大语言模型结合知识图谱构建专业汉语教材的路径探索和技术实现

赵毅飞

（华侨大学华文学院，福建 厦门 361021）

摘要：为解决专业汉语教材编写面临的内容专业性强、细分度高、资源匮乏等问题，本文探讨了大语言模型结合知识图谱在构建专业汉语教材方面的路径探索和技术实现。首先利用机器语言思维转化多类型需求，并将知识图谱融入其中，构建专业汉语教材的提示语模板，以提高教材编写效率和质量。其次利用大语言模型强大的自然语言理解和生成能力，结合知识图谱提供的结构化、丰富且准确的知识资源，生成具有专业性、准确性和多样性的专业汉语教材内容。在计算机专业汉语教材构建示例中，提出了详细的架构设计和课文模块设计，强调了基于知识图谱的课文内容生成、词汇列表、语法注释和多层次练习的设计，以及结合人类反馈进行的教材合规审核机制。未来的研究应关注优化提示语模板，加强人工智能辅助审核系统的研发，构建全方位教学资源库，并关注原创性和知识产权问题。

关键词：专业汉语教材；大语言模型；知识图谱；路径探索

一、引言

专业汉语作为专门用途汉语（CSP）的关键分支，与“业务汉语”和“中文 +”并列，是专门用途汉语的重要组成部分。专业汉语教材的编写，不仅要求涵盖词汇、语法、阅读等基础语言技能，更需针对特定专业领域，提供专业性话题和案例分析，提升学习者的专业情境感知素养和语言实践应用能力，满足学习者针对特定领域的个性化需求，提高学习效果。然而，目前关于专业汉语教材的研究尚显不足，成熟的专业汉语教材也十分有限。这主要是因为专业汉语教材的内容专业性强，细分度高，不仅要求编者对专业有精深的理解，还要熟悉汉语作为第二语言的教学规律。因此，能够结合国际汉语的教学经验，深刻理解专业知识，成为解决专业汉语教材建设的关键途径。

大语言模型（LLM）是一种使用海量文本数据训练的深度学习模型，不仅可以生成自然语言文本，还可以深入理解文本含义，处理各种自然语言任务。知识图谱是一种结构化的知识表示方式，它可以将不同领域的知识组织起来，以便机器和人类进行理解和应用。将大语言模型和知识图谱结合是目前的发展趋势，二者结合望为自然语言处理领域带来革命性的变革。首先，知识图谱为 LLM 提供了丰富的结构化知识，增强了模型对语言含义和上下文的理解能力。通过整合知识图谱中的实体、关系和三元组信息，LLM 可以更准确地捕捉文本中的语义信息和潜在含义，使 LLM 在处理复杂语言现象和特定领域任务时更具优势。其次，知识图谱中的实体和关系为 LLM 提供了丰富的词汇和表达方式，扩展了模型的词汇库和表达方式，能使其生成更加多样化和自然的文本，从而产生更具创造性和个性化的文本输出。总之，LLM 负责文本生成和分析，而知识图谱则提供结构化的知识表示和推理能力。这种结合支持更复杂的知识处理和应用场景，使得我们可以更加高效地处理和管理知识，从而实现全链条的知识管理与协同应用。

在专业汉语教材领域，结合 LLM 与知识图谱的应用具有巨大的潜力和价值，其核心是应用路径的探索。一方面，针对专业汉语教材编写的特点，以机器语言的思维将多类型需求转换成提示语，这些提示语将作为教材编写的重要参考，帮助编写者更加明确和系统地整理教材内容。另一方面，将知识图谱融入提示语，可以为教材编写提供更加丰富和准确的信息资源，增强教材的专业性和准确性，提高编写效率和质量。要实现这一应用路径，关键挑战就是教材生成的提示语设计。这一挑战需要我们结合专业汉语教材编写的特点，使用大语言模型和知识图谱的技术手段，创新设计出提示语模板。该探索将有助于推动专业汉语教育的数字化转型和专业发展，为留学生提供优质的教材资源，也将为未来的教学创新和教材编写提供新的思路和方法。

二、专业汉语教材建设研究

近年来，专门用途汉语研究综述较多，但相较于通用汉语，其在教材开发、资源建设及学术研究上均显滞后。文献计量分析显示，2014 年 1 月至 2023 年 7 月，专业汉语期刊论文仅 153 篇，且研究领域发展不均衡。目前教材以商务、科技为主，其他专业领域教材亟待开发。专业汉语基于内容教学，需求巨大但资源匮乏。教育部发布《“中文 + 职业技能”行动计划》，推动“中文 +”系列教材扩展，但权威品牌缺乏，内容宽泛不深入，忽视语体特色，鲜有教材兼顾专业性、真实性与时代性。大语言模型与知识图谱技术为教材建设提供新思路，通过二者的结合，可以提高教材内容的质量，实现知识的跨界整合和协同创新。先进的自然语言处理技术在保持专业汉语教材的时效性　和科学性方面具有关键的作用，也能有效解决国际汉语教师在编写专业汉语教材时专业知识缺乏的问题。但其在专业汉语教材编写中的应用尚待探索。

研究表明，中国专门用途汉语教学始于 20 世纪 50 年代，与国际中文教育同步，但发展滞后。鉴于专业汉语需求激增，现有教学存在缺口，发展潜力大。目前专业汉语教材稀缺，尤其缺乏铁路、航海、旅游、计算机等领域教材，急需加强建设以满足国际交流需求。相较于专业英语，专业汉语在国际中文教育中比重不足，构建专业汉语教材对提升国际中文教育专业度至关重要。专业汉语教材编写需结合二语教学理念和专业知识，促进学科交融，提升教师素养，推动国际中文教育。

综上，专业汉语教材建设对于满足国际社会对专业汉语的需求、推动国际文化交流、提升国际中文教育的专业度、促进跨学科融合以及提高国际中文教师的专业素养等方面都具有重要的意义。然而专业汉语教材的国际需求迫切但资源相对匮乏，发展相对滞后，其研究和出版资源均不足。尽管国家已出台相应计划以推进其发展，但当前教材内容及编写方式均面临挑战，仍处于探索阶段，缺乏高质量的权威品牌教材。针对此，将自然语言处理技术，特别是大语言模型和知识图谱技术，应用于专业汉语教材

编写的研究尚显不足。因此，探索这些技术在专业汉语教材建设中的应用路径，将有助于推动国际中文教育的持续和深入发展。

三、大语言模型结合知识图谱构建专业汉语教材的路径

（一）大语言模型结合知识图谱构建专业汉语教材的架构设计

为实现知识图谱增强的大语言模型专业汉语教材生成，需重点探索专业知识图谱在大模型提示语设计与生成中的机制。

大语言模型结合知识图谱构建专业汉语教材的架构 1 所示，架构的底层分成两部分，一部分是将知识图谱作为核心知识库，梳理并结构化专业领域和国际中文教育领域的多元教学内容，建立知识点间的逻辑关系，为知识点添加适当的标签和分类，形成逻辑严密、脉络清晰的知识体系；另一部分是设计出结构化交互引导提示语，即利用机器语言思维转化多类型需求，并将知识图谱融入其中，构建专业汉语教材的提示语模板，以提高教材编写效率和质量。架构的中间层是大语言模型生成，利用大语言模型强大的自然语言理解和生成能力，结合知识图谱提供的结构化、丰富且准确的知识资源，生成具有专业性、准确性和多样性的专业汉语教材内容。架构的上层是 AI 生成的专业课文模块，包括标题、课文文本、计算机专业词汇表、词汇练习题、课文语法点、课文阅读理解题、拓展阅读文本以及拓展阅读文本练习题等。此外，专业汉语教材的评估与质量控制包括 AI 审核和专家审核两个模式。其中 AI 审核运用自然语言处理和知识图谱技术对生成教材进行初步的质量筛查。同时，专家审核由学科领域专家对教材进行人工评估，确保教材的学术严谨性和教学有效性。

（二）专业课文模块设计

教育部中外语言交流合作中心于 2022 年制订发布的《国际中文教材质量评测体系》为国际中文教材的内容构建提供了指导框架，强调教材应综合包含原创性内容、系统化语言知识讲解、具有代表性的课文单元，并配以生词注解、语法解析及实例展示，旨在实现语言知识教育与语言技能培养的有效融合，同步配备针对性的语言技能训练习题。据此，本研究遵循科学化原则，设计了一套适用于专业中文教学的课文模板体系，该体系由八个核心模块构成，具体阐述如下：

1. 课文标题和课文文本模块：给定任务相关的条件文本作为输入内容，模型生成与之相符的课文标题、内容丰富且结构完整的课文文本。专业汉语课文依据国际通用的《国际中文等级标准大纲》设定语言难度，紧密围绕计算机专业知识领域，尤其以“计算机专业 101 知识图谱”为基础，选取具有科普性质的话题和内容进行编写。

2. 计算机专业词汇表模块：词汇列表汇聚了与课文主体内容紧密相关的关键词汇，是学生学习的核心焦点。鉴于本研究教材面向计算机专业留学生群体，词汇列表所收录的词汇严格限定为计算机专业术语范畴，致力于拓展留学生的专业词汇储备，并对照计算机领域知识图谱及专业主题词典进行遴选与整合。

3. 课文语法点模块：本模块精炼萃取自《国际中文等级标准大纲》中针对留学生阶段应掌握的各项语法要点、语言结构、句型模式以及惯用表达等元素。本研究教材中的语法注解部分，系统梳理了课文内对应等级的语法知识，并通过精心编排的例句进行简洁明了的阐释说明。

4. 词汇练习题模块：词汇练习模块作为辅助巩固语言知识和提升读写技能的重要实践环节，围绕词汇列表中的关键词汇设计技能训练题。本研究所创设的词汇练习包括两组选词填空题目，旨在强化学生对核心词汇含义和应用情境的深度认知。

5. 课文阅读理解题模块：课文理解练习是深化学生对课文内涵领悟的关键途径。本研究所构思的此类练习基于课文多元内容层面设计而成，用于客观评估学生是否真正透彻理解了课文的主旨和细节。

6. 拓展阅读文本模块：扩展阅读部分旨在拓宽学生视野，通过提供与课文主题紧密关联但内容新颖独立的文章，对课文所涉知识点形成有益补充。本研究所策划的扩展阅读材料篇幅适度短于课文，保持内容的连贯性与重点突出，以避免分散学生对主课文内容的关注。

7. 拓展阅读文本练习题模块：为有效测评学生对扩展阅读文章的领悟能力，本研究设计了一系列包含多种考查角度的扩展阅读理解练习题，共计 5 道题目。这些练习旨在全面锻炼和提高学生的篇章阅读理解能力。

（三）结构化提示语模板设计

提示语是大模型的输入文本，是表达意图和需求的指令，可以是简单的单个词，如“画图”或“搜索”，也可以是一整句，一个好的提示语能够打开 AI 大模型的无限潜力。提示语是有要件和节奏的，呈现结构化特性。本研究设计的提示语模版框架为 RACWI（Role Ability Constraint Workflow Initialization），该框架将用户输入分解为五个主要部分：角色、能力、约束、工作流和初始化。这种分解方法有助于提高人机协作效率，减少误解与误差，充分发挥 AI 大模型的潜力。RACWI 的结构化特性说明如下：

1. 设定角色

在提示语中明确指定 AI 大模型在执行任务时所扮演的角色或身份。这有助于模型理解其在特定场景下的定位，如“作为设计师”“作为法律顾问”“作为数据分析师”等。

2. 明确能力

列举模型需要具备的核心技能或能力，以便其在执行任务时能充分发挥作用。比如，“具备图像识别能力”“精通法律条文检索与解读”“熟练掌握统计分析与预测方法”等。此外，还明确指出模型应具备的知识背景，如“熟悉心理学理论”“了解全球金融市场动态”“精通中西烹饪技法”。

3. 约束规则

规定模型在执行任务时应遵循的任务边界，即限制条件或不应逾越的边界，如“仅使用公开可用数据源”“避免涉及个人隐私信息”“在法律法规允许范围内提供建议”。此外，还设定模型的输出规范，即生成结果应满足的标准或格式要求，如“生成报告需包含摘要、分析过程、结论与建议”“绘制图表应遵循公司 VI 标准”“生成代码应符合 Python PEP8 编码规范”。

4. 工作流程

将复杂任务进行多步拆解，即拆解为一系列明确的子任务或步骤，并在提示语中逐一列出，如“首先进行数据清洗，然后执行特征工程，最后训练并验证机器学习模型”。此外，对于涉及判断或选择的环节，明确给出决策逻辑，包括指导原则或决策树，如“根据用户偏好和历史行为数据，按以下优先级推荐商品：1）用户已收藏；2）用户近期浏览过；3）高用户评分”。

5. 初始化

指示模型进行数据预载，即在开始任务前加载必要的初始数据或状态信息，如“加载用户 A 的历史购买记录和偏好设置”“从云端获取最新更新的模型权重文件”。此外，还设定模型的环境配置，即运行所需的软硬件环境参数或依赖服务的状态，如“使用 GPU 加速计算”“确保与 API 接口 Y 的连接正常”。

（四）基于 AI 和人类反馈的教材合规审核

对语言模型生成的内容加以人类反馈，将进一步提升内容的质量，因此我们将对教材的每一具体模块进行合规审核。审核规范设计如下：

（1）对课文标题和课文文本的审查过程涵盖了多个维度的精密评估，其中包括但不限于字数适宜性、国际中文词汇的难易程度量化分析、语法构造的复杂度测定以及目标专业知识内容在文本中的覆盖率统计。为了实现教材高效精确的审核目标，本研究计划部署基于人工智能技术的自动化审核系统。

（2）对于词汇列表部分的审查机制，旨在验证列表中所有词汇是否严格契合目标专业领域，并确认它们在实际课文中均有据可查。此项工作将不再依赖单一的人工智能算法，而转向权威专家的严谨审核程序。

（3）在语法注释方面，审核的核心任务是对标注的语言知识点进行严格鉴定，确认其难度级别与预设教学要求相符，若不符则需进一步归类至相应难度层次。本研究所采用的方法同样倚重专家团队的专业评估，以精准识别各个语言点的教学难度层级及其在教材中的分布情况。

（4）针对词汇练习模块，审核工作聚焦于题目的语言流畅性检验、排查潜在的语言错误、核实问题表述与提供的答案之间的一致性和唯一性。这一环节亦将运用专家审核机制，确保每一项练习的质量达到标准。

（5）在课文理解练习的评判上，重点在于验证题目答案的出处确实源自课文内容，本研究坚持采用专家评审方法，确保每一道理解题目的合理性。至于扩展阅读部分，则涉及字数控制与主题相关性的双重考量，同时，对扩展阅读配套练习的审核着重于考察答案是否能在原文中得到确切支撑。此部分的审核同样会依托专家力量，对上述各项指标进行全面细致的核查。

四、大语言模型结合知识图谱构建专业汉语教材 – 以计算机专业汉语为例

（一）计算机专业汉语教材生成的过程展示

以下展示根据 RACWI 框架进行专业汉语生成的实现方法。专业汉语课文需要生成的模块较多，为确保每一模块按照需求自动生成，初始提示语需要简洁全面地设计工作流程和规则，大语言模型会按照预期逐步交互，生成所需模块。因此，按照工作流程需要人机协同进行多步交互，该交互过程呈现多轮对话的特点，且以大模型主动引导为特色。完整示例请见网盘链接：（https：//pan.hqu.edu.cn/share/ff06b4fbea0e56d71d063091ab）

（二）计算机专业汉语教材的审核与修正

AI 能够对教材中的客观要素进行高效的自动审核，在本研究中，课文的总字数、目标关键词覆盖率、题目数量、选项字数、题目排列顺序、阅读理解题目的选项字数等能够通过 AI 进行准确的审核。人类专家审核能够解决 AI 不能完美解决的专业知识科学性、文本表述的正确性、流利性与连贯性、主题匹配性、内容完整性、标点符号正确性等方面。AI 审核和专家审核的联合审核能够反映出大语言模型结合知识图谱生成专业汉语教材的质量。

AI 审核和专家审核的联合审核过程展示完整示例请见网盘链接：（https：//pan.hqu.edu.cn/share/22e90425aeaeb7af1dbc7de21c）。

五、结语

专业汉语教材建设对于满足国际社会对专业汉语日益增长的需求、推动跨文化交流、提升国际中文教育的专业度以及促进教师专业素养的发展具有深远的意义，结合大语言模型与知识图谱技术的创新应用，为专业汉语教材的编写提供了崭新的解决方案。大语言模型与知识图谱的深度融合，不仅能够为教材内容注入丰富、准确且结构化的专业知识，增强教材的专业深度与广度，还能通过智能化的提示语设计，辅助编者系统化地整理教材内容，提高编写效率。这种结合实现了知识的跨界整合与协同创新，使得教材能够紧跟时代步伐，保持高度的时效性与科学性，有效弥补国际汉语教师在专业知识方面的不足。尽管这一技术路径的应用尚处于初级阶段，其展现出的潜力预示着它将在未来专业汉语教材建设中发挥关键作用。

本研究仍有一些工作需要优化，首先，应进一步优化提示语模板设计，使之更加贴合专业汉语教学需求。例如在提示语中注重体现教学目标，使教材内容能够有针对性地帮助学生掌握相关知识和技能；将学生的认知特点考虑进提示语，采用易于理解和接受的语言和表达方式，有效引导大语言模型产出符合专业规范、教学目标与学生认知特点的高质量教材内容。其次，加强人工智能辅助审核系统的研发与应用，推进人机协同编写与审核流程标准化。建立和完善人机协同编写教材的工作流程，设计更加精细的结构化提示语模板，并结合人工智能自动审核与专家人工审核相结合的方式，在确保教材质量的同时提高编写效率。这将涉及对 AI 审核算法的持续优化，以及制定一套适合不同专业领域的教材内容审核标准。另外，如何利用大语言模型和知识图谱的优势，构建覆盖各类专业领域的全方位汉语教学资源库，为不同专业背景的学习者提供系统化的课程体系和教材资源，如何处理大语言模型生成内容的原创性、知识产权保护等问题，都应得到足够的重视。

参考文献：

[1] 邓凯尤 . 面向理工类本科学习的预科留学生词汇需求分析 [D]. 北京：北京外国语大学，2022.

[2] 乔胜男 . 基于近代海关洋员汉语教材的现代航海汉语教材编写设计研究 [D]. 山东：山东师范大学，2021.

[3] 李亚男 .《航海汉语》课程建设与教材编写——以大连海事大学为例 [J]. 现代语文，2020（01）：110-114.

[4] 熊珈玄 . 近十年专门用途汉语教学研究述评 [J]. 汉字文化，2023（19）：83-86.

[5] 刘皓文 . “中文 + 职业技能”培养模式下专门用途汉语教材分析 [D]. 南京：南京信息工程大学，2023.

[6] 孙莹，李泉 . 专门用途汉语教学研究综论（1980—1999）[J]. 国际汉语教学研究，2022（04）：28-37.

项目：中外语言合作中心 2023 年国际中文教学实践创新项目《大语言模型结合知识图谱构建专业中文教材的路径探索》编号：YHJXCX23-063