



BSI Standards Publication

Language resource management — Semantic annotation framework (SemAF)

Part 15: Measurable quantitative information extraction (MQIE)

National foreword

This British Standard is the UK implementation of ISO 24617-15:2025.

The UK participation in its preparation was entrusted to Technical Committee TS/1, Terminology.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

© The British Standards Institution 2025
Published by BSI Standards Limited 2025

ISBN 978 0 539 30245 5

ICS 01.020

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 May 2025.

Amendments/corrigenda issued since publication

Date	Text affected
------	---------------



International
Standard

ISO 24617-15

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

**Part 15:
Measurable quantitative
information extraction (MQIE)**

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

*Partie 15: Extraction d'informations quantitatives
mesurables (MQIE)*

**First edition
2025-05**

Reference number
ISO 24617-15:2025(en)

© ISO 2025  **Documents**
Document Sharing Hub



COPYRIGHT PROTECTED DOCUMENT

© ISO 2025

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 General framework of MQIE	2
4.1 Overview	2
4.2 Primary requirements of MQIE	2
4.3 Framework	3
4.4 Preprocessing	4
4.5 Basic element identification	4
4.6 Link identification	5
4.7 Measure normalization	6
4.8 Verification and filtering	7
5 Examples	7
5.1 General	7
5.2 Sample data	7
5.3 Procedure of extraction	8
5.3.1 Overview	8
5.3.2 Preprocessing	8
5.3.3 Basic element extraction	8
5.3.4 Link identification	8
5.3.5 Measure normalization	9
5.3.6 Verification and filtering	9
Annex A (informative) Examples of applications extended based on MQIE	11
Annex B (informative) Informal statements of MQI during extraction	14
Bibliography	15

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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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) describes one of basic properties that is associated with the magnitude aspect of quantity, and is very common in ordinary language. The main characteristics of MQI, as described in ISO 24617-11, is that quantitative information is presented as measures expressed in terms of a pair of a numerically expressed quantity and a unit. 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 an accurate extraction of MQI have increased.^[1] For example, business investment companies frequently need to identify and aggregate various 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 text to analyse the dose of medicine, the eligibility criteria of clinical trial, the phenotype characters of patients, the laboratory tests in clinical records, etc.^{[2][3]} All these demands either in industry or in medical research require the effective extraction of MQI for automated identification, aggregation, computation and analysis.^[4]

However, in the information retrieval and natural language processing areas, there is no standardized way of extracting measurable quantitative information currently available. Each application system developed in industrial sectors has hitherto used common NLP models or their own models to identify measurable quantitative information from unstructured text. There is no standard extraction procedure for ensuring the quality of the extraction currently. A general, interoperable and standardized measurable quantitative information extraction scheme for IR and NLP tasks to work with many different application systems is called for.

This document formulates a general extraction scheme while following the basic requirements of semantic annotation laid down in ISO 24617-11, which facilitates the annotation of MQI in scientific and technical language and makes it interoperable with other semantic annotation schemes such as those given in the parts of the ISO 24617 series. The extraction scheme also utilizes various International Standards on lexical resources and morpho-syntactic annotation frameworks. It aims at being compatible with other existing relevant standards such as ISO 24617-9.

NOTE ISO 24617-11 provides a standardized schema of annotating measurable quantitative information from unstructured text.

Focusing on measurements in scientifico-technological language, this document is expected to contribute to information retrieval (IR), question answering (QA), text summarization (TS) and other natural language processing (NLP) applications.^{[5][6][7]}

