



བདུད་རྩི་སྒྲིང་པོ་ཡན་ལག་བརྒྱད་པ་གསང་བ་མན་  
ངག་གི་རྒྱུད་ཞེས་བྱ་བ་བཞུགས་སོ།།

《甘露精要八支秘诀续》--读书报告

西北民族大学  
冷本扎西



随着深度学习不可解释的局限性，知识资源的构建对于NLP研究工作显得尤为重要，因此，知识图谱构建工作获得格外青睐。如何将知识进行数据化、结构化是目前藏医药领域本体构建过程中的瓶颈之一。

- 《全国藏医药古籍名录》统计，藏医药图书总数达1057本，占《全国中医图书联合目录》总数之11.5%。已完成数字化的图书仅492本。
- 藏医药学文献因其文献篇幅长、体系健全、医法奇特、疗效绝佳而著名于世。但，就其知识表示形式而言，多采用诗体书写，形式化程度较低。
- 藏文本身形态特征较弱，词间无空格，格助词使用繁多，语法位置不甚稳定。
- 目前，藏语文本自动处理技术较弱，处理工具应用到具体小众领域时，其准确性直线下降。



知识本体构建思路:

- 1) 《四部医典》数字化工作
- 2) 完成多层次文本标注工作
- 3) 整理《四部医典》知识分类
- 4) 《四部医典》语言特征解析
- 5) 抽取实体及实体关系...

本次报告内容:

- (1. 《四部医典》及其版本介绍
- (2. 《四部医典》文献考源
- (3. 《四部医典》内容概要
- (4. 常用本体知识表示方法
- (5. 《四部医典》树喻及其知识分类...

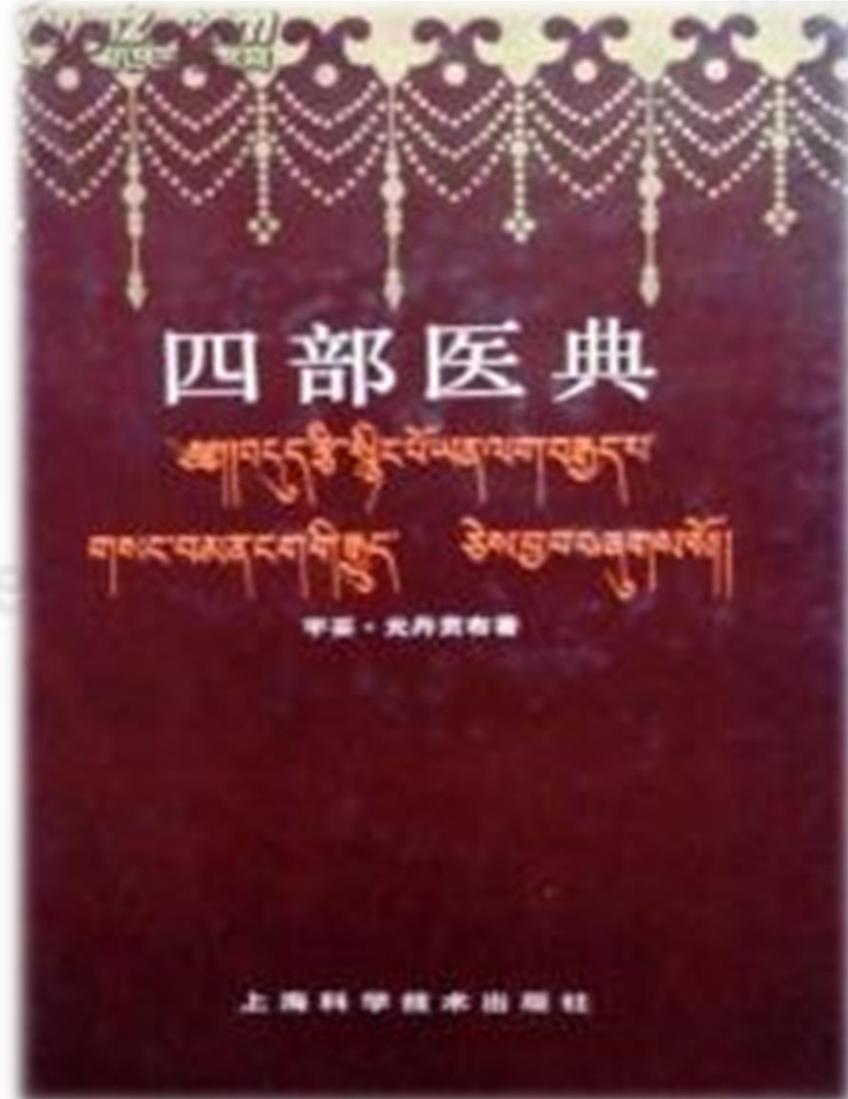




# 《四部医典》及其版本介绍



biomechanics



《四部医典》是藏医学的主要医典，全名为《甘露精要八支秘诀续》(བདུད་རྩི་སྒྲིབ་པོ་ཡན་ལག་བརྒྱད་པ་གསང་བ་མན་ངག་གི་རྒྱུད་ཞེས་བྱ་བ་བརྒྱུགས་སོ།།)，简称《四部医典》(གསོ་རིག་དཔལ་ལྷན་རྒྱུད་བཞི།)。《四部医典》是一部集藏医药医疗实践和理论精华于一体的藏医药学术权威工具书，被誉为藏医药百科全书，是藏医药学中最系统、最完整、最根本的一套理论体系。

甘露	精要	八支	秘	诀续	
བདུད་རྩི་	སྒྲིབ་པོ་	ཡན་ལག་བརྒྱད་པ་	གསང་བ་	མན་ངག་གི་རྒྱུད་	ཞེས་བྱ་བ་བརྒྱུགས་སོ།།

“四部医典”（续）藏语简称中，“rgyud”一词在藏语中有传统、相续、一脉传承之义。“rgyud”本是一个佛教用词，即密宗、密乘，又有译成续、本续者，与梵文坦特罗（Tantra）相对应，指密乘佛教及其有关的经典著作。





# 《四部医典》及其版本介绍

- 《四部医典》共出现宇妥宁玛元丹贡布版，却拿版，总噶版，扎塘版、达旦版、药王山版、甘丹平措林版等。然而，多个版本今佚。
- 藏医院现珍藏的《四部医典》有4种木刻版，分别为扎塘版、达旦版、药王山版、甘丹平措林版及一部金汁手写版。其中，最早的木刻版为“扎塘版”（1546年），字迹清晰，保存完好，没有破损，中间附有彩色手绘插图，以上4种木刻版《四部医典》是全世界目前保存最完好的版本。金汁手写版是存世的唯一一部金汁手抄医学文献。

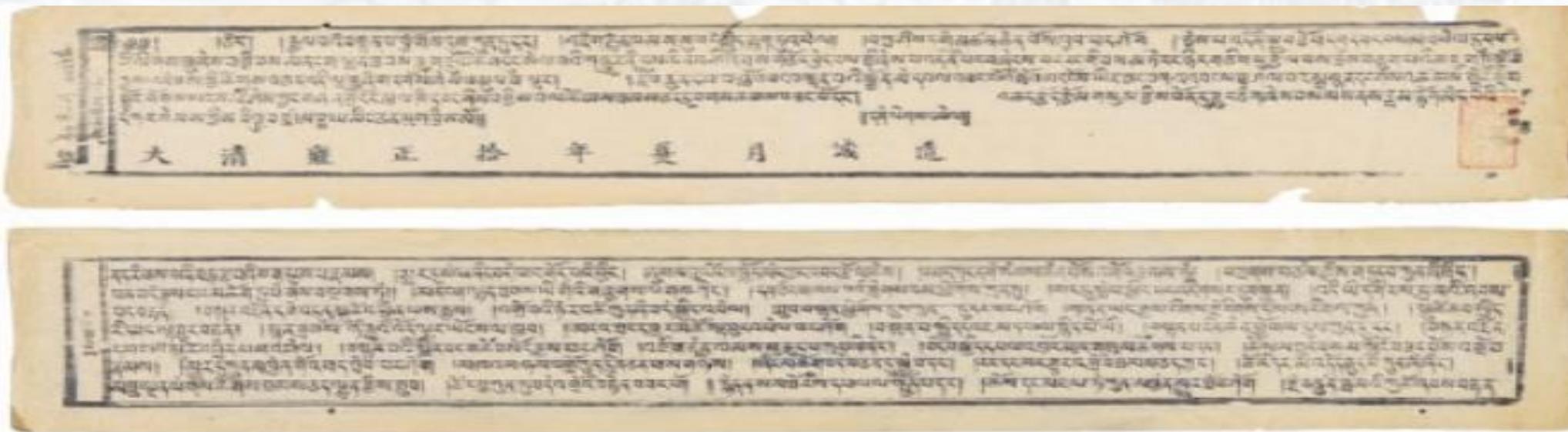


图1 四部医典·诀窍部 藏文 宇妥·云丹贡布撰 十四代字 妥·云丹贡布修订 清雍正十年(1732)





# 《四部医典》及其版本介绍



《四部医典》译本有：

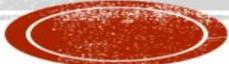
1. 《四部医典》（汉译本），李永年译，人民卫生出版社出版。
2. 《四部医典》（汉译本），马世林等译，上海科学技术出版社出版。
3. 《四部医典》（蒙译本），敏珠尔道尔吉译，木刻版。
4. 《四部医典》（俄节译本），巴得玛也夫（P·A·Badmaev）。
5. 《四部医典》（俄节译本）波茨涅耶夫（A·Pozdneev）。
6. 《四部医典》（日节译本），芳村修基。
7. 《四部医典》（英节译本），热琼仁波切·甲白衮桑（རི་ཚུང་རིན་པོ་ཆེ་འཇམ་དཔལ་ཀུན་བཟང་།）。
8. 《四部医典》（英节译本），艾默瑞克（REEmmerick）。
9. 《四部医典》（英节译本），伊丽莎白·芬克（ElisabethFinckh）。
10. 《四部医典》（德节译本），伊丽莎白·芬克（ElisabethFinckh）。

另有32本全释本和节释本及24本作者不详的注释本。





# 《四部医典》来源、作者、成书年代



首先，就关于《四部医典》是属于“经”（佛说经，甘珠尔）还是属于“论”（后人编著的论典，丹珠尔）问题，有许多不同的观点。多因宗教流派而异。

- ① 《四部医典》被认为是属于经者以佛祖释迦牟尼在印度药林居四年时所言有关医药学方面的谈论作为所引之依据。经可分口传、随许、加持等三种。尤其在贡曼贡曲彭达所著《医学如意宝》中说道：“医药传经显见学，水晶之镜属口传。三怙主著属随许。医诊四续属加持”，认为《四部医典》属经类之加持类。

《四部医典》属于佛经的论据具体体现在书名和偈语体中包含的佛教色彩、“极乐净土药王城”等词汇、文本叙述方式以佛祖与道徒的问答形式、治疗咒语、古印度“八支”理论）。



# 《四部医典》来源、作者、成书年代

《四部医典》源自“佛经”的说法恐难成立。该论著在名称、叙述方式、咒语、八支理论等与“论”中典籍也多有耦合之处，不能依此断言。

纵然如此，《四部医典》中仍有不少内容是移植或借用佛学文化的内容的，所以与“古印度佛教”的关系非常密切。如藏文化“大小十明”也是沿用了古印度的学问分类体系。“七岁之后，渐授五明大论：一曰声明，释诂训字，诠目疏别；二工巧明，伎术机关，阴阳历数；三医方明，禁咒闲邪，药石针艾；四曰因明，考定正邪，研核真伪；五曰内明，究畅五乘因果妙理。

玄奘，《大唐西域记》卷二，上海出版社，1997年，第36页。

# 《四部医典》来源、作者、成书年代

另外，多数学者认为《四部医典》属于“论”，如苏卡·洛追杰布（བུར་མཁར་སྒོ་གྲོལ་རྒྱལ་པོ།）在《四部医典之经论分析·驱暗明灯》（རྒྱུད་བཞིའི་བཀའ་དང་བསྟན་བཅོས་རྣམ་པར་དབྱེད་པ་ལུག་སེལ་སྒྲོན་མེ་ལེས་བྱ་བ་བཞུགས་སོ།།）中说道：“那么《四部医典》被认为是佛说经，这样是否正确呢？现予以阐述。倘若不认为是经类，西藏的上下贤愚人士难以置信。故对那些只重表面词句而不视其义的人，不得不说成是经类。其实从饮茶、瓷碗、验便等占卜理论和“杜鹃鸣”等词语看很明显是西藏本地的典籍”。

- ② 印度医学：《宁妥·云丹贡布》记载，《四部医典》为译师白若杂纳翻译的作品，云丹贡布是在后期对它进行了补充和注释。目前见到的几种版本，已不是原来版本了。法国作者P·Huard也认为《四部医典》为生命吠陀（Ayurveda，亦译长寿吠陀、阿输吠陀）的回忆录（6世纪）。

# 《四部医典》来源、作者、成书年代

- ③ 苯教医学：雍仲苯教的第一本“医典”记载敦巴辛饶之子杰布楚西后来成为敦巴辛饶的医学继承人，杰布楚西将他父亲的用象雄文所撰写的医学论著（约公元前1800年）编纂命名为《苯医四续》。这部书就是现在我们使用的《四部医典》中的宗源书之一。如今与《གསལ་རིག་རྒྱུད་བཞི།》对立，称《གསལ་རིག་འབྲུམ་བཞི།》。内容基本一致，文体不同。
- ④ 汉地医学：松赞干布迎娶文成公主入藏时。当时由汉地的摩诃弟瓦（མ་ཏཱ་དེ་བཤ།）和尚及西藏的大译师达磨郭卡把文成公主携带的《医法大论》译成藏文并予以厘定，这是藏文译成的第一部汉医医典，称其为《医学大典》（མཁན་དཔུང་ཚེན་མོ།）可惜此书早已散失。藏王与金城公主联姻后，由将其所带去的医药典籍经藏汉医僧合作，译成《月王药诊》（公元8世纪中期，མཁན་དཔུང་ཟླ་བའི་རྒྱལ་པོ།），其精华部分均收辑于《四部医典》之中。

# 《四部医典》来源、作者、成书年代

- ⑤ 多方医学精华：普顿·唐杰钦巴、达仓译师等认为。该典籍是宇妥·云丹贡布吸收了印度的《八支要论》，由汉地传入并藏译的《索玛惹扎》（སོ་མ་རྩ་ཇ།）、象雄的《轮王小册》等典籍的精华而编撰的。
- ⑥ 宇妥·云丹贡布：宇妥萨玛之弟子松顿·益希所著《史说·善业之钩（德格版）》第7页中认为《四部医典》是由宇妥萨玛编撰的。《四部医典》在8世纪由著名藏医学家宇妥·宁玛云丹贡布所著。相传到了十一世纪时，云登贡布的后裔玉妥·萨玛吸收《月王药珍》（公元8世纪中期，ལྷན་དབྱེད་ལྷ་བའི་རྒྱལ་པོ།）的精华，对《四部医典》的内容作了进一步充实。五世达赖喇嘛时期，著名学者第司·桑杰嘉措对《四部医典》进行了校对和修订，后期西藏地方政府又多次对其进行修订。最终形成现今的《四部医典》。

- 《四部医典》的内容十分丰富，《四部医典》吸收了《医学大全（藏名《སྐབ་ལྷོ་ཆེན་མོ།》）》、《无畏的武器》（藏名《མི་འཇིགས་མཚོན་ཆ།》）》、《月王药诊》（藏名《སྐབ་དབྱེད་ལྷ་བའི་རྒྱལ་པོ།》）等著作的精髓，并参考了中医学、天竺和大食医学的理论，但是并非搬抄任何一部医学理论而作。《四部医典》中的多项内容在古印度医学体系中并不存在，如脉诊、尿诊方法在“古印度”医学中并未涉及（达仓译师）。据考察，“八支理论”首次出现在古印度吠陀时期（前2000-1000年）长寿吠陀（Ayurveda）的医学理论中，唐代称“八医”，《四部医典》虽然采用了古印度“八支”的名称为书名，然而其实际内容和结构与八支理论并不协调，基本不再具有“八支”的轮廓。

故此认为，《四部医典》是从古代经验医学的经验中著成，是一本由宇妥·云丹贡布编撰的多方精华的藏医学集成体。

婆罗贺摩（又译大梵天Brahma）创造的吠陀经中将八支分别为：外科（Salya），头颈部疾病治疗（Salakya），毒物治疗（Agada），妇科及小儿病治疗（KumaraBhrtya），内科病治疗（Kayacikitsa），魔鬼病治疗（BhutaVidya），强精科（Vajikarana），返老还童科（Rasayana）。





# 《四部医典》内容

《四部医典》首次系统阐明维持人体正常的生理功能，主要围绕：

- 三大因素：即“龙”，“赤巴”，“培根”。
- 七大物质基础：即饮食精微、血、肉、脂肪、骨、骨髓、精。
- 三种排泄物：即大、小便、体汗。
- “三诊理论”：望、闻、切。
- “四友原则”，包括饮食、起居、内服药物、外治等四个方面，说明治病除服用内服药物和外治法外，饮食起居也很重要。

《四部医典》共分四部，即《根本医典》、《论述医典》、《秘诀医典》、《后续医典》。

- 第一部《根本医典》，(རྩ་བའི་རྒྱུད།) 共六章，概括性地介绍了人体生理、病理、诊断及治疗的基本知识。本部用比喻阐述法高度概括了其他三部的内容，是四部医典的总纲性部分。





# 《四部医典》内容



- 第二部《论述医典》（བཤད་པའི་རྒྱུད།）共三十一章，详细介绍了人体生理、解剖，疾病发生的原因及规律、卫生保健知识、药物性能、诊断方法和治疗原则等内容。本部是全面阐述藏医学理论的理论性部分。
- 第三部《秘诀医典》（མན་ངག་གི་རྒྱུད།）共九十二章，阐述各科疾病的临床诊断和治疗方法、老年养生等内容，是临床医学部分。
- 第四部《后续医典》（བྱི་མའི་རྒྱུད།），亦称《诊治医典》，共二十五章，主要论述脉诊和尿诊，各种方剂和药物的配伍、功能、药材的炮制、开方途径及外治法（放血、艾灸、火灸、外敷、拔罐）等内容。本部是阐述诊法、用药、施治的诊治部分。

《四部医典》最后另附有《四部医典结束要义》一章和《诸续概说》一章。全书共有一百五十八章。







# 本体知识表示方法

人工智能是以知识为基础的。知识包括事实、规则和控制策略等不同类型。如何让计算机能够理解人类用自然语言表达的知识，以象人类思维过程一样在机器上进行推理、判断、学习、决策等各种信息加工，并将加工后的结果以人类能够理解的形式呈现给用户是人工智能要解决的首要问题。

在本体构建中，知识的表示形式和手段决定了知识推理的形式和难度。知识表示，可以理解为知识图谱逻辑层面的表示，即数据模型，是将知识符号化和形式化的过程，从人的角度对本体知识进行描述。为了建立完善的知识系统，以便更好的应用人类的知识成果。科学家们提出了许多种知识表示方法，为此科学家们相继提出了逻辑表示法、语义网络、框架表示法、脚本表示法、基于语义网的知识描述体系等等多种知识表示方法。

文本将知识表示方法简单分为经典知识表示理论、语义网中的知识表示方法、本体知识的数值化表示方法三种。限于篇幅，本文重点介绍语义网知识表示方法：





# 本体知识表示方法



## 1. 逻辑表示法

- 命题逻辑
- 一阶谓词逻辑
- 高阶逻辑

## 2. 语义网络表示法

## 3. 知识框架表示法

## 4. 语义网的知识表示方法

- XML (eXtensible Markup Language, 可扩展标记语言)
- RDF: 全称资源描述框架 (Resource Description Framework)
- RDFs: RDF框架 (RDFschema)
- OWL: 全称网络本体语言 (Web Ontology Language)





# 常用知识表示方法



## <逻辑表示法>

### (1) . 命题逻辑

命题逻辑具有最简单的语法，它定义了具有真假值的原子命题，并可以通过"与"、"或"、"非"、"蕴含"、"当且仅当"等逻辑连接符将多个原子命题组合成复合命题，而推理过程则根据逻辑连接词的真值表进行自动推导。

### (2) . 一阶谓词逻辑

谓词逻辑是一种形式语言，用谓词逻辑可以表示事物的状态、属性、概念等事实性的知识。

如：COMPUTERScience(zhaxi)  $\wedge$   $\sim$ LIKE(zhaxi, programming)。

首先定义谓词，第二步将个体代入谓词中，第三步：根据语义，用逻辑联接词将它们联接起来，就得到表示上述知识的谓词公式。





# 本体知识表示方法



## (3). 高阶逻辑

高阶逻辑亦称“广义谓词逻辑”、“高阶谓词逻辑”。谓词逻辑有一阶逻辑和高阶逻辑之分。在一阶逻辑中，量词只能用于个体变元，取消这一限制条件，允许量词也可用于命题变元和谓词变元，由此构造起来的谓词逻辑就是高阶逻辑。二阶逻辑可以量化集合，三阶逻辑可以量化集合的集合，依次类推。

命题逻辑和（一阶）谓词逻辑是人工智能领域使用最早的和最广泛的知识表示方法之一。但由于这种产生式表示方法难以表示过程性知识和不确定知识，而且当表示知识中的属性、谓词和命题数量增大时，其推理过程因为符号的组合爆炸问题，计算复杂性呈指数级增长态势。因此，基于谓词逻辑的推理过程比较耗时，工作效率较低。





# 语义网络



语义网络是一个通过语义关系连接的概念网络，它将知识表示为互相连接的点和边的模式，其中，节点表示实体、事件、值等，边表示对象之间的语义关系。也就是说，语义网络其实是一种有向图表示的知识系统，节点代表的是概念，而边则表示这些概念之间的语义关系。

语义网络与一阶谓词具有相同的表达能力，不同的是，它用最简单的一种统一形式描述所有知识，非常有利于计算机的存储和检索。

语义网络的缺点：

- 它仅用节点及其关系描述知识，推理过程不像谓词逻辑表示方法那样明了，需要针对不同关系做不同处理，推理方法还不完善
- 语义网络中，并没有对节点和边进行标准的定义，也就是完全用户自定义的。没有标准，一方面，会增加知识分享的难度；另一方面，知识描述和知识实例难以分隔。





# 知识框架表示法



框架是一种描述固定情况的数据结构，一般可以把框架看成是一个节点和关系组成的网络。框架的最高层次是固定的，并且它描述对于假定情况总是正确的事物，在框架的较低层次上有许多终端--被称为槽 (Slots)。在槽中填入具体值，就可以得到一个描述具体事务的框架，每一个槽都可以有一些附加说明--被称为侧面 (Facet)，其作用是指出槽的取值范围和求值方法等。

## 框架的基本形式：

<框架名>

槽名1:

侧面名11: 侧面值11

侧面名12: 侧面值12

.....

侧面名1n: 侧面值n

槽名k:

.....

## Disaster Description Frame System

框架名: <灾害事件>

地点: 字符串

日期: 单位(年, 月, 日)

死亡人数: 单位(人)

财产损失: 单位(亿美元)

框架名: <地震>

ISA: <灾害事件>

强度: 单位<里氏级>

断层: 字符串

1976年3月18日，一次强度为里氏8.5级的地震发生在下斯洛文尼亚地区，造成25人死亡和5亿美元的损失，多年来靠近萨蒂亚豪金斯断层的地区一直是一个地震多发区。

框架名: <地震1>

地点: Low Slabovia

日期: 1976,3,18

死亡人数: 25

财产损失: 5

强度: 8.5

断层: Sadie Hawkins





# 知识框架表示法



框架知识表示方法其最突出的特点是善于表示结构性知识，能够把知识的内部结构关系以及知识之间的特殊关系表示出来，并把与某个实体或实体集的相关特性都集中在一起。

该方法的优点体现在：

- 框架系统的数据结构和问题求解过程与人类的思维和问题求解过程相似；
- 框架结构表达能力强，层次结构丰富，提供了有效的组织知识的手段，只要对其中某些细节作进一步描述，就可以将其扩充为另外一些框架；
- 可以利用过去获得的知识对未来的情况进行预测，而实际上这种预测非常接近人的知识规律，因此可以通过框架来认识某一类事物，也可以通过一系列实例来修正框架对某些事物的不完整描述（填充空的框架，修改默认值）。

其不足点主要体现在：

缺乏形式理论，适应能力不强，子框架间的数据结构一致性要求较高、很难实现知识体系的自动化构建、统一框架体系的设计难度太大等不足。



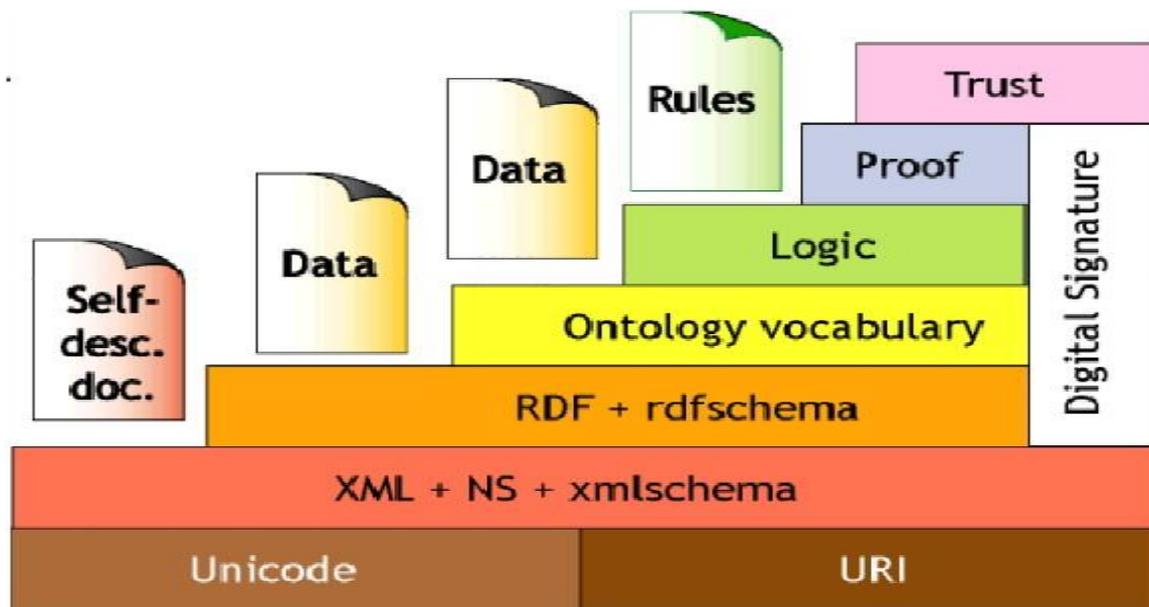


# 语义网表示法方法



与语义网络不同，语义网的概念来源于万维网，其最初的目的就是为了对万维网的功能进行扩展以提高其智能程度，通常被称为 Web 3.0。

## 语义网的技术架构



第七层：信任层，提供信任机制，保证信息交互的安全可靠

第六层：论证层，提供认证机制，验证信息的有效性和可靠性

第五层：逻辑层，提供推理规则的描述手段，推导新知识

第四层：本体层，提供共享的语义概念模型，描述资源间的语义关系，建立计算机可理解的资源概念的语义空间

第三层：数据层，提供资源的通用描述框架和语义描述机制，描述资源及类型

第二层：语法层，使用标准格式语言标识数据的内容和结构

第一层：基础层，对资源进行统一编码和唯一标识，解决资源标准格式和资源定位问题





# 语义网表示法方法



XML是最早的语义网表示语言，是从网页标签式语言向语义表达语言的一次飞跃。XML以文档为单位进行表示，不能显示地定义标签的语义约束。

XML文档的结构用其扩展版本XML Schema定义，指出XML文档元素的描述形式。

XML的内容通过元素来记录，元素带有标签，它可以有嵌套结构，嵌套深度不受限制。XML源于HTML，相比于HTML，XML可扩展性更强，结构性更强。右侧为一段藏文大藏经中《戒律》的基本信息描述：

<文献位置>:པོད་རྟེན་གསལ། འདུལ་བ། ། ཅ ། (5)</文献位置>  
 <页码>: ། ཅ ། 5-1-1a1~5-1-21b4</页码>  
 <藏文名>:སོ་སོ་ཐར་པའི་མདོ།</藏文名>  
 <藏文名拉丁>::so sor thar pa'i mdo</藏文名拉丁>  
 <别称>:དག་སྟོང་པའི་སོ་མཐར་གྱི་མདོ།</别称>  
 <梵文名>:प्रतिमोक्षसूत्रा</梵文名>  
 <梵文名拉丁>:prātimokṣa-sūtra</梵文名拉丁>  
 <中文名>:《根本說一切有部戒經》唐 義淨譯</中文名>  
 <归敬颂> :ཐམས་ཅད་མཁུན་པ་ལ་ཕུག་འཚལ་ལོ།</归敬颂>  
 <作者>:སངས་རྒྱལ་ལུག་ཐུབ་པ།</作者>  
 <译师>འཕགས་པ་གཞི་ཐམས་ཅད་ཡོད་པར་སྐྱབ་པའི་འདུལ་བ་འཛིན་པ་སྟོབ་  
 དཔོན་ཇི་ན་མི་ཏུ་དང། ལུ་ཆེན་གྱི་ལོ་རྒྱུ་བ་བན་དེ་ཅོག་རོ་གྱུའི་རྒྱལ་མཚན་གྱིས་  
 བསྐྱར་ཅིང་ལུས་ཏེ་གཏན་ལ་ཕབ་པ།།</译师>





# 语义网表示法方法



资源描述框架 (RDF) 也是W3C提出的一种语义网描述框架，它假定任何复杂的语义都可以通过若干个三元组的组合来表示，并定义这种三元组的形式为“对象-属性-值”或“主语-谓语-宾语”。其实该方法与语义网络并没有本质的差别，差别在于，RDF规范化了所有的属性和概念，避免了语义网络不便于分享和难以区隔知识描述和知识实例的缺点。



但是RDF本身有其问题：Rdf只能表示实体与实体之间的关系、属性。但是没法定义实体的概念、概念具有的属性。但是不论是在智能的概念上，还是在现实的应用当中，概念都是相当重要的，也是知识本体本身十分强调的。





# 语义网表示法方法



为了解决RDF表示法无法表示概念等缺陷，W3C又推出了RDF schema (RDFS)。RDFSchema定义了概念，概念的属性，属性值，概念与概念之间的层级关系以及约束关系。

RDFS中有两个非常重要的词汇就是domain和range。

- rdfs:domain 用来表示属性的域
- rdfs:range 用来表示属性的取值类型
- rdfs:Class: 所有RDF类型所属的类型
- rdfs:subClassOf: 用来描述一个类的父类
- rdfs:Resource: RDF描述的所有内容都称为资源，并且都是类rdfs:Resource的实例。





# 语义网表示法方法



为了解决RDF表示法无法表示概念等缺陷，W3C又推出了RDF schema (RDFS)。RDFS定义了概念，概念的属性，属性值，概念与概念之间的层级关系以及约束关系。

RDFS中有两个非常重要的词汇就是domain和range。

- rdfs:domain 用来表示属性的域
- rdfs:range 用来表示属性的取值类型
- rdfs:Class: 所有RDF类型所属的类型
- rdfs:subClassOf: 用来描述一个类的父类
- rdfs:Resource: RDF描述的所有内容都称为资源，并且都是类rdfs:Resource的实例。





# 语义网表示法方法



使用RDFs方法的好处在于，当一个类型被定义以后，其他人可以很容易地向其中添加额外的属性，因为添加属性可以在不需要重新修改这些类的原始描述的情况下完成。这正是以属性为中心来构建类型系统的优势所在，允许任何人扩展现有资源的描述。

比如我们已经有一条定义Person这个类的描述 `<Person, rdf:type, rdfs:Class>`，如果向这个Person类添加另一条属性时候并不需要改动原来Person类的定义。

```
<hometown,rdf:type,rdf:Property>
```

```
<hometown,rdfs:domain,Person>
```

```
<hometown,rdfs:range,Place>
```





# 语义网表示法方法



RDFs 的表达能力比较弱，于是又提出网络本体语言 OWL (Web Ontology Language)，为网络本体的构建提供了丰富的建模原语

。OWL 是对 RDFs 的一个扩展，添加了额外的预定义词汇。使其具备更强的推理能力。OWL 是声明性的 (declarative)，也就是说，它用逻辑的方式描述一个事务的状态。OWL 除了拥有像 RDFs 一样灵活的数据建模能力之外，还提供了一套可以帮助计算机进行自动推理的功能强大的词汇。

## Property Declarative Lexical:

- owl:TransitiveProperty. 声明属性的传递性质
- owl:inverseOf. 声明两种关系之间是互反关系
- owl:SymmetricProperty. 声明两种关系之间是对称关系
- owl:FunctionalProperty. 声明属性具有函数性 (唯一性)





# 语义网表示法方法



类型关系声明词汇 (OWL中还提供了一些词汇来表示在不同类型做AND, OR 和 NOT操作), 此外, 还提供了:

## Class Declarative Lexical:

- owl: oneof 声明枚举类型
- owl: disjoint with 声明两个类型不相交
- owl: union of 声明某个类型为其他类型的并集运算

**RDFS**Schema 特征: class, **rdfs:subclassOf**, rdf:property, rdfs:subpropertyOf, **rdfs:domain**, **rdfs:range**, individual;

等同性: equivalentClass, equivalentProperty, sameAs, differentFrom, AllDifferent, distinctMembers;

属性特征: objectProperty, DatatypeProperty, **inverseOf**, **transitiveProperty**, SymmetricProperty, **FunctionProperty**, **InverseFunctionalProperty**;

局限限制: Restriction, onProperty, allValuesFrom, someValuesFrom;

基数限制: minCaidinality, max Caidinality, cadinality;

类相交: intersectionOf;

头信息: Ontology, imports;

版本信息: versionInfo, priorVersion, backwardCompatibleWith, incompatibleWith, DeprecatedClass, DeprecatedProperty;

评注性质: rdfs:label, rdfs:comment, rdfs:seeAlso, rdfs:isDefinedBy, AnnotationProperty, OntologyProperty;





# 语义网表示法方法



OWL在RDF的基础上定义了自己独有的语法，主要包括头部和主体两个部分。

**头部 (head) :** OWL描述一个本体 (概念) 时, 会预先定制一系列的命名空间, 并使用命名空间中预定义的标签来形成某个概念的头部, 例如电影明星本体的头部可以这样表示:

```
<owl:Ontology rdf:about="">  
<rdfs:comment>一个概念的例子  
</rdfs:comment>  
<rdfs:label>电影明星</rdfs:label>  
</owl:Ontology>
```

其中<owl:Ontology rdf:about="">表示模块当前描述的概念。

**主体 (body) :** OWL的主体是用来描述概念的类别、实例、属性之间相互关联的部分, 它是OWL的核心。例如上面电影明显的概念可以有如下的组成部分:

```
<owl:Ontology rdf:ID="电影明星">  
<rdfs:Name rdf:resource="张国荣">  
<rdfs:Birthplace rdf:resource="中国香港">  
</owl:ObjectProperty>
```





# 语义网表示法方法



本体知识的表示方法，总体来说，就是以本体为核心，以RDF的三元组模式为基础框架，但更多的体现实体、类别、属性、关系等多颗粒度、多层次的语义关系。

然而，经典的三元组模型不能解决复杂关系的知识表示：（1）多元关系，比如典型的事件描述场景（2）时空知识表示，需要五元组来表示。例如：公元前407年柏拉图在雅典跟随苏格拉底学习。为了表达这个，把三元组拓展成5元组<亚里士多德，导师，柏拉图>。被拓展成<亚里士多德，导师，柏拉图，（北纬38，东经23），公元前407年>的五元组。（3）多模态的表示等。

领域知识图谱中的知识表示有知识定义（知识体系）与知识实例两个层面。知识实例是本体的一个一个实例，对应的是真实的数据存储层。知识定义（知识体系）描述了本体以及本体之间的关系，是上层建筑。由此可知，《四部医典》中的知识结构的整理对于藏医药领域的本体构建工作极其重要。





# 语义网表示法方法



在通常情况下，一个知识本体主要涵盖以下几个方面的内容：

- 事物：客观世界中的实体或对象
- 概念：具有相似本体特征的一类事物，也称类型
- 属性：事物或概念具有的特征和特性等
- 关系：概念与实体之间的关联方式
- 函数：事物或概念之间进行转化的形式表达，如：国籍（莫泊桑） = "法国"中的国籍（X）则为函数
- 约束：某项断言成立的限制条件的形式化描述
- 规则：依据某项断言得到逻辑推论的因果关系知识的形式化描述
- 公理：永远为真的断言

目前大部分知识图谱主要是对前四部分内容（即事物、概念、属性和关系）进行建模，只有很少的知识图谱建模了简单的规则结构，这也反映了不同层次知识在表示上的复杂程度是不同的。而对于藏医药领域本体的知识构建而言，约束、规则、公理等均不可少。





# 《四部医典》树喻图



17世纪末，五世达赖喇嘛时期第司·桑杰嘉措根据《四部医典》并参考其他医学相关的著作撰写了《四部医典·蓝琉璃》一书。从此，《蓝琉璃》便成为了《四部医典》的标准注解本。以《四部医典·蓝琉璃》为蓝本，第司·桑杰嘉措召集了全藏区权威的藏医药专家及顶级的藏族绘画艺人，在整合前人绘制的图谱基础上，结合了其他医学典籍，补充数幅穴位图，最终绘制成80幅成套曼唐（医药彩色树喻唐卡）而厘定为权威系列挂图，成为藏医药学理论和实践发展史上重要的里程碑。

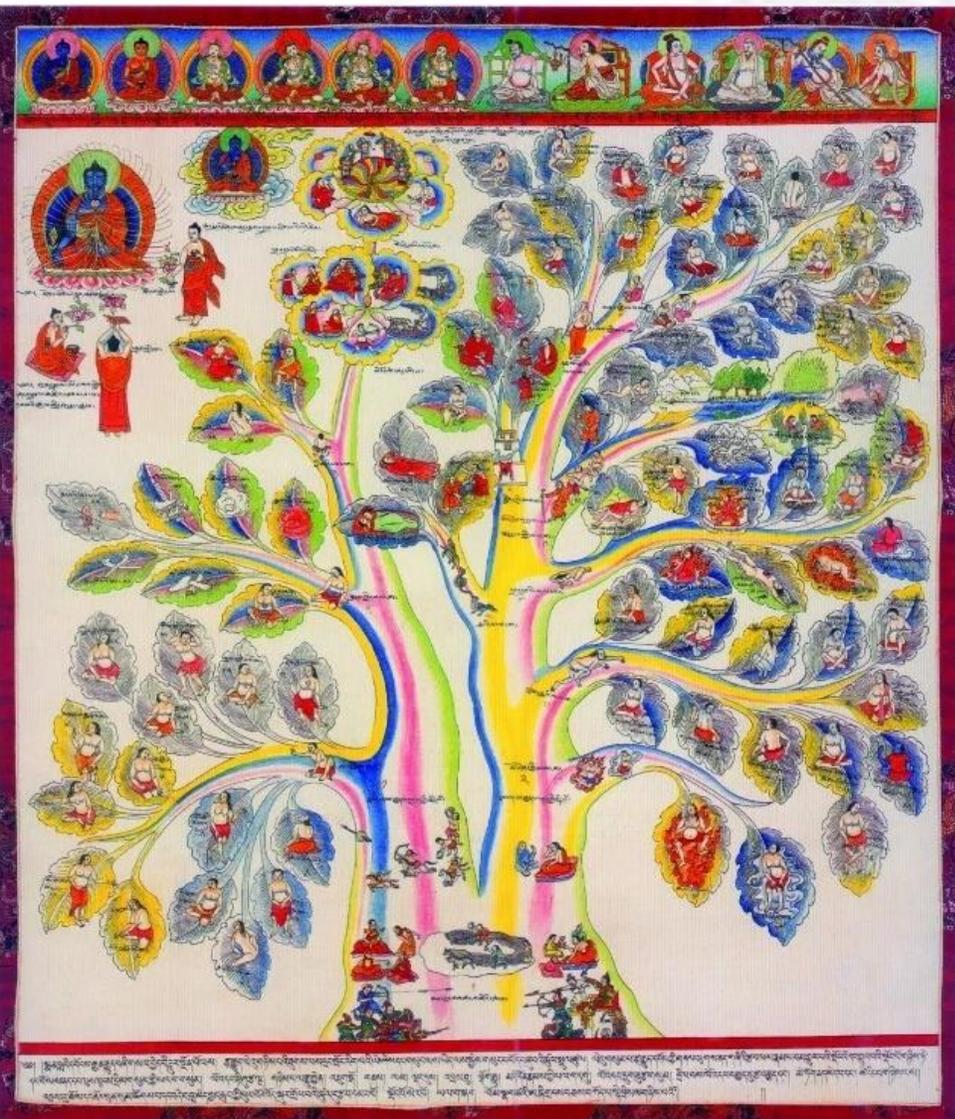
全套曼唐是以《四部医典·蓝琉璃》的文字内容为主要依据，其中根据《根本医典》的内容绘制的共四幅；根据《论述医典》的内容绘制的共三十五幅；根据《秘诀医典》的内容绘制的共十六幅；根据《后续医典》的内容绘制的共二十四幅。加上最后一幅历代名医图，总共是八十幅。

“医理树喻图”是藏医最具特色的曼唐，它以一种特殊的“愿望树”的形式，形象生动地表达了藏医学的全部内容，来表达藏医在人体生理稳态和病理状态、诊断原则和治疗原则的全部内容。在世界传统医学中具有无与伦比的特殊性。





# 《四部医典》树喻图



生理病理树有两个树干，一干表示正常生理，一干表示病态的生理。正常生理干有三枝，分别表示三大因素，每一个因素又分为五种，其表示隆的叶子有五片；表示赤巴的叶子有五片；表示培根的叶子有五片。七种物质，三种排泄物。正常生理树干共计二十五片叶子。

此树干还有两朵花和三个果实。这部分内容是藏医理论的一种哲学，即长寿之花及所结的无限安乐之果和健康之花结出的信仰（真理）和财富之果。这虽然是一种原始的朴素哲学观念，但是其中却蕴含着现代的一些医疗观念和思想。现代人所认识的健康、安乐、长寿、真理、财富之间相互影响制约和生成的关系，在远古的藏医理论中就得到了高度的认可和体现。





# 《四部医典》树喻图

在右侧的病理树干上分出九个树枝，分别代表疾病的根源、发病的原因、发病的途径、发病部位、发病通道、病变的具体位置、发病的规律和年龄、与地方和季节的关系、疾病的转归和致死的原因、疾病的转换、疾病的种类。每一枝又分出若干树叶。具体如下：

- 七大物质：食物精微、血液、肌肉、脂肪、骨骼、骨髓和精液，被称为人体七精华。）
- 疾病的根源有三：藏医理论认为贪（贪欲）、嗔（愤怒）、痴（痴呆）是引起人体疾病的总根源，其中贪欲能引起隆病的发生；嗔怒能引起赤巴病的发生；愚痴能引起培根病的发生。
- 发病的原因有四：分别为季节影响、饮食不当、起居不当、其他因素如外力，鬼邪作祟。
- 发病的途径有六：分别为皮肤、肌肉、经脉、骨骼、五脏、六腑。
- 发病的部位有三：隆、赤巴、培根。培根病好发于上部，近头、脑。赤巴病好发于中部，近肝、膈。隆病好发于下部，近腰、髌。
- 发病的通道有三：隆道、赤巴道、培根道，三者流通人体表里的十五种道路。其中，隆的通道：骨骼、耳、作为触觉门户的汗孔、心及作为命根的大肠；赤巴通道：血、汗、眼、肝、胆及小肠；培根通道：肉脂髓精、大小便、鼻舌、肺脾肾、胃膀胱，共有十五片叶子来表示）。





# 《四部医典》树喻图

- 发病的规律和年龄、地区和季节的关系：隆、赤巴、培根三个内容：分别为老年易患的隆病，壮年易患的赤巴病，儿童易患的培根病；高寒地区易患隆病，干热地区易患赤巴病，潮湿地区易患培根病；隆病在夏季、夜晚、拂晓易发，赤巴病在秋季、中午、午夜易发，培根病在春季、黄昏、清晨易发。
- 致死的原因有九个：分别为阳寿终结、三因相克、用药失当、伤中要害、呼吸停止、发烧过度、体温耗尽、虚弱不堪、其他因素。
- 疾病的转换有十二种：分别为赤巴病愈后转培根病，隆病愈后转培根病，赤巴病愈后转隆病，隆病未愈又添培根病，赤巴病未愈又添隆病，培根病未愈又添赤巴病，赤巴病未愈又添培根病，隆病愈后转赤巴病。培根病未愈又添隆病，隆病未愈又添赤巴病；培根病愈后转赤巴病，培根病愈后转隆病。
- 疾病的归类有两个：寒病和热病。

最底下的藏文为对该幅树图的内容来源及内容简要部分。这样，通过一棵两干、十二枝、八十八片叶子、两花、三果的“愿望树”，清楚地将藏医的生理、病理内容表达出来了。





# 《四部医典》树喻图

“医理树喻图”是藏医最具特色的曼唐，以一种特殊的“愿望树”的形式，来表达藏医在人体生理稳态和病理状态、诊断原则和治疗原则的全部内容。除了上图（生理病理图）外，还有：

- ② 诊断树
- ③ 治疗树
- ④ 组织胚胎发育图
- ⑤ 人体白脉图
- ⑥ 饮食、饮食禁忌和饮食进量图
- ⑦ 医师品德图
- ⑧ 养生法图
- ⑨ 药物图



# 《四部医典》知识结构整理

藏医学距今已有千年历史，主要研究人体的生理和解剖、疾病的病因、诊治及预防疾病和保健养生、藏药等。

《四部医典》中的知识分类结构与藏因明学耦合。其中的理论基础、诊治思维、藏药基础等分别以因明学中的四行学说、因果辩、对立与联系、认知法等为其哲学基础。如：

- ① 藏医病机学认为所有的疾病皆由无明（贪嗔痴）产生。无明产生贪嗔痴，贪嗔痴产生龙、赤巴、培根，龙、赤巴、培根间的不平衡是所有疾病的病因。（因明因果关系中的近因、远因思想）
- ② 藏医诊治学中的概念、判断、推理的思辨方法；藏医预防学中的非共存对立思想（ལྷན་ཅིག་མི་གནས་འགཞུགས།）；藏药学中的药性、药味、药效与四行的关系（土坚硬凝结成形，水润滑滋润，火万物生长，风清扬流动）；炮制方法中的对立思想等。

以上作为例证可发现，藏医学和因明学在很多内容上相互贯穿、相互渗透，尤其在《四部医典》的知识分类上有所体现。（因明学分类体系）



# 《四部医典》知识结构整理



## 《四部医典》知识结构：

第一部《根本医典》（རྩ་བའི་རྒྱུད།）：饮食35，习惯6，药物50，诊断7。共98片叶子。

དེ་ལྟར་བྲན་ལ་སོ་སྤྱོད་ལམ་ལ་དག་སྐྱབ་ལ་ལྟ་བུ་དབྱེད་ལ་བདུན་བསྐྱེད་ལ་གསོ་བྲབས་ཀྱི་ལོ་མ་དགུ་བཅུ་ཚ་བརྒྱད་པོ་འདི་ཉིད་བར་བར་དུ་ཡིངས་མེད་པར་སྐྱབ་པས་སྤྱར་ཞིང་། བད་པས་གུས་པ་དང་བཅས་ཏེ་བརྩོན་པས་བསྐྱེན་ན། བད་པས་ཀྱི་འདམ་ལས་རྒྱར་དུ་གྲོལ་བར་འགྱུར་པོ།

第二部《论述医典》（བཤད་པའི་རྒྱུད།）：4桩根，1颗树，40根树枝，150片叶子，2个果实。

དེ་ལྟར་ཚ་བ་བཞི་ལ་སྤྱོད་པོ་བཅུ་གཅིག་ལམ་ག་བཞི་བརྒྱལ་ལོ་འདབ་བརྒྱ་དང་ལྟ་བུ་བཞི་འབྲས་བུ་གཞིས་བཅས་བཤད་པ་སྤྱི་རྒྱད་ཀྱི་བསྐྱེས་དོན་ས་བཅད་སྤོང་འགྲེམས་སྤྱིང་པོའི་ཚུལ་དུ་བཀོད་པའོ།

第三部《秘诀医典》（མན་ངག་གི་རྒྱུད།）：16桩根，92颗树，408根树枝，5467片叶子，3个果实。

དེ་ལྟར་རྒྱད་ཆེན་འདི་ཉིད་ཀྱི་གཞུང་དོན་ཆམས་མུ་རན་མེད་པའི་སྤོན་ཤིང་གི་ཚུལ་དུ་དཔེ་དང་སྤྱར་ནས་བཀོད་པ་འདི་ཉིད་ལ་ཚ་བ་བཅུ་དུག་སྤོང་པོ་གོ་གཞིས་ལམ་ག་བཞི་བརྒྱ་དང་བརྒྱད་ལོ་འདབ་ལྟ་སྤོང་བཞི་བརྒྱ་པོ་བདུན་མེ་ཏོག་གཞིས་འབྲས་བུ་གསུམ་བཅས་ལེགས་པར་བཀོད་ཟེན་ཏོ།

第四部《后续医典》（བྱི་མའི་རྒྱུད།）：药物1200，医治方法366.1桩根，4颗树，25根树枝，190片叶子，1多花，1个果实。

སྐབས་དོན་གསོ་བྱེད་གཞིན་པོ་སྤོང་དང་ཚ་བཞིས་དང་གསོ་བྲབས་ལག་ལེན་སྐྱེས་བརྒྱ་དུག་ཚ་བུ་ཀྱི་རྩམ་གྲངས་ནང་གསེས་བཅས་པ་གཙོ་བོར་བརྒྱད་བ་བཅས་དེ་ལྟར་བྱི་མ་འཕྲིན་ལས་རྒྱད་འདིར་ཚ་བ་གཅིག་ལ་སྤོང་པོ་བཞི་ལམ་ག་ཞེས་སྤྱོད་པོ་མ་བརྒྱ་དང་དགུ་བཅུ་ལོ་འདབ་དེ་དག་གི་རྩེར་མེ་ཏོག་དང་། འབྲས་བུ་གཅིག་





# 《四部医典》药物分类



药剂	数量	能净	数量	缓峻法	数量	饮食	数量
汤药	77	泻药	82	熏烟	30	谷物类	13
粉末	165	呕吐药	16	浸浴	24	肉类	15
丸药	22	鼻药	7	涂敷	28	酥油类	12
糖浆	20	缓导剂	9	针刺	77	蔬菜类	6
脂剂	23	浣肠	14	火灸	71	米粥类	29
煨灰	13	根泻	2	刀针	110	牛奶类	11
膏剂	17					水类	10
药酒	19					酒类	5
珍宝	20						
草药合剂	42						
<b>药剂类：共418</b>		<b>能净类：共130</b>		<b>缓峻法：共340</b>		<b>饮食：共101</b>	





# 《四部医典》治疗法分类

切脉查尿	数量	药剂	数量	能净	数量	缓峻法	数量
脉	27	汤药	5	清油疗法	7	针刺	28
尿	21	粉末	11	泻药	36	火灸	10
错乱	34	丸药	5	呕吐药	20	熏烟	5
治法	47	糖浆	2	鼻药	11	浸浴	5
		脂剂	2	缓导剂	9	涂敷	4
		煨灰	8	浣肠	11	刀针	16
		膏剂	1	根泻	18		
		药酒	5				
		珍宝	11				
		草药合剂	7				

切脉尿诊法：共**129**

药剂治疗法：共**57**

能净治疗法：共**112**

缓峻治疗法：共**68**





# 知识表示形式与语言知识表示

《四部医典》中将近有2000种生理、心理疾病分为15类，三因病259种，痼疾病293种，热疫病338种，头部和五官病170种，脏腑疾病166种，隐疾22种，零杂疾病283种，自然疮94种，儿童疾病62种，妇女病支54种，中邪病支83种，创伤支134种，中毒支70种，返老支5种，壮阳支17种。其中，三因病，痼疾病，热疫病，头部和五官病，零杂疾病，脏腑疾病等的分类过多，可做更为详细的分类。

知识表示的形式决定了知识获取的形式和难度。跟普通语句相比，《四部医典》中的知识表示形式较为独特。《四部医典》多以诗体形式书写，又参杂散体、问答形式的语句。在一定意义上，它的语言相对精炼，术语用词规范。但是，在另外一层意义上，《四部医典》知识表示形式的研究具有一定的挑战性。《四部医典》中表示实体的语言形式单位不一致，或以短语形式，或以词语形式，其中稀用词、梵文词汇较多，且领域性较强，训练数据稀少，在一定程度上提升了藏文分词、标注难度。因此，《四部医典》中的知识表示形式的研究更为重要。



# 半结构数据与非结构数据对比 ( “བླ་སའི་བདེ་བྱེད་ཀྱི་པམ།” 与 “གཅིན་སྩི་བ་ལུ” )

**【药物成分】** 红花、诃子、余甘子、乌龙胆、巴夏嘎、波棱瓜子、渣驯膏、骨碎补、猪血粉等。

**【功能主治】** 清热解毒，调和龙，开胃，止血。用于“木布”病引起的胃肠溃疡出血，亦可用于陈旧热症，久病不愈的身倦体重，肝区疼痛，胃腕疼痛，腹胀，食欲不振，消化不良，呕逆、泄泻等急慢性胃炎，及新旧肝病等。

**【适应症状】** 急慢性胃炎，萎缩性胃炎，胆汁返流性胃炎，胃、十二指肠溃疡、出血，返酸，胃胀，胃灼热，食欲不振，呕吐，胃癌术后修复，胃疼，胃下垂，胃窦炎。

**【专药专治】** 胃病治疗彻底:夏萨德西丸专治胃溃疡、十二指肠溃疡、浅表性胃炎、肥厚性胃炎、结肠炎等胃肠病变;针对病因。

**【服药方法】** 嚼碎口服，一次5-7丸，--日2-3次。

**【规格】** 每丸重0.2克。

**【禁忌】** 生冷、辛辣、酸腐食物。

**གསོ་བྱེད་ཐབས། ( 治疗方法 ) :** གཅིན་སྩི་ཡིན་མིན་གཅིན་པའི་གྲངས་ལ་བརྟེན། ཟླ་ལྷུ་ཡིན་མིན་གཅིན་པའི་མདོག་ལ་བརྟེན། ཚེས་པ་དང་། གཅིན་སྩི་བ་ལུ་བདེ་ཚོལ་གཙོ་བོར་བཅས། །འབྲུ་རྩིང་སྐྱམ་སའི་ཤ་ལ་སོགས་པ་ཡིས། །ཡང་ཞིང་རྩུབ་པའི་བས་སྐྱོམ་སྦྱོང་ལམ་བཟོ་ན། །ཞིས་གསུང་བ་ལྟར་ཐོག་མར་ལུས་རྩུངས་གསོ་བ། བར་དུ་མཐལ་རུས་གསོ་ཞིང་འཇག་པ་གཅོད་ཟུ། མཇུག་དུ་ནད་རོ་འདོན་པ་དང་མཐལ་དོད་བསྐྱེད་ཟུ།

**■ ལྷན། ( 药物治疗 ) :** གཅིན་སྩི་བདེ་བྱེད་དང་ཨ་རུ་བརྩུ་བའི་སྦྱང་དོམ་མཁྲིས་དང་གི་ལང་། གསེར་བྱེད་བ་ཤ་ཀ །ཡུང་བ་བཞི་(ཡུང་བ། ལྷུར་བ། ལྷུ་རུ་བ། གཟེ་མ།) ཐང་གིས་དབུལ། ཡང་ན་ལོང་སྐྱོན་ལུགས་ཀྱི་ཨ་རུ་བཙོ་བརྩུད་དང་འཇག་སྐྱོམ་ཨ་རུ་བདུན་པ་སོགས་ལྷ་ཕྱི་རིས་མོས་ལུ་བཏང་ན་རྩུང་མཁྲིས་བད་ཀན་ལས་གྱུར་བའི་བ་ལུ་འི་རིགས་ལ་པམ། གྲང་བ་ཤས་ཆེ་བའི་རིགས་ལ་ལུག་སྐྱེལ་བརྩུ་བ་དང་། འབྲས་རྩ་བཙོ་སྐྱ་གར་ཆང་གིས་བཏང་བས་ཐོག་ལེ་འཇག་པ་གཅོད་ཅིང་མཐལ་རྩ་འགྲམས་པ་སེལ། གཞན་ཡང་སེ་འབྲུ་ཉེར་སྐྱ་སེ་འབྲུ་ཉི་དཀྱིལ། ལོ་ལྷ་ཉེར་བརྩུད་སོགས་གཏོང་ཚོག

**■ དཔྱད། ( 有待确症 ) :** ལུག་གཟེར་ཆེན་དེུ་ཆང་བཟན་དང་སྦང་སྐྱམ་ཆང་བཟན་བཙོས་པས་དུགས་པར་བྱ། ཚོགས་པ་བརྩུ་གཉིས་བཙོ་བརྩུད་སོགས་ནད་དམིགས་གར་ཡོད་སར་མེ་བཙོའ་གདབ།





# 《四部医典》阅读体验简述



《四部医典》是集成多方地域的医学理论与临床经验的医学基础理论著作。知识结构清晰、系统健全。但是，仍有知识分类不够适当之处，如：

- 现有《四部医典》中的药物是按照治疗方法进行归类，而未形成基于形状、药性的药物分类。适宜于知识图谱构建的药物分类体系、治疗法分类体系、病因分类体系、症状分类体系等有待进一步完善。
- 《四部医典》内容是藏医药理论、临床的基础文本，多项知识在后期不断被突破、翻新。知识更新与理论渊源标记工作在图谱构建中显得尤为重要。
- 《四部医典》中的知识自成体系，但未按照国际疾病分类WHO-ICD10进行对应。藏医药知识分类体系与国际接轨是未来藏医药发展的必然趋向。
- 考虑目前藏文文本处理技术相对滞后、数据稀疏，《四部医典》知识基本呈非结构化现象，对医药文本中语言知识的表示方法研究极其重要。



# 附录：

(通用分词、标记标准与藏医药领域分词、标记体系，分类层级问题)

ICS 35.240.01  
L 70



中华人民共和国国家标准

GB/T 36452—2018

信息处理用藏文分词规范

Specification on Tibetan segmentation for information processing

2018-06-07 发布

2019-01-01 实施



国家市场监督管理总局  
中国国家标准化管理委员会

发布

ICS 35.240.01  
L 70



中华人民共和国国家标准

GB/T 36338—2018

信息处理用藏文文献文本信息标记规范

Specification on Tibetan literature text markup for information processing

2018-06-07 发布

2019-01-01 实施



国家市场监督管理总局  
中国国家标准化管理委员会

发布



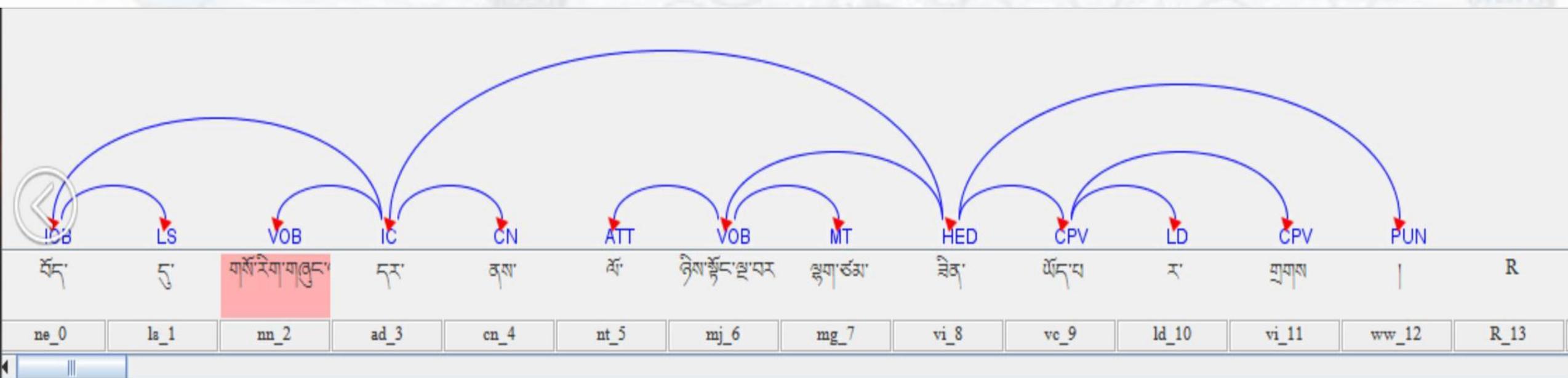
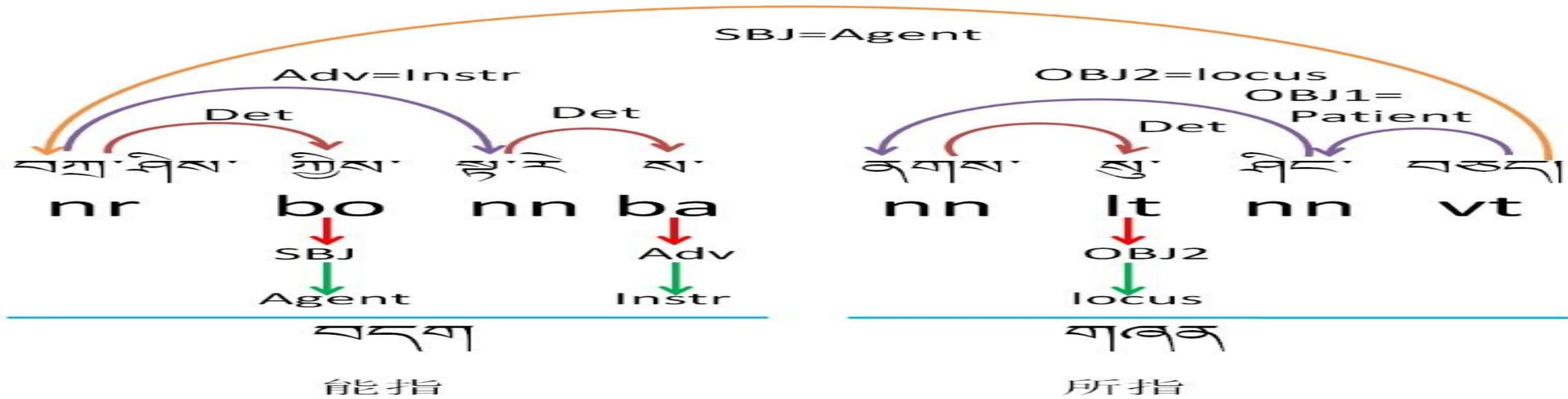


- བདུད་ཚི་\_nn རྩིང་པོ་\_nn ཡན་ལག་བརྒྱད་པ་\_nz གསང་བ་མན་ངག་\_nz གི་\_gi རྒྱུད་\_nn ལས་\_jg དུམ་ལུ་\_nn དང་པོ་\_mx ཚ་བ་\_nn རི་\_gi རྒྱུད་\_nn ཅེས་\_ce བྱ་བ་\_nv བཞུགས་\_vi རོ་\_uz |\_ww
- འདི་སྐད་\_rz བདག་\_rr གིས་\_bo བཤད་པ་\_vt རི་\_gi དུམ་\_nt གཅིག་\_mj ན་\_cv |\_ww དྲང་སྲོང་\_nn གི་\_gi གནས་སྐྱོན་\_nn རྒྱུ་གི་\_gi རྒྱུང་སྲུང་\_nn རྩ་ན་སྲུང་\_ns ཅེས་\_ce བྱ་བ་\_nv རིན་པོ་ཆེ་\_nn རྩ་མུ་\_mj ལས་\_jg རྒྱུ་བ་པ་\_vi རི་\_gi གཞུལ་ཡས་ཁང་\_nn ཡོད་\_vc དེ་\_cg |\_ww ཁང་པ་\_nn དེ་\_rz རི་\_gi རྒྱན་\_nn རྩི་\_ci རྩན་\_nn རྒྱུ་གི་\_gi རོར་ལུ་རིན་པོ་ཆེ་\_nn རྩ་མུ་རིན་པོ་ཆེ་\_nn རྩི་\_ci རྩན་\_nn གི་\_gi རྩད་\_nn དང་\_cd མཁྲིས་པ་\_nn རི་\_gi རྩད་\_nn དང་\_cd བད་ཀན་\_nn རྒྱུ་གི་\_gi རྩད་\_nn དང་\_cc རྩན་པ་\_vc རི་\_gi རྩད་\_nn དང་\_cd འདུས་པ་\_vi རི་\_gi རྩད་\_nn ལས་\_jg རྒྱུར་པ་\_vi རི་\_gi རྩད་\_nn བཞི་བརྒྱ་ཚ་བཞི་\_mj རིལ་བ་\_vt ར་\_ld རྩེད་པ་\_vu |\_ww ཚ་བ་\_nn རི་\_gi རྩད་\_nn ལ་\_ls བསིལ་\_vi དུ་\_ld རྒྱུར་པ་\_vi |\_ww རྒྱང་པ་\_as རི་\_gi རྩད་\_nn ལ་\_ls རྩད་\_nn དུ་\_ld རྒྱུར་པ་\_vi |\_ww བཞེགས་རིགས་\_nn རྩོང་སྲུང་\_mg བརྒྱད་ཅུ་\_mj རི་བ་\_vi ར་\_ld རྩེད་པ་\_vu |\_ww དགོས་འདོད་\_nn ཐམས་ཅད་\_dw ཡིད་བཞིན་\_nn དུ་\_ld རྒྱུ་བ་པ་\_vi ར་\_ld རྩེད་པ་\_vt ཡོད་\_vc དོ་\_uz |\_ww |\_ww

问题：

- 1) 专业词汇标记分类不够细致，分词错误
- 2) 实体与词对应，又与短语对应
- 3) 实体名称的指代和藏匿现象普遍





བཀྲ་སྒྲིལ་བདེ་ལེགས།

敬请各位不吝赐教，扎西德勒！

