

Des accents dans les noms de domaines Internet

(Auteur : Alain LaBonté, Direction de l'architecture et de l'infrastructure, secrétariat du Conseil du trésor)

Des accents dans les noms de domaines Internet	1
Introduction – les problèmes mondiaux actuels et l'expertise québécoise	1
Le système des noms de domaines	1
Les brevets, un frein au développement ?	2
L'ouverture des noms de domaines aux langues du monde : aboutissement	3
Le « cyber-squat » : enregistrements multiples de noms similaires	4
Comment passera-t-on à la nouvelle pratique d'accentuation des noms ?	4
Publicité et éducation	5
Saisie des caractères composant ces noms	6
Conclusion	6

Introduction – les problèmes mondiaux actuels et l'expertise québécoise

Le vendredi 6 décembre 2002, dans le cadre des travaux visant à procurer au gouvernement du Québec un standard sur les noms de domaines¹, quelques privilégiés ont eu la chance de venir écouter Marc Blanchet, de chez Hexago (société née d'un projet de commercialisation de la R&D de la société Viagénie), nous entretenir de l'internationalisation de ces noms. Pour nous, cela se traduit surtout par la simple possibilité d'accentuer les noms de domaines Internet.

Marc Blanchet est co-président, avec le Singapourien James Seng, du groupe de travail de l'IETF (« Internet Engineering Task Force ») qui s'est donné pour mandat de trouver une solution permettant d'exprimer les noms de domaines dans toutes les langues du monde.

Depuis les débuts d'Internet, en effet, tous les noms de domaines sont limités à un jeu de caractères restreint, soit celui des 26 lettres non accentuées de l'alphabet latin (52 caractères si l'on inclut les minuscules), des dix chiffres arabes, du trait d'union et du trait de soulignement. Techniquement, le codage sous-jacent à cet ensemble est l'ASCII (« American Standard Code for Information Interchange »), un code limité mathématiquement et structurellement à 96 caractères imprimables (notamment les 64 précités, plus quelques signes de ponctuation et caractères spéciaux) qui ne permet pas sans restriction le codage des 32 lettres accentuées ou ligatures propres au français (si l'on inclut les majuscules), par exemple, et à plus forte raison les dizaines de milliers de caractères chinois ou les milliers d'autres caractères nécessaires à l'écriture des diverses langues du monde.

Le système des noms de domaines

Aujourd'hui, presque tous ceux qui font usage d'Internet savent ce qu'est un nom de domaine. Et si ce n'est pas le cas, comme M. Jourdain qui faisait de la prose sans le savoir, tous en ont utilisé un. Un nom de domaine est en effet nécessaire tant pour envoyer un courriel que pour avoir accès à une page web. Dans les adresses de courriel, le nom de domaine est tout ce qui se trouve à la droite du signe

« @ », cette arrobe au nom officiel si méconnu. Dans une référence de page web, le nom de domaine est tout ce qui se retrouve généralement à la droite des sept caractères « http:// » et à la gauche du prochain « / », le cas échéant. Un tel nom de domaine est toujours composé d'un certain nom de termes (minimum de deux termes) séparés par des points.

Comme les machines sont très mal adaptées à ces noms destinés à un usage plus ou moins mnémorique, on a recours à des serveurs, appelés « serveurs DNS » (DNS : « Domain Name System »), pour les convertir en adresses utilisables directement par le protocole IP, ce qui permet de joindre un ordinateur (ou une machine quelconque) sur le réseau des réseaux. Ces adresses sont, dans le système actuel, composées d'une série de quatre nombres de valeur comprise entre 1 et 254 (pour une possibilité de plus de quatre milliards d'adresses de « machines » joignables simultanément).

Comment cela fonctionne-t-il ? Au moyen d'un certain nombre de serveurs DNS qui se « parlent » entre eux à partir de votre ordinateur qui leur demande leur aide à votre insu, par un système d'indexage hiérarchique, à partir du dernier terme du nom de domaine et à rebours pour les autres termes, une adresse est finalement obtenue. Le dernier terme est soit un nom réservé par les concepteurs antédiluviens d'Internet, un nom comme « org », « com » ou « net », notamment, soit le code du pays dans lequel se trouve l'ordinateur que l'on veut atteindre, en vertu de la norme ISO 3166 (dans ce système, le Japon a comme code JP, la France le code FR, l'Allemagne le code DE, l'Afrique du Sud le code ZA, le Canada le code CA, et ainsi de suite).

Dans chacun des pays, il y a un registraire responsable de l'attribution des noms de domaines pour son « domaine » particulier, celui du pays. En règle générale, pour la création de tout nom de domaine, une ou plusieurs organisations ont l'insigne honneur d'encaisser des sous au passage. C'est la société à but non lucratif ICANN (« Internet Corporation for Assigned Names and Numbers » qui s'est vue attribuer par l'ensemble des utilisateurs la responsabilité d'attribuer les adresses Internet et les noms de domaines. Mais en marge de cet organisme de haut niveau à but non lucratif, les sociétés responsables de l'enregistrement dans les divers pays récoltent leur taxe privée au passage pour ce faire. On comprend assez rapidement que cette activité d'enregistrement peut s'avérer très lucrative et elle l'est. Tout changement dans ce système des noms de domaines peut donc avoir un impact financier pour ces sociétés, il va de soi. C'est d'ailleurs l'une des raisons qui ont fait que le sujet controversé de l'internationalisation des noms de domaines a pu avancer ces dernières années : certains y voient une occasion de plus de faire fonctionner la planche à billets pour la création d'une manne de nouveaux enregistrements.

Les brevets, un frein au développement ?

Certaines techniques proposées pour l'internationalisation des noms de domaines étaient protégées par des brevets, mais le standard tel que proposé à l'heure actuelle en est tout à fait exempt. Toute modification des autres techniques se heurterait inéluctablement à ce problème, que l'on a en fin de compte évacué.

Par ailleurs, les très nombreux pourparlers qui ont mené à la solution technique retenue pour l'internationalisation (pour nous l'accentuation) des noms de domaines ont conclu qu'il était plus réaliste d'effectuer un tel changement sans affecter les serveurs de noms de domaine. De toute manière, la mise à niveau de tous les serveurs DNS n'est pas quelque chose qui aurait pu se faire universellement

en un temps record. Une solution en amont, compatible avec l'existant, apparaissait la solution la plus raisonnable.

L'ouverture des noms de domaines aux langues du monde : aboutissement

On a vu que les serveurs DNS ne soutiennent depuis les débuts d'Internet que le code ASCII, et l'on a vu aussi que ce code comporte des restrictions de transmission.

Vouloir transposer l'ensemble des systèmes d'écriture du monde sans modifier les serveurs DNS, c'est-à-dire, en limitant ce qui se passe sous le capot à l'ASCII, nécessite l'intervention de spécialistes du monde du codage, pour optimiser le résultat, et minimiser aussi les restrictions.

Depuis quelques années, il existe une norme internationale, l'ISO/CEI 10646 (correspondant pour l'essentiel au standard Unicode), qui permet de coder chacun des caractères de toutes les langues du monde, passées et présentes, sur plusieurs octets. Soulignons au passage que le gouvernement du Québec a été un acteur important de l'élaboration de cette norme, et qu'il a notamment contribué à sa publication en langue française. Un des impacts les plus importants pour nous, outre l'excellente prise en compte des caractères du français intégral, est que chacun des milliers de caractères à avoir un nom (les dizaines de milliers de caractères chinois n'en ont pas dans la norme) peuvent maintenant être désignés dans notre langue, ce qui est un privilège réservé pour l'instant à l'anglais et au français, et se reflète déjà dans un certain nombre de produits informatiques qui présentent ces noms à l'utilisateur.

La transposition des caractères du « jeu universel de caractères codés sur plusieurs octets » (JUC) en ASCII, heureusement, est prévue dans la dernière édition de la norme internationale.

Les spécialistes d'Internet ont donc eu recours à cette norme. Avant d'effectuer la transposition, il faut passer par une phase de normalisation, chacun de nos caractères accentués pouvant être codés selon deux méthodes de par la norme internationale, soit intégralement, soit avec sa lettre de base accompagnée d'un accent combinatoire séparé.

Dans le processus de transposition, une fois que les caractères de chaque terme d'un nom de domaine ont été pré-codés de manière unique, « canoniquement » et au plus simple, ils sont transformés mathématiquement en caractères « cryptés », mais imprimables, du code ASCII.

Un préfixe est ajouté à chacun des termes transposés pour indiquer que l'on a affaire à un terme de nom de domaine internationalisé (ce qui rend possible la transposition inverse [la reconstitution du nom de domaine accentué correspondant, notamment] tout en minimisant les possibilités que la transposition n'entre en collision avec un nom de domaine qui aurait déjà été enregistré).

Là où le problème se gâte un peu, c'est que le système DNS a limité à 63 le nombre de caractères ASCII pour chacun des termes composant un nom de domaine et que le nom de domaine lui-même est limité à un maximum de 255 caractères ASCII. En pratique, pour des raisons mathématiques qu'il serait long d'expliquer, chaque terme d'un nom de domaine « internationalisé » peut en général comporter au minimum 20 caractères, quelle que soit la langue utilisée, bien que cela doive être validé au moment de l'enregistrement pour plus de sûreté (en fait, selon la langue et les caractères utilisés, ce minimum pourra parfois même s'étendre jusqu'à 40 caractères).

Pour éviter la possibilité pour quelqu'un d'enregistrer un nom de domaine avant tout le monde (à des fins plus ou moins orthodoxes), le préfixe utilisé pour chaque terme fait l'objet d'un embargo de la part des spécialistes jusqu'à la publication du standard Internet, publication qui devrait se faire au début de l'an 2003. Quoiqu'il en soit, et malgré cet embargo relatif, depuis deux ans, l'ICANN a conseillé de trouver un autre nom à ceux qui voulaient enregistrer des noms utilisant des termes commençant par ce préfixe, de manière, donc, à diminuer les possibilités de collision.

Le « cyber-squat » : enregistrements multiples de noms similaires

Il va de soi que le problème de l'enregistrement multiple de noms de domaines est déjà connu. En effet, il est évident que, comme aucune loi internationale ne régit l'enregistrement des noms de domaines Internet, cette activité suit celle de la jungle, celle de la nature, et que les registraires ne peuvent que suggérer un autre nom dans les cas où l'on soupçonne un problème.

Comme une société peut s'enregistrer sous le nom de domaine supérieur « .COM », « .ORG », « .NET », notamment, ou sous celui de chacun des pays (comme « .CA », « .FR », etc.), elle peut se sentir obligée de payer pour s'enregistrer plusieurs fois de manière à protéger son nom. La société ICANN elle-même, peut être jointe en utilisant indifféremment le nom de domaine « ican.com », « icann.org » ou « icann.net », par exemple. Tout oubli de tels enregistrements peut à la limite mener à l'enregistrement par un tiers d'un nom réservé, avec les conséquences les plus inattendues, les pires étant bien sûr malveillantes. Ce phénomène est connu sous le nom de « cyber-squat ». Bien que cela puisse donner lieu à des poursuites pour la violation de marques de commerces ou de noms réservés dans différents pays, on comprendra bien que cela peut s'avérer difficile dans un contexte multinational.

La création récente de nouveaux noms de domaines supérieurs, comme « .BIZ », est venue amplifier encore ce phénomène de cyber-squat, à la grande joie, au passage, comme toujours, de ceux qui impriment des billets et qui s'en lavent les mains tout comme Ponce-Pilate.

Ici, l'accentuation des noms de domaines nécessitera des enregistrements multiples, à la fois sans accents pour être compatible avec la pratique passée, et avec accents pour respecter notre caractère culturel et linguistique et permettre une utilisation éventuellement intuitive des noms de domaines du gouvernement du Québec.

Heureusement pour le gouvernement du Québec, c'est la DGT (Direction générale des télécommunications) du SSIGRI (sous-secrétariat à l'inforoute gouvernementale et aux ressources informationnelles, au secrétariat du Conseil du trésor) qui est la seule à détenir les droits permettant l'enregistrement de noms de domaines se terminant par « GOUV.QC.CA » (ce qui n'évite toutefois pas le problème des « .QC.CA », des « .CA » tout court, ou des « .COM », « .NET » et autres « .BIZ »).

Comment passera-t-on à la nouvelle pratique d'accentuation des noms ?

Comme on l'a vu, aucun changement matériel ou logiciel n'est nécessaire dans les serveurs DNS puisque les véritables noms de domaines qui seront véhiculés sur le réseau seront des noms qui pourraient théoriquement déjà être enregistrés.

Un nom éventuel comme « Été.com » devra être transposé en un nom (ici fictif, à des seules fins d'illustration) comme « xx--zstzs.com », lui-même re-transposable automatiquement et de manière réciproque en « Été.com »). Cette transposition devra être essentiellement faite dans chacun des postes de travail, dans les applications « clientes », comme votre fureteur (Netscape, Internet Explorer, notamment) ou votre courrielleur (MS-Outlook Express, Lotus Notes ou Eudora, pour ne citer que les plus populaires).

Il va donc sans dire que dès que le préfixe unique de chacun des termes internationalisés sera connu, comme l'algorithme de transposition l'est déjà, la DGT pourra d'ores et déjà commencer à enregistrer de nouveaux noms de domaines accentués, noms de domaines qui pourraient par exemple comprendre le terme « Québec » avec l'accent. Le rôle des registraires consistera, à partir d'un nom accentué, à déterminer un nom transposé, et à l'enregistrer comme tel, après avoir vérifié un certain nombre de restrictions techniques sur la longueur des termes, de même que d'éventuelles collisions, peu probables, avec des noms déjà enregistrés. Ces vérifications se feront vraisemblablement à l'aide de logiciels que les registraires devront se procurer. Le responsable de chaque registre DNS, s'il est différent du registraire, n'a rien à faire, puisque tout ce qui est enregistré dans les fichiers est un nom « crypté », qui obéit à toutes les restrictions techniques actuelles.

Pour ce qui est de l'utilisation devant chaque poste de travail, on fera face à deux situations dès le départ : ou bien l'accentuation fonctionnera, si l'application-« cliente » a été mise à jour, ou elle ne fonctionnera pas, dans lequel cas une utilisation intuitive devrait inciter même un utilisateur naïf à omettre l'accentuation.

On ne peut donc compter que sur le temps, mais un temps que l'on peut d'ores et déjà considérer comme relativement bref compte tenu de la situation quasi-monopolistique qui existe pour les diverses applications clientes, avant que la pratique d'accentuation puisse se répandre. L'éducation se fera sans doute progressivement, et de manière plus ou moins intuitive. Dans un premier temps, le monde de la publicité nous martèlera sans doute l'alternative « avec ou sans accents », jusqu'à ce que l'on n'en parle plus.

Publicité et éducation

D'ici là, que devra-t-on mettre comme référence sur nos cartes d'affaires, dans la publicité faisant référence à des sites web ? La réponse est simple : s'il y a des accents, dans un premier temps, on pourra ajouter la mention (« avec ou sans accents ») à la suite des références (alors qu'à l'heure actuelle, on entend souvent à la radio ou à la télé des mentions telles que « Québec sans accent », ce qui ne sera qu'une variante moins agressive du lavage de cerveau qui a prévalu jusqu'ici pour les noms de domaines qui devraient intuitivement comporter un accent pour des lettrés normalement constitués.

Il est par contre nécessaire de mentionner que pour les adresses de courriel, dans un premier temps, la partie individuelle de l'adresse, soit les termes à gauche du signe « @ », conserveront une restriction, dans ce qu'ils ne pourront être accentués. Le gouvernement du Québec créera un standard¹ qui devrait permettre de faciliter la vie des utilisateurs jusqu'à ce que tout puisse être accentué. Notons au passage qu'après que la question de l'internationalisation des noms de domaines aura été « réglée », ce qui est en très bonne voie de se faire de façon définitive, l'IETF se penchera aussi sur la question de l'internationalisation des adresses de courriel. Au moment d'écrire ces lignes, on apprenait de source généralement bien

informée que le démarrage de ce projet n'était qu'une question de quelques semaines.

Saisie des caractères composant ces noms

Certains pourraient manifester une certaine inquiétude à propos de ce qui arrivera quand une personne d'origine étrangère, même résidente du Québec, inscrira un jour un prénom comme « João » dans son adresse de courriel sur une carte d'affaires : pourra-t-on alors taper tous ces caractères au clavier ?

La réponse à cette question permet de faire au passage un peu de publicité sur nos orientations, car en principe, tous les claviers acquis par le gouvernement du Québec sont conformes à la norme CAN/CSA Z243.200-92, qui permet de saisir aisément l'ensemble des caractères latins utilisés en Europe de l'Ouest et dans les deux Amériques.

Il va de soi que pour les autres systèmes d'écriture, en général, les personnes à même de lire ce qui leur sera présenté et en mesure d'utiliser un système informatique, seront déjà des lettrés de ces langues et ont bien des chances de déjà disposer d'un clavier leur permettant d'écrire en chinois, en arabe, en grec, et ainsi de suite.

Pour ce qui est des Américains unilingues qui ne disposent pas tous de claviers bien élaborés leur permettant de taper nos lettres accentuées, il y a fort à parier que l'on prendra aussi chez nous les mesures pour que les adresses puissent fonctionner quand même avec ou sans accent, comme on l'a dit plus tôt. Au-delà (autres systèmes d'écriture), il va de soi aussi que les références, dans un très grand nombre de cas (la majorité, pourrions-nous dire, si nous n'étions trop prudents), font généralement l'objet d'opérations de « couper-coller » de toute manière.

Conclusion

Nous pourrions nous poser, comme beaucoup d'Américains spécialistes d'Internet le faisaient au départ, le pourquoi de ces changements. Pour ceux dont le nom ou le prénom est accentué (c'est le cas d'environ 20 % des Québécois, grosso modo), la question du respect de leur nom ne leur a jamais fait se questionner longtemps : il y va du respect de l'identité de chacun, ce qui constitue un des droits de l'homme les plus fondamentaux.

Pour les autres, la question se pose pour les noms collectifs : le non-respect d'une culture ou d'une langue constitue à la limite l'un des freins les plus grands dans le développement du commerce mondial, au-delà d'un certain seuil. Ce seul énoncé a réussi à convaincre les meilleurs commerçants de la planète, qui ne veulent être limités par aucune contrainte dans leur expansion. Or le respect des cultures et des langues constitue l'ultime barrière non-tarifaire du monde, dont aucune réglementation ne viendra jamais à bout tout à fait. En un sens, pour les Québécois, cela constitue un bienfait.

par Alain LaBonté (alb@sct1.gouv.qc.ca)

Secrétariat du Conseil du trésor du Québec
Direction de l'architecture et de l'infrastructure
1500-H, rue Jean-Talon Nord
1^{er} étage, porte 1.12, boîte n° 1
Ste-Foy, QC G1N 4T5 Canada

Téléphone : +1 418 528 6144
Télécopieur : +1 418 646 3571

1 : Le standard SGQRI 021, sur les noms de domaines gouvernementaux, qui sera éventuellement approuvé par le Conseil du trésor comme norme obligatoire pour l'ensemble du gouvernement du Québec, devrait être approuvé au consensus par les divers ministères et organismes qui voudront participer à son élaboration. Il fait partie d'une série de standards sur l'adaptabilité culturelle et linguistique des technologies de l'information (ACL) qui feront écho à un certain nombre de normes et de standards internationaux que le gouvernement du Québec a contribué à créer depuis une quinzaine d'années. L'auteur du présent article, Alain LaBonté, lui-même impliqué dans la rédaction de plusieurs normes internationales, est responsable de mener à bien ce projet d'élaboration de la série des standards gouvernementaux en ACL.

Le standard SGQRI 044 se penchera, quant à lui, sur les adresses de courriel, un sujet connexe.