

Un système interactif pour la création et l'interrogation de fichiers

Par J.-Ph. Baud et A. Jacquesson, Genève

Au cours de ces dernières années, il est apparu nécessaire que les sciences sociales et humaines disposent d'un système simple pour le stockage et la recherche d'information. La société Control Data a développé un tel langage: INFOL qui est l'acronyme anglais de « Information Oriented Language ». Il s'agit d'un langage de programmation simple, facile à apprendre (on l'enseigne en 4 demi-journées) qui a connu un grand succès auprès des utilisateurs du centre de calcul de l'Université de Genève. Mais ce langage n'est implémenté que sur la série des ordinateurs CDC 3000 upper. En vue d'y apporter des améliorations et afin de le rendre compatible avec d'autres machines, l'Institut interfacultaire de calcul électronique a entrepris d'écrire deux compilateurs INFOL en FORTRAN standard. Le premier est une version qui doit tourner en batch processing et qui doit prendre la relève de l'INFOL CDC. Le second est une version interactive qui est prévue pour fonctionner par l'intermédiaire d'une console et d'un écran de visualisation (display). Une première version batch FORTRAN nous a été fournie par CDC, elle avait été développée à la Northwestern University; un groupe de l'IICE continue la mise au point de ce compilateur.

La mise en mode conversationnel d'INFOL avait essentiellement deux buts. Le premier, de diagnostiquer les erreurs de syntaxe et de logique et de permettre la correction immédiate de ces fautes. Le second est de « souffler » à l'utilisateur certaines actions à entreprendre. Du point de vue software, le compilateur INFOL-2 se présente comme un programme FORTRAN standard, à quelques exceptions près: utilisation de certaines instructions d'entrée/sortie propres à CDC. Le langage INFOL admet des informations de différents types: des textes alphanumériques (entiers et flottants) et des codes numériques ou mnémoniques. Le stockage sur bande magnétique se fait dans des « records » de longueurs variables. On considère que le maximum d'informations pouvant être traité par INFOL doit être contenu sur une bande. Pour l'utilisateur, la préparation des données est extrêmement simple; il les perfore en champs libres, l'astérisque étant considéré comme séparateur.

Le langage peut se diviser en 3 phases principales: création du fichier, mise à jour, interrogation et édition des résultats. La création du fichier, la phase ESTABLISHMENT, a un but purement didactique; il est en effet impensable de créer un fichier comprenant l'équivalent de plusieurs milliers de cartes de données en mode interactif. Par contre nous pensons qu'il peut être intéressant d'appren-

dre à manier INFOL par une telle méthode, malgré les quelques changements de syntaxe que le mode interactif impose.

La mise à jour est une opération purement technique qui ne pose pas de problèmes particuliers.

La phase la plus intéressante est sans aucun doute l'interrogation. Toute question posée correctement va donner divers renseignements statistiques simples qui doivent permettre à l'utilisateur d'analyser les réponses à sa question. En effet cette question peut être complexe : elle est alors décomposée en sous-critères reliés par des opérateurs logiques .OR., .AND., .NOT. ; et l'introduction d'une analyse statistique simple permet d'exprimer de façon numérique ces notions d'intersection, d'union et de complément.

Le système interactif présente ici un immense avantage : à l'aide des résultats statistiques, l'utilisateur pourra poser des questions de plus en plus précises et éliminer ainsi le « bruit », erreur bien connue en sémantique documentaire, ou au contraire, élargir le domaine de recherche en cas de résultat nul. L'écran de télévision permet ensuite de déterminer rapidement la mise en page pour l'édition des résultats.

Applications

On peut traiter essentiellement deux types de renseignements :

- numériques (âge, sexe, salaire, etc.)
- réponses ouvertes (texte en langue naturelle).

Citons comme exemples les 2 importantes applications suivantes :

1) L'Institut de géographie (M. le professeur Raffestin) a créé un fichier des frontaliers travaillant à Genève (13 000 frontaliers). Il s'agit de la première étude exhaustive sur le sujet. Les renseignements enregistrés (âge, sexe, commune d'origine, salaire, etc.) ont permis de construire des pyramides d'âge, des moyennes de salaires, des statistiques par commune, par poste douanier, etc.

2) Le Département de science politique (M. David H. Handley) a entrepris une compilation d'environ 400 enquêtes d'opinion publique sur l'intégration européenne depuis 1945, soit un dépouillement de 400 000 réponses individuelles. INFOL leur a permis le stockage, l'analyse et l'édition in extenso de ces réponses.