

团 体 标 准

T/CES XXX-XXXX

移动端智能交互命名实体标注规范

Mobile Intelligent Interactive Named Entity
Annotation Specification

(征求意见稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国电工技术学会 发布

目 次

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
<u>5 总则.....</u>	<u>2</u>
6 样本基本要求.....	2
6.1 文本文件存储格式要求.....	2
6.2 文本文件命名要求.....	2
6.3 文本类样本质量要求.....	2
6.4 文本样本描述文件.....	2
7 样本标注要求.....	2
7.1 基本要求.....	2
7.2 单类词词性标注要求.....	3
7.3 多类词词性标注要求.....	3
7.4 实体抽取样本标注要求.....	3
7.5 实体关系抽取样本标注要求.....	3
7.6 事件抽取样本标注要求.....	3
7.7 标注文件命名与存储要求.....	3
8 样本标注流程.....	3
8.1 总体要求.....	3
8.2 样本检查.....	3
8.3 安全管控.....	3
8.4 标注工具.....	4
8.5 样本标注结果收集.....	4
8.6 样本标注结果检查.....	4
9 标注任务开展.....	4
9.1 基本要求.....	4
9.2 人名标注规范.....	4
9.3 职名标注规范.....	6
9.4 地名标注规范.....	7
9.5 组织机构名标注规范.....	10
9.6 产品名标注规范.....	13
9.7 品牌标注规范.....	15
9.8 事件标注规范.....	15
9.9 时间结构标注规范.....	16
9.10 数量结构标注规范.....	17
9.11 数码标注规范.....	19

前 言

为规范移动端智能交互命名实体模型开发的命名实体标注基本要求,为移动端智能交互样本标注和样本入库的统一管理工作开展提供指导规范,制定本文件。

本文件按照 GB/T1.1—2020《标准化工作导则 第 1 部分:标准化文件的结构和起草规则》给出的规则起草。

本文件由四川中电启明星信息技术有限公司提出。

本文件由中国电工技术学会标准工作委员会能源智慧化工作组归口。

本文件起草单位:国网信息通信产业集团有限公司、四川中电启明星信息技术有限公司、国网重庆市电力公司。

本文件主要起草人:李强、宋卫平、邓芷珊、周孔均、张捷、赵峰、田鹏、李炳森、李欢欢、杨帆、王刚强、杨智、谷波、胡州明、王红蕾、田富强、李立。

本文件为首次发布。

1 范围

本标准规定了移动端智能交互训练时命名实体（包括狭义和广义命名实体）标注的基本要求、流程和规范。

本标准适用于各单位进行移动端智能交互命名实体模型开发时的样本标注和样本入库的统一管理，包括样本的质量管控、样本标注的技术管控和流程管控。

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 5271.28—2001 信息技术 词汇 第 28 部分 人工智能 基本概念与专家系统

T/CESA 1040—2019 信息技术 人工智能 面向机器学习的数据标注规程

Q/GDW 12118.1~12118.3—2021 人工智能平台架构及技术要求

GB/T 18521—2001 我国地名分类与分类代码编制规则

2007.3.12 命名实体语料标注规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

命名实体 *named entity*

1) 狭义讲，命名实体是唯一一个体的专有名称。狭义命名实体具备三个充要属性。属性 a. 符号性：命名实体是一个区别性称谓，具有代号性质；命名实体并不排除蕴含义，但蕴含义不是命名实体的必然属性。属性 b. 个体性：命名实体所指称的事物不可向下分类，因为类是对群体区分的结果（但任何事物都是一个过程，都是一个和合态，所以，既可以按阶段划分，也可以分解成部分）。属性 c. 唯一性：命名实体所指称的事物是世上独一无二的，通常不以数量词修饰；一旦饰以数量词，命名实体便转义了。

2) 广义讲，除狭义命名实体之外，命名实体还包括对一类事物的非概念性称谓。对一类事物的命名可划分为概念性命名和非概念性命名两种方式。概念性命名是通过自然语义和构词法来规定事物内涵的，如：月门、拱桥；非概念性命名是通过不表明事物内涵的饰词（一般充当命名实体饰名）来指称一类事物的，如：“丁达尔现象，一二次设备”。广义命名实体一般只具备符号性，不具备个体性和唯一性。

3.2

样本数据 *sample data*

其具备的特征能够反映总体数据情况的一部分个体数据。

[Q/GDW 12118.1—2021 定义 3.5]

3.3

标注 *corpus annotation*

采用人工或计算机自动方式对样本的属性或特征进行描述，可用于实用的目的，如客户服务或资讯获取等。

[Q/GDW 1906—2013 定义 3.5]

3.4

标签 *label*

标识数据的特征、类别和属性等内容，可用于建立数据及深度学习训练要求所定义的机器可读数据编码间的联系。

[T/CESA 1040—2019 定义 3.2]

3.5

分词 participle

将连续的字符序列切分成一个个单独的词。

3.6

标注工具 annotation tool

标注人员执行标注任务生成标注结果的过程中使用的工具和软件。标注工具按照自动化程度分手动、半自动和自动三种。

[T/CESA 1040—2019 定义 3.5]

4 缩略语

下列缩略语适用于本文件。

BIOES: BIOES 标注模式 (B-begin, I-inside, O-outside, E-end, S-single)

JSON: JavaScript 对象表示法 (JavaScript Object Notation)

WAV: 窗波 (Windows Wave)

XML: 可扩展的标记语言 (Extensible Markup Language)

5 总则

本文件共分为样本基本要求、样本标注规范、标注任务流程三部分。总则内容概要参见表 1:

表 1 总则内容概要

序号	内容	概要
1	样本基本要求	体现文本类样本要求规范
2	样本标注要求	包括文本类样本词性、关系、命名等标注规范
3	标注任务流程要求	根据业务需求和给出的人名、职名、地名等常见实体分类标准进行标注

6 样本基本要求

6.1 文本文件存储格式要求

文本数据应采用 txt、csv、JSON、xls、xlsx、XML 存储格式。

6.2 文本文件命名要求

文本文件名称应由四个部分组成:

- a) 第一部分为项目命名词或文本来源;
 - b) 第二部分为当前文本文件的专业信息;
 - c) 第三部分为原始源文件生成时的日期, 日期格式: YYYY-MM-DD;
 - d) 第四部分为文件唯一性编号, 从 1 开始计数。
- e) 这四部分用下划线连接, 且文件名称长度和扩展名在内最大长度不超过 100 个字符

(包含中英文字符和特殊字符);

6.3 文本类样本质量要求

文本类样本质量应满足下述要求:

- a) 应支持计算机正常读取, 文本内容无乱码;
- b) 内容应满足相关业务需求;

c) 应使用 UTF-8 编码。

6.4 文本样本描述文件

每批次文本样本集应有一个描述文件，应满足下述要求：

- a) 存储格式应为 txt 格式；
- b) 命名应由三个部分组成：
 - 1) 项目命名词或样本来源；
 - 2) 本文件创建的日期，日期格式：YYYY-MM-DD；
 - 3) 文件唯一性编号，从 1 开始计数；
 - 4) 这四部分用下划线连接，且文件名称长度和扩展名在内最大长度不超过 100 个字符(包含中英文字符和特殊字符)；
- c) 文档内容应描述本样本集的基本信息，应包括样本所属项目、样本来源、创建日期、样本上传单位及联系人、样本量、样本用途等信息。

7 样本标注要求

7.1 基本要求

应满足标注对象范围、标注方式、标注文件命名要求。具体要求包括：

- a) 文本类样本标注应支持词、句子、整个文本等不同范围的标注；
- b) 文本类样本标注应支持序列标注、指针标注、多头标注等多种标注方式；
- c) 序列标注应采用 B、I、E、O、S 标签列表，宜采用 BIOES 标签方案进行标注。

7.2 单类词词性标注要求

对于单类词，在标注时应确定其词性。

7.3 多类词词性标注要求

对于多类词，应满足多种多类词标注规则。

7.4 实体抽取样本标注要求

实体抽取样本标注应满足下述要求：

- a) 样本标注前应定义实体语义类型，包含实体名称与层次结构；
- b) 若实体包含属性，应定义属性名称与属性值。

7.5 实体关系抽取样本标注要求

实体关系抽取应满足下述要求：

- a) 样本标注前应定义实体语义类型与关系；
- b) 关系定义应包含关系类型与相关实体信息，起始实体与目标实体，关系为单向或双向等信息。

7.6 事件抽取样本标注要求

事件抽取样本应满足下述要求：

- a) 样本标注前应定义实体语义类型与事件类型；
- b) 事件抽取样本应包含事件；
- c) 文本样本应具有正确的词性分析，事件触发词宜为名词或动词；
- d) 若事件具有属性，应定义属性名称与属性值。

7.7 标注文件命名与存储要求

标注文件应由两部分组成，第一部分与对应标注文本命名一致，第二部分为“-bz”，应保存为 txt 格式。

8 样本标注流程

8.1 总体要求

样本标注应包含样本检查、安全管控、标注工具选择、标注任务开展、标注结果收集和标注结果检查环节。

8.2 样本检查

在样本标注前应按照本文标注样本要求对待标注样本进行检查,应根据业务需求和样本的数量采用全量检查或抽样检查,方式如下:

- a) 全量检查应对指定范围内的所有样本进行逐条检查;
- b) 抽样检查可采用随机抽样或分层抽样,方式如下:
 - 1) 随机抽样,即:针对不同业务类型的数据样本采用随机抽样进行检查;
 - 2) 分层抽样,即:针对同一业务类型的样本数据,根据样本类型不同采取分层抽样的方式进行检查。

8.3 安全管控

应满足对标注环境及标注人员的安全管控要求。具体要求包括:

- a) 标注过程应在内网环境下的指定机器中进行,机器应开启防火墙,安装杀毒软件,并禁用 USB 接口功能。机器中的所有数据文件需定期做好数据备份,不得擅自拷贝、传输,防止数据丢失或泄露;
- b) 标注人员应经过标注工作培训并签署样本标注保密协议后才可上岗操作。

8.4 标注工具

应使用满足移动端智能交互命名实体模型开发要求的标注工具或与其格式相兼容的标注工具进行标注。

8.5 样本标注结果收集

标注结果收集应满足以下具体要求:

- a) 样本标注结果应由统一的人员进行回收和存放,防止文件外泄;
- b) 标注结果收集人员应对样本标注结果的相关信息(包括任务名称、任务类型、任务开始时间、任务结束时间、任务描述进行核对)进行检查,防止文件遗漏;
- c) 标注结果收集人员,宜按照样本类型(图像视频,语音和文字)和标注方式(图像标注包括图像分类、图像目标检测、图像分割;文本标注包括文本分类、文本标注;音频标注包括音频分类、音频标注)进行分类保存。

8.6 样本标注结果检查

在样本标注结果收集后应按照本文第7章要求对标注结果进行检查,应根据业务需求和样本标注的数量采用全量检查或抽样检查,方式如下:

- a) 全量检查应对指定范围内的所有样本进行逐条检查;
- b) 抽样检查可采用随机抽样或分层抽样,方式如下:
 - 1) 随机抽样,即:针对不同业务类型的数据样本采用随机抽样进行检查;
 - 2) 分层抽样,即:针对同一业务类型的样本数据,根据样本类型不同采取分层抽样的方式进行检查。

9 标注任务开展

9.1 基本要求

应根据业务需求和命名实体分类进行标注。

9.2 人名标注规范

9.2.1 人名相关定义

狭义人名即个人专有名称,如:李华。广义人名指含有姓或名的个人称谓,如:大志先生。人名全称:指应有的姓、名的全称。人名简称,指省略了应有的姓、名构件或部分的所

有人名称谓方式。人名扩展：与人名右紧邻、被人名直接或间接修饰，并与人名一起表人称的词或紧密短语。如：小明妈妈，居里夫人。

根据项目研究目的，可以按照人名的文字特征、称谓方式及人的工作信息三个标准对人名进行分类。按文字特征和称谓方式对人名分类有助于人名自动识别。

9.2.2 按文字特征对人名分类

按文字特征对人名进行分类，分为汉字人名、非汉字人名、混合人名。详细分类见附表 A.1。

9.2.3 按称谓方式对人名分类

根据称谓方式对人名分类可以从实语料中统计得到人的称谓模式。人的称谓模式是某些称谓构件的组合方式。现将汉字人名的常用称谓构件归纳为 17 种：姓、名男、名女、性、职、敬、昵、亲、讳、序、诨、化、字、帝、译、符、其他。详见附表 A.2。

9.2.4 按职业对人名分类

根据职业对人名分类可主要依据由劳动和社会保障部、国家质量技术监督局、国家统计局联合组织编制、1999 年 5 月正式颁布的《中华人民共和国职业分类大典》进行。该书将我国职业归为 8 个大类，66 个中类，413 个小类，1838 个细类（职业），每个职业都已经编码。2005 年新增 10 项职业。8 个大类分别是：第一大类：国家机关、党群组织、企业、事业单位负责人，其中包括 5 个中类，16 个小类，25 个细类；第二大类：专业技术人员，其中包括 14 个中类，115 个小类，379 个细类；第三大类：办事人员和有关人员，其中包括 4 个中类，12 个小类，45 个细类；第四大类：商业、服务业人员，其中包括 8 个中类，43 个小类，147 个细类；第五大类：农、林、牧、渔、水利业生产人员，其中包括 6 个中类，30 个小类，121 个细类；第六大类：生产、运输设备操作人员及有关人员，其中包括 27 个中类，195 个小类，1119 个细类；第七大类：军人，其中包括 1 个中类，1 个小类，1 个细类；第八大类：不便分类的其他从业人员，其中包括 1 个中类，1 个小类，1 个细类。在今后使用中，可以此为母本进行适用性改造。

9.3 职名标注规范

9.3.1 职名定义及相关说明

职名：职务、职称、职业、职位名称的合称。

职务名称：组织机构中有职权的岗位名称。如：教务长、处长、秘书长、主席、经理等。

职业名称：表明所从事行业的人的职能身份名称。如：教师、公务员、程序员、商人、秘书、运动员等。“公司职员”、“职工”“白领”等不表明所从事行业，“商业”、“教育”、“体育”等不表明人的职能身份，这些都不是职业名称。

职称名称：有统一规范的职业级别的细分名称。如：教授、讲师、研究员等。职位名称：表示某机构的某个岗位的名称。职位名规则：机构名+职务名：润量公司总经理；或机构名+职称名：润量公司研究员。

职名并不属于命名实体范畴，虽然它能充当命名实体的构件和属性；但由于它和命名实体联系密切且具有重要的情报价值，所以将职名特别地在命名实体标注中同时予以标注。

9.3.22. 职名分类、构件、代码、归类标准、标注原则与式样

职名可分为职业名称、职务名称、职称名称。根据各自定义，职务名称不必须表明所从事行业、职业，但必须标志一定的职权，如：秘书长。职业名称则必须表明所从事行业和人的职能，且不表明职权，如：秘书。职称是职业的技术等级细分，与职务差别比较明显，但有时与职业不易区分。如：图书馆员本身既是职业名称，又是一级职称名称。若不明确指出职称、职业与职称同名的情况统一视为职业名称。详见附表 A.3。

9.4 地名标注规范

9.4.1 地名相关定义

地名 (geographical names): 人们对各地理实体赋予的专有名称。

标准地名 (standard geographical names): 使用规范的语言文字书写的地名全称。

地名简称 (shortened forms of geographical names): 经过官方认可或约定俗成的地名的简单称谓。

地名别名 (alias of geographical names): 某一地理实体的标准地名及其简称以外的其他现行名称。

现今地名 (under using geographical names): 目前正在使用的地名。

历史地名 (former geographical names): 过去曾经使用过、目前已不再使用的地名。

地名扩展: 与地名右紧邻、被地名直接或间接修饰, 并与地名一起表处所的词或紧密短语。如: 长江以南、李村村西小河。

9.4.2 地名分类、构件、归类标准

根据该项目研究目的, 可以按照地名的文字特征、地名所指代地理实体的空间位置、地名所指代地理实体的地理属性、地名的使用时间、地名的表示方式五项指标作为地名类别划分的标准。按照文字特征对地名进行分类有助于地名的自动识别。地名所指代地理实体的地理属性、地名的使用时间、地名的表示方式五项指标是地名最基本、最稳定的属性, 以此作为地名类别划分的标准能保证分类体系的稳定性。这也是我国对地名分类所采用的标准。

按照地名所指代地理实体的空间位置、地名所指代地理实体的地理属性、地名的使用时间、地名的表示方式对地名进行分类可直接依据我国地名分类与分类代码编制规则 [GB/T18521—2001] 进行。其分类原则如下: A. 选择地名最基本、最稳定的属性为分类依据, 保证分类体系的稳定性。B. 以地名学、地理学的学科分类为基础, 尽可能采用相关国际、国家、行业分类标准, 充分吸收新的科研成果, 体现分类体系的科学性、继承性、兼容性。C. 在体现科学性的前提下, 按地名及相关行业对地名分类的习惯, 以及不同类别地名的数量和使用频率, 在具体分类时作适当调整, 提高分类体系的实用性、通俗性。D. 在同一层面使用统一的分类方法, 保证分类体系的系统性、完整性。其分类方法见我国地名分类与分类代码编制规则 [GB/T 18521—2001]。

以上国家对地名的分类与编码方法能够将地名按照地名通名的类别聚类, 在此基础上可以对地名专名、地名饰名进行统计分析。

9.4.3 地名分类标注原则与式样

地名内部构件可分为地名通名、地名专名和地名饰名三大类。地名通名是表示该地的地理属性、功能属性、行政区划级别、建筑类别等属性的词或短语。如地名中的“洞峡/瀑布/岛/涧/湖/高原/湿地/省/市/村/堡/公园/自然保护区/碑/塔/墙”等, 它与我国对地名的标准分类具有一定的对应关系。地名专名和地名饰名是含相同通名的地名命名实体的区别性构件; 其中, 可作为地名简称而独立运用的区别性构件称为地名专名, 如“石家庄市”; 反之称为地名饰名, 如“青县”。

地名标注原则如下:

a. 地名标注的总原则是选取有助于提高地名自动识别、自动聚类能力和检索效率的内容进行标注, 同时为信息抽取做必要准备。

b. 在不打乱原文语法关系的前提下, 将与地名相关的紧凑的处所表达标注出来。

c. 地名识别语料标注范围为地名、地名扩展及句内地名属性。地名和地名扩展分别标注。国家地名代码及句外地名属性暂不标注。

d. 地名需根据相应分类、编码、格式对地名整体及其通名、专名、饰名进行全面标注。

e. 非命名实体的地名扩展原则上不做细标。

分层整体标注原则: 当两层以上地名、组织名嵌套出现时, 通常后者隶属于前者, 此时其间隶属关系缺省不予标注, 但各层需分别标注为地名或组织机构的地名属。

按照地名实体特征对地名通名分类结果详见附表 A. 4。按照文字特征对地名专名分类结果详见附表 A. 5。按照语义关系对地名饰名分类结果详见附表 A. 6。

9.5 组织机构名标注规范

9.5.1 组织机构名相关定义

组织机构：是按照一定的宗旨和系统建立起来的集体。主要包括机关、企事业单位、社会团体、民办非企业单位、国际组织及其内部机构（其他组织定义参照 7）。组织一般侧重于整体指向，机构一般侧重于局部指向，这点区别是相对。

组织机构名：人们给组织机构赋予的专有名称。

标准组织机构名：使用规范的语言文字书写的组织机构名全称。

组织机构名简称：经过官方认可或约定俗成的组织机构名的简单称谓。

组织机构别名：某一组织机构的标准组织机构名及其简称以外的其他现行名称。现今**组织机构名：**目前正在使用的组织机构名。

历史组织机构名：过去曾经使用过、目前已不再使用的组织机构名。

组织机构名扩展：与组织机构名右紧邻、被组织机构名直接或间接修饰，并与组织机构名一起表组织机构的词或短语。如：宝钢销售部门。

9.5.2 组织机构名构件分类、归类标准及其标注式样

组织机构名内部构件可分为组织机构通名、组织机构专名和组织机构饰名三大类。组织机构通名是表示该组织机构类型的构件，如组织机构名中的“公司/集团有限公司/大学/政府/养老院/联合会/局”等，它与组织机构按功能属性分类的结果具有一定的对应关系。组织机构专名和组织机构饰名是含有相同通名的组织机构命名实体的区别性构件。其中，可作为组织机构的简称而独立运用的区别性构件成为组织机构专名。如：国家电网有限公司、清华大学。反之称为组织机构饰名，是组织机构名中通名、专名以外的语词构件。如“北京大学”。

组织机构通名可按照组织机构类型分类，详见附表 A. 7。

组织机构专名可按照其文字特征分类，详见附表 A. 8。

组织机构饰名可按语义功能分类，详见附表 A. 9。

9.5.3 组织机构标注原则

a. 组织机构名标注的总原则是选取有助于提高组织机构名自动识别、自动聚类能力和检索效率的内容进行标注，同时为信息抽取做必要准备。

b. 在不打乱原文语法关系的前提下，将紧邻组织机构名且表组织机构的文字标注出来。

c. 组织机构名识别语料标注范围为组织机构名、组织机构名扩展及句内组织机构名属性。组织机构名和组织机构名扩展分别标注。句外组织机构属性暂不标注。

d. 组织机构名需根据相应分类、编码、格式对组织机构名整体及其通名、专名、饰名进行全面标注。

e. 非命名实体的组织机构名扩展原则上不做细标。

f. 分层整体标注原则：

1. 当两层以上组织机构名嵌套出现时，通常后者隶属于前者，此时其间隶属关系缺省不予标注，但各层需分别标注为组织机构名。

2. 通名不能作为一级机构整体标注，必须与其前一级机构进行整体标注。参见组织机构名通名分类编码使用规则。

g. 组织机构名简称、别称及历史组织机构名部分分别整体标注，组织机构名简称、别称内部暂不作细标，历史组织机构名内部细标。现今组织机构名、组织机构名全称缺省不标。

h. 组织机构名作为其他命名实体构件时，将每层组织机构名整体标注为构件，内部不细标。

- i. 当其他命名实体充当组织机构名构件时，对该命名实体整体标注，内部不细标。
- j. 当句中使用的是组织机构名的组织机构属性时，该属性缺省不标。
- k. 两标原则：当组织机构饰名中的部分与其后组织机构通名构成最小指称、甚至词典词

时：

1. 如果构成的是词典词，则该词典词标为通名，将其前内容标注为组织机构饰名即可。该类词典词代日后对词典进行深化标注时统一处理。
2. 如果构成的是最小指称，则将该组织机构饰名按照最小指称的分割位置分标成两部分，组织机构通名与最小指称照常标注。

9.6 产品名标注规范

9.6.1 产品名相关定义

产品名：狭义产品名是指作为货物（可运输产品）、服务及资产（不可运输产品）名称的狭义命名实体。如：司母戊方鼎。广义产品名是指含有产品专名或品牌、商标、规格等标志性产品饰名的一类货物（可运输产品）、服务及资产（不可运输产品）的专有名称（有些产品、商品、服务的名称本身就是品牌）。

对于产品的货物、资产和服务三大门类十大部类，我们为所有产品设立产品公共属性，为每个产品门类设立门类属性，为每个产品部类设立部类属性，为产品部类下设的各类产品设立产品种类属性。产品公共属性是指所有产品通常都具备的产品属性，产品门类属性是指该门类产品通常都具备的公共属性之外的产品属性；产品部类属性是指该部类产品通常都具备的公共属性和门类属性之外的产品属性；产品种类属性是指该类产品所具备的公共属性、门类属性和部类属性制外的产品属性。

9.6.2 货物饰名分类、归类标准及其标注式样

货物饰名是指货物专名中修饰货物通名的词或短语。详见附表 A. 10。

9.7 品牌标注规范

“品牌”不是商标，更不是产品，是企业或品牌主体（包括城市、个人等）一切无形资产总和的全息浓缩，而“这一浓缩”又可以以特定的“符号”来识别；它是主体与客体，主体与社会，企业与消费者相互作用的产物。

品牌符号是区别产品或服务的基本手段，包括名称、标志、基本色、口号、象征物、代言人、包装等。这些识别元素形成一个有机结构，对消费者施加影响。它是形成品牌概念的基础，成功的品牌符号是公司的重要资产，在品牌与消费者的互动中发挥作用。

在命名实体标注研究中，品牌能够作为处理对象的只能是品牌符号，而且只能是品牌符号中的名称、口号、代言人，更实际地讲，只是名称。而组织机构名是作为独立的命名实体进行标注，商标名本身很少单独出现，多数是出现在产品名中，这样在产品名中商标已经被标注了。如果仍嫌不足，可以将商标独立作为标注对象。

9.8 事件标注规范

9.8.1 事件相关概念

事件名：主述结构的，文字连续的短语或句子。事件名具有嵌套性。

主语：具体事物，人/动物/植物/自然物及其部件：大熊猫赠港，猪肉涨价等。

机构团体：政府采购流标公司状告财政部一审胜诉，澳大利亚向中国移交没收的吉林疑犯赃款。注：不一定是命名实体，即人名/地名/组织机构名

述语：动词及其组成部分。

事件动词：在语义上表示动态行为的动词，包括下述 13 类：

- 1 变化（change）：死 病 下降 长高 缩小 变暗

- 2 气象 (weather): 下雨 刮风 打雷 起雾
- 3 身体活动 (bodily care and functions): 蹬 跳 推 笑 咳嗽 游泳
- 4 五官感觉 (perception): 看见 听到 闻着 品尝
- 5 消耗 (consumption): 吃 喝 饮
- 6 位移 (motion): 跑 走 散步 飞 过来 回去 拉来
- 7 创造 (creation): 制作 画 炒 写 创建 修筑
- 8 接触 (contact): 触摸 撞击 打中 系 挖掘
- 9 领属转移 (possession): 买 卖 赠送 给 转让 借
- 10 信息交流 (communication): 告诉 询问 请求 转达 叮嘱 说
- 11 比赛 (competition): 竞赛 赛跑 打仗 摔跤 辩论
- 12 社会活动 (social behavior): 改革 调价 开会 联欢
- 13 其他行为 (other event)

9.8.2 事件的命名方式

考虑字长的原因，事件不一定由事件名来命名。

命名方式：用文中一组强文本表示功能的词或短语来表示事件。

{参与者, 事件动词, (时间), (地点), (子事件)}

{Participants, Event Verb, (Time), (Location), (sub-event)}

Participants: 事件的参与者。名词和基本名词短语。

Event

Verb: 表示动态行为的动词

Time: 事件发生的时间

Location: 事件发生的地点

9.9 时间结构标注规范

9.9.1 时间结构相关定义

时间结构：用纪时法（广义）纪录时间的完整实词短语。具有四个特征：1. 时序性：所表时间具有指定性，指定该时间在时间之流中的位置；2. 规范性：通过通用、规范的纪时法表达时间，而不是通过一般叙事来表达时间；3. 实词短语型：是不包含虚词的短语；4. 完整性：是文中完整的时间表达。

纪时法：纪年法、纪月法、纪日法、纪时法（狭义）等纪录时间的系统、规范的方法。

时间结构与时间性数量结构的区别：时间性数量结构是表达时间长短的数量结构。时间结构所表时间具有指定性，即指定该时间在时间之流中的位置，而时间性数量结构表达时间不具有指定性，仅表示时间长短。如：3 个月、两星期。

时间结构与时间词的区别：时间词是一个实词，时间结构是一个实词短语。

9.9.2 时间结构分类及归类标准

根据表达时间的特定程度，时间结构可分为绝对时间结构和相对时间结构。绝对时间结构是表绝对时间的的时间结构。如：2007 年、2007 年 5 月 5 日凌晨 1 点。相对时间结构是表相对时间的的时间结构。该短语或者缺省绝对时间的限制，如：9 点 50 分、5 月 5 日；或者以相对时间代替绝对时间表达时间限制，如：当年 5 月、本月 12 日。

时间结构根据表达形式可分为五类：1. 单点叙述式；2. 单点符号式；3. 起止叙述式；4. 起止符号式；5. 多点缩略式。详见附表 A.11。

9.9.3 时间结构构件分类及其标注式样

时间结构构件分类及其标注式样详见附表 A.12。

9.10 数量结构标注规范

9.10.1 数量结构相关定义

数量结构：指形式完整的数量短语或含名数量短语。如：八九个人。

含数量短语：包含了被修饰核心名词的完整数量短语。如：20 个人出头。

数词：表示事物的数目或顺序的词。如：一、第一、-1.5。

量词：表示事物、动作或时空计量单位的词。如：个、位。

9.10.2 数词细分与示例

根据数词的构成特点及其与量词的搭配特点，将其分为三类：单纯数词、兼量数词、复合数词。复合数词是由两个以上相邻的单纯数词构成的完整数词短语。数词细分结果详见附表 A.13。

9.10.3 数量结构分类

根据数量结构的构成特征，将数量结构分为定指数量结构（表确定值的完整数量结构）、概指数量结构（表不确定值的完整数量结构）和含名数量结构（包含了被修饰核心名词的完整数量结构）三类。详见附表 A.14。

9.11 数码标注规范

9.11.1 数码定义

数码是根据统一编制规则产生的，由数字（必备）辅以字母、符号、汉字等组成的，表达数字以外含义的代码。

数码与数词的根本差别在于数码具有数字以外的含义或所指，可指人、指物、指事等，是一种代码。如身份证号、学号、单位代码、电话号码、专利号码、国家标准号码、图书 ISBN 号码、期刊 ISSN 号码等都是数码。

9.11.2 数码的分类

根据数码与其他命名实体间的构成关系，数码可分为两类。一类是数码本身就是其他命名实体或其构件（专名、饰名或通名），例如，数码指人的化名，如：007；指组织，如 91765（部队番号）；指地点，如：798；指产品，如 101；指时间，如：2007-05-01。我们将这类充当其他命名实体或其构件的数码称为别名数码。

另一类数码称为独立数码，是与其他命名实体不具有构成关系的数码。独立数码包括很多种类，如身份证号、学号、单位代码、电话号码、专利号码、国家标准号码、图书 ISBN 号码、期刊 ISSN 号码等。我们将根据具体应用需求来设定我们要标注的独立代码种类。

根据编制规则，典型独立数码的种类有以下三种：

1. 顺序码：又称系列码，是一种用连续数字代表编码对象的码，例如，用 1 代表男性，2 代表女性。

2. 区间码：区间码是把整个编码分成多个分组，形成多个区间，每个区间是一组，每组的码值和位置都代表一定意义。典型的区间码是邮政编码。区间码又可分为以下三种类型：

a) 多面码。一个数据项可能具有多方面的特性。如果在码的结构中，为这些特性各规定一个位置，就形成多面码。

b) 上下关联区间码。上下关联区间码由几个意义上相互有关的区间码组成，其结构一般由左向由排列。

c) 十进制码。此法相当于图书分类中沿用已久的十进制分类码，它是由上下关联区间码发展而成的。如 610.736，小数点左边的数字组合代表主要分类，小数点右边的指出子分类。

3. 自检码：由原来的代码（本体部分）和一个附加码组成。附加码用来检查代码的录入和转录过程中是否有差错，附加码又叫校验码，它和代码本体部分有某种唯一的关系，它是通过一定的数学算法得到的。

9.11.3 数码标注规则

- 1) 别名数码依据各类命名实体的标注规范进行标注。
- 2) 独立数码进行整体标注，基本标注代码为：ma。对于未被列入标注范围的独立代码，均不予以标注。
- 3) 当独立代码充当其他命名实体属性或与其他命名实体发生关系时，按照各自的标注范式进行标注。

附录 A
(资料性附录)

表 A.1 按文字特征人名分类表

类名	定义	归类标准	标注示例	备注
汉字人名	汉语人名	姓名以汉语语义命名的人名。	汉族、回族人的汉名及其他汉语汉字人名。 [陈辉]	含外国人取的中文名
	汉语化名	含化名构件的汉语汉字人名。	依定义。 [白玫瑰小姐]	
	汉语虚拟人名	超自然的汉语人物专名。	依定义。 [孙悟空]	
	类汉语人名	姓名符合汉字方式的外国人名。	包括新加坡、越南、朝鲜、韩国的非音译汉字人名。 [李惠利]	
	少数民族汉字人名	包含少数民族人汉字姓名的人名。	依定义。 [爱新觉罗·胤禛]	
	少数民族译名	包含中国少数民族汉译姓名的人名。	依定义。 [卓玛]	
	日本人名	包含日本人姓名的人名。	依定义,无论其是否包含日文假名。 [坂井泉水]	
	外国人音译名	包含外国人汉译姓名的人名。	依定义。 [爱因斯坦]	
非汉字人名	姓、名构件非汉字的真实人名	依定义。 [TOM]	含中国人取的外语名字	
	姓名、化名非汉字的非真实人名	依定义。 [SNH48]		
混合人名	姓名、化名含汉字及非汉字的人名	依定义。 [Tony 张]		

表 A.2 汉字人名称谓构件类型

类别	诠释	举例	备注
化	笔名、化名等	心吾	
性	主表性别的称谓	王女士、张小姐	
帝	帝王帝号	唐太宗	
字	古人字、号、谥号、爵号等	关云长	
讳	代替名的称谓	张氏、吴某	
敬	主表敬称	李老、吴总、王先生	
名	男	男性名	关羽
	女	女性名	关琳

昵		主表亲昵、熟识的称谓	老王、华仔、小刘、阿紫	
亲		亲族辈分的称谓	李阿姨	
序		次序性称谓	老三、二爷	
诨		外号、绰号、诨称	孙小妖	
姓		姓	慕容、司马	
译		汉译姓名	卡尔·马克思	
职	职业	就业工作所属职业	周编辑	
	职务	就业工作职务	周处长	
	职称	就业工作职称	周教授	
	学历	学历、学位	周博士	
	学职	在校就读职务	周学委	
符		人名中的标点符号	马克·吐·温	
其他		其他称谓类型	王犯人、李同志	

表 A.3 职名分类表

职名类别	职名构件		定义、标准	标注示例
职务	职务通称		表基本职务名称的词	[副主席]、[主席]
	职务 饰 词	业务饰词	表职权的业务范围	[销售主任]、[业务经理]
		等级饰词	表职务级别的接首词	[副主席]、[总经理]
		组织机构通名饰词	以组织机构通名表示职权范围的职务饰名	[销售部主任]、[大学校长]
		其他饰词	其他职务饰词	[名誉教授]、[独立主事]
职业	职业通称		表基本职业名称的词	[大学教师]、[教师]
	职业 饰 词	组织机构通名饰词	以组织机构通名表示任职单位类型的饰词	[大学教师]
		其他饰词	其他职业饰词	[兼职教师]
职称	职称通称		表基本职称名称的词	[助理教授]
	职称饰词		职称通称的饰词	[副教授]、[客座教授]

表 A.4 汉语地名通名分类表

类别	常用通名
洞域通名	涧、洋、涧沟、涧底峡谷、涧渊、涧槽、涧盆、洼地、涧底平原、涧底扇、涧底高原、涧隆、涧底平台、涧山、涧岭、涧丘、大陆架、大陆坡、海湾港湾、洞峡、沙、暗沙、滩、暗滩、礁、暗礁、岛屿、群礁、群岛、列岛、泻湖、卸角、沙嘴、洞滩、洞涂、洞岸、半岛、洞岸带
水系通名	江、河、湖、川、溪、沟、水、泉、河源、河谷、峡谷、河滩、边滩、河湾、河口、三角洲、洲、河岛沙洲、江心洲、湖岛、矾、冰川、冰盖、冰斗、冰川谷、瀑布、流域

陆地地形通名		平原、低地、洼地、盆地、高原、丘陵、山地、山口、关、垭口、山谷、谷地、山坡、山麓、崖、峰、顶、梁、火山、山系、山脉、沼泽、湿地、苔原、草原、森林、丛林、沙漠、沙地、戈壁、绿洲、洞、沼泽、湿地、苔原、草原、森林、丛林、沙漠、沙地、戈壁、绿洲、洞、(天生)桥、台地、黄土
(准)行政区域通名		省、市、自治区、行政区、地区、自治州、盟、县、区、自治县、旗、林区、工业区、特区、乡、民族乡、辖区；油田、煤田、农区、林区、牧区、渔区、开发区、边贸区、港(口岸)、军区、地片、区片、小区、村、镇、庄、庄子、屯、堡、村片
住点通名		楼、单元、门
交通设施通名	水上交通设施通名	描地、泊地、润港、河港、船闸、升船机站、渡口
	陆路交通设施通名	公路、长途汽车站、收费站、铁路、火车站
	航空与管道运输通名	(航空)港、管道、管站
	城镇交通运输通名	道、路、街、巷、胡同、地铁、轻轨、线、(公共交通)站、停车场
	交通运输附属设施通名	(公路、铁路)桥梁、(城镇内部的)立交桥、过街桥、(隧道交通用)涵洞、(城镇内部的)地下通道、道班、检查站、边境检查站、环岛、路口、加油站、灯塔、导航台、公路出入口
	其他交通设施通名	索道、扶梯
水电通信设施专名	水力设施通名	井、池塘、涧塘、水库、人工湖、蓄洪区、泻洪区、灌区、(灌溉)渠、排水沟、排水管道、排污管道、渡槽、泵站、排灌站、扬水站、(输水)涵洞、涧堤、河堤、湖堤、闸坝、拦河坝、运河
	电力设施通名	输变电线路、发电站、输变电站
	通信设施通名	通信线路、通信基站、卫星地面接收站
构筑物、场地通名		(人物、事件)纪念地、纪念馆、纪念堂、(古)战场、故居、陵园、寺、庙、教堂、公园、风景区、自然保护区、亭、台、碑、塔、广场、体育场、城堡、墙、长城、场、地
其他地名通名		基地、示范园、科技园

表 A.5 按文字特征地名专名分类表

类名	定义	归类标准	标注示例	备注	
汉字地名专名	汉语地名专名	以汉语语义命名的地名专名。	汉语及汉语义译地名专名	北京	汉族主要居住区地名均可依此标
	类汉语地名专名	新加坡、越南、朝鲜、韩国的汉字地名专名	包括其汉字原名及汉译名	汉城	1. 民族、国家代码均可暂代标为其国家汉字简称，不知者可暂代标“.”；2. 少数民族自治区地名均可依少数民族
	少数民族音译地名专名	以少数民族语言命名的汉译地名专名。	依定义	吐鲁番	

	外国音译地名 专名	非新、越、朝、韩国 的汉译地名专名。	依定义	柏林	族音译地名专名标。
非汉字地名专名		非汉字形式的地名专名。	依定义	London	

表 A. 6 地名饰名分类表

地名饰名类别		定义	归类标准	标注示例
地 名	移植地名	该地名与所指地 无地缘关系	以城镇名为饰名者 入此	南京路
	根源地名	该地名与所指地有地缘关系	以非城镇地名为饰名者入此	复兴门
	无饰建构	该建构物、地点通 名前无饰词	词义为建构物、地点类别者 入此	车站街
	有饰建构	建构物、地点通名前 有饰词		新车站街
机 构	组织机构	表组织机构名	依组织机构名定义	七建大楼
	机构通名	表组织机构通名	其饰词亦入此	省政府街
方位		表示相对方向与位置（东南西北 中前后内外心口头沿环临等）	须具有明示的相对物	环城路 湖心路
拟形		表示其形状特征	无相对物，仅描述自 身形状	月字街 日月潭
产 品	无饰产品	由无饰词的商品、产品、服务充 当	依定义	茶儿胡同
	有饰产品	由有饰词的商品、产品、服务充 当	依定义	木马街
人 名	人名	由人名充当	依定义	灰姑娘街
	人名属性	表人人名属性	依定义	尚书村
数 量	数名	由地名通名前数名结构充当	依定义	九龙山
	数量	由地名通名前数量结构充当	依定义	半步桥
编号		由直接修饰通名的编码充当	依定义	2 又 1/2 街
功用		表该地功能作用	望文生义即可	校务大楼
叙事		起叙事作用（多具政治色彩）	须含动词	反修路 登云坡巷
吉励		做祝吉、勉励用	望文生义即可	吉利大厦
颜色		表颜色	依定义	白石桥
时间		表时间	依定义	唐朝碧阁
领域		表领域	依定义	文化东路
动物		表动物	做交易用的动物按产品标， 如：骡马市	大黑虎胡同
植物		表植物	做交易用的植物按产品标， 如：花市	梅园 劲松楼

自然物	表自然现象	依定义	电子胡同
其他	其他	依定义	懒惰的 Y 街

表 A.7 组织机构通名分类表

类别	定义	归类标准	标注示例
企业单位通名	企业法人单位通名	依定义	某集团公司
非企业单位通名	非企业法人单位通名	依定义	某大学
部门通名	法人单位的下设部门通名	依定义	某单位财务处

表 A.8 按文字特征组织机构专名分类表

类别		定义	归类标准	标注示例	备注	
汉字 组 织 机 构 专 名	词典词组织机构专名	由单一词典词充当的完整组织机构专名	依定义	联想	部分组织机构专名已录入词典	
	非词典词组织专名	汉语组织机构专名	汉语或汉语义译组织机构专名。	专名词义可望文生义理解者入此	惠普	音译义译结合者均入此
		少数民族组织音译专名	中国少数民族语言组织机构专名的音译名。	依定义	略	民族、国家代码可暂代标为其国家汉字简称，不知者可暂代标“.”
		外语组织机构音译专名	外语组织机构专名的音译名。	依定义	略	
非汉字组织机构专名		非汉字形式的组织机构专名。	依定义	IBM		

表 A.9 按语义功能组织机构饰名分类表

类别	定义	分类标准	示例
地名饰名	由地名充当的组织机构饰名	依定义（注：含表组织机构属性的地名）	浙江保一
空间饰名	表空间范围、地点方位意义的组织机构饰名	依定义	国际航空公司
时间饰名	表时间、时代意义的组织机构饰名	依定义	冠福现代家用公司
服务饰名	表经营性服务内容的组织机构饰名	兼表行业、产品或服务的饰名优先归入企业产品或服务饰名。符合“该企业生产 / 提供的 A 很好”句式，则 A 为产品或服务饰名，否则为企业行业饰名	国际航空公司
产品饰名	表企业相关产品的组织机构饰名		伟业陶瓷
企业行业饰名	表企业所属行业、领域的组织机构饰名		华强照明公司

非企行业 饰名	表行业或领域的非企业组 织机构饰名	兼表行业、学科、技术的 饰名，在行业性社会团体 命名实体中优先入非企 行业饰名，其他情况优先 入学科或技术饰名	物流与采购联合会
学科饰名	表事业单位、社会团体所涉 及学科的组织机构饰名		中华美学学会
技术饰名	表专业技术性社会团体的 研究对象		遵义辣椒协会
事务饰名	表机关、事业单位或其他组织机 构下设部门所管事务的组织 机构饰名	依定义	国家防汛抗旱总指挥部
成员饰名	表组织机构主要成员的组 织机构饰名	依定义	中国电视艺术家协会
人才饰名	表组织机构服务对象或培 养目标的组织机构饰名	依定义	检察官学院
人名饰名	由人名充当的组织机构饰名	依定义	宋庆龄基金会
组织机构 名饰名	与该组织机构没有隶属关 系的组织机构饰名	依定义	黄埔军校同学会
隶级饰名	表组织机构隶属级别关系 的组织机构饰名	依定义	首钢总公司
编号饰名	作为组织机构饰名的编号	依定义	照明一厂
符号饰名	指组织机构名中出现的标 点符号	依定义	恒强（福建）鞋塑公司
其他饰名	未被列入组织机构饰名细 类的其他饰名	依定义	恒强（福建）鞋塑发展有限公司

表 A. 10 物饰名表

名 称	定 义	举 例	
货物用途饰名	表示货物使用的方向、场所、 领域的饰名	车用润滑油、民用计量表、钢筋混凝土用热轧带 肋钢筋、建筑防水卷材、食用植物油、户外照明灯 具、通用粉，专用粉、男式西服、胶版印刷纸、 内外墙乳胶漆	
货物功能饰名	表示货物自身功能、性 能特征的饰名	如 采暖散热器、建筑防水卷材、户外照明灯具、 智能机器人、即溶咖啡、无滴棚膜、数控铣锐 床	
货物化学属性饰 名	表示货物化学属性的 饰名	酸性染料、可还原染料	
质料饰名	表示货物材质、原料、配料、 成分的货物饰名	不锈钢餐具、茶饮料、高压钠灯、金属卤化物灯、 纯羊绒衫、无磷洗衣粉、复合肥、釉面砖、芳香 族类杀菌剂	
	动能饰名	表示货物动力、能源来源的饰 名	电动剃须刀、手动晾衣架、真空管太阳能热水器、电 辅助加热型热水器、全自动洗衣机、光电互补 型热水器

货物物理属性饰名	颜色饰名	表示货物颜色的饰名。	红葡萄酒、白砂糖
	包装饰名	表示货物包装形式的饰名	瓶装矿泉水
	形状饰名	表示货物物理形状、构造特征的饰名	热轧带肋钢筋、皱纹卫生纸、双桶洗衣机、 <u>波轮</u> 全自动洗衣机、 <u>嵌入式</u> 照明灯、台式电脑、球轴承、滚子轴承、双端荧光灯、水性涂料、便携式计算机
	其他物理属性饰名	表示货物其他物理属性的饰名	中密度板、软木制品、硬质纤维板、高纯阴极铜、实木木、高速平带、同步齿形带
货物指标饰名	规格饰名	表示货物规格、型号的饰名	F248A 系列节能灯、 <u>幼儿</u> 配方乳粉（400g/袋 900克/听）、5-7 号燃料油
	性能指标饰名	表示货物具体性能指标的饰名	60W、220V、50Hz 节能灯
	系列饰名	非规格、性能指标的序列饰名	<u>P</u> 系列分散染料
货物标牌饰名	货物商标饰名	货物商标（商标是生产经营者在其生产、制造、加工、拣选或者经销的商品或者服务上采用的，区别商品或者服务来源的，由文字、图形或者其组合构成的，具有显著特征的标志）	<u>飞雪牌</u> 特精粉、 <u>小天鹅</u> 滚筒洗衣机
	货物品牌饰名	货物的非商标品牌、牌号	长虹电器
货物工艺饰名		表示货物生产、加工工艺的饰名	钢筋混凝土用热轧带肋钢筋、 <u>注塑</u> 机械、 <u>璇切</u> 薄木、手工水饺
货物等级饰名		表示货物质量等级的饰名	一级菜籽油、 <u>溶解级</u> 化学木浆
货物用法饰名		表示货物使用方法的饰名	含服舒咽片
货物字母饰名		不能归入以上各类的字母型、数字饰名	建筑（ <u>PVC</u> ）型材、 <u>PET</u> 瓶装茶饮料
其他货物饰名		其他货物饰名	半化学木浆、普通板材

表 A.11 时间结构表达形式分类

类名		归类标准	标注样例
单点式	单点叙述式	同时满足单点式和叙述式时间结构定义	2007 年 2 月 10 日 8 点半
	单点符号式	同时满足单点式和符号式时间结构定义	2007-2-10 8: 30 2007.3 五·一
起	起止叙述式	同时满足起至式和叙述式时间结构定义	2007 年~2008 年 3 月~5 月

止式	起止符号式	同时满足起至式和符号式时间结构定义	8:30~10:00 2007.3~2007.4 2007.8~9
多点缩略式		据其定义	星期一、二

表 A. 12 时间结构构件分类及其标注式样

类名		归类标准	举例	
时序量词	现今时序量词	基本时序量词	依定义	2007 年
		复合时序量词	依定义	2007 赛季
古代时序量词		依定义	清代、五更	
时序数词	完整时序数词	纯汉字型	依定义	前二二一年、初五、第二周
		混合型		前 221 年
	缩略时序数词	汉字型	依定义	二零零七年、五·一、二〇〇七年、8 点半、零七年
		阿拉伯数字型		2007 年、07 年、2007.3、2007-2-10、8:30
	干支时序数词		依定义	辛亥
	地支时序数词		依定义	子时
纪时法名称		依定义	公元、星期	
历法名称		依定义	农历、阳历	
纪时专名	帝号	王公即位年次纪年法中王公、皇帝的名号	鲁僖公、赵惠文王	
	年号	年号纪年法中的年号	至和元年七月某日	
	国号	各种纪年法中的国号、朝代名	民国二十五年、清顺治二年	
	时区	时区名称	东京时间 15 点整	
	节气	二十四节气名	去年立秋	
	时节	季节与时节称谓	2004 年仲春	
	节日	非事件性、非数字型节日名称	2008 年春节	
	天色	天色纪时法中的天色名称	次日平明	
	月相	月相纪日法使用的月相称谓	七月望日	
	月份别称	十二个月份的别称	戊戌年桂月	
	生肖		马年	
地质纪年		白垩纪		
相对纪时词	相对时间词		昨天、今天、去年、前年、大前年、后年、大后年、明年	
	替代词		当地时间	
	指代词		这周五	
	时间方位		初、中、末、底、终、来（以来）、间、年间、期间、前、以前、上（上上）、下（下下）	
日区间词			凌晨、上午、晚上	
隐讳词			上月某日	
符号			“~”、“”、“!”、“.”	

非汉字	December 2003
-----	---------------

表 A. 13 数词细分表

数词细分		定义	示例
单纯数词	基数词	表示数目多少的数词	5、5.1、1/2、2%、-1
	序数词	表示顺序的数词	第一、首、二版、2nd
	约数词	不以数字、数位表数的数词，具约略型、独立性	多、若干、些微、许许多多
	助数词	自身不独立表数，只能辅助数字、量词等表示不确定数量的数词	来、余、头两个、出头、左右
	残缺数词	表数的不完整结构	三圣之一
兼量数词		独立的数词本身兼表数、量双义	俩、仁
复合数词	整数复合数词	以整数为表数主体的完整复合数词	四、五人
	分数复合数词	以分数为表数主体的完整复合数词	三分之一、二
	小数复合数词	以小数 of 表数主体的完整复合数词	一点儿倍
	序数复合数词	以序数词为表数主体的完整复合数词	第一、二名、七、八名（名次）
	兼量复合数词	以兼量数词为表数主体的完整复合数词	俩仁人
	定模复合数词	包含数字、数位的定模型数词	千千万万、成百上千
	区域复合数词	以“~”、“-”连接表示数的范围的复合数词	1~10、4-6
	乘接型数量结构	数字之间以“*”或“×”连接表特定比例关系的复合数词	180×90×45
	比例型数量结构	数字之间以冒号连接表特定比例关系的复合数词	2:0
	其他复合数词	其他复合数词	999、八八六十四

表 A. 14 数量结构细分表

数量结构细分		定义	示例
定指数量结构	基数定指数量结构	表示确定数量的完整数量结构	五十个
	序数定指数量结构	表示确定数量顺序的完整数量结构	第三位、三版、首名
概指数量结构	基数概指数量结构	以基数词为表数主体的完整概指数量结构	七八个
	序数概指数量结构	以序数词为表数主体的完整概指数量结构	第三四名；第三、四名
	约数概指数量结构	以约数词为表数主体的完整概指数量结构	多名、数年、近几年、斤把重
	定模型数量结构	以定模为表主体的完整概指数量结构	千千万万名、一片片、一片一片
含名数量结构		包含了被修饰核心名词的完整数量结构	20 人左右、30 人出头