

Troisième édition  
2011-08-15

AMENDEMENT 2  
2015-04-01

---

---

**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**

**iTeh STANDARD PREVIEW  
(standards.iteh.ai)**  
**AMENDEMENT 2**

ISO/IEC 14651:2011/Amd.2:2015  
<https://standards.iteh.ai/catalog/standards/sist/611ba19e-8324-4c2f-9733-5bb534feb14f/iso-iec-14651-2011-amd-2-2015>  
Information technology — International string ordering and  
comparison — Method for comparing character strings and  
description of the common template tailorable ordering

*AMENDMENT 2*

---

---

Numéro de référence  
ISO/IEC 14651:2011/Amd.2:2015(F)



## iTeh STANDARD PREVIEW (standards.iteh.ai)

[ISO/IEC 14651:2011/Amd 2:2015](https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015)

<https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015>



### DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO/IEC 2015

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  
Case postale 56 • CH-1211 Geneva 20  
Tel. + 41 22 749 01 11  
Fax + 41 22 749 09 47  
E-mail [copyright@iso.org](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Publié en Suisse

## 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 (IEC) en ce qui concerne la normalisation électrotechnique.

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](http://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 ne saurait être tenue pour responsable 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](http://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. <https://standards.iteh.ai/catalog/standards/sist/61aba9e-8524-4c2f-9753-5bb534fe614f/iso-iec-14651-2011-amd-2-2015>

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

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

[ISO/IEC 14651:2011/Amd 2:2015](https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015)

<https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015>

# 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

## AMENDEMENT 2

Page 1, article 1

Remplacer la 2<sup>e</sup> puce par la suivante.

Une table-modèle commune de classement utilisée par la méthode de référence. Cette table décrit un ordre de base pour les caractères correspondant au répertoire de caractères du standard Unicode 7.0, sous-ensemble substantiel des caractères codés dans l'ISO/CEI 10646:2014. 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).

Page 2, Article 3

Remplacer la référence normative par la suivante.

ISO/CEI 10646:2014, *Technologies de l'information — Jeu universel de caractères codés (JUC)*

Page 16, Paragraphe 6.5

Remplacer le Paragraphe 6.5 par le suivant:

### 6.5 Nom de la table-modèle commune et déclaration de nom

Le nom ISO 14651\_2014\_TABLE1 doit être utilisé pour désigner la table-modèle commune dans toute référence externe à cette table comme base dans un contexte donné, que ce soit un processus, un contrat ou des exigences d'approvisionnement. Si un autre nom doit être utilisé pour des raisons pratiques, une déclaration de conformité doit indiquer la correspondance entre cet autre nom et le nom ISO 14651\_2014\_TABLE1.

L'utilisation d'un nom précis est nécessaire pour gérer les différents états de développement de cette table. Ceci découle de la nature du répertoire de caractères de référence, amenée à s'étendre pendant des années, voire des décennies.

Page 17, Annexe A

Remplacer l'Annexe A par la suivante:

#### Annex A (normative) Table-modèle commune

Dans le but de minimiser les problèmes de formatage et les risques d'erreurs de reproduction, la table-modèle commune est fournie séparément, comme composante normative de la présente Norme internationale. Le nom de fichier dans cette version linguistique de la norme diffère du nom de référence normative précisé en 6.5 de la présente Norme internationale à cause de l'existence de versions du fichier avec des commentaires en d'autres langues. Le fichier pour cette version linguistique peut être récupéré du site web de l'ITTF à l'URL suivant:

[http://www.iso.org/ittf/ISO14651\\_2014\\_TABLE1\\_fr.txt](http://www.iso.org/ittf/ISO14651_2014_TABLE1_fr.txt)

Il existe une version anglaise officielle du fichier qui ne diffère que par les commentaires (le contenu technique est identique) et dont le nom est ISO 14651\_2014\_TABLE 1\_en.txt

NOTE 1 La présente norme internationale déconseille mais n'empêche pas la référence aux anciennes tables, qui contiennent les informations de tri des répertoires des versions précédentes de l'ISO/CEI 10646 et de leurs amendements. Les anciennes tables peuvent être récupérées des URL suivants:

- [informations de tri du répertoire de l'ISO/CEI 10646:1993-1 avec ses amendements 1 à 9] [http://www.iso.org/ittf/ISO14651\\_2000\\_TABLE1.htm](http://www.iso.org/ittf/ISO14651_2000_TABLE1.htm).
- [informations de tri du répertoire combiné de l'ISO/CEI 10646-1:2000 et de l'ISO/CEI 10646-2:2001] [http://www.iso.org/ittf/ISO14651\\_2002\\_TABLE1\\_fr.txt](http://www.iso.org/ittf/ISO14651_2002_TABLE1_fr.txt)
- [informations de tri du répertoire de l'ISO/CEI 10646:2003] [http://www.iso.org/ittf/ISO14651\\_2003\\_TABLE1\\_fr.txt](http://www.iso.org/ittf/ISO14651_2003_TABLE1_fr.txt)
- [informations de tri du répertoire de l'ISO/CEI 10646:2003 avec ses amendements 1 à 2] [http://www.iso.org/ittf/ISO14651\\_2006\\_TABLE1\\_fr.txt](http://www.iso.org/ittf/ISO14651_2006_TABLE1_fr.txt)
- [informations de tri du répertoire de l'ISO/CEI 10646:2003 avec ses amendements 1 à 4] [http://www.iso.org/ittf/ISO14651\\_2008\\_TABLE1\\_fr.txt](http://www.iso.org/ittf/ISO14651_2008_TABLE1_fr.txt)
- [informations de tri du répertoire de l'ISO/CEI 10646:2003 avec ses amendements 1 à 4] [http://www.iso.org/ittf/ISO14651\\_2010\\_TABLE1\\_fr.txt](http://www.iso.org/ittf/ISO14651_2010_TABLE1_fr.txt)
- [informations de tri du répertoire de l'ISO/CEI 10646:2003 avec ses amendements 1 à 4] [http://www.iso.org/ittf/ISO14651\\_2012\\_TABLE1\\_fr.txt](http://www.iso.org/ittf/ISO14651_2012_TABLE1_fr.txt)

La table-modèle commune actuelle prend en compte le répertoire de caractères du standard Unicode 7.0 compris dans l'ISO/CEI 10646:2014. <https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733->

NOTE 2 Le répertoire visé par la présente norme internationale est équivalent au répertoire du standard Unicode, version 7.0 (voir la publication intitulée *The Unicode Standard Version 7.0*, publiée par le consortium Unicode). Ce répertoire constitue un sous-ensemble substantiel de l'ISO/CEI 10646 :2014 et non le jeu complet, en raison de questions liées à la synchronisation des deux documents.

Quand des données de tri applicables à d'autres amendements à l'ISO/CEI 10646 seront disponibles, la présente Norme internationale, avec notamment sa table-modèle commune, sera amendée pour tenir compte du tri des caractères et écritures ajoutés. Pour satisfaire les exigences culturelles de communautés spécifiques, des déclarations de delta devront être appliquées à la table amendée telle que définie dans la présente Norme internationale.

La version courante de la présente Norme internationale désigne cette table sous le nom de **ISO 14651\_2014\_TABLE1**.

Page 45, après l'annexe D

Insérer une nouvelle annexe, l'Annexe E.

## Annexe E (informative) Recherches et correspondances floues

La recherche d'information textuelle utilise largement, au niveau fondamental de la comparaison de la chaîne recherchée avec le corpus cible, les mêmes concepts et les mêmes conventions que le tri.

En effet, il est très souvent utile de pouvoir effectuer une recherche sans égard à la casse (classement entre majuscules et minuscules), à l'accentuation ou aux caractères spéciaux. Ainsi, une recherche du mot-clé « contremaitre » pourra-t-elle trouver comme cibles légitimes les mots « contremaître », « contre-maître » (cette dernière étant une graphie ancienne du même mot), ou contremaitre (sans accent circonflexe sur le i, nouvelle graphie admise dans ce cas non-ambigu. Elle pourrait tout aussi

bien retrouver « CONTREMAÎTRE », puisque quelques usagers de la langue française ne mettent pas d'accents sur les majuscules, et que ce mot peut très bien se retrouver dans un corpus.

Pour assurer l'adéquation de ces recherches – floues ou exactes – avec les attentes des utilisateurs, il est approprié de considérer que les fonctions de comparaison utilisées devraient être fondées sur la même méthode de comparaison que celle du tri.

Toutefois, les notions de « flou » utilisées en recherche d'information débordent largement les simples notions d'indépendance de la casse, de l'accentuation et des caractères spéciaux. Il existe, notamment des fonctions de recherche sur les préfixes, tenant compte de la morphologie des mots (recherche de mots par leur racine), de la phonétique ou même des équivalences dans une langue étrangère.

Une recherche floue complexe – besoin qui va très au-delà du domaine d'application de la présente Norme internationale – pourrait par exemple considérer comme équivalentes deux chaînes prononcées de la même façon, comme dans le cas des homophones « cinq sens » et « Saint-Saëns ». De la même manière, l'on pourrait considérer comme équivalentes les formes grammaticales infléchies de mots comme « œil » et « yeux ». Un exemple semblable en anglais lierait les mots « man » et « men ».

Étant donné la grande variété des notions de flou possibles et le fait que les circonstances d'une recherche peuvent dicter le choix de l'une ou l'autre sans égard à leur importance relative dans le tri, nous ne discuterons ici que des simples cas basés sur les différents niveaux de pondération utilisés dans la présente Norme internationale.

Si l'on considère le champ d'application de la présente Norme internationale, la fonction de recherche devrait utiliser la même méthode de comparaison, en séparant les caractères de base, les signes diacritiques, la casse et les autres caractères (caractères spéciaux, signes typographiques et symboles). Toutefois, certains niveaux peuvent être omis lors de la comparaison, selon le niveau de précision nécessaire pour trouver une correspondance.

Cela se traduit comme suit, par exemple :

- Une recherche indépendante des signes diacritiques devrait considérer comme équivalentes (l'équivalence est marquée par « <> » dans les exemples) toutes les chaînes qui diffèrent seulement au niveau 2 dans le processus de comparaison.

Exemple : contremaître <> contremaitre

- Une recherche indépendante de la casse devrait considérer comme équivalentes toutes les chaînes qui diffèrent seulement au niveau 3 dans le processus de comparaison

Exemple : contremaître <> CONTREMAÎTRE

- Une recherche indépendante des caractères spéciaux (caractères « ignorables ») devrait considérer comme équivalentes toutes les chaînes qui diffèrent seulement au niveau 4 dans le processus de comparaison.

Exemple : contremaître <> contre-maître

- Une recherche qui veut cibler toutes les variantes vaguement similaires devrait considérer comme équivalents tous les mots qui correspondent seulement au niveau 1.

Exemple :

contremaître <> contre-maitre <> contremaitre <> CONTREMAÎTRE <> Contre-maître

[etc.]

Ce dernier comportement doit être considéré comme le cas général lors de recherches simples et est préférable en l'absence de besoins spécialisés.

Il est recommandé de faire en sorte que l'utilisateur soit informé des paramètres de recherche à l'aide d'éléments appropriés de l'interface utilisateur.

NOTE Pour plus d'information à ce sujet, voir :

- Standard sur le tri alphabétique et la recherche de chaînes de caractères, notamment les articles 4 et 6 (publication en français seulement);
- Unicode Technical Standard #10 Unicode Collation Algorithm, clause 8.2 (publication en anglais seulement).

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

[ISO/IEC 14651:2011/Amd 2:2015](https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015)

<https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015>

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

[ISO/IEC 14651:2011/Amd 2:2015](https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015)

<https://standards.iteh.ai/catalog/standards/sist/61abaf9e-8324-4c2f-9733-5bb534fe614f/iso-iec-14651-2011-amd-2-2015>