
**Language resource management —
Semantic annotation framework
(SemAF) —**

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

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

Partie 11: Informations quantitatives mesurables (MQI)



This document is a preview generated by EKO



COPYRIGHT PROTECTED DOCUMENT

© ISO 2021

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Abstract specification of QML	3
4.1 Overview	3
4.2 Characteristics of QML	4
4.3 Metamodel	4
4.4 Abstract syntax of QML (QML_as)	5
4.5 Concrete syntaxes of QML (QML_cs) and its subsets	6
5 XML-based concrete syntax of QML (QML_csx)	6
5.1 General	6
5.2 Tag names with ID prefixes	6
5.3 Attribute specification of the root <MQI>	7
5.4 Attribute specification of the basic element types	7
5.5 Attribute specification of the link types	8
5.6 Illustrations of QML_csx	8
5.6.1 General	8
5.6.2 Sample data	8
5.6.3 Procedure of annotation	9
6 TEI-based concrete syntax of QML (QML_cst)	11
6.1 Concrete syntaxes of QML (QML_cst)	11
6.1.1 Overall	11
6.1.2 Tag names with ID prefixes	11
6.1.3 Attribute specification of the basic element types	11
6.1.4 Attribute specification of the two link types	12
6.2 Illustrations of QML_cst	12
6.2.1 Overall	12
6.2.2 Sample data	12
6.2.3 Illustrations of TEI-based Concrete Syntax	13
Annex A (informative) Illustrations of QML_csx with more samples	16
Annex B (informative) Informal statements of MQI	19
Annex C (informative) The representation of units	20
Bibliography	21

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 $\langle n, u \rangle$, 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)^[2], 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.]