数字人文视域中的古籍文本标注方法研究——以MARKUS为例

于亚秀[1]，李欣[2]
1. 华东师范大学图书馆，上海 200062；
2. 华东师范大学数据科学与工程学院，上海 200062

摘要
文本标注是文本分析挖掘中的重要一步，面对大规模古籍资源，人工标注无法满足人文研究需求，且古籍语法结构和语言特点特殊，现代文本标注技术很难直接用于古籍研究。在分析人文研究者进行古籍文本标注中面临的难点和痛点的基础上，提出普适性的古籍标注标准流程，给出基于MARKUS的文本标注模型，并通过具体实践，探索基于该模型的古籍文本标注方法，旨在助推借助数字人文工具改变古籍人文研究方式，拓宽研究规模的应用深度。

关键词
数字人文；古籍；文本标注；MARKUS

中图分类号：TP391.1，G255.1　　文献标志码：A　　doi: 10.11959/j.issn.2096-0271.2022046

Research on text annotation method of ancient works from the perspective of digital humanities: a case study on MARKUS

YU Yaxiu[1], LI Xin[2]
1. East China Normal University Library, Shanghai 200062, China
2. School of Data Science and Engineering, East China Normal University, Shanghai 200062, China

Abstract
Text annotation is an important step in text analysis and mining. Manual labeling can no longer meet the needs of humanistic research faced with large-scale text resources, and due to the special grammatical structure and language characteristics of ancient works, the text annotation technology on modern corpora cannot be directly applied to the ancient works. Based on the analysis of the challenges faced by humanities researchers, a universal standard text annotation process of ancient works was proposed, and a model based on MARKUS was given. And ancient works

注：为保证本文研究的真实性和可信性，文中保留涉及原文内容的繁体字。

annotation method based on this model through specific example was explored, to promote using tools to change the research methods in digital humanities and to expand the scale of research.

Key words
digital humanities, ancient works, text annotation, MARKUS

0 引言

“古”是相对于“今”而言的，未采用现代印刷技术印制的书籍，皆可被称为古籍[1]。近些年，随着数字化项目的推进，大量的古籍数字化成果被推出[2]，除了鼎秀古籍全文平台、中国基本古籍库、中华经典古籍库等商业数据库外，还有两个线上开放平台：中国哲学书电子化计划平台和维基文库。线上开放的中国哲学书电子化计划平台收藏的文本已超过3万部，有50亿字之多，涵盖先秦两汉、汉代之后的众多资料，是历代中文文献资料库最大者。维基文库典籍和史书目录中也涵盖了经史子集等众多古籍数字化资源。借助互联网、大数据带来的便利，学者得以通过互联网查询到以往不易获得的海量文献。

数字人文视域中，数据驱动在人文学科领域的研究范式越来越多地被采用，人文学者运用计算机技术在海量文本中发现更多新材料与新问题。柯平等人[3]以Web of Science核心数据集为来源，通过文献计量法分析数字人文的研究热点，其中文本挖掘是数字人文的实践前沿之一，文本标注又是文本挖掘过程中重要的环节[4]。一个好的模型需要质量优异的数据资源做支撑，数据标注是各种算法得以有效运行的关键环节，数据标注越准确、标注的数据量越大，算法的性能就越好，能被发现的知识就越多。但在古籍研究中，大部分古籍研究者的文科背景及古籍数据的特殊性对人文学者的古籍文本标注工作提出了挑战。

古籍研究普遍面临文本量大、标注精度高、存在计算机技术短板的难点和痛点。而借助普适性的数据标注工具，可以提高标注效率、降低技术门槛，满足人文学者进行多种计量分析的研究需要。因此，本文通过对古籍的特点及常见标注需求进行分析，结合MARKUS的功能及特点，提出普适性的古籍标注标准模型，探索基于MARKUS进行古籍标注的方法，助推借助数字人文工具改变古籍人文研究方式，拓宽研究规模的应用深度。

1 古籍数据

1.1 古籍的数字化现状

在古籍数字化项目的推动下，古籍电子化的程度越来越高，各种大型的网络版、单机版的电子古籍检索系统被开发出来。以袁行霈主编的《中国文学史》中所提到的古籍为例，其电子文本大部分可以在互联网上找到。在古籍文本资源的数据库构建方面，国内外已经有了一些重要的研究工作，古籍文本的电子化已经取得了许多重大成果，古籍数字化数据库见**表1**。此外，随着近些年人工智能的发展和深度学习的兴起，学术界开展了与古籍光学字符识别（optical character recognition，OCR）相关的众多的科研实践[5]，古籍OCR的准确率有显著提升。基于人工智能的OCR技术及基于互联网协作的古籍数字化

表1 古籍数字化数据库

类型	数据库
综合类全文古籍数据库	中国基本古籍库、中华经典古籍库、四部丛刊在线全文检索、四库全书在线全文检索、二十五史全文检索系统、二十五史研习、汉籍电子文献资料库、鼎秀古籍全文平台……
专题类古籍数据库	历代宝案脉络分析系统、台湾大学佛学数位图书馆、朝鲜王朝实录、本草经集注……
语言类数据库	韵典网、古音小镜、汉字古今中外读音查询、搜韵网、东方语言学：方言字音查询、台湾闽南语常用词辞典、吴音小字典……
文字类数据库	数字说文、国学大师、说文解字综合检索、汉语多功能字库、先秦甲骨金文简牍词汇资料库、引得市、小学堂、开放古文字字形库、楚简帛字典：清华篇、战国楚简帛文字典型形体检索系统、中国古代简帛字形、辞例数据库、拓本文字资料库、汉字辞书……
版本目录类数据库	中国古籍总目在线检索、日本所藏中文古籍数据库、四库提要在线检索……

工作平台陆续被推出，如：如是古籍数字化平台，中文古籍OCR等在线开放平台，汉王、深延科技古籍字识别软件，ABBYY、Calamari等国内外商用软件。这些平台极大地方便了人文学者自主进行古籍图片的OCR，进一步方便人文学者获取古籍文本数据。

1.2 古籍权威语料库

在古籍研究中，人名、官职、地理、物质等信息的标注是人文研究者常见的标注任务。近10年来，为了满足汉语研究的需要，学术界开始尝试建设深加工的标注型语料库，具体工作包括对古代文献进行词语切分，并添加词性、义项、语法地位标注等多方面信息，形成的语料库可以辅助人文研究进行分词和标注。例如，中国历代人物传记资料库（China biographical database，CBDB）目前已系统性地收入中国历史上重要的传记资料，并将其内容毫无限制地、免费地公诸学术之用，截至2021年12月，该资料库共收录515 488人的传记资料，这些人物主要出自7世纪至19世纪，目前该资料库正致力于增录更多唐代和明清的人物传记资料。中国历史地理信息系统（China historical GIS，CHGIS）是一个免费的中国朝代地名和历史行政单位数据库，旨在通过建立连续的时间序列描述地名、行政建制和其他基础地理要素随时间的变化，记录中国从公元前221年至1911年人口稠密地和历史行政单位，并为用户提供按不同历史时期进行查询、检索和展示等功能。法鼓文理学院的佛学规范资料库整合已完成的和进行中的各专案人物与地点资料，建立时间、地点、人物与佛经目录4个规范资料库，并建立历史对照年表，面向互联网开放资源共享，供佛学研究专家进行项目研究。

随着古典文献的数字化程度不断刷新历史新高，学者似乎没有理由抱怨资料不足或者难以获取之类的问题。与此同时，数字人文学科的发展为古籍的研究带来新的方法，技术驱动研究、数据驱动研究已成为人文学科发展的一个方向[6]，不以传统的近距离方式阅读文本，而是采用远距离阅读，对大规模人文资源进行定量分析挖掘，借助统计、图表、地图等方式让文本数据形象化地成为数字人文的研究热点[7]。一个纯文字的文档对于计算机而言，是一个个平面的数值，但对于使用者来说，一篇文本中可能包含的人、事、时、地、物，甚至更多不同性质的词汇都代表不同的意义。在古籍研究领域，对大规模古籍资源进行定量分析挖掘也已成为古籍研究中的一种新的、行之有效的办法。文本标记

是研究过程中的关键一步，其利用自然语言处理（natural language processing，NLP）技术，将古籍文本数据以自动或半自动的方式进行分词、词性标注、命名实体（常见有人名、地名、机构名）识别、关系抽取、主题建模等，从而挖掘并展示古籍文本中的潜在特征和语义信息。

2 古籍文本特征和标注面临的挑战

数据特征方面，综合分析苏祺等人[8]、谢韬[9]关于古籍数字化及识别的相关文献，分析古籍文本相对于现代汉语的不同，主要体现在现代汉语分词中已有比较通用的分词标准，如MSRA标准、CTB标准、PKU标准等，并有相应的语料库，而在古籍方面尚缺乏统一的分词标准。此外，在古籍中，单体字有比在现代汉语中更加丰富的意义，从字符组合中定义“词”更加模糊。以上因素都导致现有的现代文语料上的分词标注技术无法直接应用于古籍语料，古籍分词比现代汉语更难以定义和实现。古代汉语的分词处理尚处于探索、验证阶段，国内外学者也对古籍文本分析进行了诸多研究。马海丽等人[10]对古籍的分词标注现状进行了详细分析。2014年钱智勇等人[11]将隐马尔可夫模型（hidden Markov model，HMM）应用于《楚辞》的分词标注，实验证明这种方法是可行的，但是对词性的标注方面还有待提升。2021年张琪等人[12]提出了面向多领域先秦典籍的分词词性一体化自动标注模型，实验结果准确率达到88.97%，准确度有待提高。上述研究都未能提出成熟的普适性模型，缺乏通用性。

实践方面，传统的古籍研究主要依靠人文学者博闻强识才能进行更精准的人工标注。尽管近10年NLP技术有了飞速发展，但具体到古籍领域，可用的NLP工具并不多，目前有针对古汉语处理的NLP工具包“甲言”，以及由Yasuoka K[13]开发的另一款名为UD-Kanbun的工具包。此外，邢付贵等人[14]于2021年提出了基于互联网大规模古籍语料构建古文基础词典的分词技术。但这些工具包和技术都需要一定的编程基础，对于大多数人文研究者而言，计算机技术是短板，能使用Java、Python、R等进阶的编程语言进行数字人文研究的学者只是少数，大多数人文学者不能通过编程解决文本标注问题。

虽然古籍以单字词为主，但是多字词仍占了相当大的比例，官职、人名、地名、年代等均存在大量的多字词，而名词、动词、人名、年代等细类区分的词类标注，对于古籍的研究有重要意义[15]，人以及各种人和人、人和地、人和官职等之间的关联研究，在历史研究中经常出现。目前在人名、年代、官名等方面已有成熟的语料库，普适性的工具和模型可以帮助古籍人文研究者应对挑战，改进研究方式，扩展研究规模。

3 基于MARKUS的古籍文本标注模型

人文研究者进行古籍文本标注的标准流程如**图1**所示。文本标注首先从数据采集开始，采集图片、文本等多种类型的数据，然后对采集的数据进行预处理，以获得高质量数据，清洗不合格的数据，减少无意义的标注工作，提高标注效率。随后制定标注目标，选择标注工具和标注方式，对清洗后的数据进行标注。

在数据标注阶段，针对不同的标注任务，选择不同的标注方法。对于已有成熟的全自动标注模型或者自行训练的模型，可采用全自动标注方式；半自动标注方式

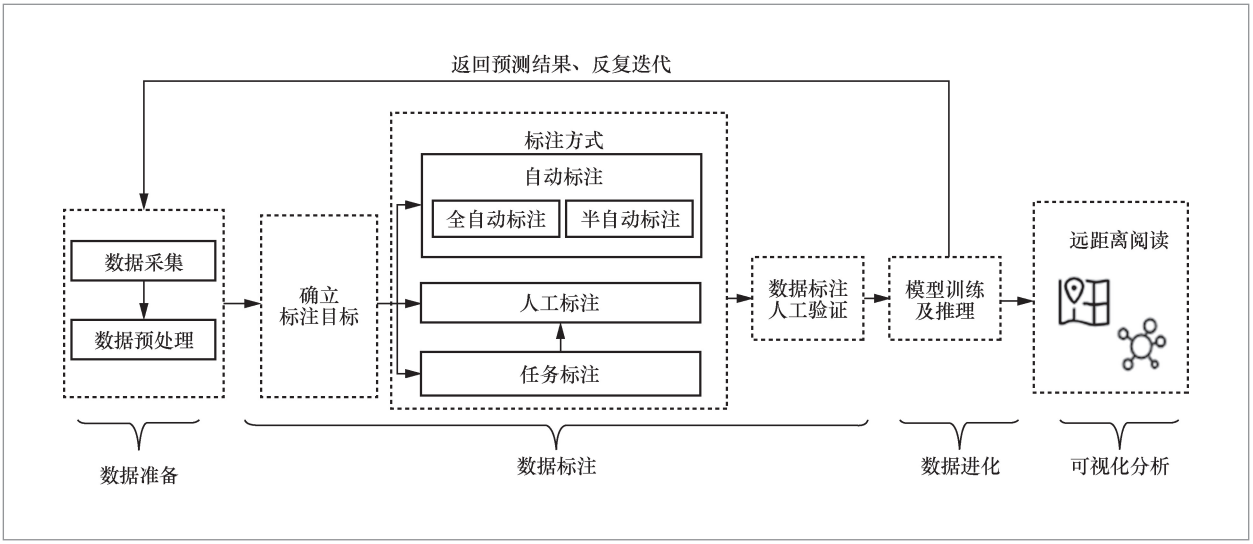


图1 古籍文本标注标准流程

则用于处理没有特定语料库或模型的标注任务，预先标注一定数量的文本，由系统基于已标注的数据进行训练，而后对剩余文本进行标注；对于标注数据量较少的标注任务，可采用人工标注方式；任务标注适用于有大量标注数据的情况，管理员将标注数据拆分，分给标注员，以减少单人标注的任务量，进而提升标注效率。

面对不同的标注任务及古籍人文研究者的技术短板，普适性的标注工具和模型尤为重要，可以帮助古籍人文研究者提高标注效率，扩展研究规模，发现新的知识。MARKUS 是数字人文浪潮下的一款具有很强应用性、普适性的古籍标注工具，其正式名称为“古籍半自动标记平台”（中文译名：码库思），由荷兰莱顿大学魏希德（Hilde De Weerdt）教授与何浩洋博士开发设计。MARKUS通过关联多个权威语料库实现古籍中历史人名、地名、官名与时间等实体的自动标注，也为研究者提供通过定义关键词列表、上下文中的关键词、正则表达式等进行半自动标注的方式，且支持以txt、excel、html 格式输出标注结果，以便做进一步分析。基于MARKUS的古籍文本标注模型如**图2**所示。

4 案例实践及分析

地方志是记载地方的建筑、地理、历史、人物、自然生态以及产业等资讯的著作，是研究地方史的重要参考资料。本文以《民國郫縣誌》为例，探讨MARKUS如何助力人文研究者进行古籍文本标注，改变古籍人文研究方式，扩大古籍人文研究规模。

设定标注任务为标注《民國郫縣誌》中的日期、官名、人名信息，以及寺庙及寺庙的朝代信息。在MARKUS中，把“民國郫縣誌.TXT”文件导入文本分析工具，通过MARKUS自动标注功能对日期、地名等信息进行快速标注，基于CBDB标记人名（全名、字号）及官名，基于中国历史地理信息系统及台湾历史地名资料库（Taiwan GIS，TWGIS）标记地名，基于法鼓文理学院的佛学规范资料库中的时间规范资料库

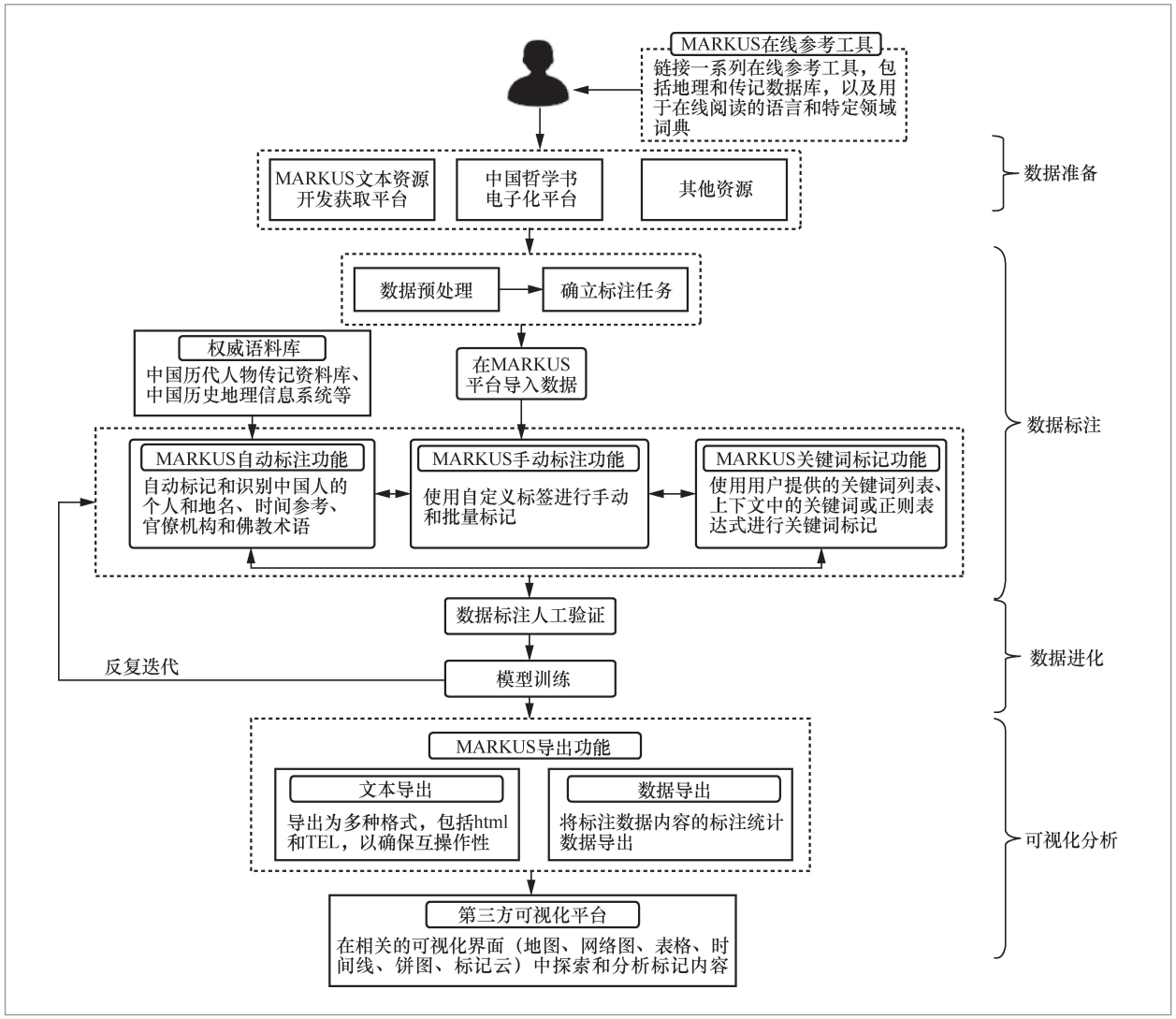


图 2 基于 MARKUS 的古籍文本标注模型

标记日期。自动标注可在短短几分钟内完成标注目标，如**图3**所示，并产生**表2**所示的标注数量。

在自动标注的基础上，使用关键词标注功能补充搜索标注《民國郸縣誌》中出现的寺庙及朝代信息。通过关键词助手定义关键词规则，系统自动进行全文检索分析查找关键词，然后利用词夹子功能反复人工扩充种子、重启演算法补充朝代信息。MARKUS关键词助手和词夹子如**图4**所示。最后基于关键词助手分析词夹子得到的词汇，进行人工筛选后定义关键词、标记名称以及字体颜色，载入正则表达式，参与全文检索标记，最终显示全文标注结果，如**图5**所示。将标注结果保存导出生成报表，具体见**表3**，通过可视化工具可直观地分析《民國郸縣誌》中寺庙的朝代分布。

在MARKUS项目论坛中也有部分古籍人文研究实践案例，如犹他大学副教授玛格丽特·万（Margaret Wan）在进行的中国小说研究项目。Margaret Wan发

Passage7

郸縣志序 三 朱[鼎臣] 清嘉慶十七年

Passage8

郸之有志創始於前邑令李君馨搜羅纂輯集有成書厥功甚偉是年 星離於重光志之後序成於玄默今上御宇之十有六年歲星復次重光適四川[總督]志勇巴圖魯[長白]常公奏修省志頒示格式下府州縣令照依塡註成册賫省備採録者其明年歲次玄默[鼎臣]奉檄署郸邑既下車詢民疾苦去其已甚而勸其不能復爲之編連保甲以安之釐止書院膏火以振興之暇時屬其紳耆父老以講求之郸雖小邑而土膏沃衍四野青葱士喜詩書家多饒裕予蒞任數月既善其風俗之醇而其吏民亦多安予之拙也檢閱舊志始知江漢炳靈何君公之所棲宅揚子雲之所釣遊其急公尚義文教彬蔚廉頑立懦蓋有自來洵乎賢人君子之所遊處其薰德而善良者風教固殊焉既爲删其繁複訂其錯誤凡自辛未以後續而輯之其先後次叙悉依省頒格式至予學植荒落謝陋寡聞後之君子網羅散失補苴罅漏綿於無盡是余之厚幸也夫郸故水鄉邑以水得名志之創纂以壬申辛未茲之續纂亦

图 3 使用 MARKUS 自动标注《民國郸縣誌》标注展示

表 2 使用 MARKUS 自动标注《民國郸縣誌》统计结果

序号	标记内容	数量/个	序号	标记内容	数量/个
1	文章分段	853	4	官名	3 070
2	日期	2 872	5	法鼓佛学人名规范资料	6 634
3	地名	3 877	6	法鼓佛学词汇	4 276

图 4 MARKUS 关键词助手和词夹子

现，使用传统方法大规模地标注并在地图上绘制小说中提到的地名非常困难，通过MARKUS可以很好地概述特定小说中提到的所有地方，并以《百家公案》为例，进行了基于MARKUS的研究，使用MARKUS自动标注《百家公案》中的地名，加上后续人工更正，最终的标注结果通过QGIS可视化。通过此研究，Margaret Wan也得到了启发，打算继续用MARKUS来调查数百部中国传统小说中提到的地理空间情况，以揭示大量被忽视的小说，并认为像MARKUS这样的数字工具可以扩

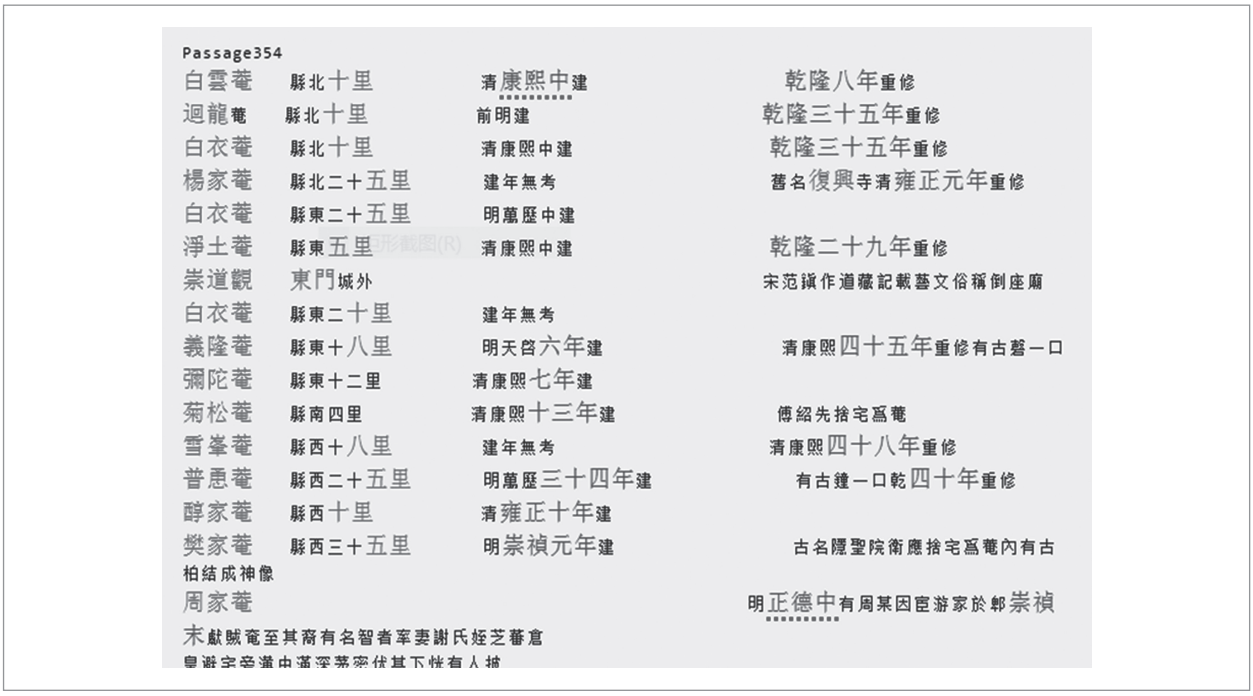


图 5 使用 MARKUS 正则表达式进行《民國郫縣誌》全文标注

表 3 《民國郫縣誌》中寺庙朝代分布

朝代	朝代	数量/个	朝代	朝代	数量/个
宋	宋景德	1	元	元始正	1
明	明万历	17		元	1
	明嘉靖	2	清	清康熙	17
	明崇祯	4		清雍正	5
	明正统	2		清乾隆	1
	明弘治	1		清顺治	1
	前明	3		清咸丰	1
	明正德	2		清光绪	2
	明天啓	1		清嘉庆	1
建年无考	建年无考	29		清初	1

大人文研究的研究规模。

MARKUS借助CBDB、CHGIS等成熟语料库实现精准标注历史人物、年代、官名、地名等信息，关键词助手和词夹子功能又增加了标注的灵活性，降低了古籍人文研究者技术门槛，节省了大量人力。此外，通过与语料库的实时关联，可拓宽人工校对和标注时参考资料的范围，提高人工矫正及标注效率。以案例中的《民國鄆縣誌》标注任务为例，基于MARKUS，半小时内可完成任务中3 877个地名、3 070个官名、2 872个日期的精准自动标注和统计，以及92个寺庙朝代信息的抽取。对于大多数没有编程基础的人文研究者而言，不借助工具，仅靠人力完成这些标注任务，不仅需要丰富的知识储备，也需要大量的时间，要完成Margaret Wan副教授提到的数百部中国传统小说地理空间标注更是不可能的。与UD-Kanbun、甲言等专注古籍标注的工具包相比，MARKUS分词和标注范围的灵活性较差，但具有技术门槛低的优点，且因其强大的语料库，在历史人名、地名、年代等方面标注更精准，应用领域更广泛。

5 结束语

数字人文项目通常旨在从大量（半）结构化文本或异构数据中发现知识，数据标注在人文学者发现新知识的研究过程中至关重要。本文介绍基于MARKUS的古籍标注模型，通过具体的标注实践案例，证明了MARKUS具有很好的应用性和普适性，可以弥补人文研究者的技术短板，为其提供一个相对容易起步的平台。但MARKUS也有其局限性，在文本类型方面，文档需为纯文本格式、UTF-8编码，其余格式不能保证正确标注，需要在数据预处理阶段进行格式转换。且MARKUS在标注范围上有局限性，古籍文本标注的常用任务是命名实体标注和文本相似性标注[16]，由于语料库的限制，MUARKUS更侧重于历史人名、官名、地名等命名实体的精准自动标注。关于古籍标注的研究，众多学者在努力进一步深入，不断提出新的研究思路，更多成熟的古籍数字人文分析和标注工具开始涌现，如DocuSky、LoGaRT等。如何深入挖掘工具的应用深度和广度，助力古籍人文研究，值得进一步深入探讨。

参考文献:

[1] 刘尚恒. 古籍概念浅谈[J]. 图书馆工作与研究, 1985(2): 49-50.
LIU S H. A note on concept of ancient works[J]. Library Work and Study, 1985(2): 49-50.

[2] 杨琳. 大陆古籍数字化的现状及存在的问题[C]//中国古籍数字化国际学术研讨会论文集. [出版地不详: 出版者不详], 2007: 46-58.
YANG L. The present situation and existing problems of the digitization of ancient books in Mainland China[C]// Proceedings of International Symposium on the Digitization of Chinese Ancient Books. [S.l.:s.n.], 2007: 46-58.

[3] 柯平, 宫平. 数字人文研究演化路径与热点领域分析[J]. 中国图书馆学报, 2016, 42(6): 13-30.
KE P, GONG P. The evolution path and hot topics of digital humanities research[J]. Journal of Library Science in China, 2016, 42(6): 13-30.

[4] 蔡莉, 王淑婷, 刘俊晖, 等. 数据标注研究综述[J]. 软件学报, 2020, 31(2): 302-320.
CAI L, WANG S T, LIU J H, et al. Survey of data annotation[J]. Journal of Software, 2020, 31(2): 302-320.

[5] YANG H L, JIN L W, SUN J F. Recognition of Chinese text in historical documents with page-level

annotations[C]//Proceedings of 2018 16th International Conference on Frontiers in Handwriting Recognition. Piscataway: IEEE Press, 2018: 199–204.

[6] 郑永晓, 段海蓉. 古籍数字化、数字人文与古代文学研究: 访中国社会科学院郑永晓教授[J]. 吉首大学学报(社会科学版), 2020, 41(2): 144–151.
ZHENG Y X, DUAN H R. Digitization of ancient books, digital humanities, and ancient Chinese literary research: an interview with professor zheng yongxiao from Chinese academy of social sciences[J]. Journal of Jishou University (Social Sciences), 2020, 41(2): 144–151.

[7] MORETTI F. Distant reading[M]. London: Verso Books, 2013: 211–221.

[8] 苏祺, 胡韧奋, 诸雨辰, 等. 古籍数字化关键技术评述[J]. 数字人文研究, 2021, 1(3): 83–88.
SU Q, HU R F, ZHU Y C, et al. Key technologies for digitization of ancient Chinese books[J]. Digital Humanities Research, 2021, 1(3): 83–88.

[9] 谢韬. 基于古文学的命名实体识别的研究与实现[D]. 北京: 北京邮电大学, 2018.
XIE T. Research and implementation of named entity recognition based on ancient literature[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.

[10] 马海丽, 王曦. 古籍数字化中计算机自然语言处理应用现状分析[J]. 古籍研究, 2020(2): 322–328.
MA H L, WANG X. Analysis on application of computer natural language processing in digitization of ancient books[J]. Reseaech on Chinese Ancient Book, 2020(2): 322–328.

[11] 钱智勇, 周建忠, 童国平, 等. 基于HMM的楚辞自动分词标注研究[J]. 图书情报工作, 2014, 58(4): 105–110.
QIAN Z Y, ZHOU J Z, TONG G P, et al. Research on automatic word segmentation and pos tagging for Chu ci based on HMM[J]. Library and Information Service, 2014, 58(4): 105–110.

[12] 张琪, 江川, 纪有书, 等. 面向多领域先秦典籍的分词词性一体化自动标注模型构建[J]. 数据分析与知识发现, 2021, 5(3): 2–11.
ZHANG Q, JIANG C, JI Y S, et al. Unified model for word segmentation and POS tagging of multi-domain pre-Qin literature[J]. Data Analysis and Knowledge Discovery, 2021, 5(3): 2–11.

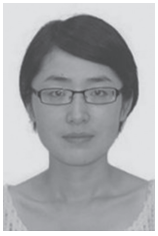
[13] YASUOKA K. Universal dependencies treebank of the Four Books in classical Chinese[C]//Proceedings of the 10th International Conference of Digital Archives and Digital Humanities. [S.l.:s.n.], 2019: 20–28.

[14] 邢付贵, 朱廷劭. 基于大规模语料库的古文词典构建及分词技术研究[J]. 中文信息学报, 2021, 35(7): 41–46.
XING F G, ZHU T S. Large-scale online corpus based classical integrated Chinese dictionary construction and word segmentation[J]. Journal of Chinese Information Processing, 2021, 35(7): 41–46.

[15] 邓三鸿, 胡昊天, 王昊, 等. 古文自动处理研究现状与新时代发展趋势展望[J]. 科技情报研究, 2021, 3(1): 1–20.
DENG S H, HU H T, WANG H, et al. Review of automatic processing of ancient Chinese character and prospects for its development trends in the new era[J]. Scientific Information Research, 2021, 3(1): 1–20.

[16] 欧阳剑, 任树怀. 数字人文研究中的古籍文本阅读可视化[J]. 图书馆杂志, 2021, 40(4): 82–89, 99.
OUYANG J, REN S H. Visualization of ancient texts reading in digital humanities research[J]. Library Journal, 2021, 40(4): 82–89, 99.

作者简介



于亚秀（1985– ），女，华东师范大学图书馆副研究馆员，主要研究方向为数字人文、知识组织与管理、智慧图书馆建设。

李欣（1961– ），女，华东师范大学数据科学与工程学院研究馆员，主要研究方向为语义网知识组织与管理、数字人文、推荐系统。

收稿日期： 2022-01-29

通信作者： 李欣，xli@dase.ecnu.edu.cn

基金项目： 中央高校基本科研业务费项目（No.2022ECNU-XWK-ZX05）

Foundation Item: Fundamental Research Funds for the Central Universities (No.2022ECNU-XWK-ZX05)