

Documentation

Les séquences de classement : du chinois à l'anglais, en passant par le français

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

Ce document a été présenté le 28 septembre 1988 à Aalborg au Danemark à la rencontre de SHARE European Association (SEAS). Il 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.

© Secrétariat du Conseil du trésor du Québec - 1998

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

Contacter : Alain LaBonté alb@sct.gouv.qc.ca

Note de navigation :

À chaque thème principal du texte correspond une figure en format gif, présentée dans l'ordre. Les figures sont aussi disponibles en format pdf grâce au lien placé à la suite du titre.

Les séquences de classement : du chinois à l'anglais, en passant par le français. Voilà un titre bien étrange pour désigner une opération qui est familière à la plupart des spécialistes de l'informatique. L'idée m'est venue de prendre comme point de départ une langue « exotique » pour en arriver à une langue presque universelle maintenant, l'anglais, afin de montrer que le classement, l'une des opérations fondamentales du traitement de l'information, n'est pas une opération nouvelle, bien qu'elle ait été négligée par les spécialistes du traitement de l'information.

En effet, à ma connaissance, il n'existe pas encore sur le marché un programme de tri qui puisse exécuter cette opération correctement dans une langue donnée, y compris en anglais. Je m'intéresse à cette question depuis un certain temps déjà, et je crois que le moment est venu de faire état de mes réflexions.

Je viens d'affirmer qu'il n'existe pas de programme de tri qui fournisse des résultats satisfaisants. Une déclaration aussi étonnante exige bien sûr certaines explications. Cette situation n'a pas eu de conséquences graves jusqu'à maintenant parce que même si le contenu d'une liste ne se présente pas de la manière souhaitée par l'être humain, celui-ci arrive à s'en accommoder si la liste contient toute l'information que renferme le fichier dont elle est issue et s'il peut

parvenir, même au prix d'une pénible recherche de temps à autres, à retrouver les enregistrements mal triés. Et tout spécialiste du traitement de l'information peut expliquer que tel enregistrement se trouve à cet endroit en raison de sa valeur binaire. C'est sans doute ce qu'on appelle la culture informatique. Mais la plupart des gens ne sont pas très cultivés dans ce domaine, et de même qu'un occidental peut tout ignorer des raffinements de la culture chinoise même s'il détient un doctorat en littérature anglaise, de même un ordinateur ASCII peut être dérouté en présence d'un système EBCDIC. Ainsi, chacun de ces systèmes peut donner des résultats bizarres si on lui demande de comparer des fichiers qui sont classés en fonction de l'autre système.

Voyons par exemple ce qui se produit lorsqu'on demande à un système ASCII et à un système EBCDIC de classer une simple liste des caractères de l'alphabet, en majuscules et en minuscules, et des dix chiffres.

figure 1 [gif](#) [pdf](#) 

On voit ici que le système ASCII place les chiffres en premier, suivis des lettres majuscules et des lettres minuscules, tandis que le système EBCDIC place d'abord les lettres minuscules en premier, puis les majuscules et les chiffres. Même si l'on effectue la traduction réciproque des jeux de caractères, les fichiers ne seront pas triés de façon adéquate, et si l'on désire extraire tous les enregistrements d'un fichier qui se trouvent dans le second système, on n'obtiendra que les dix chiffres, et rien d'autre, lorsque l'on commande un tri séquentiel. En outre, vous remarquerez que dans un même fichier, nous avons en réalité deux listes triées et non une seule, comme on pourrait s'y attendre, car toutes les minuscules sont séparées des majuscules. Et pourtant, il s'agit là du cas le plus simple. Je vous laisse deviner quels résultats on obtiendrait à partir d'une liste imaginaire où les majuscules et les minuscules seraient mélangées au gré de votre fantaisie.

Ce simple exemple m'amène à formuler une première constatation : les systèmes ASCII et EBCDIC n'offrent pas une séquence de tri naturelle convenable, même dans le cas des classements les plus élémentaires.

Supposons maintenant que les spécialistes de l'informatique fassent preuve de plus d'intelligence, ce qui est effectivement le cas.

figure 2 [gif](#) [pdf](#) 

Ils ont d'abord décidé de construire des bases de données ne contenant que des lettres majuscules, ce qui laisse croire que les tris sont appropriés la plupart du temps ; cela est vrai uniquement si l'on ne tient pas compte du fait que cette méthode ne permet toujours pas de classer correctement certains mots qui se trouvent pourtant dans le dictionnaire. Ainsi, un mot comme « VICE-PRESIDENT », qui contient un trait d'union, serait classé après « VICE VERSA », ce qui constitue une manière plutôt révolutionnaire de classer les mots, fondée sur les séquences de bits, qui demeurent invisibles pour le commun des mortels. Afin de corriger de telles anomalies, on pourrait alors concevoir une clé de tri secondaire temporaire qui ne tiendrait pas compte des caractères spéciaux. Et si l'on se sert aussi de caractères accentués, on pourrait traduire ceux-ci en leur équivalent non accentué. On pourrait également en profiter pour traduire les minuscules en majuscules. Pourtant, une telle solution ne permet pas non plus d'obtenir des tris appropriés, comme je vous le démontrerai tout à l'heure.

Pour l'instant, je vous propose de quitter le monde occidental pour faire un voyage aux antipodes, là où les gens sont tous d'excellents acrobates puisque, de notre point de vue, ils vivent constamment la tête en bas. Si nous faisons de même nous aussi de temps à autres, nous pourrions voir la réalité sous un angle différent et découvrir des choses extraordinaires dans ce qui nous apparaît familier. C'est

maintenant le moment pour moi de remplir la promesse qui vous a été faite dans la publicité qui a précédé cette communication.

Vous êtes-vous déjà demandé de quelle façon les Chinois arrivent à retrouver la définition de chacun de leurs caractères, qui se comptent par milliers, par opposition aux 26 lettres de notre alphabet ? Nous, nous apprenons cet alphabet par coeur, mais eux, que font-ils ? D'ailleurs, ont-ils des dictionnaires ?

Les Chinois, effectivement, possèdent des dictionnaires. Leur dictionnaire le plus important fut compilé au XVII^e siècle, durant le règne de Kangxi (1662-1723), de la dynastie des Ch'ing, et contient plus de 60 000 caractères. Si l'on considère que chaque caractère, qui est un mot en soi, entre aussi dans la composition de nombreux autres, on peut penser que cette ancienne civilisation, qui a créé plus de 80 000 caractères au cours des millénaires, jusqu'à la révolution chinoise et même jusqu'à nos jours, a sans doute quelque chose à nous révéler dans l'art vénérable du classement.

Est-il possible de classer tous ces caractères dans un ordre qui reste accessible aux humains ? Hé oui ! Et les Chinois ont même inventé différentes façons de les classer. Les dictionnaires chinois contemporains ont recours à au moins trois méthodes de classement. La plus élémentaire, et aussi la plus récente, repose sur les principes de la phonétique.

figure 3 [gif](#) [pdf](#) 

Ce procédé permet à un occidental de retrouver dans l'ordre alphabétique un mot dont il connaît la prononciation, en suivant des règles presque semblables à celles auxquelles il est habitué. La révolution chinoise a contribué à la diffusion d'une prononciation romanisée de chaque caractère du chinois mandarin, ce qui a d'ailleurs favorisé l'alphabétisation d'une grande partie du peuple chinois. À chacun des phonèmes de la langue chinoise officielle correspond une graphie qui fait appel à l'alphabet latin et à certains signes diacritiques servant à marquer l'intonation. Ce système, de compréhension facile, est familier aux occidentaux, à l'exception de certaines lettres qui ont une prononciation différente, comme la lettre « Q », qui se prononce « tchi », et la lettre « R » qui, à l'intérieur d'un mot, se prononce un peu comme le « J » français.

Cependant, cette représentation phonétique ne permet pas de retracer un caractère dont on ignore la prononciation. Elle ne permet pas non plus de retracer un caractère dont on connaîtrait la prononciation en cantonais ou en japonais, même si le symbole qui sert à le représenter est le même. Comme il n'y a pas de rapport logique entre les caractères chinois et leur prononciation, ceux-ci ont été repris par de nombreux peuples de l'Asie et ils sont encore en usage dans nombre de pays dont la langue parlée diffère considérablement. Les caractères chinois sont des idéogrammes, c'est-à-dire qu'ils sont la représentation de concepts ; en d'autres mots, ce sont des images. Leur classement doit donc reposer sur un système permettant de trier des images ou des formes, ce qui n'est évidemment pas le cas des langues occidentales.

La deuxième méthode permettant de repérer un mot dans un dictionnaire chinois, celle qui est utilisée le plus fréquemment, se fonde sur la décomposition formelle de chaque caractère. C'est la technique du tri par les radicaux.

figure 4 [gif](#) [pdf](#) 

Chaque caractère chinois est constitué d'un, de deux ou même de plusieurs caractères primitifs, appelés radicaux, tous tracés suivant un ordre strict, de haut en bas et de gauche à droite, trait par trait. Le caractère chinois peut être délimité par un carré imaginaire dans lequel les formes sont disposées de manière à respecter un équilibre visuel. Dans la partie gauche de ce carré peut se trouver,

par exemple, une forme qu'aucun trait ne relie au reste du caractère. C'est ce qu'on appelle le radical. Le nombre de traits qu'il contient est déterminant. Dans un dictionnaire chinois, on trouve un premier ensemble de listes contenant les radicaux primitifs, chacune de ces listes étant constituée de radicaux qui comptent le même nombre de traits.

Il faut donc en premier lieu identifier le radical primitif du caractère, compter le nombre de traits qu'il contient et consulter la liste correspondante.

figure 5 [gif](#) [pdf](#) 

Il faut ensuite consulter un autre ensemble de listes, qui obéissent au même principe pour tous les radicaux secondaires qui peuvent accompagner le radical primitif. On compte alors le nombre de traits que contient le radical secondaire, qu'on cherche ensuite dans la liste correspondante, qui renvoie généralement au mot ou à la série de mots recherchés. Dans le présent exemple, le radical primitif représente le soleil. Il est constitué de quatre traits ; pour bien comprendre, il faut savoir qu'il suffit de trois traits, et non de quatre, pour tracer une forme carrée en chinois.

La liste des radicaux primitifs à quatre traits nous permet de repérer rapidement le caractère et nous conduit dans la seconde liste, où nous pourrions repérer le radical secondaire, c'est-à-dire la lune.

figure 6 [gif](#) [pdf](#) 

La juxtaposition de ces deux radicaux produit un nouveau caractère, qui correspond à l'adjectif « clair » ou qui peut aussi correspondre, en abrégé, au verbe « illuminer » ou même au verbe « comprendre ». Pour bien montrer que ces caractères ne possèdent pas de valeur phonétique intrinsèque, soulignons que le mot « soleil » se prononce « zh' » en mandarin (« rì » en pinyin) et que le mot « lune » se prononce « yu-e » (« yuè » en pinyin), alors que la combinaison de ces deux radicaux se prononce « ming » (même chose en pinyin) dans la langue officielle de la Chine.

Cette petite incursion dans la linguistique chinoise me permet de formuler une deuxième constatation : un système fondé sur des radicaux ou sur l'analyse intrinsèque des caractères pourrait faciliter les opérations de classement. Je reviendrai plus tard sur ce sujet.

Une troisième méthode de repérage des caractères chinois, qui est aussi une méthode visuelle, s'appelle la méthode des quatre coins.

figure 7 [gif](#) [pdf](#) 

Cette méthode, inventée au début du siècle, consiste à mémoriser un ensemble de dix formes qu'il faut identifier dans les coins de chaque caractère, et à attribuer à chaque coin un chiffre qui représente chacune de ces formes, ce qui donne un nombre de quatre chiffres correspondant, en principe, au caractère recherché. En fait, les dictionnaires contiennent trois pages de règles permettant de résoudre les cas ambigus, comme l'absence d'un coin, illustrée dans l'exemple que voici.

figure 8 [gif](#) [pdf](#) 

figure 9 [gif](#) [pdf](#) 

Ces méthodes de recherche présentent un intérêt indéniable si l'on considère qu'elles permettent de repérer un grand nombre de caractères d'une manière assez simple. Il existe plusieurs autres méthodes, et je sais que des chercheurs tentent de trouver des méthodes encore plus rapides pour repérer des caractères

dans les dictionnaires. Même si ces méthodes n'impliquent pas nécessairement l'aide d'une machine, elles peuvent néanmoins nous apprendre quelque chose.

D'abord, vous remarquerez que la plupart des difficultés inhérentes au tri sont causées par les homographes. Dans l'exemple que vous voyez présentement, où l'on doit en fin de compte chercher le 6702, il faut faire une recherche séquentielle parmi 30 homographes. Nous verrons dans quelques instants que cette situation se produit également dans les tris informatisés qui sont effectués dans les pays occidentaux. Cela ne présente pas de difficulté lorsque les listes sont destinées à des êtres humains, car leur intuition leur permet de retrouver l'information dont ils ont besoin. Dans le cas d'une machine, cependant, l'ordre de tri doit être défini d'une façon rigoureuse. Nous arrivons ainsi à une troisième constatation : les tris réalisés par des machines doivent être prévisibles de façon absolue pour être vraiment efficaces.

Revenons dans notre monde occidental. Je vais maintenant tenter de vous montrer pourquoi les méthodes de classement actuelles, avec lesquelles on croit régler les problèmes des doubles listes et des caractères accentués en convertissant en minuscules et en éliminant les accents, ne sont pas parfaitement efficaces.

Voici les résultats d'une expérience que j'ai faite avec le programme SORT de PC-DOS, le premier programme de tri qui m'est tombé sous la main.

figure 10 [gif](#) [pdf](#) 

Une première liste de mots disposés dans l'ordre que vous voyez a été soumise à l'ordinateur, qui l'a « triée » d'une première manière. La même liste, disposée dans un ordre différent, a été « triée » de cette deuxième manière. Lequel des deux tris est correct? Aucun, bien sûr, si on les compare aux véritables résultats qu'on aurait dû obtenir. Ce programme de tri n'a donc rien trié du tout puisque dans les deux cas, l'ordre des mots à la sortie est le même qu'à l'entrée. On peut répéter cette expérience à volonté avec la plupart des programmes de tri supposément sophistiqués qu'on retrouve sur le marché. Mais à quoi de tels résultats sont-ils dus ? Dans ces programmes de tri, la majuscule et la minuscule d'une même lettre sont considérées comme équivalentes à cette même lettre non accentuée. Ce procédé est acceptable pour l'intuition humaine mais, comme je l'ai souligné précédemment, l'ordinateur est incapable de faire une distinction de ce genre. Alors si vous devez comparer des listes prétendument triées comme celles-ci, vous pouvez être assurés que les résultats ne seront pas prévisibles. Je crois que cette démonstration est claire pour tous.

Pourquoi est-il important de corriger cette situation ? Parce qu'avec la généralisation de l'OSI (interconnexion de systèmes ouverts), les systèmes vont s'échanger des listes triées et des fichiers dont l'ordre de tri sera présumé correct. Parce que l'usage des caractères accentués sera de plus en plus répandu. Parce que dans l'univers IBM, les architectures d'applications de systèmes devront offrir à l'utilisateur une interface cohérente qui devra toujours donner les mêmes résultats, peu importe l'environnement. Parce que les séquences de tri actuelles ne respectent jamais le classement lexical. Et enfin, pour un grand nombre de raisons liées aux applications, la première étant qu'un programme de tri doit trier dans un ordre absolument prévisible et qu'environ 70 % de toutes les applications sur ordinateurs de grande puissance comportent des opérations de tri ; je ne peux démontrer l'exactitude de ce pourcentage, mais il a été avancé lors de la dernière rencontre de SHARE et il semble assez juste, d'après ma propre expérience.

figure 11 [gif](#) [pdf](#) 

Comment pourrait-on réussir à obtenir des ordres de tri absolument prévisibles ? De toute évidence, il faut respecter toute l'intégrité de l'information et définir un

ordre de distinction précis qui permette de faire la différence entre les quasi-homographes. À ce stade de ma réflexion, mon sens de l'humour m'a laissé pressentir que le syndrome chinois pourrait nous être utile. Les caractères chinois possèdent des caractéristiques que nous pouvons distinguer selon leur ordre de signification. Il en est de même pour l'information écrite dans les pays occidentaux qui utilisent l'alphabet latin. Quelles sont les données les plus significatives dans l'information textuelle occidentale ? Ce sont les lettres de l'alphabet latin, dépourvues d'accents et de caractères majuscules. Voilà le premier radical de notre information. Qu'est-ce qui vient en second lieu ? Ce sont les signes diacritiques et les accents. Les caractères minuscules et majuscules ont un degré de signification encore moindre. Finalement, les caractères spéciaux sont les moins significatifs dans l'information de type textuel. Nous pouvons donc poser comme quatrième constatation que pour trier correctement du texte, il faut au moins quatre clés, qui vont de l'information la plus significative à la moins significative.

figure 12

En réalité, c'est la complexité de l'accentuation en langue française qui m'a suggéré la hiérarchie que je viens de présenter. Dans les dictionnaires de langue française, on ne tient pas compte des accents pour le classement, sauf dans le cas des homographes, c'est-à-dire des mots qui s'écrivent de la même manière lorsqu'on supprime les accents. Dans le cas des homographes, le classement se fait selon un ordre précis, mais inverse de l'ordre alphabétique habituel : les accents obéissent à un ordre de classement, mais la distinction se fait à partir de la fin des mots, en remontant à la première différence observable, autrement dit à la dernière différence dans les quasi-homographes. Ce procédé fonctionne bien, mais si on ne fait pas la distinction entre les accents, les minuscules et les majuscules, il présente certaines difficultés lorsqu'il faut classer des mots en majuscules par rapport à leur équivalent en minuscules, peu importe la permutation des majuscules et des minuscules. Nous avons donc dû tenir compte de ce problème et nous avons établi une distinction entre ces deux ordres de signification. Ce procédé comporte toutefois des conséquences secondaires intéressantes puisqu'il permet de délimiter avec précision les différents niveaux de signification dans le texte et de distinguer les caractères accentués et les caractères en majuscules, qui constituent deux types différents d'information.

Ce principe de décomposition du texte en niveaux hiérarchiques relève des mathématiques. En effet, quelles sont les caractéristiques d'un nombre réel ? Son ordre de grandeur, puis sa précision numérique, constituée d'une suite de chiffres qui vont du degré de signification ou de précision le plus élevé au degré le plus bas. Lorsqu'un nombre est tronqué, seule sa précision en souffre. Le même principe peut être appliqué à des données textuelles organisées selon un degré de précision allant du plus élevé au moins élevé.

Au Canada, ces considérations ont conduit à l'élaboration d'un projet de norme, qui est encore à l'étude, sur les tris en français et en anglais, deux langues qui peuvent être traitées de façon similaire. Ces considérations sont aussi à l'origine d'une requête présentée à IBM en avril dernier à Davos, demandant la mise en place de fonctions de systèmes normalisées qui permettraient de créer des clés de tri en fonction des caractéristiques propres à chaque langue. Cette structure composite servirait également à construire des clés de traitement de données qui pourraient être utilisées avec les méthodes classiques d'accès indexé, comme VSAM dans les environnements IBM. Ces dernières clés s'enchaîneraient automatiquement en fonction de la séquence de tri nationale prévue dans chaque ordinateur. La structure à quatre clés permettrait aussi de fusionner des données accentuées à des données non accentuées sans qu'il soit nécessaire d'en faire l'analyse, car le premier ordre de signification d'une clé serait toujours dépourvu d'accents. Cette structure, conçue de manière appropriée, permettrait en outre de régler le problème de la conversion des pages de codes. Le seul défaut que

présente la requête de Davos, c'est qu'elle ne comporte que trois ordres de clés ; elle a dû être révisée par la suite car, comme je l'ai mentionné il y a quelques instants, les majuscules et les accents doivent constituer deux niveaux différents. Une nouvelle requête mise à jour sera présentée à IBM¹.

figure 13 [gif](#) [pdf](#)

Laissez-moi vous expliquer les détails de la structure de cette clé à l'aide d'un exemple. Prenons le mot français « Vice-président », suivi d'un point. Dans ce mot, on trouve des lettres, un accent, des minuscules et une majuscule, et deux caractères spéciaux. Quelle clé devons-nous construire à partir de ce cas ? Le premier ordre de signification ne devrait contenir que des lettres de l'alphabet latin, sans marques d'accents ni de majuscules. Le deuxième ordre de signification est celui des données accentuées ; ainsi, des codes sont affectés à chaque accent (avec un code spécial en cas d'absence d'accent), mais en français, l'ordre doit être inversé. Le troisième ordre de signification se rapporte aux caractères majuscules et minuscules, c'est-à-dire qu'un code est attribué à chaque majuscule et un autre à chaque minuscule. Le quatrième ordre de signification est structuré de manière à pouvoir reconstituer le texte original et à établir des distinctions très précises. Il contient un 0 binaire au début, afin de pouvoir tenir compte des zones variables de tri ; comme les trois premiers ordres de signification ne peuvent, par définition, contenir de zéros binaires, on est ainsi assuré que le quatrième ordre demeure le moins significatif. À la suite de ce zéro, on retrouve des paires de codes dont chacune indique le décalage, par rapport au texte original, du caractère spécial représenté par le code qui suit, et il en est ainsi pour chacun des caractères spéciaux de la zone d'entrée.

Vous remarquerez que le jeu de codes du deuxième ordre de signification est constitué de nombres qui sont plus petits que ceux du premier ordre, que le jeu de codes du troisième ordre est constitué de nombres qui sont plus petits que ceux du deuxième, et que le quatrième ordre commence par zéro, un nombre plus petit que les précédents lui aussi. Ce procédé assure une hiérarchie cohérente entre les ordres de signification. Comme les quatre clés contiennent toute l'information nécessaire, elles permettent de reconstituer le texte original.

figure 14 [gif](#) [pdf](#)

Comment ce procédé permettrait-il de trier la liste de mots français que je vous ai déjà présentée ? Vous voyez présentement les quatre clés associées aux mots de l'exemple précédent. Lorsque ceux-ci sont triés en ordre binaire, le résultat donne forcément une liste triée.

Ce procédé peut-il s'appliquer indifféremment à toutes les langues ? Malheureusement, il n'est pas applicable tel quel ; il exigerait un changement de tables, et peut-être aussi un changement de l'ordre d'analyse de la clé du deuxième ordre et une certaine analyse contextuelle.

figure 15 [gif](#) [pdf](#)

Voici un aperçu des problèmes qui peuvent se présenter dans certaines langues. En allemand par exemple, on a souvent dit, et même à SEAS, que les voyelles infléchies, comme le « ü », doivent être triées comme si elles étaient écrites sans accent et suivies d'un « e », c'est-à-dire « ue » dans notre exemple. Mais qu'arrive-t-il ainsi ? Le nom « Müller » serait alors classé avec « Mueller » mais avant « Mulhouse », qui serait lui-même classé avant « Muller », le biologiste américain. Or, un tel classement pourrait ne pas être souhaitable, et on pourrait préférer trier plutôt « Mueller » comme s'il s'écrivait avec un « ü », ce qui serait plus naturel. Mais un tri de ce genre exigerait une analyse contextuelle. La même structure à quatre clés pourrait être utilisée, et elle serait d'ailleurs nécessaire en

allemand aussi.

figure 16 [gif](#) [pdf](#)

En espagnol, il existe certaines lettres doubles qui sont considérées comme des lettres distinctes, comme « ch » et « ll », qui doivent être classées après « cz » et « lz » respectivement. À moins que ces lettres doubles soient codées à l'aide d'un octet, ce qui n'est généralement pas le cas, une analyse contextuelle serait nécessaire ici aussi et la structure à quatre clés s'imposerait également. Par conséquent, la cinquième constatation qui se dégage de cet exposé, c'est que dans certaines langues, même le tri au niveau élémentaire exige une analyse contextuelle.

Dans la définition des tables de caractères de certaines langues, certaines lettres devront comporter deux codes dans les trois premiers ordres de clés si l'on veut pouvoir reconstituer correctement le texte original. Ainsi, le « s » dur allemand (b) devrait posséder deux codes dans les trois premiers ordres de clés, en l'occurrence « ss » dans le premier ordre et des codes correspondants dans le deuxième et le troisième ordre pour indiquer que ceux-ci se rapportent à une lettre double.

figure 17 [gif](#) [pdf](#)

Une même lettre, dans différentes langues, peut être interprétée de diverses façons; cela nécessiterait donc des codes différents dans les clés des deux premiers ordres. La lettre « ñ », par exemple, peut être considérée, dans la plupart des langues, comme un « n » accentué et elle serait codée comme un « n » avec un accent. En espagnol cependant, cette lettre est distincte du « n » ordinaire et elle est classée après « nz », entre le « n » et le « o ». Personnellement, j'estime qu'en espagnol, cette lettre devrait être codée comme une lettre double dans la structure que je vous ai présentée, c'est-à-dire comme la lettre « n » suivie d'un code supérieur au « z » dans la clé de premier ordre de tri.

Ces quelques exemples illustrent la sixième constatation de mon exposé : le niveau de classement alphabétique le plus bas dépend de la langue, et il sera impossible de mettre au point un mécanisme de tri uniforme et universel tant que nous ne parlerons pas tous la même langue. C'est pourquoi nous devons déterminer non seulement des séquences de tri de caractères, mais aussi des schèmes de classement dans les systèmes. Il faudrait qu'un registre officiel de ces schèmes de classement soit créé, de préférence par l'ISO, et je souhaite que cette proposition connaisse une large diffusion. Comme il existe déjà un registre des jeux de caractères codés, il faudrait également qu'il y ait un registre des séquences de tri codées.

Pour bien des gens, tous ces problèmes ne sont qu'une source d'ennuis et la tour de Babel est un phénomène anormal, mais en considérant la réalité de cette façon, ils ne se rendent pas compte de toute la complexité de la nature qui a créé dans l'univers des myriades de formes que les êtres intelligents, y compris eux-mêmes, prennent plaisir à collectionner et à classer. Cette complexité est une loi de la nature, et c'est bien ainsi. Si Dieu avait voulu que la nature soit uniforme, Il aurait conçu l'univers sous la forme d'un gaz sans forme, ce qui, de toute évidence, n'est pas le cas, fort heureusement...

D'autre part, le problème est rendu encore plus complexe par le fait qu'au niveau le plus élevé, les tris dépendent des applications et du contexte même s'ils devraient toujours faire appel au schème de bas niveau que je viens de présenter. Pour construire des annuaires téléphoniques, il faut allonger certains mots, en abrégier d'autres, etc. Il existe certaines normes nationales à cet effet, et celles-ci devraient être homologuées à l'ISO et désignées au moyen de conventions normalisées.

figure 18

Conseil du trésor/Technologies de l'information/Utilisation du français

En conclusion, voici une récapitulation des six constatations présentées au cours de cet exposé :

1. Les séquences de tri de l'ASCII et de l'EBCDIC ne peuvent effectuer les classements correctement, quelle que soit la langue, y compris l'anglais.
2. Des séquences de classement fondées sur l'analyse de l'information contenue dans chaque caractère permettraient d'assurer des résultats prévisibles.
3. Les classements automatisés doivent être prévisibles de façon absolue, puisque c'est leur véritable raison d'être.
4. Pour trier correctement du texte, il faut au moins quatre clés, qui vont de l'information la plus significative à la moins significative.
5. Dans certaines langues, le tri nécessite une analyse contextuelle, même au niveau de classement le plus bas.
6. Le niveau de classement le plus bas dépend de chaque langue, et il faut mettre au point un moyen d'identifier les schèmes de classement non seulement au niveau des caractères comme c'est le cas actuellement, mais aussi dans les systèmes.

J'espère que cette présentation, en dépit de son caractère assez technique, a été claire pour tous, et que vous avez apprécié la leçon que nous livre la vieille civilisation chinoise, qui déploie des merveilles d'ingéniosité pour arriver à classer ses idéogrammes fascinants d'une manière toujours plus efficace. Cela devrait tous nous inciter à rechercher sans cesse la précision dans notre travail et à améliorer la qualité des classements. Je vous remercie de votre bienveillante attention.

Liste des sigles utilisés

ASCII : « American Standard Code for Information Interchange » À l'origine, code de 7 bits pour représenter 96 caractères normalisés, et dont on connaît aujourd'hui plusieurs extensions à 8 bits.

EBCDIC : « Extended Binary Coded Decimal Interchange Code » : Code utilisé pour la représentation interne des caractères dans les grands systèmes IBM.

ISO : (Provient de la racine grecque ἴσος, isos », qui signifie « égal ») « International Organization for Standardization » Organisation internationale de normalisation.

SEAS : « SHARE European Association » : Association européenne d'utilisateurs de grands systèmes IBM. (SHARE : Raison sociale de la principale association américaine d'utilisateurs de grands systèmes IBM).

1. Le texte français de cette requête a été publié dans le document intitulé **Fonctions de systèmes - Soutien des langues nationales**, publié par la Direction générale des technologies de l'information du ministère des

Communications du Québec.

© **Gouvernement du Québec**
Dépôt légal - 4^e trimestre 1988
Bibliothèque nationale du Québec
ISBN 2-550-19307-5

[[Ministre](#) | [Communiqués et discours](#)]
[[Politique québécoise de l'autoroute de l'information](#)]
[[Fonds de l'autoroute de l'information](#) | [Réalisations et projets](#)]
[[À l'agenda](#) | [ABC des inforoutes](#) | [Services publics en ligne](#)]
[[Plan du site](#) | [Recherche](#) | [Aide](#) | [Contactez-nous](#) | [Documentation](#)]
[[Autres sites](#) | [Archives](#)]
[[Accueil](#)]



© [Gouvernement du Québec, 2000](#)

«Ordre alphabétique»: une symétrie nuisible sous différentes architectures

ASCII			EBCDIC		
0	A	a	a	A	0
1	B	b	b	B	1
2	C	c	c	C	2
3	D	d	d	D	3
4	E	e	e	E	4
5	F	f	f	F	5
6	G	g	g	G	6
7	H	h	h	H	7
8	I	i	i	I	8
9	J	j	j	J	9
	K	k	k	K	
	L	l	l	L	
	M	m	m	M	
	N	n	n	N	
	O	o	o	O	
	P	p	p	P	
	Q	q	q	Q	
	R	r	r	R	
	S	s	s	S	
	T	t	t	T	
	U	u	u	U	
	V	v	v	V	
	W	w	w	W	
	X	x	x	X	
	Y	y	y	Y	
	Z	z	z	Z	

FIGURE 1

«Ordre alphabétique» en anglais pour les lettrés informatiques

CO-OP

CO-STAR

CONTAINER

COOP

COPENHAGEN

VICE VERSA

VICE-PRESIDENT

FIGURE 2

～ | 消除～思想。

【名列前茅】 míng liè qián máo 指名次列在前面(前茅, 春秋时代楚国行军, 有人拿着茅当旗子走在队伍的前面)。

【名流】 míng liú 旧时称著名的人士(多指学术界、政治界)。

【名落孙山】 míng luò sūn shān 宋朝孙山考中了末一名回家, 有人向他打听自己的儿子考中了没有, 孙山说: '解名尽处是孙山, 贤郎更在孙山外'(见于宋范公称《过庭录》)。后来用来婉言应考不中。 「～。」

【名目】 míng mù 事物的名称, ～繁多 | 巧立

【名牌】 míng pái ①出名(货物)的牌子, ～货 | ～商品。②写着人名的牌子, 标明物品名称等的牌子

【名片】 míng piàn 相互联系时自己的姓名、职

【名气】 míng qi 的医生。

【名人】 míng rén 著名的人

【名声】 míng shēng 在社会上流, ～ | ～很坏。

【名胜】 míng shèng 有古迹或优美风景的著名的地方。

【名士】 míng shì ①旧时指以诗文等著称的人。②旧时指名望很高而不做官的人。

【名士派】 míng shì pài 旧时指知识分子中不拘小节、自由散漫的一派人, 也指这种人的作风。 「名的人。」

【名手】 míng shǒu 因文笔、技艺等高超而著

【名数】 míng shù ①带有单位名称的数, 如三斤、四尺二寸、五本等。

【名堂】 míng táng ①花样; 名目等。联欢会上～真多, 又有跳舞, 又有杂耍。②成就; 结果, 我就不相信我搞不出～来! 跟他讨论了半天, 也没讨论出个～来。③道理; 内容, 真不简单, 这里面还有～呢。

【名头】 míng tóu ①方名。②名声。

【名望】 míng wàng 好的名声, 张大夫医术高明, 在这一带很有～。

【名位】 míng wèi 名声和地位。

【名物】 míng wù 物品的名称和形状。

【名下】 míng xià 某人名义之下, 指属于某人或跟某人有关, 今天下午的话儿是小李替我

干的, 工分不能记在我的～ | 这事怎么搞到我～来了?

【名学】 míng xué 逻辑学的旧称。

【名义】 míng yì ①做某事时用来作为依据的名称或称号, 我以共青团员的～向你保证, 一定提前完成任务。②表面上; 形式上(后面多带'上'字)。

【名义工资】 míng yì gōng zī ①工人付出劳动力时所得到的以货币表现出来的工资。名义工资不能确切反映出工资的实际水平, 因为名义工资不变, 实际工资可以因物价的涨跌而降低或上升。参考[实际工资]。

【名誉】 míng yù ①个人或集团的名声, 爱惜～。②名义上的(多指赠给的名义, 含尊重意), ～会员 | ～主席。

【名正言顺】 míng zhèng yán shùn 名义正当, 道理也讲得通。 「～。」

【名著】 míng zhù 有价值的出名著作。文学

【名字】 míng zì ①一个或几个字, 跟姓合在一起代表一个人, 区别于别的人。②一个或几个字, 用来代表一种事物, 区别于别种事物。

明¹ míng ①明亮(跟'暗'相对), ～月 | 天 | 灯火通～。②明白, 清楚, 问～ | 说～ | 黑白分～ | 身分不～ | 去向不～。③公开; 显露在外; 不隐蔽(跟'暗'相对), ～沟 | 有话～说 | 不敢～言 | ～令公布 | ～枪易躲, 暗箭难防。④眼力好; 眼光正确; 对事物现象看得清, 聪～ | 英～ | 精～强干 | 开～人士 | 眼～手快 | 耳不聪, 目不～。⑤心地光明, ～人不做暗事。⑥视觉, 双目失～。⑦懂得; 了解, 深～大义 | 不～利害。

明² míng 次于今年、今天的, ～天 | ～晨 | ～年 | ～春。

明³ míng ①朝代, 公元1368—1644, 朱元璋所建。先定都南京, 永乐年间迁都北京。②姓。

【明白】 míng bai ①内容、意思等使人容易了解; 清楚; 明确, 他讲得十分～。②公开的; 不含糊的, 在知识分子中～提出思想改造的任务, 是完全必要的。③聪明; 懂道理, 他是个～人, 不用说就知道。④知道; 了解; 集体合作的好处他们是容易～的。

【明摆着】 míng bǎi zhe 明显地摆在眼前, 容易看得清楚。

部首目录

: 4 traits

1. 部首左边的号码表示部首的次序。
2. 部首右边的号码指检字表的页码。
3. 没有序号的是重见部首。

一 画	二 画	三 画	四 画	五 画	六 画	七 画	八 画	九 画	十 画
1 丨 10	7 十 10	38 干 14	80 王 25	123 尸 10	163 来 35	202 艮 38	235 雨 40	236 影 40	237 門 40
2 丨 10	8 厂 10	39 工 14	81 主 26	124 丩 10	164 耳 35	203 鹵(鹵) 38	238 高 40	239 产 40	240 黄 40
3 丿 10	9 ナ 10	40 土(土) 14	82 天(天) 26	125 心 10	165 犮 35	204 里 38	241 麻 40	242 鹿 40	243 需 40
4 丶 10	10 匚 10	41 扌 15	83 韦(韋) 26	126 巾 10	166 亞(亞) 35	205 足(足) 38	244 猓 40	245 黑 40	246 暴 40
5 一 10	11 匸 10	42 升 16	84 步 26	127 尺 10	167 亞(亞) 35	206 身 38	247 猓 40	248 猓 40	249 猓 40
6 乙 10	12 冂 10	43 大 16	85 廿(廿) 26	128 水 10	168 西(西) 35	207 谷 38	250 猓 40	251 猓 40	252 猓 40
(乚) 10	13 冂 11	44 兀 16	86 木 26	129 火 10	169 束 35	208 禾 38	253 猓 40	254 猓 40	255 猓 40
二 画	14 乚 11	45 尤 16	87 支 27	130 玉 10	170 而 35	209 菊 38	256 猓 40	257 猓 40	258 猓 40
7 十 10	15 彳 11	46 寸 17	88 不 27	131 示 10	171 頁(頁) 35	210 龜(龜) 38	259 猓 40	260 猓 40	261 猓 40
8 厂 10	16 厂 12	47 寸 17	89 犬 27	132 去 10	172 至 35	211 角 38	262 猓 40	263 猓 40	264 猓 40
9 ナ 10	17 八(ノ) 12	48 乚 17	90 歹 27	133 甘 10	173 虎 35	212 卿 38	265 猓 40	266 猓 40	267 猓 40
10 匚 10	18 人(人) 12	49 口 17	91 车(車) 27	134 艸(艸) 10	174 光 35	213 寔 38	268 猓 40	269 猓 40	270 猓 40
11 卜(卜) 10	19 X 12	50 冂 19	92 戈 27	135 石 10	175 虫 35	214 幸 38	271 猓 40	272 猓 40	273 猓 40
12 冂 10	20 勹 12	51 巾 19	93 比 27	136 戊 10	176 肉 36	215 青 39	274 猓 40	275 猓 40	276 猓 40
13 冂 11	21 匕 12	52 山 19	94 无 27	137 龙(龍) 10	177 飠 36	216 猓 39	277 猓 40	278 猓 40	279 猓 40
14 乚 11	22 儿(儿) 12	53 彳 19	95 牙 27	138 业 10	178 舌 36	217 猓 39	280 猓 40	281 猓 40	282 猓 40
15 彳 11	23 夕(見32刀) 12	54 彳 19	96 瓦 27	139 彳 10	179 竹(竹) 36	218 猓 39	283 猓 40	284 猓 40	285 猓 40
16 厂 12	24 一 12	55 彳 20	97 止 28	140 彳 10	180 白 37	219 猓 39	286 猓 40	287 猓 40	288 猓 40
17 八(ノ) 12	25 彳 12	56 夕 20	98 支 28	141 申 28	181 自 37	220 猓 39	289 猓 40	290 猓 40	291 猓 40
18 人(人) 12	26 一 12	57 夕 20	99 日(日) 28	142 田 28	182 彳 37	221 猓 39	292 猓 40	293 猓 40	294 猓 40
19 X 12	27 乚(言) 12	58 丸 20	100 日 28	143 由 28	183 彳 37	222 猓 39	295 猓 40	296 猓 40	297 猓 40
20 勹 12	28 匚 13	59 彳(彳) 20	101 中 28	144 囟 28	184 彳 37	223 猓 39	298 猓 40	299 猓 40	300 猓 40
21 匕 12	29 尸 13	60 广 20	102 貝(貝) 28	145 囟 28	185 彳 37	224 猓 39	301 猓 40	302 猓 40	303 猓 40
22 儿(儿) 12	30 冂 13	61 亡 20	103 見(見) 28	146 彳(彳) 28	186 彳(彳) 37	225 猓 39	304 猓 40	305 猓 40	306 猓 40
23 儿(儿) 12	31 冂 13	62 彳 20	104 牛(牛) 28	147 生 28	187 彳(彳) 37	226 猓 39	307 猓 40	308 猓 40	309 猓 40
24 一 12	32 刀(刀) 14	63 冂(冂) 21	105 手 28	148 矢 28	188 衣 37	227 猓 39	310 猓 40	311 猓 40	312 猓 40
25 彳 12	33 力 14	64 冂(冂) 21	106 毛 28	149 禾 29	189 彳 37	228 猓 39	313 猓 40	314 猓 40	315 猓 40
26 一 12	34 マ 14	65 彳 21	107 气 28	150 白 29	190 典 37	229 猓 39	316 猓 40	317 猓 40	318 猓 40
27 乚(言) 12	35 厶 14	66 彳 25	108 片 29	151 瓜 33	191 米 37	230 猓 39	319 猓 40	320 猓 40	321 猓 40
28 匚 13	36 又 14	67 彳 25	109 女 29	152 鳥(鳥) 33	192 艸(艸) 37	231 猓 39	322 猓 40	323 猓 40	324 猓 40
29 尸 13	37 乚 14	68 彳 25	110 斤 29	153 广 34	193 羽 37	232 猓 39	325 猓 40	326 猓 40	327 猓 40
30 冂 13		69 尸 25	111 爪(爪) 29	154 立 34	194 彳 37	233 猓 39	328 猓 40	329 猓 40	330 猓 40
31 冂 13		70 己(巳) 25	112 父 29	155 玄 34	195 彳(彳) 37	234 猓 39	331 猓 40	332 猓 40	333 猓 40
32 刀(刀) 14		71 弓 24	113 月 29	156 穴 34	196 走 37	235 猓 39	334 猓 40	335 猓 40	336 猓 40
33 力 14		72 中(中) 24	114 氏 29	157 彳 34	197 彳 37	236 猓 39	337 猓 40	338 猓 40	339 猓 40
34 マ 14		73 女 24	115 欠 29	158 民 35	198 彳 37	237 猓 39	340 猓 40	341 猓 40	342 猓 40
35 厶 14		74 小(小) 24	116 風(風) 30	159 疋 35	199 彳 37	238 猓 39	343 猓 40	344 猓 40	345 猓 40
36 又 14		75 子(子) 24	117 彳 30	160 皮 35	200 彳 37	239 猓 39	346 猓 40	347 猓 40	348 猓 40
37 乚 14		76 馬(馬) 24	118 文 30	161 彳 35	201 彳 38	240 猓 39	349 猓 40	350 猓 40	351 猓 40
		77 彳(彳) 24	119 方 30	162 彳 35		241 猓 39	352 猓 40	353 猓 40	354 猓 40
		78 彳 25	120 火 30			242 猓 39	355 猓 40	356 猓 40	357 猓 40
		79 彳 25	121 一 30			243 猓 39	358 猓 40	359 猓 40	360 猓 40
			122 彳 30			244 猓 39	361 猓 40	362 猓 40	363 猓 40

p. 28 n° 99

FIGURE 4

～|消除～思想。

【名列前茅】 míng liè qián máo 指名次列在前面(前茅,春秋时代楚国行军,有人拿着茅当旗子走在队伍的前面)。

【名流】 míng liú 旧时称著名的人士(多指学术界、政治界)。

【名落孙山】 míng luò sūn shān 宋朝孙山考中了末一名回家,有人向他打听自己的儿子考中了没有,孙山说,‘解名尽处是孙山,贤郎更在孙山外’(见于宋范公称《过庭录》)。后来用来婉言应考不中。 「～」

【名目】 míng mù 事物的名称,～繁多|巧立

【名牌】 míng pái ①出名(货物)的牌子,～货|～商品。②写着人名的牌子,标明物品名称等的牌子。

【名片】 míng piàn (～儿)旧时拜访人或跟人相互联系时所用的长方形纸片,上面印着自己的姓名、职位、地址等。

【名气】 míng qì 〈口〉名声,他是一位很有～的。

【名人】 míng rén 有名的人,～墨迹。

【名声】 míng shēng ①流传的评价,好

～|～败坏。

【名胜】 míng shèng 有名的

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

干的,工分不能记在我的～|这事怎么搞到我～来了?

【名学】 míng xué 逻辑学的旧称。

【名义】 míng yì ①做某事时用来作为依据的名称或称号,我以共青团员的～向你保证,一定提前完成任务。②表面上,形式上(后面多带‘上’字)。

【名义工资】 míng yì gōng zī 〈经〉工人付出劳动力时所得到的以货币表现出来的工资。名义工资不能确切反映出工资的实际水平,因为名义工资不变,实际工资可以因物价的涨跌而降低或上升。参考【实际工资】。

【名誉】 míng yù ①个人或集团的名声,爱惜～。②名义上的(多指赠给的名义,含尊重),～会员|～主席。

【名正言顺】 míng zhèng yán shùn 名义正当,道理也讲得通。 「～」

【名著】 míng zhù 有价值的出名著作,文学

【名字】 míng zì ①一个或几个字,跟姓合在一起代表一个人,区别干别的人。②一个或几个字,用来代表一种事物,区别于别种事物。

明 míng ①明亮(跟‘暗’相对),～月|天～|灯火通～。②明白,清楚,问～|

～|黑白分～|身分不～|去向不～。③公开,显露在外,不隐蔽(跟‘暗’相对),～沟|有话～说|不败～言|～令公布|～枪易躲,暗箭难防。④眼力好,眼光正确,对事物现象看得清,聪～|英～|精～强干|开～人士|眼～手快|耳不聪,目不～。⑤心地光明,～人不做暗事。⑥视觉,双目失～。⑦懂得,了解,深～大义|不～利害。

明 míng 次于今年、今天的,～天|～晨|～年|～春。

明 míng ①朝代,公元1368—1644,朱元璋所建。先定都南京,永乐年间迁都北京。②姓。

【明白】 míng bai ①内容、意思等使人容易了解,清楚,明确,他讲得十分～。②公开的,不含糊的,在知识分子中～提出思想改造的任务,是完全必要的。③聪明,懂道理,他是个～人,不用说就知道。④知道,了解,集体合作的好处他们是容易～的。

【明摆着】 míng bǎi zhe 明显地摆在眼前,容易看得清楚。

明

明,～墨迹。

上流的评价,好

～|～败坏。

【名胜】 míng shèng 有名的

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

【名士】 míng shì 有名的人。

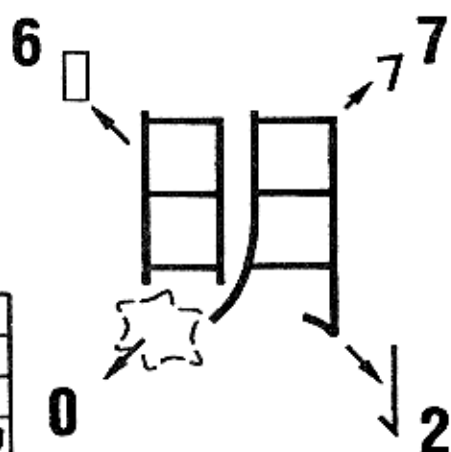
四角号码检字

四角号码查字法

甲 笔形和代号

本查字法分笔形为十种,用0到9十个号码代表,如下:

笔名	号码	笔形	字 例	说 明
点	0	丶	主病广告	点和横相结合
单	横	一	天土	横
		ノ ㄣ	活培织凡凤	挑、横上钩和斜右钩
垂	2	丨	旧山	直
		丿 丨	千顺力刚	撇和直左钩
点	3	丶	宝社军外去亦	点
		ㄣ	造瓜	捺
又	4	十	古算	两笔交叉
		ナ ㄣ ㄣ	对武虎猪	
串	5	丰	骨本	一笔穿过两笔或两笔以上
		才 ㄣ 丰	打戈森申吏	
方	6	口	另扣国甲曲曲	四角整齐的方形
		凵 口	囧 四	
角	7	ㄣ ㄣ ㄣ	刀 写 亡 表	一笔的转折
		ㄣ ㄣ ㄣ	阳 兵 又 雪	两笔笔头相接所成的角形
入	8	八	分 共	八字形
		人 入 ㄣ ㄣ	余 央 余 羊 午	八字形的变形
小	9	小	尖 家	小字形
		ㄣ 小 ㄣ ㄣ	快 本 景 当 兴 组	小字形的变形



6702

乙 查字方法

一、取角顺序 每字按①左上角、②右上角、③左下角、④右下角的次序取四个角的号码。

例 ①左上角0-----②右上角2
③左下角1-----④右下角2 -0212

：顏：-0128 ；截：-4325 ；烙：-9786

(6404.)	噤	530	6438,	噤	517	吧	1220	6612,	乳	467	噤	602
(時)	929	6409,	噤	241	6508,	6602,	噤	679	6701,	噤	1080	
時	1331	噤	986	6438,	噤	1287	噤	1050	噤	737	噤	727
(時)	138	6409,	噤	(噤)	241	噤	1211	噤	1346	741	729, 730	
6404,		噤	225	6462,	噤	6509,	噤	401	噤	1116	736, 737	
咭	1320	噤	1290	噤	1161	噤	1070	噤	6615,	噤	768	(噤)
	1362	(噤)	978	6480,	噤	噤	1343	噤	(噤)	56	噤	1050
噤	77	6409,	噤	(噤)	1069	噤	694	噤	6618,	噤	546	噤
噤	1127	噤	635	6482,	噤	600	(噤)	1191	噤	215	1379	噤
噤	73	噤	637	噤	1169	噤	980	噤	6619,	噤	743	噤
噤	616	(噤)	636	6485,	噤	噤	6603,	噤	噤	436	1056	(噤)
噤	455	6409,	噤	(噤)	1283	6509,	噤	727	噤	1286	138	(噤)
461, 758		(噤)	600	6488,	噤	830	噤	740	6621,	噤	743	6703,
6405,		6410,		(噤)	953	6510,	噤	6603,	噤	550	噤	422
(噤)	431	噤	299	6500,	噤	1352	噤	410	噤	845	1077	噤
	432	6411,		噤	412	6511,	噤	720	6621,	(噤)	758	1183
(噤)	1202	噤	1114	噤	338	噤	824	6603,	噤	1342	6701,	6703,
6405,		(噤)	824	6500,	噤	6514,	噤	1071	6624,	噤	1197	噤
(噤)	1068	6411,		噤	909	噤	138	6603,	(噤)	1180	6701,	6704,
6406,		噤	592	6501,	噤	6516,	噤	801	6632,	噤	14	噤
咭	352	6411,		噤	1127	噤	149				17	噤
噤	705	噤	103	噤	1133	6518,						6704,
噤	706	6411,		6501,	噤	噤	900					噤
咭	241	噤	120	噤	249	噤						1234
6406,		6412,		噤	249	6518,						6704,
咭	465	噤	1217	6502,	噤	(噤)						1171
噤	169	6412,		噤	1134	6580,						噤
	989	噤	585	噤	1068	噤	1					1094
噤	941	噤	765	噤	837	6581,		噤	395			146
6406,	1097	(噤)	765	噤	532	(噤)	530	6604,	(噤)	397		155
噤	756	6414,		(噤)	1134	6584,	351	6604,	噤	1381		1171
855		(噤)	138	噤	1356	6600,		噤				830
6408,		6414,		(噤)	753	噤	1177	6604,	(噤)	1161		702
噤	412	噤	803	噤	733	1186, 1202		噤	397	711		
	417	噤	811	6504,		噤	312	6605,		1002		
噤	235	噤	77	噤	138	噤	558	噤	1104	噤	465	
噤	118	6416,		6504,	1356	(噤)	593	6605,	(噤)	811	1300	6705,
噤	1010	噤	473	6504,	656	6600,	1280	6605,	噤	54	1342	噤
噤	118	6416,		6504,	658	噤	1282, 1283	6605,	噤	107	986	6705,
6408,		噤	144	(噤)	656	6601,	760	6606,	噤	994	1113	1313
噤	772	6418,		(噤)	656	(噤)	1342	噤	112	6671,	191	6706,
773		噤	1379	6505,	297	(噤)	121	噤	954	6671,	498	噤
6408,		噤	226	6506,	95	6601,	1286	噤	1157	6680,	393	噤
噤	1238	6419,		噤	452	噤	444	噤	1286	噤	116	6706,
(噤)	996	6419,		6507,	452	6601,	617	噤	798	6699,	1301	1300
噤	395	噤	635	6508,	299	噤	626, 1233	噤	483	6700,	565	6706,
6408,		6431,		噤	1220	噤	671	噤	129	6701,	330	噤
噤	892	噤	189	噤	1237	6601,					559, 653	
6408,		6432,		噤	225	噤	669				671	
噤	446	噤	1251	噤	1105	噤					667	
6409,		6436,		噤	1105	噤					6707,	748
噤	1154	噤		噤	1310	6508,					6707,	190
6409,	640	(噤)										

FIGURE 8

～ | 消除～思想。

【名列前茅】 míng liè qián máo 指名次列在前面(前茅, 春秋时代楚国行军, 有人拿着茅当旗子走在队伍的前面)。

【名流】 míng liú 旧时称著名的人士(多指学术界、政治界)。

【名落孙山】 míng luò sūn shān 宋朝孙山考中了末一名回家, 有人向他打听自己的儿子考中了没有, 孙山说, ‘解名尽处是孙山, 贤郎更在孙山外’(见于宋范公称《过庭录》)。后来用来婉言应考不中。 「～。」

【名目】 míng mù 事物的名称, ～繁多 | 巧立

【名牌】 míng pái ①出名(货物)的牌子, ～货 | ～商品。②写着人名的牌子, 标明物品名称等的牌子。

【名片】 míng piàn (～儿) 旧时拜访人或跟人相互联系时所用的长方形纸片, 上面印着自己的姓名、职位、地址等。

【名气】 míng qì 〈口〉名声, 他是一位很有～的

【名人】 míng rén 有名的人。

【名声】 míng shēng ①名声, 他是一位很有～的

【名姓】 míng xìng ①有名的人。②有名的人。

【名士】 míng shì 有名的人。

【名士派】 míng shì pài 指知识分子中不拘小节、自由散漫的一派人, 也指这种人的作风。

【名手】 míng shǒu 因文笔、技艺等高超而著名

【名歌】 míng gē 〈歌〉带有单位名称的歌, 如三斤、四尺二寸、五本等。

【名堂】 míng táng ①花样, 名目等, 联欢会上～真多, 又有跳舞, 又有杂耍。②成就, 结果, 我就不相信我搞不出～来 | 跟他讨论了半天, 也没讨论出个～来。③道理, 内容, 真不简单, 这里面还有～呢。

【名头】 míng tóu 〈方〉名声。

【名望】 míng wàng 好的名声, 张大夫医术高明, 在这一带很有～。

【名位】 míng wèi 名声和地位。

【名物】 míng wù 物品的名称和形状。

【名下】 míng xià 某人名义之下, 指属于某人或跟某人有关, 今天下午的活儿是小李替我

干的, 工分不能记在我的～ | 这事怎么搞到我～来了?

【名学】 míng xué 逻辑学的旧称。

【名义】 míng yì ①做某事时用来作为依据的名称或称号, 我以共青团员的～向你保证, 一定提前完成任务。②表面上, 形式上(后面多带‘上’字)。

【名义工资】 míng yì gōng zī 〈经〉工人付出劳动力时所得到的以货币表现出来的工资。名义工资不能确切反映出工资的实际水平, 因为名义工资不变, 实际工资可以因物价的涨跌而降低或上升。参考[实际工资]。

【名望】 míng wàng ①个人或集团的名声, 爱憎～。②名义上的(多指赠给的名义, 含尊重意), ～会员 | ～主席。

【名正言顺】 míng zhèng yán shùn 名义正当, 道理也讲得通。 「～。」

【名著】 míng zhù 有价值的出名著作, 文学

【名字】 míng zì ①一个或几个字, 跟姓合在一起代表一个人, 区别于别的人。②一个或几个字, 用来代表一种事物, 区别于别种事物。

明' míng ①明亮(跟‘暗’相对), ～月 | 天 | ～火通～。②明白, 清楚, 问～ | 晓～ | 黑白分～ | 身分不～ | 去向不～。③公开, 显露在外, 不隐蔽(跟‘暗’相对), ～沟 | 有话～说 | 不敢～言 | ～令公布 | ～枪易躲, 暗箭难防。④眼力好, 眼光正确, 对事物现象看得清, 聪～ | 英～ | 精～强干 | 开～人士 | 眼～手快 | 耳不聪, 目不～。⑤心地光明, ～人不做暗事。⑥视觉, 双目失～。⑦懂得, 了解, 深～大义 | 不～利害。

明' míng 次于今年、今天的, ～天 | ～晨 | ～年 | ～春。

明' míng ①朝代, 公元1368—1644, 朱元璋所建。先定都南京, 永乐年间迁都北京。②姓。

【明白】 míng bai ①内容、意思等使人容易了解, 清楚, 明确, 他讲得十分～。②公开的, 不含糊的, 在知识分子中～提出思想改造的任务, 是完全必要的。③聪明, 懂道理, 他是个～人, 不用说就知道。④知道, 了解, 集体合作的好处他们是容易～的。

【明摆着】 míng bǎi zhe 明显地摆在眼前, 容易看得清楚。



1. ～画迹。

好

看

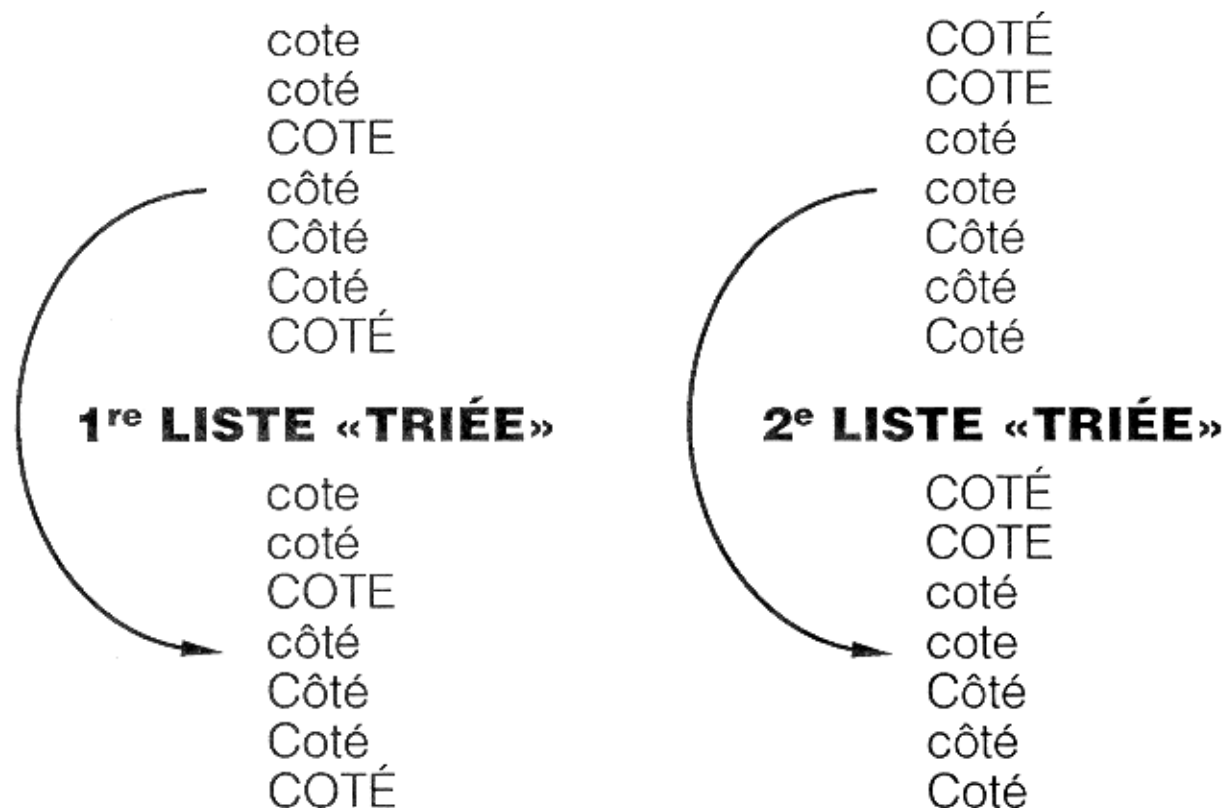
【诗文等著称的

做官的人。

1^{re} LISTE

2^e LISTE

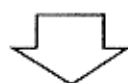
(même liste que la première, mais dans un ordre différent)



RÉSULTAT CORRECT:

cote
COTE
coté
Coté
COTÉ
côté
Côté

Convertir en majuscules et enlever les accents détruit l'intégrité



Résultats imprévisibles du classement

FIGURE 10

CLÉ DE TRI

4 SOUS-CLÉS

1^{er} ordre

2^e ordre

3^e ordre

4^e ordre

Base
alphanumérique

Accents

Casse

Information non-
alphanumérique

FIGURE 11

peau neuve, changer complètement. « Il sortait du saltim-
trait dans le lord. Changements de peau qui sont
changements d'âme » (HUO). O (1850) Dans
la vie, l'existence. Jouer, risquer, craindre pour,
lancer sa peau. « Il se sentait capable de tout pour
lui, de fuir, de demander grâce, de trahir, et pour-
rait pas tellement à sa peau » (SARTRE). « Il bran-
lings, il guralait qu'il aurait leur peau à tous »
dre cher sa peau, se défendre vaillamment. Pop.
a peau, on le tuera. § 3* (1845, « prostituée »).
peau : injure adressée à une femme. « Il culbuta
dans les cols, tout en la traitant de vieille peau »,
ze » (ZOLA). § 4* (xv). La dépouille de cer-
vux destinée à fournir la fourrure, le cuir. La
ry*. Traitement, travail de peaux. V. Cuir; tan-
ôt d'elle logeait un artisan tanneur. Il tannait
petites peaux d'animaux. Il les pendait pour les
aux volets de sa fenêtre » (Giono). Ouvriers des
ux : corroyeurs, mégissiers, tanneurs, etc. Peau
- Etal en peau de serpent. Les peaux d'un man-
ture. « Une collection de valises plates en peau
LARBAUD). Pelisse en peau de mouton. Absolt.
simple. Culotte de peau. Gants de peau. O Fam.
: diplôme, parchemin. — Peau de chagrin (d'apr.
Balzac), bien matériel ou moral qui s'amenuise.

§ 5* Peau de saumon. § 6* (1538). Enveloppe extérieure
des fruits. V. Épicarpe. Enlever, ôter la peau d'un fruit. V.
Peler. Peau de pêche. Glisser sur une peau de banane. O Fig.
Peau d'orange. Méd. Aspect en peau d'orange de l'épi-
derme, dans la cellulite. — Peau du lait, pellicule qui se
forme sur le lait au repos. § 7* Pop. (1872). Peau de balle;
et vulg. Peau de zébi : rien du tout. « C'est toute la pièce ou
peau de zébi » (CÉLINE). Absolt. La peau ! exclam. de refus,
de mépris. « Pour ce qui est des bougies... la peau !... elles
sont sous clé » (MIRBEAU). O NOM. P. F.

PEAUCIER [posaj]. adj. et n. m. (1560; de peau). Anat.
Muscle peaucier, et subst. Un peaucier : muscle superficiel qui
s'attache à la face profonde du derme. Peaucier du cou. O
NOM. Peaussier.

PEAUFINER [pofine]. v. tr. (1883; de peaufiner, 1865;
de peau, et fin). § 1* Nettoyer avec une peau de chamois.
§ 2* Fig. et fam. Préparer, orner minutieusement; signoler
(un travail). — Au p. p. « Un élégant pompadour, parfumé,
peaufiné » (J.-R. BLOCH). — Dér. PEAUFINAGE [pofinaj].
n. m.

PEAU-ROUGE [poru]. n. (1858; de peau, et rouge).
Indien d'Amérique. Les Peaux-Rouges se teignent le visage
enocre.

PEAUSSERIE [posaj]. n. f. (1723; de peau). § 1* Com-
merce, métier, travail des peaux, des cuirs. § 2* Une, des
peausseries(s), peau travaillée. V. Cuir, peau.

PEAUSSIER [posaj]. n. et adj. m. (1545; paucier, 1292;
de peau). Artisan, ouvrier qui prépare les peaux pour les
transformer en cuirs. O NOM. Proucier.

PÉBRINE [pebrin]. n. f. (1839; prov. mod. pebrino, de
pore « poivre »). Agric. Maladie des vers à soie.

PÉBROC ou PÉBROQUE [pebrak]. n. m. (1907; de
pépin 2. et suff. arg.). Arg. Paraphrase. « J'ai oublié mon
pébroque ou bistrot » (QUENEAU).

PÉCATREI [pekain]. interj. (1773; pechère, xiv; prov.
peccare « pécher », francisé en peuchère). Région. (Provenç.).
Exclamation exprimant une commisération effective ou
ironique.

PÉCARI [pekar]. n. m. (1699; pequir, 1640; mot
carabe). Sorte de sanglier (Suidés), cochon sauvage d'Amé-
rique. O Cuir de cet animal. Des gants de pécar.

PECCABLE [pekab(a)]. adj. (1050; lat. peccā, de pec-
care). Relig. Sujet à pécher. « Si Dieu a créé l'homme peccable,
il ne devait pas le punir » (FLAUB). Par ext. La nature peccable
de l'homme.

PECCADILLE [pekadil]. n. f. (Peccatille, 1539; msc.,
1660; esp. peccadillo « petit péché »). Littér. Poché, faute sans
gravité. « Sa peccadille fut jugée un cas pendable » (LA FONT.).
« La peccadille du soldat est un crime chez le général, et réci-
proquement » (BALZ.).

PECCANT, ANTE [pekā, āt]. adj. (Pechantes, 1314; du
lat. médiév. peccans, de peccare « pécher »). Vx. Humeurs
peccantes : mauvaises.

PECHBLÉNDE [pebld]. n. f. (1790; all. Pech « poix »,
et Blende). V. Blende. Miner. Minéral renfermant une forte
proportion d'uranium. V. Uraninite. P. et M. Curie ont
découvert le polonium et le radium en portant de la pechblende.

1. PÊCHE [pe]. n. f. (1671; peche, xiv; lat. pop. persica,
n. f., plur. de persicum [pomum] « fruit de Perse »). § 1* Fruit
du pêcher, à noyau très dur et à chair fine. Pêche à prau
lisse. V. Brugnon. Pêche-abricot. Pêche de vigne. — Pêche
Meibor. O Loc. fig. Peau, teint de pêche : rose et velouté.
— Fam. Rembourré avec des noyaux de pêche : très dur.
O Appos. Couleur pêche, d'un rose qui rappelle la peau

d'une pêche. § 2* Pop. Coup, gifle. Il va te flanquer une pêche.
§ 3* Fam. Visage. Loc. Se fendre la pêche. Rire. (Cf. Se
fendre la pipe). § 4* Fam. Avoir la pêche. avoir le moral, être
en forme. O NOM. Formes des v. pêcher et pêcher.

2. PÊCHE [pe]. n. f. (Pesche, 1261, « droit de pêcher »;
du v. pêcher). § 1° Action ou manière de prendre les pois-
sons. V. Halieutique. Ouverture, clôture, fermeture de la
pêche : de la période où la pêche est autorisée. Engins de
pêche : filet, ligne, nasse; trident. Pêche hauturière. Grande
pêche au large (ex. : morue, flétan). Petite pêche, côtière
(colin, merlan, raie). Pêche à la ligne (et absol. pêche).
Articles de pêche : bouchon, épuisette, flotteur, gaulle, hame-
çon, moulinet, plomb. Pêche au coup, au lancer. Pêche au
chahut (V. Chalutage), à la seine. Pêche sous-marine. Pêche
artisanale, industrielle. — Loc. La pêche miraculeuse, que
le Christ fit faire à ses disciples. § 2° Endroit où l'on pêche,
où l'on peut pêcher. Garde-pêche qui surveille une pêche
réservée. § 3* (1538). Poissons, produits pêchés. Rapporter
une belle pêche. § 4° Dr. Droit de pêche. Riverain qui a la
pêche d'un canal jusqu'au milieu du cours de l'eau. O NOM.
PÊCHE (1); formes des v. pêcher et pêcher.

PÊCHÉ [pe]. n. m. (xiv; pechit, x; lat. peccatum
« faute, crime »). Acte conscient par lequel on contrevient
aux lois religieuses, aux volontés divines. Commettre, faire
un péché. V. Pêcher. Avoir, confesser ses péchés. Expi-
er, racheter ses péchés : faire pénitence. Absolution, rémission
des péchés. À tout péché miséricorde. « Que celui d'entre vous
qui est sans péché lui jette la première pierre » (Évang.).
« Vous avez encore une vingtaine d'années de jolis péchés à
faire : n'y manquez pas; ensuite vous vous en repentirez »
(DIDER.). — Pêché de jeunesse. Pêché mignon : défaut véniel
et agréable; petite faute habituelle. La gourmandise est son
pêché mignon. V. Faible. — Pêché mortel : qui entraîne la
damnation du pêcheur (opposé à péché véniel). Les sept
péchés capitaux. V. Avarice, colère, envie, gourmandise,
luxure, orgueil, paresse. — Pêché originel : commis par
Adam et Ève et dont tout être humain est coupable en nais-
sant. O Absolt. LE PÊCHÉ : l'état où se trouve celui qui
a commis un péché mortel (opposé à état de grâce). V. Pêcher.
Tomber, vivre dans le péché. V. Mal. « L'absurde... ne mène
pas à Dieu. L'absurde c'est le péché sans Dieu » (CAMUS).
« Le péché, qui tue l'âme, repêché le corps à son affreuse res-
semblance » (MAURAC). O NOM. Pêcher (1 et 2).

PÊCHER [pe]. v. intr. conj. céder (xiv; pechier,
1120; lat. peccare). § 1° Commettre un péché, des péchés.
V. Faillir. Pêcher par orgueil, par ignorance. « C'est nous in-
spirer presque un désir de pêcher. Que montrer tant de soins
de nous en empêcher » (MOL.). O Pêcher contre qqch. : faillir
(contre une règle). V. Contrevenir, manquer (à). Pêcher contre
la bienséance, les bonnes mœurs. § 2° (xvii). Commettre une
faute, une erreur. Pêcher contre l'esprit. « Toute cette
brochure pêche par une grande obscurité et une grande confu-
sion d'idées » (STY-BEUVE). O NOM. Pêché : pêcher (1 et 2).

1. PÊCHER [pe]. n. m. (1677; pechier, 1190; de pêche).
Arbre (Rosacées) d'origine exotique, acclimaté et cultivé
pour ses fruits, les pêches. Pêcher en espalier. « Les premiers
pêchers, d'un rose un peu fétide, fleurissent en bouppes »
(COLETTE). — Couleur (de) fleur de pêcher, d'un rose assez
vif. O NOM. Pêché : pêcher, pêcher (2).

2. PÊCHER [pe]. v. tr. (1680; pecher, 1138; lat. pop.
« piscare, class. piscari »). § 1° Prendre ou chercher à prendre
(du poisson). Pêcher la morue, la truite. — Pronom. (PAR.).
L'anguille se pêche au ver de terre. V. Prendre (se). — Absolt.
Pêcher à la ligne, au filet. Pêcher à l'asticot, à la mouche.
Pêcher en mer, dans une rivièr. Loc. fig. Pêcher en eau trouble :
profiter d'un état de désordre, de confusion. O (D'autres
animaux que les poissons) « De l'ambre mi soir il pêchait des
grenouilles pour les vendre » (GENEVARD). § 2° Fig. et fam.
Chercher, prendre, trouver (une chose inattendue) d'une
manière incompréhensible. « Où diable avez-vous pêché des
radis ? demanda Godefroid » (BALZ.). Où as-tu été pêcher ce
costume ? Je me demande où il va pêcher ces histoires. V.
Imaginer. O NOM. Pêché : pêcher (n. m.); pêcher.

PÊCHÈRE [pe]. interj. Var. de Peuchère. V. PÉCATREI.
PÊCHÈRESSE. n. f. V. PÊCHEUR.

PÊCHERIE [pe]. n. f. (1606; pecherie, 1155; du v.
pêcher). Lieu aménagé pour une entreprise de pêche. Les
pêcheries de Terre-Neuve. « Nous voici au milieu des pêcheries,
des barques, des filets tendus » (LOTI).

PÊCHETTE [pet]. n. f. (1868; « petit filet » n. 1773;
de pêcher 2). Région. Petit filet à écrevisses. V. Balance.

PÊCHEUR, PÊCHÈRESSE [pe]. n. (xii; pechior, 980; fem. pecheris, v. 1130; lat. ecclési. peccator,
« celui de peccare » « pécher »). Personne qui est dans l'état
de péché, commet habituellement de graves péchés. Pêcheur
endurci, repent. — Dieu ne veut pas la mort du pêcheur : il
est indulgent.

PÊCHEUR, BUSE [pe]. n. (Peschew, 1138; 98m.

Codes utilisés:

<00> délimiteur (4e ordre)

<08> minuscule

<09> majuscule

<16> absence d'accent

<17> accent aigu

Vice-président.

1

2

3

4

<00> <04> <-> <14> <.>

v i c e p r e s i d e n t

<16> <16> <16> <16> <16> <16> <17> <16> <16> <16> <16> <16> <16>

<09> <08> <08> <08> <08> <08> <08> <08> <08> <08> <08> <08> <08>

Vice-président.

FIGURE 13

Liste triée

Sous-clés

	1 ^{er} ordre Base latine	2 ^e ordre Accents	3 ^e ordre Casse	4 ^e ordre Spécial
cote	cote	<16><16><16><16>	<08><08><08><08>	<00>
COTE	cote	<16><16><16><16>	<09><09><09><09>	<00>
coté	cote	<17><16><16><16>	<08><08><08><08>	<00>
Coté	cote	<17><16><16><16>	<09><08><08><08>	<00>
COTÉ	cote	<17><16><16><16>	<09><09><09><09>	<00>
côté	cote	<17><16><19><16>	<08><08><08><08>	<00>
Côté	cote	<17><16><19><16>	<09><08><08><08>	<00>

Codes

<16> absence d'accent
<17> accent aigu
<18> accent grave
<19> accent circonflexe
<20> tréma
... etc.

<08> minuscule
<09> majuscule

<00> délimiteur
4^e sous-clé

FIGURE 14

**Müller
Mueller
Mulhouse
Muller**

ou

**Mulhouse
Muller
Müller
Mueller**

**Problèmes potentiels
du classement phonétique
dans certaines langues**

(impliquant ici des caractères allemands)

FIGURE 15

ALPHABET. El alfabeto español

L'alphabet espagnol comprend les lettres suivantes:

a	m
b	n
c	ñ
ch	o
d	p
e	q
f	r
g	s
h	t
i	u
j	v
k	x
l	y
ll	z

Dans l'ordre alphabétique, les combinaisons **ch** et **ll** et la lettre **ñ** sont rangées comme des **lettres distinctes**.

Interprétation des mêmes lettres en différentes langues

Français et anglais

ñ



Lettre «n» avec tilde.

**Est triée comme un «n»
avec considérations
spéciales pour les
homographes.**

Espagnol

ñ



**Une lettre distincte entre
«n» et «o».**

Est triée après "nz".

- 1. L'ASCII et l'EBCDIC ne sont pas appropriés pour le tri.**
- 2. Une analyse sémantique des caractères assure des séquences de classement rationnelles.**
- 3. Le classement automatisé doit être prévisible de façon absolue.**
- 4. 4 clés sont nécessaires pour trier le texte propre aux langues occidentales.**
- 5. Quelques langues nécessitent une analyse contextuelle.**
- 6. Niveau de classement le plus bas: dépendant de la langue. Un moyen d'identifier les différents schèmes de classement est requis dans les systèmes.**