

---

# Marquage coopératif de documents

**Bérenger Arnaud, Mountaz Hascoët**

*LIRMM (UMR 5506)*  
*161 rue ADA*  
*34392 Montpellier Cedex 5*  
*{arnaud, mountaz}@lirmm.fr*

---

*RÉSUMÉ. Depuis quelques années les systèmes de bookmarking collaboratifs tels que blogmarks ou delicious ont rendu le marquage de documents par tag très populaire. Néanmoins les modèles de marquage existants rendent l'exploitation des tags créés encore limitée. Certains travaux ont permis d'améliorer la précision du tag notamment en hiérarchisant les tags. Dans cet article nous proposons une alternative aux modèles de marquage par tag actuels. Notre approche consiste à laisser chaque utilisateur créer sa structure personnelle de tags et de lui permettre de se coordonner partiellement ou totalement avec les autres membres. Ainsi, la création d'une hiérarchie de tags commune peut se faire de manière progressive, sans perte d'information ou dépréciation du travail individuel.*

*ABSTRACT. Collaborative tagging systems have become increasingly popular with social bookmarking systems such as delicious or blogmarks. However, tagging models are still at their early stage. Previous work proposed hierarchies in order to enhance tag precision; however building such hierarchies cooperatively remains an open issue. In this paper, we propose an alternative to existing tagging models. Our approach consists in letting users build their individual hierarchies of tags. Shared hierarchies are further created progressively from individual ones.*

*MOTS-CLÉS: marquage de documents par tags, coopération, hiérarchie de tag, folksonomies.*

*KEYWORDS: tagging systems, cooperative systems, tag hierarchies, folksonomies.*

---

## **1. Introduction**

Dans les systèmes de bookmarking collaboratifs les plus utilisés (par exemple delicious), le modèle de marquage reste assez primitif : les individus créent des tags dont les plus utilisés composent le vocabulaire commun sans aucun contrôle ni aucune coordination. La structure des tags est une simple liste, inadaptée au passage à l'échelle et limitant le pouvoir d'expression. Dans ce contexte, imposer un système de classification unique tel que cela a été pratiqué dans le cadre de la documentation ne paraît adapté ni aux médias ni aux publics visés.

Dans cet article nous commençons par un rapide examen des travaux récents liés à la construction et à l'exploitation de collections de tags, nous présentons ensuite notre modèle de marquage qui repose sur (1) des tags atomiques structurés hiérarchiquement, (2) un marquage polyhiérarchique collaboratif et (3) la génération de collections hiérarchisées de tags partagés.

## **2. État de l'art et problématique**

Une analyse comparative des systèmes de marquage est donnée dans (Hammond et al., 2005). Une taxonomie est présentée par (Marlow et al., 2006) et permet entre autres de distinguer 3 types de modèles de marquage : (1) le marquage aveugle où l'utilisateur n'a pas connaissance des tags des autres utilisateurs, (2) le marquage visible où l'utilisateur voit les tags entrés par les autres et (3) enfin le marquage suggéré où le système propose automatiquement des tags.

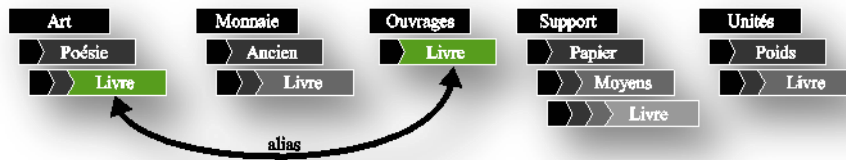
L'étude de (Golder et al., 2006) met en évidence des régularités dans l'usage des tags. Dans (Farooq et al., 2007), des critères plus précis sont proposés et mesurés dont la non-trivialité du tag ainsi que son pouvoir discriminant. Ces critères sont utilisés pour suggérer plusieurs heuristiques d'aide au marquage.

Pour l'exploitation du marquage afin d'améliorer les résultats des recherches, (Yanbee et al, 2007) proposent un moteur de recherche qui exploite des annotations extraites de systèmes de marquage collaboratif. Par ailleurs, (Zhang et al., 2006) utilisent une combinaison d'informations extraites de tags et d'ontologies pour permettre d'améliorer la recherche par tags.

Certains problèmes restent pourtant ouverts. Dans les systèmes de social bookmarking, coexistent des tags différents qui identifient pourtant les mêmes concepts, c'est un problème de vocabulaire (Furnas et al, 1987). Par ailleurs, certains tags sont polysémiques ce qui conduit à une diminution de la précision des recherches. Enfin, la construction collaborative de hiérarchies de tags reste problématique. Les approches automatiques (Wu et al, 2006) n'offrent qu'un contrôle insuffisant aux utilisateurs. Notre proposition vise à apporter quelques solutions à ces problèmes.

### 3. Tags atomiques et contextes

Construire une collection structurée de tags n'est pas aisé car il est difficile d'expliciter des notions vagues ou complexes en ayant recours à des notions canoniques, les plus simples possibles. Une hiérarchisation canonique des tags permet de résoudre en partie les problèmes d'ambiguïté de vocabulaire et de tags complexes.



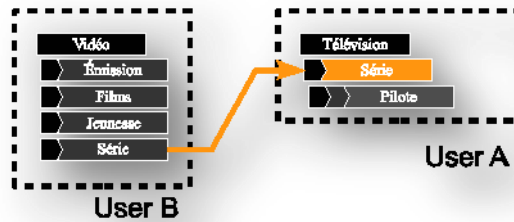
**Figure 1.** Exemple d'une structure individuelle (de l'utilisateur A) gérant plusieurs tags différents portant le même texte (gris) avec un ou plusieurs contextes (vert).

Dans les hiérarchies de tags, les nœuds sont eux mêmes des tags. Ils décrivent des notions plus fines que leurs parents mais plus générales que leurs enfants. Ces structures ne sont pas strictes : un terme peut apparaître à différents endroits de la hiérarchie (le terme multi-hiérarchie apparaît parfois pour désigner cette caractéristique).

### 4. Marquage polyhiérarchique collaboratif

Si chaque utilisateur ne se sert que de ses propres tags, le marquage perd tout son intérêt puisque confiné au seul univers de l'utilisateur. L'inverse n'est pas plus souhaitable car comme le montre une étude sur Connotea (Farooq et al, 2007) beaucoup de tags ne sont que très peu partagés.

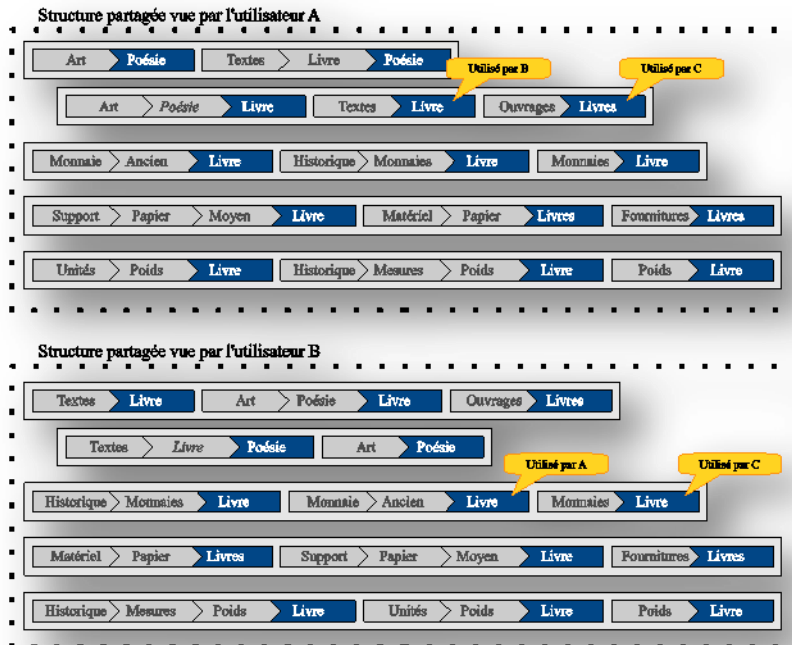
Notre approche compose donc avec ces deux constats et consiste à rendre possible la coexistence des structures hiérarchiques individuelles tout en offrant des possibilités de partage. Chaque utilisateur peut associer tout ou partie de ses tags avec ceux d'autres utilisateurs. Ces passerelles permettent de générer à terme des structures communes.



**Figure 2.** Exemple d'un partage : l'utilisateur A explicite qu'un tag de l'utilisateur B correspond à l'un des siens.

### 5. Génération de la structure commune

La structure commune de tags comprend tous les tags de tous les utilisateurs, tous leurs partages et leurs libertés structurelles. Il n'est donc plus possible de parler d'arbre, mais de graphe orienté du fait de la présence éventuelle de cycles de partages. Sa reconstruction se base sur des contextes d'apparition dans les hiérarchies des utilisateurs. Ainsi, la structure commune peut-elle prendre plusieurs visages selon le contexte utilisateur choisi pour référence lors de la génération. La figure 3 (ci-après) montre deux représentations d'une structure partagée en prenant pour référence l'utilisateur A ou l'utilisateur B, basée pour partie sur l'exemple de la figure 1, où les tags Livre et Poésie sont partagés entre les utilisateurs A, B et C.



**Figure 3.** Exemple de structures partagées vues par deux utilisateurs différents

## 6. Conclusion et perspectives

Dans cet article nous nous intéressons à plusieurs problèmes qui apparaissent dans les folksonomies aujourd'hui. Nous nous intéressons en particulier aux problèmes de vocabulaire, de la polysémie et de la construction collaborative de hiérarchies de tags. Notre approche repose sur la gestion de hiérarchies de tags individuelles, la définition explicite de passerelles entre les hiérarchies individuelles au travers de partage de tags et enfin la génération automatique de hiérarchies globales communes à partir des hiérarchies individuelles. Cette approche est actuellement en test dans un système d'annotation coopérative pour un petit groupe d'utilisateurs. Dans l'avenir nous envisageons de faire des expérimentations de plus grande taille en définissant des rôles et des relations entre les utilisateurs pour mieux cibler le partage et la génération automatique des hiérarchies communes.

## 7. Bibliographie

- Golder, S. A. and Huberman, B. A. Usage patterns of collaborative tagging systems. *J. Inf. Sci.* 32, 2 (Apr. 2006), 198-208. 2006.
- Furnas, G. W. , Landauer, T. K., Gomez, L. M. and Dumais, S. T.. The Vocabulary Problem in Human-System Communication. *Communications of the ACM*, 30(11):964–971, 1987.
- Farooq, U., Kannampallil, T. G., Song, Y., Ganoë, C. H., Carroll, J. M., and Giles, L. 2007. Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics. In *Proceedings of the 2007 international ACM Conference on Supporting Group Work (Sanibel Island, Florida, USA, November 04 - 07, 2007)*. GROUP '07. ACM, New York, NY, 351-360, 2007.
- Hammond, T., Hannay, T., Lund, B, and Scott, J., *Social Bookmarking Tools (I): A General Overview*, D-Lib Magazine, Vol 11, n°4, ISSN:1082-9873, 2005.
- Hammond, T., Hannay, T., Lund, B, and Scott, J., *Social Bookmarking Tools (II): A Case Study - Connotea*, D-Lib Magazine, Vol 11, n°4, ISSN:1082-9873 *Social Bookmarking Tools (II)*, 2005.
- Marlow, C., Naaman, M., Boyd, D., and Davis, M. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of HYPERTEXT '06*. ACM, New York, NY, 31-40, 2006.
- Yanbe, Y., Jatowt, A., Nakamura, S., and Tanaka, K. 2007. Can social bookmarking enhance search in the web?. *JCDL '07*. ACM, New York, 107-116, 2007
- Wu, X., Zhang, L., and Yu, Y. 2006. Exploring social annotations for the semantic web. In *Proceedings of the WWW '06*. ACM, New York, NY, 417-426
- Zhang, L., Wu X. and Yu, Y.. Emergent Semantics from Folksonomies: A Quantitative Study. *Journal on Data Semantics VI, LNCS 4090*, 168-186, 2006.