

标准文献结构化数据规范 第2部分：指标元数据

Structured data specifications of standard documents
Part 2: Indicator metadata

2016-05-23 发布

2016-07-01 实施

河北省质量技术监督局 发布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 指标元数据的属性 2

5 指标元数据模型 2

6 技术指标元数据描述 3

7 指标元数据字典描述 10

8 指标元数据扩展原则和方法 10

附录 A（规范性附录） 标准文献指标元数据代码集 11

附录 B（资料性附录） 标准文献结构化指标元数据-字典描述 12

参考文献 17

前 言

本部分按照GB/T 1.1-2009给出的规则起草。

DB13/T 2351《标准文献结构化数据规范》包括了3部分：

- 第1部分：全文元数据；
- 第2部分：指标元数据
- 第3部分：元数据著录规则。

本部分是第2部分，本部分与第3部分配套使用。

本部分由河北省服务业标准化技术委员会提出并归口。

本部分起草单位：河北省标准化研究院。

本部分主要起草人：马星晓、施健、卢艳芳、郭天强、闫志强、李建勋、谢燕、张园园、陈南、赵倩、安姣。

引 言

标准文献电子文档格式有DOCX、数字版PDF、扫描版PDF。在生产、科研和管理过程中，往往需要定位到标准内容、指标信息上，如：检索标准指标、指标比对及标准水平分析等。但是，现有的标准文献电子文档为非结构化文档，这一特性决定它的内容检索指向性笼统模糊，故而，以非结构化文档为单元存储无法实现小颗粒度的指标数据的检索、比对分析。只有通过数据处理将非结构化文档转换为结构化或半结构化文档方能解决上述问题。

为寻求上述瓶颈的突破，河北省标准化研究院在2012年将标准文献结构化列为重点科研项目，在此基础上，2013年又承担了国家标准化委员会标准化信息中心的《“双打”重点产品标准文献结构化内容数据库》项目，而实现标准文献结构化必须首要解决2个问题：1) 结构化标准的数据结构问题，换言之，标准结构化加工抽取出的数据是哪些数据？2) 抽取出来的数据进行加工应符合什么样的数据规范要求，方能满足与原始标准文本内容的一致性的需求。

解决第一个问题，需要建立标准结构化数据模型，并通过元数据规范来实现；

解决第二个问题，需要建立数据的著录规则，实现数据的规范性。

项目组在确定文献结构化抽取技术的基础上，对4000余项强制性国家标准的内容要素逐一进行了分析研究、归纳总结后，按UML思想确定了标准结构化数据模型，并在工程上对4000余项强制性国家标准进行了结构化数据的抽取验证，确定了结构化元数据。通过验证测试，逐步完善了加工抽取结构化数据的规范要求，形成了《标准文献结构化数据规范》。

目前，在行业内广泛使用的GB/T 22373-2008《标准文献元数据》是针对标准文献题录的元数据标准，侧重的是标准文献形式特征，而此次制定的《标准文献结构化数据规范》是关于标准内容结构化的全文元数据和指标元数据规范，希望此两个标准的结合使用，实现标准文献信息资源的统一描述，促进标准文献信息资源的开发利用。

标准文献结构化数据规范 第2部分：指标元数据

1 范围

本部分规定了标准文献结构化技术指标元数据（以下简称指标元数据）的术语和定义、属性、数据模型、数据描述以及扩展原则和方法。

本部分适用于标准文献结构化数据的处理。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 7408 数据元和交换格式信息交换 日期和时间表示法（GB/T 7408-2005，ISO 8601:2000，IDT）

GB/T 18391.3-2009 信息技术 元数据注册系统（MDR）第3部分：注册系统元模型与基本属性（ISO/IEC 11179-3:2003，IDT）

GB/T 19710-2005 地理信息元数据（ISO 19115-2003，MOD）

GB/T 26816-2011 信息资源核心元数据

DB13/T 2351.1-2016 标准文献结构化数据规范 第1部分：全文元数据

3 术语和定义

下列术语和定义适用于本文件。

3.1

信息资源 information resource

在政治、经济和社会等各领域产生和使用、具有各种载体形式的信息内容。

[GB/T 26816-2011，定义3.1]

3.2

元数据 metadata

定义和描述其他数据的数据。

[GB/T 18391.3-2009，定义3.2.18]

3.3

元数据元素 metadata element

元数据的基本单元。

注：元数据元素在元数据实体中是唯一的。

[GB/T 19710-2005，定义4.61]

3.4

元数据实体 metadata entity

一组说明信息资源相关特性的元数据元素。

注：可以包含一个或一个以上元数据实体。

[GB/T 19710-2005, 定义4.71]

3.5

核心元数据 core metadata

描述信息资源基本属性的元数据元素和元数据实体。

[GB/T 26816-2011, 定义3.5]

4 指标元数据的属性

本部分采用摘要表示的方法定义和描述标准结构化指标元数据，摘要内容包括以下八个属性：中文名称、英文名称、缩写名、定义、数据类型、值域、约束/条件、最大出现次数。

4.1 中文名称

元数据元素或元数据实体的中文名称，用第6章中各条的标题来表达。

4.2 英文名称

应符合DB13/T 2351. 1-2016中4.2的规定。

4.3 缩写名

应符合GB/T 26816-2011中4.6的规定。

4.4 定义

应符合GB/T 26816-2011中4.2的规定。

4.5 数据类型

应符合GB/T 26816-2011中4.4的规定。

4.6 值域

应符合GB/T 26816-2011中4.5的规定。

4.7 约束/条件

应符合GB/T 26816-2011中4.7的规定，约束/条件代码见DB13/T XXXX. 1-2016附录A 表A.1。

4.8 最大出现次数

应符合GB/T 26816-2011中4.8的规定。

5 指标元数据模型

标准文献结构化指标元数据模型见图1。

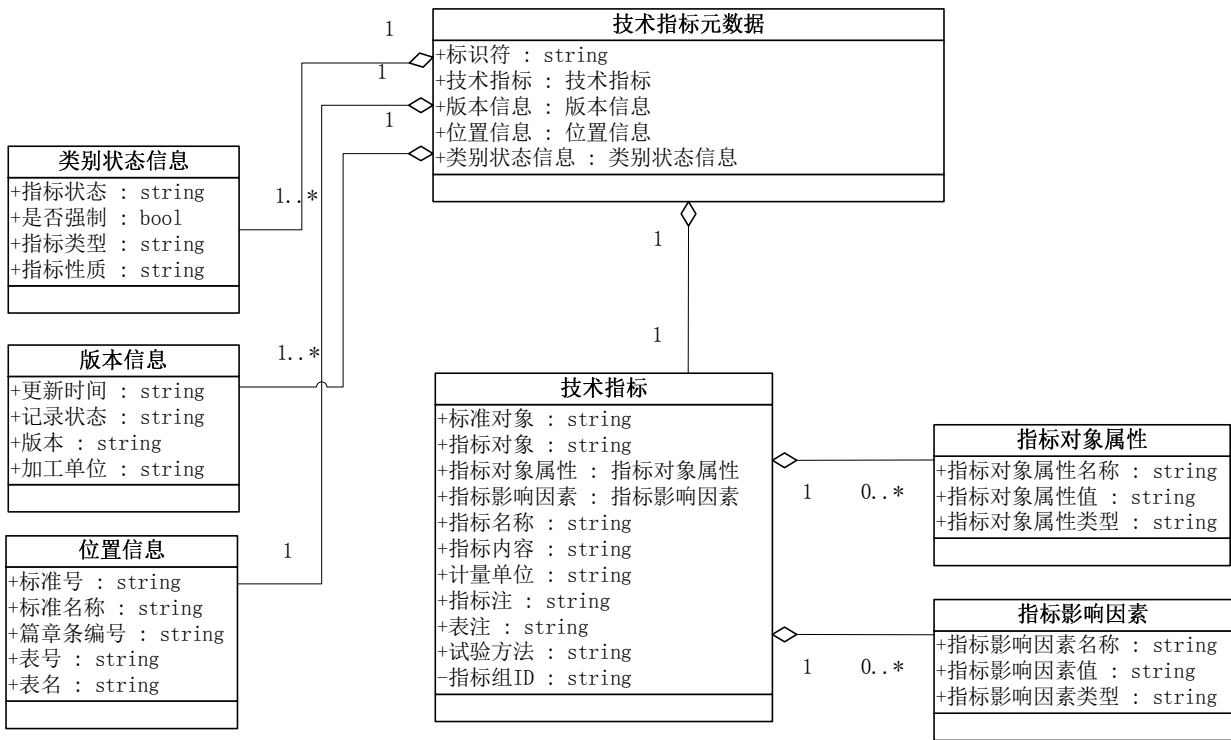


图1 指标元数据模型

6 技术指标元数据描述

6.1 技术指标元数据

中文名称:	技术指标元数据
英文名称:	Technical Indicator Metadata
缩写名:	TechIndMdata
定义:	描述结构化标准中技术指标数据的元数据。
数据类型:	复合型
约束/条件:	M
最大出现次数:	1

6.2 标识符

应符合DB13/T 2351.1-2015中6.2的规定。

6.3 技术指标

中文名称:	技术指标
英文名称:	Technical Indicator
缩写名:	TechInd

定义:	描述标准质量特性指标的一组信息。
数据类型:	复合型
约束/条件:	M
最大出现次数:	1

6.4 标准对象

中文名称:	标准对象
英文名称:	document object
缩写名:	docObj
定义:	描述一个实体的标准化活动的作用对象。
数据类型:	字符串
值域:	自由文本
约束/条件:	M
最大出现次数:	1

6.5 指标对象

中文名称:	指标对象
英文名称:	indicator object
缩写名:	indObj
定义:	描述一个实体的质量特征的作用对象。
数据类型:	字符串
值域:	自由文本
约束/条件:	M
最大出现次数:	1

6.6 指标对象属性

中文名称:	指标对象属性
英文名称:	Indicator Object Attribute
缩写名:	IndObjAttr
定义:	描述指标对象特征、特性的一组信息。
数据类型:	复合型
约束/条件:	0
最大出现次数:	N

6.7 指标对象属性名称

中文名称:	指标对象属性名称
英文名称:	indicator object attribute name
缩写名:	indObjAttrNm
定义:	描述指标对象的特性、特征的名称。如:形状、颜色、气味、优劣、用途等都是指标对象的性质。
数据类型:	字符串
值域:	自由文本
约束/条件:	M

最大出现次数: 1

6.8 指标对象属性值

中文名称: 指标对象属性值
 英文名称: indicator object attribute value
 缩写名: indObjAttrVal
 定义: 描述指标对象属性的内容。
 数据类型: 字符串
 值域: 自由文本
 约束/条件: M
 最大出现次数: 1

6.9 指标对象属性类型

中文名称: 指标对象属性类型
 英文名称: indicator object attribute type
 缩写名: indObjAttrTyp
 定义: 描述指标对象的类别。
 数据类型: 字符串
 值域: 字典
 约束/条件: 0
 最大出现次数: N

6.10 指标影响因素

中文名称: 指标影响因素
 英文名称: Indicator Influencing Factor
 缩写名: IndInfluFact
 定义: 描述与指标数据有关的一组信息。
 数据类型: 复合型
 约束/条件: 0
 最大出现次数: N

6.11 指标影响因素名称

中文名称: 指标影响因素名称
 英文名称: indicator influencing factor name
 缩写名: indInfluFactNm
 定义: 对指标数据产生有一定约束关系的因素名称。如: 环境因素、测试方法、工具、工艺、作用对象等。
 数据类型: 字符串
 值域: 自由文本
 约束/条件: M
 最大出现次数: 1

6.12 指标影响因素值

中文名称:	指标影响因素值
英文名称:	indicator influencing factor value
缩写名:	indInfluFactVal
定义:	指标影响因素的具体内容描述。
数据类型:	字符串
值域:	自由文本
约束/条件:	M
最大出现次数:	1

6.13 指标影响因素类型

中文名称:	指标影响因素类型
英文名称:	indicator influencing factor type
缩写名:	indInfluFactTyp
定义:	描述指标影响因素的类别。
数据类型:	字符串
值域:	字典
约束/条件:	0
最大出现次数:	N

6.14 指标名称

中文名称:	指标名称
英文名称:	indicator name
缩写名:	indNm
定义:	描述实体质量特征属性的技术要求名称。
数据类型:	字符串
值域:	自由文本
约束/条件:	M
最大出现次数:	1

6.15 指标内容

中文名称:	指标内容
英文名称:	indicator value
缩写名:	indVal
定义:	某项技术要求的具体内容描述。
数据类型:	字符串
值域:	自由文本
约束/条件:	M
最大出现次数:	1

6.16 计量单位

中文名称:	计量单位
英文名称:	indicator measurement unit
缩写名:	indMeasUnit

定义:	技术要求的量值单位。
数据类型:	字符串
值域:	字典
约束/条件:	0
最大出现次数:	1

6.17 指标注

中文名称:	指标注
英文名称:	indicator note
缩写名:	indNote
定义:	指标的注释。
数据类型:	字符串
值域:	自由文本
约束/条件:	0
最大出现次数:	1

6.18 表注

应符合DB13/T 2351.1-2015中6.17的规定。

6.19 试验方法

中文名称:	试验方法
英文名称:	test method
缩写名:	tstMethd
定义:	验证指标的试验方法的一组信息。
数据类型:	字符串
值域:	自由文本
约束/条件:	0
最大出现次数:	1

6.20 指标组 ID

中文名称:	指标组 ID
英文名称:	indicator group identifier
缩写名:	indGrpID
定义:	一组指标数据的唯一识别符。
数据类型:	数值型
值域:	正整数
约束/条件:	M
最大出现次数:	1

6.21 位置信息

中文名称:	位置信息
英文名称:	Location Information

缩写名:	LocInfo
定义:	指标所在标准内容中的定位信息描述。
数据类型:	复合型
约束/条件:	M
最大出现次数:	1

6.22 标准号

应符合DB13/T 2351. 1-2016中6.3的规定。

6.23 标准名称

应符合DB13/T 2351. 1-2016中6.4的规定。

6.24 篇章条编号

应符合DB13/T 2351. 1-2016中6.5的规定。

6.25 表号

应符合DB13/T 2351. 1-2016中6.14的规定。

6.26 表名

应符合DB13/T 2351. 1-2016中6.15的规定。

6.27 版本信息

中文名称:	版本信息
英文名称:	Version Information
缩写名:	VInfo
定义:	描述指标数据产生、更新的一组信息。
数据类型:	复合型
约束/条件:	M
最大出现次数:	1

6.28 更新时间

中文名称:	更新时间
英文名称:	update time
缩写名:	upTime
定义:	描述指标数据修改更新的时间。
数据类型:	日期型
值域:	按 GB/T 7408 执行, 格式为 CCYY-MM-DD hh:mm:ss
约束/条件:	M
最大出现次数:	1

6.29 记录状态

中文名称:	记录状态
英文名称:	record status

缩写名:	recStatus
定义:	关于标准文献指标元数据库中记录所处的状态（修改、删除、新增）的说明。
数据类型:	字符串
值域:	采用附录 A 表 A.1 中的名称。
约束/条件:	M
最大出现次数:	1

6.30 版本

中文名称:	版本
英文名称:	version number
缩写名:	verNo
定义:	描述指标数据更新的版本编号。
数据类型:	字符串
值域:	自由文本
约束/条件:	M
最大出现次数:	1

6.31 加工单位

中文名称:	加工单位
英文名称:	process unit
缩写名:	prsUnit
定义:	描述指标数据加工单位的信息。
数据类型:	字符串
值域:	字典
约束/条件:	M
最大出现次数:	1

6.32 类别状态信息

中文名称:	类别状态信息
英文名称:	Type State Information
缩写名:	TypStaInfo
定义:	描述技术指标状态的一组信息。
数据类型:	复合型
约束/条件:	M
最大出现次数:	1

6.33 指标状态

中文名称:	指标状态
英文名称:	indicator state
缩写名:	indSta
定义:	描述指标有效性的特征。一般分为 3 种状态：现行、作废、未实施。
数据类型:	字符串

值域:	采用附录 A 表 A.2 中的名称。
约束/条件:	M
最大出现次数:	1

6.34 是否强制

中文名称:	是否强制
英文名称:	whether mandatory
缩写名:	whethMand
定义:	描述指标所在篇章条是否为强制性条款。
数据类型:	布尔型
值域:	“0”代表推荐,“1”代表强制
约束/条件:	M
最大出现次数:	1

6.35 指标类型

中文名称:	指标类型
英文名称:	indicator type
缩写名:	indTyp
定义:	指标的分类编码(体系)。
数据类型:	字符串
值域:	字典
约束/条件:	0
最大出现次数:	1

6.36 指标性质

中文名称:	指标性质
英文名称:	indicator quality
缩写名:	indQlty
定义:	描述指标内容的类型是定性亦或定量。
数据类型:	字符串
值域:	采用附录 A 表 A.3 中的名称。
约束/条件:	M
最大出现次数:	1

7 指标元数据字典描述

标准文献结构化指标元数据字典描述见附录B,表B.1中带“*”的为核心元数据。

8 指标元数据扩展原则和方法

标准对象的多样性决定标准内容的多样性,标准文献结构化指标元数据需要扩充。扩充的原则与方法应符合GB/T 26816-2011第6章的规定。

附 录 A
(规范性附录)
标准文献指标元数据代码集

A.1 记录状态代码

记录状态代码见表A.1。

表A.1 记录状态代码

代码	名称	说明
A	修改	A-Amend
D	删除	D-Delete
N	新增	N-New

A.2 指标状态代码

指标状态代码见表A.2。

表A.2 指标状态代码

代码	名称	说明
X	现行	
W	作废	
R	未实施	

A.3 指标性质代码

指标性质代码见表A.3。

表A.3 指标性质代码

代码	名称	说明
QLTY	定性	QLTY 是quality的缩略语
QTY	定量	QTY 是quantity的缩略语

附 录 B
(资料性附录)
标准文献结构化指标元数据-字典描述

B.1 标准文献结构化指标元数据的字典描述见表B.1。

表B.1 标准文献结构化指标元数据-字典描述

序号	中文名	英文名	缩略语	定义	数据类型	值域	约束	最大出现 次数	核心元 数据
0	技术指标元数据	Technical Indicator Metadata	TechIndMdata	描述结构化标准中技术指标数据的元数据	复合型		M	1	
1	标识符	identifier	ID	名称空间中对象的唯一标识符。	字符串	UUID	M	1	*
2	技术指标	Technical Indicator	TechInd	描述标准质量特性指标的一组信息。	复合型		M	1	
2.1	标准对象	document object	docObj	描述一个实体的标准化活动的作用对象。	字符串	自由文本	M	1	*
2.2	指标对象	indicator object	indObj	描述一个实体的质量特征的作用对象。	字符串	自由文本	M	1	*
2.3	指标对象属性	Indicator Object Attribute	IndObjAttr	描述指标对象特征、特性的一组信息。	复合型		0	N	

表 B.1 (续)

序号	中文名	英文名	缩略语	定义	数据类型	值域	约束	最大出现次数	核心元数据
2.3.1	指标对象属性名称	indicator object attribute name	indObjAttrNm	描述指标对象的特性、特征的名称。如：形状、颜色、气味、优劣、用途等都是指标对象的性质。	字符串	自由文本	M	1	*
2.3.2	指标对象属性值	indicator object attribute value	indObjAttrVal	描述指标对象属性的内容。	字符串	自由文本	M	1	*
2.3.3	指标对象属性类型	indicator object attribute type	indObjAttrTyp	描述指标对象的类别。	字符串	字典	0	N	
2.4	指标影响因素	Indicator Influencing Factor	IndInfluFact	描述与指标数据有关的一组信息。	复合型		0	N	
2.4.1	指标影响因素名称	indicator influencing factor name	indInfluFactNm	对指标数据产生有一定约束关系的因素名称。如：环境因素、测试方法、工具、工艺、作用对象等。	字符串	自由文本	M	1	*
2.4.2	指标影响因素值	indicator influencing factor value	indInfluFactVal	指标影响因素的具体内容描述。	字符串	自由文本	M	1	*
2.4.3	指标影响因素类型	indicator influencing factor type	indInfluFactTyp	描述指标影响因素的类别。	字符串	字典	0	N	
2.5	指标名称	indicator name	indNm	描述实体质量特征属性的技术要求名称。	字符串	自由文本	M	1	*
2.6	指标内容	indicator value	indVal	某项技术要求的具体内容描述。	字符串	自由文本	M	1	*

表 B.1 (续)

序号	中文名	英文名	缩略语	定义	数据类型	值域	约束	最大出现次数	核心元数据
2.7	计量单位	indicator measurement unit	indMeasUnit	技术要求的量值单位。	字符串	字典	0	1	
2.8	指标注	indicator note	indNote	指标的注释。	字符串	自由文本	0	1	*
2.9	表注	table note	tabNote	标准内容中表格的注释。	字符串	自由文本	0	1	
2.10	试验方法	test method	tstMethd	验证指标的试验方法的一组信息。	字符串	自由文本	0	1	*
2.11	指标组 ID	indicator group identifier	indGrpID	一组指标数据的唯一识别符。	数值型	正整数	M	1	
3	位置信息	Location Information	LocInfo	指标所在标准内容中的定位信息描述。	复合型		M	1	
3.1	标准号	document reference	docRef	由有关标准化机构给定的用于唯一识别某一标准的注册号或登记号,标准号由标准代号、顺序号、发布年份及有关标识符组成。	字符串	标准代号+1 个空格+顺序号+1 个连字符“-”+4 位发布年份	M	1	*
3.2	标准名称	document name in Chinese	docNmCN	标准的中文名称。	字符串	自由文本	M	1	*
3.3	篇章条编号	chapter number	chapNo	标准内容中篇章条按一定规则的顺序编号(篇章条编号定义说明见 DB13/T XXXX.1-2015 中附录 B.1.1)。	字符串	应符合 DB13/T XXXX.1-2015 中附录 B.1.2 的规定。	M	1	
3.4	表号	table number	tabNo	标准内容中表格按一定规则的顺序编号。(表号定义说明见 DB13/T XXXX.1-2015 中附录 B.2.1)。	字符串	应符合 DB13/T XXXX.1-2015 中附录 B.2.2 的规定。	0	1	

表 B.1 (续)

序号	中文名	英文名	缩略语	定义	数据类型	值域	约束	最大出现次数	核心元数据
3.5	表名	table name	tabNm	描述表格主题的名称。	字符串	自由文本	M	1	
4	版本信息	Version Information	VInfo	描述指标数据产生、更新的一组信息。	复合型		M	1	
4.1	更新时间	update time	upTime	描述指标数据修改更新的时间。	日期型	按 GB/T 7408 执行，格式为 CCYY-MM-DD hh:mm:ss	M	1	*
4.2	记录状态	record status	recStatus	关于标准文献指标元数据库中记录所处的状态（修改、删除、新增）的说明。	字符串	采用表 A.1 《记录状态代码》中的名称。	M	1	*
4.3	版本	version number	verNo	描述指标数据更新的版本号。	字符串	自由文本	M	1	*
4.4	加工单位	process unit	prsUnit	描述指标数据加工单位的信息。	字符串	字典	M	1	
5	类别状态信息	Type State Information	TypStaInfo	描述技术指标状态的一组信息。	复合型		M	1	
5.1	指标状态	indicator state	indSta	描述指标有效性的特征。一般分为 3 种状态：现行、作废、未实施。	字符串	采用表 A.2 《指标状态代码》中的名称。	M	1	*
5.2	是否强制	whether mandatory	whethMand	描述指标所在篇章条是否为强制性条款。	布尔型	0 代表推荐，“1”代表强制	M	1	*
5.3	指标类型	indicator type	indTyp	指标的分类编码（体系）。	字符串	字典	0	1	

表 B.1（续）

序号	中文名	英文名	缩略语	定义	数据类型	值域	约束	最大出现 次数	核心元 数据
5.4	指标性质	indicator quality	indQlty	描述指标内容的类型是定性亦或定量。	字符串	采用表 A.3 《指标性质 代码》中的名 称。	M	1	

参 考 文 献

- [1] GB/T 1.1-2009 标准化工作导则 第1部分:标准的结构和编写
 - [2] GB/T 22373-2008 标准文献元数据
 - [3] GB/T 24639-2009 元数据的XML Schema置标规则
 - [4] GB/T 25100-2010 信息与文献 都柏林核心元数据元素集 (ISO 15836-2009, MOD)
 - [5] 崔运鹏. 基于本体论的农业知识管理关键技术研究. 北京: 中国农业科学技术出版社, 2009
 - [6] 高瓦特斯, 利维森. 正则表达式经典实例. 郭耀, 译. 北京: 人民邮电出版社, 2010
 - [7] 胡荷芬, 高斐. UML面向对象分析与设计教程. 北京: 清华大学出版社, 2012
 - [8] 李景. 本体理论在文献检索系统中的应用研究. 北京: 北京图书馆出版社, 2005
 - [9] 李景, 孟宪学, 苏晓路. 领域本体的构建方法. 北京: 中国农业科学技术出版社, 2009
 - [10] 施穆勒. UML基础、案例与应用 (第三版). 李虎, 赵龙刚, 译. 北京: 人民邮电出版社, 2012
-