

ICS 01.020
CCS A 22



**NATIONAL STANDARD OF THE
PEOPLE'S REPUBLIC OF CHINA**

GB/T 44217.11-2024

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

语言资源管理 语义标注框架 第11部分：可度量数量信息（MQI）

(English Translation)

Issued on: July 24, 2024

Implemented on: February 1, 2025

**Issued by: State Administration for Market Regulation, China National
Standardization Administration**

Table of Contents

Preface	
Introduction	
1 Scope	
2 Normative references	
3 Terms and Definitions	
4 QML abstract specification	
4.1 Overview	
4.2 Features of QML	
4.3 Metamodel	
4.4 QML Abstract Syntax (QML_as)	
4.5 Concrete syntax of QML and its subsets (QML_cs)	
5 QML XML-based concrete syntax (QML_csx)	
5.1 Overview	
5.2 Tag names with ID prefixes	
5.3 Attribute Description of Root ?MQI?	
5.4 Attribute Description of Basic Element Types	
5.5 Link Type Attribute Description	
5.6 Description of QML_csx	
6 QML (QML_cst) concrete syntax based on TEI	
6.1 Specific syntax of QML (QML_cst)	
6.2 Description of QML_cst	
Appendix A (Informative) QML_csx Example Description	
Appendix B (Informative) Irregular Expressions of Measurable Quantity Information	
Appendix C (Informative) Units	
References	

Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for standardization work Part 1: Structure and drafting rules for standardization documents" Drafting is required.

This document is Part 11 of GB/T 44217.GB/T 44217 has been published in the following

parts.

--Part 6: Principles of semantic annotation;

--Part 11: Measurable Quantity Information (MQI).

This document is equivalent to ISO 24617-11:2021 "Semantic Annotation Framework (SemAF) for Language Resource Management Part 11: Measurable Quantity Information (MQI) the following minimal editorial changes were made to this document.

--In order to coordinate with the series of standards, the standard name was changed and the abbreviation of semantic standard framework was deleted;

--The expression method of explanatory comments is changed from the original {**} to Lu Xi, Xu Shukun, Liu Xiaodong, Jia Yangli, Liu Lei, Liu Liangliang, Zhou Hongwei, Shi Jiahao, Liu Runpeng, Liu Ningchang, Qu Yingying, Zhu Xianchao, He Lili.

Introduction

Semantic annotation is one of the important technologies for deep processing of natural language by computers. It is to add semantic annotation to words or sentences in text.

The process of semantic labeling is to solve the problem. According to the general principles of annotation and the specific annotation objects, the standard is divided into different parts.

GB/T 44217 "Semantic Annotation Framework for Language Resource Management" is planned to consist of 12 parts.

--Part 1: Time and Events. The purpose is to provide a general method for describing time and events in text.

--Part 2: Dialogue Acts. The purpose is to provide a notation language for representing dialogue acts and a method for segmenting dialogue into the method of defining units.

--Part 4: Semantic Roles. The purpose is to provide a consensus annotation scheme for semantic roles.

--Part 5: Discourse Structure. The purpose is to provide a way to represent discourse realization and discourse content.

--Part 6: Principles of semantic annotation. The purpose is to define semantic annotation methods characterized by a semantic annotation framework.

--Part 7: Spatial Information. The purpose is to provide a general method to describe the spatial information related to the expression of motion in natural language text.

Spatial information and spatiotemporal information.

--Part 8: Semantic relations in discourse, core annotation framework. The purpose is to

provide a plan.

--Part 9: Reference annotation framework. Aims to provide a framework for the annotation and representation of references in natural language text and multimodal interactions.

Provide a comprehensive model.

-- Part 11: Measurable Quantity Information (MQI). The purpose is to provide a notation scheme for measurable quantity information.

--Part 12: Quantity. The purpose is to provide a general formal definition for the semantic representation of quantity information.

-- Part 14: Spatial Semantics. The purpose is to provide a method for annotating spatial information by establishing formal semantics for abstract syntax.

--Part 15: Extraction of measurable quantity information. The purpose is to provide a method for extracting measurable quantity information from natural language text.

General method of information.

Measurable quantity information (MQI), such as "165 cm" or "60 kg", can be used to describe a person's height or weight.

MQI describes one of the basic properties related to the size of the quantity. Its main feature is that the quantity information is in the form of a pair of the quantity represented by is composed of a number n and a unit u . The unit u can be a basic unit or a the units may be derived, normalized, or traditionally used. Such information is available in scientific publications or technical reports.

Therefore, any successful management of language resources must be this information needs to be processed.

In such a big data era, the industry and academia have higher and higher requirements for accurate extraction of MQI.

It is often necessary to identify and summarize the target company's net sales, gross profit, operating expenses, operating profit, interest expense, and net profit before tax from the annual report.

The rapidly developing medical informatics research also requires processing a large amount of medical text to analyze drug dosage, Inclusion criteria for clinical trials, phenotypic characteristics of patients, laboratory tests in clinical records, etc.

Research areas, These requirements require accurate and consistent representation of MQI for automatic processing, calculation and exchange.

However, in the field of information retrieval and natural language processing, there is currently no standardized method to represent measurable quantitative information.

Until now, industrial applications have usually used their own formats to annotate measurable quantities. We need a universal, Interoperable and standardized methods for representing measurable quantitative information so that it can work with different application systems.

Based on the semantic annotation principles specified in ISO 24617-6 and the basic requirements of ISO 24611, a universal annotation framework is developed to facilitate the use of scientific and technical annotations.

It also uses the ISO 24617 semantic annotation scheme to represent MQI.

It complies with various standards for morphological and syntactic annotation frameworks and is compatible with other existing relevant standards.

Note: For example, ISO 24617-1 and ISO 24617-7 propose notation methods for temporal (duration or time quantity) and spatial (distance) metrics, respectively.

ISO 24612 provides a graph annotation framework that can be used to annotate temporal or spatial metrics using these two annotation methods.

QML is standardized at an abstract level, allowing various serialization formats to represent annotations of measurable quantity information, such as this document standardizes the quantity information annotation at the abstract level of annotation and adopts the Independent annotation format.

This paper focuses on quantitative information in scientific and technical language and is expected to be helpful for information extraction (IR), question answering (QA), text summarization (TS) and other natural language processing (NLP) applications.

A semantic annotation framework for language resource management Part 11: Measurable Quantity Information (MQI)

1 Scope

This document specifies the semantic representation of measurable quantitative information for use in information retrieval, question answering, text summarization, and other natural language processing applications.

This document applies to language-related technologies or practices and is used for reference in other related technical fields.

This document also addresses temporal duration issues discussed in ISO 24617-1, and spatial measures such as distance dealt with in ISO 24617-7.

This document also includes the metrics introduced in 8.3 of ISO 24617-6:2016 or Quantity processing.

NOTE: ISO 24617-12 deals with more general theoretical issues of quantification and quantitative