深度应用驱动的医学知识图谱构建

数字医学知识中心 徐美兰
浙江数字医疗卫生技术研究院
2021.3.27

开放医疗与健康联盟
Open Medical and Healthcare Alliance

NGO 自组织 开放 开源 医疗 健康

内容

- 国内外医学知识图谱发展情况
- 医学知识图谱的领域特征和应用需求
- 数研院医学知识图谱构建
  - 模型建立
  - “七巧板” 本体术语集构建
  - “汇知” 图谱构建
- 医学知识图谱应用案例

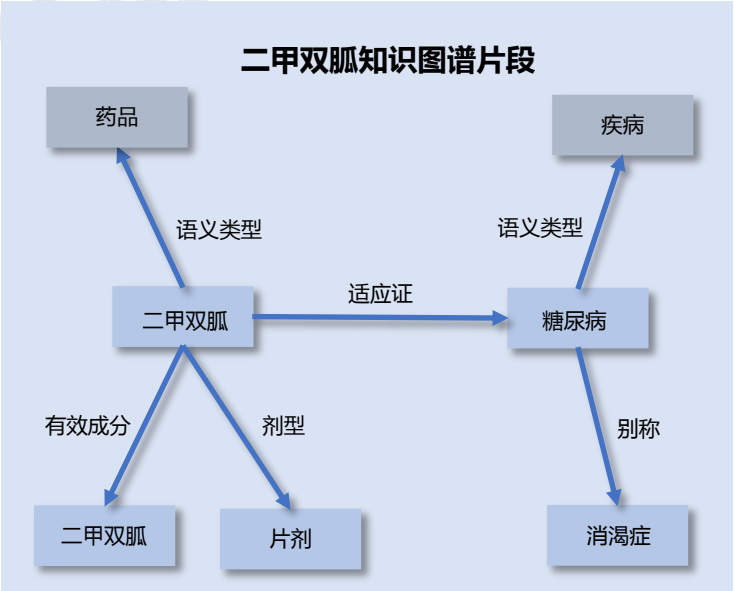
内容

- **国内外医学知识图谱发展情况**
- 医学知识图谱的领域特征和应用需求
- 数研院医学知识图谱构建
  - 模型建立
  - “七巧板” 本体术语集构建
  - “汇知” 图谱构建
- 医学知识图谱应用案例

知识图谱概念

知识图谱广义概念：作为一种技术体系，指大数据知识工程的一系列代表性技术的总称

知识图谱狭义概念：作为一种知识表示形式，知识图谱是一种大规模语义网络，包含实体、概念及其之间的各种语义关系



概念摘自《知识图谱 概念与技术》（肖仰华等编著）

国外医学知识图谱：UMLS

- **简介**
  - UMLS是美国国家医学图书馆自1986年起研究和开发的一体化医学语言系统，包含超级叙词表、语义网络、专业词典和词汇处理工具
    - 超级叙词表是一个非常庞大的术语库，集成了生物医学和健康方面的本体、叙词表、分类表、疾病编码集、专家系统、词汇表中的术语及相关信息，如有MeSH, SNOMED CT等
    - 语义网络：为超级叙词表中的概念提供统一的组织和分类，并揭示概念之间的语义关系
    - 专家词典和词汇处理工具：用于超级叙词表同义概念的自动归并
- **规模**
  - 语义网包含133种语义类型，54种语义关系。超级叙词表包含300多万概念，1300多万概念名称

www.nlm.nih.gov/research/umls

国外医学知识图谱：SNOMED CT

- **简介**
  - 2002年1月，SNOMED CT首次发布，它由两大医学术语SNOMED RT与CTV3合并而来，国际版SNOMED CT在每年的1月和7月各更新一次。SNOMED C核心构件是概念、描述（术语）和关系
- **样例**
  - Mild pre-eclampsia
- **规模**
  - 目前SNOMED CT包含19种语义类型，50多种语义关系，35万概念，120万描述（术语），110万关系

www.snomed.org

国内医学知识图谱：CUMLS

简介

- CUMLS是中国医学科学院医学信息研究所基于UMLS开发的中文一体化医学语言系统，包含医学词表、语义网、构建工具与平台
  - 医学词表涵盖了10余个生物医学领域内的主题词表、分类表、术语表及医学语料库，其中重要的来源词表包括《医学主题词表（中文版）》（MeSH中文版）、《中国中医药学主题词表》等
  - 语义网：基于美国UMLS建立，由两部分构成：语义类型、语义关系
  - 构建工具与平台：构建工具包括同义词识别工具、语义相似度计算工具、主题分类自动映射工具、主题词/副主题词自动组配工具；词表发布服务平台是中文一体化医学语言系统面向最终用户使用的平台

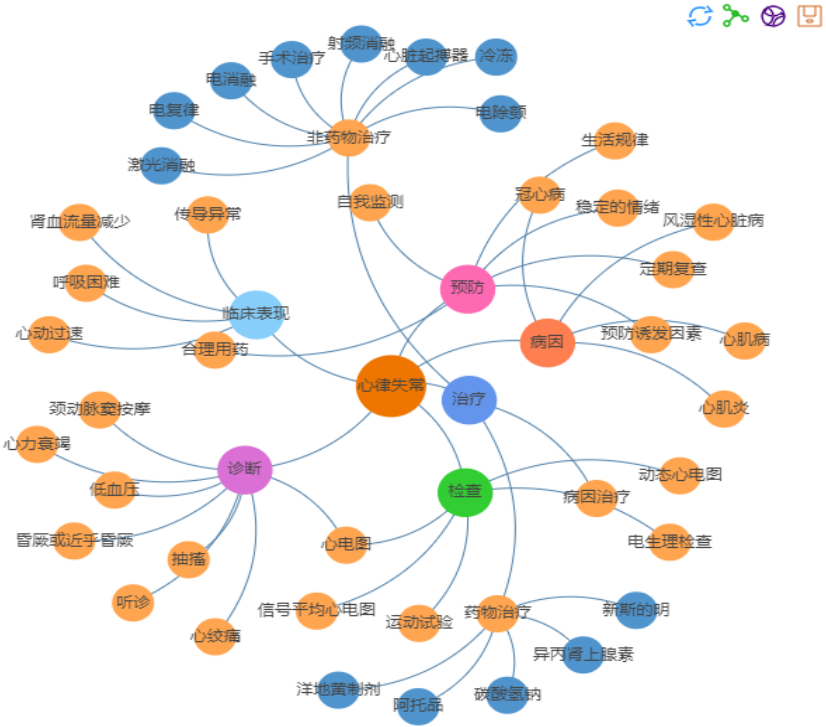
规模

- 共收录医学主题词3万余条、入口词3万余条、医学术语10万余条、医学词汇素材30万余条

国内医学知识图谱：医药卫生知识服务系统

医药卫生知识服务系统 Medical Knowledge Service System

- **简介**
  - 医药卫生知识服务系统由中国医学科学院医学信息研究所承建，通过对资源的深度挖掘和关联分析，建设了知识图谱、知识脉络分析等特色知识服务和应用
- **样例**
  - 心律失常
- **规模**
  - 医药卫生知识服务系统已发布疾病和药品领域知识图谱，其中疾病涵盖心脑血管疾病、呼吸系统疾病、免疫系统疾病、消化系统疾病、肿瘤等

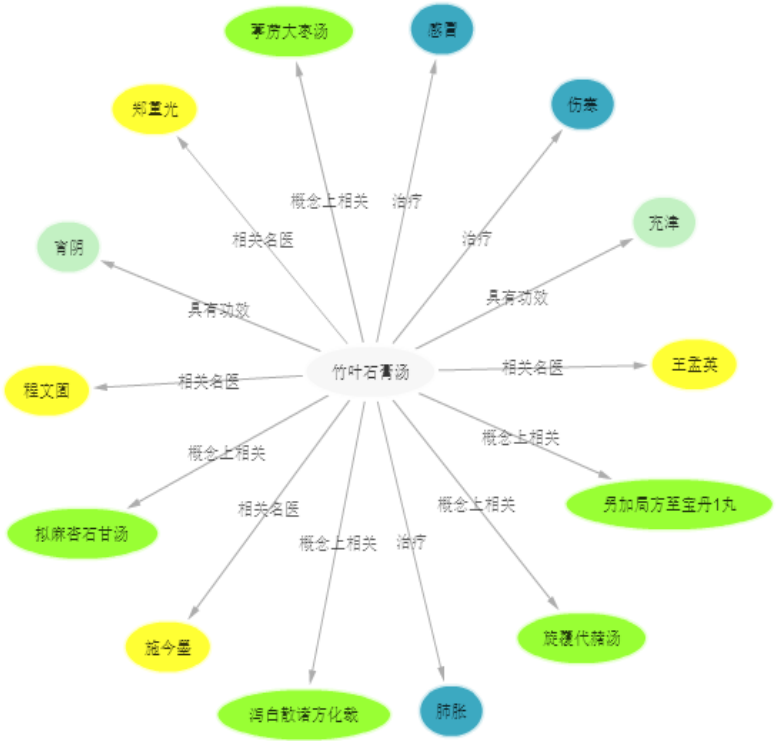


med.ckcest.cn

国内医学知识图谱：中医药知识图谱

中医药知识图谱
Knowledge Graph for Traditional Chinese Medicine

- **简介**
  - 中国中医科学院中医药信息研究依托中医药学语言系统（TCMLS）构建了中医药知识图谱。中医药知识图谱是面向中医药领域的知识图谱
- **样例**
  - 竹叶石膏汤
- **类型**
  - 共有9种知识图谱类型，包含“基于中医药学语言系统的知识图谱”、“中医美容知识图谱”、“中医养生知识图谱”、“中医临床知识图谱”等



www.tcmkb.cn/kg

国内医学知识图谱：OpenKG

- **简介**
  - OpenKG是由中国中文信息学会倡导的中文领域开放知识图谱社区项目，主要工作内容包括OpenKG.CN（开放图谱资源库）、cnSchema（中文开放图谱Schema）和Openbae（开放知识图谱众包平台）

- **主要工作内容**
  - OpenKG.CN：聚集了很多开放的中文知识图谱数据、工具、文献资源。主要有93个数据集，包括面向中文电子病历的命名实体识别数据集、病人事件知识图谱等
  - cnSchema：定义了中文领域开放知识图谱的基本类、术语、属性和关系等本体层概念
  - Openbase：以中文为核心，机器学习与众包协同；支持将知识图谱转化为Bots

openkg.cn

内容

- 国内外医学知识图谱发展情况
- **医学知识图谱的领域特征和应用需求**
- 数研院医学知识图谱构建
  - 模型建立
  - “七巧板” 本体术语集构建
  - “汇知” 图谱构建
- 医学知识图谱应用案例

医学知识的特点

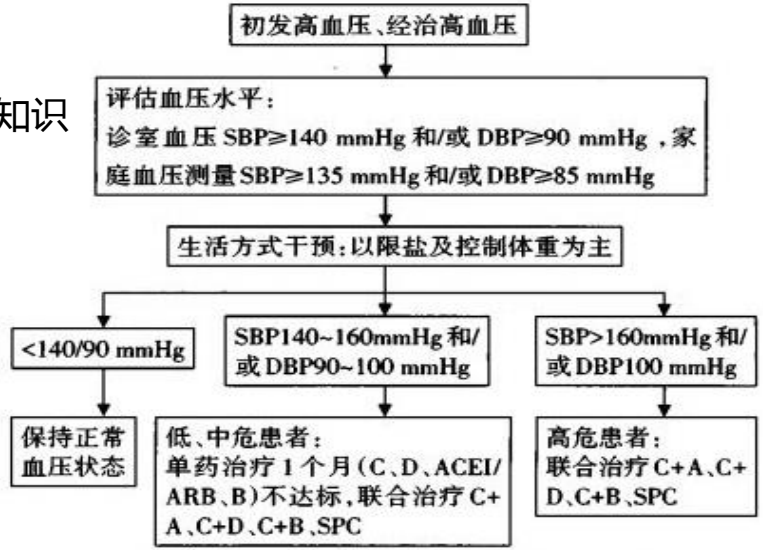
□ 医学术语多样性

不同知识源对同一个概念采用了不同术语进行表达。

医学知识的特点

- **精确度要求高**

医学知识专业性高，医学应用场景容错率低，因此医学知识图谱的精确度要求高。

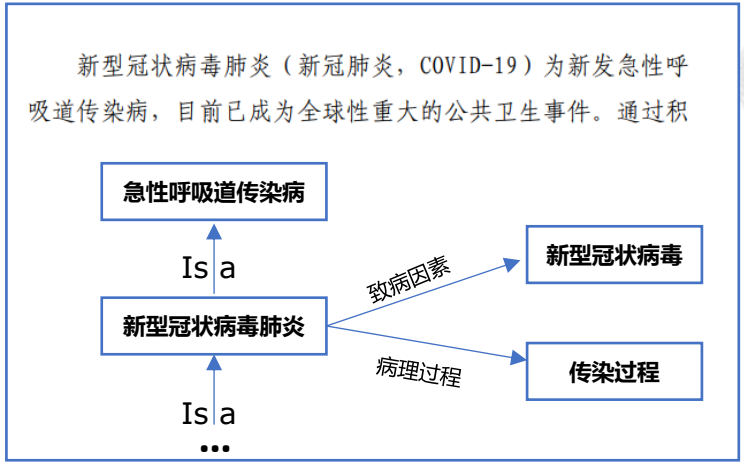


示例摘自《高血压基层诊疗指南(2019年)》

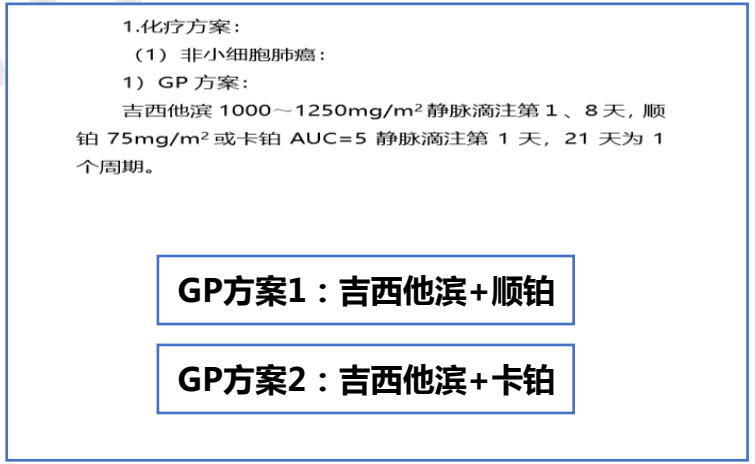
医学知识的特点

◘ 复杂程度高

医学是经验总结的科学，医学概念的内涵往往比较丰富，且有些医学知识复杂很难用简单三元组表达。



示例摘自《新型冠状病毒肺炎诊疗方案（试行第八版）》



示例摘自《原发性支气管肺癌临床路径（2019年版）》

医学知识图谱的不同应用场景需求侧重点也有所不同，需要最大化的满足才能提高图谱的适用性

1 语义搜索

- 需要建立完整的模型框架和层级体系，使关系最大化扩展，从而使搜索结果能够更加智能；
- 保证术语多样性，相同术语通过同一概念统领；

2 智能问答

- 关系类型丰富，除了概念内涵关系外，还需要更多经验性的关系；
- 除了涵盖专业的词汇外，也需要能覆盖到大众使用的口语化词汇；

3 决策支持

- 需要保证知识的严谨性、正确性和可解释性，为医生能提供准确、适当的决策支持；

4 数据分析

- 需要最大化解决语义异构问题，实现数据语义层面的分析和处理；

为满足行业深度应用需求，医学知识图谱构建时需引入更多定制化解决方案

医学知识图谱特点	实施方案	应用效果
术语多样化	• 收录尽可能多来源的术语； • 进行实体归一。	• 适用性强，满足多种场景应用需求，如病历、科研、互联网诊疗等； • 可有效解决语义异构问题，提高语义互操作能力。
精度高	• 以权威知识源为核心，如教科书、指南、文献、行业标准； • 图谱中记录知识来源； • 领域专家参与构建和审核。	• 可以按照应用场景对知识的精确度要求来选择知识源使用； • 具备可解释性、可信度高。
关系丰富	• 医学概念的内涵定义采用本体进行知识组织，对概念间的层级关系进行严谨定义； • 使用属性组，更加准确地表示复杂知识。	• 利用本体层级关系进行关系推理，获得更多关系，使知识得以最大化使用； • 计算机理解知识更准确、无歧义。

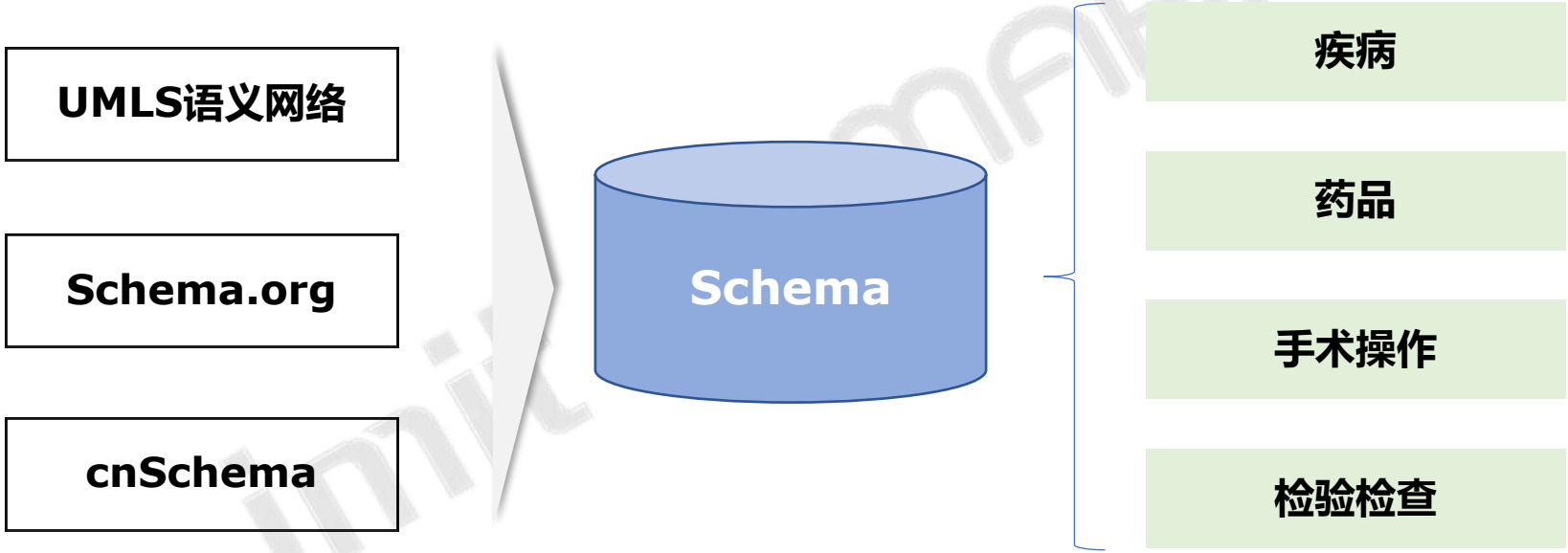
内容

- 国内外医学知识图谱发展情况
- 医学知识图谱的领域特征和应用需求
- **数研院医学知识图谱构建**
  - **模型建立**
  - “七巧板” 本体术语集构建
  - “汇知” 图谱构建
- 医学知识图谱应用案例

医学领域的知识图谱由于其知识专业性强，行业通常采用自上而下的方式，先构建Schema，再抽取知识

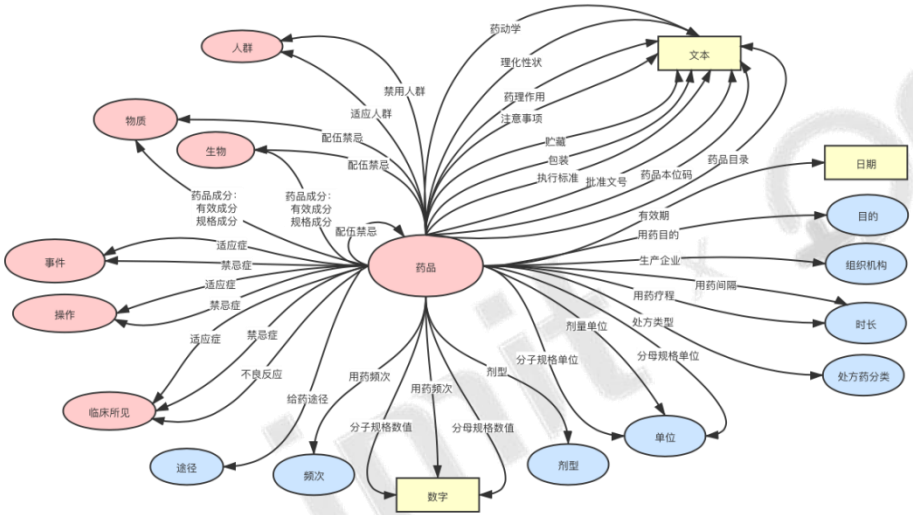
数研院医学知识图谱Schema主要参考了UMLS语义网络、Schema.org、cnschema等建立，涉及四大领域

在知识图谱的构建过程中，根据抽取和应用的实际情况，不断完善和优化Schema

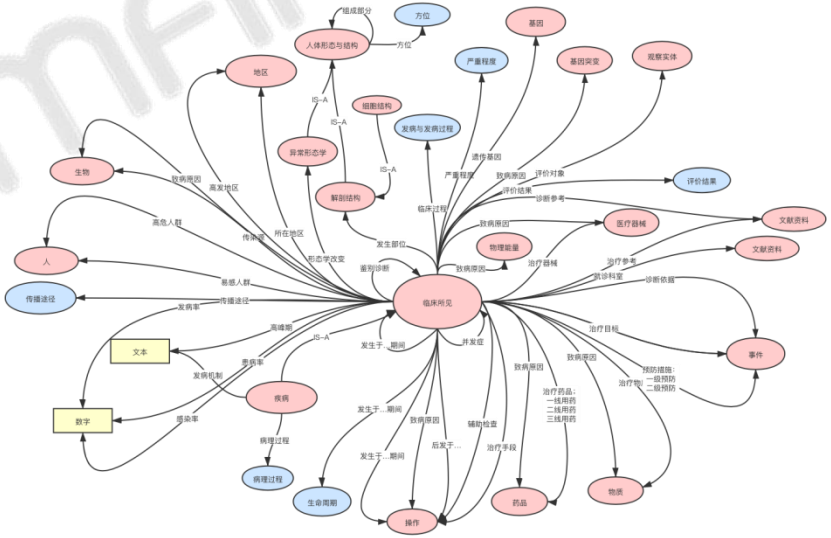


2019年8月首次发布Schema ，目前包含72种语义类型、493种语义关系

药品领域模型示例



疾病领域模型示例



Schema查询和下载网址：schema.omaha.org.cn

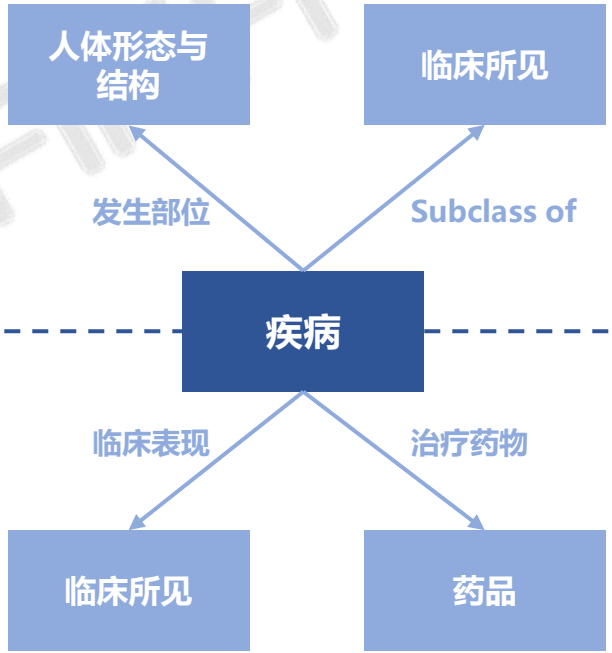
Schema分别用于指导“七巧板”医学本体术语集和“汇知”医学知识图谱的构建，完善医学知识表达的体系

OMAHA七巧板医学术语集
OMAHA Tangram Medical Terminology

- 采用本体，解决与逻辑定义（即内涵定义）相关的关系
- 层级关系

汇知
医学知识图谱

- 采用语义网络，解决可能性、经验性的关系
- 无层级关系

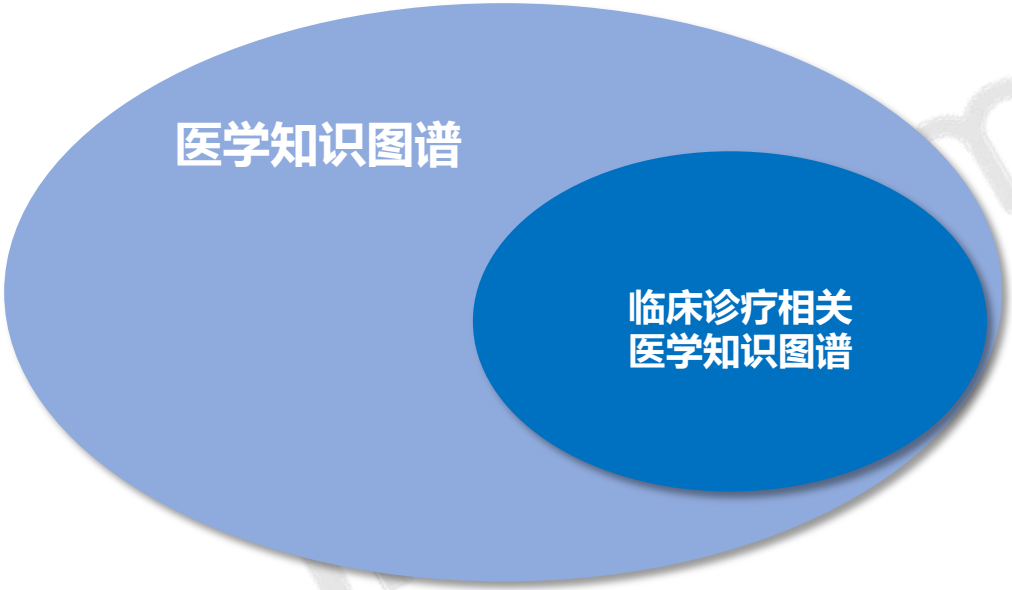


内容

- 国内外医学知识图谱发展情况
- 医学知识图谱的领域特征和应用需求
- **数研院医学知识图谱构建**
  - 模型建立
  - **“七巧板” 本体术语集构建**
  - “汇知” 图谱构建
- 医学知识图谱应用案例

一、确定领域范畴

以满足临床诊疗需求为切入点，开始尝试构建医学知识图谱



临床诊疗相关医学知识图谱

主要涉及范围：

- 疾病、症状、体征
- 手术操作、检验检查
- 药品
- 人体形态结构
- 基因
- 医疗器械
- …

二、选取合适的知识源

充分收录行业现行标准、教科书、指南等权威知识源，并同时补充临床病历、互联网诊疗中的术语

权威知识源

- **疾病、症状、体征：**内科学等书籍、指南文献、ICD-10、常用医学名词、MedDRA、元数据标准等；
- **手术操作、检验检查：**诊断学、ICD-9-CM-3、全国医疗服务项目、医疗机构临床检验项目目录、元数据标准等；
- **药品：**药品说明书、药典、医保目录、基药目录、NMPA药品注册信息、ATC等;
- **人体形态结构：**解剖学、FMA等；
- **基因：**NCBI、HGVS数据库等；
- **医疗器械：**NMPA医疗器械分类目录、各地医疗耗材目录等；

其它知识源

- 临床病历中抽取的医生习惯用术语；
- 互联网诊疗系统中梳理的术语；
- 医学类开放资源中的术语；
- 临床专家；

三、梳理重要术语

梳理领域中的重要术语，并由领域专家进行语义层面的实体归一，完成概念化

梳理术语 → **标准化处理** → **机器推荐** → **人工审核**

- **语义标签分类整理；**
- **去重处理；**

- **格式标准化：**编码类型、全半角等；
- **文本标准化：**繁体换简体、大写数字统一为阿拉伯数字、罗马数字统一、英文字符大小写转化、标点符号统一；

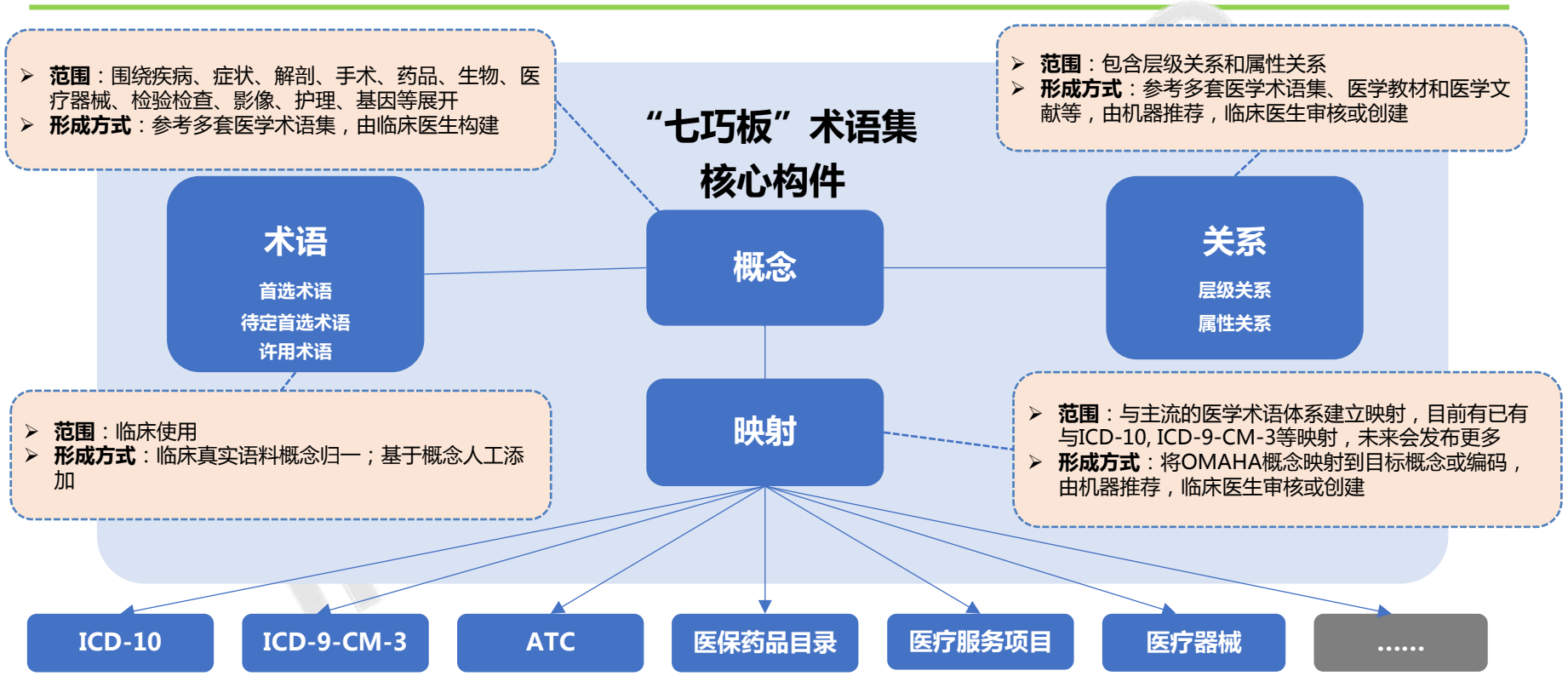
- **语形相似度：**编辑距离（Damerau-Levenshtein距离）；
- **语义相似度：**词向量（word2vec、bert）；

- **至少1次人工审核；**
- **同语义标签、不同语义标签的一词多义问题审核；**
- **由领域内专家完成，避免跨领域；**

- 以1个基础版本为中心体，尽量不打乱原层级结构；
- 同类文件可首先进行归一，提高效率，如ICD-10；
- 根据不同语料特点选择不同的方法，而非固定模式；
- 为方便后续应用，可制定规则指定概念下的首选术语、许用术语等术语类型。

四、建立关系

“七巧板”医学本体术语集的核心构件包括：概念、术语、关系及映射



充分保留知识源中的已有层级关系，通过机器推理、人工添加的方式进行优化

（一）层级关系

关系继承	• 知识源中天然已有的层级关系，可以直接继承使用；继承过程中要对某些层级关系进行进一步的调整，尤其是有**冲突**的部分。
机器推荐	• 逆向最大匹配算法在药品的层级关系建立中比较有效； • **词汇解析引擎**：从词的构成出发，快速实现父节点推荐，如组合特征词"肺炎所致的咳嗽"isA"咳嗽"，修饰词“急性肺炎”isA“肺炎”。
人工建立	• 知识源中进一步人工抽取潜在的层级关系； • 根据专家经验进一步添加、调整。
逻辑验证	• 层级关系不能成环，如A is a B, B is a C, C is a A； • 是否有多余关系，如A is a B, B is a C, A is a C。
本体推理	• 本体雏形形成后，可以用本体推理机（如HermiT）进行本体关系推理和关系校验，进推理机前需将文件先转换为推理机可用的格式。

挖掘知识源中的属性关系，并通过机器推荐、人工添加进行补充

（二）属性关系

关系继承	• 有些知识源中自带属性关系，如医保目录中的剂型。
机器推荐	• 构建属性值分词词典，对术语进行分词，如部位、急慢性、方位等。
人工建立	• 知识源中进一步人工抽取潜在的属性关系，根据专家经验进一步添加、调整； • 为某些属性关系建立“属性组”，减少理解歧义，如“皮肤脓肿伴淋巴管炎”，属性关系发生部位“皮肤”和形态学改变“脓肿”，发生部位“淋巴管”和形态学改变“炎症”需要分别成组。
逻辑验证	• 子节点的属性关系原则上比父节点更丰富，属性值颗粒度更细； • 不能存在多条相同的属性关系。
本体推理	• 利用本体推理机进行本体关系推理和关系校验，进一步优化关系。

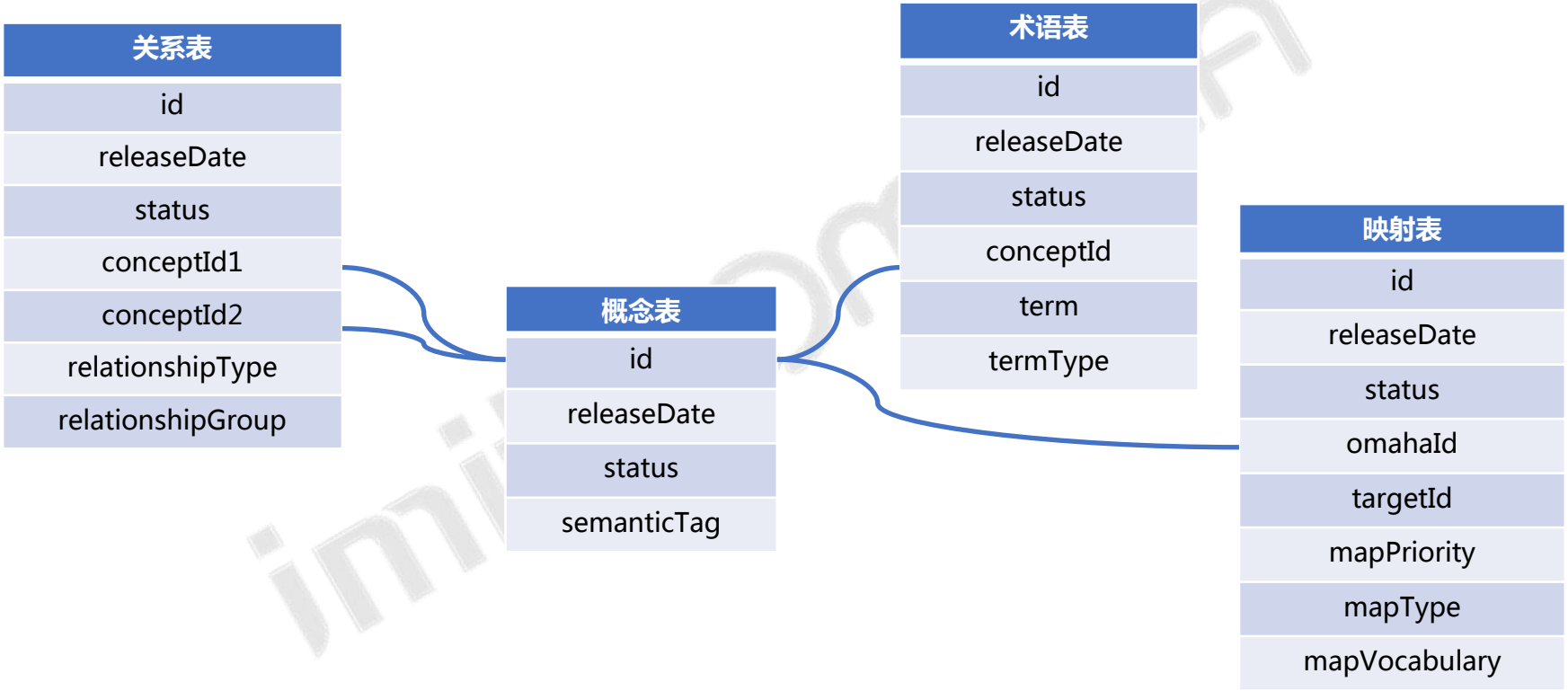
制定明确的映射规则，采用机器推荐、专家审核的方式建立映射

（三）映射关系

确定规则	• 确定映射类型：等同映射、等级映射（上位映射、下位映射）、相关映射； • 建立映射的优先级，优先建立等同映射； • 建立关系最近的映射，可推理得到的映射不重复建立，如A is a B, A equivalence mapping to C, B narrower mapping to C。
机器推荐	• 本质还是语义相似度推荐。
专家审核	• 根据专家经验进一步审核。
逻辑验证	• 主要根据映射规则验证有无冲突。

五、存储和浏览

采用关系型数据库，分为概念表、术语表、关系表、映射表进行存储，且保留历史痕迹



术语浏览器实现术语集构件的快速查找，并可按需实现子集定制

术语浏览器实现术语集构件的快速查找，并可按需实现子集定制

术语浏览器实现术语集构件的快速查找，并可按需实现子集定制

六、平台及工具支撑

研制知识库维护平台（CoWork），内嵌术语集研制规则，支持多人共同协作

“七巧板” 协作平台功能

- **概念编辑**：功能类似protégé ，可实现概念层面的新增、删除、修改。
- **术语映射**：基于语义相似度推荐相似概念，可将语料收录至术语集或建立两套术语体系间的映射。
- **术语审核**：以特定范围为任务，开展术语审核。
- **一词多义**：将存在一词多义的概念形成任务，审核相关概念。
- **历史痕迹**：对修改的历史痕迹进行记录和管理。
- **任务管理**：创建项目，给不同用户角色配置不同的任务，对任务进度进行统计展示等。

术语集编辑器可实现概念层面的编辑功能需求，并支持多人同时在线协作

1、层级浏览　　2、概念编辑　　3、前置规则、错误提醒

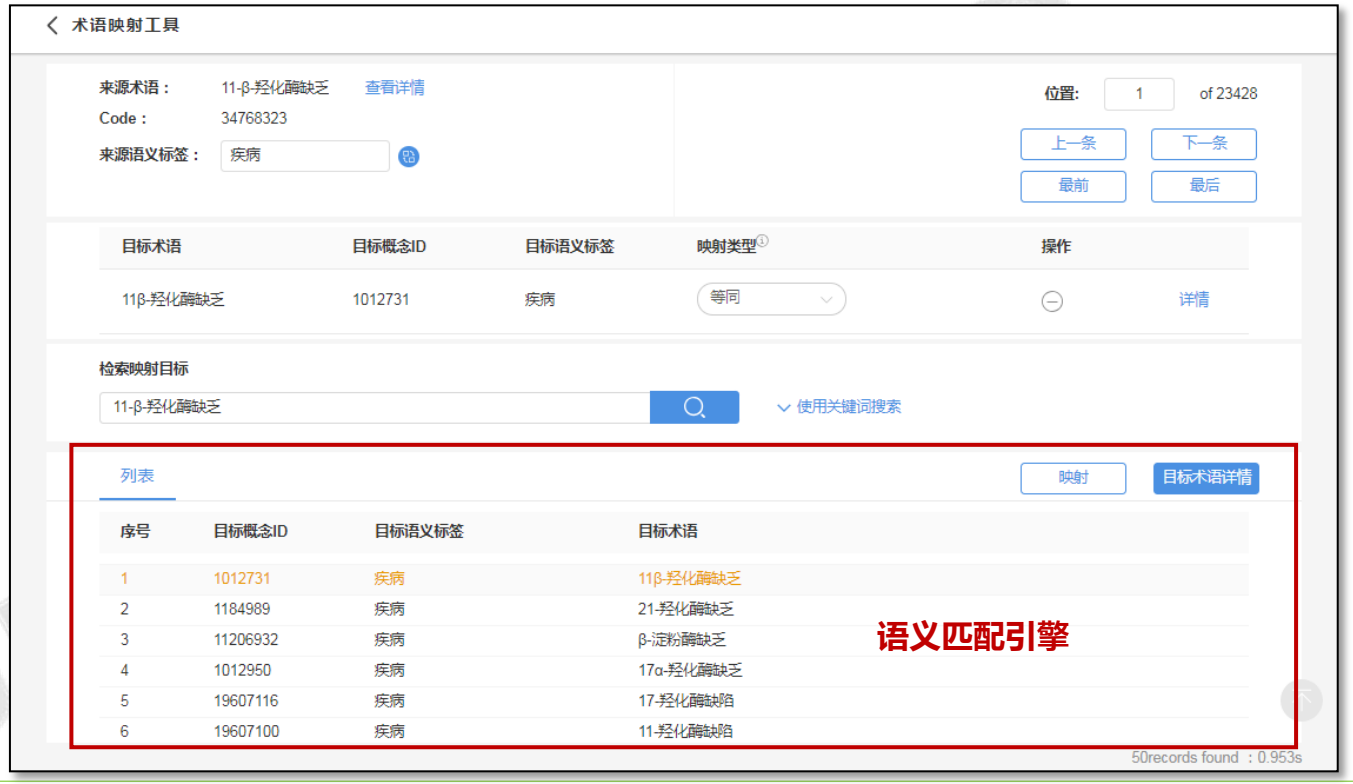
协作方式：不创建分支，采用编辑锁。

术语映射工具利用算法推荐，调高映射效率

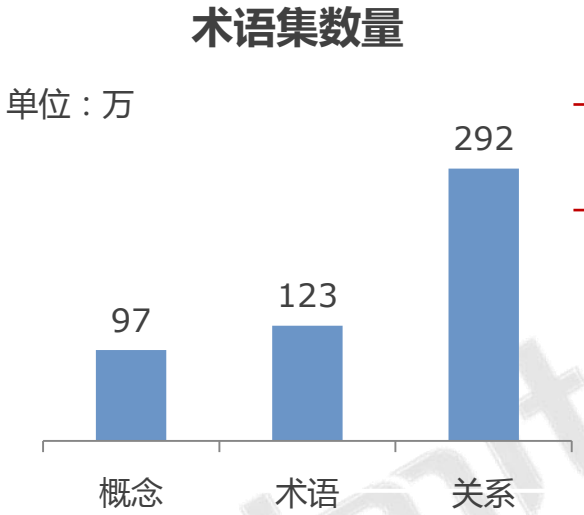
来源术语

映射结果

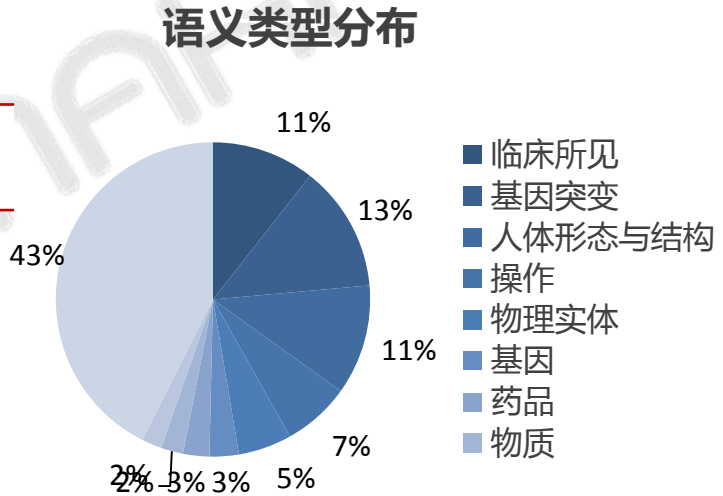
映射推荐



“七巧板”术语集目前收录97万概念、123万术语和292万关系，包含疾病、操作、药品等语义类型



持续维护更新
按季度定期发布



- 持续进行维护更新，按季度发布，每季度第一个月20号发布新版本

内容

- 国内外医学知识图谱发展情况
- 医学知识图谱的领域特征和应用需求
- **数研院医学知识图谱构建**
  - 模型建立
  - “七巧板” 本体术语集构建
  - **“汇知” 图谱构建**
- 医学知识图谱应用案例

一、选取合适的知识源

选取临床指南、临床路径、医学书籍文献等权威知识源，并同时补充医学百科类知识

非结构化知识源

- 临床指南
- 临床路径
- 医学书籍
- 医学文献
- ……

半结构化知识源

- 药品说明书
- 医学百科知识库
- 医保药品目录、基药目录
- ……

结构化知识源

- 七巧板医学术语集
- ICD-10、ICD-9-CM-3等行业标准术语集
- ……

二、知识抽取

基于规则的命名实体识别+专家审核提高标注效率，产生的标注数据用于训练深度学习模型

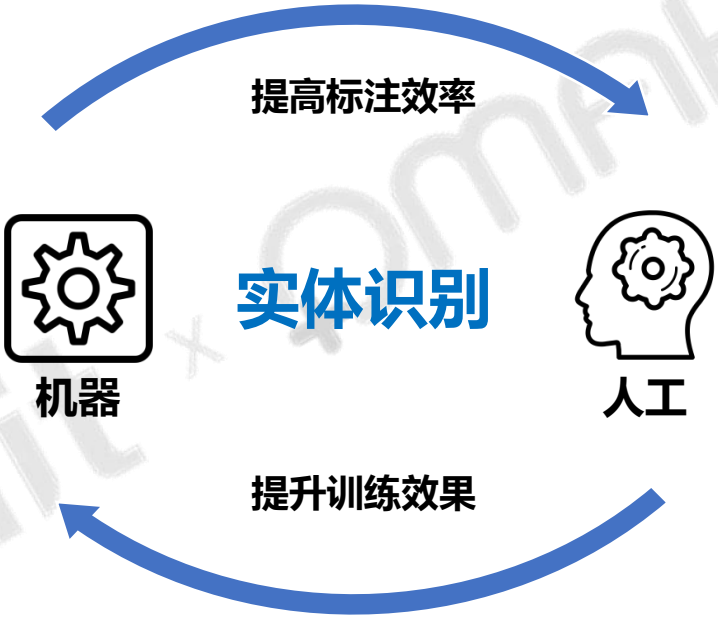
实体识别

基于规则

- 七巧板术语集为分词词典；
- 自定义词典：修饰词、否定词、无效词、连接词等
- 分词：HanLP、jieba分词等

基于深度学习

- 词向量：word2vec、bert等
- BiLSTM+CRF



专家审核

审核标注结果

- 错误标注修正
- 错别字表维护
- 文本清洗规则维护

标注新实体

- 规则词典更新
- 发现新语义类型

基于实体识别的结果，专家标注关系，产生的标注数据用于句法规则总结和半监督学习

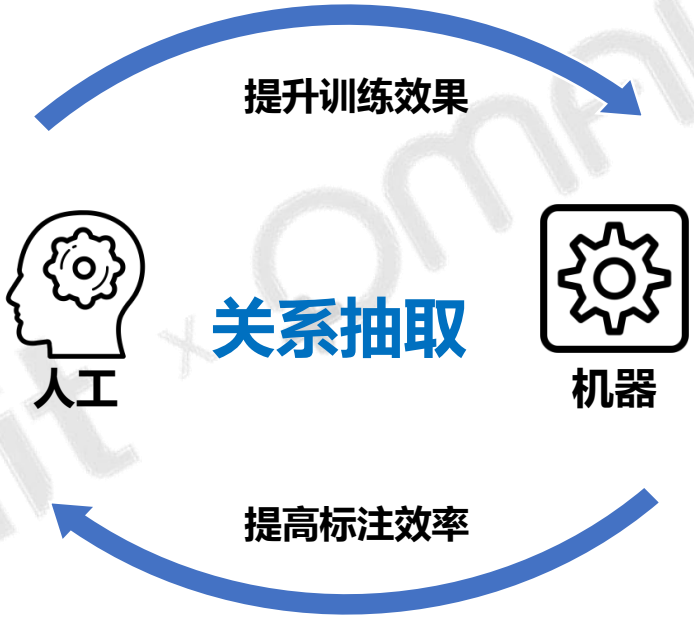
专家标注

标注关系

- 充分理解Schema中每条关系的定义
- 基于预标注实体的文本人工标注关系

发现新关系

- 完善Schema



自动关系抽取

基于规则

- 规则总结
- 句法分析

半监督学习

- Bootsrapping
- 远程监督

三、知识融合

最大化地将“汇知”图谱与“七巧板”术语集融合，可为图谱的深度应用打下基础

实体归一

- 去重处理；
- 依据语义相似度进行归一。

实体对齐

- 将抽取来的实体与“七巧板”术语集中的概念进行融合；
- 将抽取的实体作为术语集扩增的来源，丰富术语集的概念。

关系融合

- 依据实体归一结果，删除冗余关系；
- 实体对齐后，依据七巧板术语集中的层级关系，删除冗余关系。

- **保证抽取的原始三元组含义不变**
- **将“汇知”图谱与“七巧板”术语集最大程度融合**

四、知识存储和检索

除传统的三元组外，加入“属性组”和“来源”字段，使知识表达地更加准确，同时确保知识的可溯源性

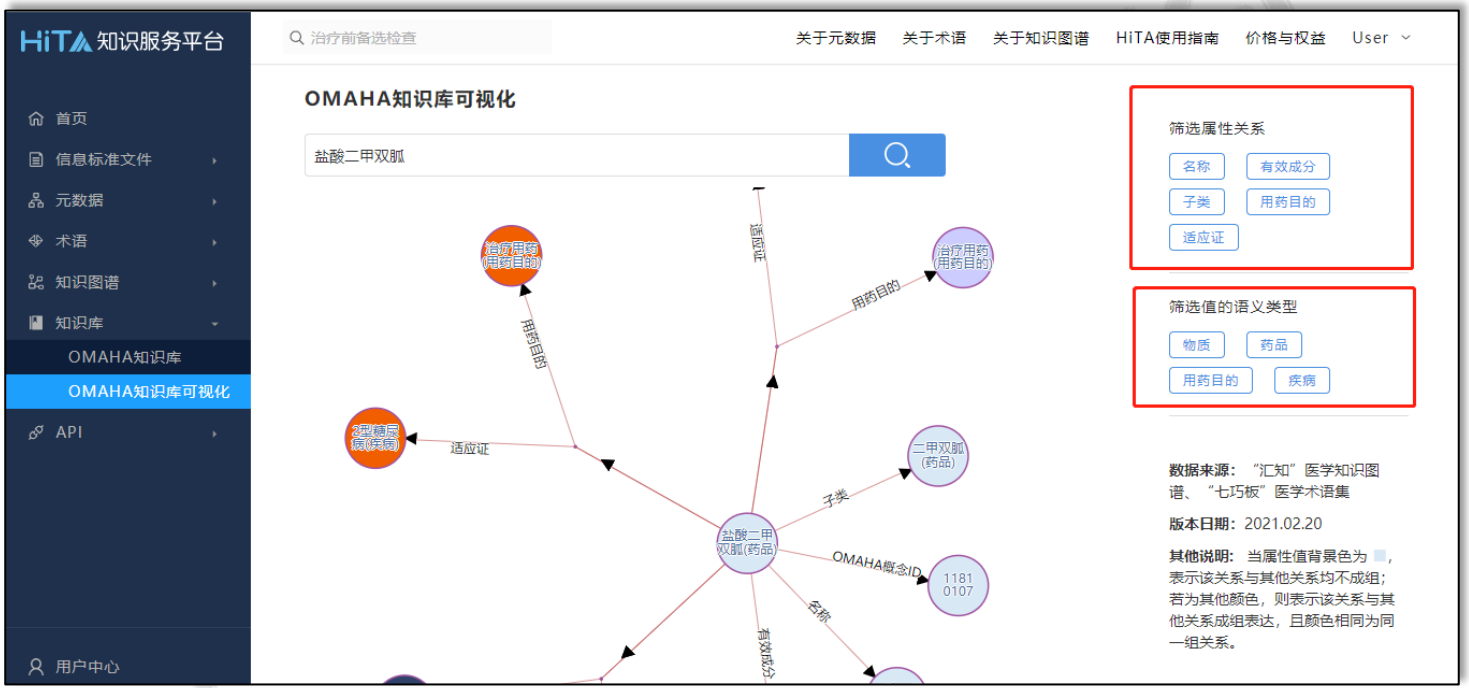
知识图谱表字段说明

字段名	字段含义	数据类型
entityId	实体ID	Longinteger
entity	实体名	String
entityTag	实体语义类型	String
property	属性名称	String
value	值	String
valueId	值ID	Longinteger
valueTag	值语义类型	String
group	关系组	String
source	来源	String

保留三元组的来源，满足三元组在不同场景应用的需求

entityID	entity	entityTag	property	valueId	value	valueTag	group	source
1567677479179763001	病毒性心肌炎	疾病	诊断相关检查	1567677492576762155	x线	观测操作	0	感染性心肌炎临床路径(2019年版)
1567677479179763001	病毒性心肌炎	疾病	诊断相关检查	1567677492927263558	心电图	观测操作	0	感染性心肌炎临床路径(2019年版)
1567677479222720145	病毒性心肌炎	疾病	治疗方式	1587707446344453773	卧床休息	事件	0	临床诊疗指南——心血管内科学分册
1567677479222720145	病毒性心肌炎	疾病	治疗方式	1576806136454450374	进富含维生素及蛋白质的食物	事件	0	临床诊疗指南——心血管内科学分册
1567677479179763001	病毒性心肌炎	疾病	治疗方式	1611149388480157512	抗感染治疗	操作	0	病毒性心肌炎临床路径(2010年版)
1567677479179763001	病毒性心肌炎	疾病	治疗相关检查	1567677492946147325	血气分析测定	观测操作	0	病毒性心肌炎临床路径(2010年版)
1567677479222720145	病毒性心肌炎	疾病	诊断相关检查	1576806136444877672	心包穿刺液检查	观测操作	0	儿童心肌炎诊断建议(2018年版)
1567677479222720145	病毒性心肌炎	疾病	诊断相关检查	1567677493030650420	活体组织检查	观测操作	0	儿童心肌炎诊断建议(2018年版)
1567677479223882683	病态窦房结	疾病	临床表现	1587707446232870828	心供血不足	临床所见	0	心动过缓临床路径(2017年县级医院版)
1567677479223882683	病态窦房结	疾病	临床表现	1576806136193850767	头晕眼花	临床所见	0	心动过缓临床路径(2017年县级医院版)

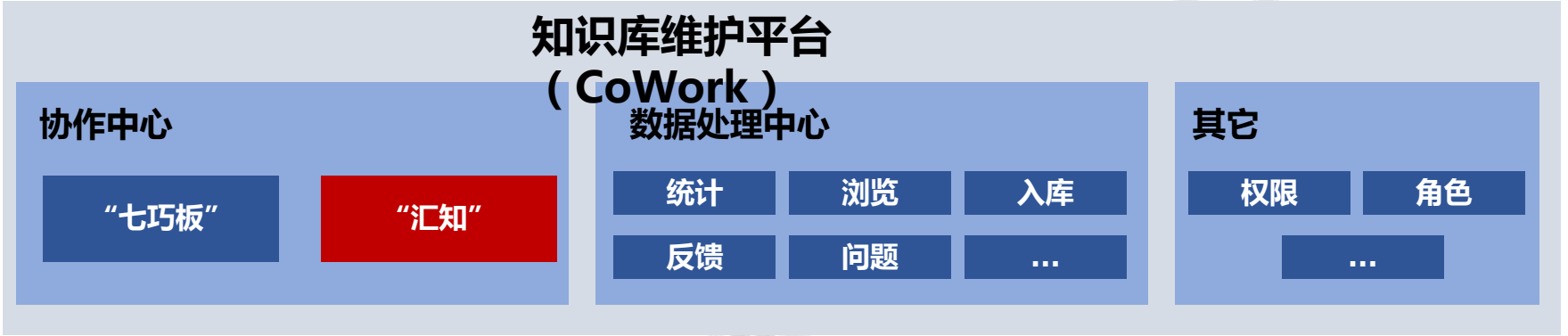
通过可视化搜索，可快速直观地查看图谱数据



HiTA知识服务平台（hita.omaha.org.cn）可查看可视化数据

五、平台及工具支撑

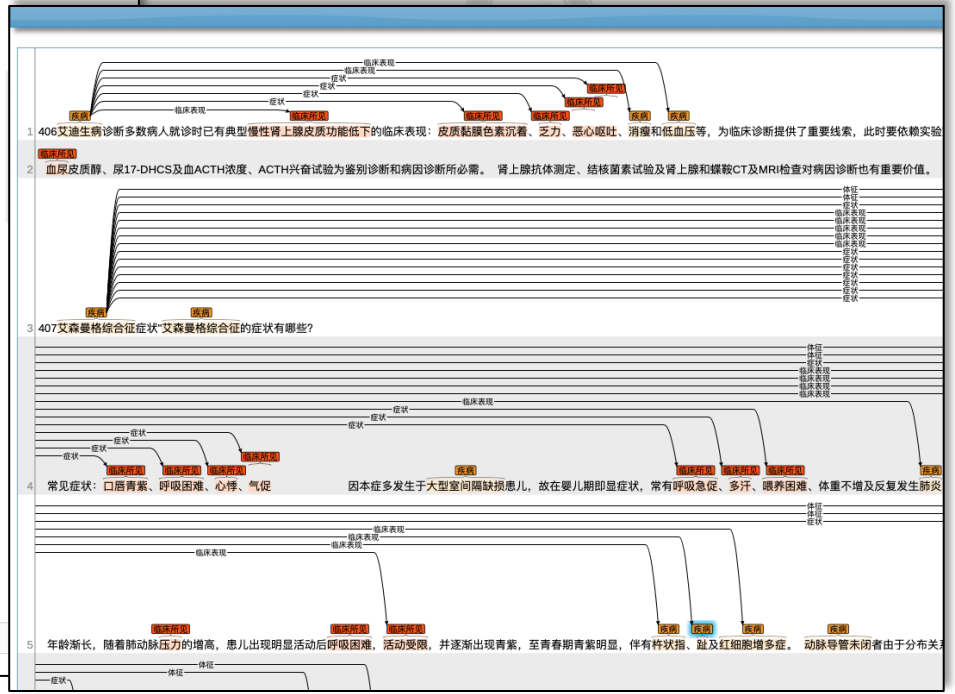
研制知识库维护平台（CoWork），内嵌知识图谱集研制规则，支持多人共同协作



“汇知” 协作平台功能

- **标注方案管理**：可选取Schema中的部分作为标注方案，同时也支持自定义方案。
- **文本上传**：需要标注的文本批量清洗后上传。
- **任务管理**：创建项目，给不同用户角色配置不同的任务，对任务进度、任务数量进行统计展示等。
- **自动标注**：对OMAHA Schema覆盖范围内的实体进行命名实体识别。
- **人工标注**：人工对标注结果进行审核和修正。
- **结果导出**：将标注好的结果导出。

创建多种自定义标注方案，批量上传和分配任务，在基于brat的文本标注工具上，各地志愿者可合作共建知识图谱



自定义标注方案

文本标注工具

“汇知”图谱目前已发布7个领域，共计约11万实体，82万三元组

药品适应证
- 3.7万实体
- 43万三元组

疾病-科室
- 3.1万实体
- 14.2万三元组

临床路径
- 1.8万实体
- 11.3万三元组

疾病-临床表现
- 1.5万实体
- 8万三元组

心血管系统疾病
- 0.6万实体
- 3万三元组

中毒
- 0.5万实体
- 2.4万三元组

新冠肺炎
- 700实体
- 3000三元组

...

- 持续进行维护更新，按季度发布，每季度第二个月20号发布新版本

数研院发起的知识图谱协作项目已持续开展5年，已有百名个人志愿者、多家优秀企业参与

企业贡献值榜单
（累计至2020年底）

排名	企业名称	总贡献值
1	杭州华卓信息科技有限公司	14600
2	上海依图网络科技有限公司	14100
3	上海柯林布瑞信息技术有限公司	12400
4	杭州朗通信息技术有限公司	10800
5	北京春雨天下软件有限公司（春雨医生）	6057
6	东软集团股份有限公司	920
7	平安医疗科技有限公司	912
8	树兰（杭州）医院	650
9	互动峰科技（北京）有限公司（好大夫）	244
10	深圳市腾讯计算机系统有限公司	68

个人贡献值榜单
（累计至2020年底）

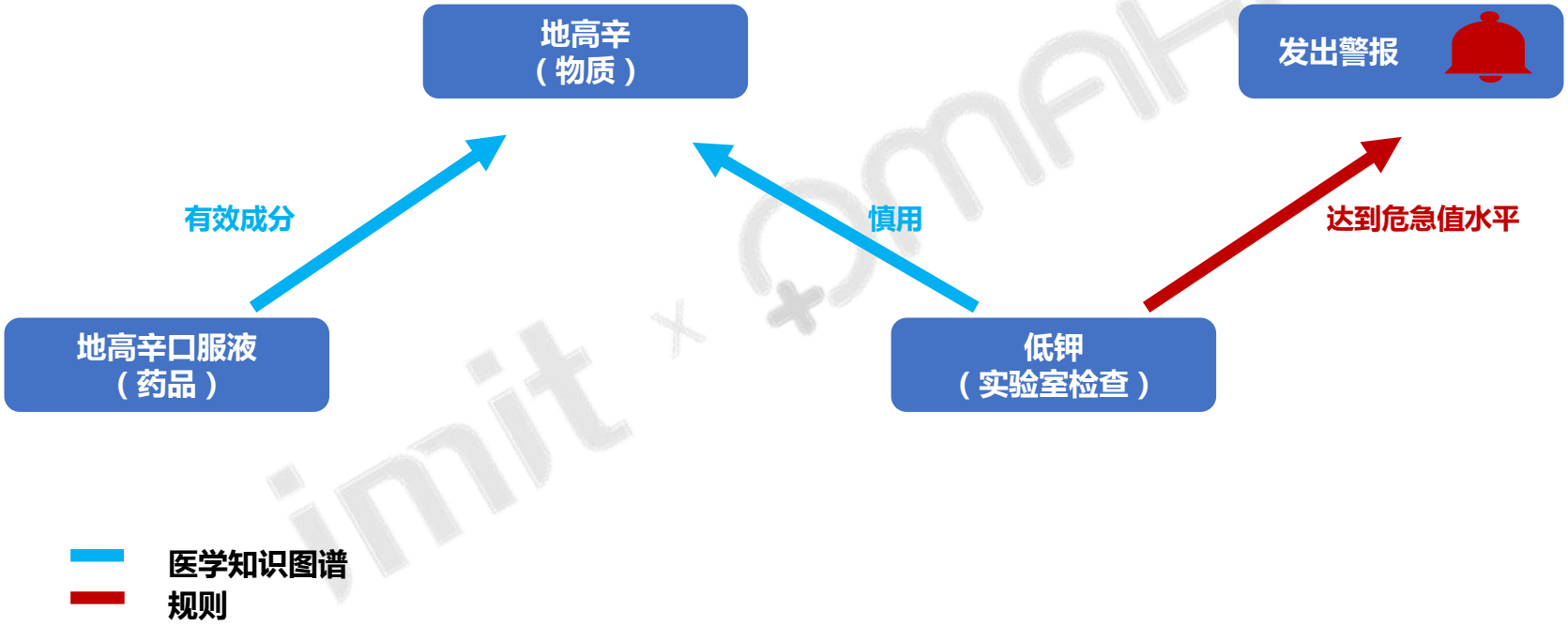
排名	志愿者	单位	总贡献值
1	胡睿瑶	中国人民解放军二六四医院	43736
2	赵晓凯	中国人民解放军二八二医院	31094
3	郑嘉堂		29522
4	唐教清	四川大学华西医院	16180
5	何姜琴	浙江大学医学院附属第四医院	14125
6	王昭予	西安交通大学医学院	13187
7	TJ		12022
8	张智星	山西医科大学第一附属医院	11895
9	XZX		10927
10	林贝		9321

参与协作项目：member.omaha.org.cn/volunteer/reg_v

内容

- 国内外医学知识图谱发展情况
- 医学知识图谱的领域特征和应用需求
- 数研院医学知识图谱构建
  - 模型建立
  - “七巧板” 本体术语集构建
  - “汇知” 图谱构建
- **医学知识图谱应用案例**

智能预警：知识图谱作为底层支撑，辅以更多规则，实现更全面的临床诊疗推理



智能预警：基于知识图谱进行推理，实现实验室危急结果的预警和处方异常预警

- 急诊科患者主诉|胸痛|

- 主治医师开具实验室检验项目医嘱，其中包括|血清钾检测项目|
- 主治医师处方|地高辛口服液|用于缓解胸痛

- 实验室信息系统（LIS）检查结果显示患者|低血钾|

- 系统从LIS中自动捕获患者|低血钾|，判断其属于危急情况
- 系统从EMR中自动捕获医嘱|地高辛口服液|
- 系统利用已有知识库推理得知|地高辛口服液|的活性成分为|地高辛|，出现|低血钾|情况要慎用 |地高辛|

系统发出警报，提示医生：
- 患者处于低钾危急值状态；
- 患者当前处于低钾状况，慎用地高辛口服液

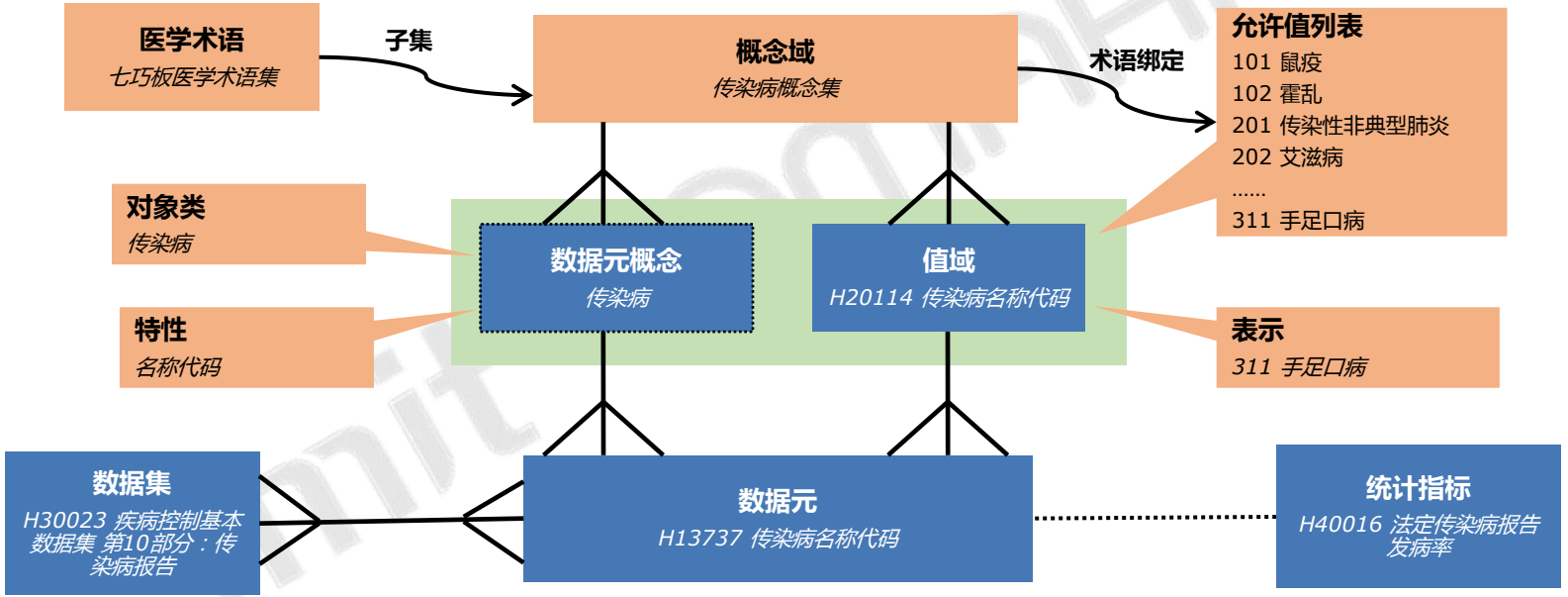
指南推荐：基于医学本体层级关系推理后进行推荐，使推荐结果更丰富

|心境障碍|、 |抑郁症|、 |科塔尔综合征|
被以上概念标识的知识文件都可以被推荐

指南推荐：根据患者信息，推荐相似病历、临床路径、指南等，辅助医生制定治疗计划、规范治疗流程

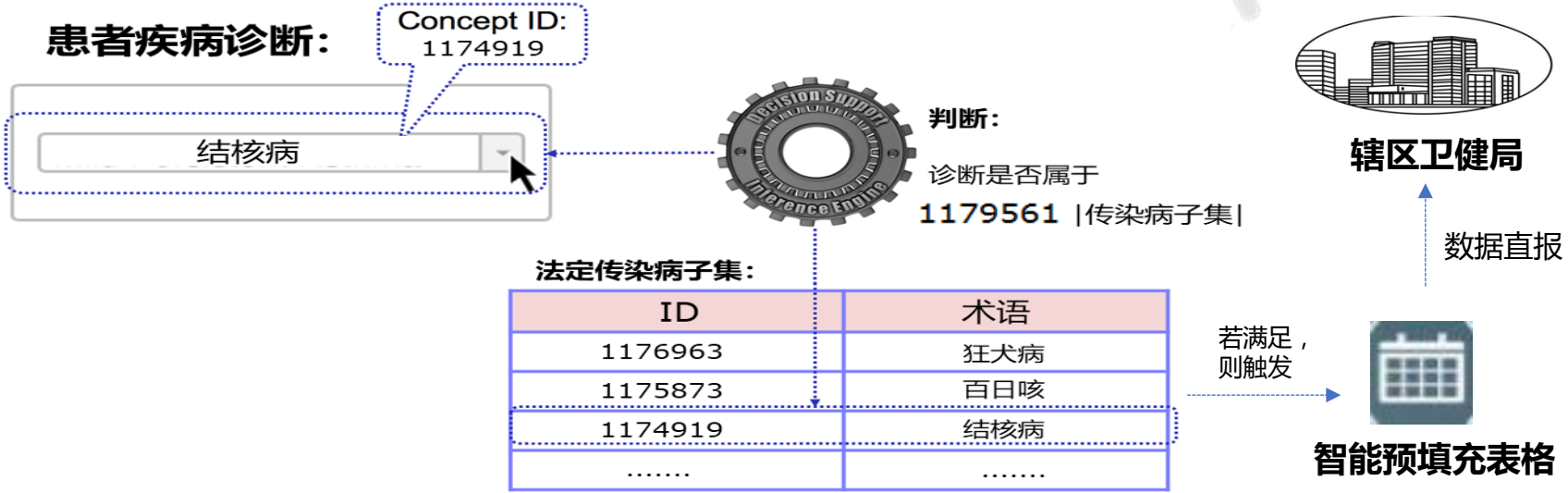
数据直报：将医学知识图谱中的部分内容作为信息模型中的值集，实现医疗数据与医学知识之间的绑定

术语绑定：将医学术语集中的概念分配临床信息模型中的具体数据单元，从而实现医学术语和临床信息模型的联系和赋予某种程度上的语义，这个过程称为术语绑定。



数据直报：信息系统中提前设定相应规则，基于“法定传染病”子集，进行传染病直报判断与提示

IF 患者的疾病诊断= 1179561 |传染病|子集（或甲类传染病子集、乙类传染病子集等）
THEN 提示：对该患者进行传染病直报



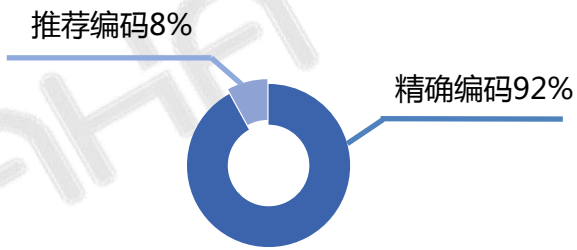
*若医院具备CDSS，可将规则放入CDSS中，并与传染病直报系统连接，第一时间从临床数据中心获取病人信息，能快速、统一、准确的进行传染病报卡智能填充，进行病情汇报

智能编码：通过术语集与ICD编码的映射，使医生关注于临床，提高编码质量、效率，为DRG的实施做好准备

智能编码：通过智能编码引擎，可快速高效开展医疗数据标准化工作，助力医疗数据分析与应用

数据清洗

ID	门诊诊断
1	筛窦炎[慢性]
2	肺上有结节（门慢）
3	多发肋骨骨折（社区）
4	1、肾结石 2、高血压
5	急性阑尾炎和眼睛结膜炎



真实世界的疾病诊断编码准确度可达97%以上，其中精确编码比例92%，推荐编码比例8%。来源术语质量越高，编码质量越高。

标准化解析

数据编码

ID	诊断名称	编码结果（医保版编码）
1	筛窦炎慢性	[J32.200 慢性筛窦炎]
2	肺上有结节	[R91.x00x001 肺部阴影]
3	多发肋骨骨折	[S22.400 肋骨多发骨折]
…	…	…
5	急性阑尾炎和眼睛结膜炎	[K35.800x001 急性阑尾炎][H10.900 结膜炎]

科研分析：在数据治理环节，医学知识图谱可辅助将医疗数据进行标准化、标注等处理，数据得以被深层挖掘

数据治理过程

数据标化

- 反复发作低血糖
- 2型糖尿病糖尿病周围血管病变糖尿病肾病
- 鼻咽癌 .
- '失眠症`慢性心肌缺血·脂肪肝·

标化后

- 19605567 |反复发作低血糖|
- 1147726 |2型糖尿病|；1184569 |糖尿病周围血管病变|；1184467 |糖尿病肾病|
- 111632640 |鼻咽癌|
- 1189440 |失眠症|；1195542 |慢性心肌缺血|；1205310 |脂肪肝|

病历后结构化

观测操作

观测操作 临床所见

【现病史】查体发现左上肢血压未测及且无脉而收入院进一步检查。

临床所见 临床所见

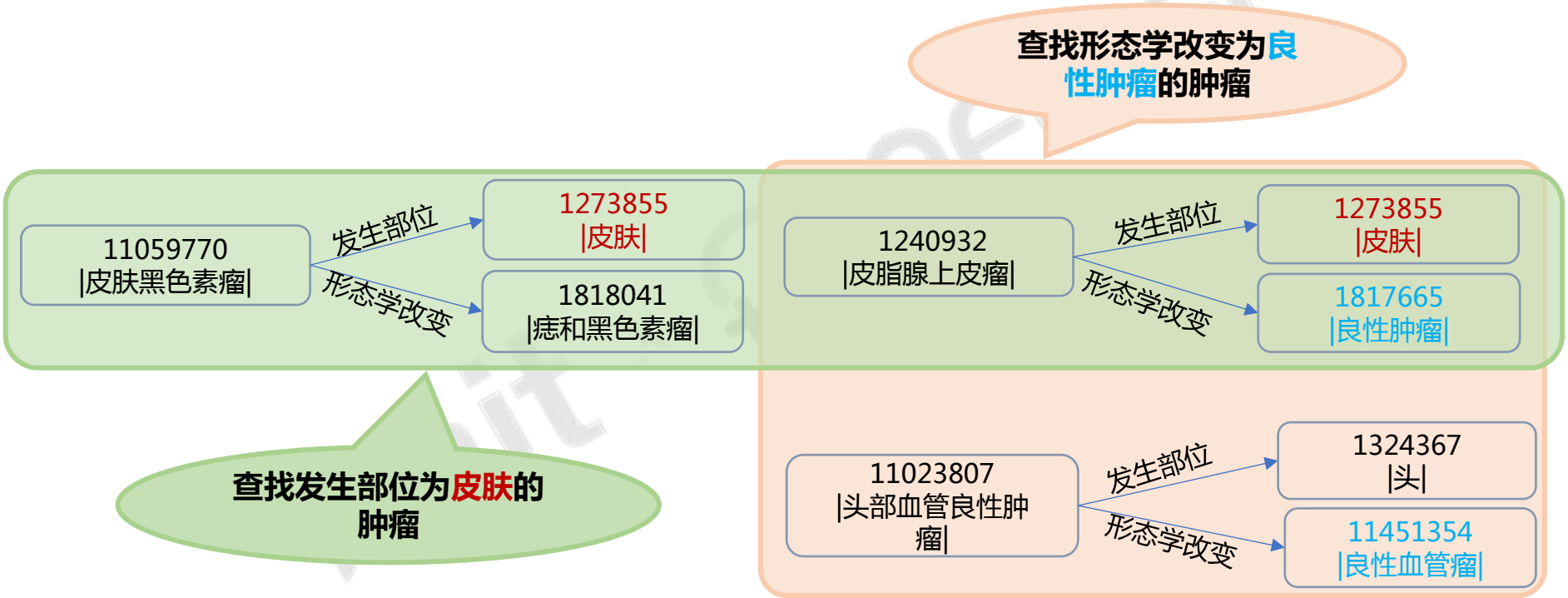
观测操作 临床所见 临床所见 临床所见 疾病

【体格检查】双上肢对称，无手指发凉、怕冷、疼痛、苍白、麻木，无疲乏无力、肌肉萎缩等。

观测操作 观测操作

【辅助检查】自身抗体（-）。

科研分析：在统计分析环节，利用知识图谱关系，可提供医生更多维度以进行数据分析



imit
浙江数字医疗卫生技术研究院
Institute of Medical-care Information Technology

OMAHA
Open Medical and Healthcare Alliance
开放医疗与健康联盟

THANK YOU!

关注我们

加入我们

地 址：浙江省杭州市余杭区良渚文化村设计路8号
微信公众号：china-omaha

联系我们：
OMAHA 市场沟通中心
电话：0571-88983625
邮箱：us@omaha.org.cn