

OMAHA 七巧板医学术语集发布文件规范

Update in 2019.04.17

Prepared by OMAHA 秘书处

一、术语集发布周期

OMAHA 七巧板医学术语集（以下称 OMAHA 医学术语集）将在每个季度第一个月 20 号通过官网发布，1 年发布 4 次。当年发布内容规划，可以详见官网 HiTA 术语中的发布计划。

二、发布文件规范

1. 文件命名

OMAHA 医学术语集发布文件的名称由五个部分组成，每个部分通过下划线“_”分开。文件名称命名如：OMAHA 七巧板医学术语集_“构件类型”_“子集名称”_“版本类型”_“发布日期”，文件名称之后为文件扩展名。

1.1 构件类型

文件名的“构件类型”用来描述文件的内容和用途。目前发布的构件类型为概念、术语、关系、映射。

1.2 子集名称

文件名的“子集名称”提供了额外的信息来描述文件的内容。若该名称存在，则说明该文件中内容为术语集子集内容；若不存在，则说明该文件不是子集内容。OMAHA 对外发布了子集定制工具（详见 HiTA 术语服务平台-子集定制），支持拉取个性化术语集子集。

1.3 文件版本类型

OMAHA 医学术语集每次发布包含三个版本类型，完整版、最新版、变更版，以下为这三个版本的详细说明。

- 完整版：包含了 OMAHA 医学术语集所有发布版本的内容。
- 最新版：包含了 OMAHA 医学术语集最新版本的内容。
- 变更版：仅包含对上一版本的补充和变更信息。

上述文件版本类型在下面的章节中会有详细的说明。

1.4 发布日期

文件名的“发布日期”元素标识该文件发布日期，用于对 OMAHA 医学术语集的版本控制。使用 GB/T 7408 标准格式为“YYYYMMDD”。

2. 文件格式

OMAHA 医学术语集对外发布文件类型分为三种：

- 文件包统一使用的压缩格式“.rar”；
- 概念、映射、关系、映射等文件使用 UTF-8 编码的文本文件“.txt”；
- 本体文件使用 UTF-8 编码的本体文件“.owl”。

3. 字段命名规范

首行为标题行是文件中每列的字段名称，字段名称使用“小骆驼拼写法”，如 conceptId。

4. 构件标识符

构件标识符为一组唯一的整数标识符，由 7-18 位数字组成，用于表示 OMAHA 构件，如概念、术语、关系和映射，不包含任何有关 OMAHA 构件具体内容的信息。

每个 OMAHA 构件标识符都包含首位构件分类符、中间 5-16 位随机数字区分符和末尾校验符。其中构件分类符分别对应 OMAHA 医学术语集构件中的概念、术语、关系、映射，依次以数字 1、2、3、4 表示；区分符是一组 5-16 位随机产生的数字，保证标识符的唯一性；校验符根据构件分类符和区分符按照一定规则生成。

例如：标识符“112345678”，首位数字“1”为分类标识符，中间数字“1234567”为区分符，末位数字“8”为校验码。

4.1 概念标识符

在 OMAHA 医学术语集里面，每个概念表示一个临床观念，且每个概念都有一个唯一的数字标识符表示，构件分类符为“1”。

4.2 术语标识符

在 OMAHA 医学术语集里面，术语是对概念的语言指称。表达相同概念的不同术语，则为同义词，且每个术语都有一个唯一的数字标识符表示，构件分类符为“2”。

4.3 关系标识符

关系是指建立一个术语集之间概念与概念之间的关系，且每条关系都有一个唯一的数字标识符表示，构件分类符为“3”。

4.4 映射标识符

映射是指建立一个术语集的编码、概念或术语与另一个术语集内有着相同或相似语义的编码、概念或术语之间的关联，且每个映射都有一个唯一的数字标识符表示，构件分类符为“4”。

三、 文件内容介绍

当前 OMAHA 医学术语集内容包括了概念表、术语表、关系表、映射表。本章将对概念表、术语表、关系表、映射表相关字段以及表示进行详细介绍。

1. 概念表介绍

1.1. 字段说明

概念表包含了 OMAHA 医学术语集发布概念，具体字段说明如表 1：

表 1：概念表字段说明

字段名	字段含义	数据类型	说明
id	概念标识符	Longinteger	7-18 位的数字
releaseDate	发布时间	Date	以 YYYYMMDD 表示
status	状态	Boolean	表示该概念是否有效，有效为‘1’，无效为‘0’

1.2. 文件示例

见概念表示例表 2，该概念的唯一标识符是“1859123”，在 2019 年 1 月 20 号为有效状态。

表 2：概念表示例

id	releaseDate	status
1859123	20190120	1

2. 术语表介绍

2.1. 字段说明

术语表包含了表达医学概念的术语，具体字段说明如表 3：

表 3：术语表字段说明

字段名	字段含义	数据类型	说明
id	术语标识符	Longinteger	7-18 位的数字
releaseDate	发布时间	Date	以 YYYYMMDD 表示
status	状态	Boolean	表示该术语是否有效，有效为‘1’，无效为‘0’
conceptId	概念标识符	Longinteger	指定该术语所表达的概念，7-18 位的数字
term	术语	String	概念的指称
termType	术语类型	Number	许用术语为‘0’，首选术语为‘1’，待定首选术语为‘2’（具体说明见表 4）
semanticTag	语义标签标识符	String	该术语所表达概念的语义类型（语义标识符以及对应的语义标签名称详见表 5）

2.1.1. 术语类型（termType）

确定术语类型的规则将在《OMAHA 七巧板医学术语集入门指南》介绍，其中术语类型具体解释如表 4：

表 4：术语类型说明

术语类型	取值	说明
首选术语	1	一个概念中通过特定的规则被指定为第一选择的术语。
待定首选术语	2	一个概念中通过特定的规则被指定为待定首选术语的术语，同一个概

		念下首选术语与待定首选术语只存在一个。
许用术语	0	某一概念的诸术语中，首选术语/待定首选术语以外的同义词。

2.1.2. 语义标签(semanticTag)

语义标签标识符以及对应的语义标签名称，如表 5 所示：

表 5：语义标签标识符与语义标签名称对应表

语义标签标识符	语义标签名称
T001	临床所见
T002	操作
T003	组织机构
T004	药品
T005	物质
T006	分子活性
T007	生物过程
T008	基因突变
T009	基因
T010	人体形态与结构
T011	生物
T012	限定词
T013	物理实体
T014	标本
T015	事件
T016	观察对象
T017	物理能量
T018	疾病
T027	发病和发病过程
T028	严重程度
T029	生命周期
T030	评价结果
T031	病理过程
T032	方位
T033	医疗器械
T034	生活制品
T035	解剖结构
T036	异常形态结构
T037	交通工具
T038	精度类型

T039	量纲类型
T040	技术
T041	单位
T042	剂型
T043	给药途径
T044	频次
T045	药品本位码编码
T046	批准文号
T047	细胞结构
T048	植介入医疗器械
T049	武器
T051	入路
T052	操作方法
T054	时间范围
T055	核苷酸位点
T056	基因组
T057	转录本编码
T058	核苷酸改变
T059	氨基酸改变
T060	基因位点
T061	数值
T062	OMAHA 医学术语
T065	观测操作
T066	身体状态
T067	体位

2.2. 文件示例

见术语表示例表 6 ,术语 A1 的唯一标识符是“21775398”,术语 A2 的唯一标识符是“21776880”。由于术语 A1 和 A2 均表达概念“1859123”,所以 A1 和 A2 为同义词。

表 6 : 术语表示例

id	releaseDate	status	conceptId	term	termType	semanticTag
21775398	20180120	1	1859123	A1	***	***
21776880	20180120	1	1859123	A2	***	***

3. 关系表介绍

3.1. 字段说明

关系表包含了概念之间的关系，表格的具体字段说明如表 7 所示：

表 7：关系表字段说明

字段名	字段含义	数据类型	说明
id	关系标识符	Longinteger	7-18 位的数字
releaseDate	发布时间	Date	以 YYYYMMDD 表示
status	状态	Boolean	表示该映射是否有效,有效为'1',无效为'0'
conceptId1	起始概念标识符	Longinteger	概念标识符, 7-18 位的数字
conceptId2	目标概念标识符	Longinteger	概念标识符, 7-18 位的数字
relationshipType	关系类型标识符	String	由字母“R”和一组数字组成(关系类型标识符以及对应的关系类型名称见表 8)
relationshipGroup	关系组	Number	非零整数相同且 conceptId1 相同则为一组,若值为 0 则不成组(具体说明见表 9)

3.1.1. 关系类型(relationshipType)

关系类型具体介绍详见《OMAHA 七巧板医学术语集入门指南》，关系类型标识符以及对应的关系类型名称，如表 8 所示：

表 8：关系类型标识符与关系类型名称对应表

关系类型标识符	关系类型名称
R001	子类
R002	是...组成部分
R003	发生部位
R004	形态学改变
R005	临床过程
R006	严重程度
R007	发生于...期间
R009	后发于
R010	评价对象
R011	评价结果
R012	致病原因
R013	病理过程
R014	方位
R015	操作部位
R016	操作形态学改变
R017	使用的器械

R018	使用的通路器械
R019	植介入器械
R020	目标物质
R021	辅助性物质
R022	入路
R023	操作方法
R024	使用技术
R025	给药途径
R026	使用的能量
R027	受检标本
R028	受检成分
R029	精度类型
R030	受检时长
R031	指标
R032	操作条件
R033	有效成分
R034	批准文号
R035	药品本位码
R036	生产厂家
R037	剂型
R038	分子规格单位
R039	分母规格单位
R040	分子规格数值
R041	分母规格数值
R042	突变核苷酸位点
R043	所在基因组
R044	突变基因
R045	转录本编码
R046	核苷酸改变
R047	氨基酸改变
R048	基因位点
R049	基因表达部位
R050	细胞定位
R051	分子活性
R052	参与过程
R053	属性组

3.1.2. 关系组 (relationshipGroup)

定义相关关系组可以将概念定义的更加清晰，OMAHA 医学术语集中 relationshipGroup 字段用于标识关系组信息，relationshipGroup 字段是一个无符号无顺序不唯一的非零整数，具体说明如表 9 所示：

表 9：关系组说明

构件	关系组值	说明
关系	N*	非零整数相同且起始概念相同的关系为一组，关系表文件示例将进一步说明
关系	0	该关系不与其他关系成组，关系表文件示例将进一步说明

3.2. 文件示例

见关系表示例表 10，每行保存一条关系，以三元组的方式起始概念通过指定的关系类型与目标概念连接。

其中关系标志符为“33728967”的关系表示概念“1000015”与概念“1315633”的关系类型是“R002”，即“鼻中隔膜部”是“鼻中隔”的组成部分。

关系标识符为“33728967”的关系其 relationshipGroup 的值为“0”，因此该条关系不成组，关系标识符为“33728967”的关系同理；关系标识符为“31126949”与关系标识符为“31126954”的这两条关系其 conceptId 相同，且 relationshipGroup 的值为相同的非零整数“1”，表示这两条关系成组定义起始概念“1000473”，即沃弗综合征[1000473]的形态学改变[R004]为出血[1264332]，其发生部位[R003]是肾上腺皮质[1340209]，这两条关系成组使该概念的定义更加清晰。

表 10：关系表示例

id	releaseDate	status	conceptId1	conceptId2	relationshipType	relationshipGroup
33396355	20190120	1	1000015	1628842	R001	0
33728967	20190120	1	1000015	1315633	R002	0
31126949	20190120	1	1000473	1264332	R004	1
31126954	20190120	1	1000473	1340209	R003	1

4. 映射表介绍

4.1. 字段说明

映射文件包含了 OMAHA 医学术语集与其它各个术语标准的映射。表格的具体字段说明如表 11：

表 11：映射表字段说明

字段名	字段含义	数据类型	说明
id	映射标识符	Longinteger	7-18 位的数字
releaseDate	发布时间	Date	以 YYYYMMDD 表示

status	状态	Boolean	表示该映射是否有效，有效为'1'，无效为'0'
omahaId	构件标识符	Longinteger	用于映射构件标识符，7-18 位的数字
targetId	映射目标	String	映射术语集中概念、词汇 id 或编码
mapPriority	优先级	Integer	当一个构件存在多个映射目标时，从“1”开始依次分配优先推荐顺序； 当一个构件仅存在一个映射目标时，默认 mapPriority 为“1”；
mapVocabulary	映射术语集	String	来源数据集的名称简写，如《ICD-10 疾病和有关健康问题的国际统计分类 中文版，不包括扩展编码》（ICD10CN）

4.2. 文件示例

见映射表示例表 12，该映射表代表 OMAHA 医学术语集与 ICD10CN 的映射，其中 OMAHA 医学术语集的概念“1014418”存在多个映射目标编码，分别是“M72.1”、“H90.5”、“Q84.4”根据 mapPriority（优先级）取值，优先选择“M72.1”为概念“1014418”的映射目标；概念“1003722”仅存一个编码是“C67.9”的映射目标，mapPriority（优先级）取值默认为“1”。

表 12：映射表示例

id	releaseDate	status	omahaId	targetId	mapPriority	mapVocabulary
42876689	20190120	1	1014418	M72.1	1	ICD10CN
43822683	20190420	1	1014418	H90.5	2	ICD10CN
43822696	20190420	1	1014418	Q84.4	3	ICD10CN
4190632	20180120	1	1003722	C67.9	1	ICD10CN

5. 构件是否有效的状态（status）

构件	状态值	说明
概念	1	该概念有效，适合使用。 引用该概念的所有事件都是有效的。
	0	该概念无效，不适合使用。 OMAHA 医学术语集不会删除任何概念，当概念被弃用时，会插入一条概念状态(status)值为 0 的数据，下面章节会详细介绍。
术语	1	该术语是对概念的有效描述。 有效的术语可以对一个失效的概念进行描述，但一般只使用有效概念的术语。
	0	该术语无效，不再用于描述相关概念。
关系	1	该关系是有效的，适合使用。
	0	该关系是无效的，在使用的时候应该忽略无效关系。

		但无效的关系不一定意味该关系是错误的，可能无效的原因是经过推断该关系是多余的。
映射	1	该映射是有效的，适合使用。
	0	该映射是无效的，在使用的时候应该忽略无效映射。

6. 历史机制介绍

发布文件中的发布时间（releaseDate）和是否有效（status）字段可以追溯每个构件的所有变更历史。一旦发布，每行数据在文件中总是保持不变。发布时间指 OMAHA 术语集内容对外发布的时间，术语集构件数据的发布时间的变更意味着在这段时间内术语发生了改变。如果构件在两次发布期间（即编辑和维护过程中）更改了多次，那么只有最近修改过的记录才会被记录到发布文件中，而不是显示维护过程中所有变化的记录。

改变当前构件的内容需要添加新的一行数据，将 status 值设为 1，发布时间字段中的时间改为该行新数据的发布日期，并更新改变的字段信息，其他信息与上一版本保持一致，以下使用术语表数据作为示例，如表 10。

- 20180120 版本中，术语 A1 的 termType 为“0”；
- 20180420 版本中，术语 A1 的 termType 由“0”变更为“2”，即从许用术语转变成待定首选术语；
- 20180720 版本中，术语 A1 的 termType 由“2”变更为“1”，即从待定首选术语变为首选术语。

表 13：历史机制-术语类型改变

id	releaseDate	status	term	termType
21775398	20180120	1	A1	0
21775398	20180420	1	A1	2
21775398	20180720	1	A1	1

失效一个构件，同样需要添加新的一行数据，将 status 值设为 0，并在发布时间字段中的时间为新版本的发布日期，代表在该发布日期的时候该构件不再有效，其他数据字段都与上一版本保持一致，如表 11。

- 20180120 版本中，术语 A1 有效，用于表达概念“1859123”
- 20180420 版本中，术语 A1 无效，不再用于表达概念“1859123”。

表 14：历史机制-术语失效

id	releaseDate	status	conceptId	term
21775398	20180120	1	1859123	A1
21775398	20180420	0	1859123	A1

7. 文件版本类型介绍

OMAHA 医学术语集每次发布包含三个版本类型，其中概念、术语、关系、映射表的最新版和变更版本均可以从完整版中生成，本体文件发布版本类型为最新版。

完整版：包含了 OMAHA 医学术语集所有发布版本的内容。如 OMAHA 医学术语集完整版包含了每个概念、术语、关系以及映射所有版本的信息，以下使用术语表数据作为示例，如表 11。

最新版：包含了 OMAHA 医学术语集最新版本的内容。如 OMAHA 医学术语集术语表最新版中，对于每一个术语来说，只包含发布时间为最近时间的那条记录，如表 12。对于术语 A1 来说，20180120 为最近发布时间，所以在最新版中，术语 A1 只有发布时间为 20180120 的记录被保留；而对于 A2, A3 来说，它们的最近发布时间为 20180420，所有最新版只保留术语 A2, A3 发布时间为 20180420 的记录。

变更版：仅包含对上一版本的补充和变更信息。具体见表 13。如在版本发布日期为 20180120 的术语 A2 在 20180420 中失效，在变更版中被记录。20180420 版本中新增术语 A3，在变更版中被记录。

表 15：2018.4.20 发布的术语表完整版

id	releaseDate	status	conceptId	term
21111111	20180120	1	112345678	A1
21234511	20180120	1	112345678	A2
21234511	20180420	0	112345678	A2
21233311	20180420	1	112345678	A3

表 16：2018.4.20 发布的术语表最新版

id	releaseDate	status	conceptId	term
21111111	20180120	1	112345678	A1
21234511	20180420	0	112345678	A2
21233311	20180420	1	112345678	A3

表 17：2018.4.20 发布的术语表变更版

id	releaseDate	status	conceptId	term
21234511	20180420	0	112345678	A2
21233311	20180420	1	112345678	A3

四、 OWL 文件

结合行业需求在 20190120 版后对外发布 OWL 文件，OMAHA 医学本体相关介绍，详见《OMAHA 七巧板医学术语集入门指南》。

1. 本体相关实体介绍

OMAMA 医学术语集基于本体方式构建，采用 W3C 建议的 OWL 2 EL 标准，其中实体为类、对象属性、注释属性。构件标识符为一组唯一的标识符，用于标识 OMAHA 医学术语集的构件，如类和对象属性，其中注释属性没有标识符。

1.1. 类及其标识符

在 OMAHA 医学术语集中，每个类表示一个临床概念，且每个类都由一个唯一的数字标识符表示，由 7-18 位数字组成，该数字标识符来自 OMAHA 术语集概念表中的概念标识符。

1.2. 对象属性及其标识符

在 OMAHA 医学术语集中，每个对象属性表示一种类和类之间的关系，且每个对象属性都由一个唯一的标识符表示，由字母“R”和一组数字组成。对象属性详见第三章中 3.1.1 关系类型(*relationshipType*)，其中的属性“R053”表示属性组，仅用于本体文件。

1.3. 注释属性

在 OMAHA 医学术语集中，注释属性用于对类和本体自身进行注释。

1.3.1. rdfs : label

在 OMAHA 医学术语集中，每个类都有且仅有一个 label，该 label 来自术语表中术语类型为“首选术语”或“待定首选术语”的术语。

1.3.2. term.chs.synonym

在 OMAHA 医学术语集中，每个类可能有若干个同义词，这些同义词来自术语表中术语类型为“许用术语”的术语。

1.3.3. concept.semanticTag

在 OMAHA 医学术语集中，每个类都有一个语义标签，该标签来自术语表中的语义标签。

1.3.4. rdfs : comment

对当前本体的注释。

1.3.5. owl : versionInfo

表示当前本体的版本。

2. OWL 文件内容介绍

2.1. 文件示例

OMAHA 医学术语集采用函数式语法 (OWL Functional Syntax)，表达式示例见表 14：

表 18：本体表达式示例

本体表达式	注释
Declaration(Class(:11679591))	“11679591”是一个类
Declaration(ObjectProperty(:R002))	“R002”是一个对象属性

SubClassOf(:1000206 :1370599)	"1000206"是"1370599"的子类
EquivalentClasses(:1000058 ObjectIntersectionOf(:1390116 :11658734)	等价类, "1000058"等价于 "1390116"和"11658734"的交集
EquivalentClasses(:1000949 ObjectIntersectionOf(:1249193 ObjectSomeValuesFrom(:R053 ObjectIntersectionOf(ObjectSomeValuesFrom(:R007:1668422) ObjectSomeValuesFrom(:R004 :1394413)))))	属性组, 该属性组以属性"R007"和 属性"R004"的合取作为它的值
AnnotationAssertion(rdfs:label :1000954 "出血性梗死")	"1000954"的 rdfs : label 是"出血 性梗死"
AnnotationAssertion(omahat:term.chs.synonym :1000954 "出血性梗塞")	"1000954"的同义词是"出血性梗 塞"。

3. OWL 文件的浏览与编辑

借助本体编辑器, 用户可对 OMAHA 医学术语集进行浏览、编辑等操作。目前已经有多种本体编辑器, 例如 Protégé、Fluent Editor、TopBraid Composer 等。

Protégé 是斯坦福大学生物医学信息研究中心基于 Java 开发的本体编辑器, 广泛应用于生物医学、电子商务等领域的本体开发。通过 Protégé, 用户可方便而直观地浏览和编辑 OMAHA 医学术语集, 因此本文档仅对使用 Protégé 本体编辑器浏览 OMAHA 医学术语集 OWL 文件的情况进行介绍。

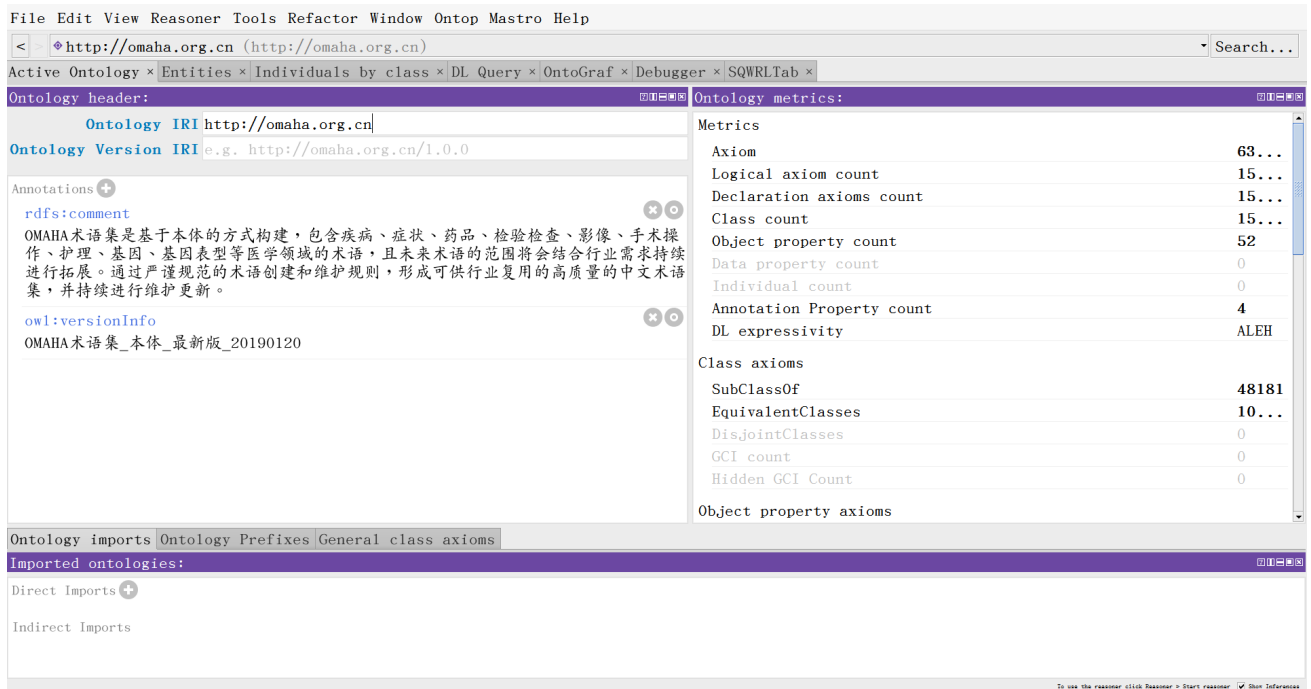
3.1. Protégé 的下载、解压和运行

Protégé 下载地址 : <http://protege.stanford.edu/products.php#desktop-protege>。解压下载的文件, 运行脚本 (run.bat, run.sh, run.command) 或者 Protege.exe 即可。

3.2. 使用简介

3.2.1. Active Ontology 标签页

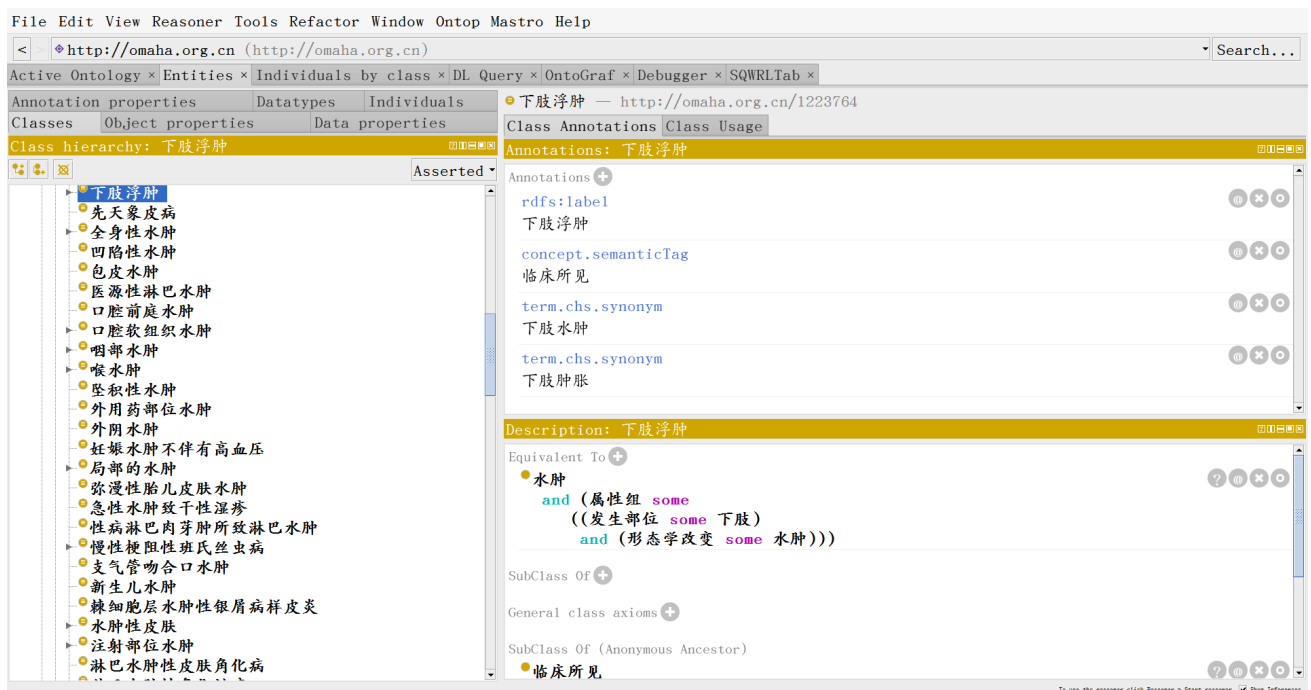
在 Active Ontology 标签页中, 左侧是对本体的注释和版本信息, 右侧是对本体的统计数据。见下图所示 :



3.2.2. Entities 标签页

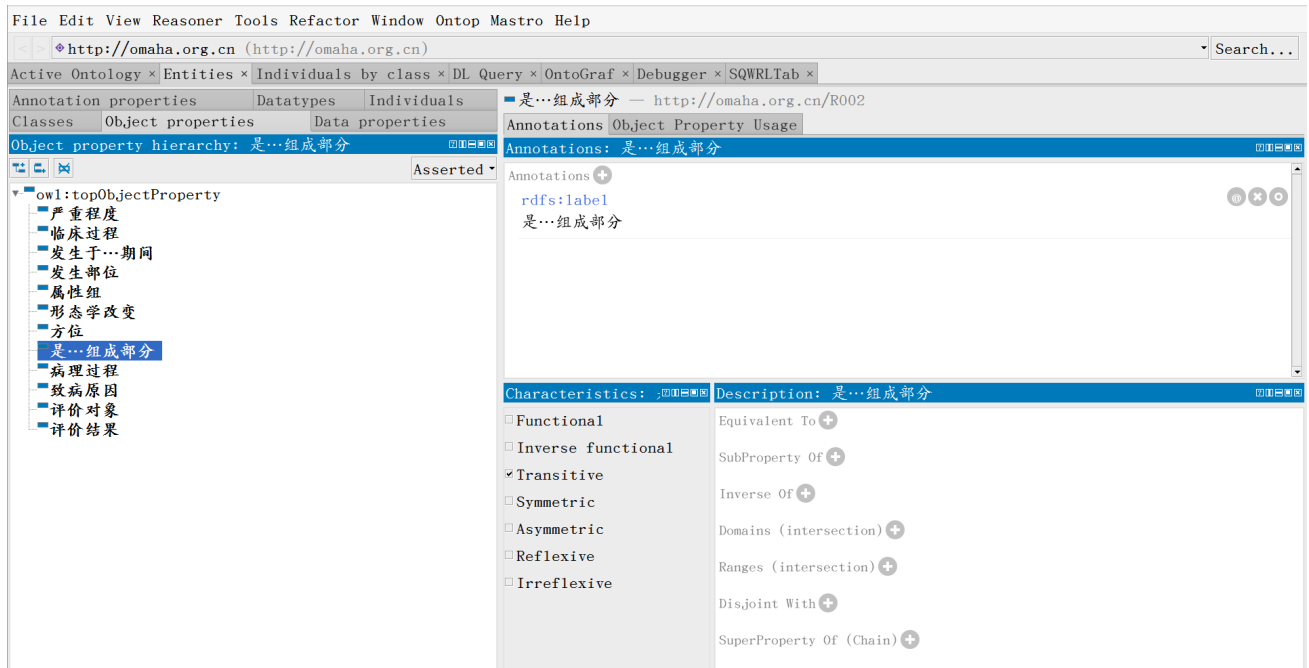
3.2.2.1. Classes

在 Classes 标签页中，可以浏览类的层级、注释属性和描述。左侧是类的层级，右侧上半部分是类的注释属性，包括其 label、同义词和语义标签。右侧下半部分显示该类的描述，例如 Equivalent To 关系和 SubClass Of 关系。见下图所示：



3.2.2.2. Object properties

在 Object properties 标签页中，左侧是对象属性的层级，右侧上半部分是对对象属性的注释属性，右侧下半部分显示对象属性的描述，例如 SubProperty Of 关系。见下图所示：



3.2.2.3. OntoGraf

OntoGraf 标签页中，可以对类及类之间的关系进行可视化展示。见下图所示：

