

# Des arbres qui cachent des forêts : remarques sur l'organisation hiérarchique de la connaissance

**Jérôme Euzenat**

*INRIA Rhône-Alpes  
655 avenue de l'Europe, 38330 Montbonnot, France  
Jerome.Euzenat@inrialpes.fr*

De nombreuses entreprises construisent leurs sites "intranets". Dans ceux-ci, elles organisent un ensemble de documents qui est destiné à évoluer. Par soucis d'ouverture, des non-professionnels du classement peuvent y insérer leurs propres contributions. La technique de classement utilisée est alors un arbre sous les feuilles (ou les nœuds) duquel les intervenants peuvent abriter leurs contributions.

La critique faite à ces systèmes est que, le premier moment d'euphorie passé, les intervenants éprouvent de la difficulté à placer certains documents dans cette hiérarchie. Cela se conçoit : un arbre est surtout utile lorsque l'on peut poser la question cruciale permettant d'aller dans une branche ou dans une autre. Lorsque le document est également éligible pour deux branches, il y a problème.

## **Justification : métaphore physique**

Utiliser des arbres pour le classement a une justification très forte : une métaphore physique. Dans une bibliothèque, les livres ne peuvent être physiquement placés qu'à un endroit. Dans la mémoire d'un ordinateur, le document occupera une place précise en mémoire et dans un bureau une feuille volante ne peut se glisser que dans un dossier.

Bien entendu, il ne faut pas confondre le fond et l'index. Ce dernier sera par exemple un ensemble de fiches et, même si la classification est un arbre, les livres peuvent être classés dans plusieurs catégories (histoire et biographie). Il est alors nécessaire de disposer de plusieurs fiches. Cette solution implique la multiplication du support matériel même si celui-ci n'est qu'un descriptif : c'est cette distance, de la référence à l'objet même, que l'ordinateur permet d'abolir.

La métaphore du bureau a eu tant de succès qu'elle est reprise par la quasi-totalité des interfaces entre le système de fichiers d'un ordinateur et l'utilisateur. Pour sa bonne compréhension par les utilisateurs novices, elle doit respecter le modèle : un fichier est dans un et un seul dossier.

## **Correction : pistes informatiques**

Tout comme dans une bibliothèque un livre peut être indexé sous plusieurs entrées, en informatique, rien ne s'oppose à ce qu'un fichier se trouve dans plusieurs dossiers. Il est étonnant qu'au lieu d'adopter ce dispositif qui ne pose pas réellement de problèmes à la conception d'un système d'exploitation, on ait opté pour des palliatifs : liens symboliques et autres alias. La présence de ceux-ci montre, s'il en était besoin, que la dérogation à la règle de l'arbre est demandée.

Au-delà des systèmes de fichiers, rien ne s'oppose à utiliser un graphe, au sens mathématique, pour organiser un ensemble de documents. Cela permettrait d'atteindre le point de rangement recherché par l'utilisateur par des cheminements diversifiés.

Le problème informatique posé par l'utilisation des graphes est l'affichage de graphes quelconques à l'écran. Il a un corollaire ergonomique : la lecture de graphes de taille importante (que ce soit sur un écran ou sur papier) n'est pas aussi aisée que l'on pourrait le penser.

## **Proposition : l'arbre de mon point de vue ne fait pas ombrage à celui des autres**

Nous avons développé une autre organisation (mathématiquement moins puissante que celle d'un graphe quelconque) mais dont l'utilisation est beaucoup plus aisée. Il s'agit de la notion de point de vue sur un ensemble d'objets (ici de documents) [3]. Un point de vue est tout simplement un arbre. Mais l'intérêt de l'approche est qu'un même ensemble de documents peut être doté de multiples points de vue permettant à chacun d'insérer un document suivant le critère de classement (l'arbre) qui lui est le plus familier. On rencontre une notion similaire dans certains logiciels d'indexation (comme les arbres de concepts de la société Verity). Un mécanisme de passerelles permet de relier ces points de vue en introduisant l'équivalence entre deux nœuds de plusieurs arbres différents et donc de propager automatiquement un classement.

Cette notion de point de vue trouve un écho dans l'étude des stratégies de classement utilisées par les archivistes [1]. Ainsi, même en partant de l'idée que les archivistes utilisent un arbre, on obtient plusieurs arbres très faiblement dépendants fondés sur la date, l'origine (l'auteur et le fournisseur), le type et la localisation du document. À côté de ces taxonomies raisonnées, on en trouve encore d'autres fondées sur la localisation des archives dans le bâtiment (alors même que les documents sont tout bonnement classés sur les rayonnements en fonction de leur ordre d'arrivée).

Un autre avantage des points de vue est de ne pas se trouver lié à une « ontologie » mais au contraire de pouvoir évoluer avec le temps (en ajoutant de nouveaux et en abandonnant les anciens points de vue). Nous avons développé depuis plusieurs années un atelier permettant de construire automatiquement les taxonomies à partir des objets même [2].

## **Références**

[1] André Bisseret, Les archives départementales : description des fonctions d'un centre et analyse de l'activité cognitive, Rapport de fin de contrat, INRIA Rhône-Alpes, Montbonnot (FR), 1996

[2] Jérôme Euzenat, Brief overview of T-TREE: the TROPES Taxonomy building Tool, dans Philip Smith, Clare Beghtol, Raya Fidel, Barbara Kwasnik (éds.), *Avances in classification research 4*, Learning information, Medford (NJ US), 1994

[3] Olga Mariño, François Rechenmann, Patrice Uvietta, Multiple perspectives and classification mechanism in object-oriented representation, Actes 9th ECAI, Stockholm (SE), pp425-430, 1990