra um diálogo entre Alice e *Humpty Dumpty*: “Quando eu uso uma palavra, ‘*Humpty Dumpty* disse em tom um tanto formal’, isto significa que eu escolho o que ela significa – nem mais nem menos”... ‘e quando Alice contestou:’ “A questão é saber se você pode fazer as palavras significarem tantas coisas diferentes”, ‘*Humpty Dumpty* respondeu:’ “A questão é saber quem manda, quem é o mestre – isso é tudo”.

Mesmo considerando a importância da linguagem, filósofos e cientistas decidiram há muito que as palavras não seriam os únicos “mestres”, como sugere o personagem. Os objetos não se alteram pelo simples uso de palavras diferentes para nomeá-los. Da observação das coisas, antes das palavras que as nomeiam, surgiu o interesse em conhecer o mundo e sua estrutura. Essa é a noção da *ontologia*: conhecer a natureza última das coisas. A presente obra aborda como o mundo é organizado e representado em sistemas de informação através de ontologias, uma atividade que levanta desafios um tanto diferentes daqueles enfrentados por profissionais da informação de 50 ou 60 anos atrás.

A presente obra é uma obra de Ciência da Informação e para a Ciência da Informação. É voltada para profissionais da informação na acepção ampla

---

1 Ilustração de Humpty Dumpty, por John Tenniel, 1871. Disponível em: [http://en.wikipedia.org/wiki/Image:Humpty\\_Dumpty\\_Tenniel.jpg](http://en.wikipedia.org/wiki/Image:Humpty_Dumpty_Tenniel.jpg)

2 Charles Lutwidge Dodgson (1832-1898), escritor inglês. Além de escritor, *Lewis Carroll* (pseudônimo de Dodgson) era matemático e sua obra é repleta de enigmas e paradoxos lógicos.

do termo. É de interesse para estudantes e pesquisadores da Ciência da Informação, para cientistas da computação e analistas de sistemas de informação, para linguistas, bibliotecários, bem como áreas afins, como cientistas de dados, programadores, administradores de bancos de dados, dentre outros. É também de interesse para executivos e gerentes de tecnologia da informação que pretendem fomentar um salto qualitativo nas práticas de informação de seus ambientes profissionais.

A obra conecta diversas teorias para alcançar o uso da ontologia aplicada a Ciência da Informação. Teorias científicas são criadas através do método científico quando a observação e a pesquisa levam a hipóteses que são testadas. Tais teorias não nascem sozinhas, mas de um emaranhado de descobertas e de extratos de outras teorias, modificadas e complementadas à medida que novas evidências são descobertas. Dessa forma esse trabalho foi concebido. A natureza interdisciplinar do assunto não sugere a possibilidade de uma abordagem exaustiva e, assim, é certa a ausência de pesquisas e de autores relevantes. Mesmo que não se admita completa, a obra pode funcionar como referência pois introduz uma variedade de temas para consulta e posterior aprofundamento. Sugestões de leituras complementares, inclusive, são apresentadas ao final de cada capítulo.

Pela extensão do assunto, um livro único seria grande e cansativo. Ao mesmo tempo, não se pretendia mais um livro de tecnologia que apenas apresentasse técnicas sem a fundamentação subjacente. Assim, surgiu a divisão da obra, onde o volume 1 traz os fundamentos. Cada volume pode ser escolhido de acordo com os objetivos de momento do leitor:

- Volume 1 – Teoria e Método
- Volume 2 – Tecnologia e Aplicações
- Volume 3 – Estudos Avançados

Aos cientistas e pesquisadores de áreas científicas das quais extratos foram tomados de empréstimo, os merecidos agradecimentos e as desculpas antecipadas no caso de falta de rigor nas interpretações ou argumentos. Agradecimentos ao Dr. *Barry Smith* da *State University of New York at Buffalo*, ao Conselho Nacional de Pesquisa e Desenvolvimento (CNPQ)<sup>3</sup>, a Companhia Energética de Minas Gerais (CEMIG) e a Fundação Centro de Pesquisa e Desenvolvimento (CPQD) que tornaram possível essa obra. O restante do Volume 1, Teoria e Prática – está organizado com segue:





O capítulo *Introdução* introduz a questão “por que ontologias” e apresenta argumentos em favor do tema e de seus desdobramentos tecnológicos. Descreve barreiras que dificultam a sociedade a gozar dos benefícios da tecnologia da informação e discute brevemente as limitações de outras modalidades de sistemas bem conhecidos. Esses argumentos são discutidos e aprofundados ao longo da obra. Discute, ainda que brevemente, a ontologia como empreendimento científico, o papel da tecnologia no escopo da Ciência da Informação e justifica a base teórica adotada.

O capítulo 1 – *Noções preliminares* – introduz elementos básicos para entender o que são ontologias. As noções e os termos interdisciplinares e, portanto, polissêmicos, são apresentados num primeiro momento de maneira informal e didática. Exemplos cotidianos explicam a representação, os artefatos e os constructos, os quais recebem definições bem fundamentadas em capítulos subsequentes. Esse capítulo também aborda o tema desde suas raízes até os dias de hoje, descrevendo o uso corrente da ontologia na ciência, em geral, assim como em Ciência da Informação.

O capítulo 2 – *Noções de ontologia metafísica* – discorre sobre ontologia e correlatos no seu campo de origem: a Metafísica. Apresenta uma breve visão de três sistemas ontológicos clássicos para depois abordar questões que estão no próprio nascimento dos estudos em ontologia como, por exemplo, a persistência e a identidade. Muitas das questões discutidas envolvem correntes filosóficas originadas de divergências oriundas dos primeiros problemas metafísicos e impactam em práticas da Ciência da Informação e na modelagem de sistemas de informação.

O capítulo 3 – *Representação em ontologia* – aborda a representação e temas afins, assunto seminal na Ciência da Informação. Discorre sobre a capacidade da mente para se dirigir a algo e como computadores são capazes de representar. Nesse aspecto, o capítulo tangencia a Filosofia da Mente e aborda problemas da intencionalidade que afligem pesquisadores em representação há mais de um século. Identifica tendências contemporâneas sobre o tema e, por fim, alcança a ontologia como artefato para representação em computadores.

O capítulo 4 – *Lógica, semântica e ontologias* – discute um assunto que gera equívocos e confunde pesquisadores e estudantes. No âmbito Web Semântica: o que é lógica, o que é semântica, de qual semântica as ontologias são capazes? Além de esclarecer questões desse tipo, discorre-se sobre o envolvimento teórico entre lógica e ontologia, além de apresentar umas poucas noções de lógica simbólica para entendimento de exemplos em linguagem formal e semiformal que aparecem ao longo da obra.

O capítulo 5 – *Categorização em ontologias* – discute a atividade de classificar e categorizar, além de retomar questões terminológicas do primeiro capítulo. Aborda o impacto de teorias de cunho ontológico em atividades de categorização e de classificação, analisando a influência de duas vertentes filosóficas. Descreve extratos de teorias consideradas fundamentais da Ciência da Informação, discutindo alguns de seus aspectos no âmbito das teses filosóficas já apresentadas. Conclui com uma reflexão sobre qual teoria seria a melhor.

O capítulo 6 – *A mobília do mundo* – traz o arcabouço teórico para aplicação da ontologia genérica – a Ontologia Formal Básica – iniciando a criação de um método para desenvolver ontologias como artefatos de representação computacionais. Fundamentado na tradição do realismo, a teoria descreve entidades do mundo em quatro eixos básicos que vão resultar em dois modos para categorizar entidades: endurantismo e perdurantismo. Além disso, introduz noções de doutrinas adicionais que complementam o arcabouço teórico.

O capítulo 7 – *Relacionamentos ontológicos* – introduz os relacionamentos entre entidades comparando-os inicialmente a propriedades. Aborda características de dois tipos de relacionamentos básicos – a subsunção e o parte-todo – na visão Linguística e Lógica. Dando sequência ao método iniciado no Capítulo 6, lista e define primitivas e tipos de relacionamentos ontológicos fundamentais, além de outros constructos relevantes para ontologias como artefatos.

O capítulo 8 – *Ontologia do social* – traz uma mudança de rumo na orientação até então adotada. O capítulo é um extrato de um conjunto de teorias que se convencionou chamar Ontologia do Social, onde se busca entender como um mundo constituído por partículas físicas abriga entidades como ética, estética, política, obrigações, direitos, governos etc. Além de complementos

a temas já mencionados em capítulos anteriores, são introduzidas ontologias computacionais para lidar também com a dimensão social: a Ontologia dos Artefatos de Informação, a Ontologia dos Atos dos Documentos, e a Ontologia Industrial.

O capítulo 9 – *Projeto de ontologias* – encerra o primeiro volume adicionando ao arcabouço conceitual desenvolvido, um conjunto de princípios e boas práticas para o projeto de ontologias. Tal projeto objetiva planejar o artefato, do ponto de vista conceitual e terminológico, antes de implementação tecnológica. Além de descrever passos de uma metodologia para construir ontologias, o capítulo traz ainda dois estudos de caso: uma iniciativa internacional em Biomedicina e um projeto pioneiro de ontologia industrial no setor de energia brasileiro.

\*\*\*



# INTRODUÇÃO

## Por que ontologias?

Questões sobre o que são ontologias e para que usá-las surgiram há mais de 30 anos. Talvez a mais famosa tentativa de resposta vem de *Thomas Gruber* em seu artigo “*What is an Ontology?*”<sup>4</sup> de 1992. *Gruber*, é um ex-professor da *Stanford University* cuja pesquisa em ontologia é responsável pela tecnologia SIRI<sup>5</sup>, um assistente virtual que vem embutido em praticamente todos os dispositivos da *Apple*. Ainda assim, a questão sobre a utilidade das ontologias se mantém e algumas considerações ajudam a entender a importância do tema.

## Considerações iniciais

As pessoas têm opiniões próprias, formadas a partir de uma infinidade de fatores, e por isso discordam sobre todos os tipos de assuntos. Muitas vezes, discordam sobre *como as coisas são*, por exemplo, se um livro é bom, se uma tese é original, se alguém agiu de forma ética. Outras vezes, discordam sobre *o que as coisas são*, por exemplo, se um vírus é um organismo, o que é a virtude, o que é a dor. Discordam até mesmo sobre *que coisas existem*. Nesse último tipo de questão, as divergências são de caráter *ontológico*, mesmo que nem todas possam receber esse rótulo.

Algumas divergências são resolvidas pela comunicação, por exemplo, como saber se existe vinho na geladeira. Certas divergências exigem verificação empírica, como saber a distância entre Rio e São Paulo; algumas exigem uma verificação matemática, como saber qual a área de um triângulo. Existem ocasiões em que não é tão simples fazer a verificação, por exemplo, quando alguém quer saber se existe vida em Marte. Existem ainda, casos que não encontram boa resposta, nos quais a verificação nem mesmo parece possível. Esses casos são chamados *questões ontológicas*. Exemplos são: existe uma alma imortal? Os números existem? O que é moral? O que é o tempo? Casos assim complexos acabam assunto de investigação filosófica e, de fato, o primeiro sentido para o termo “ontologia” é filosófico.

Ontologia no âmbito da Filosofia é uma disciplina, um ramo da Metafísica, a ciência do que existe, dos tipos de coisas e das suas estruturas, sejam elas objetos, propriedades, eventos, relações, dentre outros, em qualquer parte da realidade. O termo remete aos estudos de Aristóteles, mas surgiu apenas

---

4 Disponível em: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

5 Disponível em: <https://en.wikipedia.org/wiki/Siri>

em textos de filósofos alemães no século XVII. Ao longo da história, a disciplina da ontologia vem acumulando debates e fornecendo alternativas aos enigmas da existência.

O termo “ontologia”, entretanto, assumiu outros sentidos além da Filosofia. Um sentido distinto surgiu em campos científicos de natureza aplicada, sobremaneira, naqueles que se prestam a manipular dados, informação e conhecimento. Em Ciência da Informação – em disciplinas como Teoria da Classificação e Organização do Conhecimento – bem como em Ciência da Computação e Sistemas da Informação – em disciplinas como Representação do Conhecimento e Modelagem Conceitual – o termo “ontologia” exhibe usos específicos. Ainda que guarde como premissa a busca por entender o mundo, uma forma de explicar a diferenciação desses usos específicos para o uso filosófico é a noção de conhecimento. Não se pretende discutir aqui como definir conhecimento, dado ou informação – o que resultaria em uma longa discussão – mas apenas fazer uma distinção.

O termo “conhecimento” como utilizado em teorias filosóficas carrega um requisito básico: uma crença ou um estado da mente, para contar como conhecimento, deve estar conectada à verdade. Dito de outra forma, uma crença conta como conhecimento apenas se o seu conteúdo representacional corresponde ao mundo, à como o mundo é de fato. Conteúdo representacional é o que constitui a representação de certa entidade, como representando esta entidade e não outra qualquer. Além da conexão com a verdade, muitas teorias filosóficas exigem ainda um elo causal entre a crença e a verificação da crença como fato. Isso quer dizer que uma pessoa precisa basear sua crença em algum tipo de evidência, ou justificativa, estabelecendo uma correspondência não acidental entre a crença e o mundo.

O sentido do termo “conhecimento” usado em Ciências Aplicadas não é tão rigoroso. Denominações como “Gestão do Conhecimento”, “Representação do Conhecimento”, “Organização do Conhecimento”, dentre outras, não se referem ao conhecimento como um corpo de crenças incontestavelmente verdadeiras. Essa acepção aplicada diz respeito ao conjunto de crenças que as pessoas têm boas razões para aceitar, mas as quais não vão aderir de forma dogmática se receberem boas razões para pensar diferente. É nessa acepção que o conhecimento é guardado, processado ou manipulado em sistemas computacionais, como sugerem as denominações mencionadas.

Surge então o segundo sentido para o termo “ontologia”: um artefato representacional que, incorporado à sistemas, provê à máquinas algum tipo de estrutura de mundo. Exemplos de “estrutura de mundo” necessárias aos sistemas são declarações e fatos triviais para seres humanos, mas necessários para máquinas, como: um artigo é um tipo de documento, a hepatite só ocorre no fígado, flor é um tipo de planta, a neve é branca etc.

Ontologia, originalmente, é um termo que nomeia uma disciplina da Filosofia a qual se ocupa da existência e das formas mais gerais daquilo que existe.

Ontologia, no escopo das Ciências Aplicadas, é um termo que refere a um artefato representacional, o qual, inserido em sistemas, provê uma estrutura de mundo à máquinas permitindo-lhes um tipo básico de raciocínio.

O interesse recai aqui na representação do conhecimento em sistemas automatizados, pelo uso de terminologias e de taxonomias em ambientes computacionais, sistemas especialistas, bancos de dados digitais, dentre outros. A automação da representação do conhecimento apresenta hoje desafios de natureza bem diferente daqueles enfrentados por pesquisadores, bibliotecários ou documentalistas da era pré-informática.

## Ontologia em Ciência da Informação

O termo “ontologia” passou a ser utilizado na Ciência da Informação na década de 1990, mas não sem dificuldades. Costuma-se atribuir como marco do assunto na área o artigo do professor inglês *Brian Vickery*<sup>6</sup>. Naquele momento, a pesquisa em ontologia era vista com certa desconfiança por muitos pesquisadores que afirmavam, por exemplo, que: i) os estudos em ontologia seriam uma impostura, pois diziam respeito à classificação e classificação é algo que não se pode reinventar; ii) o uso do termo “ontologia” em outras áreas para denominar um tipo de estrutura de classificação seria apenas uma questão etimológica; iii) o termo “ontologia” seria resultado da aplicação de conceitos antigos da Ciência da Informação à novas tecnologias. Por outro lado, havia também quem fomentasse a cooperação, identificando a similaridade entre pesquisas em ontologias e em classificação bibliográfica. Com o passar do tempo, as dúvidas iniciais foram retomadas e esclarecidas.

A despeito dessa discussão e mesmo que a adoção de uma teoria possa resultar em debate, a abordagem ontológica figura entre os três modos de investigação que devem ser considerados no escopo da ciência:

- O *ontológico*, quanto se pergunta: qual é a natureza dos objetos que podem ser conhecidos? Qual é a natureza da realidade?
- O *epistemológico*, quando se pergunta: qual é a natureza da relação entre o sujeito que conhece e o objeto que é conhecido?

6 VICKERY, Brian. Ontologies. *Journal of Information Science*, v. 23, n. 4, 1997.

- *Metodológico*, quando se pretende saber: como o sujeito deve proceder para descobrir o conhecimento?

O “sujeito” e o “objeto”, duas primitivas, se correlacionam de forma transcendente: o sujeito é um ente cognoscente (uma pessoa, por exemplo) que apreende, “conhece” um objeto, por meio de sua função cognitiva. Esse tipo de questão, mesmo sendo a mais estudada, é de cunho epistemológico. O que interessa nesse momento é a abordagem ontológica a qual, mesmo que nem sempre identificada, está presente em debates e em teorias da Ciência da Informação.

Um exemplo disso é a Teoria da Classificação, um assunto elementar da área, cujos fundamentos residem originalmente em distinções de caráter ontológico da tradição aristotélica. A questão nos primórdios da teoria era ampla: como é possível a existência de uma cognição universal sobre entidades que são individuais? Na busca por respostas, Aristóteles fez uso de constructos abstratos, os quais chamou de *universais*, além de entidades concretas que denominou *particulares*, associados a um universal. Para explicar como particulares pertencem a um universal, o filósofo grego lançou mão de teorias ontológicas, como identidade, persistência, composição, dentre outras, que estão na base da categorização.

Outro exemplo é a Recuperação da Informação, um tema também fundamental da área. Se Recuperação da Informação, como explica o pesquisador croata *Tefko Saracevic*<sup>7</sup>, foi uma resposta bem sucedida da Ciência da Informação à problemas da sociedade, tal resposta trouxe novas dificuldades. Talvez a mais proeminente seja como definir a própria informação: o termo “informação”, de fato, continua polissêmico e dentre as alternativas para defini-lo não existe uma que satisfaça a todos. Duas perspectivas para abordar informação são a epistemológica e a ontológica. A perspectiva epistemológica, mais comum nas Ciências Sociais, associa informação à conhecimento, significado, semântica. Por outro lado, a perspectiva ontológica associa a informação a estrutura, a forma das coisas, a maneira de organizar a natureza. Essa última perspectiva é bem estabelecida em áreas com a Física, a Cosmologia, a Filosofia etc. Quando se considera uma única visão – a epistemológica – as descobertas da ciência moderna são ignoradas. Uma visão assim única limita a perspectiva a uma abordagem antropocêntrica da informação.

A Ciência da Informação, por vocação, tem contribuído amplamente na busca de soluções para resolver problemas relacionados a dados, informação, e conhecimento, bem como aqueles que envolvem seus registros, no

7 SARACEVIC, T. Ciência da Informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, v. 1, n. 1. p. 41-62, 1996



contexto social, institucional ou individual. Essa contribuição é proveniente também de um ramo da pesquisa que, pode-se dizer, é parte do núcleo duro da Ciência da Informação: os vocabulários controlados. A importância do assunto é ilustrada em um levantamento<sup>8</sup> que identificou, com base no idioma inglês e desde 1930, mais de 130 tipos “linguagens controladas” existentes ao redor do mundo. A pesquisa em Ciência da Informação também recebe menção nesse levantamento e, de fato, tem comprovada *expertise* no assunto.

Ainda assim, novos desafios continuam a surgir: um vocabulário controlado do século XXI está em uma máquina e o controle não é feito mais por restrições linguísticas, mas por restrições lógicas. Assim, por meio da ontologia, é possível conectar as teorias e a experiência da Ciência da Informação em vocabulários controlados à modernas linguagens de representação para computadores. De fato, o ambiente digital pervasivo impacta diretamente em campos, como a Organização do Conhecimento, um dos alicerces da Ciência da Informação herdado da Biblioteconomia. Se há 40 ou 50 anos o conhecimento era armazenado majoritariamente em bibliotecas, hoje se encontra fragmentado no mundo digital.

Muito se diz que o importante não é o meio e sim o conteúdo, mas esse argumento não é de todo verdadeiro. Imagine-se um cientista que resolva um problema e rabisque sua teoria em uma bolacha de *chopp*, compartilhando-a somente com os companheiros de bebida. Nesse caso, mesmo sabendo que a solução existe, pouquíssimos cientistas poderão incorporar as descobertas à suas pesquisas. O cientista então publica sua teoria em uma revista conhecida, mas a descreve de forma tão desleixada e ambígua que poucos podem decifrá-la. Nesse cenário, mais cientistas terão acesso à descoberta, mas serão capazes de entendê-la apenas parcialmente. Não importa quão importante seja o cientista ou quão brilhante seja sua teoria: para que a comunidade científica se beneficie de uma descoberta, ela deve ser transmitida de maneira que permita acesso imediato, bem como seu entendimento. A moral da história é que os meios pelos quais o conhecimento é transmitido podem ser tão importantes quanto o próprio conhecimento.

Os meios de acessar, coletar, armazenar e disseminar informação e conhecimento no século XXI exigem novas habilidades. Se no mundo analógico o registro é feito por uma marca física, por exemplo, da tinta sobre papel, e para ler essa marca basta saber português, no mundo digital a correspondência não é a mesma: o *software* surgiu como requisito, um intermediário necessário para

---

8 KUHN, T. A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*, v. 40, n. 1, p. 121-170, 2014.

ler as novas marcas. O filósofo francês *Pierre Levy*<sup>9</sup> alerta para esse tipo de desafio narrando um caso em que historiadores e antropólogos europeus investigavam a distribuição de comida na bacia do Mediterrâneo. Há 2000 anos, os alimentos eram armazenados em ânforas na superfície das quais eram feitas gravações indicando quem, o que, onde, quando etc. O material resultante da investigação de grupos de pesquisa, cada um cobrindo parte das descobertas, foi documentado e armazenado em sistemas variados. Para fazer avançar a pesquisa, contudo, esses recursos precisariam ser combinados e oferecidos a comunidade científica como um todo. *Levy* aponta como obstáculo o fato de que, via de regra, especialistas de Ciências Sociais e Humanas não estão familiarizados com consultas à repositórios e à recursos digitais (*queries*, não apenas acesso). Dessa forma, têm se tornado mais e mais dependentes de terceiros para executar as próprias pesquisas.

## Ontologia e tecnologia da informação

É inegável a presença dos recursos digitais na pesquisa de qualquer área científica, mesmo que alguns pesquisadores ainda confundam as perspectivas entre, por um lado, aquilo é preciso ensinar a alunos para que possam tirar proveito da tecnologia; e, por outro lado, aquilo é preciso ensinar a alunos para sejam cientistas da computação. São coisas completamente diferentes e a Ciência da Informação carece de tecnologia em temas como Organização do Conhecimento, Bibliometria, dentre outros. O sociólogo espanhol *Manuel Castells*<sup>10</sup> corrobora com a inviabilidade de se posicionar a parte do mundo digital explicando que a informação é parte integral da atividade humana e, assim, todos os processos da existência individual e coletiva são moldados pelo meio tecnológico.

A tecnologia da informação é muitas vezes vista e adotada com uma “panaceia”, capaz mesmo de substituir a mente humana. O filósofo norte-americano *John Searle*<sup>11</sup> explica que esse tipo de “fantasia” não é obra apenas do *marketing* moderno, ilustrando como a tecnologia mais atualizada foi e é usada como modelo para o cérebro. No século XVII, o filósofo alemão *Gottfried Leibniz* comparava o cérebro à moinhos; na passagem do XIX para o XX, o neurocientista britânico *Charles Sherrington* contrapunha o cérebro ao telégrafo; o psicanalista alemão *Sigmund Freud* teorizava o funcionamento do cérebro de forma similar à sistemas hidráulicos e eletromagnéticos. A metáfora da passagem do século XX para o XXI é bem conhecida: o computador digital.

9 FOLHA DE S.PAULO. Entrevista com Pierre Levy. *A tecnologia pode tirar ciências humanas da Idade Média*, 2020. Disponível em: <https://www1.folha.uol.com.br/>. Acesso 10 abr. 2020.

10 CASTELLS, M. *The Rise of the Network Society*. Oxford: Blackwell, 1996.

11 SEARLE, J. *Minds, Brains and Science*. Harmondsworth: Penguin, 1984. p. 44.

Apesar de todo apelo a soluções técnicas e, mesmo considerando que a ontologia é um artefato computacional, não se aventa aqui mais tecnologia pela tecnologia. Para resolver problemas reais, em aberto, a abordagem tecnológica em ontologias vem acompanhada de abordagem conceitual, lógica, terminológica e filosófica, sob pena de se tornar mais do mesmo. De fato, a ontologia é o único artefato dentre as tecnologias recentes ligadas a Inteligência Artificial que carece de um forte componente humano. Se isso é um benefício ou não, cada leitor formará própria opinião.

No escopo da pesquisa ou da prática profissional, sejam as ontologias artefatos estritamente computacionais, ou não, se mantém a questão: **por que ontologias?** Por que construir uma? Para que servem e quais os benefícios? Para iniciar a resposta a essa questão, que se estende por toda a obra, apresenta-se cenário onde a ontologia é sabidamente útil. Um relatório que envolve a Medicina e o atendimento clínico, da *American Hospital Association*, é um bom exemplo:

A boa notícia: [...] a maioria dos hospitais é capaz de enviar e receber dados por meio de um prontuário eletrônico [...] permitindo que informações clínicas sejam compartilhadas entre vários provedores de serviços de saúde em uma região ou estado. A má notícia: apenas 40% dos hospitais são capazes de usar as informações que recebem [...] apenas um quarto de todos os hospitais é capaz de descobrir, disseminar, receber e usar informações devido a barreiras substanciais<sup>12</sup>.

As “barreiras substanciais” citadas no destaque acima se referem a um problema mundial, não limitado a Medicina, que impacta em praticamente todos os setores: **a falta de interoperabilidade entre sistemas de informação**. De forma breve, interoperabilidade pode ser definida como a capacidade de sistemas computacionais em trocar dados sem intervenção humana.

Outro relatório internacional<sup>13</sup> dá conta que a interoperabilidade de dispositivos médicos poderia eliminar pelo menos US \$ 36 bilhões/ano em desperdícios nos ambientes hospitalares. No mesmo momento, estima-se que um hospital de 500 leitos custe cerca de US \$ 750 milhões<sup>14</sup>, o que equivale dizer que a falta de interoperabilidade consome o equivalente a construção de 50

12 AMERICAN HOSPITAL ASSOCIATION. *Why Interoperability Matters*, 2015. Disponível em: <https://www.aha.org/guidesreports/2015-10-07-aha-report-why-interoperability-matters>. Acesso: 11 mar. 2020.

13 WEST HEALTH INSTITUTE. *The value of medical device interoperability: Improving patient care with more than \$30 billion in annual health care savings*, 2013. Disponível em: <http://patientsafetymovement.org/>. Acesso: 11 mar. 2020.

14 GAMBLE, M. *200 Hospital Benchmarks*, 2012. Disponível em: <https://www.beckershospitalreview.com/lists/200-hospital-benchmarks-october-2012.html>. Acesso em: 12 fev. 2020.

hospitais por ano. Uma iniciativa da União Europeia também está em curso no mesmo sentido<sup>15</sup>. No Brasil não se tem estimativas, mas a situação não é diferente. Em 2018 o governo brasileiro negociou a compra de vocabulários médicos internacionais na tentativa de mitigar o problema em uma área vital<sup>16</sup>. O que vale para um setor importante com os cuidados à saúde, vale também para outros setores estratégicos de qualquer nação.

A falta de interoperabilidade não é uma questão nova, mas é de difícil solução por sua complexidade e pela necessidade de providências nem sempre triviais em várias frentes. Trata-se de um problema generalizado que coloca obstáculos para o avanço da ciência, pois grupos de cientistas coletam dados de forma independente, e apenas depois tentam colaborar. A interoperabilidade é um desafio para governos que desejam ver suas agências coordenadas e a ação pública racionalizada. Também é problema sério para o setor privado, onde corporações estão sujeitas a um ambiente de competição turbulento e são, a cada dia, mais pressionadas por serviços de qualidade, limpos, transparentes e de baixo custo.

As ontologias são uma das mais promissoras alternativas para lidar com questões que envolvem a interoperabilidade em cenários diversos. Antes de abordar a ontologia como alternativa, examinam-se algumas tecnologias bem conhecidas dos últimos 30 anos ou mais e suas formas de lidar com a interoperabilidade.

### Bancos de dados relacionais

O banco de dados relacional é uma tecnologia originada nos anos 1960 que evoluiu e se tornou confiável. Por sua eficiência, ainda é extensivamente adotado em todas as áreas para manipular dados. Do ponto de vista da interoperabilidade, entretanto, os bancos relacionais não apresentam o mesmo desempenho, principalmente a medida que cresce o volume de dados. Exemplos são problemas como:

- Dificuldade em realizar consultas que envolvem dois ou mais bancos de dados;
- A flexibilidade permitida ao especificar o modelo, por exemplo, deixando livre a definição de relações, gera ambiguidade e idiossincrasia;

15 SEMANTIC Interoperability for Health Network. Disponível em: <https://www.i-hd.eu/index.cfm/resources/ec-projects-results/semantichhealthnet/>.

16 SNOMED International Announces Brazil as its 33rd Member. Disponível em: <https://www.globenewswire.com/news-release/2018/04/26/1488202/0/en/SNOMED-International-Announces-Brazil-as-its-33rd-Member.html>. Acesso: 20 set. 2019.

- Não existe mecanismo de herança, de forma que para modelar entidades muito similares é preciso alterar uma tabela ou criar outra, a todo momento;
- As alterações originadas da flexibilidade e da falta de herança, acumuladas por anos, tornam a gestão complexa;
- Grandes bancos podem ter centenas de tabelas e uma consulta pode envolver milhares de linhas, tornando inviáveis algumas operações usuais.

### Sistemas de gestão integrada – *Enterprise Resource Systems* (ERPs)

Os ERPs surgiram nos anos 1980 e prometiam integrar os sistemas corporativos, o que já era visto como problema na época. A concepção estendia a ideia de sistemas de “chão de fábrica” aplicando-a à gestão administrativa. A solução de interoperabilidade dos fabricantes de ERP resumia-se em manter um único banco de dados integrado. As outras aplicações necessárias deveriam ser construídas a volta do ERP. Entretanto, tal solução também não se mostrou eficaz por diversos motivos, dentre eles:

- Os ERPs se tornavam mais e mais complexos, a cada versão, incluindo esquemas com centenas e até milhares de tabelas;
- A complexidade crescente, paradoxalmente, dificultava a integração no próprio ambiente do sistema e também com outras aplicações;
- Os projetos de implantação eram de alto risco e exibiam custos crescentes, ainda que os escopos fossem cada vez mais reduzidos;
- Com escopos reduzidos por custo e risco, os projetos de ERPs alcançavam apenas uma parcela da corporação e não se cumpria a promessa de integração.

### Armazéns de dados – *Data warehouse* (DWs)

Se ERPs sugeriam um banco de dados único como solução para integração, os DWs sugeriam integrar várias fontes em diferentes sistemas para atender as consultas das pessoas. Os DWs eram capazes de combinar dados de diferentes aplicações, permitindo acesso e consultas aos resultados combinados. Os DWs, no quesito de integração, padeciam de problemas que eles mesmos criavam:

- O processo de popular os DWs – o *Extract, Transform and Load* (ETL) – se mostrava trabalhoso e demorado, exigindo muitas horas de analistas;
- Apesar de altos custos de projeto, os DWs podiam manter apenas um subconjunto reduzido dos dados de interesse para o negócio;
- Não eram capazes de nenhum tipo de transação, apenas de gerar relatórios;
- Quanto mais fontes consultadas para atender necessidades de negócio, mais tabelas eram criadas, de forma que popular o DW poderia demorar dias ou semanas;
- O subconjunto integrado era, além de parcial, originado em um procedimento complexo que precisava ser repetido a cada necessidade.

### *Data lakes*

Os custos proibitivos dos DWs levaram a popularização de *data lakes*, os quais usam técnicas de *big-data* baseado em processamento paralelo do lado de *hardware*, e em *machine-learning* do lado do *software*. Ao invés de conformar dados a estrutura de relatórios, os *data lakes* lidavam com dados em seu estado original e depois aplicavam *big-data* à fases problemáticas dos DWs. Para fins de interoperabilidade, o *data-lakes* mantem problemas dos DW e trazem outros:

- Ao contrário do DWs, o *data lakes* tem alto custo e complexidade na recuperação por envolver *big-data*, o qual vingou apenas nas *big-techs* e que requer cientistas de dados, métodos estatísticos, *machine learning*, Inteligência Artificial etc.;
- Técnicas muito adotadas em *big-data*, como *machine-learning*, funcionam bem em cenários com enormes volumes de dados para cada tipo de problema, em geral não disponíveis em corporações; os algoritmos de *machine-learning* tem ainda pouca capacidade de lidar com cenários ligeiramente diferentes;
- As técnicas adotadas em *data lakes* são *black-boxes*, não lidam com conhecimento de *background* e nem com hierarquias, presumem um cenário estável e não distinguem causação de correlação, dentro outras limitações.

## Ontologia

Do ponto de vista conceitual, o projeto e desenvolvimento de ontologias – ontologias bem fundamentadas – depende de um elaborado arcabouço conceitual que enseja a execução de atividades diversas antes de qualquer implementação. O aspecto conceitual é tanto ou mais importante que o aspecto tecnológico porque sem ele, em grande medida, mantém-se muitos dos problemas. Esse arcabouço conceitual envolve noções de representação, terminologias e vocabulários, Metafísica e Filosofia da Linguagem, Lógica e Semântica, Ciência da Informação, classificação e categorização, gestão de projeto e arquitetura da informação.

O trabalho em ontologias começa naturalmente quando grupos interdisciplinares se envolvem no compartilhamento e na tradução da informação e conhecimento. Nesse contexto, uma das primeiras providências é reunir a informação em um único repositório onde o projeto possa evoluir organizada-mente. As inevitáveis divergências conceituais e terminológicas, num primeiro momento, são resolvidas caso a caso. Gradualmente, as pessoas percebem que as incompatibilidades são contornáveis, com vantagens significativas, pelo uso de uma estrutura comum. Essa estrutura é a ontologia, uma ontologia de referência, um conjunto compartilhado de termos que representa entidades e define, sem ambiguidade, como tais entidades se relacionam. Em etapas posteriores, a ontologia pode ser transferida para o meio computacional para aproveitar os benefícios da tecnologia da informação.

Do ponto de vista tecnológico, de forma geral, a ontologia se presta a diversos fins, mas aqui se mantém a ênfase em problemas já citados. Nesse sentido, a ontologia tem potencial para ser parte da solução para a falta de interoperabilidade, um problema que exige ação em diversas frentes. Um primeiro aspecto a considerar é que uma ontologia é bem mais simples, do ponto de vista técnico, do que a maioria as tecnologias já mencionadas. Apesar dessa simplicidade, pode funcionar como esquema mediado para o *framework* de integração, proporcionando raciocínio automático e consultas sobre expressões complexas.

Ao contrário dos sistemas baseados banco de dados, mais populares, os sistemas baseados em ontologias manipulam modelos de domínio complexos e possivelmente incompletos. Os bancos de dados trabalham com a cláusula de *mundo-fechado*, o quer dizer se um conjunto de dados não está no banco, ele simplesmente não existe. Ontologias, por sua vez, trabalham com a cláusula de *mundo-aberto*, ou seja, se um conjunto de dados não está na ontologia, pode de fato não existir, mas pode também existir e mesmo se originar de raciocínio automático. Assim, as ontologias são capazes representar

e responder consultas de uma forma que outras tecnologias não podem. Um exemplo pode ilustrar diferenças.

### O enigma de *Harry Potter*

Ao responder a uma consulta – quantos amigos tem *Harry Potter*? – uma ontologia responderia de forma diferente de um banco de dados: enquanto o banco de dados responde “dois amigos”, a ontologia responde “pelo menos um amigo”. Mas como isso é possível?

Considere dados sobre *Harry Potter*, em código legível por máquina e em português:

hasPet some owl:Thing SubclassOf: Human	Alguns humanos são donos de animais de estimação
Fluffy SubclassOf: petOf only Wizard	Fluffy é um animal de estimação que pertence apenas a mágicos
HarryPotter: Wizard	Harry Potter é um mágico
DracoMalfoy: Wizard	Draco Malfoy é um mágico
HarryPotter hasFriend RonWeasley	Harry Potter é amigo de Ron Weasley
HarryPotter hasFriend HermioneGranger	Harry Potter é amigo de Hermione Granger
HarryPotter hasPet Hedwig	Harry Potter tem animal de estimação

Para o banco de dados *Harry Potter* tem dois amigos, porque isso é o que os dados informam na quinta e na sexta linhas acima: os amigos de *Potter* são *Ron* e *Hermione*. Uma ontologia por outro lado, faz uso da cláusula de mundo-aberto de forma que não se fia apenas no que está declarado, mas no que pode ser inferido. Assim, a resposta “pelo menos um” faz sentido caso *Ron* e *Hermione* sejam a mesma pessoa. Se isso ocorre, a resposta seria “pelo menos um amigo”; caso contrário, a resposta seria “dois amigos”, como no banco de dados. Para que a ontologia respondesse o mesmo que o banco de dados, seria preciso acrescentar duas linhas ao final do modelo informando que *Ron* e *Herminoe* são entidades disjuntas.

Observa-se nesse exemplo simples que a ontologia tem mais a oferecer, ou melhor dizendo, o mesmo conhecimento com maior precisão. Isso faz diferença na construção de modelos para domínios complexos e em consultas possíveis. Na verdade, diversos aspectos da interoperabilidade entre sistemas são problemas de representação que ocorrem em artefatos computacionais, como ontologias. Além disso, outras vantagens podem ser enumeradas:

- Os princípios de classificação envolvidos na construção de ontologias proporcionam modelos bem fundamentados;



- Os sistemas comerciais, intencionalmente ou não, escondem os modelos enquanto a ontologia, um sistema declarativo, torna explícito o modelo e o conhecimento;
- Modelos simples nem sempre são úteis e por isso vão recebendo adições, em geral, de forma redundante, as quais seriam unificadas por referência à termos da ontologia;
- Ontologias podem, de forma fundamentada e sem riscos, lidar com mudanças no status do conhecimento, enquanto em bancos de dados mudanças levam a retrabalho com a criação imediata de novas tabelas;
- A ontologia é computável e útil no local onde está, além de continuar presente na forma original em aplicações distribuídas para consultas em seu escopo original;
- A cláusula de mundo-aberto permite construir modelos que têm estrutura e semântica bem definidas, ainda que não sejam limitados àquela estrutura;
- A ontologia faz uso de identificadores únicos da web – o *Uniform Resource Identifier* (URI) – que definem recursos do modelo de forma unívoca, identificando-os ao longo das três camadas (conceitual, lógica e física).

Os aspectos tecnológicos citados aqui, apenas para introduzir ontologias, são objeto do estudo no Volume II dessa obra. Naquele volume estão explicadas as tecnologias de Web Semântica, a Lógica Descritiva, as linguagens de consulta, as restrições, as ferramentas, metodologias e engenharia de ontologias, dentre diversos outros assuntos como a própria interoperabilidade.

### **Uma breve nota: ontologia e Aristóteles**

A Ciência da Informação, em alguma medida, se fundamenta em filósofos diversos, mas nem tanto em Aristóteles. Exemplos de filósofos populares na área são *Immanuel Kant* e *Ludwig Wittgenstein*, além de pesquisadores proeminentes do século XX, como por exemplo *Ingetraut Dahlberg* e *Shiyali Ranganathan*. Todos esses filósofos, estudiosos, e cientistas são apresentados em capítulos adiante, onde se encontram breves extratos de suas teorias.

Dito isso, cabe informar que grande parte da teoria da presente obra é orientada pela tradição aristotélica. Isso explica, em parte, porque pesquisadores seminais da Ciência da Informação, dentre outros contemporâneos, não figuram dentre as principais referências. Isso ocorre também pelo fato de que esses autores não pesquisaram *Ontologia Aplicada*, uma abordagem que

é explicada no primeiro capítulo desse volume. A literatura de ontologia em Ciência da Informação está em construção, de forma que, como em outras ocasiões, muito é tomado emprestado de outras ciências, principalmente de campos da Filosofia. A preferência por Aristóteles também tem suas razões.

Do ponto de vista teórico, após um longo período de esquecimento, surgiu um novo interesse nas teorias de Aristóteles a partir dos anos de 1970. Um dos mais importantes fatores para isso foi a necessidade de noções aristotélicas no âmbito de novas teorias lógicas da época – em particular, a Lógica Modal – desenvolvida por *Saul Kripke*, um dos grandes filósofos do século XX. *Kripke* reviveu e estendeu as noções aristotélicas necessárias a sua teoria e acabou assim provendo respostas e argumentos aos neo-aristotélicos. Naquele momento, os seguidores da tradição aristotélica se defendiam de duras críticas vindas, por exemplo, do filósofo norte-americano *W. V. Quine* e seus colegas.

Motivos teóricos adicionais para a renascença aristotélica, a partir dos anos 1980, foram desafios lançados contra a teoria da causalção que questionaram o que ainda estava em vigor devido *David Hume*, um filósofo inglês do Iluminismo. Além disso, os debates na Filosofia da Ciência geravam ideias, com destaque para as teorias de *Nancy Cartwright*, cujos resultados retratavam cenário tipicamente descrito na teoria de Aristóteles. Todas essas razões, dentre outras, levam filósofos contemporâneos a afirmar que categorias aristotélicas não são apenas produtos da imaginação daquele filósofo, mas sim reais manifestações da forma pela qual os seres humanos acessam o mundo.

Do ponto de vista prático, a ontologia aristotélica se adequa muito bem a uma forma de estruturar o mundo passível de ser incorporada a computadores. Filósofos contemporâneos como *Jonathan Lowe*, *Barry Smith*, *Ingvar Johansson*, *Ludger Jassen*, dentre muitos outros, em um esforço colaborativo com grupos e entidades científicas ao redor do mundo, delinearam a Ontologia Aplicada: uma conjunção entre Filosofia e tecnologia da informação que objetiva prover melhorias aos sistemas e possibilitar novos patamares de qualidade as instituições.

## **Começando a jornada**

Escolher que teoria usar na pesquisa científica suscita discussões as vezes um tanto acaloradas. Um exemplo recente é o ocorrido em grupo de discussão, reunindo um cientista da informação, uma especialista em Filosofia da Mente e uma antropóloga, dentre outros profissionais. Alguém sugeriu discutir o tema “identidade” e como o grupo discutia pela internet, várias mensagens foram trocadas, às vezes de forma um tanto impaciente, antes que cada um pudesse entender o alvo do empreendimento. Para um, identidade dizia

respeito a teorias da identidade que estudam como distinguir coisas através de sua natureza; para outra, a distinção de tipos e *tokens* parecia obsoleta e identidade seria um fenômeno mental; ainda, alguém entendeu que a discussão seria sobre identidade cultural de grupos étnicos. Mesmo que a ciência seja a princípio um ambiente em que a clareza seja quase uma obrigação, problemas na comunicação entre cientistas são uma realidade.

Não parece razoável, quando se busca a verdade, suscitar disputa sobre qual teoria é melhor. Sem entrar no mérito de cada alternativa, esse tipo de disputa vai se manter travestida em outro formato, em questões mais amplas, por exemplo, no âmbito da Filosofia da Ciência. Não parece haver uma alternativa única, e não é a toa que algumas discussões se mantem seculares. Cabe, entretanto lembrar que, como diz *Humberto Eco*, as escolhas têm consequências: “quando se trata de ir a Lua, a interpretação de Copérnico é melhor que o *Almagesto* de Ptolomeu”<sup>17</sup>.

## Simbologia adotada no livro

### a) Texto

- No geral, o texto apenas usa *itálico* ou **negrito** para diferenciar, na primeira ocorrência, palavras que podem ter significado diferente do uso cotidiano.
- Aspas são um dispositivo para indicar quando a linguagem é usada para teorizar ou para falar sobre propriedades da própria linguagem. Exemplos de situações são:
  - » “A neve é branca” é verdadeiro se e somente se a neve é branca.
  - » “Aristóteles” se refere a Aristóteles.
  - » “O” é um artigo definido em português.
  - » “Amazonas” têm oito letras.
- Onde necessário, aspas duplas são aplicadas sobre aspas simples, aspas triplas são aplicadas sobre aspas duplas, e assim por diante.
- Em casos particulares e diferenciados, o próprio capítulo provê a convenção.

### b) Quadros

Quadros como esse, com linhas duplas, contém material adicional para ajudar a entender do assunto em questão.

Quadros como esse, sem linha e sombreado claro, contém exemplos ou destacam sentenças e fragmentos do texto principal.

### c) Fórmulas

As fórmulas são grafadas em fonte diferente e estão cercadas por barras, como em:

| se X é semelhante a Y, então X representa Y |  
 | F = x é feliz |  
 |  $\exists x [ Sx \wedge Ax ]$  |

### d) Lógica

Sentenças em lógica são evitadas, pois não fazem parte da formação usual em Ciência da Informação. A preferência é a explicação em linguagem natural, mas onde isso não for possível adota-se a notação básica:

Operador	Nome técnico	Significado	Exemplo
$\neg$	Negação	Não	$\neg P$
$\wedge$	Conjunção	E	$P \wedge Q$
$\vee$	Disjunção	Ou	$P \vee Q$
$\rightarrow$	Condicional	Se, então	$P \rightarrow Q$
$\leftrightarrow$	Bicondicional	Se e somente se	$P \leftrightarrow Q$
$\forall$	Quantificador universal	Para todo	$\forall x$ se lê "Para todo x..."
$\exists$	Quantificador existencial	Alguns	$\exists x$ se lê "Existe um x tal que..."

\*\*\*