FINAL DRAFT

INTERNATIONAL STANDARD

ISO/FDIS 24617-11

ISO/TC **37**/SC **4**

Secretariat: KATS

Voting begins on: **2021-05-10**

Voting terminates on: **2021-07-05**

Language resource management — Semantic annotation framework (SemAF) —

Part 11: **Measurable quantitative information (MQI)**

Gestion des ressources linguistiques — Cadre d'annotation sémantique (SemAF) —

Partie 11: Mesurer l'information quantitative (MQI)

Reference number
ISO/FDIS 24617-11:2021(E)

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 4, *Language resource management*.

A list of all parts in the ISO 24617 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Measurable quantitative information (MQI) such as '165 cm' or '60 kg' of 'John' that applies to the height or weight of the person is very common in ordinary language. MQI describes one of basic properties that is associated with the magnitude aspect of quantity. The main characteristics of MQI is that quantitative information is presented as measures expressed in terms of a pair <*n*, *u*>, consisting of a numerically expressed quantity *n* and a unit *u*, which is either basic or derived, or either normalized or conventionally used. Such information is much more abundant in scientific publications or technical reports to the extent that it constitutes an essential part of communicative segments of language in general. The processing of such information is thus required for any successful language resource management.

In such a big data era, demands from industry and academic communities for a precise acquisition of measurable quantitative information have increased. For example, business investment companies frequently need to aggregate various sorts of information covering net sales, gross profit, operating expenses, operating profit, interest expense, net profit before taxes, net income, etc., of the target companies from their annual reports. The fast-growing medical informatics research also needs to process a large amount of medical texts to analyze the dose of medicine, the eligibility criteria of clinical trial, the phenotype characters of patients, the lab tests in clinical records, etc.[8]. All these demands either in industry or in medical research require the accurate and consistent representation of measurable quantitative information for automated processing, computation, and exchange.

However, in the IR and NLP areas, there is no standardized way of representing measurable quantitative information currently available. Each application system developed in industrial sectors has hitherto used its own format to annotate measurable quantitative information. A flexible, interoperable and standardized measurable quantitative information representation format for IR and NLP tasks to work with many different application systems is called for.

This document aims at formulating a general annotation scheme with following the principles of semantic annotation laid down in ISO 24617-6 in general and the basic requirements of ISO 24611, that facilitates the processing of MQI in scientific and technical language and to make it interoperable with other semantic annotation schemes, such as ISO 24617. The annotation scheme is designed to be interoperable with other parts of ISO 24617. It also utilizes various ISO standards on lexical resources and morpho-syntactic annotation frameworks. It aims at being compatible with other existing relevant standards.

NOTE ISO 24617-1 and ISO 24617-7, for instance, have proposed a way of annotating measures on time (durations or time amounts) and space (distances), respectively. ISO 24612 provides a pivotal form (graphic annotation framework) that makes all the annotation of temporal or spatial measures in these two annotation schemes.

QML is normalized at the abstract level that allows various serialization formats representing annotated measurable quantitative information such as an XML-based representation. The normalization of QI (quantitative information) annotation is stated at the abstract level of annotation, and the standoff annotation format is adopted at the concrete level of serialization.

Focusing on measurements in scientifico-technological language, this document is expected to contribute to information extraction (IR)[9], question answering (QA), text summarization (TS), and other natural language processing (NLP) applications[10].

Language resource management — Semantic annotation framework (SemAF) —

Part 11: Measurable Quantitative information (MQI)

1 Scope

This document covers the measurable or magnitudinal aspect of quantity so that it can focus on the technical or practical use of measurements in IR (information retrieval), QA (question answering), TS (text summarization), and other NLP (natural language processing) applications. It is applicable to the domains of technology that carry more applicational relevance than some theoretical issues found in the ordinary use of language.

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

This document also treats temporal durations that are discussed in ISO 24617-1, and spatial measures such as distances that are treated ISO 24617-7, while making them interoperable with other measure types. It also accommodates the treatment of measures or amounts that are introduced in ISO 24617-6:2016, 8.3.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 24612, *Language resource management — Linguistic annotation framework (LAF)*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
quantity
property of a measurable object referring to its magnitude or multitude

[SOURCE: ISO/IEC Guide 99:2007, 1.1, modified — Definition substantially redrafted, and Notes removed.]

3.2
base quantity
quantity (3.1) in a conventionally chosen subset of a given system of quantities, where no quantity in the subset can be expressed in terms of the other quantities within that subset

Note 1 to entry: Kinds of quantities include seven base quantities defined by the International System of Quantities (ISQ).

[SOURCE: ISO/IEC Guide 99:2007, 1.4, modified — "no subset quantity" replaced with "no quantity in the subset", "the others" replaced with "the other quantities within that subset", and Notes and Example removed.]

3.3
derived quantity
quantity (3.1), in a system of quantities, defined in terms of the *base quantities* (3.2) of that system

EXAMPLE Speed is a derived quantity defined by length (distance) over time (LT^{-1}), where length (L) and time (T) are base quantities.

[SOURCE: ISO/IEC Guide 99:2009, 1.5, modified — Example replaced.]

3.4
quantitative information
QI
measurement associated with the *quantity* (3.1) of a measurable object

3.5
measurable quantitative information
MQI
quantitative information (3.4) that can be expressed in unitized numeric terms

3.6
measurable quantitative information markup language
markup language of measurable quantitative information
quantitative markup language
QML
specification language for the annotation of *measurable quantitative information* (3.5) extractable from text or other medium types of language

3.7
measurement unit
unit of measurement
unit
scalar basis, defined and adopted by convention, of measuring objects by multiplying their quantitative values expressed in real numbers

Note 1 to entry: The expressions that are used in measurement such as “metre”, “litre”, and “μmol/kg” are units by the definition given above. The multitude expressions such as “bottles”, “boxes”, or “two” as in “two bottles of milk”, “a box of apples”, and “two coffees” sometimes fail to be regarded as units, but they can also be if they are accepted as units by convention or agreement in some communities. ISO 24617 SemAF Part 12: Quantification treats such multitude expressions as genuine units.

[SOURCE: ISO/IEC Guide 99:2007, 1.9, modified — Definition substantilly redrafted, original Notes removed, new Note 1 to entry added.]

3.8
base unit
measurement unit (3.7) that is adopted by convention for a *base quantity* (3.2)

Note 1 to entry: There are seven base units chosen by the International System of Units (SI) associated with seven ISQ base quantities to measure quantities, as shown in Table 1.

Table 1 — Base units

SI base unit (unit symbol)	Associated ISQ base quantity (base quantity symbol)
metre (m)	length (L)
kilogram (kg)	mass (M)
second (s)	time (T)
ampere (A)	electric current (I)
kelvin (K)	thermodynamic temperature (È)
mole (mol)	amount of substance (N)
candela (cd)	luminous intensity (J)

[SOURCE: ISO/IEC Guide 99:2007, 1.10, modified — Notes and Examples removes, new Note 1 to entry and Table 1 added.]

3.9
derived unit
measurement unit (3.7) for a *derived quantity* (3.3)

EXAMPLE The unit "newton" (N) is a derived unit for a derived quantity "force" (F), which is defined to be "mass times acceleration" (MLT^{-2}), where the quantity "acceleration" is a derived quantity defined by "velocity divided by time" (VT^{-1}) and "velocity" defined by "length (distance) divided by time" (LT^{-1}).

Note 1 to entry: Table 2 illustrates some of the derived units.

[SOURCE: ISO/IEC Guide 99:2007, 1.11, modified — Examples removed, new Example and Note 1 to entry added.]

Table 2 — derived units

Derived unit (unit symbol)	Associated derived quantity
kilometre per minute(km/min)	speed = length(L)/ time(T)
gram per cubic metre (gram/m^3)	density = mass(M)/volume(L^3)
kilogram metre per square second (kg x m/s^2)	force = mass (M) x length(L)/time(T^2)
lumen per square metre (lm/m^2)	Illuminance = luminous intensity (J)/area(M^2)

4 Abstract specification of QML

4.1 Overview

The quantitative markup language (QML) (3.6) is specified at two levels, abstract and concrete. Some characteristics of QML are listed in 4.2. The overall structure of QML is represented by a metamodel, as introduced in 4.3. The abstract syntax of QML as QML_as shall be a set-theoretic specification of QML in conceptual terms that are independent of ways of representing the annotation (content) of measurable quantitative information. The concrete syntax of QML as QML_cs shall be a specification of a set of representation formats, based on QML_as, for the annotation of measurable quantitative information in a computationally tractable way. The QML_as is introduced in 4.4, while QML_cs is presented in 4.5. Equivalent concrete syntaxes, including an XML-based concrete syntax QML_csx and a TEI-based concrete syntax QML_cst, are described in Clause 5 and Clause 6, respectively.

NOTE There can be many equivalent concrete syntaxes defined on a single abstract syntax.

4.2 Characteristics of QML

QML shall have the following characteristics.

a) QML shall focus on the annotation of the measurable attributes of entities. For example, "BMI between 10-20 kg/ m^2"

b) QML shall provide a way to annotate the relations of measures. For example, "age 40 or older" and "fpg>=100 mg/dl or a1c not less than 5,8 %"

c) QML shall cover the complex uses of unitized numeric quantities. For example, "14,0 × 109", "glycosylated haemoglobin (hba1c) <1,15 times the upper limit of normal".

d) QML shall facilitate the identification of normalized numeric, units, as the measurable attribute of an associated entity.

NOTE QML does not specify ways of annotating the normalization (e.g. "millimoles per litre" is normalized to "mmol/L") or complete specification (e.g. "kg/m" is "kg/m2" for BMI) of units, which will be dealt with in another part of ISO 24617 addressing automated implementation of MQI.

4.3 Metamodel

The overall structure of measurable quantitative information is represented by the metamodel in Figure 1.

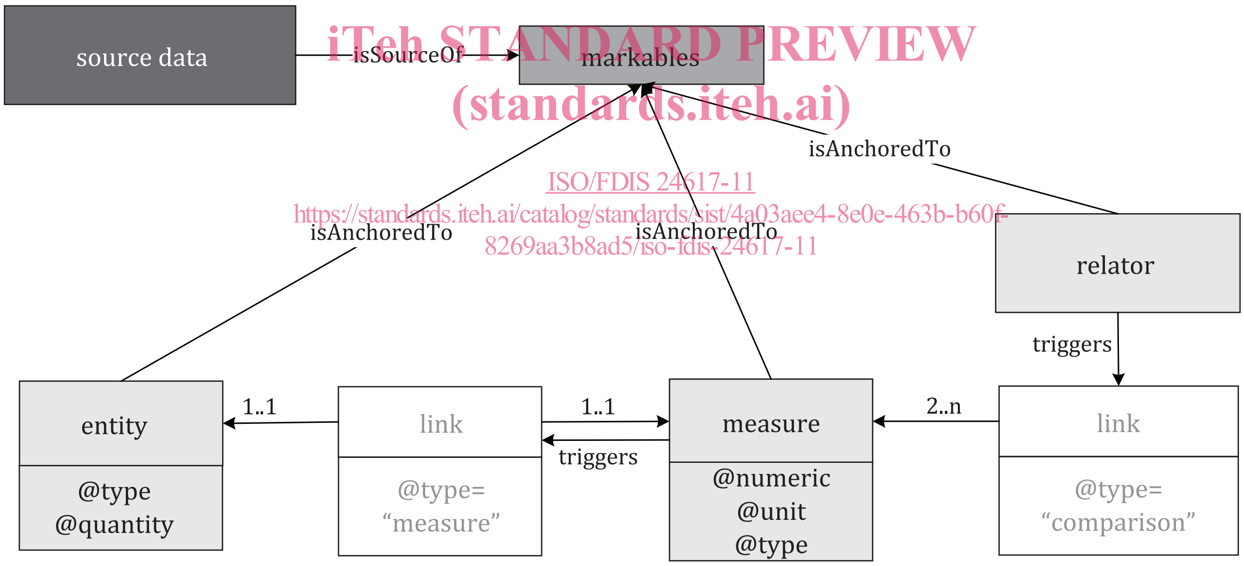

Figure 1 — Metamodel of measurable quantitative information

This metamodel shall consist of seven class components, represented as square boxes in Figure 1:

a) **source data** as input to the annotation of MQI,

b) **markables** extracted from data sources,

c) three types of basic elements: **entity**, **measure**, and **relator**,

d) two types of links: **measure link** and **comparison link**.

The element "entity" shall be any object that has the property of a measurable quantity, represented by "@quantity", as one of its properties. The "entity", as is used in this document, shall be a very general term that refers to any object, not just to individual entities, but also to their properties, such as "height" of a building or "speed" of a car, and also to any kinds of eventualities such as states, processes or transitions.

EXAMPLE 1 We drove at more than 200 kilometres per hour on a German autobahn.

The speed mentioned by "more than 200 kilometres per hour" applies to the quantitative property of a motion: e.g. the measure "over 200 kilometres per hour" applies to the motion of driving mentioned in the example.

The element "measure" represents a measurable quantity of an entity in terms of three attributes: quantity, unit, and type.

EXAMPLE 2 The height of Mt. Hall is 1 950 metres.

The measure shall consist of a quantity referred to by a numeric expression "1 950" and a unit "metre". It applies to the "height" quantity of the geographical object, named "Mt. Hall".

The element "relator" which is associated with markables such as "equal to", "greater than", "<=", "between", or "at least" has only a functional status of relating two or more measures.

EXAMPLE 3 One pound equals 16 ounces.

It is a relator of identity between two measures, "one pound" and "16 ounces".

EXAMPLE 4 1 foot is less than 1 metre, for it is exactly equal to 30,48 cm.

This example illustrates two types of links between measures: the relation of being "less than", and that of being an identity.

A link of the type "measure" shall relate a measure to the quantitative property of an entity. Such a link is triggered by a measure element.

A link of the type "comparison" shall relate a measure to another or other more measures. Such a link is often triggered by an element "comparison".

4.4 Abstract syntax of QML (QML_as)

A markup language QML shall be a specification language for the annotation of MQI. The abstract syntax of QML shall specifies an annotation scheme in set-theoretic terms based on a conceptual understanding of MQI. The abstract syntax QML_as is understood to be structured as a triple <*B*, *R*, @> such that

a) *B* is a set of three basic element types: **entity**, **measure**, and **relator**;

b) *R* is a set of two link types: **measure** and **comparison** types;

c) @ is a set of assignments that specify the list of attributes and their value types associated with each of the basic element types in *B* and each of the link types in *R*.

Every element in *B* shall have at least one attribute, @type, and so does every link. The values of @ type are CDATA associated with each of the elements. For instance, the entity of "mountain" is of the "geographical" type, and the entity named "John" is of the "person" type.

The values of @quantity for an entity are CDATA that may include values such as height, width, or weight, and so on.

The assignment of **measure** shall have three attributes: @numeric, @unit, and @type. A possible value of the attribute @numeric is a real number. A possible value of @unit is one of the units in a system conventionally accepted such as one of the SI base units or derived units. A possible value of @type is one of the quantities listed as ISQ base quantities or derived quantities, such as length, mass, voltage, and so on.