

中文新闻信息置标语言标准课题组 编

中文新闻信息 置标语言 使用指南

00101
0010010110
001001101001
001000101
0010
00101
0010010110
001000101

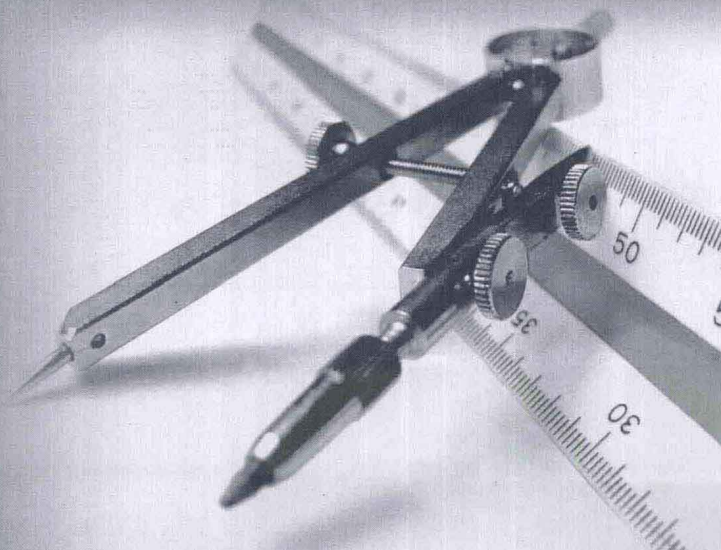
新华出版社

中文新闻信息置标语言标准课题组 编

中文新闻信息 置标语言 使用指南

00101
0010010110
001001101001
01000101
0010
00101
0010010110
01000101

新华出版社



图书在版编目(CIP)数据

中文新闻信息置标语言使用指南/中文新闻信息置标
语言标准课题组编. —北京:新华出版社,2008. 11

ISBN 978 — 7 — 5011 — 8571 — 9

I. 中… II. 中… III. 汉语—新闻语言 IV. G210

中国版本图书馆 CIP 数据核字(2008)第 168550 号

中文新闻信息置标语言使用指南

责任编辑:韩 芳 郭信峰

封面设计:冯 锴

出版发行:新华出版社

地 址:北京石景山区京原路 8 号

网 址:<http://press.xinhuanet.com> <http://www.xinhupub.com>

邮 编:100040

经 销:新华书店

印 刷:北京新魏印刷厂

开 本:710 × 1000mm 1/16

印 张:22.75

字 数:330 千字

版 次:2008 年 11 月第一版

印 次:2008 年 11 月第一次印刷

书 号:ISBN 978 — 7 — 5011 — 8571 — 9

定 价:45.00 元

本社购书热线:(010)63077122

中国新闻书店电话:(010)63072012

图书如有印装问题,请与印刷厂联系调换 电话:(010)89201322



(上) 2008年3月30日，新华社副社长、全国中文新闻信息标准化技术委员会主任委员鲁炜与国际新闻电信理事会(IPTC)董事会主席斯蒂芬·格里特先生亲切交谈。

(下) 2008年4月3日，国际新闻电信理事会(IPTC)与新华社技术局技术人员交流新闻信息技术标准的推广和应用情况。图为IPTC成员聆听技术交流讲座。



2006年5月25日，国家标准委、国家质检总局和新华社联合发布两项中文新闻信息技术国家标准。国家质检总局局长李长江、新华社社长田聪明和国家标准委主任刘平均出席会议并讲话。

(上)国家标准委主任刘平均在新闻发布会上讲话。

(下)新华社社长田聪明在新闻发布会上讲话。



（上）2004年10月23日，“中文新闻信息技术标准”研制课题组召开工作会议，项目组组长鲁炜和来自各新闻媒体的课题组成员共议新闻信息标准化大计。

（下）2005年11月7日，国家“十五”重大科技专项“中文新闻信息技术标准”通过国家标准化管理委员会组织的国家标准审查验收。



（上）2004年5月28日，新华社人员在新华社香港分社和国际新闻标准化组织IPTC（国际新闻电信理事会）理事会成员进行交流，探讨加入IPTC事宜。

（下）2004年9月12日至19日，中文新闻信息置标语言标准课题组赴上海、广州、济南等地进行考察调研，图为课题组在广州日报社调研。

序

“中文新闻信息技术标准”是“十五”期间国家重大科技专项“重要技术标准研究”的重点支持项目，“中文新闻信息技术标准”经过两年的艰苦研制，终于通过国家标准委的审查验收，形成《中文新闻信息置标语言》和《中文新闻信息分类与代码》两部具有我国自主知识产权的推荐性国家标准。它的研制成功，填补了我国乃至全球华语地区新闻信息技术领域标准的空白，是自主创新的成果，是我国新闻技术发展史上的里程碑。这是值得新闻行业庆贺的大事。

“中文新闻信息技术标准”是由新华通讯社牵头，会同国务院新闻办、新闻出版总署、国家广播电影电视总局、人民日报社、光明日报社、经济日报社、中央人民广播电台、中央电视台、中国国际广播电台、解放军报社、中国新闻社、中国新闻技术工作者联合会以及中国网、北大方正电子有限公司、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、清华大学等国内重要新闻单位、研究机构和企业联合研制。全国新闻单位联合起来帮助我们做这件事情，其重要原因就是新闻事业的发展需要有一个统一的标准，没有统一的新闻信息技术标准是不能适应当今时代发展需要的，标准对促进新闻信息的整合与利用、交换与共享、实现中文新闻信息的全球化并提高我国新闻科技产业的核心竞争力有着举足轻重的意义和作用。

研制标准归根到底是为了应用，离开了应用，标准没有任何意义。我们应带头率先使用起来，并且要面向全社会大力宣传、推广这两部国家标准。要让越来越多的人了解它，使用它，真正做到“谁用谁得益”。用实践来检验标准的优劣，以实际效用证明标准的价值。用了能得益，将会促使更多的人认可标准、使用标准，由此推动整个新闻行业的标准化，而标准化是当今时代增强行业核心竞争力的重要武器。

我们不是为了有标准而研制标准，站在时代以及新闻事业发展的高度看标准，我们的目标是：要让中文新闻信息技术标准走向全国、走向全球华语地区，乃至走向全世界。

增强标准意识，大力推广和使用标准，促进整个中文新闻行业的技术进步，是所有华文媒体的责任。这本由多名标准研制人员共同参与编制的《中文新闻信息置标语言使用指南》，提供给人们一个了解标准、学习标准，以及掌握标准基本使用方法的便捷途径。这是我们为推广《中文新闻信息置标语言》标准所做努力的起点。

“万事开头难”，《指南》能否为广大使用者正确使用标准提供有效的帮助需要时间和实践的检验。标准的使用和推广，任重道远，我们要做的工作还有很多，实际应用标准的过程中还将不断涌现新的难点和问题，这都需要我们认真思考、逐一解决，标准也将在此基础上不断完善。我们期待着社会各界的批评和帮助。

希望更多的新闻技术工作者通过该《指南》认识《中文新闻信息置标语言》标准，理解该标准的基本概念、主要内容，掌握标准的使用，明确标准应用的意义，更祝愿中文新闻信息技术标准早日遍及全世界每一个有华语新闻的地方！

新华社副社长
全国中文新闻信息标准化技术委员会主任委员



2008年4月

引言

随着信息技术特别是互联网技术的迅猛发展和经济全球化进程的加快,信息共享逐渐成为人们的普遍追求。同时,新闻的传播方式和业务模式也有了新的需求,各种新的媒体形式不断出现,新闻信息的表现形式和发布渠道日趋多样化,信息资源作为一种财富越来越被人认识,媒体资源整合的趋势也逐步明朗,多媒体已经成为新闻传媒发展的新趋势,人们对信息检索的准确性、信息资源的利用率也有了更高的要求,新闻用户的主体需求逐渐从阅读新闻转向咨询服务、数据深入挖掘和多元化的个性化服务。这些都反映了新闻媒体行业的发展趋势和新的变化。

面对新的趋势和变化,国际上众多新闻机构参加的 IPTC 组织已经推出了适应网络 and 多媒体新闻的多项标准,路透社等许多国际新闻媒体都把它作为战略性措施加以采用,对其新闻业务的发展起到了巨大的推动作用。而我国尚没有一套既与国际接轨、又具中文特色、跨媒体的全国统一的中文新闻信息技术标准。如何能够有效组织、管理、存储和发布海量多媒体新闻信息,实现中文新闻信息在华语媒体乃至更大范围内的交流和共享,提高新闻宣传报道质量,拓展新闻信息服务的新领域,提高新闻信息的交换能力,乃至进行更深层次的信息挖掘,为广大受众提供更加快捷、丰富、优质的新闻信息服务,摆在了我国新闻技术工作者的面前。

《中文新闻信息置标语言》——CNML (Chinese News Markup Language)正是在这样一个背景下在新华社、中国新闻技联等单位的积极努力下,由科技部批准立项的国家“十五”重大技术标准研制项目——“中文新闻信息技术标准研制”项目的子项目,由多家新闻单位和科研机构联合研制完成,并批准作为国家推荐性标准于 2006 年 5 月 1 日起在全国实施。它的研制成功,填补了我国乃至全球华语地区新闻信息技术领域标准的空白,是自主创新的成果,也是我国新闻技术发展史上新的里程碑。不仅对我国新闻行业的标准化、规范化发展开辟了一条崭新的道路,也为我国新闻信息技术同国际接轨奠定了基础。

《中文新闻信息置标语言》全面支持多媒体新闻处理的各个环节,创新性地提出了具有三层结构的概念模型,可方便地表达各种不同类型的内容实体和内容

组合的新闻信息稿件的结构,标准本身的定义以 XML Schema 形式为主,并借鉴 RDF 的数据建模思想,提供了基于 XML Schema 语法的融合语义特征描述功能的独特设计。标准的概念模型中的元数据及特有的关系描述机制简化了新闻信息的表示,是一个具有前瞻性、专业性、可扩展性和跨媒体的中文新闻信息技术标准。

本《指南》编制的主要目的是帮助读者解决如何从应用的角度对《中文新闻信息置标语言》的重点内容有更好的把握。在整体结构和写作风格上,通过系统的介绍和重点内容的提炼,加上丰富的应用实例,希望帮助更多的新闻技术工作者理解 CNML 标准的基本概念、主要功能,掌握标准使用方法,明确标准应用的实际意义,推动 CNML 标准在新闻媒体领域技术建设中得到更好的应用,实现新闻信息的标准化和规范化,为我国新闻媒体的信息化服务,对中文新闻信息资源的交换、整合、资源共享和挖掘起到积极的促进作用,推动多媒体新闻在全球华语地区新闻行业的应用。

参与本书编制的人员都是《中文新闻信息置标语言》标准核心研制组的成员,他们对标准的设计思想,技术路线都有深入的研究,对标准的内容理解准确、透彻,对标准的应用有独到的见解。参与编制的人员有武国卫(第1章、第5章),丁峰、李涓子(第2章、第8章、第9章、第11章),曹学军(第3章、第6章),刁毅刚(第4章、第10章),黄菁(第7章)。全书由黄菁、丁峰两位同志负责统稿,武国卫同志负责审定。

在本书的编制过程中,还得到了新华社技术局有关领导的大力支持,有关专家对本书的编写提出了宝贵意见,在此表示衷心的感谢。由于时间仓促,加之水平和经验有限,本书难免有疏漏和不足,但我们希望这种付出和执着是值得的,能给我们新闻界的标准应用带来切实的帮助,欢迎广大读者予以斧正。

编 者

2008 年 4 月

目 录

序	
引言	
第 1 章 CNML 概述	1
1.1 什么是 CNML	1
1.2 CNML 的开发背景	1
1.3 CNML 的设计思想	2
1.4 CNML 的概念模型	3
1.5 CNML 的体系结构	5
1.5.1 CNML 主体框架结构	5
1.5.2 CNML 文档层的构成	7
1.5.3 稿件层结构	9
1.5.4 内容项层结构	10
1.5.5 元数据结构	11
1.5.6 受控词表结构	13
1.5.7 关系结构	14
1.6 CNML 中信息实体的元数据	
描述	15
1.7 CNML 中 XML 组件的命名	
约定	16
第 2 章 XML Schema 基础	17
2.1 XML Schema 概述	17
2.2 XML Schema 编辑器中的主要	
图标	21
2.2.1 简单实例	21
2.2.2 XMLSPY 中的图标	22
2.3 XML Schema 文件的结构	24
2.4 XML Schema 中的元素与属性	25
2.4.1 XML Schema 中的顶层	
组件	25
2.4.2 粒度 particle	26
2.4.3 粒子	26
2.4.4 元素 element	28
2.4.5 属性和属性组	31
2.4.6 组 group	33
2.5 XML Schema 中的类型	34
2.5.1 简单类型	34
2.5.2 复杂类型	38
第 3 章 CNML 基础数据类型	45
3.1 通用属性组	45
3.2 类型属性组	45
3.3 引用属性组	46
3.4 日期时间数据类型	46
3.5 人物、人物组、机构数据类型	47
3.5.1 参与方类型	47
3.5.2 人物类型	47
3.5.3 人物组类型	49
3.5.4 机构类型	51
3.6 地点数据类型	53
3.6.1 地名	53
3.6.2 地理位置	53
3.6.3 天体	54
3.6.4 国家	54
3.6.5 地区	54
3.6.6 行政区划	54
3.6.7 邮政地址	55
第 4 章 CNML 文档	56
4.1 新闻文档定义的演变	56
4.1.1 新闻文档定义的转变	56
4.1.2 CNML 新闻文档的定义	56

4.2 CNML 文档的技术特色分析	57	5.3.2 传输标识 (<TransmissionID>)	90
4.2.1 CNML 文档的数据模型	57	5.3.3 系统标识 (<SystemId>)	90
4.2.2 CNML 文档的元素和属性 描述	59	5.3.4 发送时间 (<TransferTime>)	90
4.3 CNML 文档在新闻业务中的 含义及应用范围	59	5.3.5 定时发布时间 (<ScheduleTime>)	90
4.3.1 通过 CNML 文档组织多媒体 新闻	59	5.3.6 发送方 (<SentFrom>) ...	91
4.3.2 运用 CNML 文档实现供稿 标准化	61	5.3.7 送达方信息 (<SentTo>)	92
4.3.3 运用 CNML 文档实现新闻 业务管理自动化	63	5.3.8 发稿优先级元素 (<Priority>)	98
4.3.4 运用 CNML 文档实现对 新闻内容实施保护	66	5.3.9 服务信息元素 (<Services>)	99
4.3.5 运用 CNML 文档支持数据 挖掘业务	68	5.3.10 产品信息元素 (<Products>)	100
4.3.6 通过对多篇稿件的封装实现 新闻报道组织指挥的意志 ...	71	5.3.11 产品元素 (<Product>)	101
4.3.7 通过对多篇稿件的封装实现 多种新闻业务一体化	75	5.4 信封结构应用示例	105
4.3.8 CNML 文档作为元数据的 保值增值容器	77	5.4.1 无产品等展示信息的简单 发稿示例	105
第 5 章 CNML 文档中的通讯过程 描述	80	5.4.2 单地址发专稿示例	107
5.1 通讯过程描述信息的建模	80	5.4.3 单地址发通稿示例	109
5.1.1 通讯过程分析	80	5.4.4 组地址发通稿示例	111
5.1.2 传送过程控制信息	80	5.4.5 单地址、组地址混合发稿 示例	113
5.1.3 被传送的新闻稿件所属 产品和服务的描述	81	第 6 章 CNML 稿件	116
5.1.4 接收过程相关验证信息	82	6.1 什么是 CNML 稿件	116
5.1.5 传送过程统计信息	83	6.2 稿件基本结构	116
5.1.6 通讯过程描述信息的 建模	83	6.2.1 稿件基类型 (BaseItemType)	116
5.2 信封的概念	86	6.2.2 稿件类型体系	124
5.3 信封结构 (<Envelop>)	87	6.2.3 文档和稿件结构关系	128
5.3.1 用户扩展元数据组元素 (<Notes>)	89	6.3 稿件元数据信息	129
		6.3.1 元数据容器类型 (MetaInfoType)	129

6.3.2	元数据组 (MetaInfoType/ MetaGroup)	130
6.3.3	元数据组基类型 (BaseMetaGroupType)	130
6.3.4	元数据说明信息 (BaseMetaGroupType/ MetaItemNotes)	133
6.3.5	元元数据说明项元素 (MetaItemNotes/ MetaDescription)	134
6.3.6	标准稿件元数据组类型 ...	134
6.4	稿件操作信息	148
6.4.1	稿件操作容器类型 (OperationsType)	148
6.4.2	稿件操作元素 (Operations- Type/Operation)	148
6.4.3	稿件操作类型 (OperationType)	148
6.4.4	稿件排序操作信息类型 (SortingType)	148
6.5	实体型稿件	153
6.5.1	实体稿件内容容器元素 (EntityItemType/ Contents)	154
6.5.2	实体稿件内容容器类型 (EntityItemContentsType) ...	155
6.5.3	实体稿件内容关系容器 元素 (EntityItemType/ Relations)	155
6.5.4	实体稿件内容关系组 类型 (EntityContents- RelationType)	157
6.5.5	内容项关系类型 (Content- ItemRelationType)	158
6.5.6	内容项关系名类型 (contentItemRelation- NameType)	159

6.5.7	实体型稿件的具体 子类型	160
6.6	数字签名稿件	163
6.6.1	数字签名稿件类型 (SignatureItemType)	163
6.6.2	数字签名内容容器类型 (SignatureContentsType)	164
6.6.3	数字签名稿件使用示例 ...	164
6.7	词表稿件	165
6.7.1	词表稿件类型 (TopicListItemType)	165
6.7.2	词表稿件内容容器类型 (Top- icListItemContentsType)	167
6.8	统计稿件	169
6.9	引用稿件	170
6.9.1	引用稿类型 (ReferenceItemType)	170
6.9.2	引用稿内容容器类型 (Ref- erenceItemContentsType) ...	171
6.10	扩展 CNML 稿件体系	173
6.10.1	扩展的需求	173
6.10.2	扩展方法	173

第7章 包含多种媒体类型内容项的

CNML 稿件	178
7.1 什么是 CNML 内容项	178
7.2 内容项的构成	179
7.2.1 元数据	179
7.2.2 内容数据	182
7.3 内容项的建模	190
7.3.1 内容项的类型	190
7.3.2 内容项的元数据描述 (MetaInfo)	191
7.3.3 数据内容 (Content)	193
7.3.4 目标文件引用 (href)	196
7.4 内容项的元数据	196
7.4.1 描述元数据组	196
7.4.2 权利元数据组	203

7.4.3 物理元数据组	205	8.4 CNML 的受控词表用例	258
7.5 CNML 稿件中包含单一媒体类型 内容项的描述方法	222	8.5 CNML 词表使用规范	259
7.5.1 包含文本类型内容项的 CNML 稿件	222	第9章 CNML 中的关系	263
7.5.2 包含图片类型内容项的 CNML 稿件	224	9.1 什么是 CNML 中的关系	263
7.5.3 包含图形类型内容项的 CNML 稿件	225	9.2 CNML 中关系的种类	267
7.5.4 包含音频类型内容项的 CNML 稿件	227	9.2.1 根据关系作用对象分类	268
7.5.5 包含视频类型内容项的 CNML 稿件	228	9.2.2 根据关系建模粒度分类	268
7.5.6 包含网页类型内容项的 CNML 稿件	229	9.2.3 根据对关系的描述层面进行 分类	268
7.5.7 包含应用文档类型内容项的 CNML 稿件	231	9.2.4 根据关系的类模型进行 分类	271
7.6 CNML 稿件中的包含多媒体类型 内容项的描述方法	232	9.3 稿件层稿件之间的关系	272
7.6.1 文字配图稿	233	9.4 内容项层内容之间的关系	273
7.6.2 文字带图形稿	234	9.5 受控词之间的关系	275
7.6.3 文字带附件稿	236	9.6 关系的例子	275
7.6.4 文字带视频稿	237	第10章 CNML 数字签名基础	278
7.6.5 文字、图片带音频稿	239	10.1 数字签名技术基础	278
7.6.6 音视频稿	240	10.2 XML 数字签名技术概述	280
第8章 CNML 的受控词表	243	10.3 新闻行业使用数字签名技术的 作用和意义	283
8.1 概述	243	10.4 CNML 中的数字签名机制设计 思想	285
8.2 受控词表的建模	244	10.4.1 CNML 对数字签名的 设计	285
8.2.1 CNML 中对受控词相关概念 和类型的定义	244	10.4.2 数字签名示例：对稿件 内容项签名	286
8.2.2 受控词表的建模	253	10.4.3 数字签名示例：对稿件 签名	287
8.2.3 受控词表稿件建模	253	10.5 数字签名实例过程	288
8.2.4 受控词表内容项建模	253	第11章 CNML 的扩展	292
8.2.5 受控词的建模	254	11.1 XML Schema 本身的扩展 机制	292
8.2.6 具有嵌套结构的受控词表的 建模	254	11.1.1 命名空间级扩展机制	292
8.3 CNML 受控词表的使用方法	254	11.1.2 文件级扩展机制	297
		11.2 CNML 中的 import 和 include	

关系	300
11.3 标准结构的扩展机制	302
11.3.1 关系基类型及其扩展 结构	302
11.3.2 受控词表及其扩展 机制	304
11.3.3 稿件基类型及其扩展 结构	305
11.3.4 内容项基类型及其扩展 结构	307
11.4 元数据扩展机制	308
11.4.1 通用元数据扩展机制	308

11.4.2 元数据项扩展机制	308
11.4.3 元数据类型扩展	309
附录 A 完整 CNML 示例	311
附录 B CNML 元数据描述 定义	333
附录 B.1 CNML 的文档层元 数据	333
附录 B.2 CNML 的稿件层元 数据	334
附录 B.3 CNML 的内容项层元 数据	339

第 1 章 CNML 概述

1.1 什么是 CNML

CNML 是中文新闻信息置标语言 (Chinese News Markup Language) 标准的英文简称,它是用于新闻信息采集、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使用的与国际接轨、又有中国特色的多媒体新闻信息元数据规范。该技术标准采用了 XML 和 RDF 技术,支持文字、图片、图形、音频、视频等多种媒体类型以及多媒体混合的新闻信息的描述,支持对新闻信息间关系和语义的描述,为新闻信息的自动解释、交换和处理等新业务的发展留有余地。CNML 标准具有良好的扩展性和实用性,可用于多媒体新闻的组织和报道。

CNML 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。

CNML 是从对中文新闻信息处理的应用需求和大中华地区各类新闻媒体单位业务特点的分析出发,针对新闻媒体单位内部的新闻采编发系统、不同类型新闻媒体之间的新闻数据交换系统、新闻信息浏览系统,以及其它与中文新闻信息数据处理相关的各类信息系统而设计的一套完整的、基于 XML 技术的数据标准,其主要任务是使相关系统能够协同工作,通过完成对新闻内容数据和相关业务数据的处理,从而实现对新闻数据的加工和利用。

CNML 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和服务商。

1.2 CNML 的开发背景

随着新闻行业竞争的日益激烈以及信息网络技术、特别是互联网技术的飞速发展,新闻信息需求不断迅猛增长,新闻信息的表现形式和发布渠道日趋多样化,多媒体已经成为新闻传媒发展的新趋势,如何有效地组织和管理日益庞大的结构化、非结构化数据,充分整合、挖掘并利用丰富的新闻信息资源,拓展新闻