

数字与文学联袂：中国外国文学研究索引（CFLSI）数据库的研制¹

Digital Meets Literature: Development of the China Foreign Literature Studies Index (CFLSI) Database

王永（Wang Yong） 杨恽恒（Yang Yiheng） 张克俊（Zhang Kejun）

内容摘要：我国外国文学的译介与研究已有 100 多年的历史，成果丰硕。全面总结这些研究成果，对其做出科学的分析，可以对外国文学译介与研究提供正确科学的引导，这是中国外国文学研究索引（CFLSI）数据库研制的价值所在。本文重点阐述数据库研制过程中探索的几个问题：1）从外国文学学科的特殊性出发，通过分析公共网络资源现状，论证构建学科领域数据库的必要性；2）从外国文学研究者对数据库的需求出发，探讨 CFLSI 数据库的数据构成、功能设计及其实现途径；3）基于数据构成及功能设计需求，规划数据库系统构架并完成技术开发。该数据库的研制是数字技术与外国文学学科知识深度融合的探索与实践。

关键词：外国文学研究；文献；索引；数据库；数字技术

作者简介：王永，文学博士，浙江大学外国语学院教授、博士生导师，主要从事俄罗斯文学的计量研究、计量语言学、语义学；杨恽恒，浙江大学外国语学院博士生，主要从事知识工程研究；张克俊，浙江大学计算机科学与技术学院教授、博士生导师，主要从事数据挖掘、机器学习研究。本论文为国家社科基金重大项目“中国外国文学研究索引（CFLSI）的研制与运用”【项目批号：18ZDA284】的阶段性成果。

Title: Digital Meets Literature: Development of the China Foreign Literature Studies Index (CFLSI) Database

Abstract: The translation and research of Chinese foreign literature have a rich and distinguished history that spans over a century, yielding numerous and substantial achievements. These research outcomes for thoroughly synthesized and scientific analyses can provide authoritative and accurate guidance for foreign literature translation and research, highlighting the significance of the China Foreign Literature Studies Index (CFLSI) database. This article focuses on three issues

¹ 浙江大学计算机科学与技术学院硕士生李一非为本数据库的研制做出了很大贡献，论文作者在此深表感谢。

encountered in database development. 1) An examination of the limitations of existing public network resources based on the unique nature of the foreign literature discipline, thereby establishing the necessity for the creation of a discipline-specific database; 2) An exploration of the composition, functional design, and implementation approaches of the CFLSI database in response to the requirements of foreign literature researchers; 3) The establishment of the database framework and development technology of the system aligned with the data composition and functional design. The creation of the CFLSI database represents an exploration and application to merge digital technology with a comprehensive understanding of the foreign literature discipline.

Keywords: foreign literature studies; literature; index; database; digital technology

Authors: **Wang Yong**, Ph. D., is Professor at School of International Studies, Zhejiang University (Hangzhou 310058, China). Her major fields of inquiry include quantitative studies in Russian literature, quantitative linguistics, semiotics (Email: wangyongzju@163.com). **Yang Yiheng** is a Ph. D. student at School of International Studies, Zhejiang University. Her major fields of inquiry include knowledge engineering (Email: yvonneyoung@zju.edu.cn). **Zhang Kejun**, Ph. D., is Professor at College of Computer Science and Technology, Zhejiang University. His major fields of inquiry include Data Mining and Machine Learning (Email: zhangkejun@zju.edu.cn).

我国外国文学的译介与研究已有100多年的历史。大致可分为晚清至新中国成立、新中国成立至改革开放、改革开放至今三个阶段。第一个阶段的外国文学发展，以译介为主，尤其是在新文化运动中，鲁迅、郭沫若、茅盾、梁实秋、卞之琳、穆旦等作家和翻译家译介了大量英、法、德、俄等世界文学经典，构成了“中国现代文学的有机组成部分”（陈众议 15）。新中国成立后，俄苏文学的译介与研究构成了外国文学研究的重点。改革开放后，西方文学思潮的涌入，促进了我国学者对西方文学及文论的译介与研究，外国文学研究呈现出多元化的发展趋势。

100多年来，外国文学研究成果丰硕。据不完全统计，文献总量大致为译著4万多部，专著5千多部，期刊论文近20万篇。这些文献散布在不同资源库中，研究者无法快速检索到全部结果，更无法了解外国文学研究的整体面貌。早在20年前，就有学者倡导利用互联网实现外国文学研究的现代化，认为将“传统的学术研究过程同信息技术结合，就可以大大提高研究工作的效率，拓宽研究视野，有力地推动和促进外国文学的研究”（童小念 聂珍钊 35）。在数字人文发展迅速的今天，构建中国外国文学研究索引（CFLSI）数据库（以下简称CFLSI数据库）可谓顺应时代之势，以解研究者之需。对国外

文学研究者而言，CFLSI数据库的研制目的不仅在于减轻其面对海量数据进行文献查阅的负担，更在于全面总结研究成果、掌握学术动态、把握学术研究前沿，满足其检索专题研究文献、科学分析专题研究成果、发掘研究课题的需求。因为数据库的研究方式既可以“运用数据建设和数字信息综合处理作为批评分析的形式”，还可以“分析文献学和其他元数据，探讨文学趋势”（霍伊特朗等 45）。

从目前我国对外国文学研究成果的数据分析论文来看，尚缺乏数据准确、科学性较强、研究结果较有价值的成果。在数据来源上，研究者或直接从知网及其他引文资源库上获取相关数据，或采用人工方法收集数据，前者的数据庞杂，专业性不够高；后者费时费力，只能做有限数据的分析。从研究问题看，较多的论文主要聚焦于对引文、作者及机构等信息的分析，较少涉及学术研究本身。从研究方法看，有的采用人工统计方法，统计数量有限，效率较低；有的采用CiteSpace做可视化分析，缺少专业性，分析结果也较为片面。

要开展外国文学数据统计分析研究，其数据通常须具备三个条件：数据较完整、数据有较高的准确度、数据有相关标注。而这，正是CFLSI数据库研制的目标。本文将阐述数据库研制过程中重点研究的三大问题：1）公共网络资源现状及其不足；2）CFLSI数据库功能及构成设计；3）CFLSI数据库的系统构架及其技术实现路径。

一、公共网络资源现状及其不足

大数据时代，研究者查找文献的首选方式是搜索电子资源数据库。目前，可以搜索到中国外国文学研究文献的数字资源不少，但大部分网站均为综合性数据库，搜索所需文献需要化较多时间。如1998年开通的龙源期刊网，虽然有各类人文期刊4200种，但外国文学研究成果甚至须从“学术研究”入口搜索，检索体验不佳；此外，该期刊网以大众期刊为主，不收录外国文学研究类的学术性期刊。2013年上线的NSSD（国家哲学社会科学学术期刊数据库）所收录的“文学”类期刊仅《当代外国文学》《国外文学》《外国文学评论》《外国文学研究》《文艺理论研究》《文学评论》《文艺研究》等11种；维普中文期刊服务平台须走“文学”类期刊入口，与外国文学有关的期刊仅45种，且只能检索到1989年后发表的论文；2000年开通的超星中文在线图书馆以及“读秀”可以搜索到较多的期刊论文和著作类信息，但没有方便的索引功能。

目前，受众面最广的是中国知网（CNKI）。CNKI以其便捷性、信息量大、文献检索功能强等优势得到了众多研究者的青睐。截至2022年8月11日，CNKI“世界文学”分类下的论文数已达207088篇，论文资源丰富。但知网的著作类成果收录工作刚刚起步不久，而译著尚未收录。此外，CNKI目前收

录的论文最早为1955年，且此后20年内的信息也不够完整。比如，《世界文学》创刊于1953年，但CNKI仅收录了该刊1959年后的论文。据笔者所知，CNKI正在不断提升其数据库的完整性和规模，但毕竟知网面对的是全国所有学科的研究资源，因此，“世界文学”类别的文献要补充完整，恐怕不是短时间内能够完成的。最后，尽管研究者可以通过CNKI的“关键词搜索”等方式便捷地获得相关数据及自动统计分析结果，但我们发现某些数据的准确性难以得到保证，甚至导致自动统计分析的结果有误。究其原因，或许在于CNKI的数据量极其庞大，即便持续开展优化和去噪工作，但不断扩大的数据规模¹，使得其校验和清洗工作越发困难、耗时。

知网可以导出的可视化分析有总体趋势、主要主题、次要主题、学科、研究层次、期刊、来源类别、作者、机构、基金等。但这些分析，无论是整体检索还是限定范围检索，其结果都存在不同程度的问题。

在整体检索方面，经本文作者校验，无论是总体趋势还是主题、期刊等，自动导出的可视化分析较难用于外国文学研究，某些数据甚至有误差²。如总体趋势分析，其自动生成的只有1983-2022年的可视化分析结果（图1），要看完整的趋势分析图，需要另行检索1955-1982（图2）；此外，按可视化分析结果所标数据计算，论文总数为209934篇，多于世界文学检索结果的207088篇，有两千多篇的误差。而这个总数中，有一定比例的论文并不属于外国文学研究，这点，我们在下文再做说明。

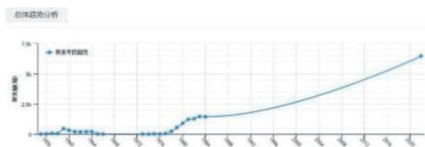


图1 1983-2022 总体趋势图

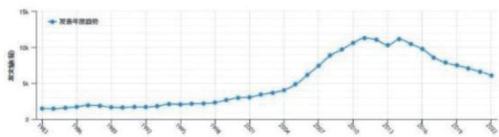


图2 1955-1982 总体趋势图

其次，期刊的可视化分析无法揭示期刊的重要性。分析结果显示，位列前10的期刊分别是《解放军外国语学院学报》《外语教学》《西南民族大学学报（人文社科版）》《外语研究》《西安外国语大学学报》《外语与外语教学》《天津外国语学院学报》《学术交流》《江西社会科学》《东北师大学报（哲学社会科学版）》，而作为外国文学研究重要期刊的《外国文学评论》《外国文学研究》《当代外国文学》《外国文学》《国外文学》《俄罗斯文艺》《外国文学动态研究》，无一进入前20榜单。（图3）

1 2018年8月18日检索的世界文学类期刊论文是185609篇，与本论文写作时检索到的207088篇相比，两年内增加了21479篇。

2 以下问题分析针对的数据，系本论文写作时的2022年8月11日采集，鉴于CNKI数据不断扩充，自动统计技术也有迭代，实时检索结果会有所不同。

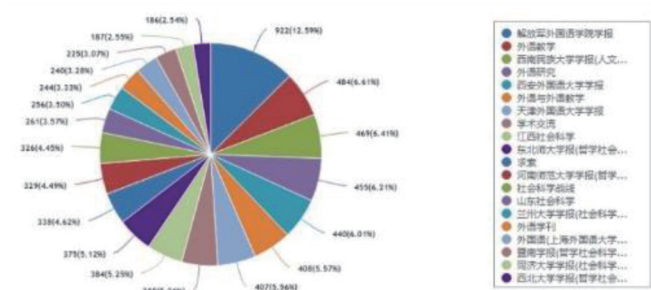


图3 期刊分布图

学科的可视化分析功能将部分非世界文学标签的论文也纳入统计，导致总比例中世界文学的占比减少了4.63%。（图4）这从另一个侧面说明知网数据库在准确度上有欠缺。

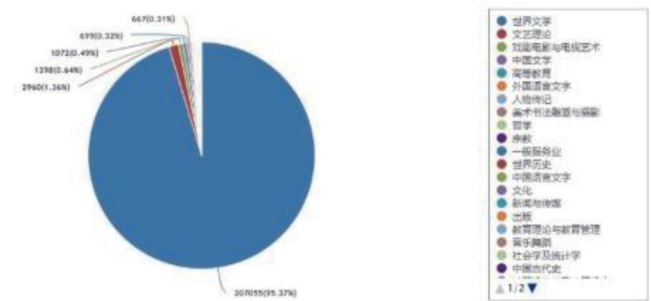


图4 学科分布图

研究层次的可视化分析无论是数据还是分类都不完整，所分析的数据总数仅7078篇，分别为：应用研究6423篇，应用基础研究562篇，学科教育教学78篇，开发研究-业务研究8篇，开发研究、应用研究-政策研究各2篇，工程与项目管理、技术研究-临床研究、技术研究各1篇。（图5）这些研究层次的分

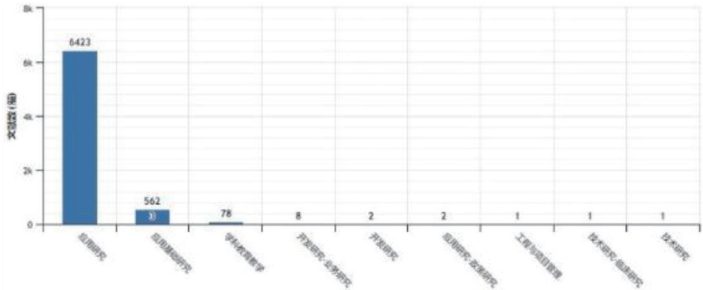


图5 研究层次分布图

与此相比, 知网在有限定条件的检索途径下, 可视化分析方面具有较高的可信度, 但在学科针对性上依然存在一些问题。

以陀思妥耶夫斯基研究为例。通过“世界文学”文献分类, 检索主题为“陀思妥耶夫斯基”的文献, 结果有1297篇。但在可视化分析中, 总体趋势分析仅统计了2003年以后的文献, 研究者无法了解到完整的总体趋势; 期刊分布图统计的论文数, 不及检索到的文献总数的一半。此外, 检索到的论文中, 有部分论文同主题基本无关或完全无关。比如: “《俄国短篇小说选》编者序”“邦达列夫答记者问”“斯拉夫主义: 西方视野中的普希金”“刘小枫教授与北京师范大学俄语系师生的座谈”“裹在‘外套’里的‘死魂灵’——读果戈理小说《外套》兼巴什玛奇金与维林比较”“俄罗斯作家协会贺词”“今日俄国屠格涅夫研究概观”“欧美作家论托尔斯泰——在上海纪念伟大作家逝世七十周年学术讨论会上的发言”“在鲁艺的《名著选读》讲授提纲(三)”“一幅色彩斑斓的‘马赛克镶嵌画’——试评《静静的顿河》的叙事结构”“屠格涅夫研究简论”“编后记”等。因此, 可视化分析结果也难以为该专题研究发挥其应有的作用。

综上所述, 现有的电子资源数据库难以满足外国文学研究的需求。主要原因在于: 1) 文献分散且不完整。论文类数据虽多, 但缺少早期的研究成果; 专著类文献较少, 且散布在不同资源库中; 译著类资源缺乏, 这不能不说是外国文学研究资源的明显不足; 2) 可视化分析模型针对所有学科, 尤其偏重应用研究, 对于外国文学的整体研究而言并不适用; 3) 后台缺乏适用于外国文学领域的人工智能识别模式, 只注重字面的匹配, 忽略了语义层面的检索, 导致提取数据既不完整, 也会出现不少错误, 最终使得可视化分析结果缺乏科学性, 无法显示实际研究情况。

二、CFLSI 数据库功能及构成设计

如前所述, 在并非所有研究成果都数字化的情况下, 专业领域的研究者要掌握研究现状, 尚无便捷且有效的途径, 仍须花费较多的时间多渠道检索并进行人工统计。因此, CFLSI数据库的数据构成及功能设计须从外国文学研究的需求出发, 以数据库的完整性及功能的针对性为研制目标。

作为以外国文学研究者为使用对象的数据库, CFLSI的功能应主要为研究所需掌握的研究现状提供依据。研究者检索数据库通常有以下需求: 1) 整体文献检索; 2) 优质文献检索; 3) 把握研究动态。据此, CFLSI在分析现有数据库的优点与不足的基础上设计主要功能。最终目标在于不仅为研究者提供全面系统的文献信息, 成为其文献搜索的便捷而可靠的途径, 还能通过统计分析把握不同研究方向的研究趋势, 挖掘有价值的研究问题。

为服务第一个需求, 建立的数据库须尽可能全面完整。因此, CFLSI数据库可以检索到的文献类别应涵盖译著、专著、论文三大类, 时间跨度自晚清

至今，并逐年更新。

第二个需求则要求数据库可以从中析出优质文献。在数据统计过程中，研究团队发现，有几种期刊发文量惊人，如《名作欣赏》有七千多篇，《文学教育》《安徽文学》均有三千多篇。究其原因，在于这些期刊或者刊期较多，如《名作欣赏》是旬刊，《文学教育》虽标明是月刊，却分为（上）（中）（下）；或者每期刊发的文章数量多，这就意味着每篇论文的字数偏少，比如《安徽文学》（下半月刊）有很多论文都只有2页。这些期刊上的论文对某些读者或许有益，但对外国文学研究而言，显然难以称其为优质文献。为此，数据库单独划分并建立了优质文献数据库子库，包括《外国文学评论》《外国文学研究》《当代外国文学》《外国文学》《国外文学》《俄罗斯文艺》六大期刊¹，以及《文学跨学科研究》（A&HCI）和《世界文学研究论坛》（ESCI）。

第三种需求在外国文学研究中最重要，可以为研究者把握研究动态服务，但也是现有资源库中最为缺少的功能，是研究团队重点攻关的难点。与研究动态相关的信息较多，目前的开放数据库，这方面功能最丰富的依然是知网。在数据库检索界面上，可以查到篇名、作者、刊名、发表时间、被引、下载等信息，借助数据库自带的分析器还可以获取总体趋势、主题、作者、机构、期刊等方面的可视化分析结果（参见本文第一部分）。如果仅选取部分论文，还可以获取关系网络，如文献互引网络、关键词共现网络、作者合作网络等。

然而，从外国文学研究文献的研究目标来看，上述信息的分析固然不可缺少，但更重要的是与论文研究选题相关的信息分析。比如文学史研究，需要按国别检索各国文学史上主要作家作品的研究情况；作家研究需要检索该作家被研究了哪些作品，采用了哪些理论和方法；作品研究需要检索该作品研究有哪些视角等；而国别、作家、理论和方法，也是趋势研究、热点研究依据的重要信息。这些信息，仅靠自动提取的主题误差率较大（如前文提到的陀思妥耶夫斯基研究成果的可视化分析）。因此，不仅要对关键词和摘要进行主题词提取，还须对主题词做细化处理。

综上所述，针对外国文学研究而研制的CFLSI数据库，既要具有论文篇名、作者、期刊等基础数据的查询功能，也要具备可以从中挖掘国别、作家、作品、理论和方法等研究信息的功能。这种功能设计也决定了数据库的整体构成。

为了满足数据库发挥上述功能的需求，其整体构成设计分为总库（所有相关研究成果）以及论文子库（含优质文献数据库子库）、专著子库和译著子库；按三大类别文献分别设计信息标签；建成一个具有可视化统计分析功能的数据库系统。

1 未来可根据实际情况添加学界认可度较高的期刊和辑刊。

三大类别文献信息标签设置如下：



与知网相比，除了基本信息，三大类文献均设置了获奖标签，可以从一个角度说明研究成果的质量。论文部分设置国别、作家、作品、理论和方法标签，有助于研究者较准确地把握国别文学、作家作品的研究动态，以及文论和方法方面的发展趋势。专著和译著部分，还设置了书评和网评标签。专家书评来源于公开发表的文章或相关著作中的评价，具有较高的权威性；网评则主要采用豆瓣等社区网站上的评价，可显示公众的欢迎度。

系统构成的设计，还考虑了其可实现度。在CFLSI数据库的研制过程中，我们认识到，数据的人工校对依然是必不可少的。虽然在大数据条件下，概率比精确度更为重要，但这有其相对性，“标注错误和不精确不会充分降低其他诸多语境中的信息提取，但对文学就会”（笄章难 38）。在外国文学研究成果的数据处理问题上，我们比较认同的观点是，“纷繁而小有瑕疵的大数据所得出的结论较之无瑕疵的小数据得出的结论更为可靠和科学”（郑永晓 142）。在数据规模较大的情况下，做到无瑕疵是不可能的。但这并不等于完全依靠机器操作，因此，需要借助人工校对尽可能减少瑕疵。

三、CFLSI 数据库的系统构架及其技术实现路径

根据CFLSI数据库的整体设计，系统构架分为前端界面层、数据层和系统层三大模块。图6展示了这三大模块的内部构成及其技术实现路径。

如图所示，CFLSI数据库系统的功能和构架由界面层、数据层及系统层组成。

界面层作为外国文学数据库与用户交互的服务界面，围绕数据库的核心需求进行功能细化，即多条件组合检索、信息浏览、可视化等。基于Vue前端框架，可为外国文学研究者提供分类浏览、多维度查询和深度检索外国文学

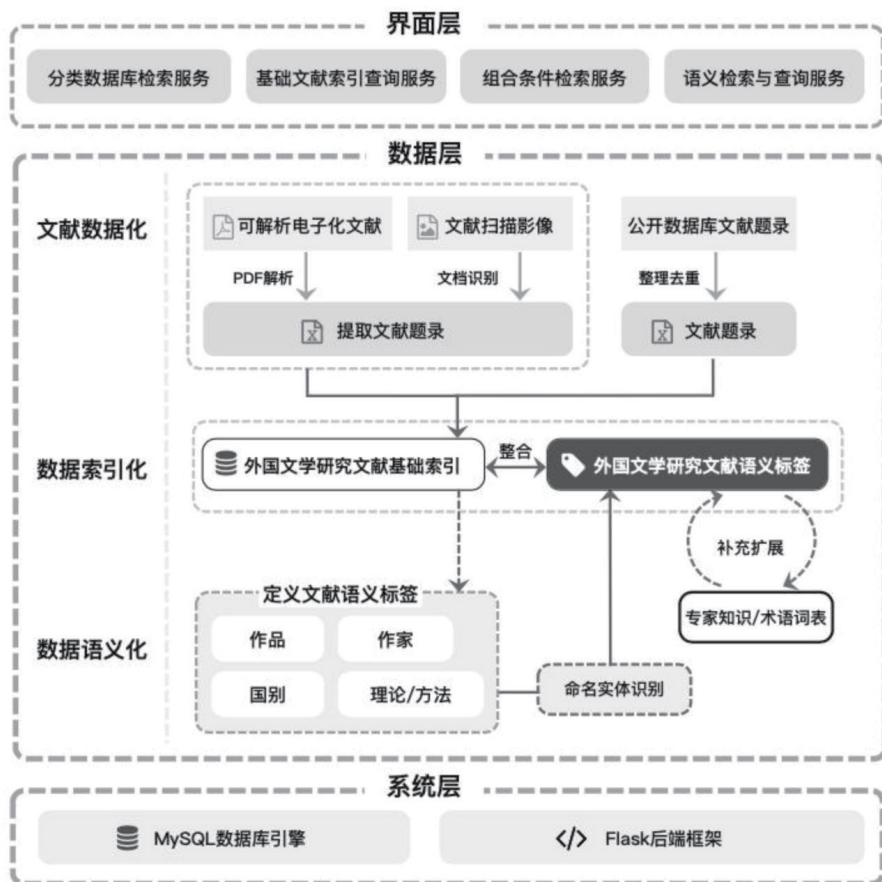
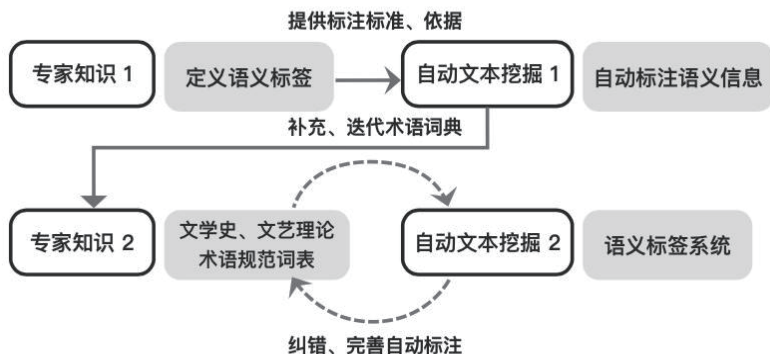


图6 CFLSI 数据库构架及技术方案

研究数据的功能。此外，本系统还设计了众包形式的数据校对功能，允许使用者在浏览数据的同时，进行纠误和补充。这不仅有助于不断提高数据的准确度，也符合数字人文发展的趋势，因为“数字人文平台还有一个特质是要利用众包让用户参与到系统的建设中来”（刘圣婴等 20）。

在数据层，CFLSI数据库的数据涉及结构化文本和非结构化文本，前者较易获取和处理，形成外国文学研究文献的基础索引；后者需要采用数据挖掘技术甚至人工辅助处理，即对“海量的、杂乱的、无序的、非结构性的”数据，“通过挖掘，寻绎出数据中间隐含的、先前未知的并有潜在价值的信息”（刘京臣 184）。经自动文本挖掘技术与专家知识相融合，即将计算机智能与人工编纂的学科领域术语、叙词表相结合，可以最大限度地发掘研究文献中高价值的信息。因此，我们的工作遵循从专家知识到自动文本挖掘，再返回专家知识的原则，既实现了专家知识的不断迭代、更新，也不断提升自动文本挖掘准确度，具体流程如下：



经过多次迭代，数据库系统才能得以不断完善。

系统层作为整体架构的底层支撑，不仅负责数据的维护、管理，也处理用户请求、访问数据库并返回数据等功能。为了能满足资料查询和数据、体系结构不断迭代的需求，我们采用了技术架构和内容架构分离的设计，后端基于MySQL数据库引擎和Flask框架为核心技术栈开发实现。该技术栈的优势在于支持数据双向数据绑定、组件化、开发快且轻量级，有利于外国文学研究数据库系统的不断优化和迭代。

在这三大模块中，数据层是数据库研制的核心。在数据库采集阶段，通过网络采集题录信息和直接挖掘本地已有文献的方式，尽可能全面、准确地覆盖研究文献基础信息的对应字段。研究团队通过四年时间，汇聚了大量公开的文献原始数据，包括以题录形式存储在公开数据库的数据和可解析的PDF文件形式或影印图像形式的原文献，后者因年代久远往往需要数字化解析。迄今为止，数据库已收集并解析的数据约为论文类15万余条、专著类2千余条、译著类2万余条。

为实现对多种复杂格式电子资源的数据整合、数据标准化和数据检索，我们将数据层的主要任务分解为文献数据化、数据索引化及数据语义化三项。

首先是文献的数据化，即将文档转换为机器能够识别、分析及计算的数据。这是实现文档智能，即辅助人类自动阅读及理解分析的首要步骤。因人工提取处理十分耗时耗力，研究团队设计了一套多层次渐进的、逐步技术融合的半自动化提取流程以实现外国文学研究文献的数据化。

此方法将待进行数字化的文献按照信息的完整程度分为三类，分别是：

- 1) 直接收集到的文献题录信息，这些信息往往已大致具备结构化的要求；
- 2) 能够直接解析的PDF格式文献数据；
- 3) 布局复杂、多为影印资料，无法直接解析的图像格式文献数据。

对于以上三种数据，我们采取了三种不同的处理策略和流程。第一种格

式的数据信息,我们对获取的数据错误和缺失的作者单位、关键词、摘要等信息进行统一处理,对照原文献修改讹处,尽可能补充缺失信息。对于第二种数据格式,先通过PDF解析将其读取为可处理的文本文件,再借助正则表达式等文本模式匹配方法提取信息;对于难以获取匹配标识的文件,则将原文件转为图像后,通过OpenCV库将图像作二值化处理,运用投影计算方法定位和分割包含关键信息的固定区域图像,最后利用OCR识别题录信息。对于第三种数据,鉴于其多为早期影印资料,文字图像模糊、排版特殊,难以直接进行分析、识别和处理,我们基于整理归纳出的十余种常见论文布局模板,利用文本注释工具,选取一定数量的具有代表性布局的文献并对实体边界加以标注。然后,借助计算机深度学习框架PyTorch完成版面分析和信息提取,采用多模态文档智能模型识别图像文档的关键信息。最后,对上述流程处理后的数据进行清洗,如去重、勘误等,不断补充、更新和完善数据库。

其次是数据的索引化,索引构建是搜索引擎的核心部分,其目的是将预处理后的文本组织成一种易于检索的数据结构。具体到外国文学研究数据库的索引,就是以同一标准描述文献数据的基本信息,并将语义信息映射到关系数据库里的概念关系中。根据这一目的,在三类信息标签的设计基础之上,对文献基础信息字段进行了标准化定义。以论文数据子库为例,索引字段包括:标题、作者、作者单位、关键词、摘要、期刊、年份、发表时间及中图法分类号等,针对关键字段创建索引并存储到数据库中,建立数据库基础数据查询表,并结合不同的文献类别构建多个关系型数据表。

数据层的第三项任务是研究文献的语义化路径。索引化的数据可以满足基础检索、查找和引证需求,但“还不是结构化的能进行统计分析和再生知识的数据库”(刘石 30)。为了让数据库更具研究价值,需要发挥“精细化语料库的功能,使其从‘字联网’形态进入到更深层的‘意联’脉络”(刘石 26-27),析取文献中的高维实体语义信息,梳理百年间中国外国文学研究发展演变轨迹。本文借助知识抽取等技术,从文献中提取出国别区域、作家作品、研究主题、文艺理论、研究方法等独有的特征信息,便于进行具体而为的研究。首先,根据学科领域定义的语义标签对文献关键字段进行命名实体识别,从中抽取出概念实体和属性。然后,通过主要国家的作家和作品表、文艺理论关键词表构建领域术语词典,以此为依据,运用实体对齐、自动与人工消歧结合的方式对提取到的信息加以补充和修改,并将识别到的未登录的新术语加入术语词典中。最终,将抽取到的字段作为高维语义信息与基础信息组合,由这些语义网络组成外国文学研究数据库的全部检索标签系统。

结语

如今,在人文学科加强数据库建设,开展基于大数据的相关研究这一观念,正逐渐被越来越多的人文学者所接受。近期国内外热议的ChatGPT,借助

大规模数据预训练和生成的方式，能够对多种领域的问题给出较合理的答案，甚至帮助用户自动撰写文章摘要和论文大纲等。这展示了海量数据在自然语言处理领域的巨大潜力，也让人们更深刻地认识到人工智能给传统型学习和研究带来的挑战与机遇。CFLSI数据库的研制在这条道路上迈出了第一步，对数字技术与外国文学学科知识的深度融合做了探索与实践。研制过程中，在借鉴其他数据库的模式和经验基础上，结合外国文学学科自身的特点设计数据库功能和系统构架。受技术手段和人力、物力、财力的限制，目前数据库已经实现的基本功能存在诸多不足，诸如精确语义检索、文献全文阅读等功能距离真正落地还需要凭借强有力的技术团队支撑，有待今后继续完善。我们期待，所建成的数据库不仅可以为外国文学研究者服务，还将在未来有助于研究者认识数字人文的目标及其本质，推进外国文学研究的智能化进程。

Works Cited

- 陈众议编：《当代中国外国文学研究（1949-2009）》。北京：中国社会科学出版社，2011年。
[Chen Zhongyi, ed. *Foreign Literature Studies in Contemporary China (1949-2009)*. Beijing: China Social Science Press, 2011.]
- 笪章难：“以计算的方法反对计算文学研究”，汪衡译，《山东社会科学》8（2019）：24-39。
[Da Zhangnan. “The Computational Case against Computational Literary Studies.” Translated by Wang Heng. *Shandong Social Sciences* 8 (2019): 24-39.]
- 霍伊特朗等：“推进计算文学研究——对笪章难《以计算的方法反对计算文学研究》一文的讨论”，汪衡译，《山东社会科学》8（2019）：40-55。
[Hoyt Long, et al.. “Advancing Computational Literary Studies: A Discussion of Nan Z. Da’s Article ‘The Computational Case against Computational Literary Studies’.” Translated by Wang Heng. *Shandong Social Sciences* 8 (2019): 40-55.]
- 刘京臣：“大数据时代的古典文学研究——以数据分析、数据挖掘与图像检索为中心”，《文学遗产》3（2015）：182-190。
[Liu Jingchen. “Classical Literature Research in the Era of Big Data: Centered on Data Analysis, Data Mining and Image Retrieval.” *Literary Heritage* 3 (2015): 182-190.]
- 刘圣婴等：“数字人文的研究范式与平台建设”，《图书情报知识》1（2022）：6-29。
[Liu Shengying, et al. “Research Paradigm and Platform Construction of Digital Humanities.” *Documentation, Information & Knowledge* 1 (2022): 6-29.]
- 刘石：“大数据技术与古代文学经典文本分析研究”《数字人文》0（2020）：24-31。
[Liu Shi. “Study of Big Data Technology and Text Analysis of Classical Ancient Literature.” *Digital Humanities* 0 (2020): 24-31.]
- 童小念、聂珍钊：“互联网与外国文学研究的现代化”，《外国文学研究》1（2003）：32-37。
[Tong Xiaonian and Nie Zhenzhao. “Internet and the Modernization of Foreign Literature Studies.” *Foreign Literature Studies* 1 (2003): 32-37.]