

# Standard sur le tri alphabétique et la recherche de chaînes de caractères (SGQRI 004)



**Alain La Bonté**

Journée-rencontre de WebÉducation, 18 janvier 2007

# Plan de la présentation

1. **Problématique**
2. **Solution générale**
3. **Définitions**
4. **Exigences**
5. **Dispositions transitoires et finales**
6. **Questions et réponses**

# 1. Problématique

# Problématique

- **Listes triées par ordre alphabétique : incapacité de toujours trouver ce qu'on cherche pour une donnée comportant un accent ou pour un nom de famille comportant une espace, un trait d'union ou une apostrophe**
  - ➔ Classement souvent incorrect sur les plateformes gouvernementales. Exemples :
    - ◆ z minuscule classé avant A majuscule
    - ◆ é classé après z
    - ◆ « été » classé après les « z »
    - ◆ « L'Heureux » classé avant « Labadie »
  - ➔ Classement incohérent entre plateformes
    - ◆ Problème d'interopérabilité lors de la fusion de fichiers
  - ➔ Complication pour les citoyens qui doivent s'y retrouver dans des listes triées, particulièrement sur Internet

# Problématique (suite)

- **Retard des MO à implanter la norme canadienne (1992)**
  - ➔ Timidité pour en exiger l'implantation auprès des fournisseurs
  - ➔ Fournisseurs locaux moins au courant de ce qui est disponible que leurs centres de recherche américains...
  - ➔ Résultat :  
« pollution informationnelle à la vitesse de l'éclair »



# 1. Solution générale

# Standard proposé

**Cohérence du classement sur toutes les plateformes**

**Conformité avec les normes canadienne et internationales**

# 3. Définitions



# Définitions

→ **caractère**

un élément d'un ensemble utilisé pour organiser, commander ou représenter des données

→ **delta**

un profil, soit l'expression d'une différence par rapport à une référence généralement admise

→ **donnée textuelle**

toute donnée informatique exprimant du texte sous forme de caractères codés

→ **logiciel embarqué**

un logiciel qui est intégré à un dispositif, une machine ou un autre système et qui pilote ce dispositif, cette machine ou ce système

# Définitions

## → recherche de chaînes de caractères

une opération par laquelle une chaîne de caractères recherchée est comparée aux chaînes d'un corpus, en fonction de critères d'équivalence entre les chaînes

## → tri

l'opération par laquelle une série de données textuelles est mise en ordre selon des critères ne dépendant que de la nature des données elles-mêmes

## 4. Exigences

# Article 5

Une fonction de tri alphabétique doit être conforme :

- a) à la norme ISO/CEI 14651:2001 et à son amendement 1, auxquels s'ajoute le delta décrit dans le document intitulé *Delta québécois à la norme internationale ISO/CEI 14651*, du ministère des Services gouvernementaux en date du 3 août 2006, traité jusqu'au quatrième niveau inclusivement ; ou
- b) à la norme CAN/CSA Z243.4.1 de l'Association canadienne de normalisation (CSA).

# Article 6

Une fonction de **recherche** de chaînes de caractères **ne** doit tenir compte **ni** **des accents**, **ni de la forme** majuscule ou minuscule des caractères, **ni des caractères spéciaux**, tant pour le mot clé de recherche que pour la chaîne recherchée.



## 5. Dispositions transitoires et finales

# Article 15

**Ce standard entre en vigueur le  
11 décembre 2006.**

# Article 7

**Toute application informatique utilisée avant la date d'entrée en vigueur du présent standard [11 décembre 2006] doit être conforme aux exigences de la section II au plus tard cinq ans après cette date.**

## 6. Questions et réponses

# A-t-on un exemple qui permettrait de comprendre le problème du tri informatique ? N'est-ce pas trivial ?

## Le bottin SuperPages truffé d'erreurs

**Claudette Samson** **LE SOLEIL**

18 décembre 2003 - 11h42

**Il n'y a pas à dire, un annuaire téléphonique, c'est bien pratique. Mais lorsque ledit annuaire ne respecte pas l'ordre alphabétique, ça devient embêtant...**

Ces dernières semaines, TELUS Québec (T) a distribué, dans la région de Québec, son premier bottin SuperPages, aussi gros que la brique Pages jaunes, de Bell.



Or, lorsqu'on le feuillète un peu, on réalise que Dubé vient après Dubuisson, Dupéré après Dupuis, Fréchette après Frongillo, Bédard après Bzdera, Déa après Dzomo-Silinou, Lépine après Lytwyn, etc.

Bref, le bottin recèle un lot de coquilles qui en limitent considérablement l'intérêt. C'est pourtant avec beaucoup de fierté que la compagnie présente son nouveau produit comme «un outil de recherche utile et fiable»... et en fait la promotion à la télévision.

Produit à 319 000 exemplaires, il a été distribué à tous les foyers et entreprises de la région.

Services  
gouvernementaux

Québec





# Pourquoi ne pas tenir compte des accents dans les recherches ? N'est-ce pas contraire à la langue ?

Le cas de la recherche est très différent du respect de l'orthographe. Exemple :


- Si l'on cherche le nom « Lévesque », on s'attend à trouver le nom « LEVESQUE » inscrit, peu importe la raison, sans accent
- Et vice-versa...

Une recherche stricte ne permettrait pas de trouver avec les accents et de mauvais moteurs de recherche sont, hélas, ainsi conçus qu'ils collent aux accents alors qu'ils ne collent pas aux majuscules ou aux minuscules

Ce n'est qu'un exemple très simple. Il y en a beaucoup d'autres mettant en jeu les formes infléchies des verbes français (**relever**, **relève**, **relevé**, etc.)

**Autres questions ?**

Services  
gouvernementaux

Québec 

**Merci de votre attention !**

Services  
gouvernementaux

Québec 