



# 中华人民共和国国家标准

GB/T 44217.11—2024/ISO 24617-11:2021

## 语言资源管理 语义标注框架 第 11 部分：可度量数量信息 (MQI)

Language resource management—Semantic annotation  
framework—Part 11: Measurable quantitative information (MQI)

[ISO 24617-11:2021, Language resource management—  
Semantic annotation framework (SemAF)—Part 11:  
Measurable quantitative information (MQI) IDT]

2024-07-24 发布

2025-02-01 实施

国家市场监督管理总局  
国家标准化管理委员会 发布

## 目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 QML 的抽象规范	3
4.1 概述	3
4.2 QML 的特点	3
4.3 元模型	3
4.4 QML 的抽象语法(QML_as)	4
4.5 QML 及其子集的具体语法(QML_cs)	5
5 QML 基于 XML 的具体语法(QML_csx)	5
5.1 概述	5
5.2 带 ID 前缀的标签名称	5
5.3 根(MQI)的属性说明	6
5.4 基本元素类型的属性说明	6
5.5 链接类型的属性说明	6
5.6 QML_csx 的说明	7
6 QML(QML_cst)基于 TEI 的具体语法	9
6.1 QML(QML_cst)的具体语法	9
6.2 QML_cst 的说明	10
附录 A (资料性) QML_csx 示例说明	13
附录 B (资料性) 可度量数量信息的不规则表达形式	15
附录 C (资料性) 单位表示	16
参考文献	17

## 前 言

本文件按照 GB/T 1.1—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》的规定起草。

本文件是 GB/T 44217 的第11部分。GB/T 44217 已经发布了以下部分：

——第6部分：语义标注原则；

——第11部分：可度量数量信息(MQI)。

本文件等同采用 ISO 24617-11:2021《语言资源管理 语义标注框架(SemAF) 第11部分：可度量数量信息(MQI)》。

本文件做了下列最小限度的编辑性改动：

——为与系列标准协调，更改了标准名称，删除了语义标准框架的简称；

——解释性注释的表示方法由原文的{\*…\*}，改为#..；

——改正4.5中印刷错误，将 ISO 24617-11:2021 原文中的“csf”更正为“cst”。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国语言与术语标准化技术委员会(SAC/TC 62)提出并归口。

本文件起草单位：南方电网科学研究院有限责任公司、中国标准化研究院、华南师范大学、湖北省标准化与质量研究院、厦门三行电子有限公司、中国科学技术信息研究所、北京信息科技大学、中国质量标准出版传媒有限公司、聊城大学、北京工业大学、上海对外经贸大学、中国中医科学院中医药信息研究所、广州智语信息科技有限公司、贵州电网有限责任公司兴义供电局、四川语言桥信息技术有限公司、北京集贤弘文文化传媒有限公司。

本文件主要起草人：郝天永、王海涛、王昕、陈炎明、曹馨宇、魏洁、周育忠、黄景明、刘耀、吕学强、鲁曦、徐术坤、刘晓东、贾仰理、刘磊、刘亮亮、周洪伟、石嘉豪、刘润鹏、刘宁畅、瞿瑛瑛、朱宪超、贺莉丽。

## 引 言

语义标注是计算机对自然语言深层次处理的重要技术之一,是对文本中的词语或句子添加可供理解的语义标签的过程。依据标注的一般原则和具体标注对象的不同,标准被划分为不同的部分,GB/T 44217《语言资源管理 语义标注框架》拟由 12 个部分构成。

- 第 1 部分:时间和事件。目的在于提供一种通用的方法来描述文本中的时间和事件。
- 第 2 部分:对话行为。目的在于提供一种表示对话行为的标注语言以及一种将对话分割为语义单元的方法。
- 第 4 部分:语义角色。目的在于为语义角色提供一个协商一致的标注方案。
- 第 5 部分:篇章结构。目的在于为话语实现和话语内容提供一种表示方式。
- 第 6 部分:语义标注原则。目的在于确定以语义标注框架为特征的语义标注方法。
- 第 7 部分:空间信息。目的在于提供一种通用的方法来描述自然语言文本中表达运动相关的空间信息和时空信息。
- 第 8 部分:篇章中的语义关系,核心标注框架。目的在于为话语关系的表示和标注提供一个方案。
- 第 9 部分:引用标注框架。目的在于为自然语言文本和多模态交互中所指现象的标注和表示提供一个综合模型。
- 第 11 部分:可度量数量信息(MQI)。目的在于为可度量数量信息提供一种标注方案。
- 第 12 部分:数量。目的在于为数量信息语义表示提出一般形式化定义。
- 第 14 部分:空间语义。目的在于通过为抽象语法建立形式语义提供标注空间信息的方法。
- 第 15 部分:可度量数量信息抽取。目的在于提供一种从自然语言文本中抽取可度量数量信息的一般方法。

可度量数量信息(MQI),如“165 cm”或“60 kg”可用来描述人的身高或体重,其在常见语言表述中普遍存在。MQI 描述的是与量的大小方面相关的基本属性之一,它主要特点是,数量信息是以一对  $\langle n, u \rangle$  表示的数量来呈现的,由一个数字表示的量  $n$  和一个单位  $u$  组成,单位  $u$  可以是基本单位,也可以是派生单位,还可以是规范化单位,也可以是传统使用的单位。此类信息在科学出版物或技术报告中更为丰富,以至于构成了一般语言交际片段的重要组成部分。因此,任何成功的语言资源管理都需要对这些信息进行处理。

在这样一个大数据时代,产业界和学术界对准确抽取 MQI 的要求越来越高。例如,商业投资公司经常需要从年报中识别并汇总目标公司的净销售额、毛利润、营业费用、营业利润、利息支出、税前净利润、净收入等各种信息。快速发展的医学信息学研究也需要处理大量的医学文本,以分析药物的剂量、临床试验的纳排标准、患者的表型特征、临床记录中的实验室检查等。无论是在工业领域还是在医学研究领域,这些需求都需要准确、一致地表示 MQI,以便进行自动处理、计算和交换。

然而,在信息检索和自然语言处理领域,目前还没有标准化的方法来表示可度量数量信息。迄今为止,工业领域开发的应用系统通常使用自己的格式来标注可度量数量信息。我们需要一个通用的、可互操作的和标准化的可测量定量信息表示方法,以便与不同应用系统协同工作。本文件旨在根据 ISO 24617-6 规定的语义注释原则和 ISO 24611 的基本要求,制定一个通用标注框架,以使用科学的技术语言表示 MQI,并使其与 ISO 24617 等其他语义标注方案具有互操作性。它还利用了 ISO 有关词法资源和形态句法标注框架的各种标准,并与其他现有相关标准兼容。

注:例如 ISO 24617-1 和 ISO 24617-7 分别提出了时间(持续时间或时间量)和空间(距离)度量的标注方法。

ISO 24612 提供了一种图形标注框架,可使用这两种标注方法对时间或空间度量进行标注。

QML 在抽象层面进行了规范化,允许采用各种序列化格式来表示可度量数量信息的标注,如基于 XML 的表示法。本文件在注释的抽象层面对数量信息标注进行了规范,在序列化的具体层面采用了独立标注格式。

本文件聚焦科技语言中的数量信息,预计将有助于信息提取(IR)、问题解答(QA)、文本摘要(TS)和其他自然语言处理(NLP)应用。

# 语言资源管理 语义标注框架

## 第 11 部分：可度量数量信息 (MQI)

### 1 范围

本文件规定了信息检索、问答、文本摘要和其他自然语言处理应用中可度量数量信息的语义表示。本文件适用于与语言相关的技术或实践，其他相关技术领域参照使用。

本文件还涉及 ISO 24617-1 中讨论的时间持续问题，以及 ISO 24617-7 中处理的距离等空间度量，同时使它们与其他度量类型具有互操作性。本文件还包括 ISO 24617-6:2016 的 8.3 中引入的度量或数量的处理。

注：ISO 24617-12 涉及量化和定量信息的更一般理论问题。

### 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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。

ISO 24612 语言资源管理 语言标注框架 (LAF) [Language resource management—Linguistic annotation framework (LAF)]

### 3 术语和定义

下列术语和定义适用于本文件。

国际标准化组织 (ISO) 和国际电工委员会 (IEC) 在以下地址维护标准化工作中使用的术语数据库：

——ISO 在线浏览平台：<https://www.iso.org/obp>

——IEC 电子百科全书：<https://www.electropedia.org/>

#### 3.1

##### 量 quantity

涉及度量对象大小或数量的属性。

[来源：ISO/IEC Guide 99:2007, 1.1, 有修改]

#### 3.2

##### 基本量 base quantity

在给定量制中约定选取的子集中的量 (3.1)，其中子集中的任何量均不能用该子集中的其他量来表示。

注：量的种类包括国际量制 (ISO) 中定义的七种基本量。

[来源：ISO/IEC Guide 99:2007, 1.4, 有修改]

#### 3.3

##### 派生量 derived quantity

量制中由基本量 (3.2) 定义的量 (3.1)。