

Jean-Marie PINON, Professeur
Laboratoire d'Ingénierie des Systèmes d'Information
Bâtiment 502 – 20 av. Albert Einstein – 69621 Villeurbanne CEDEX
Tél. (33) 04 72 43 84 81 - Fax . (33) 04 72 43 85 18
e-mail : pinon@if.insa-lyon.fr

Andréa IACOVELLA
Ecole Française d'Athènes
6, rue Didotou - 106 80 Athènes – Grèce
Tél. (301) 36 12 518 / 521 – Fax. (301) 36 32 101
e-mail : iacovella@efa.gr

**Proposition de sujet de thèse
pour Aurélien BENEL**

**Modélisation opératoire de la documentation de consultation
pour experts : application à l'archéologie**

Encadrement : Jean-Marie Pinon¹ (LISI), Andréa Iacovella (EFA)

1. Problématique

Il existe deux manières d'exprimer des connaissances en vue d'un raisonnement ultérieur : la première fondée sur l'utilisation d'un langage formel, la seconde fondée sur l'utilisation d'un langage naturel (textes, schémas, images...). La supériorité des connaissances exprimées dans un langage formel réside dans le fait que les raisonnements qui s'y appliquent puissent être effectués de manière totalement automatique. Si l'on a pu croire un instant, devant les réussites de l'informatique de gestion, que les raisonnements humains pourraient tous être formalisés, on doit reconnaître aujourd'hui, à la lumière de l'incursion du documentaire (donc de l'informel) en informatique et de la déception envers les systèmes experts que, dans la plupart des professions intellectuelles, le document demeure l'unique support valide de la connaissance (plus particulièrement en consultation) : du chercheur et ses articles, à l'ingénieur et ses normes, en passant par le juriste et ses articles de lois.

L'enjeu est maintenant d'offrir des outils de consultation de cette connaissance informelle (sur support informatique) plus appropriés que ceux hérités de la tradition du livre de Gutenberg. La première méthode revient à formaliser la structure logique des documents, c'est à dire à représenter formellement les relations entre des « îlots » informels pour obtenir des « hyperdocuments ». La seconde consiste à décrire dans un langage formel, appelé ici langage documentaire, le contenu sémantique de ces îlots informels [PINO 97]. Traditionnellement, ces deux approches sont identifiées comme appartenant à deux domaines différents : la navigation interactive, et la recherche d'informations, basée sur les principes du traitement par lot (une question, des milliers de réponses triées par pertinence). De nombreux chercheurs ont cependant tenté d'intégrer la structure logique pour optimiser la recherche d'informations (cf. [FOUR 98]). Nous considérons pour notre part que, si la structure logique et la description dans un langage documentaire sont liés, c'est parce que l'une exprime les relations explicites et l'autre les relations implicites entre des textes. Ainsi, on peut définir la recherche de documents comme une navigation d'un ordre supérieur où chaque application ou relâchement d'une contrainte se traduit par un saut d'un corpus à un autre. C'est parce que nous sommes convaincus que l'intelligence sera toujours du côté de l'humain et non de l'automate qu'il nous semble indispensable de mettre l'interactivité au cœur de la recherche de documents.

Une autre limite des systèmes documentaires actuels est l'application du modèle issu de l'informatique de gestion, dans lequel le cadre des informations est fixé par la direction de l'entreprise, les informations sont saisies par des employés et sont consultées par des clients. Si ceci reste acceptable au niveau d'une bibliothèque « grand public », ceci est tout à fait inadapté dans le cas qui nous intéresse des documents pour experts. Dans une profession où la consultation de documents occupe une place centrale, imposer à l'expert une description des documents, c'est nier son expertise. En effet, la problématique change quelque peu, il ne s'agit pas tant de trouver des documents que de les « retrouver », en effet, l'expert lit, lors de leur parution, la plupart des documents susceptibles de l'intéresser un jour, et, face à une question ultérieure, devra s'orienter dans son raisonnement par rapport à ses raisonnements précédents et à ceux de ses collègues.

2. Description

Nous proposons d'indexer les documents en construisant, « au-dessus » de leur identifiant, un ordre partiel de descripteurs (à rapprocher dans [SOWA 92] de la relation « sorte de »). Cette entorse à l'habituelle structure arborescente permet de représenter un descripteur comme la composition d'autres descripteurs. Ce mécanisme s'il est

¹ Jean-Marie Pinon sera secondé par Sylvie Calabretto (LISI).

intéressant pour des concepts ou des relations entre concepts est primordial pour des régions spatiales ou des intervalles temporels (cf. [PRED 99]).

Nous concevrons le processus interactif qui, intégré à une interface homme-machine, permettra à l'expert de consulter des portions de documents, de prendre des notes de lecture et d'enrichir son ontologie du domaine, de comparer et d'échanger ses notes ainsi que son ontologie avec d'autres experts, et enfin, de retrouver, des notes de lectures et les portions de documents associées.

Il s'agira d'adapter la complexité temporelle des calculs aux contraintes des interactions homme-machine. La solution choisie pourra s'inspirer de celle de H. Aït-Kaci (cf. [AITK 89]) en calculant à l'avance les opérations récursives, coûteuses en temps de calcul, et à faire « à la volée » les opérations ensemblistes, opérations dont les systèmes de gestion de bases de données s'acquittent fort bien.

3. Validation et applications

Nous pourrions valider nos modèles en les appliquant au raisonnement archéologique, particulièrement intéressant pour sa position entre science formelle et science humaine (cf. [GARD 96]) et pour la position centrale qu'il donne à l'espace et au temps (cf. [GALL 86]) ; dans des activités aussi représentatives que la consultation d'un corpus de chroniques de fouille, celle d'une bibliothèque spécialisée, celle d'une photothèque scientifique ou enfin celle de typologies d'artefacts. Sera réalisé un prototype opérationnel du système informatique illustrant en particulier son nouveau mode d'interaction homme-machine et son utilisation via internet (répartition des données et des traitements, système de cache...).

Ce travail s'inscrit dans la continuité des stages d'Aurélien BENEL (quatrième année d'école d'ingénieur² et DEA³), effectués en collaboration avec l'Ecole Française d'Archéologie d'Athènes et l'équipe de recherche Documentation et Aide à la Décision [DAD] du Laboratoire d'Ingénierie des Systèmes d'Information [LISI], de l' INSA Lyon . Le premier stage visait à étudier la faisabilité et définir les grandes orientations d'un projet de mise « en ligne » de la *Chroniques des fouilles* du *Bulletin de Correspondance Hellénique*. Notre étude a mis en évidence, entre autres, la spécificité des documents de consultation, la similarité entre un corpus documentaire et une base de connaissances, et l'intérêt de stocker des informations signifiées plutôt que des informations signifiantes et ce à différents niveaux (glyphe/caractère, typographie/structure logique, phrase/proposition). Le second stage avait pour but de modéliser la consultation de la *chronique des fouilles* et de définir un langage documentaire adapté, basé sur les relations entre concepts et permettant d'exprimer l'intersubjectivité des connaissances ainsi que les dimensions de l'archéologie (espace, temps, artefacts).

4. Bibliographie

- [AITK 89] AÏT-KACI H., et al. Efficient Implementation of Lattice Operations. In: *ACM Transactions on Programming Languages and Systems, Vol. 11, No 1 (Jan. 1989)*. p.115-146.
- [FOUR 98] FOUREL, F. *Modélisation, indexation et recherche de documents structurés*. Thèse de doctorat en informatique, Université de Grenoble 1, 1998.
- [GALL 86] GALLAY, A. *L'archéologie demain*. Belfond, 1986.
- [GARD 96] GARDIN, J.-C. « Formalisation et simulation des raisonnements ». In : *Une école pour les sciences sociales*. Ed. du Cerf, Ed. de l'Ecole des Hautes Etudes en Sciences Sociales. Paris, 1996. p.185-208.
- [PINO 97] PINON J.M, CALABRETTO S, POULLET L. « Document Semantic Model : An Experiment with Patient Medical Record ». In: *Proceedings of ICCS/IFIP Conference on Electronic Publishing : EP'97, New Models and Opportunities*. The University of Kent at Canterbury, Great Britain. April 14-16, 1997. Edited by Fytton Rowland and Jack Meadows. ICCS Press, Washington, pp.262-272. ISBN 1-891365-00-2.
- [PRED 99] PREDIGER S., WILLE R. The Lattice of Concept Graphs of a Relationally Scaled Context. To be published in: *Knowledge Science and Engineering with Conceptual Structures*, Lecture Notes in Artificial Intelligence, Springer, Berlin, 1999.
- [SOWA 92] SOWA, J.F. *Semantic Networks*. IBM Systems Research, 1992.

² Institut National des Sciences Appliquées [INSA] de Lyon.

³ DEA Informatique de Lyon [DIL] : Université Lyon 1, INSA Lyon, ENS Lyon.