

INTERNATIONAL  
STANDARD

**ISO**  
**1087-2**

NORME  
INTERNATIONALE

First edition  
Première édition  
2000-03-01

---

---

**Terminology work — Vocabulary —**

Part 2:  
**Computer applications**

**Travaux terminologiques — Vocabulaire —**

Partie 2:  
**Applications logicielles**

[ISO 1087-2:2000](https://standards.iteh.ai/catalog/standards/sist/9694d9c1-caf2-4626-b603-c72335b55fd5/iso-1087-2-2000)

<https://standards.iteh.ai/catalog/standards/sist/9694d9c1-caf2-4626-b603-c72335b55fd5/iso-1087-2-2000>



Reference number  
Numéro de référence  
ISO 1087-2:2000(E/F)

© ISO 2000

**PDF disclaimer**

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

**PDF – Exonération de responsabilité**

Le présent fichier PDF peut contenir des polices de caractères intégrées. Conformément aux conditions de licence d'Adobe, ce fichier peut être imprimé ou visualisé, mais ne doit pas être modifié à moins que l'ordinateur employé à cet effet ne bénéficie d'une licence autorisant l'utilisation de ces polices et que celles-ci y soient installées. Lors du téléchargement de ce fichier, les parties concernées acceptent de fait la responsabilité de ne pas enfreindre les conditions de licence d'Adobe. Le Secrétariat central de l'ISO décline toute responsabilité en la matière.

Adobe est une marque déposée d'Adobe Systems Incorporated.

Les détails relatifs aux produits logiciels utilisés pour la création du présent fichier PDF sont disponibles dans la rubrique General Info du fichier; les paramètres de création PDF ont été optimisés pour l'impression. Toutes les mesures ont été prises pour garantir l'exploitation de ce fichier par les comités membres de l'ISO. Dans le cas peu probable où surviendrait un problème d'utilisation, veuillez en informer le Secrétariat central à l'adresse donnée ci-dessous.

iTeh STANDARD PREVIEW  
(standards.iteh.ai)

ISO 1087-2:2000

<https://standards.iteh.ai/catalog/standards/sist/9694d9c1-caf2-4626-b603-c72335b55fd5/iso-1087-2-2000>

© ISO 2000

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester. / Droits de reproduction réservés. Sauf prescription différente, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'ISO à l'adresse ci-après ou du comité membre de l'ISO dans le pays du demandeur.

ISO copyright office  
Case postale 56 • CH-1211 Geneva 20  
Tel. + 41 22 749 01 11  
Fax + 41 22 734 10 79  
E-mail [copyright@iso.ch](mailto:copyright@iso.ch)  
Web [www.iso.ch](http://www.iso.ch)

Printed in Switzerland/Imprimé en Suisse

**Contents**

Page

Foreword.....	v
Introduction .....	vii
1 Scope .....	1
2 General concepts.....	1
3 Data organization.....	6
4 Filtering of terminological data .....	7
5 Characters .....	9
6 Storing data .....	11
7 Information retrieval .....	13
8 Data handling .....	15
9 Data validation .....	16
10 Operations .....	18
11 Related fields of application .....	20
Bibliography .....	21
Alphabetical index .....	23

ISO 1087-2:2000

<https://standards.iteh.ai/catalog/standards/sist/9694d9c1-caf2-4626-b603-c72335b55fd5/iso-1087-2-2000>

## Sommaire

Page

Avant-propos.....	vi
Introduction .....	viii
1 <b>Domaine d'application</b> .....	1
2 <b>Concepts généraux</b> .....	1
3 <b>Organisation des données</b> .....	6
4 <b>Filtrage des données terminologiques</b> .....	7
5 <b>Caractères</b> .....	9
6 <b>Stockage des données</b> .....	11
7 <b>Recherche documentaire</b> .....	13
8 <b>Manipulation des données</b> .....	15
9 <b>Validation des données</b> .....	16
10 <b>Opérations</b> .....	18
11 <b>Domaines d'application associés</b> .....	20
Bibliographie .....	21
Index alphabétique .....	25

[ISO 1087-2:2000](https://standards.iteh.ai/catalog/standards/sist/9694d9c1-caf2-4626-b603-c72335b55fd5/iso-1087-2-2000)

<https://standards.iteh.ai/catalog/standards/sist/9694d9c1-caf2-4626-b603-c72335b55fd5/iso-1087-2-2000>

## 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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this part of ISO 1087 may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

International Standard ISO 1087-2 was prepared by Technical Committee ISO/TC 37, *Terminology (principles and coordination)*, Subcommittee SC 3, *Computer applications*.

This part of ISO 1087 cancels and replaces ISO 1087:1990, clause 7. ISO 1087-1, which is being prepared, will constitute a technical revision of the other clauses of ISO 1087:1990.

ISO 1087 consists of the following parts, under the general title *Terminology work — Vocabulary*:

— *Part 1: Theory and application*

— *Part 2: Computer applications*

## Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (CEI) en ce qui concerne la normalisation électrotechnique.

Les Normes internationales sont rédigées conformément aux règles données dans les Directives ISO/CEI, Partie 3.

Les projets de Normes internationales adoptés par les comités techniques sont soumis aux comités membres pour vote. Leur publication comme Normes internationales requiert l'approbation de 75 % au moins des comités membres votants.

L'attention est appelée sur le fait que certains des éléments de la présente partie de l'ISO 1087 peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence.

La Norme internationale ISO 1087-2 a été élaborée par le comité technique ISO/TC 37, *Terminologie (principes et coordination)*, sous-comité SC 3, *Terminotique*.

La présente partie de l'ISO 1087 annule et remplace l'ISO 1087:1990, article 7. L'ISO 1087-1, actuellement en cours d'élaboration, constituera une révision technique des autres articles de l'ISO 1087:1990.

L'ISO 1087 comprend les parties suivantes, présentées sous le titre général *Travaux terminologiques — Vocabulaire*:

— *Partie 1: Théorie et application*

— *Partie 2: Applications logicielles*

## Introduction

The terms of this international terminology standard are given in a clustered ordering subsumed under a few general headings. Usually, terms necessary for the definition of other terms are preceding the entries of the latter. For convenience of use, however, an alphabetical index is to be found at the end of this part of ISO 1087.

The layout is designed according to ISO 10241. Thus, the elements of an entry appear in the following order:

entry number

preferred term (bold face), word class

short form (bold face)

admitted term

deprecated term: labelled "(deprecated)"

subject field (between angle brackets)

definition

examples

notes

} reference to another entry in bold face followed by entry number in brackets, when it is first mentioned

Except for entry number, preferred term and definition, elements appear only where appropriate.

Entries are nouns except where stated otherwise.

## Introduction

Les termes de la présente Norme internationale terminologique sont classés par thème dans plusieurs sous-groupes généraux. En principe, les termes qui entrent dans la définition d'autres termes précèdent les entrées de ces derniers. Néanmoins, pour des raisons pratiques, un index alphabétique est fourni à la fin de la présente partie de l'ISO 1087.

La structure est établie conformément à l'ISO 10241. Les éléments de chaque entrée sont affichés dans l'ordre suivant:

- le numéro d'entrée
  - le terme privilégié (en gras), la classe de mots
  - l'abréviation (en gras)
  - le(s) terme(s) admis
  - le terme rejeté, indiqué par «(rejeté)»
  - le domaine <affiché entre ces signes>
  - une définition
  - des exemples
  - des notes
- ITIL STANDARD PREVIEW  
(standards.iteh.ai)  
ISO 1087-2:2000  
<http://standards.iteh.ai/catalog/standards/sist/9694d9c1-caf2-4626-b603-c72335b55615/iso-1087-2-2000>
- une référence à une autre entrée, en gras et suivie par le numéro d'entrée entre parenthèses, quand elle est mentionnée la première fois

À l'exception du numéro d'entrée, toujours affiché, le terme privilégié et les différents éléments de la définition ne sont indiqués que lorsque cela est nécessaire.

Sauf indication contraire, les entrées sont des noms.



## Terminology work — Vocabulary —

### Part 2: Computer applications

## Travaux terminologiques — Vocabulaire —

### Partie 2: Applications logicielles

#### 1 Scope

This International Standard defines terms for language and information processing for applications in terminology work and terminography.

#### 2 General concepts

##### 2.1 information

⟨information processing⟩ knowledge concerning such things as facts, concepts, objects, events, ideas

NOTE 1 Adapted from ISO 2382-1:1993.

NOTE 2 Information is defined differently in other application areas.

##### 2.2 data

representation of **information** (2.1) in a formalized manner suitable for communication, interpretation, storing or processing

NOTE Adapted from ISO/IEC 2382-1:1993.

##### 2.3 data processing DP

systematic performance of operations upon **data** (2.2)

EXAMPLE Arithmetic or logic operations upon **data** (2.2), **merging** (8.5) or **sorting** (8.4) of data, assembling or compiling of programs, or operations on **text** (2.6), such as **text editing** (11.3), sorting, merging, storing, retrieving, **displaying** (8.6), or printing.

#### 1 Domaine d'application

La présente Norme internationale définit les termes utilisés pour le traitement du langage et de l'information dans les applications liées à la terminologie et à la terminographie.

#### 2 Concepts généraux

##### 2.1 information

⟨traitement de l'information⟩ connaissance d'éléments tels que des faits, des concepts, des objets, des événements et des idées

NOTE 1 Adapté de l'ISO 2382-1:1993.

NOTE 2 Le terme information est défini différemment dans d'autres domaines d'application.

##### 2.2 donnée

**informations** (2.1) représentées sous une forme conventionnelle convenant à la communication, à l'interprétation, au stockage ou au traitement

NOTE Adapté de l'ISO/CEI 2382-1:1993.

##### 2.3 traitement des données

déroulement systématique d'opérations sur des **données** (2.2)

EXEMPLE Opérations arithmétiques ou logiques sur des **données** (2.2), **fusion** (8.5) ou **tri** (8.4) de données, assemblage ou compilation de programmes, ou opérations sur du **texte** (2.6) comme l'**édition de texte** (11.3), le tri, la fusion, le stockage, la recherche, l'**affichage** (8.6) ou l'impression.

NOTE The term should not be used as a synonym for **information processing** (2.4).

[ISO 2382-1:1993]

## 2.4 information processing

systematic performance of operations upon **information** (2.1) that includes **data processing** (2.3)

NOTE 1 The term must not be used as a synonym for **data processing** (2.3).

NOTE 2 Adapted from ISO 2382-1:1993.

## 2.5 language processing information processing (2.4) on language

EXAMPLE Proofreading: proofreading can be carried out intellectually and by using a spell-checker. Both operations are complementary.

## 2.6 text data (2.2) in the form of **characters** (5.1), symbols, words, phrases, paragraphs, sentences, tables, or other character arrangements, intended to convey a meaning, and whose interpretation is essentially based upon the knowledge of some natural or artificial language

NOTE Adapted from ISO 2382-1:1993.

## 2.7 text corpus corpus

⟨language engineering⟩ systematic collection of machine-readable texts (see also 6.4) or parts of text prepared, coded and stored according to predefined rules

NOTE 1 A text corpus may be limited according to aspects of subject fields, size or time, e.g. mathematical texts, certain periodicals from 1986 onwards. It is used as source material for further linguistic analysis or terminology work.

NOTE 2 See also ISO 1087-1.

NOTE Il convient que ce terme ne soit pas utilisé comme un synonyme du terme **traitement de l'information** (2.4).

[ISO 2382-1:1993]

## 2.4 traitement de l'information déroulement systématique d'opérations sur des **informations** (2.1), incluant également le **traitement de données** (2.3)

NOTE 1 Il convient que ce terme ne soit pas utilisé comme un synonyme du terme **traitement des données** (2.3).

NOTE 2 Adapté de l'ISO 2382-1:1993.

## 2.5 traitement du langage traitement de l'information (2.4) sur le langage

EXEMPLE La correction d'un texte peut être réalisée intellectuellement ou en utilisant un correcteur orthographique. Ces deux opérations sont complémentaires.

## 2.6 texte données (2.2) sous la forme de **caractères** (5.1), de symboles, de mots, de phrases, de paragraphes, de tableaux, ou de tout autre disposition de caractères, dont l'objectif est de véhiculer du sens, et dont l'interprétation repose essentiellement sur la connaissance d'un langage naturel ou artificiel

NOTE Adapté de l'ISO 2382-1:1993

## 2.7 corpus de textes corpus

⟨ingénierie linguistique⟩ regroupement systématique de textes exploitables par une machine (voir également: 6.4) ou de fragments de texte préparés, codés et stockés selon des règles prédéfinies

NOTE 1 Un corpus de textes peut ne regrouper que les textes correspondant à certaines des caractéristiques du domaine étudié, telles que la taille ou la période, par exemple textes mathématiques, périodiques spécifiques parus depuis 1986. Un corpus est utilisé comme le matériau source à partir duquel peuvent être réalisés des analyses linguistiques ou des travaux de terminologie.

NOTE 2 Voir aussi l'ISO 1087-1.

## 2.8 type

(language engineering) linguistic unit in a **text** (2.6) representing a defined class

NOTE Such linguistic units are usually **character strings** (5.13) uninterrupted by **delimiters** (2.10).

## 2.9 token

(natural-language processing) occurrence of a **type** (2.8)

EXAMPLE If the class is defined as all **word forms** (2.11) of “good”, then all the occurrences of the word forms “good”, “better” and “best” are tokens.

## 2.10 delimiter

separator  
one or more **characters** (5.1) used to indicate the beginning or the end of a **character string** (5.13)

[ISO 2382-4:1987]

NOTE Blanks or punctuation marks often function as delimiters.

## 2.11 word form

morphosyntactical variant of a given word

EXAMPLE Indicate: indicates, visitor: visitors, visitor's.

NOTE In inflecting languages, word forms are often equivalent to inflected forms e.g. “go”, “goes”; but also: “go”, “went”, “gone”.

## 2.12 paradigm

(natural-language processing) class of **word forms** (2.11) belonging to a given word or **multi-word term** (2.14)

EXAMPLE Sell, sells, sold, selling

## 2.13 base form

reference form  
**word form** (2.11) chosen according to lexicographical conventions representing the forms of a **paradigm** (2.12)

## 2.8 type

(ingénierie linguistique) unité linguistique dans un **texte** (2.6) représentant une classe définie

NOTE Ces unités linguistiques sont généralement des **chaînes de caractères** (5.13) non interrompues par des **délimiteurs** (2.10).

## 2.9 occurrence

(traitement en langage naturel) nombre de fois où apparaît un **type** (2.8)

EXEMPLE Si la classe est définie comme l'ensemble des **formes lexicales** (2.11) du mot «bon», alors toutes les occasions où apparaissent les formes lexicales «bon» et «meilleur» sont des occurrences.

## 2.10 délimiteur

séparateur  
un ou plusieurs **caractères** (5.1) utilisés pour marquer le début ou la fin d'une **chaîne de caractères** (5.13)

[ISO 2382-4:1987]

NOTE Les blancs ou les signes de ponctuation sont souvent utilisés comme délimiteurs.

## 2.11 forme lexicale

variante morphosyntaxique d'un mot donné

EXEMPLE Indiquer: indiquons, indiquant; instituteur: instituteurs, institutrices

NOTE Dans les langages conjugués, les formes lexicales correspondent souvent aux formes conjuguées, par exemple «être», «est», mais aussi «être», «fut», «été».

## 2.12 paradigme

(traitement en langage naturel) classe de **formes lexicales** (2.11) se rapportant à un mot simple ou à un **terme composé** (2.14)

EXEMPLE Dormir, dort, dormi, dormant

## 2.13 forme de base

forme de référence  
**forme lexicale** (2.11) choisie selon des conventions lexicographiques et représentant les formes d'un **paradigme** (2.12)

NOTE 1 Adapted from ISO 1087:1990.

EXAMPLE bind: bind, bound, binds, binding

NOTE 2 The term “base form” is also applied to **multi-word terms** (2.14).

**2.14 multi-word term**

term consisting of more than one **character string** (5.13) (5.13)

**2.15 compressed form**

reduced form (deprecated) **character string** (5.13) having undergone **compression** (7.4)

EXAMPLE The string “input/output-algorithm” becomes the compressed form: “inputoutputalgorihtm”.

NOTE The term “reduced form” may be misleading in this context.

**2.16 deinflection**

deletion of inflectional elements in **word forms** (2.11)

**2.17 deinflected word form**

**word part** (2.18) remaining after **deinflection** (2.16)

**2.18 word part**

word segment **character string** (5.13) taken from a **word form** (2.11) for a specific application

**2.19 lemmatization**

process of deriving the **base form** (2.13)

EXAMPLE “go” becomes the **base form** (2.13) of “goes”, by **deinflection** (2.16); “go” also serves as the **base form** (2.13) of “went” by virtue of irregular verb conventions that do not conform to standard inflectional practice.

NOTE The result may also be called lemmatization.

NOTE 1 Adapté de l'ISO 1087:1990.

EXEMPLE Prendre: prends, pris, prenant

NOTE 2 Le terme «forme de base» s'applique également aux **termes composés** (2.14).

**2.14 terme composé**

terme comprenant une ou plusieurs **chaînes de caractères** (5.13)

**2.15 forme comprimée**

forme réduite (rejeté) **chaîne de caractères** (5.13) ayant subi une **compression** (7.4)

EXEMPLE Une fois comprimée, la chaîne «algorithme-entrée/sortie» devient: «algorithmeentréesortie».

NOTE Le terme «forme réduite» peut induire en erreur dans ce contexte.

**2.16 suppression des désinences**

suppression d'éléments flexionnels dans des **formes lexicales** (2.11)

**2.17 forme lexicale sans désinence**

**partie de mot** (2.18) restant à la suite de la suppression des **désinences** (2.16)

**2.18 partie de mot**

segment de mot **chaîne de caractères** (5.13) tirée d'une **forme lexicale** (2.11) pour une application donnée

**2.19 lemmatisation**

processus consistant à dériver la **forme de base** (2.13)

EXEMPLE «naître» devient la **forme de base** (2.13) de «naît» par suppression des **désinences** (2.16), mais «naître» est également la forme de base de «naquit» en vertu des règles régissant les verbes irréguliers, qui diffèrent des pratiques courantes de conjugaison.

NOTE Le résultat peut également être appelé lemmatisation.