



SLOVENSKI STANDARD
oSIST ISO/DIS 24623-2:2021

01-marec-2021

**Upravljanje jezikovnih virov - Lingua franca za korpusne poizvedbe (CQLF) - 2. del:
Ontologija**

Language resource management -- Corpus Query Lingua Franca (CQLF) - Part 2:
Ontology

iTeh STANDARD PREVIEW

Gestion des ressources linguistiques -- Corpus Query Lingua Franca (CQLF)
(standards.itih.ai)

Ta slovenski standard je istoveten z: ISO/DIS 24623-2

oSIST ISO/DIS 24623-2:2021
<https://standards.itih.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021>

ICS:

01.020	Terminologija (načela in koordinacija)	Terminology (principles and coordination)
01.140.20	Informacijske vede	Information sciences
35.240.30	Uporabniške rešitve IT v informatiki, dokumentiranju in založništvu	IT applications in information, documentation and publishing

oSIST ISO/DIS 24623-2:2021

en

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[oSIST ISO/DIS 24623-2:2021](https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021)

<https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021>

DRAFT INTERNATIONAL STANDARD

ISO/DIS 24623-2

ISO/TC 37/SC 4

Secretariat: **KATS**

Voting begins on:
2021-01-07

Voting terminates on:
2021-04-01

Language resource management — Corpus Query Lingua Franca (CQLF) —

Part 2: Ontology

ICS: 01.020

iTeh STANDARD PREVIEW (standards.iteh.ai)

[oSIST ISO/DIS 24623-2:2021](https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021)

<https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021>

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENT 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.

This document is circulated as received from the committee secretariat.



Reference number
ISO/DIS 24623-2:2021(E)

iTeh STANDARD PREVIEW (standards.iteh.ai)

[oSIST ISO/DIS 24623-2:2021](https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021)
<https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021>



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 Motivation and aims	3
5 CQLF Ontology	4
5.1 OWL DL formalism	4
5.2 Structure of the ontology	5
5.3 CQLF Metamodel	7
5.4 Functionalities	9
5.5 Frames	11
5.6 Use Cases	12
5.7 CQLs	12
6 Conformance statements	13
6.1 Positive conformance statements	13
6.2 Negative conformance statements	14
7 RDF/XML serialization	14
Annex A (informative) Illustrative examples of non-normative elements in the CQLF Ontology	15
Annex B (informative) CQLF Ontology: Moderated community process	18
Bibliography	19

<https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021>

ISO/DIS 24623-2:2021(E)

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 24623 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

Technical Committee ISO/TC 37, *Language and Terminology*, Subcommittee SC 4, *Language Resource management* has developed several families of standards codifying various aspects of representation of language data. These standards describe general corpus-oriented data models in the Linguistic Annotation Framework (LAF, ISO 24612) family, various aspects of the semantic representation in the family of the Semantic Annotation Framework (SemAF, ISO 24617-1 and others), the representation of lexical data in the Lexical Markup Framework family (LMF, ISO 24613-1 and others), as well as the representation of metadata in the Component Metadata Infrastructure (CMDI, ISO 24622-1 and others). Complementary to the standards concerning the representation of language data, the Corpus Query Lingua Franca (henceforth CQLF) family of standards (ISO 24623) focuses on the exploitation of language data and ways to satisfy various kinds of information needs targeting these data.

The CQLF Metamodel, described by ISO 24623-1 (CQLF-1), is a maximally permissive construct that establishes means of describing the scope of corpus query languages (CQLs) at a general level and with a focus on various kinds of data models assumed by query systems, with conformance conditions meant to be satisfied by a wide range of CQLs. The Metamodel provides a “skeleton” for a CQL taxonomy by setting up basic categories of corpus queries (encoded as CQLF-1 levels and modules) as well as the dependencies among them.

Consequently, the task of a more concrete characterization of CQLs falls to other members of the CQLF standard family. This document (ISO 24623-2, “CQLF-2” for short) establishes an ontology which focuses on the generalized information needs satisfied by corpus queries, and which is structured as a multi-layer taxonomy against which individual CQLs can make positive and negative conformance statements.

Establishing this ontology allows, on the one hand, a fine-grained comparison of the expressive power of CQLs, and, on the other hand, it is going to serve a practical purpose: as a foundation for a platform where developers can enter conformance statements, and where end users can see which CQL to turn to in order to ensure that their search needs get satisfied.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[oSIST ISO/DIS 24623-2:2021](https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021)

<https://standards.iteh.ai/catalog/standards/sist/8f12802d-a801-4c81-86b5-7b983c8a1074/osist-iso-dis-24623-2-2021>

Language resource management — Corpus Query Lingua Franca (CQLF) —

Part 2: Ontology

1 Scope

This document defines an ontology for a fine-grained description of the expressive power of CQLs in terms of search needs. The ontology consists of three interrelated taxonomies of concepts: a) the CQLF Metamodel (a formalization of CQLF-1), b) the Expressive Power taxonomy, which describes different facets of the expressive power of CQLs, and c) a taxonomy of CQLs.

The normative parts of this document comprise a) the taxonomy of the CQLF Metamodel, b) the Functionality layer of the Expressive Power taxonomy, c) the structure of the layers of the Expressive Power taxonomy and the relationships between them, in the form of subsumption assertions, as well as d) the formalization of the linkage between the CQL taxonomy and the Expressive Power taxonomy, in the form of positive and negative conformance statements.

This document does not provide a normative listing of the middle and bottom layer of the Expressive Power taxonomy (called Frames and Use Cases, respectively). An exhaustive inventory of the concepts at these two layers is not possible due to the fact that existing CQLs differ widely in the complexity of the supported combinations of Functionalities and that new CQLs can be created offering additional combinations or subtypes of Functionalities. Frames and Use Cases are expected to be filled in through a moderated community process, driven by CQL developers as well as end users. An informative annex to this document contains a sample of Frames and Use Cases together with conformance statements linking them with the CQP [5] and ANNIS [Z] query languages.

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/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*

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

ISO 24623-1, *Language resource management — Corpus query lingua franca (CQLF) — Part 1: Metamodel*

МОТИК B. Patel-Schneider, Peter F., and Parsia, B. (2012). *OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (Second Edition)*. W3C Recommendation, 11 December 2012. (Latest version available at <http://www.w3.org/TR/owl2-syntax/>.)

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 24612, ISO 24623-1 and the following 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>

ISO/DIS 24623-2:2021(E)

— IEC Electropedia: available at <http://www.electropedia.org/>

3.1
CQLF module
subcomponent of the CQLF Metamodel, defined with reference to a specified data model characteristic

Note 1 to entry: The CQLF Metamodel currently distinguishes three modules within CQLF Level 1, Linear (plain-text, segmentation, and simple annotation), and three modules within CQLF Level 2, Complex (hierarchical, dependency, and containment).

[SOURCE: ISO 24623-1:2018, 3.8, modified – “a CQLF level” was replaced with “the CQLF Metamodel” in order to improve clarity outside the context of ISO 24623-1.]

3.2
Functionality
label for a concept in the CQLF Ontology that represents a family of capabilities contributing to the expressive power of a CQL, formulated at a general level and linked to one or more CQLF modules

3.3
Frame
label for a concept in the CQLF Ontology that represents a typical search need of end users, understood as one facet of the expressive power of CQLs

Note 1 to entry: Most Frames arise from the specialization of a Functionality and/or the combination of multiple Functionalities.

3.4
Use Case
label for a concept in the CQLF Ontology that represents a concrete instantiation of a Frame, for which it can be determined unambiguously whether a given query expression satisfies the search need or not

Note 1 to entry: Use Cases are often parameterized, i.e. they contain variable elements. Parameterized Use Cases are satisfied by parameterized query expressions.

3.5
CQL
corpus query language
formal language designed to retrieve specific information from (large) language data collections, and thereby incorporate certain abstractions over commonly shared data models that make it possible for the end user (or user agents) to address parts of those data models

Note 1 to entry: A CQL defines a syntactic notation for query expressions and the corresponding search semantics, i.e. an intensional specification of the intended result set. For most current CQLs, semantics are implicitly defined by a particular implementation.

[SOURCE: ISO 24623-1:2018, 3.4, modified – “user” was replaced with “end user” in the definition, the abbreviation CQL was added as preferred term, and Note 1 to entry was added.]

3.6
search need
information pattern that an end user wants to locate in a corpus, based on the primary data stream and/or simple or complex annotation

3.7
end user
agent who uses a CQL to satisfy his or her search needs

Note 1 to entry: This can be done via an interactive GUI, a command-line tool, programmatically via some API, or by a software program developed by the end user.

3.8**query expression**

string that is syntactically valid in a given CQL and can be executed to return a result set

Note 1 to entry: Query expressions are often parameterized with variable elements. No formal specification of the parameter substitution procedure is attempted, but entries for parameterized query expressions in the ontology are required to include informal descriptions of the range of admissible values and any transformations required.

3.9**parameter**

variable element in a query expression or in the description of a search need

3.10**positive conformance statement**

assertion that a given CQL supports a given Use Case by means of a query expression

3.11**negative conformance statement**

assertion that a given CQL cannot support a given Use Case, Frame or Functionality

Note 1 to entry: Negative conformance is due to technical unavailability of specific capabilities in the respective CQL or limitations on the complexity of query expressions.

3.12**CQL capability**

corpus query language capability
facility provided by CQLs to meet a specific aspect of search needs

3.13**layer**

totality of concepts at the same level of abstraction in the CQLF Ontology

EXAMPLE Functionalities, Frames, Use Cases

3.14**token**

non-empty contiguous sequence of graphemes or phonemes in a document

[SOURCE: ISO 24611:2012, 3.21, modified — The note was deleted.]

4 Motivation and aims

CQLs differ widely in their basic sets of capabilities. Whereas some are restricted to rather specific application scenarios, others are able to cover a wider variety of applications and search needs. It is therefore both the quality and the quantity of CQL capabilities – as well as the degree of their combination – that determine the expressive power of a CQL. The CQLF Ontology is not intended to articulate all the possible combinations of capabilities unless these are justified by genuine usage. Its aim is to provide representative categories for typical search needs within a taxonomy of CQL capabilities.

Yet another important aspect is the degree of explication that a CQL delivers. Some CQLs are able to express a particular search need in a condensed and highly specialized manner, while others rely on complex combinations of a few elementary capabilities. The CQLF Ontology leverages information from CQLs with a more explicit formalization of capabilities in order to create a systematic taxonomy of search needs and thus be able to classify CQLs of rather implicit formalization. Their respective degree of explication is visible in the parameterized query expressions included in all positive conformance statements.

This document defines the structure of an ontology representing CQL capabilities and search needs in a taxonomy consisting of three layers of varying degrees of abstraction. It also provides formal means for stating the conformance of individual CQLs to this central taxonomy.