



Norme
internationale

ISO/IEC 14651

Septième édition
2025-07

Technologies de l'information —
Classement international et
comparaison de chaînes de
caractères — Méthode de
comparaison de chaînes de
caractères et description du modèle
commun et adaptable d'ordre de
classement

*Information technology — International string ordering and
comparison — Method for comparing character strings and
description of the common template tailorable ordering*



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO/IEC 2025

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, 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, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos	iv
Introduction	v
1 Domaine d'application	1
2 Références normatives	2
3 Termes et définitions	2
4 Symboles et conventions	3
5 Conformité	3
6 Comparaison de chaînes	4
6.1 Prétraitement des chaînes de caractères avant comparaison	4
6.2 Construction des clés et comparaison	5
6.2.1 Considérations préliminaires	5
6.2.2 Formation de clés de tri de référence	6
6.2.3 Méthode de comparaison de référence pour le tri des chaînes de caractères	8
6.2.4 Définition de l'ordre des clés	8
6.3 Table-modèle commune: Formation et interprétation	9
6.3.1 Généralités	9
6.3.2 Règles de syntaxe BNF pour les tables-modèles communes de l' Annexe A	10
6.3.3 Contraintes de forme	12
6.3.4 Interprétation des tables adaptées	13
6.3.5 Évaluation des tables de poids	15
6.3.6 Conditions d'équivalences de tables spécifiques	15
6.3.7 Conditions d'équivalence des résultats	15
6.4 Déclaration d'un delta	15
6.5 Noms des tables-modèles communes et déclaration de nom	17
Annexe A (normative) Tables-modèles communes	18
Annexe B (informative) Exemple de deltas d'adaptation	20
Annexe C (informative) Prétraitement	27
Annexe D (informative) Annexe didactique sur les solutions apportées par le présent document aux problèmes de tri lexical	43
Annexe E (informative) Recherche et correspondances floues	47
Bibliographie	49

Avant-propos

L'ISO (Organisation internationale de normalisation) et l'IEC (Commission électrotechnique internationale) forment le système spécialisé de la normalisation mondiale. Les organismes nationaux membres de l'ISO ou de l'IEC participent au développement de Normes Internationales par l'intermédiaire des comités techniques créés par l'organisation concernée afin de s'occuper des domaines particuliers de l'activité technique. Les comités techniques de l'ISO et de l'IEC collaborent dans des domaines d'intérêt commun. D'autres organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO et l'IEC participent également aux travaux.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de document ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives ou www.iec.ch/members_experts/refdocs).

L'ISO et l'IEC attirent l'attention sur la possibilité que la mise en œuvre du présent document implique l'utilisation d'un ou plusieurs brevets. L'ISO et l'IEC ne se prononcent pas sur la preuve, la validité ou l'applicabilité de tout droit de brevet revendiqué à cet égard. À la date de publication du présent document, l'ISO et l'IEC n'avaient pas reçu notification de brevets pouvant être nécessaires à la mise en œuvre du présent document. Toutefois, les utilisateurs sont avertis que ces informations ne sont pas nécessairement les plus récentes et qu'ils peuvent obtenir des informations plus récentes dans les bases de données de brevets disponibles à l'adresse www.iso.org/patents et <https://patents.iec.ch>. L'ISO et l'IEC ne peuvent être tenues responsables de l'identification de tout ou partie de ces droits de brevet.

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la nature volontaire des normes, la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir www.iso.org/avant-propos. Pour l'IEC, voir www.iec.ch/understanding-standards.

Le présent document a été élaboré par le comité technique mixte ISO/IEC JTC 1, *Technologies de l'information*, sous-comité SC 2, *Jeux de caractères codés*.

Cette septième édition annule et remplace la sixième édition (ISO/IEC 14651:2020) qui a fait l'objet d'une révision technique.

Les principales modifications sont les suivantes:

- les tables-modèles communes sont désormais référencés de manière externe, éliminant la nécessité d'une mise à jour fréquente du présent document.

Il convient que l'utilisateur adresse tout retour d'information ou toute question concernant le présent document à l'organisme national de normalisation de son pays. Une liste exhaustive desdits organismes se trouve aux adresses www.iso.org/fr/members.html et www.iec.ch/national-committees.

Introduction

Le présent document fournit une méthode universelle de mise en ordre des données textuelles. Il référence les tables-modèles communes qui, lorsqu'elles sont adaptées, peuvent satisfaire aux exigences de tri d'une langue donnée, tout en triant de manière raisonnable les autres écritures.

Les tables-modèles communes sont conçues de sorte qu'une adaptation s'avère nécessaire pour chaque environnement local. C'est pourquoi la conformité au présent document requiert que les modifications à ces tables communes, appelées «deltas», soient déclarées de manière à documenter les différences dans les résultats.

Le présent document décrit une méthode pour classer l'information textuelle de manière indépendante du contexte.

L'ISO/IEC 30112 contient des dispositions pour le tri complémentaires à celles du présent document; on y trouve aussi des renseignements complémentaires sur les mots-clés définis dans le présent document et utilisés pour le tri.

Technologies de l'information — Classement international et comparaison de chaînes de caractères — Méthode de comparaison de chaînes de caractères et description du modèle commun et adaptable d'ordre de classement

1 Domaine d'application

Le présent document définit une méthode de comparaison de référence. Cette méthode est applicable à deux chaînes de caractères ou plus pour déterminer leur ordre de classement dans une liste triée. La méthode peut être appliquée aux chaînes contenant des caractères du répertoire complet de l'ISO/IEC 10646. Cette méthode est également applicable aux sous-ensembles de ce répertoire pour produire des résultats de tri valides (après adaptation) pour un ensemble donné de langues pour chaque script. Cette méthode de référence utilise des tables de tri dérivées soit des tables-modèles communes de classement définies dans le présent document, soit d'une de leurs adaptations. Le format de la table-modèle commune est décrit en notation BNF (*Backus-Naur Form*, *Forme de Backus-Naur*). Son emploi est normatif dans le présent document;

Le présent document définit également les éléments de syntaxe pour adapter ces tables-modèles communes utilisées par la méthode de comparaison de référence.

De plus, il définit les exigences relatives à une déclaration des différences (delta) entre une table de tri et une table-modèle commune donnée, y compris les éléments d'adaptation.

Ces tables-modèles communes décrivent un ordre pour tous les caractères encodés dans les éditions actuelles et passées de l'ISO/IEC 10646, y compris les amendements. Elles permettent de spécifier un ordre complètement déterministe. Ces tables constituent le point de départ permettant de préciser un ordre de classement adapté aux règles de classement locales, sans qu'il soit nécessaire de connaître tous les systèmes d'écriture repris dans le jeu universel de caractères codés (JUC).

Toutes ces tables-modèles communes comportent des noms de référence qui sont liés à un stade particulier de développement de l'ISO/IEC 10646 relative au jeu universel de caractères codés ou d'une version particulière du standard Unicode. Ces noms et leur relation avec l'ISO/IEC 10646 ou le répertoire du standard Unicode sont spécifiés par un document de référencement externe: Unicode Technical Standard, UTS #10, Unicode Collation Algorithm.

Le présent document n'impose pas ce qui suit:

- une méthode particulière de comparaison; toute méthode équivalente conduisant aux mêmes résultats est acceptable;
- un format précis pour décrire ou pour adapter les tables dans une mise en œuvre donnée;
- des symboles spécifiques à utiliser par les mises en œuvre;
- un format interne particulier pour les clés intermédiaires utilisées dans les comparaisons ou pour la table de tri. L'utilisation de clés numériques n'est pas spécifiée non plus;
- un ordre dépendant du contexte;
- un prétraitement particulier des chaînes de caractères avant comparaison.

NOTE 1 Bien que ceci ne soit pas spécifié par le présent document, il s'avère courant de préparer les chaînes de caractères avant leur comparaison (voir l'[Annexe C](#)).

NOTE 2 L'[Annexe D](#) décrit les problèmes qui ont donné lieu au présent document avec leurs solutions anticipées.