



**HAL**  
open science

## Construction de profils folksonomiques pour la Recherche d'Information

Jonathan van Puymbrouck, Adeline Nazarenko, Antoine Rozenknop

► **To cite this version:**

Jonathan van Puymbrouck, Adeline Nazarenko, Antoine Rozenknop. Construction de profils folksonomiques pour la Recherche d'Information. Semaine du document numérique et de la recherche d'information, Mar 2012, Bordeaux, France. pp.12. hal-00703373

**HAL Id: hal-00703373**

**<https://hal.science/hal-00703373>**

Submitted on 1 Jun 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Construction de profils folksonomiques pour la Recherche d'Information

**Jonathan van Puymbrouck, Adeline Nazarenko,  
Antoine Rozenknop**

LIPN – UMR 7030  
Université Paris 13  
99 Avenue Jean-Baptiste Clément  
93430 Villetaneuse  
{nom.prénom}@lipn.univ-paris13.fr

---

*RÉSUMÉ.* Cet article présente une méthode qui exploite les folksonomies pour prendre en compte les profils de l'utilisateur lors de sa recherche d'information. Une folksonomie est constituée d'un ensemble de relations associant un utilisateur, une ressource et le mot-clé que le premier a utilisé pour tagger la seconde. L'idée consiste à construire un recouvrement de ces relations folksonomiques par un ensemble de micro-folksonomies qui relient des groupes d'utilisateurs, un vocabulaire de tags et un ensemble de ressources et qui généralisent ainsi les relations folksonomiques initiales. Nous introduisons l'algorithme de construction des micro-folksonomies, nous expliquons son fonctionnement sur un exemple jouet puis nous présentons des expériences qui montrent son apport en termes de recherche d'information.

*ABSTRACT.* This paper presents a method that exploits a folksonomy to take user profiles into account in information retrieval. A folksonomy is made of a set of triples associating a user, a resource and the keyword that the former has used to tag the latter. Our approach consists in building a cover of the initial set of triples. It is a set of micro-folksonomies that link a group of users, a group of tags and a group of resources and thus generalize the triples of the initial folksonomy. The paper introduces the algorithm that builds micro-folksonomies, illustrates it on a toy example and presents its impact on information retrieval.

*MOTS-CLÉS :* folksonomie, profil, Recherche d'Information

*KEYWORDS:* Folksonomy, user profile, information retrieval

---

## 1. Introduction

Cet article présente une méthode permettant d'exploiter les folksonomies pour prendre en compte le profil de l'utilisateur lors de sa recherche d'information. Notre approche repose sur l'idée qu'un utilisateur a en réalité autant de profils que de centres d'intérêts, qu'il passe rapidement de l'un à l'autre et que ses centres d'intérêts évoluent dans le temps. Il faut donc pouvoir rendre compte de la pluralité des profils des utilisateurs.

Une folksonomie indique qui a consulté quels documents et associe une sémantique à la relation utilisateur-document à travers les annotations qui leurs sont associées. Notre idée consiste donc à organiser ces traces que les utilisateurs laissent sur les documents sous la forme de "tags" de façon à permettre ensuite de répondre de manière profilée aux requêtes des utilisateurs, plutôt que d'extraire de la base documentaire des ensembles indifférenciés de ressources partageant des mots-clés avec les requêtes.

Concrètement, notre analyse de la folksonomie permet de faire émerger des profils de taggeurs, sous la forme de groupes d'utilisateurs qui ont tendance à tagger de manière similaire des documents similaires, un même utilisateur pouvant appartenir à plusieurs groupes et donc avoir plusieurs profils de taggeur. Cela revient à construire un recouvrement des relations folksonomiques initiales (utilisateur-tag-ressource) par un ensemble de micro-folksonomies qui relient chacune un groupe d'utilisateurs, un vocabulaire de tags et un ensemble de ressources.

La prise en compte de ces micro-folksonomies permet de généraliser les réponses aux requêtes d'un utilisateur en retournant, au-delà des documents qu'il a lui-même taggés avec les mots-clés de la requête, les documents que les gens de son groupe ont étiquetés de cette manière-là. Les micro-folksonomies permettent aussi de documenter l'ensemble des ressources retournées en montrant les sous-classes associées à différents ensembles de mots clés.

Cet article met l'accent sur la méthode de construction des profils de taggeurs, et plus précisément des micro-folksonomies, à partir d'un jeu de données folksonomiques. La section 2 situe cette approche par rapport aux recherches, aujourd'hui nombreuses, contribuant à l'émergence du web social. Les sections 3 et 4 présentent la méthode de construction des groupes d'utilisateurs et montrent son déroulement sur un exemple. Les expériences de la section 5 présentent les premiers résultats obtenus sur un jeu de données de grande taille.

## 2. Système social de partage de ressources

Le champ de la Recherche d'Information (R.I.) sur le Web a connu ces dernières années plusieurs évolutions importantes avec l'accroissement massif du nombre de ressources disponibles et l'émergence du Web social dont les utilisateurs ne sont plus seulement lecteurs mais aussi auteurs.

L'information et les ressources étant pléthoriques, l'un des principaux problèmes des utilisateurs est de sélectionner l'information. Beaucoup de travaux en recherche d'information ont cherché à prendre en compte les profils des utilisateurs, pour améliorer la précision de leurs résultats en utilisant une analyse plus ciblée. Les profils sont calculés à partir d'informations de différentes natures : certains reposent directement sur des descriptions ontologiques (Gauch *et al.*, 2003), d'autres sont fonction de la place que les utilisateurs occupent dans un réseau social (détection de communautés) (Blondel *et al.*, 2008), d'autres encore reposent sur le partage de propriétés entre les utilisateurs (Yan *et al.*, 2005) et une dernière approche exploite la manière dont les utilisateurs manipulent des objets en contexte de collaboration et les traces qui en découlent (Griffith *et al.*, 2007).

Les systèmes qui exploitent ces différents types de profils cherchent en général à identifier les utilisateurs pour leur proposer une aide ou un service adapté. Nous poursuivons un même objectif mais en cherchant à tirer parti de l'information que les utilisateurs fournissent sous la formes de tags et de folksonomies. Les systèmes sociaux de partage de ressources, ou "folksonomies", donnent des informations sur le lien que les utilisateurs entretiennent avec des ressources pour une tâche donnée. Le site *delicious* (Wetzker *et al.*, 2008), qui propose un système social de partage de ressources étiquetées à l'aide de tags, en est un exemple bien connu. C'est sur ce système que reposent les expériences présentées plus loin.

(Hotho *et al.*, 2006) définit une folksonomie de la manière suivante :

$$F := (U, T, R, Y)$$

où  $U, T$  et  $R$  sont des ensembles finis, respectivement appelés *utilisateurs*, *tags* et *ressources* et où  $Y$  est une relation ternaire entre des ressources, des tags et des utilisateurs ( $Y \subseteq U \times T \times R$ ). Les composants élémentaires de la folksonomie sont donc les triplets  $\{u, t, r\}$  de  $Y$ , tels que  $u \in U, t \in T, r \in R$  et peuvent être organisés sous la forme d'un graphe non orienté en reliant par des arcs ceux qui partagent des éléments communs dans  $U$  et/ou  $T$  et/ou  $R$ . Dans le cadre de la R.I., (Hotho *et al.*, 2006) exploite alors ce graphe d'annotations en y appliquant son algorithme FolkRank, inspiré du célèbre PageRank (Brin *et al.*, 1998) qui, lui, repose sur le graphe des hyperliens.

(Zhou *et al.*, 2008) introduit, avec la notion de "facettes", le fait qu'un utilisateur dans le web social puisse être caractérisé par plusieurs profils différents. Pour démontrer l'aspect sémantique des folksonomies, (Jäschke *et al.*, 2006) propose une méthode dérivée de l'Analyse Formelle de Concepts, qui permet d'identifier des sous-ensembles cohérents de la folksonomie initiale et de les organiser en un graphe orienté. (Angeletou *et al.*, 2007) propose une autre approche qui en utilisant les ressources du Web Sémantique et les tags des folksonomies permet de faire émerger une sémantique capable de générer des ensembles de termes cohérents.

Notre approche s'inspire des mêmes intuitions selon lesquelles les utilisateurs taggent en suivant différents profils et que la folksonomie véhicule une information

sémantique. Mais au-delà du tri des résultats (Hotho *et al.*, 2006), nous exploitons les folksonomies dans le processus même de recherche d'information.

### 3. Méthode

#### 3.1. Principe

Nous souhaitons utiliser les profils des utilisateurs dans la recherche d'information mais en tenant compte de leur pluralité et de leur plasticité, les utilisateurs ayant différents centres d'intérêts, passant de l'un à l'autre pour revenir ensuite au premier et en changeant au fil du temps (Yeung *et al.*, 2008).

Les approches déclaratives, dans lesquelles les utilisateurs saisissent eux-mêmes les caractéristiques de leurs profils, permettent difficilement de prendre en compte la plasticité de ceux-ci. Inversement, les approches par induction représentent de manière uniforme les différents profils des utilisateurs et ne permettent pas d'en mesurer la diversité. L'intérêt de l'approche folksonomique vient de ce qu'elle exploite les traces des utilisateurs et la sémantique de ces traces, laissée sous forme de tags, pour faire ressortir différentes facettes d'un même utilisateur.

L'objectif de la méthode de détection des micro-folksonomies consiste à découvrir dans la folksonomie des groupes d'utilisateurs qui ont tendance à tagger de la même manière (avec des mots-clés identiques ou similaires) des documents similaires. On substitue aux relations de taggage de la folksonomie initiale  $\{u, t, r\}$  où  $u$ ,  $t$  et  $r$  sont respectivement un utilisateur, un tag et une ressource, des relations  $\{Groupe_U, Vocabulaire_T, Collection_R\}$  dites "micro-folksonomies", associant un groupe d'utilisateurs, un vocabulaire de tags et une collection de ressources.

Il faut souligner tout d'abord qu'un utilisateur, un tag ou une ressource peut appartenir à plusieurs micro-folksonomies, lesquelles ne sont donc pas disjointes les unes des autres. Cela représente le fait qu'un utilisateur peut avoir plusieurs centres d'intérêts, qu'un tag peut être utilisé de diverses manières et qu'une ressource est elle-même une source d'informations variées. Il faut également remarquer que l'ensemble de ces micro-folksonomies généralise la folksonomie initiale dans la mesure où l'existence de la micro-folksonomie  $\{Groupe_U, Vocabulaire_T, Collection_R\}$  n'implique pas que toutes les relations de taggage élémentaires  $\{u, t, r\} \in Groupe_U \times Vocabulaire_T \times Collection_R$  soient exprimées dans la folksonomie initiale.

#### 3.2. Etapes

La construction des micro-folksonomies se fait en trois étapes qui étudient à tour de rôle un couple d'éléments en relation dans la folksonomie par rapport à la troisième dimension soit la folksonomie  $F := (U, T, R, Y)$ . La relation  $Y \subset U \times T \times R$  est une relation ternaire qui lie un utilisateur de  $U$ , un tag de  $T$  et un document de  $R$ . Pour tout sous-ensemble  $E$  de  $Y$ , on note  $P_X(E)$  la projection de  $E$  sur  $X$ , où  $X$  vaut  $U$ ,

$T$ ,  $R$  ou un produit de ces ensembles. Notre algorithme repose sur la construction de sous-ensembles  $Y_i$  de  $Y$ , qui en constituent un recouvrement. A l'issue de la dernière étape, on considérera que pour chaque sous-ensemble  $Y_i$ , on a construit une micro-folksonomie  $\{P_U(Y_i), P_T(Y_i), P_R(Y_i)\}$ .

### 3.2.1. Etape 1

L'étape 1 rassemble les utilisateurs qui ont taggé similairement un même document. Formellement, on considère séparément chaque couple  $(t, r) \in (P_{T \times R}(Y))$ , et on définit pour chacun un sous-ensemble  $Y_{(t,r)}^1$  de  $Y$ , constitué de tous les triplets de  $Y$  contenant  $(t, r)$  :

$$Y_{(t,r)}^1 = P_{T \times R}^{-1}(\{(t, r)\})$$

### 3.2.2. Etape 2

A l'étape 2, on étend chaque ensemble  $Y_{(t,r)}^1$  indépendamment des autres, pour obtenir un ensemble  $Y_{(t,r)}^2$ . Cette extension se fait en deux temps. On ajoute d'abord (étape 2a) à  $Y_{(t,r)}^1$  tous les triplets de  $Y$  dont le couple (utilisateur, tag) est contenu dans  $Y_{(t,r)}^1$  ; autrement dit, on construit l'ensemble  $Y_{(t,r)}^{2,\alpha}$  suivant :

$$Y_{(t,r)}^{2,\alpha} = \{(u', t', r') \in Y \setminus (u', t') \in P_{U \times T}(Y_{(t,r)}^1)\} = P_{U \times T}^{-1}(P_{U \times T}(Y_{(t,r)}^1))$$

On ajoute ensuite (étape 2b) au nouvel ensemble obtenu tous les triplets de  $Y$  dont le couple (utilisateur, ressource) est contenu dans  $Y_{(t,r)}^{2,\alpha}$ , i.e. :

$$Y_{(t,r)}^2 = \{(u', t', r') \in Y \setminus (u', r') \in P_{U \times R}(Y_{(t,r)}^{2,\alpha})\} = P_{U \times R}^{-1}(P_{U \times R}(Y_{(t,r)}^{2,\alpha}))$$

L'étape 2 ne crée pas de nouveaux sous-ensembles par rapport à ceux de l'étape 1 mais elle étend le nombre de ressources et de tags de chaque ensemble.

### 3.2.3. Etape 3

La troisième étape fusionne les sous-ensembles issus de l'étape 2 et considérés comme proches. Deux sous-ensembles sont fusionnés lorsqu'ils partagent strictement le même vocabulaire de tags et qu'ils ont en outre des utilisateurs et des ressources en commun. Ce processus est répété tant que des fusions sont possibles.

Formellement, on considère l'ensemble  $\mathcal{E}$  des  $Y_{(t,r)}^2$  comme les sommets d'un graphe non-orienté  $\mathcal{G} = (\mathcal{E}, \mathcal{A})$ , où  $\mathcal{A}$  est défini ainsi :

$$(Y_{(t,r)}^2, Y_{(t',r')}^2) \in \mathcal{A} \Leftrightarrow \begin{cases} P_T(Y_{(t,r)}^2) = P_T(Y_{(t',r')}^2) \\ \text{et } P_U(Y_{(t,r)}^2) \cap P_U(Y_{(t',r')}^2) \neq \emptyset \\ \text{et } P_R(Y_{(t,r)}^2) \cap P_R(Y_{(t',r')}^2) \neq \emptyset \end{cases}$$

Pour chaque composante connexe  $\mathcal{C}_i$  de  $\mathcal{G}$ , on définit l'ensemble  $Y_i$  ainsi :

$$Y_i = \cup_{Y_{(t,r)}^2 \in \mathcal{C}_i} Y_{(t,r)}^2$$

L'ensemble des  $Y_i$  obtenus forme le *recouvrement* que l'on cherche à étudier. Chacun des  $Y_i$  constitue une micro-folksonomie  $\{P_U(Y_i), P_T(Y_i), P_R(Y_i)\}$ .

#### 4. Exemple

Pour illustrer les différentes étapes de cet algorithme, nous considérons la folksonomie représentée par l'ensemble des triplets figurant en ligne dans le tableau 1. Formellement, cette folksonomie  $F$  est définie par  $F = \{U, T, R, Y\}$ , avec  $U = \{u1, u2, u3, u4\}$ ,  $T = \{php, zend, design, art, peinture\}$ ,  $R = \{doc1, doc2, doc3, doc4\}$  et  $Y$  l'ensemble des triplets de départ.

id	Utilisateur	Tag	Ressource
1	u1	php	doc1
2	u2	php	doc1
3	u2	php	doc2
4	u2	zend	doc2
5	u2	design	doc3
6	u3	php	doc3
7	u3	art	doc4
8	u4	peinture	doc4

**Tableau 1.** Ensemble  $Y$  des triplets de la folksonomie-jouet

Le tableau 2 liste les ensembles obtenus à l'issue de l'étape 1 de l'algorithme : les triplets 1 et 2 ont été regroupés dans un même ensemble du fait qu'ils partagent le même tag (php) et la même ressource (doc1) ; les autres triplets se retrouvent isolés dans leurs ensembles respectifs.

Ensemble	id	Utilisateur	Tag	Ressource
$Y_{(php,doc1)}^1$	1	u1	php	doc1
	2	u2	php	doc1
$Y_{(php,doc2)}^1$	3	u2	php	doc2
$Y_{(zend,doc2)}^1$	4	u2	zend	doc2
$Y_{(design,doc3)}^1$	5	u2	design	doc3
$Y_{(php,doc3)}^1$	6	u3	php	doc3
$Y_{(art,doc4)}^1$	7	u3	art	doc4
$Y_{(peinture,doc4)}^1$	8	u4	peinture	doc4

**Tableau 2.** Ensembles créés à partir de la l'exemple de folksonomie à l'issue de la première étape de l'algorithme

Le tableau 3 liste les ensembles obtenus après l'étape 2a : le triplet 3 a été ajouté au premier ensemble, car il partage l'utilisateur u2 et le tag php avec le triplet 2, qui en faisait déjà partie. Réciproquement, le triplet 2 a été ajouté au second ensemble, auquel le triplet 3 appartenait déjà.

Ensemble	id	Utilisateur	Tag	Ressource
$Y^{2,\alpha}$ ( <i>php,doc1</i> )	1	u1	php	doc1
	2	u2	php	doc1
	3	u2	php	doc2
$Y^{2,\alpha}$ ( <i>php,doc2</i> )	2	u2	php	doc1
	3	u2	php	doc2
$Y^{2,\alpha}$ ( <i>zend,doc2</i> )	4	u2	zend	doc2
$Y^{2,\alpha}$ ( <i>design,doc3</i> )	5	u2	design	doc3
$Y^{2,\alpha}$ ( <i>php,doc3</i> )	6	u3	php	doc3
$Y^{2,\alpha}$ ( <i>art,doc4</i> )	7	u3	art	doc4
$Y^{2,\alpha}$ ( <i>peinture,doc4</i> )	8	u4	peinture	doc4

**Tableau 3.** Ensembles créés à partir de l'exemple de folksonomie à l'issue de l'étape 2a de l'algorithme

Lors de l'étape 2b, le triplet 4 est ajouté aux deux premiers ensembles, et le triplet 3 est ajouté au troisième, du fait qu'ils partagent le même utilisateur (u2) et le même document (doc2). Le résultat est illustré par le tableau 4.

Enfin, lors de l'étape 3, illustrée par le tableau 5, les trois premiers ensembles de l'étape 2 fusionnent car ils ont le même ensemble de tags ( $\{\text{php, zend}\}$ ) et partagent deux à deux un utilisateur (u2) et une ressource (doc2).

L'algorithme construit donc un recouvrement de la folksonomie initiale et permet de construire les 5 micro-folksonomies du tableau 6.

Sur cet exemple-jouet, le pouvoir de généralisation peut être illustré dans  $F_1$  par l'apparition de nouveaux liens, comme le lien entre l'utilisateur u1 et le document doc2. Les expériences sur des jeux de données de taille réelle permettent de mieux mesurer cet effet.

## 5. Expériences

Les expériences présentées dans cette section visent à analyser la structure du recouvrement obtenu et son impact en termes de recherche d'information. Un jeu de données a été constitué par sélection aléatoire de 4 997 triplets dans la collection de *delicious.com* (Wetzker *et al.*, 2008). La sélection aléatoire a pour conséquence que la densité de la folksonomie-échantillon obtenue (en nombre de triplets par utilisateur,



Ensemble	id	Utilisateur	Tag	Ressource
$Y_{(php,doc1)}^2$	1	u1	php	doc1
	2	u2	php	doc1
	3	u2	php	doc2
	4	u2	zend	doc2
$Y_{(php,doc2)}^2$	2	u2	php	doc1
	3	u2	php	doc2
	4	u2	zend	doc2
$Y_{(zend,doc2)}^2$	3	u2	php	doc2
	4	u2	zend	doc2
$Y_{(design,doc3)}^2$	5	u2	design	doc3
$Y_{(php,doc3)}^2$	6	u3	php	doc3
$Y_{(art,doc4)}^2$	7	u3	art	doc4
$Y_{(peinture,doc4)}^2$	8	u4	peinture	doc4

**Tableau 4.** Ensembles créés à partir de l'exemple de folksonomie à l'issue de l'étape 2b de l'algorithme

Ensemble	id	Util.	Tag	Ress.
$Y_1 = Y_{(php,doc1)}^2 \cup Y_{(php,doc2)}^2 \cup Y_{(zend,doc2)}^2$	1	u1	php	doc1
	2	u2	php	doc1
	3	u2	php	doc2
	4	u2	zend	doc2
$Y_2 = Y_{(design,doc3)}^2$	5	u2	design	doc3
$Y_3 = Y_{(php,doc3)}^2$	6	u3	php	doc3
$Y_4 = Y_{(art,doc4)}^2$	7	u3	art	doc4
$Y_5 = Y_{(peinture,doc4)}^2$	8	u4	peinture	doc4

**Tableau 5.** Ensembles créés à partir de l'exemple de folksonomie à l'issue de l'étape 3 de l'algorithme

tag ou ressource) est moins grande que celle de la collection complète, ce qui limite *a priori* le pouvoir de généralisation de la construction des micro-folksonomies.

### 5.1. Pouvoir de généralisation des micro-folksonomies

En tant que service de partage de ressources, une simple folksonomie ne permet de retrouver en réponse à une requête  $R$  d'un utilisateur  $u$  que les ressources qui

Micro-folksonomies	Utilisateurs	Tags	Ressources
$F_1$	{u1, u2}	{php, zend}	{doc1, doc2}
$F_2$	{u2}	{design}	{doc3}
$F_3$	{u3}	{php}	{doc3}
$F_4$	{u3}	{art}	{doc4}
$F_5$	{u4}	{peinture}	{doc4}

**Tableau 6.** *Micro-folksonomies créées à partir de l'exemple-jouet*

ont été taggées avec des mots-clés de  $R$  soit par  $u$  lui-même (résultat trivial) soit par l'un quelconque des utilisateurs (résultat généralisé). Selon la façon de considérer les micro-folksonomies produites par notre algorithme, on peut définir d'autres niveaux de généralisation : si l'on considère le recouvrement de la folksonomie initiale construit par notre algorithme (ensembles  $Y_1$  à  $Y_5$  de l'exemple-jouet), on peut définir un niveau de généralisation intermédiaire : la requête php de l'utilisateur u1, qui n'est que dans  $Y_1$ , renverra doc1 et doc2, mais pas doc3 ; si l'on considère la micro-folksonomie  $F_1$  plutôt que le sous-ensemble  $Y_1$ , on définit au contraire un niveau de généralisation plus élevé : la requête zend aura pour résultat les documents doc1 et doc2, alors qu'aucun utilisateur n'a jamais utilisé ce tag pour doc1. C'est ce niveau de généralisation le plus élevé que nous tentons de mesurer ici.

Pour observer le pouvoir de généralisation des micro-folksonomies, nous avons appliqué notre algorithme sur un échantillon de triplets tirés aléatoirement de la collection de *delicious.com* de 2003.

**Données de départ** La folksonomie-échantillon initiale est composée de 4 997 triplets et comporte ;

- 557 utilisateurs, qui apparaissent en moyenne dans 8,9 triplets, au minimum dans 1 triplet et au maximum dans 264 triplets ;
- 1 464 tags, qui apparaissent en moyenne dans 3,4 triplets, au minimum dans 1 triplet et au maximum dans 391 triplets ;
- 2 271 urls (ressources), qui apparaissent en moyenne dans 2,2 triplets, au minimum dans 1 triplet et au maximum dans 30 triplets.
- 2 944 relations binaires (utilisateur, tag) ;
- 2 375 relations binaires (utilisateur, ressource) ;
- 4 906 relations binaires (tag, ressource).

**Données résultat** A partir de la folksonomie ci-dessus, l'algorithme produit un total de 1 229 micro-folksonomies comprenant :

- 1 229 groupes d'utilisateurs comptant entre 1 et 59 utilisateurs et en moyenne 2,5 utilisateurs par groupe ;

Jonathan van Puymbrouck, Adeline Nazarenko, Antoine Rozenknop

- 1 263 groupes de tags comptant entre 1 et 189 tags et en moyenne 3,5 tags par groupe ;
- 1 225 groupes de ressources comptant entre 1 et 10 ressources et en moyenne 1,5 ressources par groupe.

Pour mesurer le degré de généralisation obtenu, nous avons pris chacun des tags de la folksonomie initiale comme requête et mesuré le nombre de documents retournés en moyenne quand ces requêtes sont lancées sur la folksonomie initiale et sur les micro-folksonomies calculées à partir de cette dernière. Sur les 1 464 requêtes ainsi lancées, nous obtenons les valeurs moyennes suivantes :

- sur la folksonomie initiale, un tag est utilisé 3,4 fois par 1,7 utilisateurs et permet de retrouver 3,3 ressources distinctes ;
- avec les micro-folksonomies, un tag est utilisé par 3,5 groupes d'utilisateurs pointant vers 7,9 ressources distinctes.

Exploiter les micro-folksonomies permet donc d'étendre significativement le nombre de ressources retournées, puisqu'on passe de 3,3 ressources retournées en moyenne à 7,9<sup>1</sup>.

Considérons par exemple le mot-clé *computers*. Il est utilisé 44 fois par 6 utilisateurs pour tagger 44 ressources différentes dans la folksonomie initiale, mais lancé comme requête sur les micro-folksonomies, il retourne 92 ressources différentes. On observe ainsi que, sur la base folksonomique considérée, *computers* est un terme assez générique<sup>2</sup> qui permet d'extraire de nombreux documents.

## 5.2. Typage sémantique

Quand on considère l'ensemble des ressources retournées en réponse à une requête  $R$ , on observe généralement qu'elles proviennent de différentes micro-folksonomies qui contiennent toutes les mots-clés de  $R$  dans leur ensemble de tags.

Ces différentes micro-folksonomies représentent un recouvrement de l'ensemble des ressources retournées et on peut étiqueter chacun des sous-ensembles de ressources ainsi constitués par le groupe des tags qui lui est lié dans la micro-folksonomie dont il relève. S'il est illusoire de chercher à évaluer la qualité des ensembles de ressources (en pratique, il s'agit d'url dans nos expériences sur *delicious.com*), on peut néanmoins observer les groupements de tags qui sont construits.

Le tableau 7 en donne quelques exemples pour les requêtes *Linux* et *Comics*. Malgré des intersections non vides deux à deux, on observe que les différents ensembles de tags reflètent des volets sémantiques assez différenciés.

---

1. Rappelons que le degré de généralisation est *a priori* plus faible sur un échantillon de triplets tiré aléatoirement que sur une folksonomie "naturelle".

2. On obtient près de 1,8 millions de résultats en l'utilisant comme requête dans google.

Linux			Comics	
a	b	c	c	d
firefox	dev	email	character	comics
linux	email	filtering	comics	entertainment
mozilla	linux	guides	generator	humour
plugin	mail	linux	police	
software	php	mail	publisher	
	programming	procmai	rpg	
	sendmail	software	sas	
		spam	topten	
		tips		

**Tableau 7.** Exemples de groupements de tags associés à différents sous-ensembles de ressources retrouvées à partir des requêtes *Linux* et *Comics*.

## 6. Conclusion

Cet article présente une méthode qui prend en compte les différents profils des utilisateurs pour retrouver des ressources dans une folksonomie. Cela permet de retrouver non seulement les ressources qu'un utilisateur  $u$  a taggées mais aussi celles que les autres utilisateurs de même profil ont taggées. Techniquement, nous proposons de calculer un recouvrement de l'ensemble des relations folksonomiques initiales et nous obtenons un ensemble de micro-folksonomies qui généralisent les relations de départ à des groupes d'utilisateurs de profils similaires, à des vocabulaires de tags sémantiquement proches et à des ensemble de ressources similaires.

L'analyse de la structure des micro-folksonomies obtenues sur un jeu de données construit à partir de *delicious.com* met en évidence le pouvoir de généralisation de cette approche : en tenant compte des similarités folksonomiques entre utilisateurs, entre tags et entre ressources, on peut généraliser les requêtes sur une base sémantique et doubler en moyenne le nombre de résultats obtenus par requête. Enfin l'analyse des groupements de tags obtenus dans différentes micro-folksonomies montre que celles-ci reposent sur des différences sémantiques.

## 7. Bibliographie

- Angeletou S., Sabou M., Specia L., Motta E., « FolksOntology : An Integrated Approach for Turning Folksonomies into Ontologies », *Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007)*, p. 57-70, 2007.
- Blondel V. D., Guillaume J.-L., Lambiotte R., Lefebvre E., « Fast unfolding of communities in large networks », *Journal of Statistical Mechanics : Theory and Experiment*, vol. 2008, n° 10, p. P10008+, July, 2008.

Jonathan van Puymbrouck, Adeline Nazarenko, Antoine Rozenknop

- Brin S., Page L., « The Anatomy of a Large-Scale Hypertextual Web Search Engine », *Computer Networks and ISDN Systems*, Elsevier, p. 107-117, 1998.
- Gauch S., Chaffee J., Pretschner A., « Ontology-Based User Profiles for Search and Browsing », *Web Intelligence and Agent Systems*, vol. 1, n° 3-4, p. 219-234, 2003.
- Griffith J., O’Riordan C., Sorensen H., « Identifying User and Group Information from Collaborative Filtering Datasets », *IJPRAI*, vol. 21, n° 2, p. 291-310, 2007.
- Hotho A., Jäschke R., Schmitz C., Stumme G., « Information Retrieval in Folksonomies : Search and Ranking », in , Y. Sure, , J. Domingue (eds), *The Semantic Web : Research and Applications*, vol. 4011 of *LNAI*, Springer, Heidelberg, p. 411-426, 2006.
- Jäschke R., Hotho A., Schmitz C., Ganter B., Stumme G., « TRIAS - An Algorithm for Mining Iceberg Tri-Lattices », *ICDM '06 : Proceedings of the Sixth International Conference on Data Mining*, IEEE Computer Society, Washington, DC, USA, p. 907-911, 2006.
- Wetzker R., Zimmermann C., Bauckhage C., « Analyzing Social Bookmarking Systems : A del.icio.us Cookbook », *Mining Social Data (MSoDa) Workshop Proceedings*, ECAI 2008, p. 26-30, July, 2008.
- Yan X., Cheng H., Han J., Xin D., « Summarizing itemset patterns : a profile-based approach », *In Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, p. 314-323, 2005.
- Yeung C. M. A., Gibbins N., Shadbolt N., « A Study of User Profile Generation from Folksonomies », *Social Web and Knowledge Management, Social Web 2008 Workshop at WWW2008*, March, 2008.
- Zhou C., Béné A., « From the crowd to communities : New interfaces for social tagging », in , P. Hassanaly, , A. Ramrajsingh, , D. Randall, , P. Salembier, , M. Tixier (eds), *Proceedings of the eighth international conference on the design of cooperative systems (COOP'08), Carry-le-Rouet, May 20-23, 2008*, Institut d’Études Politiques d’Aix en Provence, p. 242-250, 2008.