



- 2.2 De plus, les mots-clés sont hiérarchisés, ainsi "enquête" contient "enquête longitudinale", "enquête rétrospective",... mais est contenu dans "rassemblement des données". En outre, il y a des renvois à des termes voisins, ces renvois sont très subjectifs aux auteurs du thésaurus.

Exemple :

ERREUR DE SONDAGE

UF : ERREUR D'ECHANTILLONNAGE

BT : ERREUR

RT : SONDAGE

ERREUR D'ECHANTILLONNAGE

USE : ERREUR DE SONDAGE

- 2.3 USE : suit un mot interdit et renvoi au mot obligatoire  
 UF : inversement, introduit un synonyme interdit  
 BT : est un terme générique plus large (BROADER TERM)  
 NT : est un terme spécifique plus étroit (NARROWER TERM)  
 RT : renvoi à un terme relié  
 SN : introduit une définition

Ce sont les six abréviations utilisées dans le thésaurus multilingue de population édité par le CICRED, mais ces six fonctions se retrouvent dans tous les thésaurus avec bien entendu des abréviations différentes.

- 2.4 Ainsi, le thésaurus contient une liste thématique, une liste alphabétique, une liste hiérarchique. Un terme spécifique peut être relié à plusieurs termes génériques, c'est une poly-hiérarchie.
- 2.5 Un thésaurus est spécifique à un domaine de documentation, cependant, il est possible d'englober les mots-clés communs à plusieurs disciplines dans un macro-thésaurus, chaque discipline conservant en plus un micro-thésaurus spécifique.  
 Un exemple de macro-thésaurus est le thésaurus géographique, où le mot-clé "Afrique Noire" renvoie à tous les pays qui constituent l'ensemble une fois que l'accord s'est fait sur les pays relevant de l'Afrique Noire. Ce thésaurus géographique peut être utilisé par toutes les disciplines.
- 2.6 La constitution d'un thésaurus requiert l'intervention de documentalistes et de spécialistes du domaine défini, une fois établi, les relations entre termes sont vérifiées par informatique.

### 3 L'industrie de l'information automatique

#### 3.1 Définitions

Un centre producteur

constitue une ou plusieurs base de données bibliographiques.

Un centre distributeur

rend diverses bases de données accessibles par un logiciel.

#### 3.2 Un exemple de centre producteur

- 3.2.1 CDSH (Centre de documentation en Sciences Humaines du CNRS) produit 17 bases appelées FRANCIS comprenant notamment :

- sociologie
- géographie
- ethnologie
- économie de la santé (RESHUS)

Les lacunes concernent le droit, l'économie, l'histoire, les sciences politiques.

3.2.2.1. Le CDSH est aussi un centre distributeur avec les logiciels SPLEEN2 (interrogation en différé) et SPLEEN3 (interrogation en conversationnel). Voir la note sur le système CNRS.

3.2.1.2. Chaque base est alimentée par environ deux documentalistes spécialisés dans le domaine correspondant. Cela est nettement insuffisant et certaines bases accusent un retard d'un à deux ans. Face à cette situation, des responsables de bases se contentent des résumés produits avec chaque article, si la revue ne livre pas de résumés comme souvent les cahiers Sciences Humaines de l'ORSTOM, il n'en est pas fait.

3.2.1.3. Normalement, le CDSH doit pouvoir diffuser les articles référencés, au début le microfichage était systématique, cela a été abandonné. Toutes les bases sont bonnes pour ce qui concerne les revues, pour les ouvrages, seules les bases de géographie, sociologie et ethnologie sont bonnes.

3.2.1.4. Il faut noter que le CDSH a deux fonctions :

- produire un bulletin bibliographique
- constituer une base de données.

Ces deux productions sont faites simultanément avec le bordereau harmonie ; cela conduit pour une meilleure présentation du bulletin bibliographique à introduire des mots-clés inutiles ; cette présentation nécessite de plus une opération de chaînage des mots-clés. Cette dernière opération est rendu nécessaire par suite de l'absence d'un thésaurus qui pourrait automatiser le chaînage. Un exemple, le bulletin prévoit une référence à mortalité, un article sur la surmortalité masculine comprendra les mots-clés mortalité, surmortalité masculine afin d'être édité ainsi.

- mortalité
- .....
- surmortalité masculine
- .....

Un thésaurus ferait automatiquement figurer tous les articles indexés par des termes plus spécifiques à mortalité avec le terme spécifique.

3.2.2 Un autre exemple est le système britannique BLAISE qui dépend de la British library et qui répertorie une quantité importante d'ouvrages parus en Europe; il n'y a ni résumé, ni mots-clés, mais la cote de classement dans plusieurs systèmes de classement.

3.2.3 Un autre exemple est constitué par Sociological Research qui entre les références de l'article et les titres et références des articles cités, cela évite le recours à une main-d'oeuvre qualifiée et permet une saisie très rapide. La recherche s'effectue sur les mots constituant le titre de l'article et/ou les titres des articles cités. Cela permet également à partir d'un article de trouver des articles postérieurs qui le citent et peuvent constituer la mise à jour d'une bibliographie.

3.3 Les centres distributeurs

3.3.1 En France, TELESYSTEMES, (40 rue du cherche midi, Paris 6e), c'est une filiale des P et T qui dispose des bases du CNRS (FRANCIS pour les Sciences Humaines, PASCAL pour les Sciences dites exactes) et de nombreuses autres bases dont RESEDA (Réseau de documentation socio-économique en agriculture avec l'INRA, et l'Institut Agronomique Méditerranéen). D'autres bases concernent la chimie, l'industrie textile, le cancer, les industries agro-alimentaires, l'urbanisme, ... La plupart de ces bases ont donné l'exclusivité à TELESYSTEMES.

Le logiciel utilisé s'appelle MISTRAL, le langage semble proche de celui de SPLEEN3 avec des possibilités supplémentaires. TELESYSTEMES donne accès à des documents iconographiques : photographies et coupures de presse. Le logiciel MISTRAL peut être utilisé pour gérer un fichier privé propre à un utilisateur.

Le coût est de 300 à 350 F/heure pour l'utilisation d'une base de donnée sauf pour un fichier privé, 160 F/heure. Cela nécessite un numéro d'abonnement (incluant un code secret) fournit gratuitement et la disposition d'un terminal.

3.3.2 Aux Etats-Unis, LOOCKHEED est le plus important centre distributeur ; il gère 87 bases de données dont 30 bases intéressant les sciences sociales LOOCKHEED utilise le logiciel DIALOGUE.

(Voir description des bases bibliographiques en sciences sociales de Lockheed informatic system interrogeable au CDSH, 1978, 136 p, 12 F).

#### 4 Le système du CNRS

4.1 La saisie s'effectue avec le bordereau appelé "harmonie". Ce bordereau comprend les zones suivantes :

- références de la revue
  - titre de l'article, (éventuellement titré translittéré en caractères latins pour les langues n'utilisant pas notre alphabet)
  - titre traduit (sauf anglais, espagnol, italien) ou titre construit, il s'agit alors d'un titre plus explicite que le titre de l'auteur
  - les auteurs avec les initiales des prénoms, des traducteurs, préfaciers, ...
  - la pagination
  - les compléments (figures, cartes, photos, ...)
  - le nombre de pages de références bibliographiques
  - la localisation où le document est disponible
  - la langue du texte et celle des résumés
  - le type de l'article : statistique, carte, ...
  - le résumé
  - la cote du plan de classement
  - les mots-clés
  - le chaînage des mots-clés pour l'édition du bulletin bibliographique.
- Le bordereau est saisi et transformé dans le format ABACUS par le logiciel SPLEEN qui se charge aussi de l'édition des bulletins bibliographiques qui incluent les index auteurs, pays, thèmes et donnent les références selon le plan de classement.

4.2 Le logiciel de traitement en différé : SPLEEN 2

On dispose de 14 types de recherche (I) :

- TYPE 1 : MOTS CLES
- TYPE 2 : CODES DE LANGUES
- TYPE 3 : AUTEURS, AUTEURS OUVRAGES RECENSES, ORGANISMES
- TYPE 4 : REFERENCES BIBLIOGRAPHIQUES
- TYPE 5 : ANALYSE
- TYPE 6 : CODE D'ORIENTATION
- TYPE 7 : TITRE
- TYPE 8 : TITRE, MOTS CLES, ANALYSE
- TYPE 9 : TITRE, REFERENCES BIBLIOGRAPHIQUES
- TYPE 10 : TITRE, MOTS CLES, ANALYSE, REFERENCES BIBLIOGRAPHIQUES
- TYPE 11 : TITRE, MOTS CLES
- TYPE 12 : MOTS CLES (chaîne par chaîne)
- TYPE 13 : CODE D'ORIENTATION, ZONE TRI, AUTEUR COMPLEMENTAIRE
- TYPE 14 : N° signalement du Bulletin.

(I) On m'a assuré qu'il était possible de faire des recherches sur la période concernée par un article (et non pas la date de parution). Seul un test concret permettra d'en tester la possibilité, cela est pourtant essentiel pour les variables démographiques dont l'évolution temporelle est importante.

Tous les types portant sur des mots se font caractère par caractère sauf le type I2.

Exemple (ou b indique un blanc)

/b MIGRATb/ conduit à sélectionner tous les mots contenant les 6 lettres : immigration, émigration, migration, migratoire, ...  
/EMIGRb/ sélectionne les mots commençant par les 5 lettres :  
émigration, émigrant  
/MIGRATION/ sélectionne le mot migration, mais inclu tous les mots-clés contenant le mot migration : migration collective, de travail, ...  
En type I2, seul le mot-clé migration est retenu, à l'exclusion des autres mots-clés.

Exemple

A I2/ENQUETE RETROSPECTIVE/  
A I2/ENQUETE LONGITUDINALE/  
B I /MORTALITE/  
C I2/AFRIQUE NOIRE/  
D 2/ENG/  
LO = A \* B \* C \* 7D

On définit le groupe A : enquête longitudinale ou enquête à passages répétés, le groupe B : mortalité avec tous les mots clés l'incluant, le groupe C : Afrique noire, le groupe D : l'anglais, la logique LO sélectionne les références qui satisfont aux quatre critères :

la mortalité dans les enquêtes rétrospectives et longitudinales en Afrique noire, références en anglais exclues

les opérations sont :

\* = et

+ = ou

7 = non

( ) = parenthèses

ainsi on peut avoir  $LO = 7D * C * B * A$  ou une formule plus complexe du type  $LO = 7E * D * (A + (B * C))$ .

C'est un langage très imple à apprendre et à manier.

Une option de pondération permet de classer les références par exemple toutes les enquêtes rétrospectives avant toutes les enquêtes longitudinales. Le logiciel travaille sur la bande de donnée et l'explore du début à la fin pour sélectionner les références qui satisfont aux critères choisis. On ne sait donc pas combien de références on va obtenir.

#### 4.3 Le logiciel de traitement en conversationnel SPLEEN3

En plus du fichier des références, il dispose d'une table donnant pour chaque mot les numéros de références utilisés. Cette table s'appelle fichier inverse.

Exemple

- EXb. SPLEEN3

- Qb1,C = DEMOGRAPHIE

C signifie que l'interrogation porte sur les mots-clés  
le système répond 158 documents

- Qb2,C = METHOD +

+ signifie tous les mots qui commence par METHOD : Methode, methodologie, ...

le système répond qu'il y a 42 mots et 1243 documents.

On peut obtenir la liste des mots et sélectionner ceux qui nous intéressent.

Eb3 = 1 & 2

On croise les critères de sélection 1 et 2, le système répond qu'il y a 25 documents

Lb3

permet de lister des documents sur l'écran

Bb3

permet de lister les documents sur l'imprimante avec une présentation sous forme de fiche

On a donc la possibilité d'affiner la question ou de l'élargir en fonction des réponses. Pour lister les documents, le programme va les chercher en accès direct sur le disque d'après le numéro de référence. Cela suppose donc de mettre toutes les références sur disques. C'est de cette façon que fonctionnent les centres distributeurs, ainsi l'ordinateur IRIS 80 de télé systèmes dispose de 6 unités de bandes magnétiques et 64 unités de disques de 200 M octets (millions d'octets).

#### 4.4 Les coûts

4.4.1 Le coût de production d'une référence est de 100 à 200 F, cela inclut notamment les coûts de saisie (incluant les corrections) qui sont de 25 F pour 1000 signes (un bordereau harmonie comporte en moyenne 600 signes soit 15 F).

Il est admis que le coût de constitution d'une base de données n'est pas payé par l'utilisateur.

4.4.2 Pour les bulletins, le coût de production est de 100 à 200 F/page pour 1000 exemplaires (réalisation en imprimerie à partir de la bande magnétique).

Pour 200 exemplaires, l'offset de liste coûte environ 20 F la page, mais il n'y a plus que 3 ou 4 analyses par page au lieu d'une vingtaine pour un bulletin imprimé. L'abonnement au bulletin signalitique varie entre 100 et 200 F pour une base.

4.4.3 Le coût du traitement informatique est de 6 F par référence et 10 F pour l'imprimerie. Le reformatage d'une base de données coûte environ 2000 F par an. Pour le conversationnel, la location de deux disques de 200 millions d'octets revient à 200 000 F, somme partagée entre les utilisateurs ce qui conduit à facturer le prix en conversationnel à 200 F/heure (incluant 30 F/heure pour TSO) (+ 0,25 F par référence pour l'édition en différée). Pour les bases US le coût est de 200 à 400 F/heure et la référence revient entre 0,25 et 0,40 F selon les bases.

Une recherche d'un quart d'heure avec l'édition de 200 références est possible entre 100 F et 180 F

4.4.4 La constitution d'un bordereau de type harmonie revient à 6 mois d'ingénieur et la programmation à 2 ans environ.

### 5 La démographie dans les bases bibliographiques

#### 5.1 FRANCIS au CDSH

5.1.1 La base sociologie comporte 3 rubriques qui peuvent nous intéresser :  
 - Ecologie humaine et démographie comprenant la sociologie des migrations  
 - Sociologie de la famille  
 - Sociologie de la médecine

5.1.2 La base ethnologie comporte la rubrique bases morphologique <sup>qui</sup> comprend la démographie et les migrations et la rubrique structures et relations sociales qui comprend la sous-rubrique famille.

5.1.3 La base géographie inclut une rubrique géographie humaine

5.1.4 La base RESHUS (réseau d'information en sciences humaines de la santé) inclut les rubriques :

- 210 Natalité
- 220 Mortalité
- 230 Maladie (morbidity, *pléaux sociaux, ...*)

- 5.2 DOCPAL sous l'égide du CELADE système de documentation latino-américain utilisé avec le logiciel ISIS
- 5.3 POPLINE est la mise sur ordinateur des références de population index depuis 1969. utilisé avec le logiciel ISIS (le nouvel index est paru dans population index d'Avril 76).
- 5.4 PIDSA (Population Information Documentation system for Africa) localisé à Accra
- 5.5 POPINFORM
- 5.6 POPIN (réseau d'information sur la population) qui est un projet de l'ONU devant fonctionner en anglais, français et espagnol
- 5.7 CARDAN Le bulletin du cardan qui est une liste de titres référencés dans un plan de classement avec un index couvre l'Afrique Noire, mais n'est pas informatisé.

6 Les possibilités pour une bibliographie en démographie africaine

- 6.1 1ère solution est de se contenter d'interroger les fichiers existants
- 6.2 2ème solution est d'intervenir dans la constitution d'un fichier pour y faire entrer nos centres d'intérêts, ce peut être la base de sociologie du CDSH, le réseau documentaire ORSTOM-GERDAT, la base PIDSA d'Accra, POPIN de l'ONU, POPLINE de population index.
- 6.3 3ème solution est de constituer une base de données séparée, mais en liaison avec une ou plusieurs des bases mentionnées (il est possible d'acheter dans ces bases ce qui correspond à notre intérêt). Dans ce cas se pose le problème du logiciel plusieurs sous-solutions existent, utiliser celui du CNRS, celui de télé systèmes, ce qui impose de localiser la base dans ce centre, acheter ou créer un logiciel adapté à un mini-ordinateur.
- 6.4 Ces solutions ne sont pas obligatoirement exclusives, l'une de l'autre, on peut participer à une base plus générale et simultanément constituer une base plus spécialisée.