NORME INTERNATIONALE

ISO/IEC 14651

Sixième édition
2020-12

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



Numéro de référence
ISO/IEC 14651:2020(F)

© ISO/IEC 2020



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO/IEC 2020

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 Préliminaires ... 5

6.2.2 Méthode de référence de construction des clés ... 6

6.2.3 Méthode de comparaison de référence pour le tri des chaînes de caractères ... 8

6.2.4 Méthode de comparaison de référence pour le tri des chaînes de caractères ... 9

6.3 Table-modèle commune: composition et interprétation ... 10

6.3.1 Généralités ... 10

6.3.2 Règles de syntaxe BNF pour la table-modèle commune de l'Annexe A ... 10

6.3.3 Contraintes de forme ... 12

6.3.4 Interprétation des tables adaptées ... 14

6.3.5 Évaluation des tables de poids ... 15

6.3.6 Conditions d'équivalence de tables spécifiques ... 15

6.3.7 Conditions d'équivalence des résultats ... 16

6.4 Déclaration d'un delta ... 16

6.5 Nom de la table-modèle commune et déclaration de nom ... 18

Annexe A (normative) **Table-modèle commune** ... **19**

Annexe B (informative) **Exemples de deltas d'adaptation** ... **21**

Annexe C (informative) **Prétraitement** ... **31**

Annexe D (informative) **Annexe didactique sur les solutions apportées par le présent document aux problèmes de tri lexical** ... **47**

Annexe E (informative) **Recherches et correspondances floues** ... **51**

Bibliographie ... **53**

© ISO/IEC 2020 – Tous droits réservés

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 documents 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).

L'attention est attirée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO et l'IEC ne sauraient être tenues pour responsables de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets) ou dans la liste des déclarations de brevets reçues par l'IEC (voir patents.iec.ch).

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/iso/avant-propos.

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

Cette sixième édition annule et remplace la cinquième édition (ISO/IEC 14651:2019), qui a fait l'objet d'une révision technique.

Les principales modifications par rapport à l'édition précédente sont les suivantes:

— le présent document ajoute les données de classement des nouveaux caractères normalisés dans la sixième édition de l'ISO/IEC 10646 (2020) ;

— le contenu de 6.2.2 a été révisé pour le rendre plus complet ;

— les poids du caractère U+A9B5 (DIACRITIQUE VOYELLE JAVANAISE TOLONG) ont été modifiés, car ce dernier est considéré comme une variante du caractère U+A9B4 (DIACRITIQUE VOYELLE JAVANAISE TARUNG). Cela se devait d'être corrigé.

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 à l'adresse www.iso.org/members.html.

 © ISO/IEC 2020 – Tous droits réservés

Introduction

Le présent document fournit une méthode universelle de mise en ordre des données textuelles. Elle fournit également une table-modèle commune qui, lorsqu'elle est adaptée, peut satisfaire aux exigences de tri d'une langue donnée, tout en triant de manière raisonnable les autres écritures.

La table-modèle commune est conçue 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 cette table commune, 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 TR 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.

© ISO/IEC 2020 – Tous droits réservés

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 ce qui suit.

— Une méthode de référence pour la comparaison de deux chaînes de caractères ayant pour but de déterminer leur ordre de classement dans une liste triée. La méthode s'applique à des chaînes utilisant le répertoire complet de l'ISO/IEC 10646, des sous-répertoires tels que ceux des divers jeux normalisés ISO/IEC à 8 bits ou tout autre jeu de caractères, normalisé ou non, et permet de produire des résultats de tri valables (après adaptation) pour un ensemble de langues de chaque système d'écriture. Cette méthode de référence utilise des tables de tri dérivées soit de la table-modèle commune de classement définie dans le présent document, soit d'une de ses adaptations. La méthode procure un format de référence de la table-modèle commune. Ce format est décrit en notation BNF (forme de Backus-Naur, *Backus-Naur Form*). Son emploi est normatif *dans* le présent document.

— Une table-modèle commune de classement utilisée par la méthode de référence. Cette table décrit un ordre de base pour tous les caractères du standard Unicode 13.0[27] compris dans l'ISO/IEC 10646:2020. Tout cela permet de spécifier un ordre complètement déterministe. Cette table constitue 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).

NOTE 1 Cette table-modèle commune de classement est destinée à être modifiée pour satisfaire aux besoins d'environnements locaux. L'avantage principal de cette pratique, sur le plan mondial, réside dans le fait que, pour d'autres systèmes d'écriture que celui de l'utilisateur, aucune modification n'est nécessaire et cet ordre demeurera aussi cohérent que possible et prévisible dans un contexte international.

NOTE 2 Le répertoire de caractères utilisé dans le présent document est équivalent à celui du standard Unicode, version 13.0[27].

— Un nom de référence représentant cette version particulière de la table-modèle commune, à utiliser comme point de départ à toute adaptation. Ce nom implique notamment que la table est liée à un stade de développement particulier du jeu universel de caractères codés (ISO/IEC 10646).

— Des exigences pour la déclaration de différences (delta) entre une table de tri et la table-modèle commune.

Le présent document *ne* spécifie *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 précis à utiliser par les mises en œuvre, sauf pour ce qui est du nom de la table-modèle commune de classement.

— Une interface utilisateur particulière destinée à choisir les options.

© ISO/IEC 2020 – Tous droits réservés

— 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 souvent nécessaire de préparer les chaînes de caractères avant leur comparaison (cf. l'Annexe C).

NOTE 2 L'Annexe D décrit les problèmes qui ont donné lieu à la présente Norme internationale avec leurs solutions anticipées.

2 Références normatives

Les documents suivants sont cités dans le texte de sorte qu'ils constituent, pour tout ou partie de leur contenu, des exigences du présent document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document de référence s'applique (y compris les éventuels amendements).

ISO/IEC 10646:2020, *Technologies de l'information — Jeu universel de caractères codés (JUC)*

© ISO/IEC 2020 – Tous droits réservés