



MINISTERE DES AFFAIRES ECONOMIQUES

NUMERO DE PUBLICATION : 1003890A6

NUMERO DE DEPOT : 9200174

Classif. Internat.: G06F

Date de délivrance : 07 Juillet 1992

---

Le Ministre des Affaires Economiques,

Vu la loi du 28 Mars 1984 sur les brevets d' invention, notamment l' article 22;

Vu l' arrêté royal du 2 Décembre 1986 relatif à la demande, à la délivrance et au maintien en vigueur des brevets d' invention, notamment l' article 28;

Vu le procès verbal dressé le 20 Février 1992 à 14h25  
à l' Office de la Propriété Industrielle

## ARRETE :

ARTICLE 1.- Il est délivré à : INISHASHE LIMITED  
Fitzwilliam Street 41, DUBLIN 2(IRLANDE)

représenté(e)(s) par : GOEGEBEUR Erik, BUGNION S.A., Rue de Namur, 43 bte 3 - B  
1000 BRUXELLES.

un brevet d' invention d' une durée de 6 ans, sous réserve du paiement des taxes annuelles, pour : PROCEDE DE CONVERSION DE BASE DE DONNEES.

INVENTEUR(S) : Biddle Martin, c/o Response House, Whitefriars, Aungier Street, Dublin 2 (IE)

ARTICLE 2.- Ce brevet est délivré sans examen préalable de la brevetabilité de l' invention, sans garantie du mérite de l' invention ou de l' exactitude de la description de celle-ci et aux risques et périls du(des) demandeur(s).

Bruxelles, le 07 Juillet 1992  
PAR DELEGATION SPECIALE :

WUYTS L.  
Directeur

## PROCEDE DE CONVERSION DE BASE DE DONNEES

1           La présente invention concerne un procédé pour  
la conversion d'une base de données d'entrée ayant des  
enregistrements avec des zones de données alphanumériques  
en une base de données de sortie qui facilitent le  
5           fonctionnement d'un processeur pour la reprise des données  
et l'impression de documents en utilisant les données  
d'une manière efficace et sans erreur.

          Un procédé de conversion de ce genre est  
nécessaire par exemple lorsqu'une base de données d'entrée  
10           avec des noms et des adresses doit être utilisée pour  
imprimer des lettres dans un mailing. Des noms et des  
adresses dans la base de données d'entrée peuvent être  
orthographiés de manière incorrecte, peuvent être en  
position incorrecte ou peuvent être absents.

15           Jusqu'à présent, la méthode pour la conversion  
d'une base de données d'entrée de ce genre consistait à  
utiliser un appareil de vérification d'orthographe tel  
celui décrit dans le brevet britannique n° 2,201,274 B  
(Brother), qui vérifie séquentiellement l'orthographe  
20           de chaque mot. Pour un procédé de conversion de base de  
données, cette approche demande beaucoup de temps, vu  
que chaque mot particulier doit être vérifié par rapport  
à un dictionnaire de grande capacité. D'autre part, des  
défauts d'exactitude dûs au conditionnement incorrect  
25           des mots dans les zones de la base de données d'entrée  
ne sont pas éliminés.

          L'invention vise à proposer un procédé de

1 conversion de base de données qui fonctionne efficacement  
et qui transforme une base de données d'entrée de façon  
à ce que le contenu des données et leur structure soient  
cohérents et corrects et puissent être utilisés pour  
5 l'impression de documents.

Conformément à la présente invention, il est  
prévu un procédé pour la conversion d'une base de données  
d'entrée ayant des enregistrements avec des zones de  
données alphanumériques en une base de données de sortie  
10 ayant un contenu de données et une structure facilitant  
le fonctionnement d'un processeur pour la reprise de  
données et l'impression de documents en utilisant les  
données d'une manière efficace et sans erreurs, le procédé  
comprenant les étapes suivantes :

15

enregistrer un ensemble de lexiques dans un  
circuit de mémoire, l'ensemble de lexiques étant  
associé à une zone particulière de cet enregistrement  
de la base de données d'entrée et chaque lexique  
20 ayant un ensemble de mots de données de référence  
et de règles associées;

25

assigner un ordre de priorité à l'ensemble des  
lexiques pour chaque position de mots, l'ordre de  
priorité étant basé sur la probabilité que le lexique  
contienne un mot de référence pour chaque position  
de la zone;

30

inscrire séquentiellement chaque enregistrement  
de la base de données d'entrée dans le circuit de  
mémoire et filtrer la zone particulière; et

35

pour chaque mot de données de chaque zone, effectuer  
les étapes suivantes :

1                   comparer le mot avec au moins certains des  
                  lexiques dans l'ordre de priorité, jusqu'à  
                  ce qu'une correspondance apparaisse;

5                   lorsqu'une correspondance apparaît, vérifier  
                  le mot par comparaison des paramètres du mot  
                  avec les règles du mot de référence dans le  
                  lexique et inscrire à nouveau le mot dans la  
10                  zone suivant les règles; et

                  si aucune correspondance n'est trouvée, insérer  
                  un indicateur dans la zone.

15                  Dans un mode de réalisation, le procédé comprend  
                  l'étape supplémentaire consistant à déterminer d'avance  
                  un indicateur de traitement et, en fonction de l'état  
                  de l'indicateur, si un indice est inséré dans une zone  
                  soit interrompre le procédé et demander l'introduction  
20                  manuelle par l'utilisateur d'une instruction ou  
                  enregistrer la zone dans un fichier d'impression pour  
                  imprimer ultérieurement une liste d'erreurs.

                  L'invention sera mieux comprise à l'examen  
                  de la description ci-après de certains de ses modes de  
                  réalisation particuliers donnés uniquement à titre  
25                  d'exemple et en se référant à la figure en annexe, qui  
                  est un diagramme synoptique représentant un procédé de  
                  conversion de base de données selon la présente invention.

                  Si on examine la figure, on peut y voir un  
30                  procédé de conversion de base de données de la présente  
                  invention, désigné généralement par le numéro de référence  
                  1. Le procédé 1 peut être effectué dans tout système  
                  de base de données informatisée approprié ayant un  
                  processeur, un dispositif de mise en mémoire, un circuit  
                  de mémoire, une unité d'affichage et une imprimante.  
35                  Quel que soit le type de système de base de données

1 informatisée utilisé , les caractéristiques techniques  
du procédé permettent, comme expliqué ci-après,  
d'améliorer l'efficacité de la conversion de la base  
de données et de diminuer de manière significative le  
5 nombre d'erreurs dans la base de données de sortie. En  
effet, il est possible d'éliminer quasi complètement  
les erreurs en utilisant le procédé 1. La première étape  
du procédé est l'étape 2, qui comprend la mise en mémoire  
dans le dispositif de mise en mémoire d'un certain nombre  
10 de lexiques. Chaque lexique est un ensemble de mots de  
référence tel qu'on peut le trouver dans un dictionnaire,  
avec également un ensemble de règles associées à chaque  
mot de référence. Il existe un lexique pour chaque "type"  
15 de données, qui peut se trouver dans une base de données  
d'entrée. Par exemple, un lexique peut concerner les  
formules de politesse pour les noms dans une base de  
données. Dans un lexique de ce genre, le mot "Esq." sera  
mis en mémoire comme mot de référence, avec une règle  
indiquant qu'il doit être enlevé de la base de données  
20 et que la formule "Mr." doit être insérée comme premier  
mot dans la zone de noms. D'une manière générale, tous  
les noms de famille ou prénoms possibles ne seront pas  
inclus dans un lexique. Il est plus efficace de créer  
25 de nombreux lexiques différents ayant chacun une partie  
du nombre total des noms possibles classés par ordre  
alphabétique. Ceci est possible parce que, si la base  
de données d'entrée est traitée, elle sera dans l'ordre  
alphabétique des noms des enregistrements de données.  
D'autres exemples de lexiques comprennent un lexique  
30 des numéros de districts postaux et un lexique des noms  
de rues. Dans ces exemples, les règles du lexique de  
districts postaux indiqueront un certain nombre de noms  
de rues possibles pour ce district postal. La création  
et la mise en mémoire de lexiques à l'étape 2 constitue  
35 une part très importante du procédé, et elle doit être

1 effectuée en tenant compte de la base de données d'entrée.  
D'une manière générale, il existera un ensemble de  
lexiques créés pour chaque type de données d'entrée.

5 L'étape 3 du procédé est également fort  
importante vu qu'elle implique la reprise de l'ensemble  
de lexique créé pour chaque base de données d'entrée  
particulière et l'attribution d'un ordre de priorité  
à celui-ci pour chaque position de données dans les zones  
10 d'enregistrement de bases de données d'entrée. Par  
exemple, l'ordre de priorité pour le premier mot dans  
une zone de noms indiquera que le lexique des formules  
de politesse a la priorité la plus haute et que le lexique  
des noms de famille correspondants a la priorité de  
15 deuxième ordre. Chaque mot "position" dans les zones  
de base de données d'entrée reçoit un ordre de priorité.  
Par exemple, l'ordre de priorité pour la troisième  
position dans une zone de noms attribuera la priorité  
la plus élevée à un lexique de prénoms. Par conséquent,  
20 pour chaque ensemble de lexique (correspondant à une  
base de données d'entrée) il existe un certain nombre  
d'ordres de priorité différents, chaque ordre de priorité  
étant associé à un mot ou à une position de données dans  
une zone des enregistrements de bases de données d'entrée.

25 A l'étape 4, on accède à la base de données  
d'entrée mise en mémoire dans le dispositif de mise en  
mémoire et on reprend un enregistrement qui est introduit  
dans le circuit de mémoire. A l'étape 5, le processeur  
filtre la zone appropriée concernée par cette partie  
spécifique du procédé. Par exemple, la zone peut être  
30 la zone de noms dans chaque enregistrement. A l'étape  
6, le processeur lit le premier mot de la zone et le  
compare au lexique qui a reçu la priorité d'ordre le  
plus élevé pour la position du premier mot dans la zone.  
Dans cet exemple, le lexique de formules de politesse  
35 est le premier choisi. A l'étape 7, le procédé détermine

1 s'il existe ou non une correspondance avec le lexique,  
c'est-à-dire s'il existe ou non un mot de référence qui  
est identique au premier mot qui a été lu dans la zone.  
Si ce n'est pas le cas, le processeur détermine à l'étape  
5 8 s'il existe un autre lexique dans l'ordre de priorité  
et, à l'étape 6, répète le procédé en comparant le mot  
au lexique ayant l'ordre de priorité suivant. Un exemple  
où il n'existe pas de correspondance pour le premier  
lexique se présente si la formule "Esq." est incluse  
10 à la fin de la zone de noms et qu'il n'existe par de  
formule de politesse au début. Dans ce cas, le premier  
mot sera lu comme étant un nom de famille et celui-ci  
est susceptible de se présenter dans le deuxième lexique.  
Dans la plupart des cas toutefois, la formule de politesse  
15 "Mr.", Ms." ou "Mrs." sera utilisée et c'est pourquoi le  
lexique de formules de politesse a reçu la priorité  
d'ordre la plus élevée. Si une correspondance est trouvée  
dans un lexique, c'est-à-dire si le mot qui a été lu est  
le même que le mot de référence dans un lexique, le  
20 processeur vérifie le mot à l'étape 9 en examinant les  
règles de contrôle associées au mot de référence. Ces  
règles de contrôle peuvent indiquer que les mots doivent  
être dans une position particulière dans la zone de noms.  
Par exemple, si la formule de politesse "Esq." est  
25 comprise à la fin d'une zone de noms dans la base de  
données d'entrée, le premier mot sera le nom de famille  
et la règle de contrôle indiquera qu'il doit se trouver  
dans la position du deuxième mot de la zone de noms et  
non dans la première. A l'étape 10, le processeur  
30 détermine, en fonction des règles de contrôle, si le  
mot dans la base de données d'entrée est correct et, à  
l'étape 11, le mot est à nouveau inscrit dans la zone,  
par exemple le nom de famille est inscrit à nouveau dans  
la deuxième position de mot de la zone.

35

Si on revient maintenant à l'étape 8 dans

1 laquelle le processeur détermine s'il existe un autre  
lexique dans l'ordre de priorité pour cette position  
de mot, si tous les lexiques dans un ordre de priorité  
ont été examinés et qu'aucune correspondance n'a été  
5 trouvée, un indice est inséré dans la zone à l'étape  
12. Lorsque ceci a été effectué, le processeur vérifie  
l'état de l'indicateur de procédé qui a été réglé avant  
le début du procédé. Cet indicateur comporte deux niveaux  
seulement, désignés par "1" et "0". Dans cet exemple, si  
10 l'indicateur est "1", le traitement est interrompu et le  
processeur produit un message à afficher sur l'écran  
vidéo pour indiquer à l'utilisateur qu'aucune  
correspondance n'a été trouvée et que des instructions  
manuelles doivent être introduites avec le clavier pour  
15 continuer l'opération. D'autre part, si l'état de  
l'indicateur est "0", le processeur inscrit la zone dans  
un fichier d'impression et continue ensuite le traitement  
du mot suivant de la zone. Ultérieurement, le fichier  
d'impression est utilisé pour imprimer une liste de zones  
20 à éditer manuellement.

A l'étape 16, on vérifie s'il existe ou non  
un autre mot dans la zone, soit après les étapes 10 (si  
le mot précédent est correct), après l'étape 11 (s'il  
est incorrect) ou après l'étape 15. S'il existe un autre  
25 mot dans la zone qui a été introduit dans le circuit  
de mémoire, les étapes 6 à 16 sont répétées pour ce mot.  
Si tous les mots dans une zone ont été vérifiés, le  
processeur vérifie certains paramètres de la zone à  
l'étape 17. Dans un exemple, la formule de politesse  
30 est le paramètre qui est vérifié pour servir d'indication  
afin de savoir si la personne dont le nom est dans la  
zone est du genre masculin ou féminin. A l'étape 18,  
on crée une autre zone dans l'enregistrement et, dans  
cet exemple, la zone supplémentaire serait un indicateur  
35 masculin/féminin. Par conséquent, non seulement le procédé

1 implique la correction d'une base de données et la  
modification de la structure pour obtenir un format  
correct mais il implique également des additions à la  
base de données, si bien que celle-ci peut être traitée  
5 plus facilement à une date ultérieure, éventuellement  
pour d'autres applications.

A l'étape 19, le processeur vérifie s'il existe  
un autre enregistrement dans la base de données d'entrée  
et, si c'est le cas, les étapes 4 à 19 sont répétées  
10 pour l'enregistrement suivant.

Lorsque tous les enregistrements ont été traités,  
le processeur vérifie, à l'étape 20, les instructions de  
l'utilisateur pour savoir si un autre passage est requis  
ou non, c'est-à-dire si une autre zone dans les  
15 enregistrements dans la base de données d'entrée doit  
être traitée ou non. Par exemple, un passage peut  
concerner la correction de la zone de noms dans une base  
de données d'entrée tandis qu'un autre passage visera  
à corriger la zone d'adresses. S'il existe un autre  
20 passage, les étapes 4 à 20 sont répétées pour chaque  
enregistrement. Si un autre passage n'est pas requis,  
le procédé se termine à l'étape 21.

On comprendra que la manière de créer des  
ensembles de lexiques, dans lesquels des ordres de  
25 priorité sont attribués à ceux-ci et qui sont utilisés  
par le processeur pour vérifier des éléments de données  
en utilisant les lexiques choisis, permet d'améliorer  
fortement l'efficacité du processeur pour la conversion  
d'une base de données d'entrée. D'autre part, la précision  
30 de la base de données de sortie est considérablement  
améliorée parce que le processeur peut réagir de manière  
intelligente pour chaque élément différent de données  
en utilisant les règles de contrôle incorporées qui  
correspondent à chaque mot de référence dans les lexiques.  
35 D'autre part, des mots de référence qui se présentent

1 rarement ne sont pas inclus dans les lexiques, et ceci  
permet de gagner du temps pour chaque passage de ce  
lexique déterminé. Lorsqu'un mot de ce genre se présente  
dans une base de données d'entrée, le traitement peut  
5 être interrompu à l'étape 14 et une inscription manuelle  
peut être fournie. On a constaté que cette méthode est  
beaucoup plus efficace qu'une tentative pour traiter  
automatiquement toute éventualité possible. On a également  
constaté que la vérification des paramètres de zone et  
10 l'insertion d'une zone supplémentaire dans  
l'enregistrement permettent d'améliorer considérablement  
la base de données de sortie et que celle-ci peut  
facilement être utilisée pour le traitement ultérieur,  
qui peut impliquer la séparation des enregistrements  
15 en différents ensembles en fonction du paramètre choisi.  
Le paramètre choisi peut être l'âge, le district postal,  
le sexe ou tout autre paramètre.

On a constaté qu'il est avantageux de régler  
l'état de l'indicateur de traitement sur "1" pour le  
20 traitement en cours de la journée lorsqu'un utilisateur  
est disponible pour introduire des instructions manuelles  
et sur "0" pour le traitement pendant la nuit, auquel  
cas un utilisateur devrait introduire des instructions  
provenant d'un document qui a été imprimé à l'aide d'un  
25 fichier d'impression. Ceci confère une souplesse  
remarquable au procédé de la présente invention.

L'invention n'est pas limitée aux modes de  
réalisation décrits ci-dessus mais peut subir des  
30 variations de construction et de détail.

1

## R E V E N D I C A T I O N S

5  
10  
1. Procédé de conversion d'une base de données d'entrée ayant des enregistrements avec des zones de données alphanumériques en une base de données de sortie ayant un contenu de données et une structure qui facilite le fonctionnement d'un processeur pour la reprise des données et l'impression de documents en utilisant les données de manière efficace et sans erreurs, le procédé comprenant les étapes suivantes :

15  
enregistrer un ensemble de lexiques dans un circuit de mémoire, l'ensemble de lexiques étant associé à une zone particulière de cet enregistrement de la base de données d'entrée et chaque lexique ayant un ensemble de mots de données de référence et de règles associées;

20  
assigner un ordre de priorité à l'ensemble des lexiques pour chaque position de mots, l'ordre de priorité étant basé sur la probabilité que le lexique contienne un mot de référence pour chaque position de la zone;

25  
inscrire séquentiellement chaque enregistrement de la base de données d'entrée dans le circuit de mémoire et filtrer la zone particulière; et

30  
pour chaque mot de données de chaque zone, effectuer les étapes suivantes :

35  
comparer le mot avec au moins certains des lexiques dans l'ordre de priorité, jusqu'à ce qu'une correspondance apparaisse;

1            lorsqu'une correspondance apparaît, vérifier  
le mot par comparaison des paramètres du mot  
avec les règles du mot de référence dans le  
lexique et inscrire à nouveau le mot dans la  
5            zone suivant les règles; et

si aucune correspondance n'est trouvée, insérer  
un indicateur dans la zone.

10

2. Procédé selon la revendication 1, comprenant  
l'étape supplémentaire de régler à l'avance un indicateur  
de procédé et, selon l'état de l'indicateur, si un indice  
est inséré dans une zone soit d'interrompre le traitement  
15            et de demander l'introduction manuelle d'une instruction  
par l'utilisateur, soit l'inscription de la zone dans  
un fichier d'impression pour imprimer ultérieurement  
une liste d'erreurs.

20

3. Procédé selon la revendication 1 ou 2,  
comprenant les étapes supplémentaires d'enregistrer une  
zone supplémentaire dans un enregistrement en réponse  
à la vérification d'un paramètre d'une zone dans  
l'enregistrement, le paramètre étant vérifié en se  
référant à un mode de référence et à une règle de contrôle  
25            associés à un mot et à une zone.

25

4. Procédé essentiellement tel que décrit ci-  
dessus en se référant à la figure en annexe et tel que  
représenté par celle-ci.

30

