NORME INTERNATIONALE

ISO/IEC 14651

Quatrième édition
2016-02-15

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 14651:2016(F)

© ISO/IEC 2016



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO/IEC 2016, Publié en Suisse

Droits de reproduction réservés. Sauf indication contraire, 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, l'affichage sur l'internet ou sur un Intranet, sans autorisation écrite préalable. Les demandes d'autorisation peuvent être adressées à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Sommaire

Page

Avant-propos ... **iv**

Introduction ... **v**

1 Domaine d'application ... **1**

2 Conformité ... **2**

3 Références normatives ... **2**

4 Termes et définitions ... **2**

5 Symboles et abréviations ... **3**

6 Comparaison de chaînes ... **4**
6.1 Prétraitement des chaînes de caractères avant comparaison ... 4
6.2 Construction des clés et comparaison ... 4
6.2.1 Préliminaires ... 4
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 ... 7
6.3 Table-modèle commune: composition et interprétation ... 8
6.3.1 Règles de syntaxe BNF pour la table-modèle commune de l'Annexe A ... 8
6.3.2 Contraintes de forme ... 11
6.3.3 Interprétation des tables adaptées ... 12
6.3.4 Évaluation des tables de poids ... 13
6.3.5 Conditions d'équivalence de tables spécifiques ... 13
6.3.6 Conditions d'équivalence des résultats ... 14
6.4 Déclaration d'un delta ... 14
6.5 Nom de la table-modèle commune et déclaration de nom ... 16

Annexe A (normative) **Table-modèle commune** ... **17**

Annexe B (informative) **Exemples de deltas d'adaptation** ... **19**

Annexe C (informative) **Prétraitement** ... **28**

Annexe D (informative) **Annexe didactique sur les solutions apportées par la présente Norme internationale aux problèmes de tri lexical** ... **44**

Annexe E (informative) **Recherches et correspondances floues** ... **48**

Bibliographie ... **50**

© ISO/IEC 2016 – 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. Dans le domaine des technologies de l'information, l'ISO et l'IEC ont créé un comité technique mixte, l'ISO/IEC JTC 1.

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 appelée 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 la CEI 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).

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 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'OMC concernant les obstacles techniques au commerce (OTC), voir le lien suivant: Avant-propos — Informations supplémentaires.

Le comité chargé de l'élaboration du présent document est l'ISO/CEI/JTC 1, *Technologies de l'information*, sous-comité SC 2, *Jeux de caractères codés*.

Cette quatrième édition annule et remplace la troisième édition (ISO/CEI 14651:2011), qui a fait l'objet d'une révision technique.

La présente version française de l'ISO 14651:2016 correspond à la version anglaise corrigée du 2016-05-01.

© ISO/IEC 2016 – Tous droits réservés

Introduction

La présente Norme internationale fournit une méthode universelle de mise en ordre des données textuelles. Elle fournit également une table-modèle commune qui, lorsque 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 telle sorte qu'une adaptation s'avère nécessaire pour chaque environnement local. C'est pourquoi la conformité à la présente Norme internationale requiert que les modifications à cette table commune, appelées «deltas», soient déclarées de manière à documenter les différences dans les résultats.

La présente Norme Internationale décrit une méthode pour classer l'information textuelle de manière indépendante du contexte.

L'ISO/CEI/TR 30112 contient des dispositions complémentaires pour le tri à celle de la présente Norme internationale; on y trouvera aussi des renseignements complémentaires sur les mots-clés de tri définis dans la présente Norme internationale.

© ISO/IEC 2016 – 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

La présente Norme internationale 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/CEI 10646, des sous-répertoires tels que ceux des divers jeux normalisés ISO/CEI à 8 bits ou tout autre jeu de caractères, 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 la présente Norme internationale, 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). Son emploi est normatif *dans* la présente Norme internationale.

— 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 8.0 compris dans l'ISO/CEI 10646:2014 et son amendement 1. 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 la présente Norme internationale est équivalent à celui du standard Unicode, version 6.0.

— 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/CEI 10646).

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

La présente Norme internationale *ne* prescrit *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 mise 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.

— 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 prescrite non plus.

© ISO 2016 – Tous droits réservés

— Un ordre dépendant du contexte.

— Un prétraitement particulier des chaînes de caractères avant comparaison.

NOTE Bien que ceci ne soit pas prescrit par la présente Norme internationale, il s'avère souvent nécessaire de préparer les chaînes de caractères avant leur comparaison (cf. l'Annexe C).

Bien que l'on ne prescrive aucune interface utilisateur destinée à choisir les options ou à adapter la table-modèle commune, la conformité exige de toujours déclarer un delta, c'est-à-dire l'ensemble des différences par rapport à cette table. Il est fortement recommandé que l'application présente à l'utilisateur les options et adaptations disponibles.

2 Conformité

Un processus est conforme à la présente Norme internationale s'il produit des résultats identiques à ceux qui résultent de l'application des spécifications décrites en 6.2 à 6.5.

Toute déclaration de conformité à la présente Norme internationale doit être accompagnée, directement ou par référence, d'une déclaration de ce qui suit.

— Le nombre de niveaux de tri que le processus peut utiliser; ce nombre doit être égal ou supérieur à trois.

— Si le paramètre de traitement forward,position est permis.

— Si le paramètre de traitement backward est permis et à quel niveau.

— Le *delta* d'adaptation décrit en 6.4 et le nombre de niveaux définis dans ce delta.

— Si un processus de prétraitement est utilisé, la méthode utilisée doit être déclarée.

Il incombe au producteur de montrer en quoi sa déclaration de delta est reliée à la syntaxe de la table décrite en 6.3, et comment la méthode de comparaison utilisée, si elle est différente de celle mentionnée à l'Article 6, peut être considérée comme produisant les mêmes résultats que ceux prescrits par la méthode décrite à l'Article 6. L'usage d'un processus de prétraitement est optionnel et ses détails ne sont pas précisés dans la présente Norme internationale.

3 Références normatives

Les documents de référence suivants sont indispensables pour l'application 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/CEI 10646:2014, *Technologies de l'information — Jeu universel de caractères codés (JUC)*

ISO/CEI 10646:1/Amd 1:2015, *Technologies de l'information — Jeu universel de caractères codés (JUC)/ Amendment 1.*

4 Termes et définitions

Pour les besoins du présent document, les termes et définitions suivants s'appliquent.

4.1
chaîne de caractères
suite de caractères considérée comme un objet simple

4.2
symbole de tri
symbole (4.12) utilisé pour préciser les poids attribués à un *élément de tri* (4.4)

 © ISO/IEC 2016 – Tous droits réservés

4.3
table (de poids) de tri
table reliant les *éléments de tri* (4.4) aux *éléments de poids* (4.14)

4.4
élément de tri
suite constituée d'un ou de plusieurs caractères considérés comme une seule entité aux fins du *tri* (4.7)

4.5
delta
liste des différences que présente une *table de classement* (4.3) donnée par rapport à une autre

Note 1 à l'article: Une table de tri donnée associée à un delta donné forme une nouvelle table de tri.

Note 2 à l'article: Sauf mention contraire, le terme «delta» désignera les différences par rapport à la table-modèle commune définie dans la présente Norme internationale.

4.6
niveau (de tri)
numéro d'une *sous-clé* (4.11) dans la série de sous-clés formant une clé

4.7
tri
procédé par lequel on détermine si, de deux chaînes, la première est plus petite, égale ou plus grande que la seconde

4.8
clé de tri
série de *sous-clés* (4.11) utilisée pour déterminer un ordre

4.9
prétraitement
procédé par lequel des *chaînes de caractères* (4.1) sont transformées en d'autres chaînes avant le calcul de la *clé de tri* (4.8) de chaque chaîne

4.10
méthode de comparaison de référence
méthode de détermination de l'ordre relatif de deux *clés de tri* (4.8)

Note 1 à l'article: Voir l'Article 6.

4.11
sous-clé
suite de poids calculée pour une *chaîne de caractères* (4.1)

4.12
symbole
élément de tri (4.4)

4.13
poids (de tri)
entier positif, utilisé dans les *sous-clés* (4.11), pour indiquer l'ordre relatif des éléments de tri

4.14
élément de poids
liste d'un certain nombre de poids séquentiellement ordonnés par niveau

5 Symboles et abréviations

Selon l'ISO/CEI 10646, les caractères se représentent à l'aide de U*X*, où *X* correspond à une série d'un à huit chiffres hexadécimaux (où toutes les lettres de la série de chiffres hexadécimaux sont en

© ISO/IEC 2016 – Tous droits réservés