

基本形式化本体及其中文版介绍*

朱彦

郑捷

(中国中医科学院中医药信息研究所
北京 100700)

(宾夕法尼亚大学医学院 美国宾夕法尼亚州 19104)

李晓璇

杨啸林

何勇群

(中国医学科学院/北京协和医学院医学
信息研究所/图书馆 北京 100005)

(中国医学科学院/北京协和医学院
基础医学研究所 北京 100005)

(密歇根大学医学院
美国密歇根州 48109)

〔摘要〕 分析基本形式化本体 (Basic Formal Ontology, BFO) 来源、特点, 介绍中文版 BFO (BFO-zh) 翻译和推广工作情况, 包括背景、翻译原则、工作流程, 详细阐述 BFO 中类的层次结构, 为国内本体研究提供参考。

〔关键词〕 基本形式化本体; 中文翻译; OntoChina; 开放生物/生物医学本体工场

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.01.006

Introduction to Basic Formal Ontology and Its Chinese Version ZHU Yan, Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China; ZHENG Jie, University of Pennsylvania Perelman School of Medicine, PA 19104, USA; LI Xiaoying, Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China; YANG Xiaolin, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China; HE Yongqun, University of Michigan Medical School, MI 48109, USA

〔Abstract〕 The paper analyzes the sources and characteristics of Basic Formal Ontology (BFO), introduces the translation and promotion of Chinese version of BFO (BFO-zh), including the background, translation principles and workflow, expounds the class hierarchy of BFO in detail, so as to provide references for domestic ontology study.

〔Keywords〕 Basic Formal Ontology (BFO); Chinese translation; OntoChina; Open Biological & Biomedical Ontology (OBO) Foundry

〔收稿日期〕 2020-07-24

〔作者简介〕 朱彦, 副研究员, 发表论文 10 余篇; 通讯作者: 何勇群, 副教授。

〔基金项目〕 国家重点研发计划资助“疗效导向下中医辨证论治能力提升数字化关键技术及平台构建”(项目编号: 2019YFC1710400、2019YFC1710401); 中央级公益性科研院所基本科研业务费专项资金资助“上市中成药信息标准化探索性研究”(项目编号: ZZ13-YQ-126); 中国医学科学院医学与健康科技创新工程项目资助“生物医学本体支持的适用数据元素表示和应用系统建设”(项目编号: 2018-I2M-AI-009); 国家重点研发计划资助“疾病表型组-实验组学数据分析、注释与整合”(项目编号: 2017YFC0908404)。

1 引言

随着大数据时代的到来，海量数据管理和集成为知识发现带来大量机会。本体是用人和计算机都可以理解的术语 (terms) 及关系 (relations) 来描述实体 (entity) 以及实体之间相互关系，提供对事物本质的统一认识，能有力支持多源异构数据的集成并最终实现知识发现。基因本体 (Gene Ontology, GO)^[1] 的巨大成功激励了越来越多研究者投身到生物学领域的本体研究和开发中。开放生物与生物医学本体 (Open Biological & Biomedical Ontology, OBO) Foundry 是其中的典型代表^[2]，成为目前生物与生物医学本体领域最有影响力的社区。中国本体学者借鉴其先进理念和成功模式成立了本体中国 (OntoChina) 社区，致力于国内生物医学信息本体研究与推广^[3]，完成了 OBO Foundry 中的部分核心本体的本地化，其中包括 OBO Foundry 中多数本体都采用的顶层本体基本形式化本体 (Basic Formal Ontology, BFO)^[4]。本文将系统介绍 BFO 的来源、特点以及中文版 BFO (BFO-zh) 的翻译和推广工作并梳理其层次结构，以期为国内本体研究提供参考。

2 BFO 基本情况介绍

2.1 概述

BFO 起源于 2002 年德国大众基金会 (Volkswagen Foundation) 资助项目，最初由 Barry Smith 和 Pierre Grenon 发起，后专家参与持续改进更新，目前最新版本于 2019 年 8 月发布^[5]。BFO 已提交国际标准化组织 (International Organization for Standardization, ISO) 的数据管理和交换技术委员会投票审核，于 2020 年 3 月份正式发布^[6]。

2.2 BFO 特点

2.2.1 在生物医学领域取得巨大成功 依托于

OBO Foundry 的发展，目前以 BFO 作为顶层框架的本体已经超过 300 个^[7]，尤其在生命医学领域得到广泛应用，见图 1。中国的研究者尝试基于 BFO 构建领域本体，如传统中药本体 (Traditional Chinese Drug Ontology, TCDO)^[8] 和中国植物物种多样性领域本体^[9]。



图 1 以 BFO 为顶层本体的中层本体及领域本体

2.2.2 体量小而精 BFO 最初构建时借鉴了部分 SUMO 和 DOLCE 的核心概念，但 BFO 是为支持集成、分析、整合科研数据而开发的顶层本体，设计更加简明，方便不同领域和粒度级别科学家复用。

2.2.3 基于本体实在论的哲学基础 BFO 以本体实在论 (Ontological Realism) 为哲学基础，即认为本体所表征的是现实本身的普遍特征。这一点与部分学者尤其是信息学研究领域部分专家所持概念取向不同，他们认为本体表征的是科学家关于现实的看法。在表述上最明显的区别是 BFO 及相关论著中尽量避免使用“概念 (concept)”这一术语，取而代之的是用哲学意义的“实体 (entity)”、“类 (class)”和“共相 (universal)”等直接指称现实的普遍特征。

3 中文版 BFO 的翻译与推广

3.1 背景

3.1.1 引入翻译 OBO Foundry 本体 为了推进国内生物医学本体研究与应用，2017 年在国家人口

与健康科学数据共享服务平台（现为国家人口与健康科学数据中心）支持下，OntoChina 社区系统引入并翻译了 OBO Foundry 的一些重要本体，其中包括 BFO^[10]、Ontology for Biomedical Investigations (OBI)^[11]、Relation Ontology (RO)^[12]、Ontology for General Medical Science (OGMS)、Human Phenotype Ontology (HPO)^[13]和 Cell Line Ontology (CLO)^[14]，开发了 MedPortal 本体资源库^[15]以存储和共享。

3.1.2 BFO 中文翻译 受 Barry Smith 教授委托，OntoChina 于 2017 年底正式开始 BFO 的中文翻译工作，还同步翻译了麻省理工出版社出版的 Barry Smith 等撰写的 BFO 本体著作 *Building Ontologies with Basic Formal Ontology*^[10]，该书即将由人民卫生出版社在 2020 年出版^[16]，将是第 1 本向国内读者系统介绍 BFO 及基于 BFO 构建本体的理论、方法和技术的书籍。

3.2 翻译基本原则

3.2.1 制定翻译基本原则 根据 BFO 特点及工作组的实际情况制定翻译工作基本原则，除了力求做到信、达、雅以外，要尽可能了解作者在原语言环境及学科背景下所要实现的效果，在新的语言环境和文化背景下重建这种效果。同时从中国传统文化尤其是传统哲学中寻找灵感。

3.2.2 术语选择 遵循大道至简原则，因为 BFO

中本身的实体就是高度抽象和概括的，翻译成中文务必简洁直白，尽量采用常见字词。一方面尊重术语哲学根源，同时考虑用户主要是具体领域专业人员，避免过于艰涩难懂。对于平级关系术语尤其是一分为二的情形，尽量使用对偶形式词语，同时兼顾父子关系术语，体现从属关系。

3.3 翻译工作流程

3.3.1 人员筹备 组织不同背景专业人员以志愿者形式加入工作组，邀请 Barry Smith 教授担任顾问。

3.3.2 平台搭建 由于参与者来自全球各地，使用 WebProtege 作为协同编辑和讨论平台并进行问题跟踪。

3.3.3 初步翻译 工作组利用近半年时间对 BFO 中的术语逐个进行溯源、讨论和翻译。

3.3.4 征求意见 通过 OntoChina 和微信公众号等平台广泛公开征求意见，基于意见进行讨论调整，加入相应中文同义词。

3.3.5 定稿公开 定稿后将 BFO-zh 上传到 MedPortal 平台 (<http://medportal.bmicc.cn/ontologies/BFO-ZH>) 和 GitHub (<https://github.com/zhuy-an166/BFO-zh>) 公开共享。

4 BFO 类层次结构 (图 2)

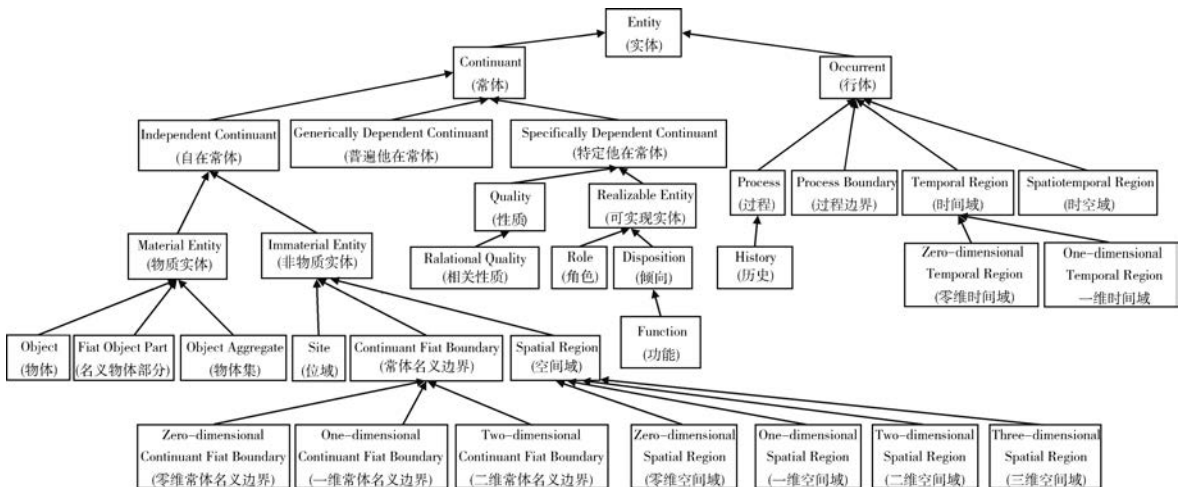


图 2 BFO 类层次结构

4.1 实体及其基本类型

BFO 首先从现实运行方式出发,认为现实是由实体 (entity) 组成的。哲学家和科学研究者常用“实体 (或存在体)”一词指称以任何方式存在的任何事物。基于两种不同世界观, BFO 将实体分为为两种基本类型: 一是常体 (continuant), 即历经时间而能持续或保持不变的实体; 二是行体 (occurrent), 即发生或出现的实体。两者最直观地区分就是常体没有时间部分 (temporal parts), 而行体有。翻译采用汉字中常见、起源较早且对偶存在的“常”和“行”, 依据儒家典籍《荀子·天论》“天行有常, 不为尧存, 不为桀亡”和佛家典籍《宗镜录 - 卷第四十二》中的“诸行无常, 是生灭法”的表述。“常”字能较好地概括历经时间不变的特征; “行”字能直观地体现四维时空中伸展穿行的时空蠕虫 (spacetime - worms) 意象^[17]。

4.2 常体

4.2.1 基本特征 一是能历经时间持续、持久或连续存在并能保持其同一性; 二是没有时间部分。具体可分为自在常体、特定他在常体、普遍他在常体 3 类。

4.2.2 自在常体 (independent continuant) 即不依赖于其他实体自己能独立存在的实体, 是作为性质 (quality) 持有者 (bearer) 的一类常体。(1) 物质实体 (material entity)。指有某一份物质作为其组成部分的自在常体, 其可以在三维空间中延展且持续存在一段时间。进一步可以分为物体 (object)、物体集 (object aggregate)、名义物体部分 (fiat object part)。物体即现实存在的基本单元或颗粒, 其特性包括: 在三维空间中延展; 自成一体; 各部分最大程度地自连接。例如原子、细胞、器官、生物体、恒星等。物体集即由物体的集合所组成的物质实体, 例如一堆石头、你血液中的一团细菌、一群鹅、某家医院的所有患者。名义物体部分即某个更大物体的真子部分 (proper

part), 但没有任何物理间断将其与该物体的其余部分分割开 (因此其本身不是物体)。“名义 (fiat)”一词表示边界规范性并非物理间断, 其存在性也不依赖于认知主体。例如人体上躯干、实心金属勺的手柄、地球的西半球等。(2) 非物质实体 (immaterial entity)。指不包含任何物质实体作为其部分的自在常体。包括常体名义边界 (continuant fiat boundary)、位域 (site)、空间域 (spatial region)。其中常体名义边界是某种物质实体的边界, 其精确地存在于物体与其周边相遇的地方。例如地球表面或细胞膜表面。BFO 识别了 3 种不同子类型, 即二维、一维和零维常体名义边界。位域即由某物质实体所包含的孔, 包括人体鼻腔、静脉 (血液流过的空腔)、胃肠道的内腔以及苏伊士运河 (沟渠) 等。空间域 (spatial region) 是空间的某部分, 独立于物质实体而存在。BFO 识别了 4 种不同的空间域, 即三维、二维、一维和零维空间域 (又称为空间体、面、线和点)。需要注意的是, 常体名义边界和位域可随其宿主的移动、形状或大小的变化而改变其位置、形状和大小。而空间域相对于参照系永远不会发生移动。

4.2.3 特定他在常体 (specifically dependent continuant) 依赖于一个或多个特定的自在常体而存在的常体, 分为两种子类型: 性质 (quality) 和可实现实体 (realizable entity)。性质与可实现实体的不同之处在于, 性质在其完全依附的实体中得到充分展现 (exhibited)、表现 (manifested) 或实现 (realized)。而可实现实体可以依附不被实现, 或者不同程度地实现。(1) 性质。例如这个肾脏的质量、这一份血液的颜色、这只手的形状等。在本体构建过程中制定定义时通常将物体及其部分的性质称为种差 (differentiae)。(2) 可实现实体。例如医生的角色、金属导电的倾向、生殖器官的功能等。对应其更具体的类型为角色、倾向和功能。角色是一种外部决定性的 (externally grounded) 可实现实体, 其存在是因为持有者处于特定的物理、社会或制度环境中且这些环境对于持有者来

说是非必需的。而倾向则不同,是内部决定性 (internally grounded) 的,如果一个实体在物理上以某种方式存在,那么就会拥有某特定倾向,如果这种方式不存在就会失去这种倾向。功能是基于持有者的物理构造而存在的一种倾向,该特定物理构造是因为自然选择(在生物学实体情况下)或者有意设计(在人为情况下)而产生的。

4.2.4 普遍他在常体 (generically dependent continuant) 其受到所谓的非迁移公理 (axiom of non-migration) 约束,不能从一个持有者迁移到另一个持有者。然而信息实体(如 pdf 文件)一方面依赖于某持有者,必须保存到某种物理存储设备上;另一方面数值上等同的信息实体又可以存在多个副本,即可以从一台电脑复制到另外一台电脑。为了处理这种特殊的客观存在, BFO 引入了普遍他在常体范畴,定义为依赖于一个或多个自在常体作为其持有者的一种常体。

4.3 行体

4.3.1 概述 指发生或出现的实体,如“事件”、“过程”或“发生的事情”等。BFO 的观点是将整个时空视为四维整体,过程在其中开始、持续和结束。每个过程可看作是一个在时间上延展的连续体,像时空蠕虫一样在统一的容器(即整个时空)中伸展。行体的子类型包括过程、过程边界、时间域和时空域。

4.3.2 过程 (process) 因为在时间上发生或出现而存在的一种行体,具有时间部分并依赖于一个或多个物质实体。包括这个有机体的生命过程、这个疾病的病程、那只鸟的那次飞翔、这个细胞分裂过程等。其中历史 (history) 是过程的一个重要子类型,是物质实体或位域所占据的时空域内发生的所有过程总和。

4.3.3 过程边界 (process boundary) 指过程的瞬时时间边界。过程边界是其所属过程的开始和结束,包括突触的形成、快速的眼球运动、睡眠的开始等。

4.3.4 时间域 (temporal region) (整个) 时间的一部分,时间域是将时空域投影到该时间维度上的结果。可分为零维时间域 (zero-dimensional temporal region) 和一维时间域 (one-dimensional temporal region)。其中零维时间域也称为时刻(或瞬间),是没有长度的时间域;一维时间域也称为时间间隔,是在时间上延展的时间域。

4.3.5 时空域 (spatio-temporal region) 指行体所在的时间和空间范围。时空域是时空的一部分(即整个时空的一部分),每个时空域都是相对于某四维坐标系统参照系所定义的。包括某个人生命、一个癌症肿瘤的发展、一个细胞减数分裂的过程、一场战争所占据的时空域等。

5 结语

在本体理论研究及技术应用上,我国与国际先进水平尚有不小差距。近年来国内越来越多研究者开始关注和应用先进的本体构建理念, BFO 本地化工作正是本体理论在中国推广和应用的基础性探索。未来将遵循可误论 (Fallibilism) 原则,随着 BFO 更新, BFO-zh 将更新修正;同时在使用过程中广泛听取中文用户意见建议,修正中文翻译共识,将 BFO 应用到实际领域本体开发中。

参考文献

- 1 Ashburner M, Ball C A, Blake J A, et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium [J]. Nature Genetics, 2000, 25 (1): 25-29.
- 2 Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration [J]. Nature Biotechnology, 2007, 25 (11): 1251-1255.
- 3 Pan H, Zhu Y, Yang S, et al. Biomedical Ontologies and Their Development, Management, and Applications in and beyond China [J]. Journal of Bio-X Research, 2019, 2 (4): 178-184.

(下转第 60 页)

系统引进知识图谱技术,从电子病历中抽取临床知识,通过本体建模方法构建病历图谱,可将病历数据归纳为患者、疾病、症状、医生、科室、医院、治疗、医嘱、检验、检查、手术、药品、住院事件、门诊事件、体检事件等概念。通过与医疗知识库相关联为实现医学知识检索、疾病预测、用药推荐和相似病历等奠定良好基础,为临床科研工作者提供新的结论以供参考,达到更好地辅助临床医学研究的目的。临床电子病历知识图谱是专科知识图谱的基础,从临床病历中抽取专科电子病历知识图谱,结合专科知识库构建专科知识图谱。

5 结语

电子病历智能检索系统从2018年下半年开始在海南省人民医院上线运行,已经完成近50万份电子病历的后结构化处理,整合临床医嘱和检验检查结果约650万条数据。系统集成业务信息系统的

院内数据以及电子健康档案、全员人口库等院外数据,重构数据模型,以患者为中心重新整合,实现患者360全景视图,为患者和医生提供全方位的数据查看。电子病历智能检索系统的应用提升信息化对临床科研的支撑保障能力,为医生深入挖掘医疗数据价值提供可能,为最终实现精准医疗和智慧医疗服务体系奠定坚实基础。

参考文献

- 1 甘霖. 基于云计算的电子病历全文检索系统 [J]. 中国数字医学, 2016, 11 (12): 41-43.
- 2 赵景越, 徐松青, 蔡连忠. 全文检索引擎技术在电子病历中的应用 [J]. 数字技术与应用, 2014 (12): 89-90.
- 3 王晓, 罗二平, 张健. 基于语义的电子病历智能全文检索 [J]. 医疗卫生装备, 2008, 29 (4): 45-46.
- 4 阮彤, 高炬, 冯东雷, 等. 基于电子病历的临床医疗大数据挖掘流程与方法 [J]. 大数据, 2017, 3 (5): 83-98.

(上接第28页)

- 4 Smith B, Grenon P. Basic Formal Ontology (BFO) Home [EB/OL]. [2020-03-24]. <http://basic-formal-ontology.org/>.
- 5 Smith B. OBO Library - Basic Formal Ontology [EB/OL]. [2019-08-26]. <http://purl.obolibrary.org/obo/bfo.owl>.
- 6 ISO/IEC. Information Technology - Top - level Ontologies (TLO) - Part 2: Basic Formal Ontology (BFO) [EB/OL]. [2020-07-04]. <https://www.iso.org/standard/74572.html?browse=tc>.
- 7 Smith B. BFO Users [EB/OL]. [2020-07-06]. <http://basic-formal-ontology.org/users.html>.
- 8 Zhu Y, Liu L, Gao B, et al. TCDO: a community-based traditional Chinese drug ontology [C]. Shanghai: Proceedings of the 11th International Biocuration Conference (Biocuration-2018), 2018.
- 9 段宇锋, 黄思思. 基于BFO构建中文植物物种多样性领域本体的研究 [J]. 现代图书情报技术, 2015, 265 (12): 80-87.
- 10 Arp R, Smith B, Spear A D. Building Ontologies with Basic

Formal Ontology [M]. Cambridge: MIT Press, 2015.

- 11 Bandrowski A, Brinkman R, Brochhausen M, et al. The Ontology for Biomedical Investigations [J]. Plos One, 2016, 11 (4): e0154556.
- 12 Smith B, Ceusters W, Klagges B, et al. Relations in Biomedical Ontologies [J]. Genome Biology, 2005, 6 (5): R46.
- 13 Groza T, Kohler S, Moldenhauer D, et al. The Human Phenotype Ontology: semantic unification of common and rare disease [J]. American Journal of Human Genetics, 2015, 97 (1): 111-124.
- 14 Samtivistijai S, Lin Y, Xiang Z, et al. CLO: the cell line ontology [J]. Journal of Biomedical Semantics, 2014 (5): 37.
- 15 郭靖文, 杨晟, 史涪仁, 等. MedPortal: 面向精准医学的生物学本体资源存储和应用平台 [J]. 中国生物医学工程学报, 2017 (5): 557-564.
- 16 朱彦, 何勇群, 郑捷, 等. 基于基本形式化本体的本体构建 [M]. 北京: 人民卫生出版社, 2020.
- 17 Smith, Barry. Classifying Processes: an essay in applied ontology [J]. Ratio, 2012, 25 (4): 463-488.