

国外多语言公共数字文化服务平台的特点与思考

刘 莉 邵 波

【摘 要】[目的/意义]对国外多语言公共数字文化服务平台进行调研与分析,探寻适合我国多语言公共数字文化服务平台的构建策略。[方法/过程]在国内外已有相关平台调研的基础上,选取国外7个公共数字文化服务平台作为研究案例,利用5S模型构建分析框架:社会-资源-服务-技术,并在此框架下进行多案例的对比与研究分析。[结论/结果]提出联合政府建立联盟性质的多语言公共数字文化服务平台,借助丰富的多语言资源提供多模态的多语言服务,以及摆脱技术应用的固定思维抓住新技术带来的机遇的建议。

【关键词】多语言服务;公共数字文化服务平台;5S模型

【作者简介】刘莉,南京大学信息管理学院博士研究生;邵波(通信作者),南京大学信息管理学院副馆长,南京大学图书馆教授,博士,博士生导师,E-mail:shao@nju.edu.cn(南京 210023)。

【原文出处】《图书情报工作》(京),2023.21.124~135

【基金项目】本文系南京大学技术服务项目“智慧图书馆建设与测评”(项目编号:2021201204)研究成果之一。

1 引言

公共数字文化建设是实现加快构建现代公共文化服务体系的重要举措,也是首要任务。我国从“十二五”期间即开始大力推进公共数字文化建设,数字图书馆推广工程以及全国文化信息资源共享工程等重点公共数字文化工程相继实施,取得了显著的成效。但是,仍存在服务不完善、服务效能不高等众多问题。“十三五”期间,中共中央办公厅、国务院办公厅印发《关于加快构建现代公共文化服务体系的意见》,提出推进国家公共文化数字支撑平台建设,汇聚分散在图书馆、博物馆、档案馆(以下简称“三馆”)等文化机构的资源,提供“一站式”服务^[1]。但随着社会经济的发展以及公共需求的多层次化,对公共数字文化建设提出了更高的要求^[2]。2021年作为“十四五”的开局之年,文旅部即发布《关于推动公共文化服务高质量发展的意见》以及《“十四五”公共文化服务体系规划建设规划》,提出要聚焦网络化、数字化、智能化的公共数字文化服务平台建设并拓展公共文化服务^[3-4]。随后,2022年中共中央办公厅、国务院办公厅

印发《关于推进实施国家文化数字化战略的意见》,明确到“十四五”末期,基本建成文化数字化基础设施和服务平台,融合多领域数据信息,形成线上线下一体化^[5]。因此,公共数字文化服务平台建设不仅是时代发展的必然之路,也是国家政策导向。

现有的相关研究普遍认为公共数字文化服务平台是以技术为支撑^[6-8]、将三馆资源或者其他文化机构资源进行“一站式”整合的开放式平台。结合已出台的公共数字文化服务平台建设的重要政策文本,政策与研究认知之间的共性是公共数字文化服务平台具备三馆属性,同时又具备数字化、一体化、普及化以及开放共享化等特征。公共数字文化服务平台涵盖数字图书馆,且与传统单一的数字文化机构(数字图书馆)相比,公共数字文化服务平台最突出的特点是一站式,即通过统一入口获取多源异构的信息资源^[7]。并且,随着技术发展的进步,众多的公共数字文化服务平台均以数字图书馆的形式展现。

在“一带一路”倡议,开放合作、共建共享等理念的影响下,我国渐渐开始意识到要构建一个多元文

化融合的综合体平台。文旅部印发《“十四五”“一带一路”文化和旅游发展行动计划》^[9],提出关注全平台,推动中国文化“走出去”,全面吸收国外文化。国际图联也在2022年的《公共图书馆宣言》指出,图书馆必须要满足多语种用户需求,帮助用户跨越语言障碍,获取、阅读与理解全球多语种信息,提供多语言服务迫在眉睫^[10]。目前,我国较为典型的多语言公共数字文化服务平台主要体现于国家数字文化资源共享服务平台、国家数字图书馆、丝路图书馆、中国民族数字图书馆、国家文化云、文化e家和国家数字文化网等国家型公共数字文化服务平台^[11]。通过调研发现,我国现有的多语言公共数字文化服务平台具有中国发展特色以及地域特色,服务对象更内向化。存在多语言门户网站缺乏多样性,信息语种资源缺乏普及性,缺乏国际影响力等问题,是在经济文化需求带动下的被动产物。而与此同时,国外在多语言公共数字文化服务平台上却拥有更加灵敏的嗅觉。在联合国教科文组织(United Nations Educational, Scientific and Cultural Organization, UNESCO)以及国际图书馆协会联合会(International Federation of Library Associations and Institutions, IFLA)等组织的共同推动下,国外多语言公共数字文化服务平台的建设、运行效果以及国际影响力要优于国内平台。拥有一批如世界数字图书馆、欧洲数字图书馆等具有影响力的跨国合作的多语言公共数字文化服务平台,以及被反复参考研究的美国数字公共图书馆、澳大利亚国家图书馆等国家级多语言平台。上述平台在多语言的资源整合、信息存取以及平台建设等方面都具有一定的经验,并且成效显著。

日前,我国中国工程院国际合作局联合产学研界众多专家学者为服务“一带一路”相关建设构建的“丝路”多语言机器翻译平台,从技术角度驱动并深化了我国多语言公共数字文化平台多语言服务的开展^[12]。但是,通过对上述国内多语言平台进行线上调研发现,目前我国总体的公共数字文化服务平台的多语言服务水平还略显稚嫩。因此,本文聚焦国外具有影响力的多语言公共数字文化服务平台,针对多语言,从多维度探究其平台建设特点,以期为我国多语言公共数字文化服务平台提供参考意见。

2 相关研究回顾

多语言公共数字文化服务平台是在公共数字文化服务平台的基础上提供多语言服务。受到国家一系列相关政策文本的支持,我国对公共数字文化服务平台的建设一直都十分重视。具体研究方向包括平台建设现状、各要素驱动下的平台架构构建、平台标准指标构建以及平台资源整合问题等。华芳园^[8]首先从上层视角,对我国相关公共数字文化服务平台的建设情况进行分析,认为在虽然我国公共数字文化服务平台数量正在不断增加,但是相对于国外发达国家,我国的平台建设较晚,还有较大的上升空间。王丹、张树臣、完颜邓邓^[13-15]等从区块链以及大数据技术等角度,探讨云平台构建的可能性。孙国焯、陈则谦、朱莉^[16-19]等则探讨公共数字文化平台的相关标准指标构建从而驱动平台标准化的构建。并提出相应优化策略促进平台建设。吴丹、韦楠华^[20-21]等则对平台面临的多源异构资源整合问题进行研究,并提出一系列解决对策。

除此之外,吴丹、孙国焯^[16,20,22]等则在此基础上调研分析公共数字文化服务平台的多语言服务需求。吴丹首先对数字图书馆用户的多语言需求进行调研,研究结果证明了多语言服务对于用户的重要性。随后,其从顶层设计角度,提出构建标准规范以及资源获取制度保障图书馆的多语言服务。孙国焯则从国家文化政策以及国家发展的角度出发,分析肯定多语言公共数字文化服务平台的必要性。

上述两个方向的研究为多语言公共数字文化服务平台的研究提供了研究基础。国内外对于多语言公共数字文化服务平台的研究主要可以分为三个层面:①案例研究。曹玲、杜慧平^[23-24]等分别围绕多语言资源建设资源利用对欧洲数字图书馆(Europeana)以及国际儿童数字图书馆(International Children's Digital Library, ICDL)的进展情况进行跟踪研究,从而启发我国多语言数字图书馆的建设。张彦文^[25]利用文献调研手段,对学术界多语言数字图书馆的研究进行概括性论述。A. Wu^[26]则对世界数字图书馆和加勒比数字图书馆进行对比研究,提出了一个支持多语言的公共数字文化服务平台框架。T. Budzise-

Weaver^[27]拉丁美洲开发档案馆以及美国古腾堡计划等四个图书馆作为调研对象,调研其多语言的馆藏资源以及多语言服务,认为多语言数字图书馆可以利用优势新技术开展提供更好的多语言服务。②多语言检索功能分析与探索。司莉^[28-30]针对用户多语言检索需求、多语言用户检索行为以及多语言检索平台功能进行分析,肯定多语言检索的重要性,并提出完善平台检索功能的建议。H. M. Alsalmi^[31]研究公共数字文化平台多语言检索时用户的交互和行为,认为需要进行用户研究,以记录信息检索系统能在多大程度上满足多语言用户的需求和期望。J. Mizera-Pietraszko^[32]为了提高公共数字文化服务平台的多语言检索效率,提出一个基于粗糙集理论的模型,限制制约因素的数量,从而提高检索准确性。③多语言技术研究。刘莉^[33-35]提出利用机器翻译技术赋能图书馆多语言信息服务,并构建以机器翻译为主导的多语言自动翻译平台整合服务方式。国外B. Kituku^[36]也认为机器翻译相较于其他技术方法是更合适多语言服务的方法,其更具性能优势。章成志^[37]研究多语言领域本体的潜在价值,探讨本体在多语言数字图书馆中的应用前景。M. Mozgovoy^[38]则提出在词典中添加本体,提高服务效能。A. Vasiljevs^[39]提出创建欧洲在线多语言服务平台,以满足数字单一市场的多语言需求。并将平台分为三层:解决方案层、基础设施层和研究层。基础设施层将在自动翻译云、人机交互云、多语言知识管理云、欧洲语言云四个云中结合成熟的语言技术,包括基础服务和语言资源。

综上所述,国内外对于多语言公共数字文化服务平台均进行了一系列研究,确定了多语言服务对于公共数字文化服务平台的重要性。分析国外优势平台的建设情况是目前国内外研究的重点内容之一,也是一众学者达成的共识。但是,现有的相关案例研究还存在以下几点问题:①研究案例较少。上述相关案例研究选择的案例维持在2-4个左右,不是多案例研究的适宜数量。②研究维度单一。已有案例研究单一的多语言服务,例如跨语言信息检索、多语言元数据描述,缺乏对多语言平台的全面性研究。③研究案例重合度较高。现有的研究案例集中

于欧洲数字图书馆、世界数字图书馆,缺乏在基础上进行拓展延伸的案例研究。本文将规避现有研究的不足,在已有研究的基础上,拓宽选择范围,选择多个国外成熟的多语言公共数字文化服务平台进行全面的探索,探索挖掘具有优势的多语言服务以及建设方法,从而为我国的多语言公共数字文化服务平台建设提供参考。

3 国外多语言公共数字文化服务平台的案例选择

有研究表明,多案例研究一般保持在6-10个是为最适宜的数量^[40]。而国外多语言公共数字文化服务平台众多,包括各国家或国际合作建设的多语言数字文化服务平台。因此,本文汇聚现有相关研究中重点关注、影响力较大的几大国外多语言公共数字文化服务平台,并在此基础上,选取具有特色的研究较少的多语言平台,探究其特点。最终,本文筛选出国外7个多语言公共数字文化服务平台作为研究对象,具体的多语言数字文化平台选择依据为:①平台在多语言服务领域较为成熟,相关优势服务或者举措能够为我国的多语言公共数字文化平台建设提供参考。②平台要涵盖国家级、全球级层次平台,能够代表国际上顶层的多语言公共数字文化服务平台建设情况。③平台的各方面发展比较完善,能够进行多方面的横向比较。根据该选择依据,本文选取欧洲数字图书馆(Europeana)、世界数字图书馆(The World Digital Library, WDL)、国际儿童数字图书馆(International Children's Digital Library, ICDL)、美国数字公共图书馆(Digital Public Library of America, DPLA)、澳大利亚国家图书馆(National Library of Australia, NLA)、加勒比海数字图书馆(The Digital Library of the Caribbean, dLOC)以及日本国立国会图书馆(National Diet Library, NDL)7个多语言数字文化服务平台作为研究对象。WDL、ICDL、Europeana,以及dLOC是国际上起步较早的跨国的多语言公共数字文化服务平台,其多语言资源馆藏相对丰富,多语言服务也较为成熟。DPLA作为一家整合了全美“三馆”资源的全数字化图书馆,是数字图书馆浪潮中的典型代表。NLA的馆藏与服务特色跻身于世界级国家图书馆的前列^[41]。NDL也在世界性的数字资源建设与服务中占据领先优势^[42]。上述7个多语言公共数字文化服

务平台的概况如表1所示。

4 多语言公共数字文化服务平台的分析框架构建

信息网络的泛在延伸、智能技术的普及应用以及数据资源的蓬勃发展给多语言公共数字文化服务平台的运行带来了更多的挑战与机遇。为了更加深入地了解多语言公共数字文化服务平台,需要构建统一的分析框架,系统、正规、全面的厘清并分析平台运行与服务现状。M. A. Goncalves^[50]等研究者提出的5S模型是数字图书馆领域普遍接受的基础理论,美国的NDLTD项目应用5S模型对数字图书馆进行分析,取得了较好的结果,证明了该模型的适用性^[51]。本文以5S模型作为理论指导,并针对多语言服务,提取、聚合核心模块构建公共数字文化服务平台的分析框架,针对性的分析平台的多语言服务特征与优势。

4.1 5S模型

5S模型认为数字图书馆由数据流(streams)、结构(structures)、空间(spaces)、情境(scenarios)以及社会(societies)五大要素组成。数据流指向图书馆的所有信息对象,包括文本以及数字音频等信息,结构则是信息对象以及所有实体的一种表征形式,数据与结构从某种程度上构成了一个完整的信息资源集合体。空间是信息与社会发生操作和关系的形式化的要素,包括作用于其他要素的服务。而情境则是在空间下,其他要素发生的变化,表征着数字图书馆功

能。社会是意指数字图书馆各个流程的人、物以及软文化等,是图书馆的最高层次表示。数字图书馆作为一个公共数字文化服务平台,为用户提供服务是其核心目标,五要素的相互协作保障了数字图书馆功能的完善度。其具体结构关系模型如图1所示。

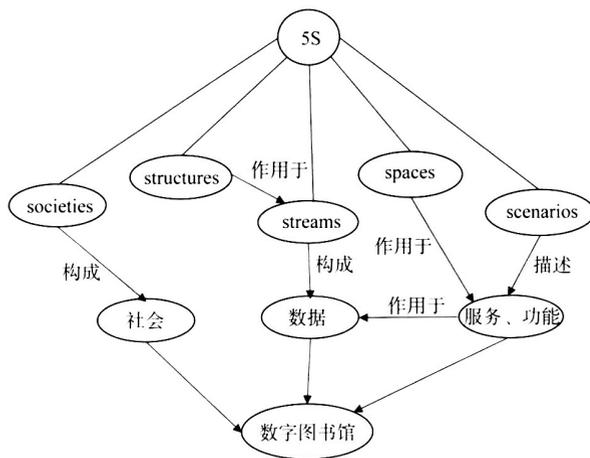


图1 5S模型^[52]

4.2 平台分析框架

5S模型的五要素完整的描述了数字图书馆的构成,维护着图书馆的运行。但随着图书馆智慧化的转型、多元文化的渗透以及人工智能技术的飞速发展,图书馆要素的内涵更加深层且丰富。本文在5S模型的基础上,针对公共数字文化服务平台的多语言服务,对其要素进行整合与延伸,构建社会、资源、服务、技术四大模块,并作为多语言公共数字文化服

表1

国外7个多语言公共数字文化服务平台概况

开始年份	平台名称	发起组织	简介
2009	世界数字图书馆(WDL)	美国国会图书馆以及联合国教科文组织等组织	整合了联合国会员国国家型平台的多种文化材料,服务面向国际公众 ^[43]
2002	国际儿童数字图书馆(ICDL)	马里兰大学和互联网档案馆合作研发,美国国家科学基金赞助	收录了不同语言版本的儿童文学作品,面向全球的儿童提供优秀作品 ^[44]
2008	欧洲数字图书馆(Europeana)	受欧盟委员会委托,欧洲基金会主办	汇聚了欧洲27个成员国多语言、多类型的数据资源,并用数十种语言对资源进行整合,是公众了解欧洲文化的窗口 ^[45]
2008	加勒比海数字图书馆(dLOC)	海底、牙买加等36家单位联合创建	拥有多种语言的有关于加勒比海文化的馆藏资料 ^[46]
2013	美国数字公共图书馆(DPLA)	美国国会	聚合了美国公立与私立文化机构的多语种数据资源,致力于让公众平等自由的获取知识 ^[47]
1901	澳大利亚国家图书馆(NLA)	澳大利亚国家图书馆	对全国图书馆进行调控,还提供其他国家的联合目录以及服务入口,实现图书馆之间的资源整合与共享,传播民族文化 ^[48]
1947	日本国立国会图书馆(NDL)	日本国立国会图书馆	以日英中韩四大语言为主,为用户提供跨语言检索服务的平台 ^[49]

务平台分析框架,如图2所示。

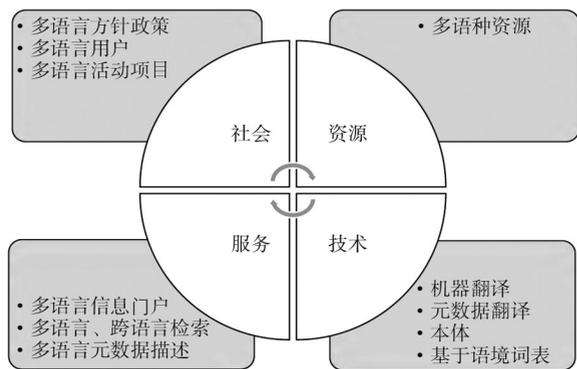


图2 多语言公共数字文化服务平台分析框架

(1)社会模块。社会模块指代社会(societies)要素,由多语言方针政策、多语言用户以及多语言活动项目组成。该模块对多语言公共文化服务平台起到一个顶层设计的作用,指导平台运行,决定平台的未来发展走向。

(2)资源模块。资源模块包含数据流(streams)以及结构(structures)要素,针对性探讨多语言公共数字文化服务平台为用户提供的多语种数据资源,包括网络数据资源、图书馆数据库资源以及第三方数据商资源等数据资源集合。丰富的多语种馆藏资源是利用资源为用户提供多语言服务的前提。

(3)技术模块。技术是支撑服务与功能的一种手段,公共数字文化服务平台的多功能的实现需要采用一定的技术方法。各大多语言公共数字文化服务

平台利用的技术不一,例如机器翻译、元数据翻译等,技术支撑着多语言公共数字文化服务平台的运行,技术的优劣决定着平台多语言服务的性能,其中优势技术具有重要的参考价值。

(4)服务模块。服务模块包含空间(spaces)以及情境(scenarios)要素。一是在平台整个空间里提供的服务;二是包括为提供服务涉及的操作。多语言服务包括多语言信息门户、跨语言检索、多语言检索以及多语言元数据描述。服务模块是多语言公共数字文化服务平台的核心模块,也是重点研究模块。对各个平台提供的多语言服务进行调研与评估,有利于了解其所提供服务的多样性、优劣势以及对资源的利用度,继而对我国多语言数字文化服务平台的建设提供参考。

5 多语言公共数字文化服务平台的比较研究

本研究按照上文提供的分析框架对平台进行系统、详细的解析,挖掘优势服务与技术,指导、启发我国多语言公共数字文化服务平台的建设。

5.1 社会模块

根据5S模型 societies 要素衍生而来的社会模块由多语言方针政策、多语言用户以及多语言活动项目组成,多语言公共数字文化服务平台按照方针政策,通过举办活动创建项目等手段为多语言用户提供服务。7个多语言公共数字文化服务平台在政策方针、活动项目以及多语言用户情况如表2所示。

表2 国外7个平台多语言方针政策、活动项目以及用户情况对比

平台名称	多语言方针政策	多语言活动项目	多语言组织联盟
WDL	《公共图书馆宣言》 ^[10]	“WDL 合作伙伴会议和执行委员会会议” ^[53]	面向全球吸纳了近百个国家的机构合作,机构包括图书馆、档案馆以及其他共198个机构
ICDL	——	——	马里兰大学与美国基金会、图书馆、微软、博物馆共同合作组织
Europeana	i2010 数字图书馆倡议 ^[54]	欧盟第五期架构计划 ^[55]	面向欧盟,涉及三十多个国家,合作机构包括档案馆、博物馆、图书馆等3000多个机构
dLOC	《2030年可持续发展议程》 ^[56]	国际倡导计划(International Advocacy Programme, IAP) ^[57]	海地国家档案馆、阿鲁巴国家图书馆、牙买加国家图书馆以及部分大学图书馆合作联盟
DPLA	《美国数字公共图书馆2019-2022年战略规划》 ^[58]	大学数字图书馆国际合作计划(China Academic Digital Associative Library, CADAL) ^[59]	传播美国文化,与美国国家档案和记录管理局、纽约公共图书馆、密歇根服务中心等4000多家文化机构合作
NLA	《超越优质服务,巩固社会结构:澳大利亚公共图书馆服务标准指南》 ^[60]	Trove ^[61]	国家图书馆与766个机构的合作,通过合作联盟实现了Trove项目
NDL	《2021-2025 国立国会图书馆数字化转型战略》 ^[62]	NDLSearch ^[63]	50多个机构,100多个国内外数据库的联合,通过合作联盟构建NDL Search综合门户

5.1.1 多语言方针政策

方针政策是 societies 要素中不可或缺的一部分。是图书馆根据自己发展的需求与特点,综合各种因素制定的符合馆情的愿景目标,展现图书馆的定位,也为未来的项目开展提供指导方向^[64]。国外7个多语言公共数字文化服务平台在多语言方针政策方面的特性主要可分为:在数字化转型中兼顾多语言服务、在国际战略方案中融和多语言方针以及在全国性服务政策中涉及多语言指示。

首先,近年来,数字化转型一直是公共数字文化服务平台的工作重心,而多语言则是在数字化的基础上丰富资源的多样性以及提高信息资源的可理解性。Europeana 以及 NDL 均将多语言服务作为数字化转型的一部分,在数字化建设中保证服务质量。其次,IFLA 以及 UNESCO 作为国际公共文化服务领导组织,一直以来从顶层对全球的公共数字文化服务事业进行把控,指导国际图书馆的发展方向。WDL 以及 dLOC 在全球化政策的指引下,提供多语言服务。NLA 以及 DPLA 在众多国外公共数字文化服务平台中在本土化政策中多语言服务表现亮眼。其中 NLA 为了服务新移民以及多元化社区的构建,针对性的出台了系列多语言方针政策,以此保障各级公共图书馆的语言服务多样化^[60]。除此之外,NLA 也十分重视多元化馆藏建设,鼓励多样化以及个性化的文献资源整合。而 DPLA 作为美国的数字中心,在多语言服务方面有着更高的敏感性以及先进性,在国家的战略方案中将多语言服务作为愿景之一,以此适应不同社区的多元化。

5.1.2 多语言活动项目

在多语言政策方针的引导与支持下各平台的多语言活动项目应运而生,多语言活动项目是 Societies 要素作用下的产物,反映出对政策的响应支持。7个国外多语言公共数字文化平台针对多语言政策提出的项目包括多语言检索项目以及进一步的计划部署。其中,Trove、NDL Search 是 NLA 以及 NDL 较为

成熟且成功的多语言检索项目,通过促成多语言国家的合作联盟,实现多语言资源的聚合,并构建统一的门户网站,为用户提供获取多语言信息的“一站式”渠道。Europeana、WDL、DPLA 以及 dLOC 则进一步从政策中开展计划实施,为多语言服务的战略部署提供更详细的布局,并且在计划中更加注重合作,在合作中倡导多语言、多元化活动。Europeana 提出的欧盟第五期架构计划即包含 14 个多语言文献项目^[55],WDL 以及 DPLA 等通过合作联盟,共同探讨多语言活动的组织实施。

5.1.3 多语言组织联盟

多语言组织联盟是 Societies 要素中组织关系的体现。多语言组织联盟不仅是馆际间的合作,更可以将范围扩展至与各机构的合作联盟,国外多语言公共数字文化服务平台更是将合作联盟作为诸多事务开展的关键一环。7个多语言数字文化服务平台在多语言组织联盟上展现出的趋势为:三馆联合保障基础、政府介入促进合作灵活有效、合作联盟体结构日趋庞大。由调研可知,三馆联合已是趋势,7个平台均将三馆作为联合的基础,并在此基础上进行拓展,是组织联盟的基底。而政府的介入则是从一定程度上给予合作保障,推动项目活动的进行,以及给予相关合作支持。通过对合作组织联盟的构成进行深度分析,可以其发现组织结构较为庞杂。文化机构、公共图书馆以及大学图书馆等的参与从一定程度上丰富了组织构成,并且聚合了更多的资源,丰富了资源多样性,给联盟创造了更多可能性。

5.2 资源模块

资源模块由 streams 要素以及 structures 要素构成,是国外多语言公共数字文化服务平台的内容核心,也是进一步提供多语言服务的基础。平台提供的多语言资源越丰富则代表在多语言服务方面拥有更多的可能性。表3体现了国外7个多语言公共数字文化服务平台资源语种的数量情况。

从表3可以看出,DPLA 提供的资源语种数量最

表3 国外7个多语言公共数字文化服务平台多语种资源对比

	WDL	ICDL	Europeana	dLOC	DPLA	NLA	NDL
资源语种	马来、英、泰、下索布、日等134种语言	英、法、德等多种语言	德、法、丹麦、荷兰、芬兰、中等37种语言	中、英、法、德4种语言	英、德、法、日、意大利、中等400多种语言	英、中、日、澳大利亚、荷兰、西班牙等不少于20种语言	日、英、法、俄、德、西班牙等不少于10种语言

多,高达400多种语言。dLOC提供的资源语种数量最少。而造成二者之间的差异的主要原因则在于地区之间的发达程度不同以及平台组织性质不同,而有时,二者之间又相互影响。美国作为全球经济文化的发达中心,以其最大程度上的开放包容进行信息资源的开放获取。而加勒比海地区则较为落后,在信息获取以及信息传播等方面都还有很大的进步空间,正在通过构建期刊网络的方式加强与外界联系^[65]。其他几大平台则在地方特色、常见语种的基础上,提供基本上能够满足需求的服务。除此之外,Trove、DPLA、Europeana以及WDL还借助Facebook、wiki和twitter让世界各国的人参与图书馆的建设以及实现资源共享,给多语言信息资源挖掘了更多的共享渠道。

5.3 服务模块

服务模块包含spaces要素以及scenarios要素,两要素在服务模块中体现为在多语言公共数字文化服务的平台上提供的场景服务。是多语言公共数字文化服务平台的核心任务模块,其目的在于满足用户需求,提高多语言资源利用率以及促进多元文化的交流等。多语言服务模块一般包含多语言信息门户,跨语言、多语言检索以及多语言元数据描述^[66],7个国外多语言公共数字文化服务平台的多语言服务情况如表4所示。

5.3.1 多语言信息门户

多语言信息门户即多语言的门户界面,是平台对外交流以及面向世界的窗口,也是多语言公共数字文化服务平台的必须具备的基本属性。平台网站根据用户需求定制相应的语言界面,助力用户实现无障碍访问,英语作为世界通用语言,是各大平台门户均提供的语言界面,从而保障平台基本的用户友好性。dLOC、NDL、Europeana以及WDL在此基础上还提供其他语种,丰富门户网站语种呈现的多样

性。其中,Europeana语言种类最多,给用户创造了更加多语的环境。除此之外,7个平台普遍采用一键切换的方式实现多语言门户的转换,给不同母语用户最熟悉的语言环境进行信息的检索以及获取。但是,总体而言,国外7个平台的多语言信息门户依旧缺乏多样性的完善,现如今的多语言门户更注重本土用户的语言习惯以及其所惯用语种的普及度。

5.3.2 多语言、跨语言检索

多语言、跨语言检索是用户获取多语言信息的必经手段,也是平台为用户提供多语言服务的必经渠道。跨语言指用户可以通过输入一种语言信息从不同语言资源中检索获取相关信息^[67]。跨语言检索功能的强弱直接影响用户获取信息的完整性以及信息的利用率。而跨语言检索功能的强弱则可以用检索语种数量、检索信息的相关性以及全面性来衡量。国外7个多语言公共数字文化服务平台在多语言、跨语言检索功能方面的差异受平台联盟体构成、政治文化以及多语言技术等因素影响。例如DPLA作为美国在数字化建设方面的重大项目,多语言、跨语言检索功能也处于相对领先的地位,提供50多种语言的检索,大大提高了平台信息的利用率。而NDL由于其国家文化属性,只提供日语与中韩英三种语言之间的相互检索。从跨语言信息检索的相关性以及全面性来看,目前,检索信息受限于翻译功能的影响,还无法完全将相关信息完整、完全地呈现于用户检索的界面,需要进行更进一步深层的挖掘,需要更加具有技术优势的模型增强跨语言检索的水平。总而言之,国外7个多语言公共数字文化服务平台的跨语言检索水平并未随着整体人工智能技术的进步而有更加深入的突破,仍然具有较大的提升空间。

5.3.3 多语言元数据描述

多语言元数据描述通常是指平台的所有数据都

表4 国外7个平台的多语言信息门户、跨语言检索、多语言元数据描述对比

平台名称	多语言信息门户	跨语言、多语言检索	多语言元数据描述
WDL	提供中、英、俄、法、阿拉伯等7种语言的界面	提供中、英、俄、法、阿拉伯等7种语言的跨语言检索	√
ICDL	提供英语界面	提供英、汉、法、德等18种语言的检索	√
Europeana	提供30种语言的门户界面	具备英、德、意大利、荷兰、法等10种语言的跨语言检索功能	√
dLOC	提供英、西班牙语、法语三种语言界面	提供英语、法语、帕皮亚门托语三种语言的跨语言检索	√
DPLA	提供英语界面	提供英、日等50多种语言的检索	√
NLA	提供英语界面	—	√
NDL	提供英、中、韩、日四种语言界面	支持日韩、日中、日英三种语言的双向翻译与跨语言检索	√

能够用不同的语言来进行表述与介绍,便于用户能够检索与目标信息相关的多语言信息^[66]。早期,受限于技术水平的发展,多语言元数据描述通常被当作衡量多语言服务优劣的一大特征,众多平台将元数据描述作为提供多语言服务使用的一种服务手段。国外7个多语言公共数字文化服务平台也普遍对元数据进行多语言描述,其中,Trove只存储元数据而不存储内容的特性直接决定了其元数据多语言描述的特征^[68]。但是,随着科技的进步,多语言元数据描述将会被替代,由翻译工具完成这一工作环节,通过引用翻译工具即可以帮助用户实现对所有多语言信息进行最直接的获取,提供更加全面、高效且优质的服务。

5.4 技术模块

技术模块是智能时代下衍生的产物。随着人工智能技术的飞速发展,技术在平台建设中逐渐占据主导地位。技术模块针对性研究多语言技术,多语言技术是支撑多语言公共数字文化服务平台提供多语言服务的重要一环,决定着多语言服务的质量与效率。目前,国外7个平台应用的多语言技术可归纳为机器翻译、命名实体识别、元数据翻译等,如表5所示。

从表5可知,国外7个公共数字文化服务平台大多使用两种或两种以上的翻译技术保障多语言服务的质量。并且利用机器翻译、元数据翻译技术实现多语言服务是7个公共数字文化服务平台普遍采用的技术手段。在经历了基于实例—基于规则—基于统计—基于神经网络的4个阶段的发展变化,目前主流的基于神经网络与深度学习的机器翻译技术已经能够在大部分场景下拥有不俗的翻译表现,并且灵活度较高,能够以插件、软件等方式嵌于各种应用载体。但是,受限于机器翻译技术自身固有的局限性,其在低资源以及语言环境较为复杂等场景下性能较弱,需要采用其他辅助翻译技术从一定程度上保障多语言服务的质量。元数据翻译将元数据用不同语言进行表述,帮助用户从海量的多语言数据中获取

相关信息。一般而言,元数据翻译依赖人工对元数据进行标注,耗时长,从长远的角度看,并不符合时代快速发展的需求。除此之外,多语言本体在机器翻译以及跨语言检索中发挥重要的作用,依靠多语言本体将处理信息进行转换或标注^[69-70],从而实现对不同语言翻译以及在不同语言环境中的信息检索,这与元数据标注具有异曲同工之处。国外7个多语言公共数字文化服务平台面对不用场景的多语言需求,通过针对性的利用不同技术的优势,继而维持一定的多语言服务。

6 我国多语言公共数字文化服务平台建设思考

在对5S模型进行解析、结合与延伸的基础上,通过对国外7个多语言公共数字文化服务平台进行社会、资源、服务、技术4个模块的分析研究发现,国外在早期阶段即对多语言服务平台进行了关注与构建,并通过一系列政策与活动以此提高多语言服务的质量以及用户的活跃度。但是,受限于技术发展的约束,新技术落地的滞后性以及应用上的保守性等因素,多语言服务也始终无法达到一个较为全面且完善的层次。尽管如此,相对于国内来说,国外的多语言公共数字文化服务平台建设较早,在某些建设经验上值得国内多语言公共数字文化服务平台思考。

6.1 联合政府建立联盟性质的多语言公共数字文化服务平台

由调研发现,国外多语言公共数字文化服务平台均是由各个机构合作共建,合作机构类型多样,数量基底较为庞大,并利用合作联盟构建了数字文化资源共享中心,聚合多样性数字资源于开放平台,最大程度上满足了用户对于资源多样性以及资源获取的需求。在智慧城市的背景下,公共数字文化服务逐渐被多螺旋式协同模式所影响,合作机构也亦趋于多元化^[71],企业、政府等多维度的集合体更能够有益于多语言服务的开展。并且,从国外多语言公共数字文化服务平台的联盟构成可以看出,政府是联盟构成体中不可或缺的部分,借助政府颁布与实施

表5 国外7个多语言公共数字文化服务平台的技术对比

服务平台	WDL	ICDL	Europeana	dLOC	DPLA	NLA	NDL
多语言技术	元数据翻译、翻译记忆工具、机器翻译、人工翻译等	元数据翻译、机器翻译、人工翻译	机器翻译、命名实体识别、翻译歧义、基于语境词表等	元数据翻译	元数据翻译	元数据翻译、基于实例、基于语境词表	元数据翻译、基于实例、基于语境词表的机器翻译

一系列多语言服务相关的政策与活动,能够给平台的运行提供了更多的支持与保障,此举给国内多语言公共数字文化服务平台的构建更多的参考与借鉴。国内多语言公共数字文化服务平台也需要将政府纳入合作体,并利用政策、活动等软文化手段,提高多语言服务的范围与质量。

6.2 借助丰富的多语言资源提供多模态的多语言服务

从整体上看,国外公共数字文化服务平台的多语言资源语种种类较为丰富,甚至,部分平台的存储语种资源已经达到数百种,为多语言服务构建了坚实的基底。当然,丰富的多语言资源只是平台提供多语言服务的基础,如何更好地利用多语言资源提供多样性的多语言服务才是更进一步的目标。通过对门户界面、跨语言、多语言检索以及多语言元数据描述三个方面多语言服务的调研看,国外平台提供的多语言服务与语种资源并不能够达到一定程度上的匹配。总体而言,多语言服务还处于一个比较初步的状态,并且局限于文本的多语言服务。而随着时代的进步与发展,多模态的多语言服务有着更为广泛的需求。例如,用户可以直接通过图片或者语音的输入,获取不同语言的相关信息。因此,对于国内多语言服务平台来说,首先保障多样性语种资源是基础,其次,需要在此基础上充分发挥多语言资源的能动性。在跨语言、多语言检索服务等场景中,如多语言服务仍然局限于文本将无法最大程度上满足用户的需求。语音、图片、视频等多模态信息的涌入促使多语言平台的信息服务向多模态的多语言服务升级。多语言公共数字文化服务平台必须在现有基础上,深化多语言服务内容,紧跟时代发展满足更多用户的需求。

6.3 摆脱技术应用的固定思维抓住新技术带来的机遇

从国外7个多语言公共数字文化服务平台所使用的多语言技术的来看,其所应用的技术在先进性方面依旧过于保守,而这也正是国内多语言公共数字文化服务平台的通病。机器翻译在国外多语言公共数字文化服务平台的普遍应用标志着将机器翻译技术引入多语言公共数字文化服务平台的必然性。

近期,ChatGPT的爆火展现出预训练模型的惊人潜力,其利用代码理解任务和人类反馈对齐技术,获得了超乎寻常的复杂推理能力和泛化到新任务的能力,而这给以训练模型为核心的机器翻译技术也带来了更大的冲击,预示着机器翻译的翻译质量可能也远远未达到上限,对于国外多语言公共数字文化服务平台来说,其多语言服务理应拥有更多的可能性。但是,现状却不尽如人意。目前,国内多语言公共数字文化服务平台正在尝试通过自建多语言翻译系统提供多语言服务,如构建“丝路”多语言公共数字文化平台服务于“一带一路”基础文化建设与传播。面对国外ChatGPT^[72-73]、PaLM^[74]等大模型在技术上的突破,国内在多语言公共数字文化服务平台的建设上应该及时的作出调整,避免出现技术滞后,轻易被取代,继而导致服务质量停滞不前的局面。

7 结语

公共数字文化服务平台提供多语言服务是促进社会经济发展、国家文化交流传播的必经之路。本文利用5S模型构建多语言公共数字文化服务平台的分析框架,在该框架下对国外7个多语言公共数字文化服务平台进行分析发现,国外多语言公共数字文化平台建设较早,在政策、方针、资源等方面具有一定的前瞻性,但在多语言技术的先进性以及多语言服务的完备性方面还有待加强,对我国多语言公共数字文化服务平台的构建具有一定的思考价值。

作者贡献说明:

刘莉:资料与数据收集、整理,论文撰写与修改;
邵波:研究思路指导,论文修改。

参考文献:

- [1]中华人民共和国文化和旅游部.《“十三五”时期公共数字文化发展规划》[EB/OL].[2023-10-08].https://zwgk.mct.gov.cn/zfxxgkml/ghjh/202012/t20201204_906376.html.
- [2]王元荣,潘颖,郑建明.我国公共文化事业的地区发展战略[J].图书馆论坛,2023,43(8):4-12.
- [3]中华人民共和国文化和旅游部.《“十四五”公共文化服务体系建设规划》[EB/OL].[2023-10-08].http://www.gov.cn/zhengce/zhengceku/2021-06/23/content_5620456.htm.(Ministry of Culture

and Tourism, PRC. The "14th Five-Year Plan" for the construction of a public cultural service system[EB/OL].[2023-10-08]. https://www.puershi.gov.cn/info/egovinfo/1001/xxgk_content/1021-12_A/2017-1011002.htm.)

[4]中华人民共和国文化和旅游部.《关于推动公共文化服务高质量发展的意见》[EB/OL].[2023-10-08].http://zwgk.mct.gov.cn/zfxxgkml/ggfw/202103/t20210323_923230.html.

[5]国务院办公厅.《关于推进实施国家文化数字化战略的意见》[EB/OL].[2023-10-07].<http://www.gov.cn/zhengce/2022-05/22/content5691759.htm>.

[6]于兴尚,谭洪,李保东.公共数字文化服务平台构建的价值议题与逻辑进路[J].数字图书馆论坛,2023,19(2):65-72.

[7]陈则谦,郑娜静,李亚灿,等.一站式公共数字文化服务平台支持用户参与的服务现状、主要问题与优化建议[J].图书馆情报工作,2021,65(7):78-89.

[8]华方园,陈思任,余安琪.国内公共数字文化服务平台建设现状调查分析[J].图书馆研究,2018,48(1):37-45.

[9]中华人民共和国文化和旅游部.《“十四五”“一带一路”文化和旅游发展行动计划》[EB/OL].[2023-10-08].https://www.mct.gov.cn/whzx/whyw/202107/t20210719_926507.htm.

[10]IFLA. IFLA- UNESCO public library manifesto 2022[EB/OL].[2023-10-08]. <https://repository.ifla.org/bitstream/123456789/2006/1/IFLA-UNESCO%20Public%20Library%20Manifesto%202022.pdf>.

[11]赵生辉,胡莹.多语言数字图书馆信息生态链的结构、类型及启示[J].图书馆理论与实践,2020(3):73-78.

[12]鹏城实验室.聚焦“一带一路”,鹏城实验室举办多语言机器翻译国际大科学计划专题报告会并发布“丝路”多语言机器翻译平台[EB/OL].[2023-10-08].<https://www.pcl.ac.cn/html/943/2021-06-09/content-3810.html>.

[13]张树臣,陈伟,高长元.大数据环境下公共数字文化服务云平台构建研究[J].情报科学,2021,39(4):112-118.

[14]王丹,陈雅.基于区块链技术的公共数字文化服务平台架构研究[J].图书馆学研究,2021(11):32-39.

[15]完颜邓邓,王子健.大数据环境下公共数字文化服务模式创新研究[J].图书与情报,2020(5):59-66.

[16]孙国焯,吴丹.多语言公共数字文化服务平台信息组织标准规范体系构建[J].图书馆建设,2022(1):56-66.

[17]陈则谦,余晓彤,郑娜静,等.公共文化云服务的评价指标构建及应用[J].图书馆情报知识,2020(6):54-66.

[18]郭春侠,吴裕鸿.长三角地区省级公共数字文化服务平台优化策略[J].图书馆学刊,2023,45(1):80-87.

[19]朱莉.公共数字文化平台服务效能评价体系构建及提升策略研究[J].情报科学,2023,41(5):107-114.

[20]吴丹,樊舒.面向多源异构资源融合的公共文化数字

化建设路径[J].西安交通大学学报(社会科学版),2021,41(5):136-143.

[21]韦楠华,吴高.公共数字文化资源共建共享现状、障碍及对策研究[J].图书馆建设,2018(9):18-26.

[22]吴丹,古南辉,何大庆.数字图书馆用户的多语言信息需求调研[J].图书馆情报工作,2011,55(2):6-10.

[23]杜慧平.数字图书馆的多语言信息存取——European项目进展与启示[J].图书馆杂志,2012,31(4):20-23,32.

[24]曹玲.国际儿童数字图书馆(ICDL)建设特点及其启示[J].图书馆杂志,2006(3):65-67.

[25]张彦文.多语言数字图书馆及其研究浅析[J].河南图书馆学刊,2013,33(9):97-99.

[26]WU A, CHEN J. Sustaining multilinguality: case studies of two multilingual digital libraries[J]. The electronic library, 2022, 40(6): 625-645.

[27]BUDZISE-WEAVER T, CHEN J, MITCHELL M. Collaboration and crowdsourcing: the cases of multilingual digital libraries[J]. The electronic library, 2012, 30(2): 220-232.

[28]司莉,封洁,杨君正,等.多语言信息组织与检索用户需求调查分析[J].图书馆论坛,2016,36(2):7-12.

[29]司莉,潘秋玉,庄晓喆.基于实证的多语言信息检索用户行为分析[J].国家图书馆学刊,2015,24(5):81-90.

[30]司莉,贾欢.多语言信息检索平台 WorldWideScience 功能及测评研究[J].图书馆学研究,2015(17):49-53.

[31]ALSALMI H M. Information-seeking in multilingual digital libraries: comparative case studies of five university students [J]. Library hi tech, 2020, 39(1): 80-100.

[32]MIZERA-PIETRASZKO J, TANCULA J. Rough set theory for supporting decision making on relevance in browsing multilingual digital resources[C]//Asian conference on intelligent information and database systems. Cham: Springer, 2017: 129-139.

[33]刘莉,王怡,邵波.机器翻译在图书馆中的研究现状及应用趋势分析[J].图书馆学研究,2021(24):2-8,41.

[34]刘莉,王怡,邵波.面向智慧图书馆的多语言自动翻译平台架构设计研究[J].图书馆学研究,2022(6):37-44.

[35]刘莉,王怡,邵波.基于机器翻译的图书馆多语言自动翻译平台构建策略[J].图书馆学研究,2022(1):32-40.

[36]KITUKU B, MUCHEMI L, NGANGA W. A review on machine translation approaches[J]. Indonesian journal of electrical engineering and computer science, 2016, 1(1): 182-190.

[37]章成志,王惠临.面向数字图书馆应用的多语言领域本体学习研究[J].图书馆情报工作,2011,55(2):11-15,94.

[38]MOZGOVOY M. Towards Word Bricks—a virtual language lab for computer-assisted language learning[C]//Proceedings of the federated conference on computer science and information

systems. Wrocław: IEEE, 2012: 229-232.

[39]VASILJEVS A, HAJIC J, HUMMEL J, et al. European platform for the multilingual digital single market: conceptual proposal[M]//Human language technologies—the Baltic perspective. IOS Press, 2016: 20-27.

[40]YIN R K. Case study research: Design and methods[M]. London: Sage, 2009.

[41]刘莎莎. 澳大利亚国家图书馆与国家图书馆法[J]. 山东图书馆季刊, 2006(4): 27-29.

[42]庞淑杰. 日本国立国会图书馆大规模馆藏数字化工程[J]. 图书馆理论与实践, 2014(9): 86-88.

[43]WDL. About the World Digital Library[EB/OL].[2023-10-08]. <https://www.wdl.org/zh/about/>.

[44]曹玲. 国际儿童数字图书馆(ICDL)建设特点及其启示[J]. 图书馆杂志, 2006(3): 65-67.

[45]Europeana. Europeana[EB/OL].[2023-10-08]. <https://www.europeana.eu/en>.

[46]dLOC.dLOC[EB/OL].[2023-10-08]. <https://www.dloc.com/>.

[47]DPLA. STRATEGIC PLAN: 2015 through 2017[EB/OL].[2023-10-08]. https://dp.la/info/wpcontent/uploads/2015/01/DPLAstrategicPlan_2015-2017-Jan7.pdf.

[48]Trove. Above Trove[EB/OL].[2023-10-08]. <https://trove.nla.gov.au/>.

[49]NDL. ヘルプ[EB/OL].[2023-10-08]. http://iss.ndl.go.jp/help/help_ja/help.html.

[50]GONCALVES M A, FOX E A, WATSON L T. Towards a digital library theory: a formal digital library ontology[J]. International journal on digital library, 2008(8): 91-114.

[51]徐大芳, 王亚军. 数字图书馆5S模型的构成与应用[J]. 情报科学, 2008(6): 872-875.

[52]张文萍, 邓仲华, 黎春兰. 基于5S模型的数字图书馆本体模型研究[J]. 图书馆杂志, 2013, 32(7): 57-62.

[53]世界数字图书馆. 世界数字图书馆章程[EB/OL].[2023-10-08]. <https://www.loc.gov/collections/world-digital-library/about-thiscollection/>.

[54]FORSTER H. The 2010 digital libraries initiative: Europe's culture and scientific information at the click of a mouse[J]. Information services and use, 2007, 7(10): 155-159.

[55]European. The European commission under the fifth framework programme[EB/OL].[2023-10-08]. <http://ec.europa.eu/research/fp5.html>.

[56]黄小平. 可持续发展背景下的图书馆社会责任——IFLA《图书馆与可持续发展目标故事讲述指南》的解读与启示[J]. 图书与情报, 2018(3): 103-108.

[57]IFLA. IFLA[EB/OL].[2023-10-08]. <https://www.ifla.org/takeaction/>.

[58]DPLA. Strategic Roadmap, 2019-2022[EB/OL].[2023-10-08]. <https://pro.dp.la/about-dpla-pro/strategic-plan>.

[59]刘博宇. DPLA对公共图书馆数字化项目发展的启示[J]. 图书馆学刊, 2013, 35(11): 141-143.

[60]张涛. 澳大利亚公共图书馆多元文化服务述略[J]. 山东图书馆学刊, 2013(6): 14-23.

[61]澳大利亚国家图书馆推出新的搜索引擎 Trove[J]. 现代图书情报技术, 2010(4): 71.

[62]日本国会. 国立国会図書館法[EB/OL].[2023-10-08]. https://dl.ndl.go.jp/view/download/digidepo_1000780_po_a1102.pdf?.contentNo=1&alternativeNo=.

[63]陈瑜, 尹铭莉. 日本国立国会图书馆新检索服务“NDL Search”介绍及分析[J]. 图书馆杂志, 2013, 32(5): 81-84.

[64]杨士丽. 英国国家图书馆资源数字化国际合作项目分析与启示[J]. 图书馆学刊, 2022, 44(5): 105-110.

[65]万益嘉, 顾立平. 全球开放获取学术交流联盟平台政策、资源与服务研究[J]. 农业图书情报学报, 2022, 34(6): 24-35.

[66]曾粤亮, 肖希明. 国外公共数字文化资源整合平台信息组织调查与分析[J]. 国家图书馆学刊, 2016, 25(5): 39-46.

[67]SALTON G. Experiments in multilingual information retrieval[J]. Information processing letters, 1973, 2(1): 6-11

[68]王世慧, 杜伟. 澳大利亚搜索引擎 Trove 对 CALIS 的启示[J]. 图书馆学研究, 2011(24): 61-64, 10.

[69]ESPINOZA M, MONTIEL-PONSODA E, GÓMEZ-PÉREZ A. Ontology localization[C]//Proceedings of the fifth international conference on knowledge capture. New York: ACM, 2009: 33-40.

[70]ABDELALI A, COWIE J, FARWELL D, et al. Cross-language information retrieval using ontology[C]//Proceedings of traitement automatique des langues naturelles. Batz-sur-Mer: ATALA, 2003: 72-86.

[71]晁亚男, 胡莹. 智慧城市背景下国外公共图书馆发展现状研究——以新加坡、澳大利亚、加拿大三国为例[J]. 国家图书馆学刊, 2021, 30(4): 86-98.

[72]BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

[73]LONG OUYANG, JEFF WU, XU JIANG, et al. Training language models to follow instructions with human feedback[J]. ArXiv: 2203.02155, 2022.

[74]AAKANKSHA C, SHARAN N, JACOB D, et al. PaLM: scaling language modeling with pathways[J]. ArXiv: 2204.02311, 2022.