

GB

中国

国家

标准

汇编

2013年 修订-24



中国标准出版社

中国国家标准汇编

2013 年修订-24

中国标准出版社 编

中国标准出版社

中国标准出版社

北 京

图书在版编目(CIP)数据

中国国家标准汇编:2013 年修订. 24/中国标准出版社编. —北京:中国标准出版社, 2014. 9
ISBN 978-7-5066-7643-4

I. ①中… II. ①中… III. ①国家标准-汇编-中国-2013 IV. ①T-652.1

中国版本图书馆 CIP 数据核字(2014)第 187400 号

中国标准出版社出版发行
北京市朝阳区和平里西街甲 2 号(100029)
北京市西城区三里河北街 16 号(100045)

网址 www.spc.net.cn
总编室:(010)64275323 发行中心:(010)51780235
读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

*

开本 880×1230 1/16 印张 38 字数 1 165 千字
2014 年 9 月第一版 2014 年 9 月第一次印刷

*

定价 220.00 元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68510107

出 版 说 明

1. 《中国国家标准汇编》是一部大型综合性国家标准全集。自 1983 年起,按国家标准顺序号以精装本、平装本两种装帧形式陆续分册汇编出版。它在一定程度上反映了我国建国以来标准化事业发展的基本情况和主要成就,是各级标准化管理机构,工矿企事业单位,农林牧副渔系统,科研、设计、教学等部门必不可少的工具书。

2. 《中国国家标准汇编》收入我国每年正式发布的全部国家标准,分为“制定”卷和“修订”卷两种编辑版本。

“制定”卷收入上一年度我国发布的、新制定的国家标准,顺延前年度标准编号分成若干分册,封面和书脊上注明“20××年制定”字样及分册号,分册号一直连续。各分册中的标准是按照标准编号顺序连续排列的,如有标准顺序号缺号的,除特殊情况注明外,暂为空号。

“修订”卷收入上一年度我国发布的、被修订的国家标准,视篇幅分设若干分册,但与“制定”卷分册号无关联,仅在封面和书脊上注明“20××年修订-1,-2,-3,……”字样。“修订”卷各分册中的标准,仍按标准编号顺序排列(但不连续);如有遗漏的,均在当年最后一分册中补齐。需提请读者注意的是,个别非顺延前年度标准编号的新制定的国家标准没有收入在“制定”卷中,而是收入在“修订”卷中。

读者配套购买《中国国家标准汇编》“制定”卷和“修订”卷则可收齐由我社出版的上一年度我国制定和修订的全部国家标准。

3. 由于读者需求的变化,自 1996 年起,《中国国家标准汇编》仅出版精装本。

4. 2013 年我国制修订国家标准共 1 979 项。本分册为“2013 年修订-24”,收入新制修订的国家标准 1 项。

中国标准出版社

2014 年 8 月

目 录

GB/T 20092—2013 中文新闻信息置标语言	1
----------------------------------	---



中华人民共和国国家标准

GB/T 20092—2013
代替 GB/T 20092—2006

中文新闻信息置标语言

Chinese news markup language

2013-12-31 发布

2014-07-15 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

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准代替 GB/T 20092—2006《中文新闻信息置标语言》。

本标准与 GB/T 20092—2006 相比的主要变化如下：

- 删除了引言；
- 修改了第 1 章范围里的部分内容；
- 修改了第 2 章规范性引用文件的引导语，删除了本标准中未直接引用的文件；
- 修改了第 3 章的部分术语定义；
- 修改了第 6 章 CNML 体系结构中的图 1、图 2、图 4、图 7，以及附录 C 中的 CNML 类继承总图；
- 修改了 6.4，删除了原稿件结构中与管理有关的内容，并删除了 12.1.2；
- 删除了 7.3，简化了关系的表达机制，并修改了第 8 章；
- 修改了第 9 章，删除了受控词的关系机制，修改了受控词的基类型，增加了对受控词表简化使用的机制和受控词语义描述的功能；
- 增加了 11.3，词表编目结构；
- 修改了 11.4，修改了标准的信封结构的部分定义以及有关应用示例；
- 修改了 11.5.2.3，修改了引用稿件类型定义；
- 增加 12.2.4.4，定义了稿件专题信息的相关类型；
- 修改了 12.2.5.8，增加了权利元数据组中授权使用信息下的元数据；
- 增加了 13.2.2.8，增加了对内容项分组引用和描述的结构；
- 修改了 13.3.4.4、13.3.4.5，增加了音频、视频元数据组中的元数据项；
- 增加了 13.3.4.6，新增了音视频合一的元数据组；
- 删除了附录 A.2，修改附录 A.1 内容并作为新的附录 A；
- 修改了附录 B，增加了部分元素；
- 在附录 D 中增加了关系词表等受控词表；
- 修改原有的附录 F，改为定义内联智能标记的 XML Schema；
- 针对以上修改和标准应用中提出的其他合理修改建议，对标准 Schema 作了相应修改，并根据由此带来的变化，对相应的图表和内容进行了修改。

本标准由全国中文新闻信息标准化技术委员会(SAC/TC 352)提出并归口。

本标准起草单位：新华通讯社、清华大学、北大方正电子有限公司、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、黄菁、李涓子、张鹏、曹学军、王付生。

本标准所代替标准的历次版本发布情况为：

- GB/T 20092—2006。

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体系结构、公共结构和基本数据类型、关系机制以及受控词描述机制,并给出了新闻信息的数字签名和内联智能标记的使用方法。

本标准适用于新闻信息电子数据的“采集、编辑、生成、发布、交换、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与研究机构。

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 4880.2—2000 语种名称代码 第2部分:3字母代码

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/T 18793—2002 信息技术 可扩展置标语言(XML)1.0

GB/T 20093—2006 中文新闻信息分类与代码

3 术语和定义

GB/T 7408—2005 和 GB/T 18793—2002 中界定的以及下列术语和定义适用于本文件。

3.1

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,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。

3.5

内容项 content item

实体稿件类型包含的单一内容,可以包含任意一种文字、图片、图形、音频、视频、视音频、多媒体、应用文档等不同类型的数据内容及其元数据。

3.6

信封 envelop

完成新闻信息交换功能的组件结构。

3.7

产品 product

新闻提供商按照特定的目的分类组织新闻的一种形式,以便提供给新闻消费者订购,一般与订购它的一组特定的新闻用户对应。“产品”是针对新闻合约和新闻分类而产生的一个概念。

3.8

专稿 special product

个别用户根据自己需要就特定内容稿件向供稿方进行预约,要求专门提供的供稿形式。

3.9

栏目 column

对内容的一种细分方式。栏目可以是稿件的一种内容属性也可以是展示属性。

4 缩略语

下列缩略语适用于本文件

CNML	中文新闻信息置标语言(Chinese News Markup Language)
NewsML	新闻置标语言(News Markup Language)
UML	统一建模语言(Unified Modeling Language)
URI	统一资源标识符(Uniform Resource Identifier)
URL	统一资源定位器(Uniform Resource Locator)
URN	统一资源名称(Uniform Resource Name)
UTF-8	8位统一码转换格式(8-bit Unicode Transformation Format)
XML	可扩展置标语言(Extensible Markup Language)

5 规则和约定

5.1 拼写规则

每个单词的首字母大写,其余字母均小写,不使用任何连字符,在名称中不使用下划线“_”、句点“.”和连字符“-”。单词应完整使用,在名称中不使用缩略语,以保证语义的清晰,提高可读性。

5.2 元素表达及描述要求

在本标准 Schema 的不同位置出现的元素如有不同的语义,一般要求采用不同的元素名。标准文档中出现元素名称时要求用〈〉括起来作为提示,便于同其他类型的名词出现相区别,如,信封元素在本文档中出现时应表示为〈Envelop〉,而不是 Envelop。本标准中元素标记的声明通常不以该元素的相关父元素或祖先元素的路径表达式作为前缀。当相同语义的元素名在 Schema 中出现不止一次时,需要使用语境描述才能区别时,则应采用 XPath 的标准写法,以便于了解其出现的具体位置,如信封结构中的当日发稿序号,可根据其不同语境表达为 SentTo/SingleRecipient/RecipientInfo/DailySequence 或

SentTo/GroupRecipient/DailySequence。

5.3 XML 组件的命名

本标准中与元素、属性、简单类型和复杂类型概念相关的命名基于如下约定。

如果名称基于多个单词,每个单词的首字母应是大写的。第一个单词的首字母是否大写是根据描述的概念的类型决定的。具体规定如下:

- a) 元素的命名:第一个单词的首字母应是大写(如表示电子地址的元素 ElectronicAddress);
- b) 属性的命名:第一个单词的首字母应是小写(如 Column 元素的属性 howPresent);
- c) 复杂类型的命名:第一个单词的首字母应是大写,类型名后添加后缀 Type(如表示受控词引用的复杂类型 TopicUseType);
- d) 简单类型的命名:第一个单词的首字母应是小写,类型名后添加后缀 Type(如表示受控词引用的简单类型 topicRefType)。

5.4 CNML XML Schema 的版本和变更约定

CNML XML Schema 的版本号及变更遵循如下原则约定:

- a) CNML XML Schema 采用“m.n”的格式表示 XML Schema 的版本信息,其中 m,n 均为非负整数,m 表示大版本号取值,n 表示小版本号取值;
- b) CNML XML Schema 的初始版本号为“1.0”;对应的目标名字空间 URI 为: <http://www.news-standards.org/2012/CNMLSchema>,本次修订以后的版本号“2.0”,对应的目标名字空间 URI 为: <http://www.news-standards.org/2012/CNMLSchema>;
- c) 如果先前版本的任何实例文档针对新修订的 CNML XML Schema 继续保持合法,则新版本的 XML Schema 只需要递增小版本号,并在 XML Schema 的 Version 属性中指定新的版本号信息,换言之,CNML XML Schema 的小版本变更严格遵循向前兼容原则;
- d) 如果先前版本的任何一篇实例文档在新修订的 CNML XML Schema 校验下不合法,则新版本的 XML Schema 需要递增大版本号 m 的取值,同时需要变更新版本 CNML XML Schema 的目标名字空间为一个新的 URI,换言之,CNML XML Schema 的大版本变更并不遵循向前兼容原则;
- e) CNML XML Schema 新修订的版本颁布后,各个先前版本的文件仍将保留并对外公布。

6 CNML 体系结构

6.1 原则

本章定义中文新闻信息置标语言的总体结构,该结构用于新闻生产各个过程对新闻信息的描述。

CNML 结构的建立原则是:

- a) 以新闻领域业务对内容的需求分析为基础;
- b) 独立于新闻样式的表现;
- c) 独立于特定新闻内容;
- d) 独立于新闻信息标准的表示语言。

本标准对新闻内容本身、新闻内容管理和不同新闻内容之间关系以及对反映新闻内容的各类元数据进行建模,并定义这些内容的描述方法。

6.2 总体结构

CNML 总体结构如图 1 所示。

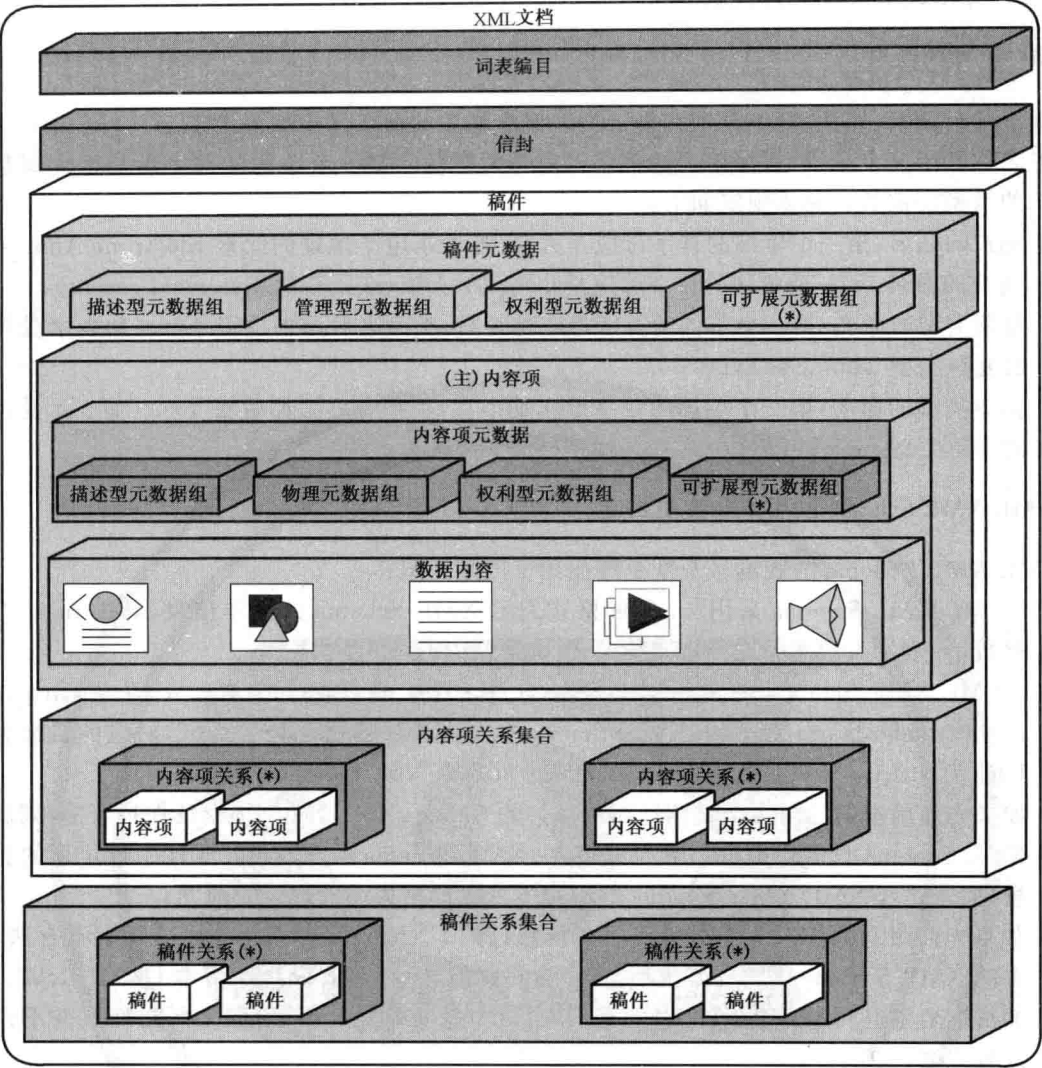


图 1 CNML 总体结构图

该模型规定了 CNML 对文档的描述方式, CNML 文档是一个包含文档层、稿件层和内容项层的三层结构描述模型。一个 CNML 文档可以包含多个稿件, 一个稿件可以由多个内容项组成。属于同一文档的不同稿件组成一个稿件序列, 稿件之间的关系描述与稿件的内容描述分离。属于同一内容项层的内容组成一个序列, 内容项之间的关系描述与内容项的描述分离。

CNML 在文档层中定义了用于文档通讯控制过程中的信息模型—信封结构。在稿件层和内容项层的描述中, 定义了对稿件层和内容项层的元数据描述模型。同时, 稿件还定义了关系描述和受控词表的定义和引用机制。

一个 CNML 文档的结构包含文档层结构、稿件层结构、内容项层结构、元数据的结构、受控词表定义和引用结构和关系结构。

6.3 文档层结构

文档层结构由四部分组成, 即词表编目、信封、稿件集合和稿件关系集合, 如图 2 所示。具体定义如下:

- a) 词表编目: 在 CNML 文档开始位置集中对使用的受控词表进行简化的别名描述;

- b) 信封:完成新闻信息交换功能的组件结构,描述了如下信息:
- 1) 传输标识:定义 CNML 文档传输的一种标识符;
 - 2) 系统标识:定义生成发布 CNML 文档的系统名称;
 - 3) 发送时间:定义为稿件的实际发送时间;
 - 4) 定时发布时间:定义为稿件的计划发送时间;
 - 5) 发送方:定义为发送稿件的人或机构;
 - 6) 送达方:定义为一个或多个稿件接受者和机构;
 - 7) 发稿优先级:定义为稿件的优先级;
 - 8) 服务信息:定义为新闻单位提供的新闻服务;
 - 9) 产品信息:定义为从新闻内容的角度出发,按照特定的分类方法具备相同内容特征的新闻数据;
- c) 稿件集合:一个 CNML 文档可以包含一个或多个稿件;
- d) 稿件关系集合:一个 CNML 文档的稿件关系集合包含对各个稿件之间的关系的描述。

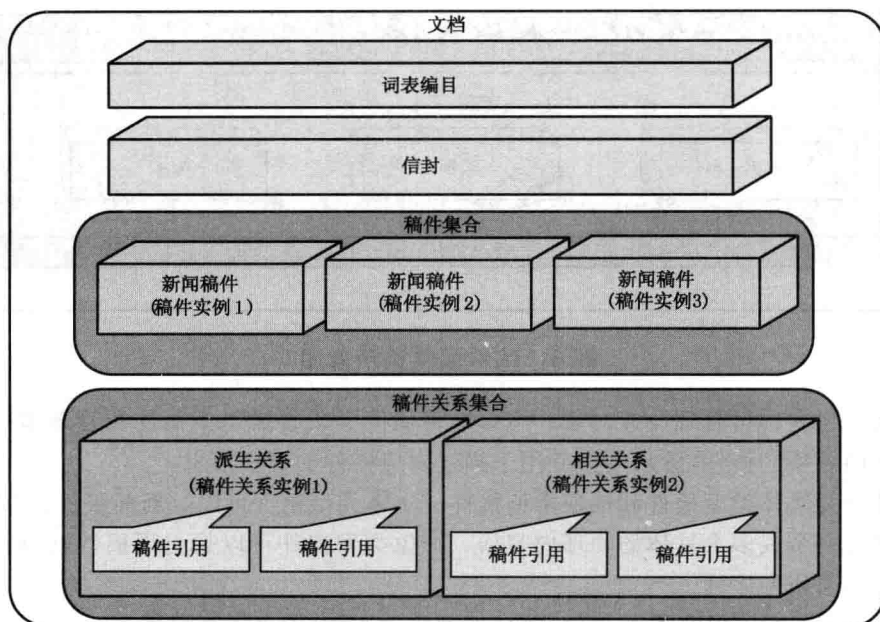


图 2 文档层结构示意图

词表编目结构提供了简化 CNML 受控词表使用的可选机制。CNML 中有些元素的取值来自受控词表,每个内容提供商均可以自由地使用自己定义的或别的组织定义的受控词表。由于 CNML 中规范的受控词表的语法描述较为冗长,为了简化 CNML 中受控词表的描述和方便程序的处理,CNML 在文档开始位置通过词表编目提供了对使用的受控词表进行简短的别名定义和外部引用的机制。

6.4 稿件层结构

稿件层结构由三部分组成,即稿件元数据集合、稿件内容集合、稿件内容关系集合,如图 3 所示。具体定义如下:

- a) 稿件元数据集合包含一个稿件管理型元数据组,一个可选的描述型元数据组、一个可选的权利型元数据组;还可以包含多个可扩展的其他元数据组实例;
- b) 稿件内容集合:包含一个或多个具体内容项,不同类型的稿件拥有不同类型的内容;
- c) 稿件内容关系集合:包含一到多个组成稿件的多个内容项之间的关系。

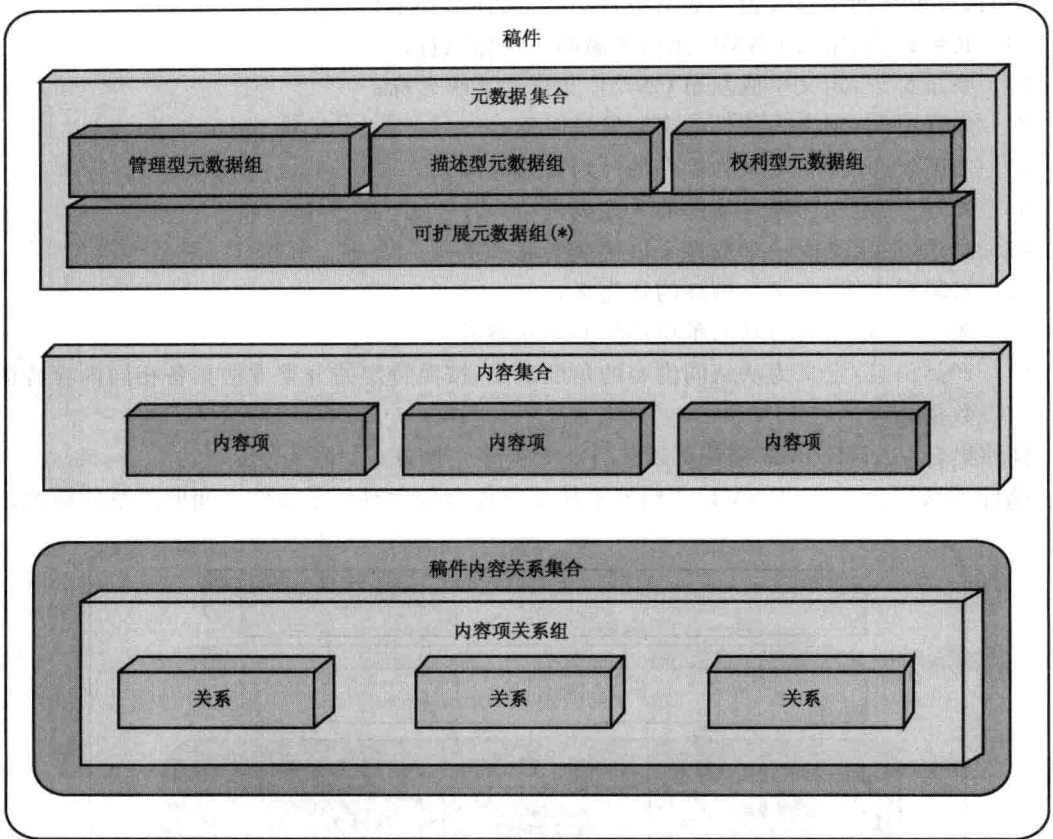


图3 稿件层结构示意图

稿件是文档中包含的可管理业务对象。本标准将稿件定义为描述型稿件和管理型稿件,描述型稿件定义具有实际内容项的稿件,管理型稿件用于描述对应稿件的管理信息。

描述型稿件分为实体类型稿件和抽象类型稿件。实体类型稿件可以是新闻稿件、广告稿件等,每个实体型稿件中含有一个或多个具体的物理内容项。抽象类型稿件可以是引用稿件,引用稿件中内容项只定义对内容项的引用的序列。

管理型稿件可以是数字签名类型稿件、统计类型稿件、受控词表类型稿件等。

为了方便对各类型稿件的管理和描述,本标准定义了如图4所示的多层次的稿件分类。

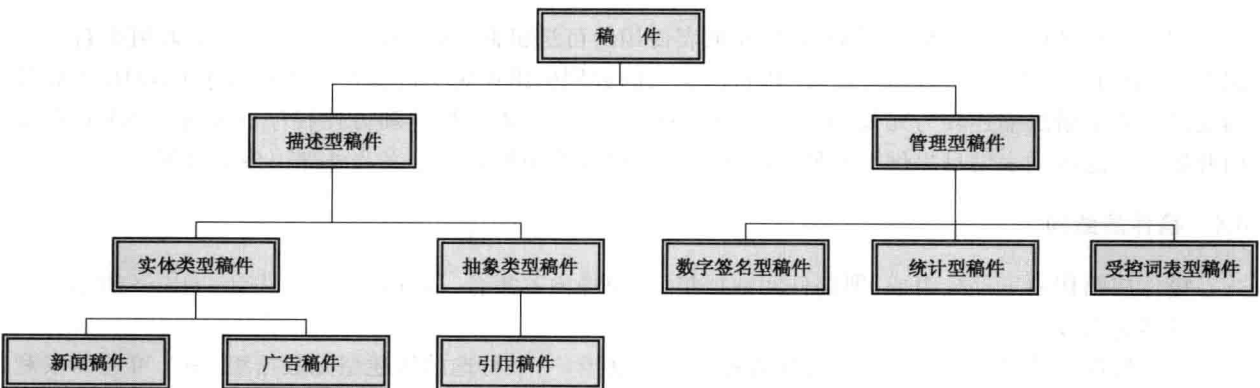


图4 稿件分类图

6.5 内容项层结构

内容项是实体稿件类型包含的具体内容，一个实体类型稿件既可以由多种不同类型的内容项组成，也可以包含同一类型的不同内容项。

本标准定义了以下类型的内容项：

- a) 文本类型内容项：内容项中的内容只包含文字；
- b) 图片类型内容项：内容项中的内容是图片类型的数据；
- c) 图形类型内容项：内容项中的内容是图形类型的数据；
- d) 音频类型内容项：内容项中只包含音频数据；
- e) 视频类型内容项：内容项中只包含视频数据；
- f) 视音频类型内容项：内容项同时包含视频和音频数据；
- g) 多媒体类型内容项：内容项中是多媒体类型的数据；
- h) 应用文档类型内容项：内容项是嵌入的一个应用文档。

内容项层结构由两部分组成，即内容项元数据集合和数据内容，如图 5 所示。

具体定义如下：

- a) 内容项元数据集合包含一个可选的物理元数据组、一个可选的描述型元数据组、一个可选的权
利型元数据组，还可以包含多个其他类型的内容项元数据组；
- b) 数据内容是内容项包含的数据，可以以外部引用的方式存在，也可以嵌入到 CNML 文档中，如
果内容项数据是二进制数据，则嵌入时需要进行编码。

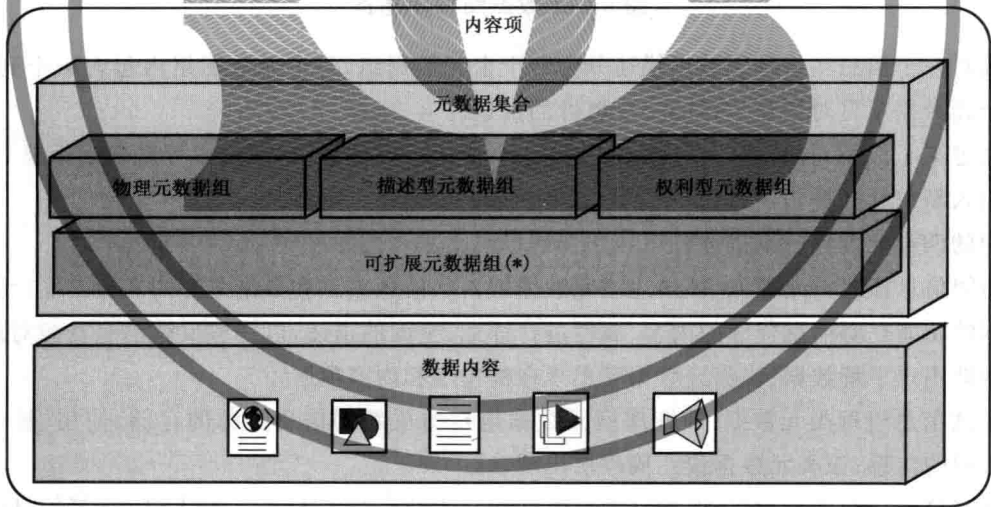


图 5 内容项结构示意图

6.6 元数据结构

元数据结构如图 6 所示。

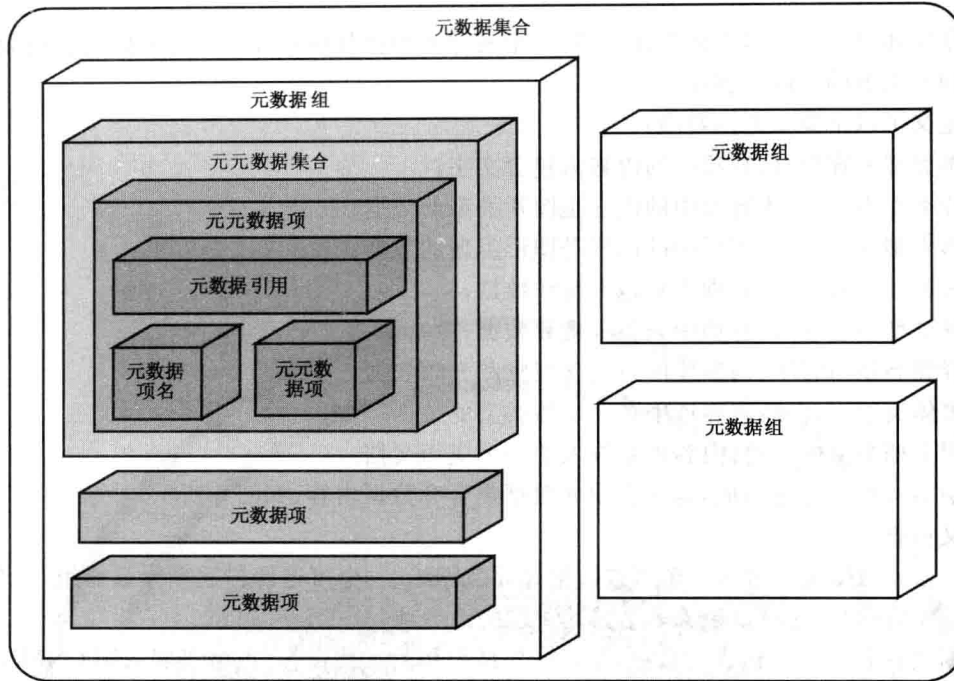


图6 元数据结构示意图

本标准对一个信息实体的元数据描述分为若干个元数据组，每个元数据组内包含一个到多个元数据项。每个元数据项是对信息实体的一个属性的描述。

本标准定义了四类新闻信息元数据，每一类定义为一个元数据组，用户也可以基于标准中元数据组基础结构引入新的元数据组。

标准中的四类元数据分别描述如下：

- a) 新闻信息描述类元数据(即描述型元数据组)：从信息共享和数据交互的需要出发，元数据最基本的功能是对信息资源的内容、属性进行详细、全面的描述，以完整反映信息资源对象的全貌，帮助用户了解数据；该类元数据可出现在稿件层和内容项层；
- b) 新闻信息管理类元数据(即管理型元数据组)：提供新闻信息资源的存储、使用、管理、传输等方面的信息；该类元数据仅在稿件层出现；
- c) 新闻信息权利类元数据(即权利型元数据组)：提供新闻信息其拥有者、版权等信息内容；可出现在稿件层和内容项层；
- d) 新闻信息物理属性类元数据(即物理元数据组)：对新闻信息物理属性、实体文件信息进行全面的描述；该类元数据仅在内容项层出现。

本标准支持对元数据的元数据信息(即元元数据)的描述；除上述四类元数据外，用户可以对标准元数据组进行扩展。

本标准根据元数据项取值特性，将其分成以下类型：

- a) 简单元数据项：该类元数据项的取值可以通过字符串进行表示；
- d) 基本元数据项：该类元数据项的取值可以通过字符串进行表示，而取值的具体数据类型是XML Schema 规范定义的基本数据类型；

- c) 自定义取值类型的简单元数据项：该类元数据项的取值为简单数据类型，取值最终可以表示为一个字符串。但数据类型不是标准 XML Schema 规范定义的基本数据类型，可以由扩展者定义；
- d) 通过受控词表限定其取值的元数据项：该类元数据项的取值通过字符串进行表示，但是字符串的取值应为受控词表中定义的一个词表项。有关受控词表的定义见第 9 章；
- e) 复杂元数据项：该类元数据项的取值相对复杂，无法通过一个简单的字符串进行描述，而应通过具备一定结构的数据模型进行描述。

本标准规定除了第一类元数据项以外，其他类型的元数据项均适用于复杂类型的元数据模型。

6.7 受控词表结构

受控词表是新闻信息交换过程中一类重要的实体，它在新闻信息交换过程中定义了新闻领域和新闻部门对某个具体事物描述的统一认识，是对数据的可理解的语义描述的基础。受控词表结构如图 7 所示。

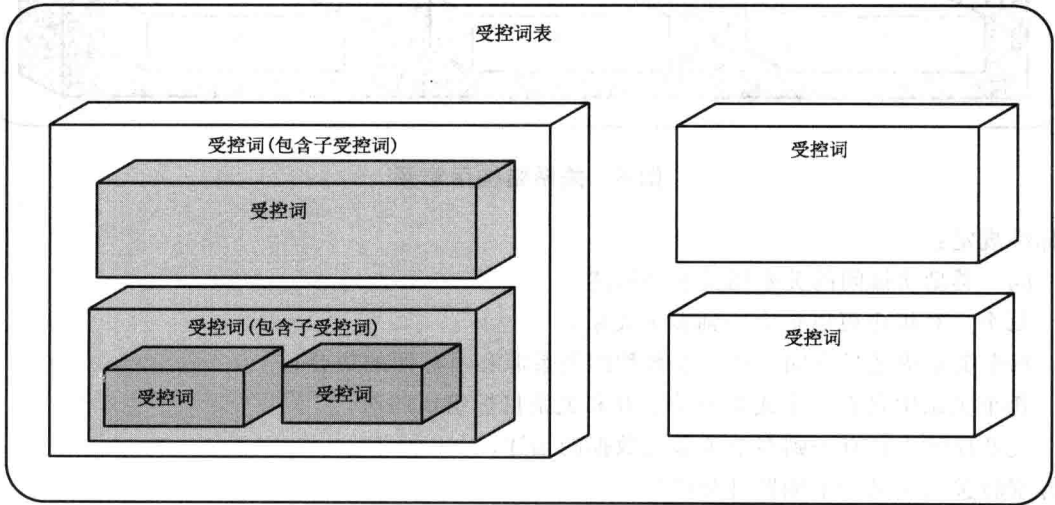


图 7 受控词表结构示意图

本标准提供了受控词表定义以及受控词引用的描述机制，见 9.2 和 9.3。

本标准提供了受控词表稿件类型来定义受控词表，支持具有层次结构的受控词表描述，此外还提供机制可以进一步描述受控词的剖面语义信息，例如，与其他受控词之间的关系等。

6.8 关系结构

本标准定义了用于描述同一类型信息实体之间的相互作用的关系结构，如图 8 所示。

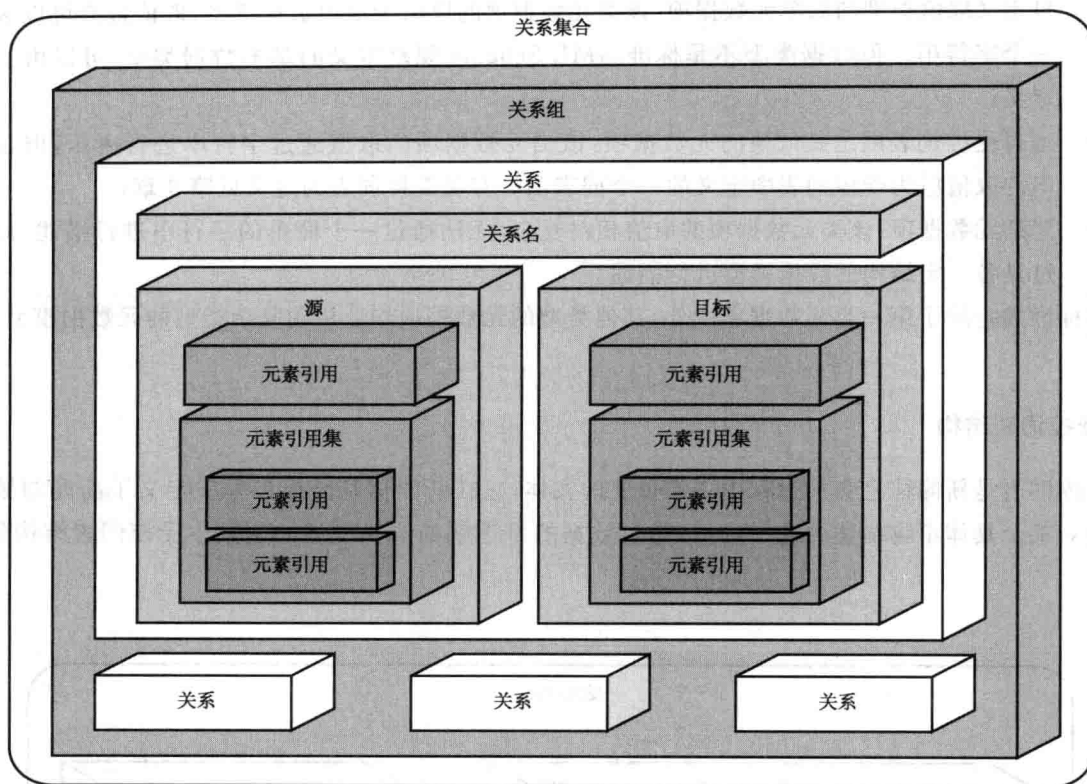


图 8 关系结构示意图

本标准规定：

- a) 同一类型实体间的关系用关系组描述；
- b) 每个关系组中可以包含一到多个关系；
- c) 每个关系描述具有同一种关系名称的关系源和目标源的集合；
- d) 每个关系中具有一个或多个关系源和关系目标组成的对；
- e) 关系描述中具有一到多个关系元数据的描述。

本标准将关系定义为下面四部分内容：

- a) 关系名称：关系的语义概念；
- b) 关系源：关系的发源体；
- c) 关系目标：关系的接受体；
- d) 关系的元数据：描述关系的元数据，每一项元数据都为简单元数据项类型。

目前，本标准定义的实体关系包括稿件之间、内容项之间的关系。

7 CNML 公共结构和基本数据类型

7.1 目标命名空间声明

本版标准规定如下 URI 作为 CNML 的目标命名空间：

<http://www.news-standards.org/2012/CNMLSchema>

7.2 基本简单类型

针对基于包括时间、日期等常见数据类型的简单元数据项，本标准定义了五类基本数据类型：