1**50种中医古籍知识元自动抽取项目技术要求**

本项目拟对150种中医古籍基于知识元理论以及表示方法，利用自然语言处理等信息处理技术，辅助以人工审核与处理，分析四类不同文献的150种中医古籍各类知识元核心要素，并自动抽取出各类知识元内容，为后续人工标引中医古籍提供基础数据，提高人工标引的工作效率。

1. **名词解释**

* 知识元

知识元是具有完备[知识](https://baike.baidu.com/item/%E7%9F%A5%E8%AF%86/74245)表达的知识单位。在中医古籍的文字内容中，它是描述某个特定对象的属性。比如描述病证对象的病证名、病因病机、证候表现等属性；描述方剂对象的组成，功用主治等属性。

* 知识元的核心要素

知识元的核心要素是指可以充分表达某一类知识元的特征的语义成分，比如对于方剂的“组成”知识元，各种药名和炮制方法名即是“组成”知识元的核心要素。对于病证的“病因病机”知识元，“脾胃亏损、虚火上炎等语义成分即是其核心要素。

1. **项目目标**
2. 分析各类知识元的核心要素

利用人工标引的中医古籍样例数据，通过计算机技术自动分析并提取各类知识元的核心要素。然后人工对计算机自动提取的结果进行审核和分析，并基于人工审核与分析的结果，改进提取知识元核心要素的算法。经过反复迭代，获得比较准确的各类知识元的核心要素。最后将自动抽取出的核心要素按如下逻辑格式输出。具体输出文件格式可以是json或Excel文件。

|  |  |  |  |
| --- | --- | --- | --- |
| **知识元类型** | **核心要素** | **样例原文** | **样例出处** |
| 病因病机 | 虚火上炎 | 亦有肾经***虚火上炎***者，加味地黄汤 | 女科心法 |
| 舌象 | 舌白 | ***舌白***者，痰也。 | 王旭高临证医案 |

**表1: 知识元的核心要素的输出格式**

1. 自动抽取各类知识元的内容

利用计算机技术，按一定规则和算法，从150种中医古籍原文中抽取出各类知识元的内容。每条自动抽取出的知识元记录由三个字段构成，知识元类型（病证名、方名等），知识元内容（出现在原文中的，描述知识元内容的单一名词或短句），知识元出处（知识元的来源书目名称）。最后将自动抽取出的知识元按如下逻辑格式输出。具体输出文件格式可以是json或Excel文件。

|  |  |  |
| --- | --- | --- |
| **知识元类型** | **知识元内容** | **知识元出处** |
| 病证名 | 温病 | 温病条辨 |
| 方名 | 桂枝汤 | 温病条辨 |
| 组成 | 桂枝(六钱)　芍药(三钱，炒)　炙甘草(二钱)　生姜(三片)　大枣(二枚，去核) | 温病条辨 |

**表2: 知识元的内容的输出格式**

1. **需处理的知识元类型**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 知识元类型 | 知识元说明 | 处理要求 | |
| 分析核心要素 | 抽取知识元的内容 |
| 1 | 病证名 | 疾病的名称或证候的名称。 | ✓ | ✓ |
| 2 | 方名 | 方剂的名称。 | ✓ | ✓ |
| 3 | 药名 | 中药的名称。 | ✓ | ✓ |
| 4 | 病因病机 | 发病的原因和病证发生、发展以及传变的机理。 | ✓ | ✓ |
| 5 | 证候表现 | 病证表现出来的症状体征，包括舌象、脉象等。 | ✓ | ✓ |
| 6 | 舌象 | 证候表现中对舌象的描述。 | ✓ | ✓ |
| 7 | 脉象 | 证候表现中对脉象的描述。 | ✓ | ✓ |
| 8 | 辨证 | 对病证的病因病机、证候表现的综合分析，以及相应的治则治法、处方、用药。 | ✓ | ✓ |
| 9 | 治则治法 | 治疗病证的原则和方法。 | ✓ | ✓ |
| 10 | 疗法 | 治疗病证的方法，包括外治、饮食、祝由、推拿、导引、炼养、针灸等 | ✓ | ✓ |
| 11 | 预后 | 预测病证的发展转归情况。 |  | ✓ |
| 12 | 宜忌 | 病证治疗过程中容或使用方剂时需要注意适用和避忌的内容。 |  | ✓ |
| 13 | 组成 | 组成方剂的药物及其炮制方法和剂量（包括药物、炮制和剂量） | ✓ | ✓ |
| 14 | 功用主治 | 方剂或药物的功效和治疗的病证。 | ✓ | ✓ |
| 15 | 制法 | 方剂的制备方法。 | ✓ | ✓ |
| 16 | 用法 | 方剂的服用或使用方法。 | ✓ | ✓ |
| 17 | 加减 | 根据兼夹证情况在本方的基础上添加或删减某些药物。 |  | ✓ |
| 18 | 方解 | 解析方剂组成的原理以及功用主治的由来。 |  | ✓ |
| 19 | 鉴别用方 | 与其他方剂的区别应用。 |  | ✓ |
| 20 | 生境 | 动植物药生长及矿物药形成的自然环境。 |  | ✓ |
| 21 | 性味 | 药物的四气五味、毒性、升降浮沉等。 | ✓ | ✓ |
| 22 | 归经 | 药物作用归属的脏腑经络。 | ✓ | ✓ |
| 23 | 加工炮制 | 药物的修治等粗加工过程、炮制过程及炮制品名等。 | ✓ | ✓ |
| 24 | 剂量 | 药剂的用药量 |  | ✓ |
| 25 | 五脏 | 胸腹腔内能贮存、分泌或制造精气的脏器 | ✓ | ✓ |
| 26 | 六腑 | 胆、胃、小肠、大肠、膀胱、三焦的总称 | ✓ | ✓ |
| 27 | 疗效 | 对疾病或证型治疗效果的描述。 |  | ✓ |
| 28 | 性别 | 关于患者性别的描述。 | ✓ | ✓ |
| 29 | 年龄 | 关于患者年龄的描述。 | ✓ | ✓ |
| 30 | 职业 | 关于患者身份或从事职业的描述。 | ✓ | ✓ |
| 31 | 体质 | 关于患者素体禀赋的描述。 | ✓ | ✓ |
| 32 | 时令 | 与患者疾病的发生发展有密切关系的时间、气候、物候等因素。 | ✓ | ✓ |

**表3: 需要处理的知识元类型列表**

1. **需处理的古籍书目**

|  |  |  |
| --- | --- | --- |
| 序号 | 古籍书目名称 | 古籍书目类型 |
| 1 | 女科心法 |  |
| 2 | 王旭高临证医案 |  |
| … | … |  |
| 150 |  |  |

**表4: 需要处理的古籍书目列表**

**医案古籍：（23种）**

石山医案、韩氏医通、名医类案、孙文垣医案、马培之医案、醉花窗医案、静香楼医案、古今医案按、吴鞠通医案、南雅堂医案、程杏轩医案、薛氏医案、三家医案合刻、张聿青医案、柳选四家医案、读医随笔（周学海）、花韵楼医案、洄溪医案、旧德堂医案、邵兰荪医案、王孟英医案、王氏医案译注、王旭高医案

**妇科古籍：（46种）**

经效产宝、坤元是保、卫生家宝产科备要、妇人大全良方、万氏女科、广嗣纪要、产宝百问、陈素庵妇科补解、女科证治准绳、产鉴、妇人规、女科百效全书、妙一斋医学正印种子编、祈嗣真诠、女科正宗、傅青主女科、女科切要、胎产指南、济阴近编、女科指掌、胎产大法、胎产心法、妇科心法要诀、树蕙编、盘珠集胎产症治、沈氏女科辑要、妇科玉尺、妇科冰鉴、宁坤秘籍、竹林寺女科秘书、彤园妇科、坤中之要、济阴宝筏、妇科指归、女科要旨、妇科备考、求嗣指源、秘诊济阴、广生编、仁寿镜、妇科良方、广嗣要语、邯郸遗稿、叶氏女科证治、郑氏家传女科万金方、种子心法

**温病古籍（33种）**

痧胀玉衡书、温热暑疫全书、广瘟疫论、温热论、痢疾论、湿热条辨、伤寒温疫条辨、松峰说疫、疫疹一得、温病条辨、温热病指南集、类伤寒集补、感证集腑、霍乱论、温热经纬、广温热论、 温病指南、六因条辨、南病别鉴、时病论、温热逢源、六气感证、伤暑论、伤燥论、伏邪新书、时疫辨、 鼠疫汇编、 时疫白喉捷要、 说疫全书、暑疟发原、伤寒经正附、燥气总论、痘疹心法

**儿科古籍：（36种）**

颅囟经、幼幼新书、小儿卫生总微论方、陈氏小儿病源方论、陈氏小儿痘疹方论、仁斋直指小儿方论、活幼心书、活幼口议、类证注释钱氏小儿方诀、全幼心鉴、婴童百问、幼科发挥、幼科种痘心法要旨、片玉心书、育婴秘诀、幼科心法要诀、痘疹心法要诀、片玉痘疹、保婴撮要、幼科证治准绳、婴童类萃、保幼新编、幼科要略、幼幼集成、幼科辑要、幼科释谜、福幼编、小儿诸热辨、幼科医学指南、幼科指归、慈幼便览、保赤新编、《抱乙子幼科指掌遗稿》（亦名《幼科指掌》）、慈幼新书、幼科铁镜、鬻婴提要说

**其他（12种）**

外科大成、外科精义、外科枢要、外科真诠、外科精要（陈自明）、外科正宗、外科准绳、普济本事方、串雅内编、串雅外编、杨氏家藏方、扶寿精方