肝病本体知识库构建

1. 肝病知识库构建目的（可以补充调整，此项提出是为了基于应用构建知识库，由肝病中心最后确定）：

支持医学决策支持系统的构建，支持医生个人知识库构建，支持医学教学，支持基于语义的查询搜索等等，肝病领域本体知识库是各种应用的核心。

1. CBR-RBR相结合的肝病知识库构建思路

肝病知识库构建框图如图1所示，所面临问题如下：



图1 肝病临床知识库构建框图

* 1. **肝病领域本体构建**

肝病知识库构建涉及到很多内容，如需要肝病领域本体，目前中医临床本体标准只有参考UMLS模型构建的TCMLS，所以肝病知识库的构建工作将更加艰难，需要自行构建肝病领域本体。

* 1. **知识表达方式选择**
     1. 肝病知识表达（怎样的应用采用怎样的表示方法，所以应用目的必须清晰）：
        1. 案例知识表示法
        2. 产生式规则表示
        3. 语义网表示法
        4. 框架理论表示法
     2. 规则知识表达及案例知识表达相结合（可改进）

目前还没有肝病诊断国标，搜索肝病诊断标准研究相关文献发现同一证诊断规则是有差异的，如肝郁脾虚证的诊断**：**

* 中华中医药学会内科肝病专业委员会天津会议 (1991年12月11～14日) 讨论，特制订辨证标准 (试行)中关于肝郁脾虚证诊断描述如下（下面2个表中内容都是复制了文献的内容）：

|  |
| --- |
| **肝郁脾虚证：** 临床表现为胁肋胀满疼痛，胸闷太息，精神抑郁，性情暴躁，纳食减少，口淡乏味，脘痞腹胀，午后为甚，少气懒言，四肢倦怠，面色萎黄，大便溏泄或食谷不化，每因进食生冷油腻及不易消化的食物而加重，舌质淡有齿痕，苔白，脉沉弦。 　　主证： (1) 胁肋胀痛 (2) 腹胀便溏 　　次证： (1) 抑郁烦闷 (2) 身倦乏力 (3) 舌淡有齿痕 　　辨证要求： 　　1。 具备主证 (1) (2) 者，即属本证。 　　2。 具备主证 (1) 及次证 (2) (3) 两项者，即属本证。 　　3。 具备主证 (2) 及次证 (1) 者，即属本证。 |

* 世界中医药学会联合会肝病专业委员会标准( 编号： SCM － C0003 － 2014)在慢性乙型肝炎( ALT≥2 × ULN) 中医证候诊断标准一文中对肝郁脾虚证诊断描述如下：

|  |
| --- |
| **肝郁脾虚证：**  主要症状： 1) 情志抑郁； 2) 腹胀； 3) 纳呆。  次要症状： 1) 乏力； 2) 烦躁易怒； 3 ) 胁胀； 4) 便溏； 5) 脉弦细。  辨证要求： 具备所有主症者，即属本证； 具备主症任意1 项与次症中任意1 项者，即属本证； 具备主症1) 与次症1) 、4) 两项者，即属本证。 |

从上可知：肝病诊断标准研究中对同一病证诊断过程是不一致的：（1）主症和次症内容不同；（2）诊断标准不同。从文献中也可以发现临床知识具有复杂性、模糊性、不确定性和不完全性，以此有如下知识表达方式选择：

知识表示内容包括：

* 精确、确定和完全的知识——〉规则知识表达方式表达
* 复杂、模糊、不确定和不完全的知识——〉案例知识表达方式表达

1. 肝病知识库构建前期工作准备
   1. **肝病中心需要完成的前期工作**

根据医学文献、诊断标准、临床指南、临床路径等完成如下2项工作：

* + 1. **肝病临床术语规范化**

此项可以参考临床科研一体化系统中已有的数据字典或叙词表，然后根据已发布的TCMLS系统及相关标准进行规范，为领域本体构建做准备。

* + 1. **肝病临床诊断规范化**

此项可以根据相关标准文献研究获得相关内容，再根据目前临床中对各证诊疗效果归纳出规范化的诊断过程，并采用肝郁脾虚证的描述形式进行描述。

* 1. **知识工程相关人员需要完成的前期工作**
     1. **本体生成**

**本体的定义**

本体是一个形式化的、共享的、明确化的、概念化规范，其中， “概念化”是指区别于其他相关概念的抽象模型；“明确”是指所用概念的类型及其约束必须是明确定义的；“形式化”是指本体必须是机器可读的；“共享”是指本体中体现的共同认可的知识， 即反映的是相关领域中公认的概念集。

1。2　本体的建模元语

一般来说，一个本体可由概念、关系、函数、公理和实例5种元素组成。

(1)概念：指任何事务， 如任务、功能、行为、策略、推理过程等。从语义上讲， 它表示的是对象的集合。

(2)关系：概念之间的关联。形式化的定义为R：C1 ×C2 ……×Cn表示概念C1 ， C2 ，……， Cn之间存在n元关系R。如子类关系subclass of。

(3)函数：函数是一种特殊的关系。形式化的定义为F：C1 ×C2 ……×Cn-1× Cn表示R的前n-1个元素可以唯一决定第n个元素。

(4)公理：表示永真式。

(5)实例：表示元素。是指属于某概念的基本元素。

另外， 从语义上讲， 基本的关系有4种：part of(部分与整体关系)、kind of(继承关系)、instance of(实例与概念间关系)、attribute of(属性关系)。

对有明确概念的实体，根据实体之间的关系产生领域本体。如上例中肝郁脾虚证的主症和次症与证的关系可以是atribute\_of 关系等等。

* + 1. **知识库中知识表达技术研究和应用**

1. **实体的识别与自然语言处理技术**

以上工作的进行需要使用到自然语言处理技术，对于中文来件，需要完成分词，根据语义获得临床实体，这样才可以将文本形式的电子病历转化为知识库中的案例知识。

中文分词在中医临床病例上面的应用需要对临床病历书写格式进行分析，对中医临床术语进行词性标注，这种做法是为了更好的抽取临床术语，否则应用中医术语词典进行抽取的话，对新词的识别（未登录词）是非常差的，不利于知识库的动态更新。

1. **知识表达与知识库存储方式与技术**

不同的应用目的采用的技术是不一样的，所以知识表达方式和存储方式是需要仔细研究的。如：

对于临床决策支持系统的应用，则需要采用数据挖掘的方法建立临床病证知识规则，则知识库的存储最好是关系数据库。

应用目的是为了搜索和可视化，则ＸＭＬ的表示方法比较好！