

Application des techniques de banques de données à une thèse d'histoire des mentalités

Muriel VERBEECK-VERHELST

Si l'informatique a fait de grandes conquêtes dans le domaine de la science historique, notamment en histoire économique et sociale et, plus particulièrement, dans le champ de la période contemporaine, l'histoire des mentalités et celle des idées restent insensibles à ses charmes, non moins que rétives à ses méthodes, — si ce n'est, peut-être, celles de la statistique intégrée. Le temps est passé, cependant, où *quantification* traduisait seul ou à peu près, le terme d'*automatisation*. Les traitements se diversifient, et les lecteurs optiques révolutionneront bientôt, là aussi, l'appréhension des sources dans leur littéralité.

Ce n'est pas ici notre propos. Notre recherche, menée en solitaire, avec des ressources limitées, et surtout dans un laps de temps très restreint, ne nous permettait pas d'envisager une heuristique ni une herméneutique informatisées. Notre projet, au risque de passer pour de peu d'ambition, visait à réduire l'informatique au rang d'outil de l'historien et, principalement, d'outil de gestion d'une thèse. Celle-ci, portant sur *La féminité dans le discours religieux*, impliquait une heuristique élargie à différents genres littéraires, de l'exégèse à la lyrique; l'aire chronologique, qui couvrait en gros la période des Temps Modernes, contribuait à prendre en compte d'autres facteurs de diversité; l'ampleur des dépouillements auxquels se prêtait la matière, enfin, acheva d'alourdir la gestion manuelle des informations collectées, en multipliant les possibilités de classements et les duplications qui leur sont corollaires.

Le but que nous nous propositions était de pallier ces inconvénients par l'intégration des données. En devaient résulter leur sécurité, la facilité de mise à jour, la suppression des duplications, l'accélération dans la consultation des fichiers et le transfert des références bibliographiques dans le corps même de la thèse. Le traitement des fichiers par banque de données se révéla d'emblée le mieux approprié à ce programme. Notre choix se porta sur le logiciel dBASE III, dont la souplesse de

programmation et de mise à jour répondait à nos besoins tant présents que futurs, tout en s'implantant sans difficulté sur le matériel PC compatible dont nous disposions.

1. Structuration en domaines et applications

La conception de nos programmes a été précédée d'une analyse de l'existant, d'après laquelle nous avons dressé le catalogue des informations, puis un tableau croisé des tâches-informations. Au terme de cette première approche, deux domaines apparaissaient clairement au sein de la banque de données. Le premier concernait la gestion bibliographique, et se décomposait en deux applications, l'inventaire bibliographique et la recherche bibliographique. Les tâches se structuraient en saisie et mise à jour, et, d'autre part, en consultation interactive, extraction de références, listage sélectif selon critères.

Le second domaine, celui de la gestion documentaire, gérait le contenu textuel de la thèse, c'est-à-dire les citations de textes sources, et comprenait deux applications : l'inventaire des textes et la recherche documentaire. La première de ces applications recouvrait les tâches de saisie, mise à jour et caractérisation. La seconde, la consultation interactive, l'extraction ponctuelle et le listage sélectif selon différents critères. Si la gestion bibliographique rassemblait le plus d'informations, son traitement demeurait d'une grande simplicité. Par contre, le domaine documentaire se montrait d'une plus grande complexité.

2. Conversion des fichiers

Les fichiers usuels de l'historien se prêtent aisément au transfert sur un support informatique. Pour les fichiers bibliographiques (FBS.DBF = Fichier bibliographique des sources; FBA.DBF = Fichier bibliographique des articles; FBT.DBF = Fichier bibliographique des travaux), la séquence des attributs respecte l'ordre du support traditionnel, et seul un identificateur unique vient s'ajouter (NOL = numéro de livre; NOA = numéro d'article)¹. Les index FBS1.NDX et FBS2.NDX permettent un classement alphabétique sur nom d'auteur, concaténé avec l'initiale du prénom, ou sur l'identificateur unique de l'ouvrage. Un simple accès aux fonctions permet d'établir des classements chronologiques (DATE), géographiques (MAISON, LIEU) ou selon le genre littéraire de l'ouvrage (CLASSE).

¹ Les traitements des fichiers travaux et articles étant strictement parallèles à ceux du fichier sources, nous ne les aborderons plus ici. Ajoutons qu'à l'origine, un fichier BIO.DBF, reprenant les informations relatives à l'auteur, était couplé à FBS.DBF. De peu d'usage, il a été abandonné.

Tableau 1 : Affichage des structures de fichiers.

Structure de la base de données: B:fps.dbf
 Nombre total d'enregistrements : 0
 Date de la dernière mise à jour: 06/07/87

Champ	Nom champ	Type	Dim	Dec
1	NOM	Caractère	20	
2	PRENOM	Caractère	3	
3	TITRE	Caractère	222	
4	LIEU	Caractère	15	
5	MAISON	Caractère	25	
6	DATE	Numerique	4	
7	PP	Numerique	4	
8	FORMAT	Caractère	2	
9	BIBLIOTH	Caractère	3	
10	COTE	Caractère	12	
11	NOL	Numerique	4	
12	CLASSE	Caractère	5	
**	Total **		320	

Structure de la base de données: B:dt.dbf
 Nombre total d'enregistrements : 0
 Date de la dernière mise à jour: 06/07/87

Champ	Nom champ	Type	Dim	Dec
1	NOD	Numerique	4	
2	DESC	Caractère	20	
**	Total **		25	

Structure de la base de données: B:inter.dbf
 Nombre total d'enregistrements : 0
 Date de la dernière mise à jour: 06/07/87

Champ	Nom champ	Type	Dim	Dec
1	NOD	Numerique	4	
2	NOT	Caractère	5	
**	Total **		10	

Structure de la base de données: B:ftxt.dbf
 Nombre total d'enregistrements : 0
 Date de la dernière mise à jour: 12/09/91

Champ	Nom champ	Type	Dim	Dec
1	NOT	Caractère	5	
2	NOL	Numerique	4	
3	TXT	Mémo	10	
4	PP	Caractère	7	
5	E	Caractère	1	
6	P	Caractère	7	
**	Total **		35	

Le fichier texte (FTXT.DBF) a quant à lui exigé une plus ample réorganisation. Le fichier matériel comprenait des citations recopiées ou des photocopies, les unes et les autres à chaque fois référencées. Le fichier informatique a dû s'alléger de cette information redondante, qui lui reste accessible par le lien avec FBS.DBF, sur l'attribut-clé (NOL). FTXT.DBF possède un identificateur unique (NOT), le numéro de texte sur lequel table l'index FTXT1.NDX. Une zone mémo (TXT) permet l'introduction du texte. Celle-ci reste facultative. Un champ (E) signale en effet si l'encodage textuel a eu lieu. Sinon, un autre champ (P) permet de référencer le fichier matériel où le texte se trouve conservé.

Le traitement documentaire exige, pour la mise en relation du fichier texte (FTXT.DBF) avec le fichier descripteurs (DT.DBF), un fichier intermédiaire, nommé INTER.DBF, et décomposant les relations n à m entre FTXT.DBF et DT.DBF. Ses deux attributs clés seront les identificateurs uniques de ces deux fichiers, NOT (numéro de texte) et NOD (numéro de descripteur) (index INTER1 et INTER2.NDX). DT.DBF comprend, outre la clé, un seul autre champ, celui du descripteur ou mot-clé (index DT1 et DT2.NDX).

3. Les programmes

Un menu présente à l'utilisateur les programmes bibliographiques (FICHE.PRG = listage; NOM.PRG = interrogation sur nom d'auteur; EXPORT.PRG = transfert du fichier sous traitement de texte (MSWORD) ou documentaires (ENCODE.PRG = encodage et caractérisation des textes; MAJ.PRG = mise à jour de la liste d'autorité; THESAUR.PRG = édition de la liste d'autorités; MOT.PRG = interrogation simple sur descripteur; BOOL.PRG = interrogation booléenne sur descripteurs; SORTDT.PRG = édition des co-descripteurs; SORTXT.PRG = édition des textes)².

Le programme FICHE.PRG permet la consultation séquentielle de la bibliographie ou la consultation partielle, à partir d'une lettre de l'alphabet. NOM.PRG guide quant à lui l'interrogation sur nom d'auteur. L'un et l'autre utilisent l'index FBS1.NDX, sur l'attribut auteur, concaténé avec l'initiale du prénom. Ils présentent, par rapport au simple listage à partir des fonctions du logiciel, l'avantage de la convivialité avec l'utilisateur et de la présentation, sous forme de fiche traditionnelle.

EXPORT.PRG permet le passage des fichiers du logiciel dBASE au traitement de texte MSWORD, retenu pour la mise en forme et l'impression de la thèse.

² Nous ne traitons pas ici le fichier COPCODE.PRG, (encodage et caractérisation des photocopies), répondant à ENCODE.PRG, et qui lui est strictement parallèle, non plus que les différents modules intermédiaires, les fichiers d'impression et de réinitialisation.

Il s'arrête plus particulièrement aux fichiers bibliographiques, qui comprennent plusieurs milliers de références. L'utilisateur indique tout d'abord le fichier à exporter, puis le mode d'exportation : partiellement, en introduisant numéro par numéro les enregistrements à transférer, ou par tranche, en complétant la fourchette qui lui est proposée. Ce dernier mode permet l'exportation de tout le fichier, par introduction d'un 1 au premier terme de la fourchette, et d'un blanc au second. Le transfert effectué, l'utilisateur peut rejoindre, en suivant les instructions du système, le logiciel MSWORD, rester sur dBASE ou revenir au DOS.

Le programme ENCODE.PRG permet la caractérisation des textes et leur saisie sur support informatique. Il présente à l'utilisateur un masque que celui-ci est amené à remplir, puis demande l'introduction, avec contrainte d'existence, d'un descripteur. Après cet encodage, le programme propose soit la poursuite de la caractérisation du même texte, soit le passage à un nouveau, soit de revenir au menu. Le choix de l'utilisateur est signifié dans une variable, que le système interprète.

Tableau 2 : Affichage du programme ENCODE.PRG

```
Record No   384
NOT
NOL
TXT      memo      (Masque proposé à l'utilisateur)
PP
E
P
```

```
Record No   384
NOT          385
NOL          319
TXT      memo
PP          39
E          o
P          480
```

Edit TXT

Elle s'estoit abandonnée pendant quelques temps de sa plus florissante jeunesse aux plaisirs sensuels de l'amour prophane, qui s'empara de son âme, qui établit sa demeure dans ses sentimens, et prit plaisir d'exercer son empire sur ses passions.

Descripteur :

Votre choix : Madeleine

Autre descripteur ?

Votre réponse : o

Descripteur :

Votre choix : amour profane

Autre descripteur ?

Votre choix : n

Voulez-vous continuer ?

Votre réponse : n

*

* * *

MAJ.PRG permet à tout moment d'intervenir dans la liste d'autorités afin d'en limiter l'extension, soit par le regroupement de descripteurs hiérarchiquement ou associativement liés, soit par suppression de ceux qui présenteraient, au fil de la constitution de la banque, un taux d'occurrences non significatif. Le menu présente deux opportunités à l'utilisateur : soit le regroupement, soit l'effacement de descripteurs. En cas de regroupement, le choix du descripteur majeur (hiérarchiquement supérieur) et mineur incombe à l'utilisateur. L'opération peut être répétée plusieurs fois, et la réorganisation des fichiers ainsi que des index associés n'intervient qu'au terme du programme.

Le programme THESAU.PRG travaille à l'impression, en colonne, de la liste d'autorités, et fournit en regard la liste des occurrences pour chaque terme dans l'ensemble de la base de données. Il n'est pas interactif.

MOT.PRG permet une interrogation simple sur descripteurs, l'impression des numéros de textes, leur affichage, celui des références, l'impression sélective des textes retenus comme les plus significatifs et de leur référence ou, si le texte n'a pas été encodé, le renvoi au fichier matériel. Le programme demande à l'utilisateur l'introduction d'un descripteur, affiche son taux d'occurrences dans la banque de données, ou éventuellement son absence de la liste d'autorités.

Tableau 3 : Liste d'autorités

Page 1

ABIGAIL	63	AXA	1	CONSTANCE	10	DISCERNEMENT	4
ABISHAG	17	AZA	2	CONTEMPLATION	7	DIVERTISSEMENT	3
ABSTINENCE	6	BAISER	11	CONTINENCE	2	DORCAS	1
ACCOUCHEMENT	126	BEAUTE	267	CONTRITION	2	DOUCEUR	16
ADULTERE	4	BETHSABEE	83	CONVENANCES	14	DOULEUR	157
AFFLICTION	2	BIBLE	109	CONVERSION	16	ECONOMIE	18
AGAR	88	BIENSEANCE	10	CONVERSION	62	EDIFICATION	100
AGE	5	BIJOUX	19	COQUETTERIE	14	EDUCATION	73
ALLAITEMENT	62	BONNES OEUVRES	9	CORPS	19	ELISABETH	44
ALLEGORIE	47	BOUCHE	16	COURAGE	50	ELOQUENCE	5
AMBIITION	27	CALOMNIE	2	COUTURE	114	EMULATION	2
AMOUR CONJUGAL	59	CALVIN	9	CREATION	28	ENVIE	19
AMOUR DE DIEU	119	CANANEENNE	66	CREDULITE	2	EPOUSE	9
AMOUR FILIAL	1	CANON	4	CROIX	55	ERUDITION	1
AMOUR MATERNEL	48	CANTIQUES	10	CRUAUTE	1	ESTHER	205
AMOUR PROFANE	16	CAPRICE	2	CUISINE	7	ETAT	10
ANCIEN TESTAMENT	36	CARESSE	7	CULTE DES SAINTS	24	EVE	277
ANNE	71	CATECHISME	14	CULTURE	3	EVODIE	1
ANNE FEMME D'ELQANA	49	CETHURA	12	CURIOSITE	77	EXEMPLA	51
ANNE LA PROPHETESSE	48	CHAIR	15	DALILA	55	EXEMPLE	390
ANNONCIATION	1	CHARITE	78	DANSE	29	EXTASE	14
APOTRE	20	CHASTETE	96	DEBORA	87	FAIBLESSE	87
APPYA	1	CHEVEUX	50	DEFAUT	7	FAMILLE	3
ARDEUR	8	COIFFURE	5	DESHONNEUR	12	FARD	43
ATHALIE	24	COLERE	8	DESIR	29	FECONDITE	6
ATTITUDE	17	COMPASSION	2	DESOBEISSANCE	15	FEMME	137
AUMONE	13	CONCEPTION	12	DEVOIRS	41	FEMME ADULTERES	345
AUSTERITES	58	CONCUPISCENCE	2	DEVOTION	49	FEMME COURBEE	4
AUTORITE	15	CONFIANCE EN DIEU	6	DIGNITE	1	FEMME DE JOB	5

L'impression des numéros de textes permet de scinder la lecture, notamment si les réponses fournies sont trop nombreuses. L'affichage se fait au rythme voulu par l'utilisateur, qui en commande la séquence. Les références des textes retenus pour impression sont stockées dans un fichier intermédiaire. L'édition a lieu en fin de programme. La variable *occurrence* enfin, est dégressive : l'utilisateur sait à tout moment combien de textes restent à consulter.

Le programme `BOOL.PRG`, le plus complexe, reste très convivial. Il requiert de l'utilisateur l'introduction successive d'un descripteur, d'un opérateur, enfin d'un deuxième descripteur, l'ensemble constituant l'équation minimale. Celle-ci peut être complexifiée. Le programme réaffiche alors sa demande d'opérateur, puis de descripteur, jusqu'à ce que l'utilisateur lui signale la fin de l'équation. Les différents termes sont rangés dans une table intermédiaire. Les étapes suivantes, gérées par des programmes modulaires, ne sont pas interactives. Le programme affiche seulement l'état de déroulement des opérations, constitutions des tables, calcul des équations intermédiaires, compactage des différents fichiers, fin des opérations. L'équation est alors réaffichée, ainsi que son taux d'occurrence, puis les textes sont présentés tout comme sous `MOT.PRG`.

Tableau 4 : Affichage du déroulement du programme `BOOL.PRG`

INTRODUCTION DE L'EQUATION

Entrez votre équation :

Descripteur : prostituée

Opérateur : ou

Descripteur : libertinage

Equation complète ? n

Opérateur : et

Descripteur : grâce

Equation complète ? n

Opérateur : pas

Descripteur : Madeleine

Equation complète ? o

PREPARATION DE LA TABLE
 PREPARATION DE L'EQUATION
 COMPACTAGE FINAL

* EQUATION TERMINEE *

(Prostituée ou libertinage)

et

(grâce pas Madeleine)

Nombre de textes à afficher : 1

Souhaitez-vous imprimer l'ensemble des numéros de textes ?

Votre réponse : o

212

Ainsi dans le sac de toute une ville,
 où apparemment Raab n'estoit pas la
 seule femme de mauvaise vie, Dieu la
 choisit d'entre toutes les autres pour
 luy faire grâce.

BAILLET A.,

Vie des saints de l'Ancien
 Testament, Paris, 1684, p.123

Impression finale ? n

Il reste 0 texte(s) à visualiser

Voulez-vous introduire une nouvelle interrogation

booléenne ? n

*

* *

SORDT.PRG sert la mise en évidence des co-descripteurs d'un texte. Il permet de mieux discerner les thèmes corollaires, et donc de signaler les différents points à aborder soit dans un plan général, soit aux paragraphes d'un même chapitre. L'interaction avec le programme se réduit à l'introduction d'un descripteur. Le système prend alors en charge l'édition sur imprimante du descripteur examiné, des numéros de textes, suivis à chaque fois des co-descripteurs. Une étape ultérieure

permettrait de cerner la fréquence des co-occurrences dans la banque de données, et de signaler ainsi l'importance des corrélations entre thèmes.

Tableau 5 : Affichage du déroulement du programme SORDT.PRG

Descripteur desirez-vous examiner ?

Descripteur : pureté

Texte numero : 3

DOULEUR

MARIE

PLAISIR

PURIFICATION

Texte numero : 24

CHASTETE

HEROISME

SUSANNE

Texte numero : 25

CHASTETE

CONTINENCE

HEROISME

JUDITH

RETRAITE

VEUVAGE

Texte numero : 26

MARIE

VIRGINITE

Texte numero : 132

ABISHAG

JEUNESSE

SAGESSE

(...)

Voulez-vous continuer ?

Votre réponse : n

*
* *

Le programme SORTXT.PRG enfin, permet l'édition sur imprimante de textes intégrés, selon diverses séquences : tous les textes d'un même ouvrage; tous les textes des différents ouvrages d'un même auteur³; ou encore des textes ponctuels, déterminés par leur numéro.

5. Conclusions

Au terme de ce travail, nous mettrons plus particulièrement en évidence la souplesse de ce système de gestion, largement ouvert aux réorganisations en cours de recherche, soit par indexation sur critères nouveaux, soit par programmation plus élaborée. Un autre de ses atouts réside dans les multiples possibilités de classement des textes, qui se trouvent en outre allégés des fastidieuses références bibliographiques. L'interrogation simple est accélérée. L'innovation majeure reste néanmoins l'interrogation booléenne, qui peut être complexifiée jusqu'à fournir des réponses extrêmement pointues. Le calcul des occurrences et co-occurrences peut enfin être considéré comme "valeur ajoutée" à l'information initiale. En dévoilant les thèmes connexes, il structure la réflexion du chercheur et, souvent, laisse son empreinte jusqu'au stade de rédaction de la thèse.

³ Par la création de "filtres", qui permettent par ailleurs une interrogation simple ou booléenne limitée à un seul auteur.