ISO 5078:2025-02 (E)

Management of terminology resources - Terminology extraction

Contents

Page

Foreword iv

Introduction v

1 Scope 1

2 Normative references 1

3 Termsanddefinitions 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