

ICS 35.240.30
L 67



中华人民共和国国家标准

GB/T 20092—2006

中文新闻信息置标语言

Chinese news markup language

2006-01-05 发布

2006-05-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

目 次

前言	VII
引言	VIII
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 规则和约定	2
5.1 拼写规则	2
5.2 元素表达及描述要求	2
5.3 XML 组件的命名	3
5.4 CNML XML Schema 的版本和变更约定	3
6 CNML 体系结构	3
6.1 概述	3
6.2 总体结构	3
6.3 文档层结构	4
6.4 稿件层结构	5
6.5 内容项层结构	7
6.6 元数据结构	7
6.7 受控词表结构	9
6.8 关系结构	9
7 CNML 公共结构和基本数据类型	10
7.1 目标名字空间声明	10
7.2 基本简单类型	10
7.2.1 XPath 字符串类型(xPathType)	11
7.2.2 基础日期时间类型(baseTimePointType)	11
7.2.3 日期时间类型(timePointType)	12
7.2.4 基础时间段类型(baseDurationType)	13
7.2.5 时间段类型(durationType)	13
7.3 基本关系名类型	13
7.3.1 关系名基类型(baseRelationNameType)	14
7.3.2 稿件关系名基类型(itemRelationNameType)	14
7.3.3 内容项关系名基类型(contentItemRelationNameType)	17
7.3.4 受控词关系名基类型(topicRelationNameType)	20
7.4 复杂元数据项类型	21
7.4.1 参与方类型(PartyType)	21
7.4.2 人物类型(PersonType)	32
7.4.3 机构类型(OrganizationType)	46
7.4.4 地点类型(PlaceType)	56

7.4.5	时间元数据项类型(DateTimeType)	67
7.5	公共结构类型	67
7.5.1	公共属性组(commonGroup)	67
7.5.2	类型属性组(kindGroup)	68
7.5.3	计数模式属性组(sequenceModeGroup)	69
7.5.4	引用属性组(referenceGroup)	69
7.5.5	本地字符串类型(LocalStringType)	70
7.5.6	引用类型(ReferenceType)	71
7.5.7	电子地址类型(ElectronicAddressType)	71
7.5.8	人名类型(PersonNameType)	74
7.5.9	名字部件类型(NameComponentType)	79
7.5.10	地点坐标类型(GeographicPointType)	80
7.5.11	色别元素(⟨ColorType⟩)	82
7.6	元数据相关的 Schema 定义	82
7.6.1	不同类型元数据项的 Schema 定义	82
7.6.2	简单元数据模型	86
7.6.3	复杂元数据模型	87
8	CNML 的关系机制	93
8.1	关系定义	93
8.1.1	概述	93
8.1.2	关系组概念的复杂类型(RelationGroupType)	94
8.1.3	关系类型(RelationType)	94
8.1.4	关系节点类型(RelationNodeType)	95
8.1.5	关系源元素(⟨Source⟩)	96
8.1.6	关系目标元素(⟨Target⟩)	96
8.1.7	角色元素(⟨Role⟩)	97
8.1.8	角色集元素(⟨RoleSet⟩)	98
8.2	稿件层关系的定义	98
8.2.1	定义	98
8.2.2	稿件关系组类型(ItemRelationGroupType)	99
8.2.3	稿件关系元素(⟨ItemRelationGroupType/Relation⟩)	100
8.2.4	稿件关系类型(ItemRelationType)	101
8.2.5	稿件关系元数据元素(⟨ItemRelationType/Notes⟩)	102
8.2.6	稿件关系源元素(⟨ItemRelationType/Source⟩)	102
8.2.7	稿件关系目标元素(⟨ItemRelationType/Target⟩)	102
8.3	内容项关系	103
8.3.1	内容项关系组类型(ContentItemRelationGroupType)	103
8.3.2	内容项关系元素(⟨ContentItemRelationGroupType/Relation⟩)	104
8.3.3	内容项关系类型(ContentItemRelationType)	105
8.3.4	内容项关系元数据元素(⟨ContentItemRelationType/Notes⟩)	106
8.3.5	内容项关系源元素(⟨ContentItemRelationType/Source⟩)	106
8.3.6	内容项关系目标元素(⟨ContentItemRelationType/Target⟩)	107
8.3.7	内容关系组类型(ContentsRelationGroupType)	107

8.4 受控词关系	108
8.4.1 受控词关系组类型 (TopicRelationGroupType)	108
8.4.2 受控词关系元素 (<TopicRelationGroupType/Relation>)	109
8.4.3 受控词关系类型 (TopicRelationType)	110
8.4.4 受控词关系元数据元素 (<TopicRelationType/Notes>)	110
8.4.5 受控词关系源元素 (<TopicRelationType/Source>)	111
8.4.6 受控词关系目的元素 (<TopicRelationType/Target>)	111
8.4.7 受控词关系名类型 (topicRelationNameType)	112
9 受控词描述机制	112
9.1 概述和分析	112
9.2 受控词表的建模	113
9.2.1 受控词表稿件建模	113
9.2.2 受控词表内容项建模	113
9.2.3 受控词的建模	113
9.2.4 具有嵌套结构的受控词表的建模	113
9.2.5 受控词表的 schema 定义	113
9.3 受控词引用的建模	126
9.3.1 受控词引用的建模	126
9.3.2 受控词引用的 schema 定义	127
9.4 受控词表的例子	129
9.4.1 简单的单层受控词表的示例	129
9.4.2 具有层次结构的受控词表稿件的示例	131
10 数字签名	132
10.1 概述	132
10.2 元数据描述	133
10.2.1 数字签名容器类型 (SignatureContentsType)	133
10.2.2 数字签名稿件类型 (SignatureItemType)	134
10.2.3 数字签名稿件内容容器元素 (<SignatureItemType/Contents>)	135
10.3 数字签名样例	136
10.3.1 对稿件进行签名	136
10.3.2 对稿件内容项进行签名	137
10.3.3 对新闻的元数据项进行签名	139
11 CNML 文档层结构	141
11.1 概述	141
11.2 CNML 元素 (<CNML>)	141
11.3 信封结构	143
11.3.1 概述	143
11.3.2 信封元素 (<Envelop>)	144
11.3.3 信封类型 (EnvelopType)	144
11.3.4 信封结构应用示例	170
11.4 稿件容器	177
11.4.1 稿件容器元素 (<Items>)	177
11.4.2 CNML 标准中可实例化的稿件类型	179

11.4.3	〈Items〉的完整应用示例	186
11.5	稿件关系容器	190
11.5.1	稿件关系容器元素 (〈Relations〉)	190
11.5.2	〈Relations〉的完整应用示例	191
12	CNML 稿件层	192
12.1	稿件层定义	192
12.1.1	稿件概念的 Schema 定义	192
12.1.2	操作概念的 Schema 定义	201
12.2	标准元数据组	207
12.2.1	概述	207
12.2.2	元数据容器元素 (〈MetaInfo〉)	208
12.2.3	管理型元数据组	209
12.2.4	描述型元数据组	235
12.2.5	权利型元数据组	277
13	CNML 内容项层	296
13.1	CNML 内容项层结构	296
13.1.1	描述型稿件相关内容项类型 ContentItemType	296
13.1.2	具备对应物理文件的描述型稿件相关内容项元素〈ContentItem〉	298
13.2	内容项层实体内容	300
13.2.1	概述	300
13.2.2	内容项类型 ContentItemType 的扩展类	300
13.2.3	数据内容元素	310
13.3	内容项层的元数据	311
13.3.1	内容项层元数据结构	311
13.3.2	内容项层与稿件层可共有的元数据组	312
13.3.3	内容项特有的物理元数据组	315
13.3.4	物理特性元数据组类型 CharacteristicMetaGroupType 的扩展类型	328
14	CNML 内联智能标记	376
14.1	概述	376
14.2	内联智能标记的命名空间	376
14.3	规范说明	376
14.4	内联智能标记类型 (cim: IntelligentMarkupType)	376
14.4.1	分类标记元素 (〈cim: Classifier〉)	378
14.4.2	事件标记元素 (〈cim: Event〉)	379
14.4.3	职务标记元素 (〈cim: Function〉)	380
14.4.4	行业标记元素 (〈cim: Industry〉)	381
14.4.5	地点标记元素 (〈cim: Location〉)	382
14.4.6	对象名标记元素 (〈cim: ObjectTitle〉)	384
14.4.7	机构标记元素 (〈cim: Organization〉)	385
14.4.8	人物标记元素 (〈cim: Person〉)	386
附录 A (资料性附录)	CNML 应用实例	388
A.1	CNML 完整用例	388
A.2	CNML 稿件的轻量级实例	405

附录 B (资料性附录) CNML 标准元素中英文对照表	407
附录 C (资料性附录) CNML 类继承总图	413
附录 D (资料性附录) 受控词表	414
D.1 受控词表的完善和更新	414
D.2 CNML XML Schema 标准受控词表	414
D.2.1 CNML 标准推荐受控词表总列表	414
D.2.2 发稿优先级词表 (topicschema, Priority. xml)	415
D.2.3 紧急度词表 (topicschema, Urgency. xml)	415
D.2.4 重要性词表 (topicschema, Importance. xml)	415
D.2.5 密级词表 (topicschema, ConfidentialLevel. xml)	415
D.2.6 受众词表 (topicschema, Audience. xml)	416
D.2.7 媒体类型词表 (topicschema, MediaType. xml)	416
D.2.8 文件格式词表 (topicschema, Format. xml)	416
D.2.9 MIME 类型词表 (topicschema, MIMEType. xml)	418
D.2.10 编码方式词表 (topicschema, Encoding. xml)	420
D.2.11 符号表示法词表 (topicschema, Notation. xml)	420
附录 E (资料性附录) CNML XML Schema	422
附录 F (资料性附录) CNML RDF Schema	494
参考文献	560
索引	561

前 言

本标准的附录 A、附录 B、附录 C、附录 D、附录 E、附录 F 为资料性附录。

本标准由新华通讯社提出。

本标准由全国信息技术标准化技术委员会归口。

本标准起草单位：新华通讯社、中国新闻技术工作者联合会、中国新闻社、人民日报、解放军报、中国网、中央人民广播电台、中国国际广播电台、中央电视台、北大方正电子有限公司、清华大学、北京拓尔思信息技术有限公司、中科软科技股份有限公司、中科大洋科技发展股份有限公司。

本标准主要起草人：武国卫、刁毅刚、曹学军、李涓子、徐鹏、丁峰、黄菁、金昉霞、林道进、周于文、王建平、鲁倩、高沁、李庆国、陈宁、郑光伟、曹晓玲、王杰中、孙丽华、徐如、刘娜、刘颖、王仲。

引 言

本标准是中文新闻信息置标语言研制单位根据新闻行业数字化、网络化的发展要求,结合新闻数据跨行业交换共享和媒体数字资产管理应用的需要,在总结了国内通讯社、报刊、新闻网站、电视台、广播电台等部门及媒体应用软件开发和服务商关于中文新闻交换和共享的经验基础上,参考国际 IPTC (国际新闻电信理事会)NewsML 等元数据标准后提出的。

本标准规定了新闻信息的结构框架,制定了统一的新闻元数据项,为实现新闻信息交换和应用的标准化管理奠定了基础。

中文新闻信息置标语言

1 范围

本标准规定了新闻行业中文新闻信息的元数据和结构,以实现新闻信息的内容描述、交换和再利用。

本标准适用于新闻信息电子数据的“采集、编辑、生成、发布、存储、检索、评估反馈”等新闻信息生命周期的各个环节。本标准也适用于新闻行业同非新闻单位之间的数据交换。

本标准的使用对象包括报刊、广播、电视、通讯社、新闻网站等新闻内容提供商及媒体应用软件开发商和服务商。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准,然而,鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本标准。

GB/T 2659 世界各国和地区名称代码(GB/T 2659—2000,eqv ISO 3166-1:1997)

GB/T 4880.2—2000 语种名称代码 第2部分:3字母代码(eqv ISO 639-2:1998)

GB/T 4881—1985 中国语种代码

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

GB 18030 信息技术 信息交换用汉字编码字符集 基本集的扩充

GB/T 18793—2002 信息技术 可扩展置标语言(XML)1.0(W3C RFC-xml:1998,NEQ)

GB/T 20093—2006 中文新闻信息分类与代码

3 术语和定义

GB/T 18793—2002 中确立的以及下列术语和定义适用于本标准。

3.1

元元数据 meta metadata

定义和描述元数据的数据。

3.2

受控词表 controlled vocabulary

从专业的角度将同一范畴的主题进行归类后形成的词表,并允许以受控的方式进行词条修改和扩展。

3.3

CNML 文档 CNML document

符合本标准的 XML 实例文档。

3.4

稿件 item

文档中包含的可管理业务对象,体现为在 Items 容器元素中包含的各个 Item 项,它一般有完整的元数据和内容,表示一个完整的新闻信息或其他定义信息。