**技术参数及要求**

|  |
| --- |
| **（一）项目概述** |
| **项目建设任务、用途和应用场景**项目建设目标：通过购置教学科研实验仪器设备，深化教学改革，创新人才培养模式，切实改善教学环境，提升基本办学条件，促进学校高质量人才培养工作；此外，通过信息化及数字资源等公共服务体系建设，提高公共服务体系服务能力。项目具体建设任务、用途和应用场景如下：

|  |  |  |
| --- | --- | --- |
| **任务序号** | **主要建设任务** | **用途和应用场景** |
| 1 | 多功能语料库检索平台建设 | 科研 |
| 2 | 舆情大数据分析模块二期功能建设 | 科研 |
| 3 | 语料对齐模块二期功能建设 | 科研 |
| 4 | 智能翻译模块二期功能建设 | 科研 |

 |
| **（二）项目建设方案****1.项目建设内容及技术参数**

|  |  |
| --- | --- |
| **项目建设内容序号** | **项目建设内容** |
| 1 | 多功能语料库检索平台建设 |
| 2 | 舆情大数据分析模块二期功能建设 |
| 3 | 语料对齐模块二期功能建设 |
| 4 | 智能翻译模块二期功能建设 |

 |
|

|  |  |  |
| --- | --- | --- |
| **序号** | **建设内容** | **功能需求** |
| **1** | **多功能语料库检索平台** | 一、单语语料库检索与管理模块1、支持英语、日语、俄语、哈萨克语、乌克兰语、白俄罗斯语、捷克语、德语、法语、西班牙语、意大利语、葡萄牙语、波兰语、罗马尼亚语、希腊语、匈牙利语、阿拉伯语、波斯语、土耳其语、斯瓦希里语、印地语、乌尔都语、朝鲜语、泰语、印度尼西亚语、马来语、菲律宾语等。2、支持导入50M以内的txt、doc、docx、xls、xlsx、可编辑pdf等格式的语料文件，txt类型的文件需支持UTF-8、ANSI、GB2312等编码格式。★3、支持语料库共建、共享，满足师生教学科研需要。4、管理员可以管理所有语料库，普通用户可自建语料库上传、共享、检索、分析。5、支持导入已进行分词或者已标注词性的常用语料库，对未进行分词或者词性标注的数据，系统支持中文、英语、日语、法语、德语、俄语、阿拉伯语、韩语、意大利语、西班牙语、葡萄牙语、泰语、越南语、缅甸语等语言采用自动分词和自动词性标注方法进行分词和词性标注。6、支持按语料库维度和文件维度统计和展示形符数量、类符数量、库容、类符形符比、标准化类符形符比、平均句长、平均词长、元信息等信息。7、支持索引功能，索引支持精确检索、通配符、正则表达式等检索模式，可区分大小写和添加忽略词。8、索引检索结果支持上下文关键词，展现方式和全句展现方式，在结果列表中可查看序号、文本内容、文件名称、语料库名称和更新时间。支持设置是否显示或仅显示检索结果的词性标注信息。9、索引检索结果支持按照邻接词的位置进行排序，排序范围可支持L1到L10、R1到R10，支持按照词频大小和单词字母顺序对检索结果进行排序。10、支持对索引检索结果中出现错误的句子进行编辑并填写修改理由，公有库编辑后的内容需通过管理员审核通过后更新至语料库中。11、支持搭配分析，支持精确检索、通配符、正则表达式等检索模式，可区分大小写和添加忽略词，支持设置搭配邻接词范围从L10~R10。★12、支持对检索关键词和邻接词进行Log-likelihood、MI、T-SCORE、Z-SCORE、DICE等多种搭配的统计分析，检索结果可按照相关统计值进行排序展示。13、支持词频统计分析，系统可提供默认停用词，用户可以在默认停用词的基础上进行停用词修改。14、词频统计时，系统需提供默认词元表，用户可以选择启用词元表进行词频统计；支持在词频统计结果中按照单词长度、频次和关键词进行二次检索；支持对词频统计分析结果生成词云图并下载。15、支持词分布统计分析，支持精确检索、通配符检索、正则表达式检索等检索模式，可以按命中词所在的文件进行词分布统计，统计结果包含文本词数、关键词命中频次等统计结果，可按照相关统计指标进行排序展示。16、支持通过分布图展示命中词在文件中的位置信息，支持保留历史检索结果，用以查看不同词语在同一个文件中分布图的堆叠效果。17、支持词簇功能，支持生成含特定检索词的多词序列，并可根据检索词在序列中的位置、词数、词频等条件过滤结果。18、支持生成指定元数的N元并统计分析，支持设置N元的模糊匹配词数，并统计N元的频次、出现文件范围数量等信息。19、支持设置语料库的元信息字段的内容和格式，方便用户在上传文件到该语料库时填写相关元信息字段内容。20、索引、搭配、词频统计、关键词分析、词分布、词簇、N元统计在检索分析时，可通过元信息内容筛选过滤语料库文件，以便快速找到需要检索的目标语料库或文件。21、支持导出索引、搭配、词频统计、关键词分析、词分布、词簇及N元统计结果，导出格式支持txt和xlsx。二、多语平行语料库检索与管理模块1、支持汉语、英语、法语、意大利语、德语、日语、韩语、葡萄牙语、俄语、西班牙语、泰语、藏语、缅甸语、越南语、阿拉伯语、荷兰语、南非语、阿尔巴尼亚语、阿拉贡语、亚美尼亚语、孟加拉语、白俄罗斯语、布林语、波斯尼亚语、保加利亚语、科西嘉语、克罗地亚语、捷克语、丹麦语、不丹语、世界语、爱沙尼亚语、斐济语、菲律宾语、芬兰语、甘达语、格鲁吉亚语、希腊语、格陵兰语、瓜拉尼语、夏威夷语、希伯来语、印地语、苗语、匈牙利语、冰岛语、印尼语、拉丁国际语、爱尔兰语、爪哇语、克什米尔语、哈萨克语、高棉语、库尔德语、吉尔吉斯语、老挝语、拉丁语、拉脱维亚语、立陶宛语、卢森堡语、马其顿语、马拉加西、马来语、马耳他语、毛利语、蒙古语、尼泊尔语、北库尔德语、挪威语、普什图语、波斯语、波兰语、罗马尼亚语、罗曼什语、萨哈语、萨摩亚语、桑戈语、梵语、苏格兰盖尔语、塞尔维亚语、斯洛伐克语、斯洛文尼亚语、索加语、索马里语、瑞典语、叙利亚语、塔吉克语、泰米尔语、鞑靼语、汤加语、茨瓦纳语、土耳其语、土库曼语、乌克兰语、乌尔都语、维吾尔语、乌兹别克语、威尔士、西马里语、科萨语、彝语、壮语等语种。2、支持导入100M以内的tmx、xls、xlsx或csv等格式的语料库文件。3、支持语料库共建、共享，满足师生教学科研需要。4、管理员可以管理所有语料库，普通用户可自建语料库上传、共享、检索、分析。5、支持选择精确匹配、模糊匹配查询方式，可区分大小写进行检索。6、支持通过语种过滤查询语料库，可在检索页查看语料库条目数和简介等基本信息。7、支持检索结果展示检索结果数量、多语译文、更新时间、所在语料库等信息。8、支持对检索结果中出现错误的词条进行编辑，编辑后的内容需通过管理员审核通过后更新至语料库中。9、支持导出语料检索结果，导出格式支持xlsx、tmx和txt。三、术语检索与管理模块1、支持汉语、英语、法语、意大利语、德语、日语、韩语、葡萄牙语、俄语、西班牙语、泰语、藏语、缅甸语、越南语、阿拉伯语、荷兰语、南非语、阿尔巴尼亚语、阿拉贡语、亚美尼亚语、孟加拉语、白俄罗斯语、布林语、波斯尼亚语、保加利亚语、科西嘉语、克罗地亚语、捷克语、丹麦语、不丹语、世界语、爱沙尼亚语、斐济语、菲律宾语、芬兰语、甘达语、格鲁吉亚语、希腊语、格陵兰语、瓜拉尼语、夏威夷语、希伯来语、印地语、苗语、匈牙利语、冰岛语、印尼语、拉丁国际语、爱尔兰语、爪哇语、克什米尔语、哈萨克语、高棉语、库尔德语、吉尔吉斯语、老挝语、拉丁语、拉脱维亚语、立陶宛语、卢森堡语、马其顿语、马拉加西、马来语、马耳他语、毛利语、蒙古语、尼泊尔语、北库尔德语、挪威语、普什图语、波斯语、波兰语、罗马尼亚语、罗曼什语、萨哈语、萨摩亚语、桑戈语、梵语、苏格兰盖尔语、塞尔维亚语、斯洛伐克语、斯洛文尼亚语、索加语、索马里语、瑞典语、叙利亚语、塔吉克语、泰米尔语、鞑靼语、汤加语、茨瓦纳语、土耳其语、土库曼语、乌克兰语、乌尔都语、维吾尔语、乌兹别克语、威尔士、西马里语、科萨语、彝语、壮语等语种。2、支持导入10M以内的tbx、xls、xlsx或csv等格式的术语文件。3、支持语料库共建、共享，满足师生教学科研需要。4、管理员可以管理所有语料库，普通用户可自建语料库上传、共享、检索、分析。5、术语库导入时支持自动术语去重和覆盖全部已有术语两种文件导入模式。6、支持选择精确匹配、模糊匹配查询方式。7、支持通过语种过滤查询术语库。8、支持检索结果展示检索结果数量、多语术语、各语言对应词性、术语类型、是否大小写敏感、描述、来源、备注、所在术语库等信息。9、支持对检索结果中出现错误的词条进行编辑，编辑后的内容需通过管理员审核通过后更新至术语库中。10、支持导出术语检索结果，导出格式支持xlsx。四、基础模块1、支持防爬功能，通过统计账号请求次数、IP地址的请求数量以及设定指定时间的访问次数，观测账号是否异常，从而防止网站被批量访问和外部爬虫定期批量爬取。2、支持管理员对账号进行新增、删除、更改、查询等操作。3、支持查看用户登录及检索次数趋势、检索词排行等数据概况。4、提供平台详细的使用说明，包括功能介绍及相关术语概念的解释说明。5、平台基于B/S 架构（无需安装客户端），用户打开浏览器即可访问，支持 Chrome，Firefox，Safari，搜狗等主流浏览器。 |
| **2** | **舆情大数据分析模块二期功能** | 1、根据各语种老师实际需求新增数据来源，新增不少于100个网站，以英文网站为主。网站内容需涵盖政治、经济、金融等有重要研究价值的领域。 2、检查已有网站爬取情况，优化爬取逻辑，提高爬虫爬取成功率，确保目标网站的数据能全部爬取成功。3、优化爬取网页内容的自动分类方案，支持对数据信息进行更详细的分类，并提升分类的准确率，在搜索界面提供分类过滤选项。4、支持用中文关键词搜索各语种相关新闻的功能，可以将中文关键词调用多语言机器翻译引擎后实现该需求，支持≥1000万字符的机器翻译调用额度。 5、在搜索结果展示后，可以提供对标题的一键翻译功能。 |
| **3** | **语料对齐模块二期功能** | 1、基于跨语言的神经网络术语识别模型，融合句子的词性和语法信息，进行术语提取。2、使用神经网络文本匹配模型，根据上下文信息，将术语原文和术语译文相匹配。3、涵盖汉语、英语、日语、韩语、俄语、德语、西班牙语、法语、葡萄牙语、阿拉伯语、缅甸语、印尼语、柬埔寨语、菲律宾语、泰语、越南语、老挝语、马来语、希腊语、意大利语、土耳其语、乌克兰语、瑞典语、希伯来语、斯洛伐克语、拉丁语、捷克语、孟加拉语、亚美尼亚语、波兰语、丹麦语、荷兰语、芬兰语、匈牙利语、罗马尼亚语、印地语、藏语、维语、彝语、保加利亚语、斯洛文尼亚语、克罗地亚语、爱沙尼亚语、立陶宛语等。★4、单文档对齐下，当中日双语在一个文档中时，支持选择源语言为奇数段或偶数段，以保证语料对齐工具准确拆分中日双语的文本内容。★5、对齐模块支持将对齐结果同步到智能翻译管理模块的记忆库管理中。对齐模块支持将对齐结果同步到多功能语料库检索模块中。 |
| **4** | **智能翻译模块二期功能** | 1、支持管理员开启用户机器翻译字符调用量限制，并为各用户设置使用额度；管理员可在设置中开启/关闭用户机器翻译字符用量限制功能；功能开启后，管理员可在用户管理界面查看用户已使用机器翻译字符量，并调整机翻字符余量；管理员亦可在用户管理界面批量进行字符余量的设置；当机器翻译调用量达到所设余量上限时，用户可在编辑器中收到提醒并联系管理员修改。2、若用户调用机器翻译字符用量超出管理员的设置额度后，用户可以自行设置API Key继续调用机器翻译服务。★3、支持自定义词典功能：项目管理者可创建自定义词典，并在项目中可以设置是否启用对应语种的词典、是否允许译员在编辑器中对词条进行增加、修改、删除操作；添加到自定义词典的单词，在用户运行QA拼写检查时将会被忽略。4、支持设置权限功能，项目管理员可设置允许审校和译员导出译文和允许用户在编辑器中检索用户所有术语库和记忆库的功能。5、支持调用多类机器翻译引擎。6、支持单语术语提取功能，支持通过词频和术语中原文译文的单词数量对术语提取的结果进行过滤筛选，并支持对筛选的结果以excel文件格式下载。7、支持术语干预功能，新建项目时启用术语干预功能按钮，系统即可在调用机器翻译结果时，根据已启用的术语库，将术语翻译自动应用到机器翻译译文中。8、编辑器中支持对译文区的内容执行句首字母大写、小型大写、全部大写、全部小写、切换大小写等操作。★9、支持自动填充标记功能。10、支持自定义项目流程功能，用户根据翻译要求自定义项目翻译流程，设置5个以上的流程环节。11、支持原文更新功能。12、支持编辑器样式可视化功能。 |

**说明：**（1）上述软件系统部署在学校的虚拟化平台上，采用H5开发技术。（2）系统连接拓扑如下： |
| **2.系统集成要求**实现与国际舆情与国际传播研究院已部署的产品数据对接和账号和应用集成。具体要求如下：(1) ★中标方须按照学校要求提供所需接口并永久免费开放。(2) ★ a)要求所有系统实现与学校统一身份认证对接，实现单点登录。b)要求按照学校的统一要求和标准规范免费完成与我校各公共平台的深度对接集成和适配工作，包括但不限于于我校移动校园、掌上西外、网上办事大厅、统一消息中心、统一短信网关的数据集成、消息集成、服务集成、应用集成，实现统一信息管理。c)本项目内所有移动端应用均要求采用H5开发，并集成到学校移动校园微信企业号、掌上西外app和小程序，H5页面需遵循学校移动应用视觉设计规范和交互设计规范。(3) ★要求完成与学校公共数据中心的数据采集集成，包括系统的所有结果数据，提供系统数据库的只读账号、数据字典、代码表、API接口、集成文档等相关文档。(4) ★本项目所涉及对接费用均包含在此次采购预算中。(5)支持**（实现）**与国际舆情与国际传播研究院已部署的一期平台数据对接和账号**和应用**集成,支持导入历史数据库。(6)系统中的多功能语料库检索平台、翻译管理平台、舆情分析平台之间可实现数据无障碍导入或者对接。(7)中标方须提供本项目正常运转所必须的全部系统软件、中间件、插件等基础环境，确保系统正常稳定运行。产品中采用的第三方软件由中标方自行购买安装，如出现软件侵权行为，由中标方负责。 |
| **3.安全性要求：**（1）认证授权：保证用户的合法性和用户使用应用信息资源的权力，避免内部敏感信息泄漏和服务所提供的信息资源被非法访问，造成严重的安全事件。（2）信息安全性：确保应用系统源代码安全，无漏洞；提供明晰的身份鉴别和访问控制机制，按业务要求实现功能分级，并对用户分级授权；（3）信息保密：充分利用密码技术，对于需要保密的信息，采用密码技术进行加解密处理，防止信息的非授权泄漏，确保涉密信息在产生、存储、传递和处理过程中的保密,提供较完善的数据加密机制，确保数据存储和数据传输安全；（4）数据完整性：建立数据完整性检验机制，保证收发双方数据的一致性，防止信息被非授权修改。（5）可审计：系统具备日志跟踪与分析功能，提供访问、修改、删除等的用户操作日志，提供丰富的查询方式，对事件进行分析，并能提供预警信息,供追溯和追责，系统日志支持留存6个月以上。（6）数据备份：系统具备数据的本地备份机制，每日备份至本地；系统中存在的核心关键数据，具备异地数据备份功能，可备份至指定的服务器或存储系统上；每周制作数据的备份在指定服务器，确保系统一旦发生故障时能够快速恢复，备份数据不得更改；业务数据支持定期、完整、真实、准确地转储到不可更改的介质上，并支持集中保存。（7）隐私保护和数据安全 本项目建设过程中，项目承建方不可避免的会接触到一部分含有隐私内容的数据，要求在项目承建方法律层面承担相应的保密责任。承建方需从数据采集、处理、加载、发布等各个环节设计和执行严格的隐私保护和和数据安全保障体系，确保不会因为个人主观和操作失误等因素导致隐私泄露和数据外流，否则将承担相应的法律责任。（8）支持https安全传输协议，支持兼容TLSv1.2以上版本使用；（9）系统具备数据保密机制，数据存储经过密码保护，针对项目中的数据信息安全等内容会签署保密协议，防止信息泄露；（10）中标单位确保系统运行环境的安全运维和防护；（11）平台数据来源保证合法合规。 |
| **4.其他技术要求****（1）技术总体要求**1. 符合国家有关规定：系统建设必须符合我国相关部门制订的标准，安全策略、密码与安全设备选用、网络互联、安全管理等方面必须符合我国信息安全法律法规。
2. 统一信息标准：支持学校智慧校园统筹规划的应用需求和未来发展，符合并遵守学校制定的教育信息化技术规范、软件设计与开发规范、软件设计开发标准等标准规范要求，系统的设计和开发遵循学校统一的信息标准。
3. 集成对接要求：与学校已建成或待建的公共数据中心、统一身份认证；提供和第三方数据交换、集成规范和标准、开放并提供数据接口。
4. 具备良好的兼容性：须跨主流浏览器运行，支持IE10以上系列、Google chrome、360浏览器等主流浏览器，并保证各浏览器网页显示效果的一致性。
5. 系统须部署在学校的虚拟化平台上，采用分布式高可用性部署；系统用户端和管理端均应采用完全的B/S架构。

**（2）先进性和稳定性** 支持高并发，能够支持同时1000人以上在线使用平台。支持亿万级数据量存储，可以满足大容量、安全性、稳定性要求。**（3）系统部署** 系统部署在学校的虚拟化平台上；采用H5开发技术；所提供的软件平台需支持IPV6网络协议。 |
| **5.项目实施与进度计划**

|  |  |  |
| --- | --- | --- |
| **序号** | **项目任务** | **计划完成月份** |
| 1 | 设备安装部署 | 2023年11月 |
| 2 | 设备验收 | 初验：2023年12月终验：2024年6月 |
| 3 | 资金支付 | 自合同签订后，支付合同总价的40%为预付款；产品交付后进行初验，初验通过后支付合同总价的30%；试运行150日后进行终验，终验验收合格后，支付合同总价的30%。  |

 |
| **6.项目成果交付清单**在本项目的开发过程中和交付使用后，交付各个阶段的成果和文档资料并提供明确的交付清单。成果和文档资料必须符合软件工程的相关要求。要交付的成果和文档资料主要包括但不限于以下部分：（1）可执行代码和源代码：项目开发过程中为用户方定制功能的源代码；（2）项目文档：项目开发中的各种技术文档，包括用户手册、管理员手册、开发环境配置说明、需求分析说明、方案设计、参数设定、软件说明书、数据库结构、数据结构（数据字典）、代码清单、软件工具清单、需求及实施变更说明、需求跟踪矩阵、API接口手册以及有关系统集成接口的技术说明文档、元数据模型说明、系统设计说明、系统维护说明、系统培训资料，系统测试报告、试运行报告、《信息系统安全等级保护项目计划书》、《系统等级测评方案》、《系统等级测评报告》、《信息系统安全等级保护定级备案证明》和《信息系统整改建议书》等；（3）管理文档：包括项目开发中的一些工作文档，如实施计划、阶段报告、讨论纲要、会议记录，培训资料、竣工文档（含项目实施总结、系统配置情况等）、验收测试报告（测试计划、测试用例和测试结果等）等；（4）提供项目实施过程中的过程记录，如文档管理、版本管理、变更记录等；（5）项目完工后提供验收报告、实施方案、进度安排、项目周报等。交付的所有成果应汇集成册交付给学校，并提供成果的电子化版本。 |

**四、验收及售后服务**

|  |
| --- |
| **1.项目验收要求**（1）项目初验：系统开发、部署、调试完成后，由学校组织项目初验。系统初验的基本条件是：全面完成系统的设计、开发、测试和集成工作，系统安装调试，并进行相关的配置和系统优化的调试，达到功能、性能、使用等方面的要求；完成系统部署实施并上线运行；系统运行稳定；完成实施过程中所有文档的提交（设计说明书、测试文档、用户使用手册等）。由采购单位组织项目初验，初验通过后，即进入试运行阶段。（2）项目终验：试运行150日后，由学校组织专家按照公开招标文件、合同条款、软件工程要求和实际应用效果对项目进行验收。在本项目履约完成后，对学校后期提出的软硬件系统扩展、升级、兼容、对接等需求，供应商须承诺实现互联互通。终验验收费用由中标单位承担。 |
| **2.付款方式**自合同签订后，支付合同总价的40%为预付款；产品交付后进行初验，初验通过后支付合同总价的30%；试运行150日后进行终验，终验验收合格后，支付合同总价的30%。 |
| **3.质保及售后服务**1.质保期:自验收合格之日起5年。2.在质保期内，中标方应每年免费提供1-2次系统跨版本升级或系统维护服务，保障产品最新功能与中标方官方网站发布内容一致和负责保障系统安全运行。2.在质保期内，中标方应负责保障系统安全运行，并免费、及时提供系统跨版本升级或系统维护服务，保障产品最新功能与中标方官方网站发布内容一致。3.在质保期间若出现故障，中标方应在接到采购方通知后1小时以内响应，2小时内维修人员到场，12小时内排除故障。质保期结束后，承诺给予必要的服务保证系统正常运行，在软件使用过程中若发现安全漏洞，及时修复，并在3天内完成漏洞修补。4.所提供的服务内容应严格按国家最新发布的规范标准执行，符合采购人宣传理念及要求，如发生质量问题由供应商承担全部责任。5.培训服务应包括产品使用操作、简单维护等培训内容。供应商需按采购人要求的时间为采购人免费培训技术人员若干名，培训服务以受培训人员熟练掌握相应技能为原则，每年需提供2次以上的线下或线上免费培训。在产品投入使用初期进行必要的跟踪指导，保障产品的稳定运行。软件产品需在甲方现场培训的，供应商应按要求履行，供应商培训人员在培训过程产生的交通费、食宿费、培训费等均由供应商承担。 |

**五、商务要求**

|  |
| --- |
| 1、提供西安外国语大学多语种智能翻译与舆情大数据分析平台二期完整和一期所有接口对接承诺函；2、中标单位完成网络安全等级保护测评（二级）工作并协助学校完成等级保护测评备案，测评费用由中标方承担。 |