QUID ET SON APPLICATION AU RECENSEMENT DES D.O.M.

Lionel VIGLINO

Le système QUID (abrégé de QUestionnaires d'IDentification) est un système de chiffrement automatique conçu et développé à l'Institut National de la Statistique et des Études Économiques (INSEE), depuis les années 1979-1980, par J. Lorigny et son équipe.

Rappel du problème

Le problème consiste à classer automatiquement un individu enquêté dans un poste défini d'une nomenclature existante (par exemple, la nomenclature des professions). Pour cela, le système utilise principalement la réponse en clair à la question posée directement (par exemple : "Quelle profession ou quel métier exercez-vous actuellement?"), et accessoirement d'autres informations figurant dans le formulaire d'enquête et supposées préalablement codifiées (par exemple, le code "Activité économique de l'entreprise employant l'individu").

Dans notre terminologie, la réponse directe en clair est appelée "intitulé". Les informations codifiées complémentaires sont désignées sous le terme générique de "variables annexes".

Nous présentons dans la prochaine section l'approche de base du système QUID, et donnons des résultats de son application à l'INSEE. Dans la section suivante nous examinons le traitement mis en œuvre pour le RPDOM. La version future de la méthode "QUID2", est évoquée dans la dernière section.

Le principe de la méthode

L'approche de base

L'approche de base du système QUID consiste à élaborer une base de données très importante constituée d'intitulés typiques des répondants, accompagnés du code correspondant attribué par un expert.

Dans notre terminologie, la base de données s'appelle "base d'apprentissage", ou "Fichier d'Apprentissage " (FA).

À partir du fichier d'apprentissage, le système constitue une arborescence de questionnement, en interrogeant les libellés pour identifier un code.

Les libellés sont tout d'abord normalisés dans des zones de caractères fixes partitionnées en zones fixes de deux caractères appelées "bigrammes".

Les bigrammes sont les questions posées par l'arborescence QUID pour identifier un code.

QUID utilise la théorie de l'information pour construire une arborescence de longueur moyenne minimale, afin de réduire au maximum le temps d'identification d'un code.

Par la suite nous utiliserons comme vocabulaire, indifféremment questionnaire d'identification ou arborescence (de questionnement).

Constitution d'un fichier d'apprentissage

Pour construire le FA, nous partons le plus souvent de l'enquête d'une année antérieure, déjà chiffrée manuellement ou par une méthode interactive. Chaque intitulé de la base est accompagné de son code (supposé *a priori* exact), et de sa "fréquence d'occurrence" observée dans le FA, c'est-à-dire du nombre d'individus ayant répondu par cet intitulé.

La constitution d'un fichier d'apprentissage suffisamment complet et juste est une opération longue et coûteuse, mais ce travail est un investissement durable.

La normalisation des libellés

Les intitulés subissent un traitement automatique de normalisation commandé par un jeu de paramètres externes choisis par l'utilisateur.

Les mots sont séparés et cadrés dans des zones fixes dont la longueur (unique pour tous les mots) et le nombre maximum (unique pour tous les intitulés) sont paramétrés.

Les "mots vides", considérés comme sans importance par l'utilisateur sont éliminés. Ces mots éliminés sont souvent des articles ou des prépositions, mais dépendent du langage traité.

L'utilisateur paramètre les séparateurs des mots, la liste des mots vides, le nombre des mots gardés et le nombre de caractères conservés par mot.

Contruction de l'arborescence optimisée

À chaque sommet du questionnaire, le système détermine le bigramme (i.e. la question) qui maximise la réduction d'incertitude sur le code. L'algorithme mis en œuvre par QUID est fondé sur la théorie de l'information et utilise l'entropie de Shannon.

Notons : T = (T1, T2,.....,Tj....Tn) le code et l'ensemble des modalités du code.

Q = (q1, q2,....,qi,....,qm) l'ensemble des bigrammes résultant de la normalisation des intitulés

X = l'arborescence à construire

A un sommet quelconque Xo, posons :

N(Xo) la fréquence d'occurrence totale des intitulés associés au sommet Xo.

N(Xo,j) la fréquence d'occurrence du code Tj parmi les intitulés associés au sommet Xo.

Hypothèse :

En supposant que la population d'apprentissage soit statistiquement représentative de la population à chiffrer, on peut estimer la probabilité de trouver le code Tj dans la population à chiffrer par :

$$P(Tj/Xo)=N(Xo,j)/N(Xo)$$

Choix du bigramme :

Au sommet Xo, l'incertitude sur le code est mesurée par l'entropie de shannon H(T/Xo).

$$H(T/Xo) = \sum_{j} P(Tj/Xo) * LOG(1/P(Tj/Xo))$$

Le bigramme choisi est celui qui maximise la réduction d'incertitude sur le code.

Le critère à maximiser est donc:

$$I(T/qi, Xo) = H(T/Xo) - \sum_{Y \in \Gamma(X_0)} P(Y) * H(T/Y)$$

(Entropie conditionnelle au sommet Xo)

avec Γ (Xo) l'ensemble des sommets Y, successeurs du sommet Xo, définis par le choix de qi.

$$P(Y) = N(Xo, a_{i,k}) / N(Xo)$$

$N(Xo, a_{i,k})$ est la fréquence d'occurrence de la modalité $a_{i,k}$ du bigramme qi parmi les intitulés associés au sommet Xo.

La construction de l'arborescence commence au sommet racine auquel est associé le fichier d'apprentissage en entier. La construction se poursuit jusqu'aux sommets terminaux.

Un sommet est terminal quand l'entropie conditionnelle de tous les bigrammes des libellés associés à ce sommet est nulle.

Il existe deux types de sommets terminaux:

- Les sommets de décision

Un sommet de décision est un sommet terminal pour lequel les libellés associés possèdent tous le même code.

- Les sommets d'indécision

Un sommet d'indécision est un sommet terminal associé à des libellés possédant des codes différents qu'aucun nouveau bigramme ne permet de discriminer.

Chiffrement, contrôle de redondance

Le questionnaire d'identification est utilisé pour déterminer le code associé à un libellé à chiffrer.

Le libellé à chiffrer est normalisé avec les mêmes paramètres que ceux utilisés pour construire le questionnaire d'identification.

Ce libellé est découpé en bigrammes et l'arborescence interroge les bigrammes du libellé pour déterminer un code.

Le contrôle de redondance est effectué lors du chiffrement au niveau d'un sommet terminal ; ce contrôle consiste à vérifier la concordance du contenu de quelques bigrammes qui n'ont pas été interrogés par le questionnaire avec celui des libellés associés au sommet terminal pendant la phase de construction de l'arborescence.

La liste des bigrammes à vérifier est un paramètre à la disposition de l'utilisateur de QUID.

Au moment du chiffrement quatre issues sont possibles :

- **un écho unique non douteux**

L'interrogation des bigrammes du libellé à chiffrer a conduit à un sommet de décision du questionnaire (écho unique) et le contrôle de redondance est satisfaisant.

- **un écho unique douteux**

Le parcours de l'arborescence est arrivé à un sommet de décision mais le contrôle de redondance signale au moins une différence parmi les bigrammes contrôlés entre le libellé à chiffrer et tous les libellés associés au sommet de décision.

- **un écho multiple**

Le libellé à chiffrer est associé à un sommet d'indécision.

- **une réponse inconnue**

Le contenu d'un bigramme interrogé par l'arborescence n'est pas connu.

ENCADRÉ

Exemple pédagogique QUID

Afin d'illustrer le mécanisme de QUID considérons un cas d'école de chiffrement de la catégorie professionnelle:

1re phase : apprentissage

FICHIER D'APPRENTISSAGE BRUT

64	CHAUFFEUR DE TAXI
64	CHAUFFEUR DE TAXI
31	DOCTEUR
31	DOCTEUR EN MÉDECINE
52	FACTEUR
54	EMPLOYÉ
52	EMPLOYÉ
34	CHIRURGIEN
31	CHIRURGIEN DENTISTE

PARAMÈTRES DE NORMALISATION

SÉPARATEURS DES MOTS : " "

MOTS VIDES : "EN", "DE"

NOMBRE DE MOTS CONSERVÉS : 2

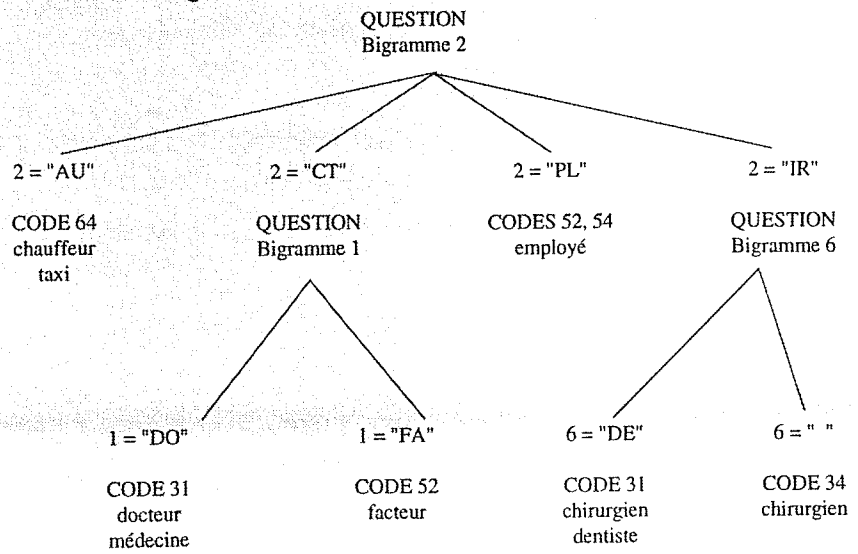
NOMBRE DE CARACTÈRES CONSERVÉS PAR MOT : 10

FICHIER D'APPRENTISSAGE NORMALISÉ :

CODE	LIBELLÉ	FRÉQUENCE
64	CHAUFFEUR TAXI	2
31	DOCTEUR	1
31	DOCTEUR MÉDECINE	1
52	FACTEUR	1
54	EMPLOYÉ	1
34	CHIRURGIEN	1
31	|CH|IR|UR|GI|EN| |DE|NT|IS|TE|	1

1 2 3 4 5 6 7 8 9 10
Bigramme

QUESTIONNAIRE D'IDENTIFICATION :



PENDANT LA CONSTRUCTION DE L'ARBORESCENCE

2 = "CT" EST UN SOMMET DE QUESTIONNEMENT

LE PROCESSUS S'ARRÊTE

QUAND L'ENTROPIE CONDITIONNELLE EST NULLE.

2 = "CT" ; 1 = "DO" EST UN SOMMET DE DÉCISION

2 = "PL" EST UN SOMMET D'INDÉCISION

2e phase : le chiffrement

AU MOMENT DU CHIFFREMENT

"CHAUFFEUR DE TAXI" DONNE UN ÉCHO UNIQUE NON DOUTEUX

"CHAUFFEUR ROUTIER" DONNE UN ÉCHO UNIQUE DOUTEUX

Le caractère douteux dépend du contrôle de redondance

"EMPLOYEE" CONDUIT A UN ÉCHO MULTIPLE 2 codes sont proposés: 52 et 54

"AGRICULTEUR" CONDUIT À UN INCONNU

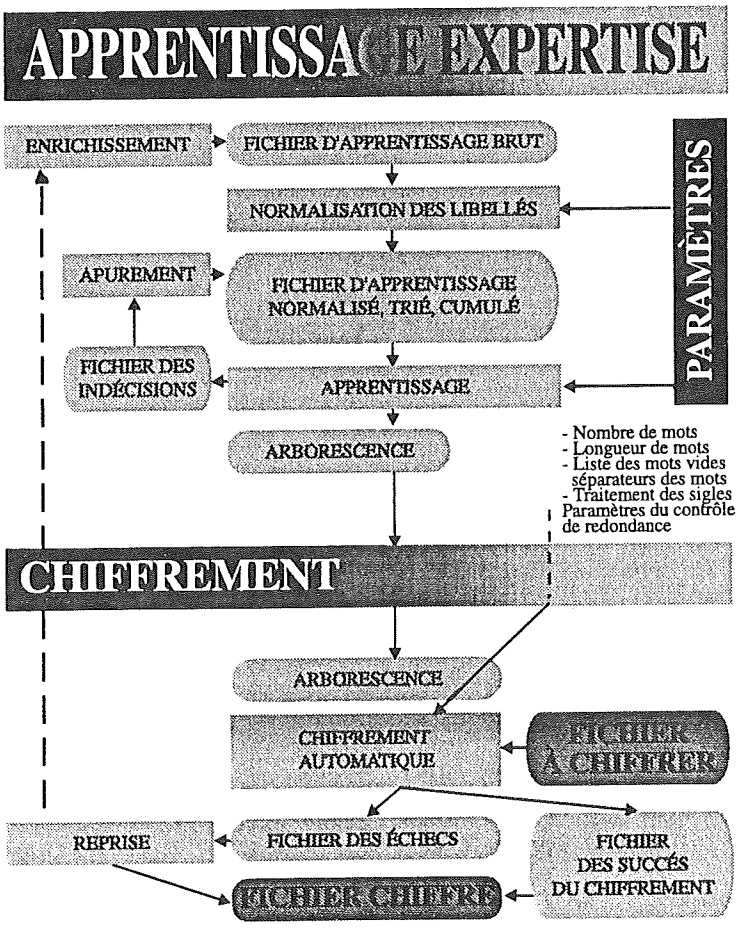
La réponse 2=RI n'est pas connue du questionnaire

LES SUCCÈS DU CHIFFREMENT AUTOMATIQUE

SONT LES ÉCHOS UNIQUES NON DOUTEUX.

Mise en œuvre du système QUID

Le logiciel QUID est une bibliothèque de programmes généraux. La mise en oeuvre du système sépare deux phases : la phase d'APPRENTISSAGE de celle du CHIFFREMENT proprement dit.
L'apprentissage utilise des programmes écrits en PL1 (IBM). Les programmes de chiffrement sont écrits en deux versions: l'une en PL1, l'autre en Cobol.
QUID est une méthode générale qui peut être utilisée sur des nomenclatures différentes avec des fichiers d'apprentissage et des paramètres de normalisation appropriés.
QUID est un système expert car, d'une part il reproduit un chiffrement d'expertise, d'autre part les connaissances (FA) sont séparées des programmes de mise en oeuvre (logiciel QUID).



La justesse du chiffrement

L'expertise du fichier d'apprentissage

À sa création, le fichier d'apprentissage n'a que la justesse de la méthode utilisée pour coder les exemples qu'il contient. Mais à chaque apurement les améliorations apportées au FA restent pérennes et la justesse de la base d'exemples augmente progressivement jusqu'à une qualité d'expertise.

La justesse du chiffrement risque d'être dégradée par la faculté de QUID d'induire une codification à la seule vue des bigrammes interrogés par l'arborescence. Cette faculté est bridée par le contrôle de redondance. Cependant l'élargissement du contrôle de redondance diminue la proportion d'échos uniques non douteux en augmentant le taux de justesse de ceux-ci.

Le paramétrage du contrôle de redondance doit donc tenir compte de l'objectif visé : soit accepter une qualité moindre en diminuant les rejets et en conséquence les coûts de traitement, ou au contraire augmenter la précision, au risque d'élever le taux de rejets.

Les mesures de justesse

Pour le moment, peu de mesures du taux de justesse ont été faites et l'on en reste à l'impression très favorable des responsables d'enquête sur la qualité obtenue par ce mode de chiffrement.

Deux tests effectués, l'un en 1985 l'autre en 1990, ont donné des taux d'erreur de 3 à 4 fois moindres que ceux obtenus avec un chiffrement manuel de qualité courante.

L'utilisation de variables annexes

Pour un certain nombre de nomenclatures, le chiffrement de libellés peut être précisé par des variables codées appelées variables annexes, obtenues à partir de questions fermées dans le questionnaire d'enquête.

Par exemple, pour chiffrer la catégorie socioprofessionnelle à partir du bulletin du recensement, une dizaine de variables annexes doivent être considérées (activité, statut salarié ou non, taille de l'entreprise, qualification, fonction, secteur public ou privé, spécialité de l'exploitation agricole, etc). Le croisement des modalités des variables présentées ci-dessus produit environ 91 000 possibilités.

Pour traiter ces variables annexes, la solution qui consiste à les concaténer à la fin du libellé et à les structurer en bigrammes n'est pas satisfaisante dès que les modalités de ces variables sont nombreuses.

Dans le cas (réaliste) où plusieurs variables annexes apportent la même quantité maximale d'information sur le code, la convention adoptée par QUID est d'interroger la première dans l'ordre de déclaration. Le problème qui se pose, quand les variables annexes n'ont pas toutes la même utilité sur tout le champ de la nomenclature, est qu'il n'est pas possible de trouver *a priori* un bon ordre de déclaration.

QUID risque alors d'identifier un code avec une variable annexe sans signification pour la nomenclature.

Ce problème est lié à l'incomplétude du fichier d'apprentissage vis-à-vis des variables annexes; il pourrait être résolu en multipliant le fichier d'apprentissage pour chaque appellation de profession par l'ensemble des combinaisons des variables annexes. Mais le nombre trop grand de modalités de l'espace croisé des variables annexes rend cette solution physiquement impossible.

Pour ces raisons, dès que le chiffrement d'un code doit tenir compte de plusieurs variables annexes, une solution à deux niveaux a été retenue :

- le premier niveau est une identification par QUID d'un code qui est soit le code définitif, soit un pointeur appelant une table pour lever une ambiguïté sur la codification ;

- le deuxième niveau est un ensemble de tables de décision pour traiter les variables annexes et achever le chiffrement.

Cette solution à deux niveaux s'est imposée pour le traitement de la CS du RPDOM qui utilise 90 tables indépendantes.

Des économies d'échelle

Des économies d'échelle sont déjà réalisées grâce à l'utilisation sur plusieurs enquêtes de fichiers d'apprentissage peu différents.

Il est rentable d'investir dans la construction d'un fichier d'apprentissage, pour exploiter une enquête périodique, ou pour chiffrer un code utilisé par plusieurs enquêtes qui respectent une normalisation minimale de leur questionnaire.

Il est aussi possible de réutiliser un fichier d'apprentissage en sachant *a priori* que le langage utilisé par les répondants à l'enquête sera un peu différent de celui des libellés du fichier d'apprentissage.

Par exemple, le chiffrement de la catégorie socioprofessionnelle du recensement de la population dans les départements d'Outre-Mer utilise un fichier d'apprentissage provenant d'une enquête en métropole dont les questionnaires ont été remplis par des enquêteurs de l'INSEE alors que les bulletins individuels du recensement dans les DOM sont remplis par les enquêtés.

Dans ce cas il est utile d'enrichir le fichier d'apprentissage juste après le début de l'exploitation par les libellés les plus fréquemment non codés (cf. page 155)

Principaux résultats de l'utilisation de QUID à l'INSEE (avril 1991)

QUID est actuellement exploité en production sur 3 grandes enquêtes de l'INSEE : les déclarations annuelles de données sociales (DADS), l'enquête emploi, le recensement dans les départements d'outre-mer (RPDOM)

DADS : Déclarations Annuelles de Données Sociales

Nomenclature	Renseignements traités	Taux de codification	Taux d'erreur
Code commune	Libellé de commune code postal	96 %	Très faible
CS : Catégorie socioprofessionnelle 42 postes	Libellé de profession code activité (APE)	86 %	Inconnu

Les DADS représentent un traitement de 900 000 déclarations par an, l'objectif de QUID était de contribuer à résorber les retards accumulés dans l'exploitation en automatisant les chiffrements.

Le fichier d'apprentissage de la catégorie socioprofessionnelle contient 110 000 libellés, le questionnaire d'identification comporte 100 000 sommets.

Enquête emploi

Nomenclature	Renseignements traités	Taux de codification	Taux d'erreur estimé
PCS : Profession 455 postes	Libellé de profession Libellé de grade Variables annexes	68 %	Faible

L'objectif de QUID est d'augmenter la qualité des chiffrements. Dans ce but le fichier d'apprentissage de la profession a été soigneusement expertisé et un contrôle de redondance maximal a été adopté.

Recensement dans les Départements d'Outre-Mer

Nomenclature	Renseignements traités	Taux de codification	Taux d'erreur
CS	Libellé de grade Variables annexes	77,9%	4,6% *
CS antérieure	Libellé de profession	89,4%	Inconnu
Code commune	Libellé de commune Code postal	98,3%	Très faible
Code APE Activité del'établissement	Libellé d'activité Libellé de raison sociale	68,7%	Inconnu
Code nationalité Code pays	Libellé de nationalité ou Libellé de pays	98,3%	Quasi nul

* Taux estimé sur le test de recensement de 1988.
(en comparaison, le chiffrement manuel a fait 17,6% d'erreur sur ce test, mais l'importance de ce taux doit être modérée car les agents n'avaient pas été spécialement reformés avant le chiffrement).

Le recensement de la population dans les départements d'Outre-Mer traite 1 450 000 bulletins individuels. Les bulletins sont saisis par des façonniers, ensuite 86 % des bulletins sont entièrement traités automatiquement par les différents algorithmes mis en oeuvre.

Le chiffrement de la CS met en oeuvre une arborescence de 28 600 sommets dont 22 630 sommets de décision.

Le chiffrement d'un code CS nécessite en moyenne l'interrogation de 2,5 bigrammes et un temps de traitement de 0,37 ms CPU sur un IBM 390-200-E(31 MIPS).

Le chiffrement de la CS du recensement de la population dans les départements d'outre-mer.

Le traitement de la CS du RPDOM

La prise en compte, en plus du libellé de profession, du libellé de grade et des variables annexes présentes dans le bulletin individuel (BI) a nécessité de compléter la méthode générale de chiffrement par des programmes aidant QUID à trouver et à traiter les informations nécessaires pour aboutir à un chiffrement.

Le libellé découpé en bigrammes et traité par QUID résulte d'un traitement préalable qui concatène au libellé de profession le libellé de grade et une variable de synthèse. Cette variable de synthèse, à 10 modalités, est calculée à partir des variables annexes du bulletin et représente les grandes divisions en espaces sociaux de la nomenclature des CS .

Parfois la pauvreté des renseignements collectés sur le BI ne permet pas d'établir cette variable de synthèse. Deux QUID sont donc utilisés :

- le premier appelé QUID-A traite les libellés pour lesquels une variable de synthèse a pu être calculée ;

- le deuxième QUID-B est chargé des libellés sans variable de synthèse.

QUID-A

QUID-A donne une réponse qui est soit le code CS définitif, soit un pointeur appelant une table pour achever la codification. En effet dans certains cas le libellé n'est pas assez complet pour déterminer précisément un code, il faut alors interroger des variables annexes du BI pour lever l'ambiguïté sur le code.

Le fichier d'apprentissage provient d'une enquête sur les qualifications professionnelles effectuée en 1985. Ce fichier a été enrichi et apuré pour les besoins de l'Enquête Emploi. Le code profession d'origine a été transformé en code de CS. Le fichier a été enrichi des principales appellations de métiers spécifiques aux départements d'outre-mer.

Les libellés de professions sont normalisés à 5 mots de 10 caractères, soit 25 bigrammes ; la variable de synthèse est ajoutée en 26^{e} bigramme. Une étude menée en 1989

a permis de réduire, sans perte de qualité du chiffrement, le contrôle de redondance à la variable de synthèse et aux 3 premiers bigrammes des deux premiers mots. En raison de cette réduction du contrôle de redondance, les échos douteux qui restent sont faux dans 2 cas sur 3. Les échos uniques douteux ne sont donc pas retenus et seuls les échos uniques non douteux sont automatiquement acceptés pour le chiffrement.

Les tables de QUID-A

Les tables associées à QUID-A ont pour rôle de lever une ambiguïté possible sur le code définitif. Ces tables sont mises au point au moment de l'expertise du fichier d'apprentissage avant la phase de chiffrement. Les tables n'ont pas été construites systématiquement à partir de la nomenclature mais ont été créées en examinant les sommets d'indécision de l'arborescence. Ces sommets d'indécision provenaient de libellés pauvres en information conduisant à deux codes possibles lors de la constitution du fichier d'apprentissage.

Cette méthode de construction a l'avantage de faire construire en premier les tables les plus fréquemment utilisées. Elle permet d'obtenir le meilleur rapport qualité/coût pour la constitution des tables.

Il a par ailleurs été décidé de construire des tables indépendantes les unes des autres. D'une façon plus imagée les tables ne s'appellent pas entre elles au moment de leur exécution. Cette décision a pour but de faciliter la maintenance du fichier. En effet une modification dans une table reste limitée à cette table et n'a pas de répercussions en cascade sur les spécifications de traitement des autres libellés ; le coût de maintenance du FA en est ainsi réduit.

Au total 90 tables ont ainsi été implantées. Elles constituent le prolongement logique de l'arborescence de questionnement du QUID ; mais si QUID-A choisit les bigrammes à interroger dans le libellé pour optimiser la recherche, la solution retenue ici laisse aux experts humains le choix de déterminer le questionnement des variables annexes.

QUID-B

Le second QUID, appelé QUID-B s'occupe des cas où aucune modalité de la variable de synthèse n'a pu être déterminée. Le chiffrement de QUID-B tient compte de l'absence d'information dans le BI qui permettrait de calculer une variable de synthèse. Le code attribué au libellé correspond à celui prévu par la nomenclature dans ce cas. Nous avons vérifié cette option par un test qui a montré que l'application de règles par défaut apporte des résultats de meilleure qualité que le chiffrement manuel à partir du document papier. (9% contre 17% d'erreur sur cette population).

Le fichier d'apprentissage de QUID-B a été déduit de celui de QUID-A en supprimant la variable de synthèse et en faisant lever les indécisions ainsi créées par des spécialistes du chiffrement de la CS.

Les mêmes paramètres de normalisation du libellé et de contrôle de redondance ont été adoptés.

Là encore, seuls les échos uniques non douteux sont acceptés pour le chiffrement automatique.

Aucune table n'a été implantée pour QUID-B ; les règles par défaut de la nomenclature qui affectent le code le plus probable ont été systématiquement appliquées.

L'enrichissement des fichiers d'apprentissage en cours d'exploitation

Les possiblités d'apprentissage du système QUID ont été utilisées après le début de l'exploitation du recensement pour enrichir les fichiers d'apprentissage de QUID-A et de QUID-B.

Les rejets des premiers lots exploités ont été triés par fréquence décroissante ; les 3% des libellés les plus fréquents représentant 28% des cas de rejet ont été expertisés puis inclus dans les fichiers d'apprentissage.

Cet enrichissement a permis d'augmenter le taux de chiffrement de 68% à 78% sans interrompre significativement l'exploitation.

La reprise des échecs des codifications automatiques

Les rejets de la codification automatique par QUID-A et QUID-B sont traités manuellement avec les rejets des autres procédures automatiques, sur micro-ordinateur, dans chaque département d'origine.

Le traitement de ces rejets est effectué grâce à une application spécialement conçue pour le RPDOM.

Les rejets concernent 22% des 403 000 actifs pour la CS, et 14% des 1 450 000 BI collectés pour toutes les codifications automatiques.

L'évolution future du système QUID

L'évolution future du système QUID dans un logiciel appelé QUID2 prendra en compte l'existence des tables au niveau du fichier d'apprentissage, le système futur ne devra plus gérer le fichier d'apprentissage libellé par libellé mais le gérer par groupes de libellés appelant la même table. De nouveaux outils d'apurement et d'enrichissement tenant compte de la gestion des tables seront implantés.

Les services visés sont l'extension des fonctions horizontales de chiffrement pour permettre des économies d'échelle en mettant des modules de chiffrement à la disposition des utilisateurs respectant un minimum de normalisation de leur questionnaire d'enquête .

La gestion des fichiers d'apprentissage associés serait alors confiée à des experts des nomenclatures. Les fichiers d'apprentissage pourraient alors devenir des références pour l'utilisation des nomenclatures au même titre que les documents papier spécifiant les consignes de chiffrement.

BIBLIOGRAPHIE

SHANNON, C.E. (1948). "A Mathematical Theory of Communication". *Bell Systems Technical Journal*, 27, 379-423, 623-656.

PICARD, C.-F. (1972). *Graphes et questionnaires*. Paris : Gauthier-Villars.

BOUCHON-MEUNIER, B. (1978). *Sur la réalisation de questionnaires*. Thèse d'Etat, Paris.

LORIGNY, J. (1982). Mesures d'entropie et d'information pour les systèmes ouverts complexes. Thèse d'État, Paris.

LORIGNY, J. (1988). *QUID, une méthode générale de chiffrement automatique*. Technique d'enquête, vol. 10, n° 2, p. 307-316 Statistique Canada.

VIGLINO, L. (1991). *Communication sur le chiffrement automatique*. Institut international de la statistique.