Journalism and Communications 新闻传播科学, 2024, 12(6), 1559-1564
Published Online December 2024 in Hans. https://www.hanspub.org/journal/jc
https://doi.org/10.12677/jc.2024.126237

语义出版的概念及其应用简析

杨鑫慧

北京印刷学院出版学院，北京

收稿日期：2024年10月22日；录用日期：2024年11月29日；发布日期：2024年12月9日

摘　要

随着我国现代信息技术和出版实践活动的不断发展，出版领域与现代电子信息技术也开始融合发展，由此，数字出版被广泛使用。数字化出版是以传统出版业为基础的一种新形式。在数字技术广泛应用的背景下，人类信息生产与存储能力不断提升。作为数字出版的高级形态，语义出版伴随着语义技术的发展首先在STM出版领域浮现，并逐渐向大众出版领域扩散。本文在简述语义出版的概念、特点等方面的同时，介绍了语义出版所需的几项关键技术，通过对国内外语义出版实践应用现状的分析，对我国未来语义出版的发展进行了展望。

关键词

语义出版，语义技术，出版平台，数字出版

A Brief Analysis of the Concept and Application of Semantic Publishing

Xinhui Yang

School of Publishing, Beijing Institute of Graphic Communication, Beijing

Received: Oct. 22nd, 2024; accepted: Nov. 29th, 2024; published: Dec. 9th, 2024

Abstract

With the continuous development of modern information technology and publishing practices in China, the publishing sector has begun to integrate with modern electronic information technology, leading to the widespread use of digital publishing. Digital publishing is a new form built upon traditional publishing. In the context of the extensive application of digital technology, humanity's capacity for information production and storage has significantly improved. As an advanced form of digital publishing, semantic publishing emerged in the STM (Science, Technology, and Medicine) publishing field alongside the development of semantic technologies, gradually expanding into the

文章引用: 杨鑫慧. 语义出版的概念及其应用简析[J]. 新闻传播科学, 2024, 12(6): 1559-1564.
DOI: 10.12677/jc.2024.126237

general publishing sector. This paper briefly discusses the concept and characteristics of semantic publishing, introduces several key technologies required for semantic publishing, and analyzes the current state of semantic publishing practices both domestically and internationally, providing an outlook on the future development of semantic publishing in China.

Keywords

Semantic Publishing, Semantic Technology, Publishing Platform, Digital Publishing

1. 语义出版概述

(一) 概念

(1) 来源

语义出版产生于语义网，语义网的使用。语义网是一种以知识为基础的网络。语义网(Semantic Web)是在 1998 年由蒂姆·伯纳斯·李提出的一个概念，并在 2001 年提出了自己的理论框架，并以此作为新一代互联网发展的趋势。目前语义网已开始进入实际应用阶段。其核心是：通过在万维网中对文档增加能为计算机理解的语义，由此使得整个互联网变成了信息交换的通用媒介。通过采用规范、标语及有关处理工具及技巧，语义网已经成为一种更有效的、更广泛的、能够自动收集和收集信息的文件[1]。语义网是一个方便快捷、自由度高的网站，它能提供很高的可信度，从而帮助我们筛选出那些我们工作生活中所需要的有效信息和内容。

(2) 概念

语义出版这一概念最早由 David Shotton 于 2009 年提出。此后许多学者对这一领域进行研究，主要集中于对该理论的应用和实践方面。在他们看来，语义出版就是能够增强期刊中文章语义的出版，促使内容能全部自动化获取，并且让其能够链接到语义相关文章中，同时，提供了文章内部数据的获取、资料的可行性方式，让论文间的数据整合变得更方便的一种出版形式[2]。

语义出版是一种基于语义网技术和多种网络信息服务标准的新型出版形态，语义出版是一种以传统出版物资源、媒体资源、在线开放资源为基础的一种新的出版形式，通过不同的语义工具和数据挖掘方法，对文字、图片、视频等数字内容资源所蕴含的知识单元、实体、参数等内容进行语义化抽取和标注，从而探索科学知识的语义关系和知识网络。

(二) 特点

(1) 内容结构化

语义出版物能够揭示出版物中的概念、命题、论证等显性和隐性信息知识，形成知识图谱或本体，提高信息的组织和检索效率。语义出版中的内容结构化就是通过本体及标注技术，把出版物内容分解成有语义的单位，增加机器的可读性。这便于机器代理对知识单元的操作，达到较高层次知识服务。语义出版内容的结构化要求利用 XML 和其他技术来生产数字资源，以及使用机器学习等技术，实现语义自动识别与标引。

(2) 数据关联化

语义出版中的数据关联化就是使用语义技术，把出版物内容同其他有关数据资源联系起来，构成知

识网络，从而实现跨领域、跨媒体、跨平台的信息整合和知识发现。语义出版数据关联化可以为科学研究和学术交流提供更高效、更深入、更广泛的支持。这就增加了出版物内容可获得性、可操作性强，可组织性强。语义出版物能够将出版物分解为更细粒度的知识单元，如纳米出版物，实现知识的重组和再利用。

(3) 阅读多样化

语义出版物将与出版物相关的实体信息，如作者、机构、期刊、引文等与该出版物互联，构建知识网络，增强阅读交互体验。读者可根据兴趣，按需阅读，选取了不同的层面和领域、从不同视角阅读出版物。可借助数字阅读平台，出版物的个性化操作，比如增加语义标注，评论，共享。语义技术可被使用，对于出版物上的观念，主张、论证及其它知识的深度挖掘与关联分析，发现新知，新视野。

2. 语义出版的关键技术

语义出版技术最早可追溯到 20 世纪 90 年代，那时候因为数据存储与处理技术的发展，人们开始思考怎样利用数据表达含义。之后是 2000 年初期，有研究人员开始探讨如何把这一表达方式运用到文本及其他文档当中，继而构成语义出版技术。语义出版技术，其核心就是把文本，文档等信息转换成机器可读形式，并能通过计算机来理解，以达到效率更高的目的、更精确地信息检索、内容分析及决策支持。所以语义出版技术既能帮助企业对其数据进行更好的管理，也有助于提高以消费者为导向的商品与服务[3]。

语义出版技术应用领域广，上至新闻出版，下至医疗保健，由金融服务向政府机构转变，均可借助语义出版技术提高数据管理与分析能力。例如新闻出版可采用语义出版技术提高文章内容分析与检索水平，从而较好地适应了读者需要。医疗保健可采用语义出版技术提高病历管理的质量，从而为医生诊断及治疗决策提供了较好的支持。金融服务可借助语义出版技术提高风险分析与决策支持的效率，从而较好地满足顾客的要求。政府机构还可借助语义出版技术，完善政策的制定与实施，才能更好地为广大人民群众服务。

(一) 本体构造技术

随着信息网络技术与出版业的加速发展和融合，传统出版与数字出版共同发展是时代导向。在语义网环境下，可以使用本体技术对数字出版领域各个概念间的关系进行描述和阐释，从而对概念体系的形式化归纳，能够有效表达知识，知识查询以及不同领域知识语义消解，还可支持更多服务发现，配套搭配，提高了自动化程度[4]。

领域本体就是指某一学科领域中的本体，只有以建构领域本体为前提，为了从根本上促进知识组织质量的提高，更好的建设为不同用户服务的平台、不同应用场景下知识产品。

(二) 语义标注技术

语义标注指的是对内容资源及各部分内容添加语义信息，由此确定类、属性和示例，这样计算机就可以了解网络资源的信息。目前国内外已有许多学者对语义标注进行研究。不同学科领域中语义标注对象的表现形式截然不同，不仅仅是结构化数据集，还有非结构化图像，音视频等等。

语义标注技术要求有知识库或者本体的支持，本体为根基，语义标注利用领域本体作为语义工具，把文档内容作为标注对象，从所述文档内容确定与所述本体相匹配的类，属性、实例和其他要素在文档内容和本体之间建立关联关系，从而给文档增加了清晰的语义信息。

(三) 语义推荐技术

语义推荐技术能够对资源供需双方进行准确，高效的匹配，减少个体间碎片化交易成本，目前国内外的语义推荐技术主要有：基于语义的内容推荐技术、基于语义的协同过滤推荐技术、基于语义的混合

推荐技术。

基于语义内容推荐技术是一种使用语义网络技术对用户爱好进行推荐的技术、对物品描述信息进行语义化，通过对用户兴趣偏好和商品以及商品之间语义关系进行推理，计算其语义相似度，然后生成推荐结果。2012 年人民出版社推出“人民金典”语义查询系统，提供了基于语义相似度的相关词推荐服务，资源覆盖《毛泽东选集》等、《邓小平文选》及其他系列重要谈话和文章。

(四) 信息可视化技术

信息可视化技术为进行知识组织和知识服务提供了一种辅助工具，这种技术是通过把网络数据、时序数据等用图像的形式进行输出，采用用户视觉直观获取，易于发现数据中固有隐性规律。当前，信息可视化技术多用于文本数据、时空数据、图、网络和多维数据的处理，具体数据需求及呈现形式见下表 1。

Table 1. Requirements and presentation of information visualization technology
表 1. 信息可视化技术的需求与呈现

类型	简要介绍
文本数据	采用文本信息作为数据源，代表性的技术是标签云，对标签按照特定规则排序布局，通过对图形大小进行控制、颜色等，使文本信息可视化显示。
时空数据	时空数据主要关注对时间和地理空间及其关联信息对象属性构建可视化表征，发掘与论证其潜藏的规律与规律，可视化结果，例如流式地图。
网络	网络关联关系可视化呈现，包括互联网，社交网络等、层次结构的数据，网络可视化经典的表现形式就是以节点和边为单位进行可视化，例如树图技术。
多维数据	多维数据就是带有多个维度性质的数据，就是为了探讨多维数据项分布规律及规律、揭示各维度属性之间隐含关系，具有代表性的有平行坐标，投影等。

在传统出版的基础上，从数字出版到语义出版，在转型升级中，出版内容资源逐步实现网络化、数字化与结构化，它蕴涵着大量的文字、网络与多维数据等，以发掘出版资源所蕴含的规律与特点，在检索结果可视化中应用了以上信息可视化技术、将分析结果可视化等等，以信息可视化为手段，不但能直观地反映出资料的真实状况，也能发掘出隐含在数据之间更有价值知识内容。

3. 语义出版的应用实践

(一) 国外语义出版的应用

(1) 维基百科

2007 年起，维基百科从英文、德文等语言版本的维基百科“infobox”及“页面分类”中提取数据，借由 RDF 技术创建可供查询的语义网，并提供了应用接口，可以让人们自由访问其中数据。

(2) BBC (英国广播公司)

BBC Dynamic Semantic Publishing 是一种使用语义技术的媒体模型，由 BBC 的 John Donovan 和 Jem Rayfield 领导的团队开发。它可以实现动态地生成和发布基于主题的网页，而不需要手动编辑[5]。应用场景主要是在媒体领域，例如新闻、体育和娱乐。它可以帮助编辑和记者快速地创建和更新基于主题的网页，提高内容的重用、搜索引擎排名和用户体验。它还可以根据用户的兴趣和行为提供个性化的推荐和发现服务。

(3) 爱思唯尔

爱思唯尔是国际多媒体出版公司，为医学、教育、科技等领域提供数以万计的产品。它也是全球领

先的开放获取出版商之一，旗下拥有超过 500 种开放获取期刊[6]。爱思唯尔有许多与语义出版、语义技术相关的产品：医学信息平台 ClinicalKey 由爱思唯尔提出，为临床决策提供支持。该平台集可视化分析和推理于一体，有助于医生迅速得到精确，简明、国际前沿循证医学知识，从而迅速解决了临床问题、提高了诊疗水平。基于语义技术的工具 Elsevier Journal Finder，能够帮助作者找到最适合他们文章主题和范围的期刊。

(4) 英国皇家化学学会

英国皇家化学学会的语义出版宣言为“见证科学的生命力”(See Science Come Alive)，即通过集成 RSC 期刊论文，建构结构化科学，它的核心内容包括化学学科的知识环境(本体库、数据库)之间的关联链接，自动识别化学术语，生物医学术语及其他内部元素、分类与标记及全文嵌入式 HTML 展示模式[7]。用户只需单击化学名称，就可直接与数据库原始记录连接，以便找到更多的关联数据与参考资料。

(5) 哈佛大学干细胞研究所

哈佛大学干细胞研究所创办的 *Stem-Book*，是一本以干细胞生物学为主题的开放获取期刊，除了为研究者提供以 Web2.0 为标志的在线讨论功能外，还适用于科学协作框架，使用 Web3.0 技术(语义网络、文本挖掘等)，允许在线科学出版物或科学数据的交互，从而实现了通过语义共享来进行数据、知识和信息等方面的交流。同时还可以通过整合生物医学数据及科学文献资源，建立 UniProt 知识库及基因表达图谱，对糖尿病患者基因与疾病关联性进行计算与分析[8]。

(二) 国内语义出版的应用

在国家新闻出版署促进产业升级的背景下，语义出版的经济效益和社会效益日益显现，已经有一些出版机构首先进行了语义出版的实验。目前国内还没有一套成熟的语义出版流程和标准，因此各出版单位都在探索适合自身特点的发展路径。汕头出版社把重点放在发展语义编辑技术上，实现了数字信息的自动化收集，整理，归类等功能，使得内容提供商提供的信息内容可以更加快捷，准确、以结构化方式展现在用户面前，由此进一步提升了用户体验效果[9]。

《中国社会科学》杂志社打造《中国社会科学》数字图书馆，使用本体驱动方法，建设一个覆盖社会科学领域所有子领域本体库，并把期刊文献和其他类型的文献联系起来[10]。同时还通过对学术期刊资源和相关数据库进行语义标注和检索，实现了从期刊到图书之间的知识链接和智能搜索服务。《中国国家地理》杂志社开发了《中国国家地理》数码阅读器，采用自然语言处理技术，给用户带来图文并茂，声音互动的效果、地图导航和其他功能。

4. 我国语义出版的未来发展

语义出版就是采用语义技术，绘制知识图谱，将出版物内容结构化，标准化、智能化加工，达到出版物高效生产、精准传播，深入应用。语义出版在出版行业中具有创新模式，也是数字出版发展的一个重要方向[11]。当前，国内已出现一些有代表性的语义出版项目及平台，例如，《中国知网》《百科全书》平台均使用语义技术，以提升信息检索与知识服务效率。语义出版具有可以使用语义标注和本体构建等优点、推理机制和其他办法，使出版内容由文本形式变为结构化，标准化数据，从而达到深入了解内容，精准匹配。

与此同时，语义出版也面临一定的挑战，比如，怎样确保语义标注准确一致、如何应对多样数据融合问题、如何平衡知识表达的复杂度与可操作性等。在今后的发展中，语义出版会不断地和人工智能、大数据、云计算等其他技术结合，实现较高水平智能出版服务，满足了用户多样化、个性化需求[12]。语义出版是新时期中国出版业的一个主要发展方向。发展语义出版，要求我们在观念和技术上不断地创新，优化内容和服务等，提高质量和利益，为社会主义文化强国建设做出更大贡献。

参考文献

[1] Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The Semantic Web. *Scientific American*, **284**, 34-43. https://doi.org/10.1038/scientificamerican0501-34

[2] Shotton, D., Portwin, K., Klyne, G. and Miles, A. (2009) Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLOS Computational Biology*, **5**, e1000361. https://doi.org/10.1371/journal.pcbi.1000361

[3] 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): 56-60.

[4] 李芳, 纪姗姗, 周毅, 等. 语义出版本体模型的构成分析——以SPAR本体和Nature本体为例[J]. 图书情报工作, 2017, 61(24): 123-134.

[5] 王莉莉, 栾冠楠. 英国广播公司(BBC)动态语义出版模式研究[J]. 图书情报工作, 2017, 61(8): 126-132.

[6] 翁彦琴, 彭希珺. 爱思唯尔(Elsevier)语义出版模式研究[J]. 中国科技期刊研究, 2014, 25(10): 1256-1261.

[7] 翁彦琴, 李苑, 彭希珺. 英国皇家化学会(RSC)——科技期刊语义出版模式的研究[J]. 中国科技期刊研究, 2013, 24(5): 825-829.

[8] 苏静. 基于领域知识库的语义出版形态研究[J]. 出版科学, 2020, 28(5): 91-99.

[9] 李楠, 孙济庆, 马卓. 面向学术文献的语义出版技术研究[J]. 出版科学, 2015, 23(6): 85-92.

[10] 苏静, 曾建勋. 国内外语义出版理论研究述评[J]. 中国科技期刊研究, 2017, 28(1): 33-38.

[11] 许鑫, 毛璐, 江燕青. 基于数字资源聚合的学术期刊语义出版研究[J]. 数字图书馆论坛, 2019(7): 2-8.

[12] 谷紫藤. 数字出版的高级形式——语义出版领域文献的定量分析[J]. 商情, 2017(23): 271.