



SLOVENSKI STANDARD

SIST ISO 5078:2025

01-junij-2025

Upravljanje terminoloških virov - Luščenje terminologije

Management of terminology resources — Terminology extraction

Gestion des ressources terminologiques — Extraction de terminologie

Ta slovenski standard je istoveten z: ISO 5078:2025

ICS:

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

SIST ISO 5078:2025

en



International Standard

ISO 5078

Management of terminology resources — Terminology extraction

*Gestion des ressources terminologiques — Extraction de
terminologie*

**First edition
2025-02**

ISO Standards
(<https://standards.iteh.ai>)
Document Preview

[SIST ISO 5078:2025](https://standards.iteh.ai/catalog/standards/sist/8144baa1-5acc-44cf-9014-ef33030e816d/sist-iso-5078-2025)

<https://standards.iteh.ai/catalog/standards/sist/8144baa1-5acc-44cf-9014-ef33030e816d/sist-iso-5078-2025>

ISO 5078:2025(en)

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

SIST ISO 5078:2025

<https://standards.iteh.ai/catalog/standards/sist/8144baa1-5acc-44cf-9014-ef33030e816d/sist-iso-5078-2025>

**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

ISO 5078:2025(en)

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Principles and methods	5
4.1 General	5
4.2 Text corpora and terminology extraction	5
4.3 Compilation of text corpora	6
4.3.1 Text corpora used for terminology extraction	6
4.3.2 Criteria for selecting texts for a text corpus	6
4.3.3 Considerations for text corpus creation	7
4.4 Terminology extraction approaches and methods	8
4.4.1 Classification of terminology extraction approaches	8
4.4.2 Extraction method according to the number of languages	10
4.4.3 Extraction method according to the process	11
4.4.4 Extraction method according to the underlying technique	11
4.4.5 Extraction method according to the underlying technology	14
4.4.6 Extraction method according to the extracted items	16
4.5 Term extraction output	17
4.5.1 Filtering candidate term lists	17
4.5.2 Assessing term eligibility	18
4.6 Uses for terminology extraction output	19
5 Implementation of terminology extraction	19
5.1 General	19
5.2 Initial considerations for terminology extraction	19
5.3 Terminology extraction workflow	20
5.3.1 Overview	20
5.3.2 Starting the terminology extraction workflow	20
5.3.3 Building or selecting a text corpus	20
5.3.4 Preprocessing the text corpus	20
5.3.5 Identifying candidate terms	21
5.3.6 Selecting relevant terms	21
5.3.7 Allocating terms to concepts	22
5.3.8 Identifying concept relations and building concept systems	22
5.3.9 Completing terminological entries	22
Bibliography	23