

Fonctions de systèmes - Soutien des langues nationales

Alain LaBonté
informaticien-conseil
Secrétariat du Conseil du trésor
Gouvernement du Québec

Ce document a été publié originellement en 1988 par le ministère des Communications du Québec. La version originelle a été revue et corrigée en août 1998 par l'auteur, M. Alain LaBonté, maintenant à l'emploi du Secrétariat du Conseil du trésor, qui a actuellement la responsabilité de ce dossier. Ce texte a servi au développement des outils de francisation Francis.

Reproduction et traduction autorisées, à condition que la source soit citée et que l'auteur en soit avisé.

Contacteur : Alain LaBonté

Un service est requis au niveau des systèmes d'exploitation pour assurer le soutien cohérent des caractères nationaux à l'aide des huit (8) fonctions suivantes: TRI, FUSION, COMPARAISON, REPÉRAGE, TRADUCTION DE CLEF, TRADUCTION DE CLEF INVERSÉE, CONVERSION MAJUSCULES/MINUSCULES et TRADUCTION DE PAGES DE CODES. Ces fonctions sont nécessaires pour la manipulation de chaînes de caractères dans un contexte d'utilisation de données mixtes, accentuées et non accentuées, dans les systèmes informatiques réels. Elles rendront plus flexible le traitement des anciens et des nouveaux modes de représentation des données à caractère nominatif (noms, adresses, etc.). Ces fonctions impliquent le respect des règles de classement propres à chaque langue. Elles sont, de plus, indépendantes de toute page de codes (c'est-à-dire qu'elles nécessitent la « connaissance » par le système des propriétés de la page de codes active). Ces fonctions devraient être utilisées par toutes les applications sensibles à des environnements linguistiques complexes.

Note : Sauf indication contraire, les longueurs de toutes les chaînes sont connues de façon implicite à l'exécution, par un mécanisme fourni par les producteurs.

PARAMÈTRES FONCTIONNELS

LANGUE UTILISÉE

(français, anglais, allemand, danois, etc.)

FONCTIONS

1- TRI

Sous-paramètres :

LISTE1 : Liste à trier

LISTE2 : Liste triée (résultat)

Le résultat doit être conforme aux règles de classement de la langue utilisée.

Ex. d'une telle liste triée en français :

cote	<i>Veillez noter les règles subtiles en français, où les accents ne sont pas considérés sauf pour de tels</i>
Cote	<i>homographes, pour lesquels la discrimination s'effectue à</i>
COTE	<i>reculons à partir de la fin, en accordant un poids précis</i>
côte	<i>aux accents. Chaque langue peut comporter ses propres</i>
côté	<i>règles.</i>
côté	
coter	

2- FUSION

Sous-paramètres :

LISTE1 : 1^{ère} liste à fusionner

LISTE2 : 2^e liste à fusionner

LISTE3 : Liste résultante (fusion des listes LISTE1 et LISTE2)

Note : Le résultat doit être conforme aux règles de classement définies dans la fonction précédente.

3- COMPARAISON

Sous-paramètres :

CHAÎNE1 : Chaîne de référence

CHAÎNE2 : Chaîne à comparer à CHAÎNE1

TYPE DE COMPARAISON :

Précision du niveau de parité acceptée :

- 1- Égalité floue (selon règles phonétiques)
- 2- Quasi-égalité (données mixtes accentuées ou non)
- 3- Parité seulement si chaînes exactement égales

Résultats possibles :

- Parité exacte (ex. côté - côté)
- Quasi-égalité (ex. COTE = côté)

Note : Si la quasi-égalité est acceptée, elle a priorité sur un résultat « plus petit que » ou « plus grand que ».

- Égalité floue ; dans certaines langues, un autre type de quasi-égalité, plus permissif, est nécessaire. Il peut être basé sur des règles phonétiques (ce type doit cependant être explicitement précisé comme paramètre)
(ex. en allemand: MÜLLER = Mueller = MULLER)
- Plus petit

(ex.:

côte < coté
 cote < côté
 COTE < côte
 coté < coter)

- Plus grand : la réciproque de ce qui précède

4- REPÉRAGE

Sous-paramètres :

Entrée :

PATRON : Chaîne recherchée

CHAÎNE : Chaîne où l'on effectue le repérage

TYPE DE REPÉRAGE : Idem TYPE DE COMPARAISON

Résultats :

INDICE : Retourne la position numérique, à l'intérieur de CHAÎNE, de PATRON. En cas d'insuccès, retourne -1

(ex.: si CHAÎNE= « côtelette » et PATRON = « COTE » :
 Si quasi-égalité acceptée :
 INDICE = 0 est retourné
 sinon, INDICE = -1 est retourné)

LNG : Retourne la longueur de la chaîne délimitée
 (peut être différente de la longueur implicite de PATRON)

(ex.: si PATRON=« Müller » et CHAÎNE=« Klaus Mueller »
 et qu'on accepte les égalités floues,
 Alors INDICE=6
 et
 LNG=7 seront retournés)

La longueur de la chaîne repérée est plus longue que la chaîne recherchée, dans cet exemple.

5- TRADUCTION DE CLEF

(pour utilisation dans différents types d'applications)

Sous-paramètres :

ENTRÉE : Chaîne à convertir (ex.: Côté)

Résultats : Ensemble de clés pour différents types d'utilisation
(ex. : comparaisons dans les tris) :

CLEF DE PREMIER ORDRE (CPO)

(dans l'exemple précédent : COTE)

CLEF DE SECOND ORDRE (CSO)

(dans le même exemple <17><16><19><16>)

Note : En français, il s'agit de codes binaires indiquant les accents utilisés :

<16> absence d'accent	<i>Note : Cette séquence varier en fonction de la</i>
<17> accent aigu	<i>langue utilisée (ici : ordre exact du français).</i>
<18> accent grave	
<19> accent circonflexe	
<20> tréma	
<21> cédille	

CLEF DE TROISIÈME ORDRE (CTO)

(dans le même exemple <09><08><08><08>)

Note : En français, il s'agit de codes binaires indiquant la hauteur de casse :

<08> minuscule
<09> majuscule

CLEF DE QUATRIÈME ORDRE (CQO)

(par exemple, si on a entré le mot anglais « vice-président »,

CPO serait alors « VICEPRESIDENT »

CSO serait égal à 13 * <16>

CTO serait égal à 13 * <08>

CQO correspondrait quant à lui aux caractères déplacés les moins significatifs et assurerait une prévisibilité absolue des tris et des comparaisons [cette clé débute systématiquement par un 0 binaire] :

<00><position><caractère>.... soit, pour cet exemple <00><04>-)

Pour une compréhension plus claire des codes recommandés et des considérations d'implantation, voir le document intitulé Règles du classement alphabétique en langue française et procédure informatisée pour le tri (dernière révision majeure le 15 août 1988), du même auteur.

6- TRADUCTION DE CLEF INVERSÉE

(étant données les quatre clés définies plus haut, cette fonction reconstruit fidèlement l'original)

Entrée :

CPO (ex.: AUJOURDHUI)

CSO (ex. 10 * <16>)

CTO (ex. 10 * <08>)

CQO (ex. <07>')

Résultat :

SORTIE : Texte reconstitué (dans cet exemple « aujourd'hui »)

7- CONVERSION MAJUSCULES/MINUSCULES

Sous-paramètres :

OPTION :

De minuscules à majuscules

De majuscules à minuscules

CHAÎNE1 : Chaîne à convertir

CHAÎNE2 : Chaîne résultante

(ex. de conversion en majuscules : été → ÉTÉ)

8- TRADUCTION DE PAGE DE CODES

Sous-paramètres :

PGENTRÉE : Page de codes de la chaîne d'entrée

PGSORTIE : Page de codes de la chaîne résultante

CHAÎNE1 : Chaîne d'entrée à convertir

CHAÎNE2 : Chaîne résultante (traduite)

Note : Ceci est la seule fonction où l'on doit préciser les pages de codes utilisées.

Conçu par Alain LaBonté, en collaboration avec Jean-Pierre Cabanié, Philips France, à Davos, Grisons, Suisse, le 1988-04-19.

SEAS - SHARE European Association

© **Gouvernement du Québec**
Dépôt légal - 1er trimestre 1991
Bibliothèque nationale du Québec
ISBN 2-550-21778-0