

什么是语义技术

王 姝

(四川大学华西口腔医学院编辑部)

1 语义技术

1.1 语义

语义(semantic)是数据符号的含义，是在相关领域里的解释限定和逻辑表达。通俗地说，语义是数据符号所对应的现实世界中的事物所代表的含义和关系。

1.2 语义技术

语义技术(semantics technology)的定义是：期望通过在网络信息创作和发布中嵌入机器可读的、代表某类知识内在含义的语义标注，使机器能够自动地处理和集成网上可用的信息，从而能够提高信息服务质量，并开拓各种崭新的、智能化的服务。早在 2001 年，蒂姆·伯纳斯就预测：“语义网将有可能用人们无法想象的方式深刻地改变科学知识的生产和共享”^[1]。

1.3 语义技术的实质

语义技术的实质是在撰写、录入、发布过程中，将每段信息中的关键数据、图片、表格、摘要、结论、参考文献等进行标记（计算机业内称为“锚点”），通过自动化程式化的链接将这些标记关联到其他相关信息。当读者需要获取某个方面的信息时，相关的信息链接可以被快速准确地展现在读者的眼前，这就使单个的信息变成了无限延展的内容，变成了五彩斑斓、栩栩如生的动态信息世界。基于语义技术的信息发布、集成与交互设计已经逐渐成为科学交流系统发展的前沿。它增强了数据和信息之间的关联，通过提高数据和信息的获取效率，从而更方便用户发现、学习和理解复杂的知识体系。

2 语义技术与科技期刊

2.1 科技期刊的数字出版

科技期刊出版已经从传统的纸质出版模式，发展到今天的数字出版模式。后者在出版的整个过程中，将所有信息都以统一的二进制代码的数字化形式存储于光盘、磁盘等介质中，信息的处理与接收则借助计算机或终端设备进行。它强调内容格式、生产模式、运作流程、传播载体、阅读消费和学习形式的数字化。随着科研效率逐步提高，科研成果增长迅猛，科研产出的数字化程度也越来越高。海量的数字化科研数据和学术文献也给学术交流和学术出版带来了新的挑战^[2, 3]。

2.2 科技期刊的语义出版

基于语义技术及其他相关信息技术，通过语义标记丰富期刊文章表现形式及显性内容，提高文章信息可操作性和交互性，增强文章关联度，改进出版流程，实现智能化出版的方式可以称为“语义出版”^[4]。语义出版是在传统互联网技术和语义网技术基础上，结合自然语言处理、本体和信息可视化技术发展起来的一种新兴出版形态。它的目的是实现数据库中的数据、图片、结论、参考文献等可以很简单地被用户识别、提取、加工、再整合，最终达到信息再利用的目的。相比数字出版，语义出版在出版内容、形式和服务上具有以下 5 个特征：1) 内容结构化，2) 数据混合化，3) 信息可视化，4) 对象关联化，5) 阅读个性化^[1, 5]。应该说，语义出版是数字出版的更高级形态。

2.3 语义出版的发展构架

语义出版的发展构架应该分为 4 个层次：1) 数据层 (data)：是语义出版系统构架的基础。除了用于存储传统的单篇文献，还分解、转化和存储大量具有完整知识表达和二次利用潜能的语义单元。如原始数据、图片、表格、摘要、关键词、结论、参考文献等。2) 业

务层(business)：即相关的数据库或网络运营商。他们将实现更精细的语义单元的划分，并将相关语义单元进行整合，实现各个资源的自动关联。3) 交互层(interaction)：是语义出版系统的前台展示及用户交互及界面。主要用于指导用户发现、阅读和使用站内资源。4) 用户层(user)：主要用于识别用户的需求、构建用户兴趣网络和提供个性化服务，从而提高信息的再利用率和绩效^[4]。

2.4 科技期刊语义出版的实施形式

科技期刊是报道相关研究领域前沿或综述的载体。从科学的角度看，理想的出版应该实现自动、无缝、快速、广泛的信息获取。从传播的角度看，理想的出版应该实现个性化和智能化。针对以上目标，语义出版分别从发展构架的四个层次进行了语义增强，具体的实施形式包括：1) 识别文章本体，生成原始文档，2) 识别文章实体，建立外部链接，3) 识别文章主题，勾勒文章结构，4) 归类原始数据，方便二次利用，5) 开放程序接口，开展信息混合，6) 了解读者个性，开展推荐阅读，7) 借助浏览软件，进行语义增强，8) 借助专用终端，进行智能订阅^[4,6]。

虽然语义出版的研究应用才开始起步，在技术运用、表现方式和运营模式上还不完善，有待在长期实践中做进一步的改进，能否开创出版行业的新时代还有待检验。但是语义出版必将深入改变信息发布的方式和信息获取的途径，是解决数据超载和迈向语义世界的必然之选。

参考文献

- [1] 王晓光, 陈孝禹. 语义出版: 数字时代科学交流系统新模型[J]. 出版科学, 2012, 20(4): 81-86.
- [2] 袁立凡, 宋淑云, 王薇. 论数字出版与传统出版[J]. 中国传媒科技, 2012(20): 142.
- [3] 张晓林. 颠覆数字图书馆的大趋势[J]. 中国图书馆学报, 2011(37): 4-11.
- [4] Shotton D, Portwin K, Klyne G, et al. Adventures in semantic publishing: exemplar semantic enhancements of a research article[J]. PLoS Comput Biol, 2009, 5(4):e1000361.
- [5] 王晓光, 陈孝禹. 语义出版的概念与形式[J]. 出版发行研究, 2011(11): 54-58.
- [6] de Waard A. From Proteins to Fairytales: Directions in Semantic Publishing[J]. IEEE Intelligent Systems, 2010, 25(2).