



DPC: 24/30486179 DC

BSI Group Headquarters

389 Chiswick High Road London W4 4AL

Tel: +44 (0)20 8996 9000

Fax: +44 (0)20 8996 7400

www.bsigroup.com

Date: 13 August 2024

Origin: International

Latest date for receipt of comments: 06 October 2024

Project No. 2024/00138

Responsible committee: TS/1 Terminology

Interested committees:

Title: Draft BS ISO 24617-15 Language resource management - Semantic annotation framework (SemAF). Part 15: Measurable quantitative information extraction (MQIE).

Please notify admin.start@bsigroup.com if you are aware of any keywords that might assist in clarifying or identifying the standard or of the content of this standard:

- i) has any issues related to 3rd party IPR, patent or copyright
- ii) affects other national standard(s)
- iii) requires additional national guidance or information

WARNING: THIS IS A DRAFT AND MUST NOT BE REGARDED OR USED AS A BRITISH STANDARD. THIS DRAFT IS NOT CURRENT BEYOND 06 October 2024

This draft is issued to allow comments from interested parties; all comments will be given consideration prior to publication. No acknowledgement will normally be sent. **See overleaf for information on the submission of comments.**

No copying is allowed, in any form, without prior written permission from BSI except as permitted under the Copyright, Designs and Patent Act 1988 or for circulation within a nominating organization for briefing purposes. Electronic circulation is limited to dissemination by e-mail within such an organization by committee members.

Further copies of this draft may be purchased from BSI Shop <http://shop.bsigroup.com> or from BSI Customer Services, Tel: +44(0) 20 8996 9001 or email cservices@bsigroup.com. British, International and foreign standards are also available from BSI Customer Services.

Information on the co-operating organizations represented on the committees referenced above may be obtained from <https://standardsdevelopment.bsigroup.com/>.

E-mail: admin.start@bsigroup.com

Introduction

This draft standard is based on national and international discussions. Your comments on this draft are invited and will assist in the preparation of the consequent standard.

For international standards, comments will be reviewed by the relevant UK national committee before sending the consensus UK vote and comments to the international committee, which will then decide appropriate action. If the international standard is approved, it is usual for the text to be published as a British Standard.

For national standards, comments will be reviewed by the relevant UK national committee and the resulting standards published as a British Standard.

UK Vote

Please indicate whether you consider the UK should submit a negative (with supporting technical reasons) or positive vote on this draft. Please indicate if you are aware of any reason why this draft standard should not be published as a British Standard.

Submission of Comments

- **Annotated drafts are not acceptable and will be rejected.**
- All comments should be submitted online at <https://standardsdevelopment.bsigroup.com/>. You will need to register in order to comment.

Template for comments and secretariat observations

Date: xx/xx/20xx	Document: ISO/DIS xxxx
------------------	------------------------

1	2	(3)	4	5	(6)	7
MB	Clause No./ Subclause No./Annex (e.g. 3.1)	Paragraph/ Figure/Table/Note	Type of comment	Comment (justification for change) by the MB	Proposed change by the MB	Secretariat observations on each comment submitted
	3.1	Definition 1	ed	Definition is ambiguous and needs clarifying.	Amend to read '...so that the mains connector to which no connection...'	
	6.4	Paragraph 2	te	The use of the UV photometer as an alternative cannot be supported as serious problems have been encountered in its use in the UK.	Delete reference to UV photometer.	



DRAFT International Standard

ISO/DIS 24617-15

Language resource management — Semantic annotation framework (SemAF) —

Part 15: Measurable quantitative information extraction (MQIE)

ICS: 01.020

ISO/TC 37/SC 4

Secretariat: **KATS**

Voting begins on:
2024-08-13

Voting terminates on:
2024-11-05

This document is circulated as received from the committee secretariat.

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENTS AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

Reference number
ISO/DIS 24617-15:2024(en)

Full document details: 24/30486179 DC

© ISO 2024  **Documents**
Document Sharing Hub



COPYRIGHT PROTECTED DOCUMENT

© ISO 2024

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.....2

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.....6

5 Examples.....7

5.1 General.....7

5.2 Sample data.....7

5.3 Procedure of extraction.....7

5.3.1 Overview.....7

5.3.2 Pre-processing.....7

5.3.3 Basic element extraction.....7

5.3.4 Link identification.....8

5.3.5 Measure normalization.....8

5.3.6 Verification and Filtering.....8

Annex A (informative) The 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 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) 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,^[1] 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.^[2] 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 analyze the dose of medicine, the eligibility criteria of clinical trial, the phenotype characters of patients, the lab tests in clinical records, etc.^[3,4] All these demands either in industry or in medical research require the effective extraction of MQI for automated identification, aggregation, computation, and analysis ^[5].

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, named 'SemAF-MQIE', aims at formulating a general extraction scheme with 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 to make it interoperable with other semantic annotation schemes such as ISO 24617. The extraction scheme 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 24617-11 has proposed 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 ^[6-8].