CCL23-Eval任务1总结报告：古籍命名实体识别(GuNER2023)

苏祺[1,3,4]，王莹莹[2,4]，邓泽琨[2,4]，杨浩[3,4]，王军[2,3,4]✉
[1]北京大学外国语学院　[2]北京大学信息管理系
[3]北京大学人工智能研究院　[4]北京大学数字人文研究中心
{sukia, dzk, yanghao2008, junwang}@pku.edu.cn, ying_y_wang@126.com

摘要

第22届中国计算语言学大会(CCL)提出了中文信息处理方面的10个评测任务。其中，任务1为古籍命名实体识别评测，由北京大学数字人文研究中心、北京大学人工智能研究院组织。该任务的主要目标是自动识别古籍文本中事件基本构成要素的重要实体，以提供对古汉语文本进行分析处理的基础。评测发布了覆盖多个朝代和领域的“二十四史”评测数据集，共15万余字，包含人名、书名、官职名三种实体超万数。同时设置了封闭和开放两个赛道，聚焦于不同规格的预训练模型的应用能力。共有127支队伍报名参加了该评测任务。在封闭赛道上，参赛系统在测试集上的最佳性能达到了96.15%的F1值；在开放赛道上，最佳性能达到了95.48%的F1值。

关键词： 古汉语；命名实体识别；评测；古文信息处理

Overview of CCL23-Eval Task 1: Named Entity Recognition in Ancient Chinese Books

Qi Su[1,3,4],Yingying Wang[2,4],Zekun Deng[2,4],Hao Yang[3,4],Jun Wang[2,3,4]✉
[1]School of Foreign Languages, Peking University
[2]Department of Information Mangagement, Peking University
[3]Institute for Artificial Intelligence, Peking Univeristy
[4]Research Center for Digital Humanities, Peking University
{sukia, dzk, yanghao2008, junwang}@pku.edu.cn, ying_y_wang@126.com

Abstract

The 22nd China National Conference on Computational Linguistics (CCL2023) presented 10 evaluation tasks in the field of Chinese information processing. Among them, Task 1 (GuNER2023) focused on the evaluation of Named Entity Recognition (NER) for Ancient Chinese texts, organized by the Digital Humanties Research Center and the Institute of Artificial Intelligence at Peking University. The main objective of this task was to automatically identify important entities related to the basic components of events in ancient texts, thus providing a foundation for analyzing and processing Classical Chinese texts. The evaluation released the *Twenty-four Histories* dataset, which covers multiple dynasties and domains, including three types of entities: personal names, book titles, and official positions. Two tracks, restricted and unrestricted tracks, were set up to assess the capabilities of pre-trained models with different specifications. A total of 127 teams registered for this evaluation task. In the restricted track, the best-performing system achieved an F1 score of 96.15% on the test set, while in the unrestricted track, the highest performance researched an F1 score of 95.48%.

本研究得到国家自然科学基金国际重点合作项目“中国儒家学术史知识图谱构建研究”(项目号:72010107003)的支持

Keywords: Ancient Chinese , Named Entity Recognition , Evaluation , Ancient Language Information Processing

1 引言

古籍命名实体识别（Named Entity Recognition）任务的目的是自动化抽取古籍善本中的明确实体对象，实体类型包括人名、地名、机构名以及其他可定义的实体类型，例如官职名、书名等 (苏祺et al., 2021)。古籍文献的命名实体识别是正确分析处理古汉语文本的基础步骤，也是深度挖掘和组织人文知识的重要前提，对于在数字人文环境下历史人文数据库和工具的构建具有显著的学术价值和实践意义。

近年来，学界已有多项研究关注史籍、方志、诗词、中医等类目的古籍命名实体识别，并构建了一些针对特定领域的小型标注数据集。实体标注的体系和规范也有所差异，识别范围通常由三种基本实体类别扩充至人文计算研究所需的多种特殊类别，如书名、药物名、疾病名、动植物名等 (杜悦et al., 2021; 黄水清et al., 2015; 刘江峰et al., 2022; 崔竞烽et al., 2020; 李娜, 2021; 谢靖et al., 2022; 林立涛et al., 2022)。总体而言，古籍命名实体识别任务仍然缺乏可用于模型训练以及评测的公开数据资源，阻碍了技术的进一步发展。另一方面，古汉语在不同时代和不同领域的古籍文献中具有丰富的字形变化和语境含义，以及行文结构的连续性、无句读等特点，这也增加了古籍文献命名实体识别任务的复杂和困难程度。

北京大学人工智能研究院和北京大学数字人文研究中心联合组织了本次古籍文献的命名实体识别评测，基于“二十四史”建构了覆盖多个朝代的历时、跨领域数据资源，以完善古籍命名实体识别数据的扩充和任务的建立。与以往的古籍命名实体识别数据集和评测任务相比，本次评测具有以下特色：

首先，针对不同朝代和领域的古籍文献所反映出的语言和实体特征差异，本次评测选择了历史典籍“二十四史”来建立实体标注体系和数据集，以期提升古籍命名实体识别模型在不同领域的适用性。“二十四史”是中国古代各朝撰写的二十四部正史的总称，均以纪传体编撰。它上起传说中的黄帝时期，下至明朝崇祯十七年，涵盖了中国古代政治、经济、军事、思想、文化、天文、地理等各方面的内容，是各个历史时期社会各领域的缩影和记录。

其次，本次评测数据集的实体知识体系涵盖了人名、书名和官职名三种类型。在历史典籍中，与事件相关的人物、地点等实体是最为重要和易于获取的知识，同时，官职身份亦是体现事件中人物关系的重要信息，需要准确地识别和挖掘。

最后，本次评测设置了封闭和开放两个赛道，旨在比较、探索和挖掘不同规模的预训练语言模型在古籍命名实体识别任务中的应用能力。封闭赛道要求参赛队伍禁止使用大模型，而开放赛道要求必须使用大语言模型。

本文主要包含如下内容：第2节主要介绍了古籍命名实体识别的相关工作，包括数据集、实体标注和模型算法等。第3节详细介绍了本次评测的具体设置，例如数据集、评价指标、赛道要求等。第4节概述了本次评测的参赛情况。第5节展示参赛队伍所使用的方法，并进行了总结分析。最后，第6节对本次评测进行了总结。

2 相关工作

2.1 古籍实体标注数据集

实体标注是对数字化古籍文本进行概念与知识的抽取、挖掘的重要支撑，但其人工标注成本显著高于现代汉语。一方面是因为古籍文本的电子化数据资源相对较少，另一方面则是古籍实体标注对标注人员的知识背景有较高的要求，需要具备一定的古汉语专业知识。而手工标注的操作效率低下，使得标注成本不断攀升。早期的古籍数据集主要关注于史籍文本中的人名和地名等基本实体类型，例如朱晓 (2012)标注了编年体《明史本纪》中的人名，皇甫晶和王凌云 (2013)标注了西晋陈寿所著《三国志·蜀书》十五卷中的人名，黄水清等 (2015)标注了《春秋左氏传》中的地名等等。随后的研究逐渐将实体标注范围扩充至其他多种可定义的实体类型。例如，李娜 (2021)标注了《方志物产》山西卷中物产信息的别名、人名、地名、引用名、用途名等，谢靖等 (2022)利用词典资源完成了《黄帝内经》中医学概念实体的标注，林立涛等 (2022)等对25部先秦典籍语料库中的动物实体进行了标注，崔竞烽等 (2020)通过网络、论文

和书籍进行菊花古典诗词数据的采集，并对其中的时间、地点、季节、花名、花色、人物和节日等7类命名实体进行了标注。此外还有许多古籍实体标注数据集的构建研究，不再一一赘述。

综合而言，现有的古籍实体标注数据集往往聚焦于特定类目和领域的文本，不同数据集之间的语言和实体特征存在明显差异，标注的体系和方法也各有不同，因此不能统一适用于模型训练。为此，本次评测选用历史典籍“二十四史”建构了覆盖多个朝代的历时跨领域数据资源，旨在扩充古籍命名实体识别数据集，并提升识别技术的领域适应性。

2.2 基于预训练语言模型的古籍命名实体识别方法

命名实体识别模型的编码层对输入进行抽象语义表示，解码层则用于预测实体的边界和类型。2018年10月谷歌AI团队发布新的语言表征模型——BERT（Bidirectional Encoder Representation from Transformers）(Devlin et al., 2019)，刷新11项自然语言处理任务记录。其后预训练模型作为编码层并结合下游任务微调逐渐成为主流的文本挖掘方法。崔竞烽等 (2020)在古典诗词的“花”类实体抽取中引入了预训练模型，并证明了BERT (Devlin et al., 2019)在诗词实体抽取任务上具有一定的优势。

中文BERT是基于中文维基百科训练的包含简体和繁体中文的预训练模型，普适性虽强，但在面对特定领域文本的自然语言处理任务时，其功能的发挥容易受限。而古代汉语与现代汉语在语法、语义、语用方面存在较大差异。古籍命名实体识别数据集具有领域特定的特殊性，因此领域化的深度预训练语言模型成为提高古籍文本实体识别效果的关键技术。GuwenBERT[2]模型是基于殆知阁古文文献语料进行训练的，包含15,694本古文书籍，总字符数达到1.7亿。该模型对所有繁体字进行了简体转换处理，并结合现代汉语RoBERTa (Liu et al., 2019)权重和大量古文语料，将现代汉语的部分语言特征迁移到古代汉语中，在2020年“古联杯”命名实体识别评测比赛中获得了二等奖。胡韧奋等 (2021)针对古汉语句子长度较短且多数不含断句和标点信息的特点，将古文段落作为输入单位，将自动断句作为下游任务，同样基于殆知阁古文文献语料训练得到一个古汉语深层语言模型，在古汉语自动断句任务上实现了高精度，并在“古联杯”命名实体识别评测比赛中获得了一等奖。王东波等 (2022)提供的SikuBERT和SikuRoBERTa则是基于《四库全书》繁体语料在BERT (Devlin et al., 2019)和RoBERTa (Liu et al., 2019)上进行继续训练的预训练模型，其设计面向《左传》语料的命名实体识别等任务，验证了SikuBERT等预训练模型在古文词法、句法和语境学习以及泛化能力方面具有较强的能力。命名实体识别模型常用的解码层方法包括条件随机场（Conditional Random Field, CRF）、指针网络（Pointer Network）、循环神经网络（Recurring Network）等。

谢志强等 (2022)针对古汉语中的嵌套命名实体识别问题，使用全局指针网络(Global Pointer Network)作为解码器，并结合了RoBERTa-classical-chinese、SikuRoBERTa、SikuBERT、RoBERTa-wwmext、BERT-wwm-ext和GuwenBERT六个预训练语言模型，在基于《史记》标注的人名、地名、官职、书名和时间这五类实体的数据集上进行了实验。实验结果表明，RoBERTa-classical-chinese和SikuRoBERTa结合全局指针网络在古汉语嵌套命名实体识别任务上能够获得良好的性能。陈雪松等 (2023)指出了一种古汉语实名实体识别方法，称为SikuBERT-BiLSTM-MHA-CRF。他们利用SikuBERT预训练模型 (王东波et al., 2022)，结合双向LSTM（Bidirectional Long Short- Term Memory）网络和多头注意力机制，实现了性能的提升。严承希等 (2023)针对古籍命名实体识别任务中的少样本问题，利用深度主动学习算法实现了高性能的预测，并且减少了迭代次数，从而有效降低了人工成本。

本次评测基于预训练语言模型的规格限制，设置了封闭和开放两个赛道，以期展示基于预训练语言模型的古籍命名实体识别方法在评测数据集上的性能。

3 评测设置

3.1 评测数据集和评价标准

本次评测提供官方评测数据集“古籍命名实体识别2023”(GuNER2023)，由北京大学数字人文研究中心组织标注，语料来源是网络上公开的部分中国古代正史纪传文本。数据包括供参赛队伍进行模型训练与调优的训练集，以及评测参赛队伍模型性能的封闭测试数据集。同时，各参赛队伍可以自行使用其他公开的人工标注数据集和伪数据集。训练集以“二十四史”为基础语料，包含13 部书中的22 卷语料，随机截断为长度约100 字的片段，标注了人名（PER）、书名

[2]GuwenBERT https://github.com/ethan-yt/guwenbert

（BOOK）、官职名（OFI）三种实体，总计15.4 万字（计标点）。数据集标注过程如下：首先，至少两名普通标注者独立对相同的文本进行标注。如果存在标注结果的不一致，那么专业标注者将进行第二轮标注检查。对古籍中不同类型命名实体的标注规范将另外撰文详述。

评测数据集格式为文本文件，参赛队伍可根据模型需要进行转化处理。其中训练集数据样例如下所示，每行为二十四史原文中的一个段落，段中每一个实体以“ {} ”标识，“ | ”后为实体类别。测试集数据集包含原文内容，参赛队伍需要提交在测试集文本上的实体识别结果文件，格式与训练集一致。训练集数据共2347段、15万余字，三种实体的数量共10246个。测试集数据共224段、约1.5万字。

{元|PER}兄{希元|PER}，{高宗|PER}洛州{司法|OFI}，{章太子|PER}召令{洗|OFI}{言|PER}等注解{范|PER}{後|BOOK}，行於代。先{元|PER}卒。

{友|PER}幼亦明敏，通{|BOOK}、{小|BOOK}，音律。{存|PER}已死，{太祖|PER}以{友|PER}{元指使|OFI}，表{右威武|OFI}。

本次评测的测试数据集采用封闭方式给出，即仅给定原古文文本，需要参赛队伍训练模型对文本中的命名实体进行自动识别和标注，并将结果文件打包上传至在线评测平台，获取评测指标得分。本次评测使用准确率（Precision）、召回率（Recall）和F1值作为评价指标。

3.2 赛道设置

为比较、探索和挖掘不同规模的预训练语言模型在古籍命名实体识别任务中的应用能力，GuNER2023设置了开放和封闭两个赛道：开放赛道要求参赛队伍必须使用ChatGPT、文心一言、ChatGLM 等大模型；封闭赛道的参赛队伍禁止使用大模型，仅允许使用拥有开源License（如GPL、BSD、MIT、Apache等）且参数量小于10B 的预训练语言模型。两个赛道使用不同的评测提交入口，参赛队伍可以同时参加两个赛道的评测提交，也可以选择只参加其中一个赛道。

4 报名情况与评测结果

4.1 评测情况

本次评测于2023年4月10日开启报名，共吸引了127支队伍报名参与，体现了行业对古文自然语言处理技术的关注。其中，92支队伍来自国内外多所科研院校和机构，包括北京大学、中国社会科学院、北京信息科技大学、南京航空航天大学、中国科学院信息工程研究所、成都信息工程大学、澳门大学、香港中文大学、美国雪城大学等。这些院校的参赛队伍涵盖了不同的专业、学院和实验室，既包括计算机及自然语言处理等工科背景团队，也有信息管理、信息传播、语言研究、民族学与人类学研究等人文社科研究团队。另外，还有19支队伍来自字节跳动、数据方舟、杭州十域科技、中国电信、金融壹账通、元知科技、联想诺谛、水滴科技等企业，以及2支队伍是苏州大学和阿里巴巴公司的校企合作参赛，此外还有1支队伍来自中国民族图书馆。

封闭和开放赛道的评测提交入口于2023年4月28日至6月1日开放。6月5日至9日，评测榜单排名较高的参赛队伍提交了实验数据、代码等信息，供评测组织方进行复现审核。根据两个赛道的榜单排名以及复现审核结果，于6月15日公布了封闭赛道的最终排名和评测得分，详见表1。开放赛道的两支参评队伍均不符合大模型使用规则，因此奖项置空。开放赛道榜单中前两名的得分如表2所示，其中的模型信息为参赛队伍提交评测时所填入，但并未提交实验代码和技术报告，大模型使用的方法和指令无法得知。参赛队伍的单位和成员信息亦无法得知。

4.2 方法分析

本次评测共接收到封闭赛道的6份技术报告，其中有5份来自排名前5名参赛队伍。本节内容对参赛队伍在封闭赛道中所使用的基于预训练模型的实体识别方法进行分析。

Table 1: 封闭赛道排名与榜单成绩

排名	队伍	单位	榜单排名	榜单成绩	复现成绩
1	KDSec_IIE	中国科学院信息工程研究所	1	96.15	96.15
2	翼智团_TeleAI	中国电信股份有限公司数字智能科技分公司	2	95.87	95.82
3	BISTU_IIIP	北京信息科技大学	4	95.34	95.34
4	CUIT_IDSE	成都信息工程大学	5	95.08	95.08
5	JZW	个人	3	95.68	94.34

Table 2: 开放赛道榜单部分结果

榜单排名	队伍	模型	榜单成绩
1	wzjj98	ChatGPT	95.48
2	东财	ChatGLM	95.43

第一名的参赛队伍KDSec_IIE所使用的预训练模型是RoBERTa (Liu et al., 2019)，预训练参数来自于在古文数据上训练的Roberta-classical-chinese-large-char[3]，是GuwenBERT的改进版本。此外，该队伍设计了Token-wise感知的序列标注和Span-level感知的实体识别两种框架，融合两种框架的实体预测结果，集成多个结果以提升识别性能。其中，Span-level感知的框架是穷举输入句子中所有满足最大实体长度限制的实体Span，进而计算每个Span在每个实体类型标签下的概率分布。同时，从信息论的视角显式地约束实体特征的表达，即最大化实体上下文特征与标签之间的互信息，以及最小化冗余信息。

第二名的参赛队伍翼智团_TeleAI基于BERT (Devlin et al., 2019)、ERNIE (Zhang et al., 2019)、GuwenBERT和MengziBERT (Zhang et al., 2021)等预训练模型，使用未标注的“二十四史”文本进行领域持续训练，然后使用GuNER2023训练集进行任务持续预训练，再使用W2NER (Li et al., 2022)、BERT-CRF和BERT-Span (Zhao et al., 2019)进行微调，实验结果表明基于字级别特征的W2NER (Li et al., 2022)可以更好地捕获词语之间的联系，实体识别性能最好。最后基于上下文信息融合多个模型的实体识别结果，进一步提升了模型性能。

第三名的参赛队伍BISTU_III所使用的预训练模型是NEZHA-Chinese-Base模型 (Wei et al., 2019)，相较于BERT (Devlin et al., 2019)采用相对位置编码词向量，可以更好地挖掘文本中的字符关系。在其后接入两层的时序卷积神经网络用于挖掘局部时序关联语义信息，并基于未标注的“二十四史”文本进行持续预训练。解码层使用全局指针网络以更准确地识别实体边界，得到实体预测结果。同时为了增强模型的泛化能力，使用对抗学习方法中的快速梯度法（Fast Gradient Method,FGM） (Miyato et al., 2016)在模型训练过程中添加干扰信息，以提升模型性能。该队伍没有融合多个模型的识别结果，但在后处理阶段结合规则改善了漏标、错标的常见错误，矫正模型输出，提升评测结果。

第四名的参赛队伍CUIT_IDSE所采用的的方法也是基于预训练模型BERT在未标注的“二十四史”文本上进行领域持续训练和任务持续训练。同时在模型训练中也使用对抗学习方法添加干扰信息，提升模型的泛化能力。解码层使用全局指针网络，同时融合多个模型的识别结果，也使用了基于规则的后处理方式，矫正模型输出以提升性能。

第五名的参赛队伍JZW使用了预训练模型BERT (Devlin et al., 2019)获取输入文本的表征，同时提出基于提示学习思想的PromptNER模型，将与实体类别有关的提示词（人、书、职）进行串联和联合编码，增强实体与类别的语义交互。解码层采用全局指针网络，基于Span预测在每个提示词上的概率分布，即可得Span对应的实体类别。该队伍同时也使用了对抗学习方法增加干扰信息，提升模型的泛化能力和性能。

参赛队伍BIT使用了预训练模型SikuRoBERTa、SikuBERT (王东波et al., 2022)、RoBERTa-classical-chinese-base-char、bert-ancient-chinese、GuwenBERT，后接Bi-

[3]https://huggingface.co/KoichiYasuoka/roberta-classical-chinese-large-char

LSTM或LSTM，解码层使用CRF预测实体信息。实验结果表明性能最好的模型是bert-ancient-chinese，相较于SikuRoBERTa的结果有所提升。这说明预训练模型的词表大小对于古籍命名实体识别任务十分重要。

综合而言，这6支参赛队伍都采用了基于预训练模型的实体识别方法，并且更倾向于使用面向古籍文本开发的领域化预训练模型。局限于标注数据的匮乏，使用词表较大的预训练模型可以获取更优的识别性能。随后使用未标注的"二十四史"文本进行领域持续预训练，并使用评测训练集进行任务训练。解码层使用全局指针网络以获取更为准确的实体边界预测结果。在模型训练过程中采用对抗学习方法增加干扰，能够提升模型的泛化能力。此外，融合多个模型的实体预测信息以及基于规则的后处理实体矫正方式也是提升模型性能的有效策略。

此外，针对古籍命名实体识别任务的少样本学习问题，参赛队伍采用了主动学习和数据增强策略。例如，参赛队伍KDSec_IIE设计了两种利用篇章信息的数据增强策略：一种是将句子所在的章节信息拼接在句子后面，引入篇章先验信息；另一种是将来自于同一来源的句子拼接合并，然后通过滑动窗口进行采样，以获取更多数据。参赛队伍JZW通过主动学习策略来筛选特殊样本，并在数据层面进行数据增强，以提升模型性能。

5 总结

本次古籍命名实体识别评测任务（GuNER2023）由北京大学人工智能研究院和北京大学数字人文研究中心联合组织，并作为第22届中国计算语言学大会（CCL2023）的10项评测任务之一。评测发布了基于"二十四史"的历时、跨领域实体标注数据集，并设置了开放和封闭两个赛道，提供了统一的评测基准和提交入口。在评测赛事阶段，共有127支队伍报名并提交参赛系统，经过模型复现和审核后，本文展示了前5名队伍的排名与成绩。

基于6支队伍所提交的技术报告，本文总结分析了参赛队伍在封闭赛道上采用的主流方法，包括预训练语言模型、领域持续训练、任务持续训练方法、对抗学习方法、全局指针网络解码层，以及模型融合、后处理和数据增强等提升模型性能的策略。同时，针对古籍命名实体识别任务的少样本学习问题，部分参赛队伍采用了深度主动学习和数据增强的方法。封闭赛道的参赛系统在测试集上获得的最好性能为F1值96.15%，展现了当前基于预训练模型的古籍命名实体识别技术的水平。

而在开放赛道上，榜单最好性能为F1值95.48%。由于没有参赛队伍提供代码和技术报告，所以我们对大模型使用的具体技术和指令无法进行分析。从得分仍可以看出大模型虽与封闭赛道的专有小模型存在差距，但也已经展现出较好的性能。然而，仍存在结果的不确定性以及边界识别不够精准等问题。因此，如何建构更适用于古籍领域的指令是大模型研究范式下古籍命名实体识别任务的重要研究方向。

参考文献

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. 2022. Unified named entity recognition as word-word relation classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10965–10973.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Takeru Miyato, Andrew M Dai, and Ian Goodfellow. 2016. Adversarial training methods for semi-supervised text classification. *arXiv preprint arXiv:1605.07725*.

Junqiu Wei, Xiaozhe Ren, Xiaoguang Li, Wenyong Huang, Yi Liao, Yasheng Wang, Jiashu Lin, Xin Jiang, Xiao Chen, and Qun Liu. 2019. Nezha: Neural contextualized representation for chinese language understanding. *arXiv preprint arXiv:1909.00204*.

Zhiqiang Xie, Jinzhu Liu, and Genhui Liu. 2022. 古汉语嵌套命名实体识别数据集的构建和应用研究(construction and application of classical Chinese nested named entity recognition data set). In *Proceedings of the 21st Chinese National Conference on Computational Linguistics*, pages 406–416, Nanchang, China, October. Chinese Information Processing Society of China.

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.

Zhuosheng Zhang, Hanqing Zhang, Keming Chen, Yuhang Guo, Jingyun Hua, Yulong Wang, and Ming Zhou. 2021. Mengzi: Towards lightweight yet ingenious pre-trained models for chinese. *arXiv preprint arXiv:2110.06696*.

Zhe Zhao, Hui Chen, Jinbin Zhang, Xin Zhao, Tao Liu, Wei Lu, Xi Chen, Haotang Deng, Qi Ju, and Xiaoyong Du. 2019. Uer: An open-source toolkit for pre-training models. *arXiv preprint arXiv:1909.05658*.

严承希, 唐雪梅, 杨浩, 苏祺, and 王军. 2023. Hanner: 一个面向汉语古籍语料命名实体自动抽取的通用框架. 情报学报, 42(2):203–216.

刘江峰, 冯钰童, 王东波, 胡昊天, and 张逸勤. 2022. 数字人文视域下SikuBERT增强的史籍实体识别研究. 图书馆论坛, 42(10):61–72.

崔竞烽, 郑德俊, 王东波, and 李婷婷. 2020. 基于深度学习模型的菊花古典诗词命名实体识别. 情报理论与实践, 43(11):150–155.

朱晓. 2012. 古汉语编年体的人名实体识别与词性标注. 复旦大学硕士学位论文.

李娜. 2021. 面向方志类古籍的多类型命名实体联合自动识别模型构建. 图书馆论坛, 41(12):113–123.

杜悦, 王东波, 江川, 徐润华, 李斌, 许超, and 徐晨飞. 2021. 数字人文下的典籍深度学习实体自动识别模型构建及应用研究. 图书情报工作, 65(3):100–108.

林立涛, 王东波, 刘江峰, 李斌, and 冯敏萱. 2022. 数字人文视域下典籍动物命名实体识别研究——以SikuBERT预训练模型为例. 图书馆论坛, 42(10):42–50.

王东波, 刘畅, 朱子赫, 刘江峰, 胡昊天, 沈思, and 李斌. 2022. Sikubert 与sikuroberta: 面向数字人文的《四库全书》预训练模型构建及应用研究. 图书馆论坛, 42(6):31–43.

皇甫晶and 王凌云. 2013. 基于规则的纪传体古代汉语文献姓名识别. 图书情报工作, 57(03):120–124.

胡韧奋, 李绅, and 诸雨辰. 2021. 基于深层语言模型的古汉语知识表示及自动断句研究. 中文信息学报, 35(4):8–15.

苏祺, 胡韧奋, 诸雨辰, 严承希, and 王军. 2021. 古籍数字化关键技术评述. 数字人文研究, 1(03):83–88.

谢靖, 刘江峰, and 王东波. 2022. 古代中国医学文献的命名实体识别研究——以flat-lattice 增强的sikubert 预训练模型为例. 图书馆论坛, 42(10):51–60.

陈雪松, 詹子依, and 王浩畅. 2023. 融合sikubert模型与mha的古汉语命名实体识别. 吉林大学学报*(*信息科学版*,* 网络首发：*2023-05-15*.

黄水清, 王东波, and 何琳. 2015. 基于先秦语料库的古汉语地名自动识别模型构建研究. 图书情报工作, 59(12):135–140.