

[The Knowledge Bank at The Ohio State University](#)

Article Title: Medium - Conception, réalisation et exploitation d'une base de données

Article Authors: Guillaumont, Agnes, and Jean-Luc Minel

Issue Date: December 1987

Publisher: William R. Veder, Slavisch Seminarium, Universiteit van Amsterdam,
Postbus 19188, 1000 GD Amsterdam (Holland)

Citation: *Polata Knigopisnaia: an Information Bulletin Devoted to the Study of Early Slavic Books, Texts and Literatures* 17-18 (December 1987): 60-72.

Appears in:

Community: [Hilandar Research Library](#)

Sub-Community: [Polata Knigopisnaia](#)

Collection: [Polata Knigopisnaia: Volume 17-18 \(December 1987\)](#)

MEDIUM
CONCEPTION, REALISATION ET EXPLOITATION
D'UNE BASE DE DONNEES

AGNES GUILLAUMONT, JEAN-LUC MINEL
Institut de recherche et d'histoire des textes
40, Avenue d'Iéna
75116 PARIS CEDEX

L'Institut de Recherche et d'Histoire des Textes, dont l'objectif est le recensement, l'analyse et le microfilmage des manuscrits essentiellement médiévaux, s'est trouvé confronté, comme bon nombre d'institutions, à l'explosion des informations produites par les différentes sections qui le constituent: plusieurs millions de fiches dactylographiées, des "normes" propres à chaque section, etc... Cette accumulation d'informations, quelquefois difficilement manipulable, entraîne des recherches longues, complexes et coûteuses.

Aussi, dans un souci de rationalisation et de simplification, l'IRHT a entrepris d'implanter une base de données sur le manuscrit médiéval. Plusieurs objectifs et contraintes ont été définis. Ceci dans le but d'améliorer aussi bien en amont (la constitution et la mise à jour de la base de données) qu'en aval (la consultation de la base de données) l'efficacité du traitement de l'information:

- Utiliser un logiciel évolutif et doté des outils de manipulation des données de haut niveau. Nous avons donc choisi un Système de Gestion de Bases de Données (SGBD) plutôt qu'un logiciel documentaire.

Concevoir un découpage de l'information primaire (les manuscrits médiévaux) qui laisse la possibilité d'enrichir la base de données au fur et à mesure que les sections de l'IRHT décideraient de s'associer au projet.

Développer des procédures d'interrogation de la base de données qui nous permettent d'exploiter les synergies existant entre les traitements des données effectués par des équipes travaillant sur des documents différents, mais dont l'approche conceptuelle est similaire.

Simplifier la tâche de saisie des données des analystes de l'IRHT en leur procurant, à la demande et interactivement, les informations nécessaires (listes d'autorité, vérification de la cohérence, signalement des lacunes, etc...). En effet, ce travail, souvent fastidieux, reste le goulot d'étranglement de toute saisie d'informations.

Les performances du système sont conformes aux exigences premières. La valeur ajoutée est impressionnante: là où une consultation des fichiers manuels peut exiger une journée de travail, MEDIUM fournit la réponse en quelques secondes.

Exploitation

Interrogation de la base.

L'interrogation de MEDIUM ne nécessite pas la connaissance d'un langage documentaire. L'utilisateur est guidé par un système d'écrans de dialogue; plusieurs chemins d'accès lui sont offerts, lui permettant de se focaliser sur l'objet de sa recherche en plusieurs étapes. L'utilisation de terminaux à écran pleine page permet un maniement agréable et l'ergonomie sera encore améliorée dès que le mode multi-fenêtres pourra être employé.

La base de données est implantée sur les ordinateurs du CIRCE. L'interrogation peut s'effectuer soit de l'IRHT, soit de tout endroit accédant aux principaux réseaux de télécommunication internationaux (TRANSPAC, EURONET, THYMNET, ...).

Un chercheur peut ainsi avoir accès aux 37000 manuscrits microfilmés de l'IRHT.

Mise à jour de la base de données.

Les corrections et ajouts d'informations dans la base de données sont réalisés interactivement et la cohérence des informations est vérifiée. Ainsi tout nouvel auteur ou nom de bibliothèque est immédiatement détecté et une confirmation doit être donnée par l'analyste. Cette phase de validation évite d'une part un bon nombre d'erreurs de saisie, et d'autre part oblige les analystes à respecter les règles de normalisation définies pour chaque champ d'information.

Critiques et perspectives

MEDIUM représente environ 20% de l'information accessible par les fichiers manuels, et l'utilisateur se voit donc contraint de compléter sa recherche par la consultation de ces fichiers, ce qui est particulièrement décourageant. L'accroissement de la base de données est donc actuellement un des axes principaux du travail des sections de l'IRHT.

L'information ainsi disponible devrait ouvrir de nouveaux horizons aux spécialistes, ne serait-ce que par les rapprochements et confrontations que suscite la combinaison des divers critères de recherche. Plus encore si l'on songe aux traitements statistiques rendus possibles par l'existence d'une passerelle avec le logiciel SAS⁽²⁾, qui offre à l'utilisateur

- Simplifier la consultation de la base de données en supprimant toute interrogation par un langage spécialisé dont la syntaxe est souvent pénible et source d'erreurs. Nous avons opté pour l'utilisation d'écrans de dialogue où l'utilisateur fournit uniquement les critères de recherche.

- Ouvrir la consultation de la base de données aux chercheurs du monde entier. La base de données est reliée au réseau de télécommunications Transpac, qui est lui-même interconnecté aux principaux réseaux internationaux.

Réalisation: le choix d'une architecture

En nous appuyant sur les possibilités offertes par le modèle entité-relation⁽¹⁾, nous avons construit notre base autour de neuf entités:

MANUSCRIT : cote du manuscrit, datation, et mention de la section de l'IRHT concernée.

MICROFILM : référence du microfilm du manuscrit dans la filmothèque de l'IRHT, et informations techniques nécessaires à la gestion de la filmothèque.

OEUVRE : description conceptuelle d'une oeuvre par son titre, sa langue et sa nomenclature.

INCIPIT : incipit principal et incipit des pièces annexes (en test).

AUTEUR : forme normalisée du nom de l'auteur associée à une description biographique.

TRADUCTEUR : forme normalisée du nom du traducteur et titre de la traduction.

COMMENTATEUR : forme normalisée du nom du commentateur et titre du commentaire.

HORS-TEXTE : description physique du texte (illustrations, blasons, etc...).

DIPLOMATIQUE : caractérisation des documents diplomatiques (ordre religieux, lieu et institution, diocèse).

Un manuscrit, et chacun des textes qu'il contient, est décrit par un faisceau d'informations réparties dans les différentes entités. Des relations permettent de lier entre elles les informations relatives à un même manuscrit. Celles-ci seront utilisées lors de la consultation pour permettre à l'utilisateur de naviguer dans la base.

toute sa puissance statistique et graphique. Un chercheur peut extraire de la base des informations sélectionnées et appliquer à ces données une analyse multidimensionnelle aussi bien qu'un simple calcul de fréquences.

D'une manière générale, il est possible d'extraire de MEDIUM un sous-ensemble de la base et de le recopier sur un micro-ordinateur pour exploiter ces données localement, et effectuer ainsi des traitements plus spécifiques et plus fins.

Le modèle de représentation des données actuellement implanté dans MEDIUM est en fait un noyau d'informations indispensable à l'identification du manuscrit et des textes qu'il contient. L'architecture modulaire adoptée donne à ce noyau initial toutes possibilités de se développer. Une section de l'IRHT qui ne trouverait pas dans MEDIUM le type d'informations qui l'intéresse peut construire une entité, choisir sa description de données, et relier cette entité aux informations existantes. Son travail en sera enrichi, et dans le même temps, si elle le désire, elle rend ses données accessibles aux autres chercheurs. On voit comment, avec le temps, MEDIUM peut accroître la synergie existant entre les différentes disciplines qui étudient le manuscrit médiéval. Néanmoins, si riche soit-elle, cette description ne restituera jamais toute la richesse d'informations que représente un manuscrit. Le couplage de la base de données avec un disque optique numérique devrait permettre, dans un avenir qui s'approche maintenant, la reproduction immédiate à l'écran des manuscrits sélectionnés par une consultation de MEDIUM.

De nouvelles possibilités soulèvent toujours de nouveaux problèmes. Ainsi la nécessaire normalisation des données engendre-t-elle des difficultés à la consultation: difficultés orthographiques, difficultés d'inférer sur les données. L'un des objectifs de la coopération entre l'IRHT et l'Institut d'Etudes Médiévales de Montréal (dans le cadre d'un projet France-Québec) est de mettre en lumière la façon dont une telle base de données peut être exploitée dans un environnement autre que celui dans lequel elle a été conçue. Un nombre important d'informations sont implicitement contenues dans la description des manuscrits. Pour pouvoir retrouver une information complexe dans cette base, il faudrait pouvoir suivre le même cheminement intellectuel, le même raisonnement, que celui qui a inscrit cette information. Or une base de données ne conserve que les informations déduites du raisonnement. C'est pourquoi nous envisageons de greffer sur la base de données des procédures utilisées dans l'intelligence artificielle et notamment dans les systèmes experts⁽³⁾. Ces procédures, en reproduisant le savoir faire des analystes maîtrisant parfaitement MEDIUM, pourront constituer une aide à la production ou à la consultation des données.

Bibliographie

1. Delobel. - Bases de données et systèmes relationnels, Dunod 1982.

A.Guillaumont, J.L.Minel. - MEDIUM, bases de données sur le manuscrit médiéval, Automatic processing of art history data and documents, vol.1, Pise 1984

2. D.Boehm, A.Naud. - Interface SAS-ADABAS, Documentation CIRCE, Orsay, 1985

3. Sur les systèmes experts et leurs applications dans le domaine historique, voir:

Buchanan. - Rule Based Expert Systems, Addison Wesley, 1983

A.Colmerauer. Prolog, bases théoriques et développements, T.S.I. vol.2 n°4, 1983

M.Stonebraker, J.Woodfill, E.Andersen. - Implementation of rules in Relational Data Base Systems, nov. 1983

G.P.Zarri. RESEDA, Thèse d'Etat, Paris 1985

C.Bourlet, L.Fossier, A.Guillaumont, J.L.Minel. - Projet de constitution d'un système expert dans le domaine historique, Documentation IRHT, Paris 1985

C.Bourlet, J.L.Minel. - Un système déclaratif d'aide à l'identification des individus dans un corpus prosopographique, Informatique et Sciences humaines, Paris, juin 1987