

LA REVUE  
DES UTILISATEURS  
DE L'INFORMATIQUE DANS LA RECHERCHE

# U NIX & SYSTEMES

**Stations serveurs :  
les meilleurs serveurs du moment**  
page 32

# D OSSIER

**Windows 95**  
page 57

# I NTERNET & RESEAUX

**SNMP version 1**  
page 78

# P UBLIER

**Latex, un pont pour faciliter  
la communication de demain**  
page 140

**N°60**

Septembre/octobre  
1995



**I S T & G E D**

# Méthode d'évaluation bibliométrique d'un logiciel de développement d'applications : Visual Basic

Gérard Pillet, Luc Quoniam, Hervé Rostaing, Henri Dou  
(Centre de recherche rétrospective de Marseille)

*La multitude des logiciels et langages de programmation présents sur le marché pose un réel problème à l'utilisateur. Le choix et la sélection restent basés sur la lecture de leur documentation, d'articles critiques, de publicités, etc. Si des méthodes d'assistance à la sélection peuvent être utilisées, elles permettront en gardant une objectivité relative un considérable gain de temps et d'efforts.*

Le monde de la science et de la technique est en perpétuelle évolution. La masse d'information engendrée par cette mutation régulière croissant de façon exponentielle, il devient indispensable de gérer au mieux la complexité de ce monde en «ébullition» pour bien le comprendre.

Cette communication présente une méthode d'analyse de l'information documentaire : elle est différente de l'approche traditionnelle de gestion de l'information documentaire. Elle ne répond pas aux objectifs de stockage et de recherche des documents rencontrés dans les outils de gestion documentaire. L'approche, envisagée ici, a pour objet l'analyse du contenu d'un ensemble de documents. On ne se limite plus à la simple gestion du document, mais on va chercher à extraire du «sens» d'un ensemble de documents en fonction de leurs contenus. Cette approche est connue sous le nom de «bibliométrie».

## La bibliométrie

Bien que méconnue d'un grand nombre, la bibliométrie est pratiquée de longue date (début de ce siècle). Depuis plus de vingt ans, le CRRM participe activement à l'évolution de cette discipline (1) (2).

Les données analysées en bibliométrie sont le plus souvent issues des banques de données internationales accessibles en ligne (c'est-à-dire consultables par les réseaux de télécommunication ou par Internet). Ces données se présentent, pour la plupart, sous forme de références bibliographiques.

La bibliométrie peut être définie comme étant l'application de méthodes mathématiques ou statistiques sur des ensembles de références bibliographiques. L'application de ces méthodes permet d'élaborer des indicateurs de tendance traduisant la globalité de l'information contenue dans un ensemble de références bibliographiques. Ces indicateurs vont synthétiser le contenu scientifique, technique ou technologique d'un grand nombre de publications (d'une centaine à plusieurs milliers) sans même les avoir lues. La bibliométrie est aux publications ce que la démographie est aux populations (3).

La bibliométrie est l'outil idéal pour les domaines scientifique, technique ou technologique où l'acte de publication et de divulgation joue un rôle important (publications scientifiques, dépôts de brevets, articles de presse...). Elle devient d'autant plus indispensable lorsque l'ampleur de la masse de ces publications ne permet plus une analyse par simple lecture.

Ainsi, la bibliométrie est utilisée pour :

- l'élaboration des programmes de recherche institutionnelle ou industrielle,
- la veille scientifique, technique et technologique (acquisition du savoir publié par autrui),
- la veille concurrentielle (surveillance active de sa concurrence),
- la sociologie de la science.

Le travail présenté ici s'est focalisé sur l'analyse d'un ensemble de références bibliographiques d'articles traitant du logiciel Visual Basic (4). L'application de la bibliométrie à cet ensemble de références (nommé aussi corpus bibliographique) va nous permettre de visualiser les pôles d'application, les orientations et les évolutions de Visual Basic.

Ce corpus n'étant absolument pas exhaustif, cette analyse n'est qu'une illustration de la méthode plus qu'une évaluation globale de Visual Basic. L'analyse peut être facilement transposée pour d'autres logiciels ou langages puisqu'une source globale d'information sur ce domaine est accessible sur le serveur de banques de données *Dialog*.

## Méthode d'analyse

### Données brutes

Le corpus a été obtenu par une interrogation multi-bases à l'aide de la commande DIALINDEX du serveur de bases de données *Dialog* (5) avec le terme VISUAL BASIC. Les références (272) ont été téléchargées sous un format réduit (peu coûteux) contenant uniquement le champ <descripteur> et le champ <code> de chaque référence conformément à la présentation du tableau 1.

Tableau 1 : exemple de références téléchargées.

CZ-	DIALOG(R)File 256:(c) 1994 Info.sources Inc. All rts. reserv.
DE-	Multimedia^Authoring Systems^Entertainment Industries^Visual BASIC^Program Development Aids
CZ-	DIALOG(R)File 256:(c) 1994 Info.sources Inc. All rts. reserv.
DE-	Database Management^Program Development Aids^Windows^IBM PC & Compatibles^Visual BASIC
CZ-	DIALOG(R)File 256:(c) 1994 Info.sources Inc. All rts. reserv.
DE-	Visual BASIC^IBM PC & Compatibles^Draw^Graphics Tools^Image Processing

L'analyse sera faite sur les champs descripteurs <DE- > communs aux références sélectionnées. Ce champ semble indexé de manière homogène sur les sources de données, et contient des termes descripteurs représentatifs des références.

### Données reformatées et dédoublonnées

Dans un premier temps, nous avons utilisé le logiciel INFOTRANS de I+K (6) pour extraire et reformater le champ <descripteur> de ces références, et ainsi obtenir un fichier de données homogène. Ce fichier est ensuite dédoublonné. Le fichier texte des données brutes a donc été traité en utilisant une table de transfert (.trf) créée sous INFOTRANS. Cette table a permis de reformater le champ <DE- >, notamment en remplaçant les séparateurs entre descripteurs par un séparateur unique (/), puis de créer un champ <DUBC- > (dubcode) sur lequel le logiciel effectuera un dédoublonnage (7). Un exemple de résultat est donné dans le tableau 2.

### Traitement bibliométrique

Sur le fichier de données reformatées, nous avons utilisé un logiciel de traitement bibliométrique mis au point au CRRM : DATAVIEW (8) (9). Ce logiciel a d'abord permis d'extraire les formes graphiques (10) (ici les descripteurs) présentes dans les références, puis de traiter ces formes en réalisant

un codage de paires intra-champs sur les descripteurs. Enfin, nous avons édité des paires et des groupes de descripteurs en faisant varier les paramètres de codage et d'édition pour affiner l'analyse, diminuer la taille des groupes, éliminer les formes triviales et le bruit...

Tableau 2 :  
exemple de  
références  
reformatées.

CZ-	DIALOG(R)File 2:(c) 1995 Institution of Electrical Engineers. All rts. reserv.
DE-	civil engineering computing/construction industry/digital simulation/expert systems/materials handling/programming/user interfaces
DUBC-	civil engineering computing/construction industry/digital simulation/expert systems/materials handling/programming/user interfaces
CZ-	DIALOG(R)File 233:(c) 1995 Learned Inform.Inc. All rts. reserv.
DE-	object-oriented languages/visual basic/Philosophy/programming aids/programming languages
DUBC-	object-oriented languages/visual basic/Philosophy/programming aids/programming languages

## Informations produites par DATAVIEW

Le traitement a généré plusieurs fichiers de données qui ont permis d'interpréter les résultats de l'analyse.

Trois types d'informations ont été utilisés :

La liste des formes extraites éditée sous deux formats :

- par ordre alphabétique, pour identifier les indexations présentes sous différentes orthographes ;
- par fréquences d'apparition décroissantes pour voir quelles sont les formes triviales ou isolées. (11)

Un certain nombre de descripteurs était très fréquemment présent (fréquences supérieures à 25). Nous avons retrouvé aussi beaucoup de descripteurs à fréquence égale à 1. Ces formes à fréquences très fortes ou très faibles (non significatives pour l'analyse) ne seront pas considérées dans le traitement. Ces caractéristiques induisent un intervalle de fréquence de formes retenue pour la suite du traitement : de 2 à 25.

La liste des paires de descripteurs, produite par le «codage» des paires. Editée par ordre alphabétique pour identifier des paires qui ne seront pas détectées lors de l'édition des groupes, et par fréquences d'apparition décroissantes pour visualiser les paires les plus abondantes (triviales) et les plus rares (bruit).

Des fichiers de groupes, produits par le module d'édition des

groupes (12) (13) (14). Chacun de ces fichiers est produit par le logiciel en fonction de paramètres seuils définis lors du chaînage des paires de descripteurs. Pour obtenir des groupes significatifs (de tailles réduites), il a fallu varier les valeurs minimum pour les fréquences des paires du chaînage, soit respectivement >2, >3 et >4.

Nous avons ainsi réalisé plusieurs traitements successifs en travaillant sur un intervalle de fréquence des formes constitutives toujours compris entre 2 et 25.

A la vue de ces résultats, nous avons réalisé une représentation graphique sous la forme d'un schéma présentant les descripteurs et les liens qui les unissent conformément à leurs fréquences de co-apparition. Sur ce graphe des paires, apparaissent les fonctionnalités, les orientations et les évolutions de Visual Basic au regard des références considérées.

## Mode d'interprétation

Certains groupes sont immédiatement apparus intéressants, ils ont donc été restitués tels quels sur le schéma. D'autres ont été pris en compte en utilisant le traitement qui permettait la meilleure représentation de ces groupes en termes d'intérêt et de signification.

Après validation par un expert du domaine, certaines formes jugées inintéressantes pour le sujet n'ont pas été considérées dans l'interprétation graphique. L'objectif étant de mettre en valeur des orientations, des évolutions, seuls les groupes et les descripteurs présentant un réel intérêt pour l'analyse ont été conservés.

Pour obtenir une lisibilité correcte sur la figure 1, nous avons relié les descripteurs par des traits d'épaisseur variée, proportionnels aux fréquences de co-apparition des formes liées. Soit respectivement pour les fréquences inférieures à 2, comprises entre 2 et 4, et supérieures à 4.

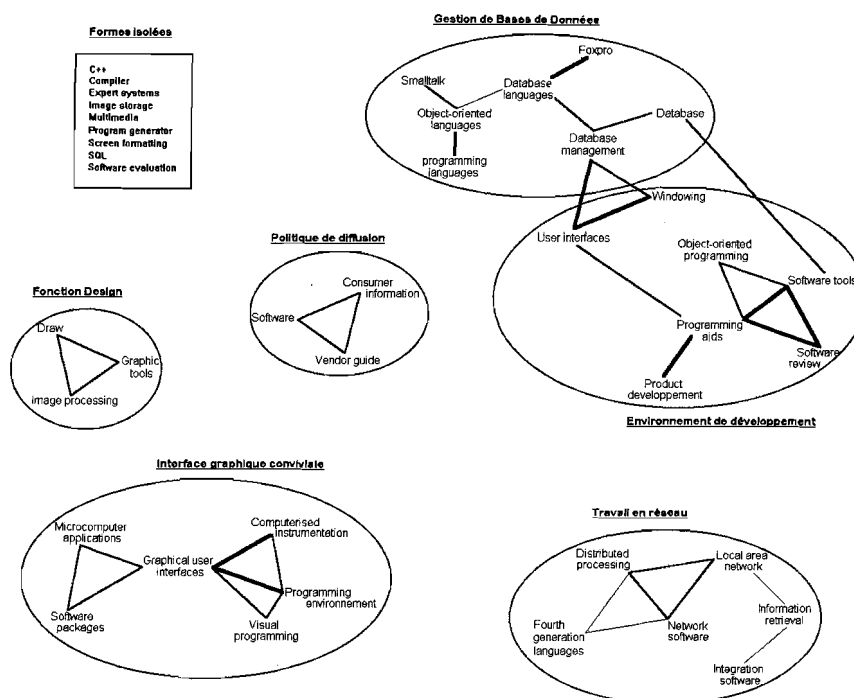
Lors de l'édition des groupes, DATAVIEW a détecté certaines formes isolées (non chaînées avec d'autres formes), mais avec des fréquences d'apparition non négligeables. Elles sont représentées dans le tableau 3. Leur marginalisation a semblé être un facteur intéressant pour les mentionner. Il est possible de penser qu'ils présentent un certain caractère «innovant».

Tableau 3 : liste des formes isolées à fréquence forte.

<C++>	<Program generator>
<Compiler>	<Screen formatting>
<Expert systems>	<SQL>
<Image storage>	<Software evaluation>
<Multimedia>	

L'ensemble des groupes et formes isolées ont été regroupés dans la figure 1.

Figure 1



## Interprétations

Un spécialiste de la programmation sous Visual Basic pourra constater que les résultats obtenus sont relativement cohérents et reflètent assez fidèlement les caractéristiques du langage. Nous remarquons sur le graphe le regroupement entre descripteurs relatifs à la gestion des bases de données (DATABASE, DATABASE MANAGEMENT, DATABASE LANGUAGES). Visual Basic a en effet produit un effort particulier dans le domaine de la gestion des données. Notamment en intégrant un nouveau module database management, qui utilise le moteur du système de gestion de bases de données *Access 1.0* de Microsoft, et en prévoyant des formats d'import-export pour les logiciels *Paradox*, *Dbase*, *Foxpro*...

Nous observons un regroupement autour de la gestion et de la création d'interfaces graphiques (GRAPHICAL USER INTERFACES), caractéristiques essentielles du logiciel. Les spécificités de Visual Basic en matière de développement d'applications conviviales sont bien retrouvées, comme la richesse et la puissance de l'environnement utilisateur interactif du type Windows.



Les fonctionnalités techniques de design de Visual Basic pour ses applications sont également représentées par un groupe isolé (GRAPHIC TOOLS, IMAGE PROCESSING, DRAW). La qualité et la diversité des outils de création et de dessin d'objets, ainsi que les possibilités de mise en forme des données de Visual Basic sont bien représentées.

Les systèmes de gestion de réseaux sont très nettement visibles et montrent que Visual Basic a aussi été conçu pour travailler sur ce support et pour organiser la distribution de ses applications.

La politique de diffusion de Microsoft apparaît également avec un groupe isolé de descripteurs où se retrouvent les termes CONSUMER INFORMATION, VENDOR GUIDE et SOFTWARE. C'est la preuve de la volonté du producteur d'accompagner son produit par des types de communications variés.

Il est possible de remarquer par contre que des thèmes essentiels tels que la gestion OLE (Object Linking and Embedding) ou DDE (Dynamic Data Exchange) ne sont pas apparus (fonctions pourtant très importantes). La communication entre applications et l'exploitation d'autres ressources Windows sont pourtant des fonctions fondamentales de la programmation et du développement sous Visual Basic 3.0.

Cependant, l'analyse avait pour objet de réaliser une évaluation globale du langage par des méthodes automatiques, et non d'en appréhender toutes les potentialités. Mais on retrouve quand même les principales orientations de Microsoft qui tente de faire évoluer Visual Basic, environnement de développement d'applications Windows, vers un outil de création d'interfaces pour des utilisateurs de tous niveaux. Utilisant un langage orienté-objet dérivé du Basic, Microsoft s'axe plus particulièrement vers le développement de systèmes de gestion de données, objectif prioritaire de beaucoup de logiciels de programmation et de développement du moment.

L'analyse a donc permis d'appréhender d'une manière assez réaliste et fidèle un langage de programmation en utilisant un traitement bibliométrique. Sans avoir eu à trouver et à lire des fiches techniques et une fastidieuse documentation, il est possible de se faire une idée assez précise des fonctionnalités, évolutions et particularités d'un tel produit. En utilisant une méthodologie type, un utilisateur pourra appliquer des techniques de gestion de l'information scientifique et technique pour évaluer les capacités d'un langage de programmation ou d'un logiciel, et ainsi orienter son choix.

*G. P., L. Q. H. R., H. D.*

## Bibliographie

- (1) Dou H., Hassanaly P., «An Original Creation of a retrospective Research Service at the University of Aix-Marseille», Proceedings : *Online Review*, 3rd International Online Information Meeting, Londres, 4-6 Déc., pp. 73-81, 1979.
- (2) Desvals H., Dou H. (éds), *La veille technologique*, Dunod, 1992.
- (3) White H.D., McCain K.W., «Bibliometrics», *Annual review of information, science and technology (ARIST)*, vol. 24, pp. 119-186, 1989.
- (4) Microsoft SARL, 18 av. du Quebec, 91957 Les Ulis Cedex.
- (5) Dialog Information Services Inc. 3460 Hillview Av. Palo Alto. CA 94303-0993. USA.
- (6) I+K France, 11 bis, rue Al.-Joly. 78000 Versailles.
- (7) Christensen F.H., Ingwersen P., «Fundamental methodological issues of data set creation online for the analyses of research publications», Proceedings : *Fifth conference of the international society for scientometrics and informetrics*, Chicago, pp. 103-112, 06/7-11/1995.
- (8) Description du logiciel sur le serveur WWW du CRRM à l'adresse :  
<http://crrm.univ-mrs.fr/commercial/software/dataview/dataview.html>
- (9) Rostaing H., *Veille technologique et bibliométrie*, Thèse : univ. Aix-Marseille III, 353 p., 13/01/1993.
- (10) Lebart L., Salem A., *Statistique textuelle*, Dunod, 1994.
- (11) Lhen J., Lafouge T., Elskens Y., Quoniam L., Dou H., «La "statistique" des lois de Zipf». Proceedings : *Les systèmes d'information élaborée*, Ile Rousse, 30/05-02/06/1995.
- (12) Haon H., Paoli C., Rostaing H., «Perception d'un programme de R. & D. à travers l'analyse bibliométrique des banques de données d'origine japonnaise», Proceedings : *IDT 93*, pp. 63-70, 22-24/06/1993.
- (13) Bordons M., Zulueta M.A., Cabrero A., Barrigon S., «Identifying research teams with bibliometric tools», Proceedings : *Fifth conference of the international society for scientometrics and informetrics*, Chicago, pp. 83-92, 06/7-11/1995.
- (14) Abd-El-Kader M., Miquel J.-F., Dore J.-C., «Méthode d'analyse des thèmes et réseaux de la coopération scientifique nationale et internationale sur les céramiques dans la base SCI», Proceedings : *Les systèmes d'information élaborée*, Ile Rousse, 30/05-02/06/1995.