
***BGR*T : une nouvelle base générique de règles d'association triadiques**

Application à l'autocomplétion de requêtes dans les folksonomies

Chiraz Trabelsi¹, Nader Jelassi¹, Sadok Ben Yahia^{1,2}

1. Faculty of Sciences of Tunis, University Tunis El-Manar, 2092 Tunis, Tunisie

2. Institut Telecom SudParis, UMR 5157 CNRS Samovar, 91011 Evry Cedex, France

RÉSUMÉ. Le tagging social s'est récemment imposé dans le paysage du web collaboratif (Web 2.0) comme un support à l'organisation des ressources partagées, permettant aux utilisateurs de catégoriser leurs ressources en leur associant des mots clefs, appelés tags. La structure ainsi créée, baptisée sous le nom de folksonomie, est assimilée à un hypergraphe triparti d'utilisateurs, de tags et de ressources. Dans ce papier, nous exploitons ces triplets pour introduire une nouvelle définition d'une base générique de règles d'association triadiques, appelée *BGR*T. Nous montrons que l'utilisation de ces règles génériques pour l'autocomplétion de requêtes permet de mettre en exergue la pertinence des folksonomies et leur intérêt réel pour la recherche d'information. Les premiers résultats obtenus sur une folksonomie réelle s'avèrent prometteurs et ouvrent de nombreuses perspectives.

ABSTRACT. The development of social tagging systems on the Web under the participatory movement often called "Web 2.0", has enabled the emergence of new forms of indexing web content, allowing users to categorize their resources by associating their keywords, called tags. Considered as tripartite hypergraphs of users, tags and resources, the resulting structures, aka folksonomies, are of great interest in information retrieval. In this paper, we exploit these triples (resources, users, tags) to introduce a new definition of a generic basis of triadic association rules, called *BGR*T. We show that the use of generic rules for autocompletion feature highlights the relevance of folksonomies and their real interest in the case of information retrieval. The first results obtained on a real folksonomy are promising and offer many opportunities.

MOTS-CLÉS : folksonomie, autocomplétion de requêtes, base générique.

KEYWORDS : folksonomy, query autocompletion, generic basis.

DOI:10.3166/DN.15.1.101-124 © 2012 Lavoisier

Document numérique – n° 1/2012, 101-124

1. Introduction

L'exploitation des réseaux sociaux et des systèmes collaboratifs dans le cadre d'un processus de recherche d'information est un phénomène qui peu à peu est intégré pratiques de la recherche sur Internet. Ces systèmes, émergents de l'ère du Web 2.0, permettent aux participants de collaborer via l'indexation libre du contenu à l'aide de mots-clés, les tags ; créant ainsi des structures triparties représentées comme des hypergraphes d'utilisateurs, de tags et de ressources, appelées folksonomies. Favorisant ainsi l'échange et le partage de ressources, l'exploitation des folksonomies s'avère d'un grand intérêt pour la recherche d'information (RI). Cependant, la démarche de recherche dans les folksonomies diffère des stratégies de recherche de la traditionnelle médiation des moteurs de recherches. En effet, un système de recherche d'information (SRI) typique retourne la même liste de résultats pour une même requête soumise par des utilisateurs ayant pourtant des besoins différents d'information. Par exemple, en soumettant la requête "Java", certains utilisateurs s'intéressent à retrouver des résultats traitant du langage de programmation "Java", tandis que d'autres s'intéressent à retrouver des résultats traitant de "l'île de java". En effet, la pertinence des résultats retournés par de tels systèmes dépend étroitement de la précision des requêtes formulées par les usagers.

Bien que la plupart des SRI modernes exploitent des mécanismes d'autocomplétion de requêtes permettant d'enrichir la requête initiale de l'utilisateur avec de l'information supplémentaire, ces derniers s'avèrent inappropriés au contexte du web collaboratif. En effet, ces mécanismes d'autocomplétion se basent généralement sur la cooccurrence relevée, à partir des recherches précédentes, du concept à rajouter à la requête initiale, *i.e.*, si le concept *language* cooccure fréquemment avec le concept *java* lors des précédentes recherches effectuées par le SRI, alors *language* sera suggéré pour l'autocomplétion de la requête *java*, sans aucune considération des relations sémantiques ou encore de l'aspect social et comportemental des utilisateurs.

Afin de pallier ces limites nous proposons dans cet article d'introduire un nouveau processus d'autocomplétion de requêtes basé sur l'exploitation des règles d'association triadiques, *i.e.*, intégrant à la fois les trois dimensions de la folksonomie (utilisateurs, ressources et tags), extraites à partir de conceptualisations partagées et consensuelles, formellement représentées, par les concepts triadiques. De manière générale, nous proposons dans ce papier une approche d'autocomplétion de requêtes qui s'articule en trois étapes :

1. Extraction des règles d'association triadiques à partir des concepts triadiques de la folksonomie. En effet, les concepts triadiques, issus de l'analyse formelle des concepts (AFC), permettent de découvrir les sous-ensembles d'utilisateurs de la folksonomie partageant la même conceptualisation sur les mêmes ressources.

2. Définition et extraction de la base générique des règles d'association triadiques. Cette base repose sur l'extraction d'un sous-ensemble générique de toutes les règles d'association triadiques par le biais d'un système axiomatique valide et complet.

3. Exploitation de la base générique des règles d'association triadiques pour l'autocomplétion de requêtes. En effet, l'objectif de l'autocomplétion est d'ajouter des informations complémentaires et utiles à la requête d'un utilisateur.

Le reste de l'article est organisé comme suit. La section 2 dresse un survol des principales approches de suggestion et de recommandation de requêtes dans les folksonomies. Nous consacrons la section 3 pour introduire les notions de l'extraction des règles d'association triadiques et leur connexion avec l'analyse formelle des concepts. La section 4 est dédiée à la définition ainsi qu'à l'extraction d'une nouvelle base informative de règles d'association triadiques, appelée *BGR_T*. Nous proposons également un mécanisme d'inférence par le biais d'un système axiomatique valide et complet, permettant de dériver toutes les règles valides pouvant être extraites à partir d'une folksonomie. Suit la section 5 qui décrit le processus d'autocomplétion de requêtes par *BGR_T* argumenté par un exemple illustratif. Les résultats des expérimentations menées sur un jeu de données réel sont présentés dans la section 6. La conclusion et les travaux futurs font l'objet de la section 7.

2. Travaux connexes

L'exploitation des folksonomies dans le cadre d'un processus de recherche d'information est un phénomène qui peu à peu s'intègre aux pratiques de la recherche sur le web. Cependant les systèmes de recherche traditionnels, bien que souvent guidés par des algorithmes du domaine de l'extraction de connaissances des bases de données, ne semblent exploiter que partiellement les apports potentiels d'une telle source d'information. En effet, les systèmes de recherche traditionnels négligent l'aspect social et comportemental des utilisateurs.

Plusieurs approches ont alors été proposées. Certaines tentent d'étendre les approches de recherche traditionnelles par l'intégration de modules complémentaires capables de garder la trace sociale et comportementale des usagers à travers la proposition de solutions prometteuses reflétées par les systèmes de recommandations (Garg, Weber, 2008 ; Karen *et al.*, 2008 ; Amer-Yahia *et al.*, 2008). L'objectif de ces systèmes est de filtrer un flux entrant d'informations (ressources) de façon personnalisée pour chaque usager, tout en s'adaptant en permanence au besoin d'information de chacun. Pour cela, les moteurs de ces systèmes gèrent des profils d'utilisateurs permettant de choisir quelles ressources à transmettre à chacun, et adaptent ces profils au cours du temps en exploitant au mieux le retour de pertinence que les utilisateurs fournissent sur les ressources reçues.

D'autres travaux proposent d'adapter les techniques de suggestion ou de complétion de requêtes sur le web à la complétion des requêtes sur les *folksonomies*. Généralement basées sur les techniques de fouille des fichiers logs, ces techniques exploitent l'hypothèse de dualité entre l'activité d'annotation des ressources par des tags et celle de la recherche des ressources via les tags. En effet, (Krause *et al.*, 2008 ; Hotho, 2010 ; Bischoff *et al.*, 2008) ont mis en exergue que le comportement de tagging d'un usager d'une folksonomie reflète étroitement son comportement de recherche. Ainsi,

si un utilisateur annote une ressource r avec un tag t , ce dernier choisira d'accéder à la ressource r si elle apparaît dans le résultat de la recherche obtenue par la soumission de t comme requête.

Par ailleurs, Schmitz *et al.*, (2006) ont proposé d'appliquer les techniques usuelles d'extraction des règles d'association (Agrawal *et al.*, 1993) par projection de la folksonomie sur deux dimensions diadiques, *i.e.*, utilisateurs-tags d'un côté et tags-ressources de l'autre. Ils ont ainsi mis en exergue la pertinence de l'utilisation des règles d'association pour la suggestion de tags et la recommandation de ressources (Schmitz *et al.*, 2006). Cependant, mis à part leur nombre considérable, les règles ainsi obtenues ne permettent pas d'exploiter la structure tripartite des folksonomies, *i.e.*, utilisateurs, ressources et tags. Le changement de dimension causera inéluctablement une perte d'information induisant un manque de sémantique des règles extraites. Considérons par exemple la dimension utilisateurs-tags, les règles extraites, *i.e.*, $utilisateur_1 \rightarrow tag_1$, ne portent aucune information sur les ressources annotées avec ces tags. Par conséquent, les requêtes ou tags suggérés amènent l'utilisateur à relancer son processus de recherche afin d'accéder aux ressources correspondantes. Néanmoins, les auteurs dans, (Nguyen *et al.*, 2010 ; 2011), ont introduit un nouvel algorithme d'extraction de règles inter-dimensionnelles, appelé GEAR, afin de généraliser le problème d'extraction des règles d'association sur une dimension n -aire. Ils proposent ainsi des implications entre des sous-ensembles appartenant à différentes dimensions. Toutefois, mis à part le grand nombre de règles redondantes, le problème de la perte d'information demeure puisque ces règles ne concernent que deux dimensions à la fois. Bien que ces approches pourraient fournir une opportunité intéressante pour l'amélioration du processus d'autocomplétion de requêtes par l'intégration de la dimension sociale à travers l'exploitation des profils utilisateurs ou l'extraction des règles d'association, aucune d'entre elles ne considère les implications ternaires entre tags qui tiennent compte de la structure triadique de la folksonomie.

Loin des promesses affichées, les règles d'association pouvant être extraites à partir de folksonomies réelles présentent un handicap de taille en vue de leur exploitation pour l'autocomplétion de requêtes. En effet, il semble évident qu'un nombre impressionnant et inexploitable de règles d'association dont un large nombre de règles redondantes sera généré (BenYahia, Nguifo, 2004 ; Pasquier, 2000). Noyé dans cette masse de connaissances, le processus d'autocomplétion ne peut tirer profit de cette connaissance qui ne présente aucune valeur ajoutée. Dans le cadre des contextes dyadiques, (Bastide *et al.*, 2000 ; Kryszkiewicz, 2002 ; BenYahia *et al.*, 2009a) ont puisé dans les résultats issus de l'analyse formelle des concepts (AFC) pour réduire "efficacement" le nombre de règles extraites. Les auteurs ont ainsi proposé l'extraction d'un ensemble minimal de règles d'association (ou base générique) présenté à l'utilisateur tout en maximisant la quantité d'information utile véhiculée. Fortement inspiré par ces travaux, nous introduisons dans cet article une nouvelle approche de sélection sans perte d'information des règles d'association triadiques. L'approche proposée repose sur l'extraction d'"un noyau irréductible" de règles d'association triadiques qui véhiculent le maximum de connaissances utiles, appelé *base générique de règles d'association triadiques*. Nous exploiterons cette base générique afin de proposer un

processus d'autocomplétion efficace couvrant au mieux les requêtes des usagers. Le fil directeur de notre base générique de règles d'association triadiques repose sur les points suivants :

- La détermination de l'ensemble minimal de règles d'association triadiques (ou base générique triadique) qui véhiculent le maximum d'information utile (Bastide *et al.*, 2000).
- La possibilité de retrouver le reste de règles d'association triadiques tout en déterminant avec exactitude leurs supports et leurs confiances (Kryszkiewicz, 2002).
- L'interprétation sémantique des règles d'association triadiques extraites (BenYahia *et al.*, 2009b).

3. Extraction des règles d'association triadiques (RAT)

Avant de définir les RAT et de décrire le processus d'extraction, nous allons tout d'abord commencer par donner les définitions afférentes, à savoir celles d'une *folksonomie* et d'un concept triadique.

DÉFINITION 1. — (FOLKSONOMIE)(Hotho *et al.*, 2006) Une folksonomie est un ensemble de tuples $\mathcal{F} = (\mathcal{U}, \mathcal{T}, \mathcal{R}, \mathcal{Y})$, avec \mathcal{U} , \mathcal{T} et \mathcal{R} représentent respectivement, l'ensemble des utilisateurs, l'ensemble de tags et l'ensemble de ressources ; $\mathcal{Y} \subseteq \mathcal{U} \times \mathcal{T} \times \mathcal{R}$ représente une relation triadique dont chaque $y \subseteq \mathcal{Y}$ peut être représenté par un triplet : $y = \{(u, t, r) \mid u \in \mathcal{U}, t \in \mathcal{T}, r \in \mathcal{R}\}$. Ce qui signifie que l'utilisateur $u \in \mathcal{U}$ a annoté la ressource $r \in \mathcal{R}$ par le tag $t \in \mathcal{T}$.

EXEMPLE 2. — Le tableau 1 illustre un exemple d'une folksonomie \mathcal{F} avec $\mathcal{U} = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$, $\mathcal{T} = \{t_1, t_2, t_3, t_4, t_5\}$ et $\mathcal{R} = \{r_1, r_2, r_3\}$.

Notons que chaque "×" représente une relation triadique entre un utilisateur appartenant à \mathcal{U} , un tag appartenant à \mathcal{T} et une ressource annotée appartenant à \mathcal{R} .

Tableau 1. Une folksonomie \mathcal{F}

U/R-T	r_1					r_2					r_3				
	t_1	t_2	t_3	t_4	t_5	t_1	t_2	t_3	t_4	t_5	t_1	t_2	t_3	t_4	t_5
u_1		×	×	×			×	×	×			×	×	×	
u_2		×	×	×		×	×	×	×		×	×	×	×	
u_3			×	×		×	×	×	×		×	×	×	×	
u_4						×			×		×				×
u_5		×	×	×	×		×	×	×	×		×	×	×	
u_6				×	×				×	×					
u_7	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

DÉFINITION 3. — (CONCEPT TRIADIQUE (FRÉQUENT)) (Jäschke *et al.*, 2008) Un concept triadique (ou un tri-concept) d'une folksonomie $\mathcal{F} = (\mathcal{U}, \mathcal{T}, \mathcal{R}, \mathcal{Y})$ est un triplet (U, T, R) où $U \subseteq \mathcal{U}$, $T \subseteq \mathcal{T}$, et $R \subseteq \mathcal{R}$ avec $U \times T \times R \subseteq \mathcal{Y}$ tel que le triplet (U, T, R) est maximal, i.e., pour $U_1 \subseteq U$, $T_1 \subseteq T$ et $R_1 \subseteq R$ où $U_1 \times T_1 \times R_1 \subseteq \mathcal{Y}$,

les ensembles $U \subseteq U_1$, $T \subseteq T_1$, et $R \subseteq R_1$ impliquent toujours $(U, T, R) = (U_1, T_1, R_1)$. L'ensemble des tri-concepts de \mathcal{F} est représenté par $\mathcal{TC} = \{\mathcal{TC}_i \mid \mathcal{TC}_i = (U, T, R) \in \mathcal{Y} \text{ est un tri-concept}, i = 1..n\}$.

Notons que quel que soit le tri-concept $\mathcal{TC}_i = (U, T, R) \subseteq \mathcal{TC}$, les cardinalités des sous-ensembles U , T et R vérifient :

$$|U| \geq u - \text{minsupp}, |T| \geq t - \text{minsupp} \times \text{et } |R| \geq r - \text{minsupp}$$

avec $u - \text{minsupp}$, $t - \text{minsupp}$ et $r - \text{minsupp}$ définissent les supports minimaux spécifiés par l'utilisateur.

Pour l'extraction des tri-concepts à partir d'une folksonomie, nous avons utilisé l'algorithme TRICONS proposé par (Trabelsi *et al.*, 2012). En effet, TRICONS prend en entrée la folksonomie $\mathcal{F} = (\mathcal{U}, \mathcal{T}, \mathcal{R}, \mathcal{Y})$ ainsi que les valeurs de supports minimaux à savoir, $u - \text{minsupp}$, $t - \text{minsupp}$ et $r - \text{minsupp}$ et retourne l'ensemble des tri-concepts. En considérant à titre d'exemple les valeurs de supports minimaux suivant : $u - \text{minsupp} = 2$, $t - \text{minsupp} = 2$ et $r - \text{minsupp} = 2$, alors l'exécution de TRICONS sur la folksonomie illustrée par le tableau 1 permet d'obtenir les trois tri-concepts suivants : $\mathcal{TC}_1 = \{(u_5, u_7), (t_2, t_3, t_4), (r_1, r_2, r_3)\}$, $\mathcal{TC}_2 = \{(u_2, u_3, u_4, u_7), (t_1, t_4), (r_2, r_3)\}$ et $\mathcal{TC}_3 = \{(u_2, u_3, u_7), (t_1, t_2, t_3, t_4), (r_2, r_3)\}$. Par exemple, le tri-concept \mathcal{TC}_1 signifie que les utilisateurs u_5 et u_7 ont tous les deux annoté les ressources r_1 , r_2 et r_3 par les tags t_2 , t_3 et t_4 .

3.1. Similarité entre tri-concepts

Afin d'éviter la redondance sémantique des tri-concepts qui entraînent la génération de règles triadiques non informatives, nous définissons dans ce qui suit, deux notions de similarité. Nous admettons que deux tri-concepts sont *similaires* s'il partagent des groupes similaires d'utilisateurs et qu'ils présentent des ensembles de ressources similaires. Un ensemble de tri-concepts similaires sera noté par l'ensemble \mathcal{TC}_C .

3.1.1. Similarité entre groupes d'utilisateurs

Pour calculer la similarité entre deux groupes d'utilisateurs, nous avons associé à chaque groupe d'utilisateurs une distribution de tags. Ainsi, à chaque groupe d'utilisateurs, correspond un vecteur d'occurrence de tags. On définit alors pour le groupe U_k , le vecteur $\mathcal{V}_{U_k} = \{(t_1, n_{U_k}), (t_2, n_{U_k}), \dots, (t_i, n_{U_k})\}$ avec (t_i, n_{U_k}) représente le nombre de fois que le groupe d'utilisateurs U_k a annoté avec le tag t_i . En considérant que deux groupes d'utilisateurs avec la même distribution de tags, partagent un intérêt commun, on peut alors estimer la similarité entre deux groupes d'utilisateurs comme suit (D.Yin *et al.*, 2010) :

$$\text{Sim}(\mathcal{V}_{U_1}, \mathcal{V}_{U_2}) = \frac{\mathcal{V}_{U_1} \times \mathcal{V}_{U_2}}{|\mathcal{V}_{U_1}| \times |\mathcal{V}_{U_2}|} \quad (1)$$

On définit ainsi, la similarité normalisée¹ par :

$$Sim_1(\mathcal{V}_{U_1}, \mathcal{V}_{U_2}) = \frac{|\mathcal{V}_{U_1} \cap \mathcal{V}_{U_2}|}{Sim(\mathcal{V}_{U_1}, \mathcal{V}_{U_2})} \quad (2)$$

Notons qu'un degré de similarité égal à 1, indique que les deux groupes d'utilisateurs sont parfaitement similaires.

EXEMPLE 4. — Considérons la folksonomie illustrée par le tableau 1 et soient $U_1 = \{u_2, u_3\}$ et $U_2 = \{u_5, u_7\}$ deux groupes d'utilisateurs. Nous avons :

$$\mathcal{V}(U_1) = \{ (t_1, 4), (t_2, 6), (t_3, 6), (t_4, 6) \}$$

$$\mathcal{V}(U_2) = \{ (t_1, 3), (t_2, 6), (t_3, 6), (t_4, 6), (t_5, 5) \}$$

$$\text{D'après (1)} \quad Sim(\mathcal{V}_{U_1}, \mathcal{V}_{U_2}) = \frac{|\mathcal{V}_{U_1}| \times |\mathcal{V}_{U_2}|}{|\mathcal{V}_{U_1}| \times |\mathcal{V}_{U_2}|} = \frac{12+36+36+36}{20} = \frac{120}{20} = 6$$

$$\text{Ainsi d'après (2)} : \quad Sim_1(\mathcal{V}_{U_1}, \mathcal{V}_{U_2}) = \frac{|\mathcal{V}_{U_1} \cap \mathcal{V}_{U_2}|}{Sim(\mathcal{V}_{U_1}, \mathcal{V}_{U_2})} = \frac{4}{6} = 0.66 \quad \square$$

3.1.2. Similarité entre ensembles de ressources

Considérons deux ensembles de ressources R_1 et R_2 avec $R_1 = \{r_1, \dots, r_i\}$ et $R_2 = \{r_1, \dots, r_j\}$. Les deux ensembles R_1 et R_2 sont dits similaires s'ils partagent des ressources similaires. Pour calculer la similarité entre r_i et r_j ², nous avons utilisé la mesure de similarité normalisée *PLSCC*³, énoncée dans (Bach, 2006), permettant de faire une approximation de la longueur de la plus longue sous-chaîne commune entre r_i et r_j .

$$Sim_2(r_i, r_j) = \frac{PLSC(r_i, r_j)}{|r_i|} \quad (3)$$

Nous admettons qu'un bon degré de similarité est atteint lorsque $Sim_2(r_i, r_j) \geq 0.75$.

3.2. Génération des règles d'association triadiques

La formalisation du problème d'extraction des règles d'association a été introduite par (Agrawal *et al.*, 1993). Il est défini comme suit : soit $\mathcal{I} = \{i_1, i_2, \dots, i_m\}$ un ensemble de m littéraux appelés *items* et $\mathcal{O} = \{o_1, o_2, \dots, o_n\}$ un ensemble de n objets dans une base contenant n transactions. Soit I un sous-ensemble d'items ou *itemset*, contenant k items, alors I est appelé *k-itemset*. Le support d'un itemset I noté $Supp(I)$ est donné par $Supp(I) = \frac{|\{o_j \in \mathcal{O}, I \subseteq o_j\}|}{|\{o_j \in \mathcal{O}\}|}$. Lorsque le support de I est supérieur ou égal à un seuil minimal *minsupp*⁴ fixé par l'utilisateur, I est dit *fréquent*.

1. Les valeurs calculées par la mesure normalisée ne peuvent varier que dans un intervalle de 0 à 1.
2. Les ressources r_i et r_j désignent des URL.
3. Plus Longue Sous-Chaine Commune.
4. Le nombre minimal de transactions auxquelles doit appartenir un *itemset* I pour être qualifié de fréquent.

Une *règle d'association* est définie comme une implication entre deux itemsets. Elle est de la forme suivante : $R : X \Rightarrow Y$ tels que $X \subseteq \mathcal{I}$, $Y \subseteq \mathcal{I}$ et $X \cap Y = \emptyset$. La règle R est dite *basée* sur l'itemset $X \cup Y$ et les deux itemsets X et Y sont appelés respectivement *prémisse* et *conclusion* de R .

Afin de vérifier la validité d'une règle d'association R , deux mesures sont communément utilisées (Agrawal *et al.*, 1993), à savoir le *Support* et la *Confiance*. Le support de la règle R , noté $Supp(R)$, correspond au nombre de transactions qui contiennent simultanément les itemsets X et Y . Il est donné par $Supp(X \cup Y)$. La confiance de la règle R , notée $confiance(R)$, exprime la probabilité conditionnelle qu'une transaction contienne Y sachant qu'elle contient X . Elle est mesurée par le ratio $\frac{Supp(X \cup Y)}{Supp(X)}$.

L'extraction des règles d'association consiste à déterminer l'ensemble de règles *valides i.e.*, dont le support et la confiance sont au moins égaux à un seuil minimal de support *minsupp* et à un seuil minimal de confiance *minconf* prédéfinis par l'utilisateur.

À la lumière de la formalisation du problème d'extraction de règles d'association ainsi introduite, nous présentons dans ce qui suit la définition d'une règle d'association triadique qui met en exergue les corrélations entre les tags d'un tri-concept.

DÉFINITION 5. — (RÈGLE D'ASSOCIATION TRIADIQUE) Soit \mathcal{F} une folksonomie et \mathcal{TC} l'ensemble des concepts triadiques. Une règle d'association triadique (ou règle triadique) est une implication $T_1 \rightarrow_{U,R} (T_2 \setminus T_1)$ où T_1 et T_2 sont des ensembles non vides de tags inclus ou égaux à l'ensemble de tags T , U un ensemble d'utilisateurs, R un ensemble de ressources et (U, T, R) un tri-concept fréquent de \mathcal{TC} . La règle triadique $T_1 \rightarrow_{U,R} (T_2 \setminus T_1)$ dénote que l'ensemble de tags de T_1 implique l'ensemble de tags de $(T_2 \setminus T_1)$ relativement à l'ensemble d'utilisateurs U et l'ensemble de ressources R .

Notons que dans la suite du papier \mathcal{RT} dénote l'ensemble de toutes les règles triadiques extraites à partir d'un ensemble \mathcal{TC} de tri-concepts fréquents.

$$\mathcal{RT} = \{ T_1 \rightarrow_{U,R} (T_2 \setminus T_1) \mid (U, T, R) \in \mathcal{TC}; T, T_1 \neq \emptyset, T_1, T_2 \subseteq T \}$$

Par ailleurs, lors de la génération des règles d'association des mesures de qualité sont nécessaires pour juger la pertinence de ces dernières (Feno, 2007). Cependant, les deux mesures usuelles utilisées dans le cas de règles d'association "diadiques", *i.e.*, *support* et *confiance*, s'avèrent inexploitable dans le cas du contexte triadique dans la mesure où elles sont inadaptées aux RAT. Nous introduisons dans ce qui suit deux nouvelles mesures de qualité des RAT, à savoir le *support triadique conditionné* et la *confiance triadique conditionnée* que nous définissons comme suit :

DÉFINITION 6. — (SUPPORT TRIADIQUE CONDITIONNÉ) Soit T un ensemble de tags, $\mathcal{TC} = (U, T, R)$ un tri-concept et \mathcal{TC}_c l'ensemble des tri-concepts similaires à \mathcal{TC} . Le support triadique conditionné (noté $supp_c$) de T est défini comme suit :

$$supp_c(T) = \frac{\prod_{t_i \in T} Supp(t_i) \mid t_i \in \mathcal{TC}_c.T}{|\mathcal{TC}_c|} \text{ où } \mathcal{TC}_c.T \text{ désigne l'ensemble de tags de } \mathcal{TC}_c \text{ et } Supp(t_i) \text{ désigne le support de } t_i \text{ dans } \mathcal{TC}_c.T.$$

PROPRIÉTÉ 7. — La mesure $supp_c$ est une mesure qui vérifie la propriété d'anti-monotonie.

PREUVE 8. — Soient T_1 et T_2 deux sous-ensembles de tags tels que $T_1 \subseteq T_2 \subseteq T$. Montrons que $supp_c(T_1) \geq supp_c(T_2)$. Nous avons : $supp_c(T_1) = \frac{\prod_{t_i \in T_1} Supp(t_i) | t_i \in \mathcal{TC}_c.T}{|\mathcal{TC}_c.T|}$ et $supp_c(T_2) = \frac{\prod_{t_j \in T_2} Supp(t_j) | t_j \in \mathcal{TC}_c.T}{|\mathcal{TC}_c.T|}$. Ainsi, $\frac{supp_c(T_1)}{supp_c(T_2)} = \frac{\prod_{t_i \in T_1} Supp(t_i) | t_i \in \mathcal{TC}_c.T}{\prod_{t_j \in T_2} Supp(t_j) | t_j \in \mathcal{TC}_c.T} = \frac{1}{\prod_{t_k \in (T_2 \setminus T_1)} Supp(t_k) | t_k \in \mathcal{TC}_c.T} = \frac{1}{supp_c(T_2 \setminus T_1)} \geq 1$ puisque $supp_c(T_2 \setminus T_1) \in]0, 1]$. Par conséquent, $supp_c(T_1) \geq supp_c(T_2)$. ■

DÉFINITION 9. — (SUPPORT TRIADIQUE CONDITIONNÉ D'UNE RÈGLE TRIADIQUE) Soient R une règle triadique de la forme $T_1 \rightarrow_{U,R} (T_2 \setminus T_1)$, $TC = (U, T, R)$ un tri-concept et $T_1, T_2 \subseteq T$. Le support triadique conditionné (noté $supp_c$) de la règle R est défini comme suit : $supp_c(R) = supp_c(T) = \prod_{t_i \in T} supp_c(t_i)$

DÉFINITION 10. — (CONFIANCE TRIADIQUE CONDITIONNÉE) Soient R une règle triadique de la forme $T_1 \rightarrow_{U,R} (T_2 \setminus T_1)$, $TC = (U, T, R)$ un tri-concept et $T_1, T_2 \subseteq T$. La confiance triadique conditionnée (notée $conf_c$) de la règle R est définie comme suit : $conf_c(R) = \frac{supp_c(T)}{supp_c(T_1)} = supp_c(T \setminus T_1)$

4. BGR_T : Une base générique sans perte d'information pour les règles triadiques

L'exploitation des règles d'association triadiques (RAT), notamment pour l'auto-complétion de requêtes, s'avère contraignante étant donné le nombre considérable de RAT pouvant être extraites à partir de folksonomies réelles. Afin de palier ce problème, nous proposons dans cette section une nouvelle approche de sélection sans perte d'information des RAT qui véhiculent le maximum de connaissances utiles. Cette approche repose sur l'extraction d'un sous-ensemble générique de toutes les RAT, appelé BGR_T . Nous présentons dans ce qui suit les propriétés structurelles de notre base générique pour énoncer par la suite l'algorithme ARTMINER dédié à l'extraction de BGR_T .

4.1. Propriétés structurelles de BGR_T

La base générique BGR_T satisfait les conditions suivantes (Kryszkiewicz, 2002) :

1. **Dérivabilité** : le mécanisme d'inférence (e.g., un système axiomatique), permettant la dérivation des règles redondantes, doit être *valide* (i.e., le système ne permet de dériver que les règles d'association valides) et *complet* (i.e., l'ensemble de toutes les règles valides peut être retrouvé).

2. **Informativité** : la base générique des règles d'association doit permettre de retrouver avec exactitude le support et la confiance des règles dérivées.

Par ailleurs, étant donné qu'un certain nombre de RAT de l'ensemble \mathcal{RT} sont redondantes, nous commençons dans ce qui suit par présenter la notion de redondance d'une RAT pour définir par la suite la base générique de RAT.

DÉFINITION 11. — (REDONDANCE DE RÈGLES TRIADIQUES) Soit \mathcal{RT} l'ensemble des règles triadiques pouvant être extraites d'une folksonomie \mathcal{F} .

Une règle $R_1 : T''_1 \Rightarrow_{U,R} T'_1 \in \mathcal{RT}$ est redondante (ou dérivable) par rapport à $R_2 : T''_2 \Rightarrow_{U,R} T'_2$ si R_1 satisfait les conditions suivantes :

1. $\text{supp}_c(R_1) \geq \text{supp}_c(R_2) \wedge \text{Conf}_c(R_1) \geq \text{Conf}_c(R_2)$
2. $T''_2 \subseteq T''_1 \wedge T'_1 \subset T'_2$.

Ainsi, étant donné que nous ne nous intéressons qu'aux requêtes d'un seul tag, nous définissons la base générique de RAT comme suit :

DÉFINITION 12. — (BASE GÉNÉRIQUE DE RÈGLES TRIADIQUES) Soit \mathcal{TC} l'ensemble des concepts triadiques. Une base générique de règles triadiques $\mathcal{BGR}\mathcal{T}$ ⁵ est l'ensemble des règles triadiques $t \rightarrow_{U,R} (T \setminus t)$ où $(U, T, R) \in \mathcal{TC}$.

$$\mathcal{BGR}\mathcal{T} = \{t \rightarrow_{U,R} (T \setminus t) \mid (U, T, R) \in \mathcal{TC}\}$$

Les règles génériques de $\mathcal{BGR}\mathcal{T}$ représentent des implications entre des prémisses minimales et des conclusions maximales (en termes de nombre de tags). En effet, (Pasquier, 2000) a mis en exergue que ce type de règles est le plus général (*i.e.*, convoyant le maximum d'information).

Dans ce qui suit, nous procédons à la vérification des conditions de non-perte d'information de $\mathcal{BGR}\mathcal{T}$, à savoir la *dérivabilité* et l'*informativité*.

1. **dérivabilité** : afin de dériver l'ensemble de toutes les règles valides, nous utilisons le système axiomatique d'Armstrong présenté dans (Beeri *et al.*, 1984) et nous prouvons qu'il est valide et complet. Par ailleurs, étant donnée la particularité des RAT considérées, *i.e.*, constituées d'un seul tag en prémisses, nous allons nous intéresser plus particulièrement aux axiomes d'augmentation et de décomposition.

Augmentation : si $\text{conf}_c(R_1 : T_1 \Rightarrow_{U,R} T \setminus T_1) \geq \text{minconf}_c$, alors $\text{conf}_c(R_2 : (T_1 \cup T_2) \Rightarrow_{U,R} (T \setminus T_2)) \geq \text{minconf}_c$; $T_2 \subset T$.

Décomposition : si $\text{conf}_c(R_1 : T_1 \Rightarrow_{U,R} T_2 \cup T_3) \geq \text{minconf}_c$, alors $\text{conf}_c(R_2 : (T_1) \Rightarrow_{U,R} (T_2)) \geq \text{minconf}_c$ et $\text{conf}_c(R_3 : (T_1) \Rightarrow_{U,R} (T_3)) \geq \text{minconf}_c$; $T_2, T_3 \subset T$.

a. Validité du système axiomatique : ceci revient à vérifier qu'il ne permet de dériver que les règles triadiques ayant des mesures de supports et de confiances (conditionnés) au moins égales aux seuils fixés par l'utilisateur.

5. $\mathcal{BGR}\mathcal{T}$ est l'acronyme de Base Générique de Règles Triadiques.

PROPOSITION 13. — Soient $\mathcal{BGR}\mathcal{T}$ la base générique et \mathcal{RT} l'ensemble de toutes les règles valides pouvant être extraites de l'ensemble des tri-concepts fréquents \mathcal{TC} . Le système axiomatique proposé est valide.

PREUVE 14. — Le support conditionné de la règle R_1 et de la règle R_2 sont égaux étant donné qu'ils concernent le même ensemble de tags T , c'est-à-dire qu'ils appartiennent au même concept triadique. Ainsi $\text{supp}_c(R_1) = \text{supp}_c(R_2)$. Par ailleurs, nous avons $\text{conf}_c(R_1) = \frac{\text{supp}_c(T)}{\text{supp}_c(T_1)}$ et $\text{conf}_c(R_2) = \frac{\text{supp}_c(T)}{\text{supp}_c(T_1 \cup T_2)}$. Puisque $T_1 \subseteq T_1 \cup T_2$ alors $\text{supp}_c(T_1 \cup T_2) \leq \text{supp}_c(T_1)$, nous déduisons que $\frac{\text{supp}_c(T)}{\text{supp}_c(T_1 \cup T_2)} \geq \frac{\text{supp}_c(T)}{\text{supp}_c(T_1)}$. Par conséquent, $\text{conf}_c(R_2) \geq \text{conf}_c(R_1)$ et $\text{supp}_c(R_2) = \text{supp}_c(R_1)$. ■

b. Complétude du système axiomatique : ceci revient à montrer qu'il permet de dériver toutes les règles valides pouvant être extraites d'un ensemble de tri-concepts fréquents \mathcal{TC} .

PROPOSITION 15. — Le système axiomatique proposé est complet : toutes les règles pouvant être extraites de l'ensemble \mathcal{TC} sont dérivables de $\mathcal{BGR}\mathcal{T}$ moyennant le système axiomatique proposé.

PREUVE 16. — Soit $\mathcal{BGR}\mathcal{T}$ la base générique de règles obtenue à partir de l'ensemble \mathcal{TC} pour minsupp_c et minconf_c donnés. Soit \mathcal{RT} l'ensemble de toutes les règles valides extraites à partir de \mathcal{TC} . Soit $R_1 : T_1 \Rightarrow_{U,R} (T_2 \setminus T_1) \in \mathcal{RT}$. Nous montrons dans ce qui suit, que R_1 peut être dérivée à partir d'une règle générique de $\mathcal{BGR}\mathcal{T}$ par l'application du système axiomatique proposé. Trois cas peuvent se présenter :

- si $|T_1| = 1$ et $T_2 = T$, alors la règle triadique R_1 appartient également à $\mathcal{BGR}\mathcal{T}$ étant donné qu'elle contient un seul tag en prémisse et $(T \setminus T_1)$ en conclusion.
- si $|T_1| = 1$ et $T_2 \subset T$, alors la règle triadique R_1 peut être retrouvée par décomposition de la règle $R_2 \in \mathcal{BGR}\mathcal{T}$ contenant le tag t de T_1 en prémisse et l'ensemble $(T \setminus T_1)$ en conclusion.
- sinon, il existe une règle $R_2 \in \mathcal{BGR}\mathcal{T}$ contenant en prémisse un tag $t \in T_1$. En conséquence, la règle R_1 peut être retrouvée par l'application de l'axiome d'augmentation des tags $t' \in (T_1 \setminus t)$ qui appartiennent à la conclusion de la règle R_2 . ■

2. **Informativité** : nous avons montré que les règles dérivées à partir de $\mathcal{BGR}\mathcal{T}$ moyennant le système axiomatique proposé possèdent des mesures de support et de confiance conditionnés supérieures respectivement à minsupp_c et minconf_c . Dans ce qui suit, nous proposons d'étudier l'informativité de $\mathcal{BGR}\mathcal{T}$ pour prouver l'exactitude de ces mesures.

PROPOSITION 17. — La base $\mathcal{BGR}\mathcal{T}$ est informative, i.e., les mesures de support et de confiance conditionnés des règles redondantes peuvent être déterminées avec exactitude.

PREUVE 18. — Soit la règle $R_1 : t \Rightarrow_{U,R} (T \setminus t) \in \mathcal{BGR}\mathcal{T}$ et $R_2 : T_1 \Rightarrow_{U,R} (T_2 \setminus T_1)$ une règle redondante appartenant à \mathcal{RT} . Peut-on déterminer les supports et confiances conditionnés de R_2 avec exactitude? D'une part, le support conditionné de R_2 peut être retrouvé par la somme des supports conditionnés des tags de $(T_2 \cup T_1)$

qui appartiennent tous à T ; ceci découle de la définition même d'une règle triadique. D'autre part, la confiance conditionnée de R' est égale à son support conditionné divisé par celui de sa prémisse. Or, le support conditionné de n'importe quelle prémisse d'une règle redondante peut être retrouvé par la somme des supports conditionnés des prémisses des règles génériques de \mathcal{BGR} contenant tous un seul tag. Par conséquent, le support conditionné et la confiance conditionnée de toutes les règles triadiques peuvent être déterminés avec exactitude. ■

EXEMPLE 19. — Soit le tri-concept $TC_4 = (\{u_5, u_7\}, \{t_2, t_3, t_4\}, \{r_1, r_2, r_3\})$ extrait à partir du tableau 1. L'ensemble total des RAT extraites à partir de ce tri-concept est égal à :

$$\mathcal{RT}_{\{u_5, u_7\}, \{r_1, r_2, r_3\}} = \{t_4 \rightarrow t_2 t_3, t_2 \rightarrow t_3 t_4, t_3 \rightarrow t_2 t_4, t_4 \rightarrow t_2, t_2 \rightarrow t_4, t_4 \rightarrow t_3, t_3 \rightarrow t_4, t_2 \rightarrow t_3, t_3 \rightarrow t_2, t_4 t_2 \rightarrow t_3, t_4 t_3 \rightarrow t_2, t_2 t_3 \rightarrow t_4\}$$

Cependant, la base générique de règles, associée à ce tri-concept est égale à :

$$\mathcal{BGR}_{\{u_5, u_7\}, \{r_1, r_2, r_3\}} = \{t_4 \rightarrow t_2 t_3, t_2 \rightarrow t_4 t_3, t_3 \rightarrow t_4 t_2\}$$

Nous remarquons à titre d'exemple que la règle $R : t_4, t_2 \rightarrow_{\{u_5, u_7\}, \{r_1, r_2, r_3\}} t_3$ est redondante par rapport à la règle $R_1 : t_4 \rightarrow_{\{u_5, u_7\}, \{r_1, r_2, r_3\}} t_2 t_3$, étant donné que la première peut être obtenue par augmentation à gauche du tag t_2 . De plus, nous avons : $t_4 \subseteq (t_4, t_2)$, $t_3 \subseteq (t_2, t_3)$, $supp_c(R) = supp_c(R_1)$ et $conf_c(R) = \frac{conf_c(R_1)}{supp_c(t_2)} \geq conf_c(R_1)$ □

Tableau 2. Notations utilisées dans l'algorithme ARTMINER

\mathcal{TC}	: l'ensemble des tri-concepts fréquents.
Group	: l'ensemble des utilisateurs dans un tri-concept.
$Tags_of_RA$: l'ensemble des tags d'une règle d'association triadique.
Ress	: l'ensemble des ressources dans un tri-concept.
\mathcal{BGR}	: la base générique de règles.
RAT	: une règle d'association triadique.

4.2. ARTMINER : Extraction de la base \mathcal{BGR}

Nous introduisons dans cette section l'algorithme ARTMINER⁶ pour l'extraction de \mathcal{BGR} . ARTMINER prend en entrée un ensemble \mathcal{TC} de tri-concepts ainsi que deux seuils définis par l'utilisateur : $minsupp$ et $minconf$. ARTMINER retourne la base générique \mathcal{BGR} . Le pseudo-code de l'algorithme ARTMINER est donné par l'algorithme 1. ARTMINER procède comme suit : il commence par initialiser les trois variables $Group$, $Tags_of_RA$ et $Ress$ à l'ensemble vide. Ensuite, dans sa boucle la plus externe (lignes 3 - 13), ARTMINER parcourt l'ensemble des concepts triadiques

6. ARTMINER est l'acronyme de Association Rule Triadic Miner.

Données :

1. \mathcal{TC} : {Tri-concepts fréquents}.
2. $minsupp$, $minconf$: deux seuils de support et de confiance.

Résultats : \mathcal{BGR} : {base générique de règles triadiques}.

```

1 début
2    $Group := \emptyset; Tags\_of\_RA := \emptyset; Ress := \emptyset;$ 
3   pour chaque Tri-Concept  $TC$  de  $\mathcal{TC}$  faire
4      $Group := TC.U; Tags\_of\_RA := TC.T; Ress := TC.R;$ 
5     pour tous les Tri-Concept  $TC'$  de  $\mathcal{TC}$  faire
6       si  $Sim_1(Group, TC'.U)$  et  $Sim_2(Ress, TC'.R)$  alors
7         si  $TC'.T \notin Tags\_of\_RA$  alors
8            $Tags\_of\_RA := Tags\_of\_RA \cup TC'.T;$ 
9         pour chaque tag  $t$  de  $Tags\_of\_RA$  faire
10           $RAT := t \rightarrow_{Group, Ress} Tags\_of\_RA \setminus t;$ 
11          si  $Supp_c(RAT) \geq minsupp$  alors
12            si  $Conf_c(RAT) \geq minconf$  alors
13               $\mathcal{BGR} := \mathcal{BGR} \cup RAT;$ 
14 retourner  $\mathcal{BGR}$ 
15 fin

```

Algorithme 1. ARTMINER

et pour chaque tri-concept TC , ARTMINER récupère et affecte chaque élément, *i.e.*, les utilisateurs, les tags et les ressources, aux variables $Group$, $Tags_of_RA$ et $Ress$ (Ligne 4).

Par la suite, pour chacun des autres tri-concepts TC' de \mathcal{TC} , ARTMINER teste la similarité avec le tri-concept courant en invoquant les deux fonctions de similarité Sim_1 (cf. formule (2) page 6) et Sim_2 (cf. formule (3) page 7) afin de vérifier si :

1. L'ensemble $Group$ est similaire au groupe d'utilisateurs du tri-concept TC' ,
2. L'ensemble $Ress$ est similaire aux ressources du tri-concept TC' .

Si ces deux conditions (Ligne 6) sont vérifiées, l'ensemble des tags du tri-concept TC' est ajouté à l'ensemble $Tags_of_RA$ (Ligne 7). Après avoir terminé la première boucle interne (lignes 5 - 8), ARTMINER procède à la génération des RAT à partir de l'ensemble $Tags_of_RA$. Dans sa deuxième boucle interne (Lignes 9-13), ARTMINER parcourt l'ensemble des tags de $Tags_of_RA$ et génère un ensemble de RAT ayant un seul tag t en prémisses et vérifiant les seuils minimaux de support et de confiance ($minsupp$ et $minconf$).

Cet ensemble de règles sera ajouté à $\mathcal{BGR}\mathcal{T}$ (Lignes 11 - 13). Une RAT générée est ainsi de la forme $t \rightarrow (Tags_of_RA \setminus t)$. ARTMINER prend fin lorsque tous les tri-concepts ont été énumérés et retourne $\mathcal{BGR}\mathcal{T}$.

EXEMPLE 20. — Nous avons appliqué notre algorithme ARTMINER sur l'ensemble des tri-concepts extraits par TRICONS à partir de la folksonomie décrite par le tableau 1. Nous avons considéré les valeurs de seuils minimaux suivantes : $minsupp = 0.66$ et $minconf = 0.8$. En outre, nous avons assumé que deux groupes d'utilisateurs \mathcal{V}_1 et \mathcal{V}_2 sont similaires si $Sim_1(\mathcal{V}_1, \mathcal{V}_2) \geq 0.3$. Nous présentons dans ce qui suit, les vecteurs associés aux différents groupes d'utilisateurs (U_1, \dots, U_5) des cinq tri-concepts, décrits par la figure 1, qui correspondent aux différents tags assignés aux différentes ressources d'un tri-concept dans la folksonomie \mathcal{F} .

$$\begin{aligned}\mathcal{V}(U_1) &= \{ (t_1, 7), (t_2, 15), (t_3, 15), (t_4, 15), (t_5, 5) \} \\ \mathcal{V}(U_2) &= \{ (t_1, 2), (t_2, 4), (t_3, 4), (t_4, 6), (t_5, 6) \} \\ \mathcal{V}(U_3) &= \{ (t_1, 8), (t_2, 6), (t_3, 6), (t_4, 8), (t_5, 2) \} \\ \mathcal{V}(U_4) &= \{ (t_1, 2), (t_2, 4), (t_3, 4), (t_4, 4), (t_5, 4) \} \\ \mathcal{V}(U_5) &= \{ (t_1, 6), (t_2, 6), (t_3, 6), (t_4, 6), (t_5, 2) \}\end{aligned}$$

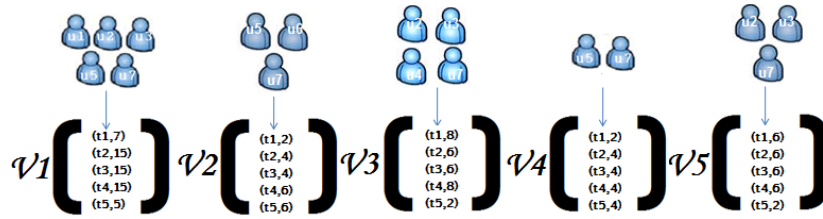


Figure 1. Les vecteurs correspondant aux différents groupes d'utilisateurs

ARTMINER procède tout d'abord à la comparaison entre le premier tri-concept de l'ensemble \mathcal{TC} avec les autres éléments. D'un côté, les vecteurs associés aux groupes d'utilisateurs de TC_1 , TC_3 et TC_5 vérifient les conditions de similarité (cf. formule (2) page 6) étant donné que :

- $Sim_1(\mathcal{V}(U_1), \mathcal{V}(U_3)) = 0.34$
- $Sim_1(\mathcal{V}(U_1), \mathcal{V}(U_5)) = 0.38$
- $Sim_1(\mathcal{V}(U_3), \mathcal{V}(U_5)) = 0.72$

D'un autre côté, ces trois tri-concepts partagent deux ressources en commun (r_2 et r_3). La condition de la ligne 6 de l'algorithme ARTMINER étant vérifiée. Les ensembles de tags associés aux trois tri-concepts TC_1 , TC_3 et TC_5 sont alors fusionnés. En conséquence, le premier tri-concept devient :

$$TC_1 = \{ \{u_1, u_2, u_3, u_5, u_7\}, \{t_1, t_2, t_3, t_4\}, \{r_1, r_2, r_3\} \}$$

De la même manière, les tri-concepts TC_2 et TC_4 fusionnent leurs ensembles de tags étant donné qu'ils partagent deux ressources en commun. De plus, les deux

groupes d'utilisateurs vérifient les conditions de similarité sur leurs vecteurs respectifs, *i.e.*, $Sim_1(\mathcal{V}(U_2), \mathcal{V}(U_4)) = 0.93$. Ce résultat est évident vu que le groupe d'utilisateurs de TC_4 est un sous-ensemble (ou sous-communauté) de celui de TC_1 .

Ainsi, le deuxième tri-concept devient :

$$TC_2 = \{\{u_5, u_6, u_7\}, \{t_2, t_3, t_4, t_5\}, \{r_1, r_2\}\}$$

ARTMINER procède ensuite à la comparaison entre TC_3 et TC_4 . Vu que $Sim_1(\mathcal{V}(U_3), \mathcal{V}(U_4)) = 0.96$, nous avons :

$$TC_3 = \{\{u_2, u_3, u_4, u_7\}, \{t_1, t_2, t_3, t_4\}, \{r_2, r_3\}\}$$

Ces trois premiers tri-concepts de l'ensemble \mathcal{TC} , dont le nombre de tags va considérablement augmenter, vont incontestablement induire le fait que les tri-concepts TC_4 et TC_5 sont désormais inclus dans l'un des trois tri-concepts modifiés. Il en résulte que ces deux derniers sont supprimés de l'ensemble \mathcal{TC} .

Parallèlement à ce "rafraîchissement" de l'ensemble des tags, ARTMINER procède pour chaque tri-concept, à l'extraction des RAT qui lui sont relatifs. Pour chaque règle extraite, ARTMINER commence par calculer le support conditionné. Si ce dernier vérifie le seuil minimal *minsupp*, ARTMINER procède alors au calcul de la confiance de cette règle qui doit également vérifier son propre seuil minimal *minconf*. Ainsi, chaque règle vérifiant ces deux seuils minimaux est ajoutée à la base générique de règles $\mathcal{BGR}\mathcal{T}$. Le tableau 3 présente les différentes règles extraites à partir des tri-concepts de \mathcal{TC} munies de leurs supports et confiances conditionnés respectifs.

Tableau 3. La base générique des règles d'association triadiques

U, R	Règle triadique	$Supp_c$	$Conf_c$	$\in \mathcal{BGR}\mathcal{T}$
$\{u_1, u_2, u_3, u_5, u_7\},$ (t_1, t_2, t_3, t_4)	$t_1 \rightarrow t_2 t_3 t_4 t_5$	0.67	0.22	non
	$t_2 \rightarrow t_1 t_3 t_4 t_5$	0.67	1.00	oui
	$t_3 \rightarrow t_1 t_2 t_4 t_5$	0.67	1.00	oui
	$t_4 \rightarrow t_1 t_2 t_3 t_5$	0.67	1.00	oui
	$t_5 \rightarrow t_1 t_2 t_3 t_4$	0.67	0.67	non
$\{u_5, u_6, u_7\},$ (t_2, t_3, t_4, t_5)	$t_2 \rightarrow t_3 t_4 t_5$	1.00	1.00	oui
	$t_3 \rightarrow t_2 t_4 t_5$	1.00	1.00	oui
	$t_4 \rightarrow t_2 t_3 t_5$	1.00	1.00	oui
	$t_5 \rightarrow t_2 t_3 t_4$	1.00	1.00	oui
$\{u_2, u_3, u_4, u_7\},$ (t_1, t_2, t_3, t_4)	$t_1 \rightarrow t_2 t_3 t_4$	1.00	0.67	non
	$t_2 \rightarrow t_1 t_3 t_4$	1.00	1.00	oui
	$t_3 \rightarrow t_1 t_2 t_4$	1.00	1.00	oui
	$t_4 \rightarrow t_1 t_2 t_3$	1.00	1.00	oui

□

Dans la prochaine section, nous mettons en exergue l'apport de BGR_T pour l'autocomplétion de requêtes soumises pour la recherche d'information dans les folksonomies. En effet, les règles triadiques de BGR_T ont la particularité d'avoir un seul tag en *prémisse* et un nombre maximal de tags en *conclusion*. Cela garantit que la complétion de requêtes se fera sans perte d'information et ce, par la suggestion à l'utilisateur de l'ensemble de tags le plus complet avec la confiance la plus élevée.

5. Autocomplétion de requêtes par la base BGR_T

L'objectif de la complétion de requêtes, domaine initialement abordé par (Motro, 1987 ; Cuppens, Demolombe, 1989 ; Frantz, Shapiro, 1991), est de suggérer des informations complémentaires et utiles à un utilisateur suite à sa soumission d'une requête dans une folksonomie ; lui évitant ainsi de reformuler la même requête jusqu'à obtention d'une réponse satisfaisante. La complétion de requêtes est alors définie comme suit :

DÉFINITION 21. — (AUTOCOMPLÉTION DE REQUÊTES) Une *autocomplétion de requête* (ou *auto-complétion de requête*) C_Q est une paire $(\mathcal{T}_c, \mathcal{U}_c)$, où \mathcal{T}_c est un ensemble tags (les tags complétant la requête) et \mathcal{U}_c un ensemble d'utilisateurs (les utilisateurs partageant les mêmes intérêts). La *complétion de requête* revient à calculer le sous-ensemble $\mathcal{U}_1 \subseteq \mathcal{U}_c$ d'utilisateurs qui ont partagé les tags de \mathcal{T}_c , ainsi que le sous-ensemble $\mathcal{T}_1 \subseteq \mathcal{T}_c$ de tags ayant été annotés par au moins un utilisateur. Notons que \mathcal{T}_c et \mathcal{U}_c sont appelés respectivement **Préfixe** et le **Suffixe** de la requête complétée.

5.1. Problématique de l'autocomplétion de requêtes

Étant donnée une requête Q soumise par un utilisateur u_s , la complétion de requêtes se base sur deux ensembles, *i.e.*, l'ensemble maximal de tags et l'ensemble maximal d'utilisateurs. Le processus d'autocomplétion de requêtes revient par conséquent à :

- Trouver l'ensemble maximal de tags partagés par un même groupe d'utilisateurs sur un même ensemble de ressources. Ce qui revient à considérer les règles triadiques ayant en prémisse la requête Q .
- Trouver l'ensemble maximal d'utilisateurs associés à cet ensemble maximal de tags. Cela revient à étudier les groupes d'utilisateurs contenant u_s et relatifs aux règles triadiques ayant en prémisse la requête Q .

Le processus d'autocomplétion de requêtes à partir de BGR_T est décrit par l'algorithme COMP_USER_QUERY donné par l'algorithme 2. COMP_USER_QUERY prend en entrée BGR_T ainsi que la requête Q soumise par l'utilisateur u_s et retourne la requête complétée C_Q . COMP_USER_QUERY procède par un parcours de l'ensemble des RAT de BGR_T afin de déterminer la règle R_c la plus confidente (*i.e.*, ayant la confiance conditionnée la plus élevée) vérifiant les deux conditions suivantes :

1. L'utilisateur u_s doit appartenir au groupe d'utilisateurs U de la règle R_c .
2. Le tag de la requête Q est contenu dans la prémisse de la règle R_c .

Ainsi, après avoir déterminé la règle la plus confidente R_c , COMP_USER_QUERY ajoute d'une part, le groupe d'utilisateurs $R_c.U$ à celui de la requête complétée C_Q et d'autre part, l'ensemble des tags qui constitue la conclusion de la règle R_c à l'ensemble \mathcal{T}_c de C_Q .

Données : - La base générique des règles triadiques BGR_T , Q et u_s .

Résultats : - La requête complétée C_Q .

```

1 début
2    $BestConfidence := 0; C_Q := Q;$ 
3   pour chaque Règle Triadique  $RT$  de  $BGR_T$  faire
4     si  $u_s \in RT.U$  alors
5       si  $Q \in RT.T.Prémisse$  alors
6         si  $Confiance(RT) \geq BestConfidence$  alors
7            $RT_c := RT;$ 
8            $BestConfidence := Confiance(RT);$ 
9       si  $RT_c \neq null$  alors
10         $C_Q.U_c := C_Q \cup RT_c.U;$ 
11         $C_Q.T_c := C_Q \cup RT_c.T.Conclusion;$ 
12  retourner  $C_Q$ ;
13 fin

```

Algorithme 2. COMP_USER_QUERY

Le processus d'autocomplétion de requêtes se base ainsi sur BGR_T afin de compléter efficacement la requête de l'utilisateur. Deux alternatives seront donc proposées à ce dernier, à savoir décider ; (i) d'explorer l'ensemble de tags définis dans le *Préfixe* de la requête complétée lui évitant de resoumettre d'autres requêtes moyennant ces tags ou ; (ii) de découvrir les utilisateurs définis par le *Suffixe* de la requête complétée avec lesquels il partage le même intérêt pour les ressources recherchées.

5.2. Exemple illustratif du processus d'autocomplétion de requêtes par BGR_T

Afin d'illustrer le processus d'autocomplétion de requêtes par BGR_T , nous prenons comme exemple un ensemble de règles extraites, décrites par le tableau 4 (les URL correspondant aux ressources du tableau 4 sont présentées dans le tableau 5), à partir d'un jeu de données collecté à partir du site de partage collaboratif *del.icio.us*. Considérons, par exemple, que l'utilisateur $u_1=Jacob$ ait soumis la requête $Q:=\{nasa\}$. Afin de compléter Q , COMP_USER_QUERY explore BGR_T pour identifier les RAT dont le groupe d'utilisateurs contient *Jacob* et dont le tag en prémisse est égale à *nasa*. Une fois que ces dernières ont été identifiées, COMP_USER_QUERY choisit la règle triadique la plus confidente pour compléter Q . Les règles, candidates, générées sont :

$nasa \Rightarrow_{(u_1, u_5), (r_2, r_3)} photos, space, images, free, stock, science, Conf_c = 0.37,$

$nasa \Rightarrow_{(u_1, u_6, u_7), (r_1, r_4)} satellite, mars, images, science, Conf_c = \mathbf{0.88},$

Puisque la deuxième règle est la plus confidente, ce sont les utilisateurs u_6 et u_7 ainsi que les tags *satellite*, *mars*, *images* et *science* qui s'ajouteront à la requête Q . Par conséquent, la requête complétée C_Q , retournée à $u_1=Jacob$, est égale à :
 $C_Q = \{(nasa, satellite, mars, images, science), (u_1, u_6, u_7)\}.$

Tableau 4. Exemple de base générique extraite à partir d'un petit jeu de données de DEL.ICIO.US

U	BGR_T	R	$supp_c$	$conf_c$
<i>Jacob, Brson</i>	space→photos,nasa,images,free,stock,science		0.571	0.653
	photos→space,nasa,images,free,stock,science	r_1	0.571	0.592
	nasa→photos,space,images,free,stock,science	r_2	0.571	0.371
	images→photos,nasa,space,free,stock,science	r_3	0.571	0.653
	science→photos,nasa,images,free,stock,space		0.571	0.592
<i>Jacob, Lynavo Sandy</i>	satellite→nasa,mars,images,science		0.819	0.452
	mars→satellite,nasa,images,science	r_1	0.819	0.726
	nasa→satellite,mars,images,science	r_4	0.819	0.887
	images→satellite,mars,nasa,science		0.819	0.717
	science→satellite,mars,nasa,images		0.819	0.213

Tableau 5. Les URL correspondant aux ressources du tableau 4

Ressource	URL du Site Web
r_1	http://answeringserviceonline.info/answer-service/
r_2	http://www.nasa.gov/multimedia/highlights/index.html
r_3	http://visibleearth.nasa.gov/
r_4	http://www.unclesamsphotos.com/

6. Étude expérimentale

Dans cette section, nous montrons en premier lieu la compacité de notre base générique. Nous prouvons ainsi qu'elle permet de réduire de manière considérable l'ensemble total des règles triadiques. Nous procédons par la suite à l'évaluation de l'autocomplétion des requêtes par BGR_T . Nos différentes expérimentations ont été menées sur deux bases extraites d'un jeu de données⁷ collecté à partir du système de marque-pages social DEL.ICIO.US et dont la taille réelle s'élève à 10 MB⁸. Décrivant des annotations ajoutées par des utilisateurs durant le mois de janvier 2010, les deux bases utilisées sont de deux type, à savoir :

7. Les noms d'utilisateurs ont été codés afin de préserver l'anonymat.

8. Le jeu de données est librement téléchargeable à l'adresse : <http://data.dai-labor.de/corpus/delicious/>.

– **Base dense** : contient un nombre élevé de tri-concepts fréquents longs et ce, pour des différentes valeurs de *minsupp* (en moyenne, jusqu'à 10 tags par tri-concept). Elle est caractérisée par un nombre réduit d'utilisateurs (6 822) et un nombre limité de tags (671), partagées par ces utilisateurs. Les utilisateurs dans une base dense tendent en effet à utiliser le même vocabulaire lors de l'annotation des ressources par des tags.

– **Base éparsé** : contient un nombre considérable de tags 18 066 mais seulement une infime partie d'entre eux cooccurrent fréquemment dans les tri-concepts (en moyenne, pas plus de 2 tags par tri-concept). Ainsi, dans une base éparsé, les utilisateurs ont tendance à utiliser différents tags pour annoter un ensemble de ressources. Les caractéristiques des deux bases sont résumées dans le tableau 6.

Tableau 6. Caractéristiques des bases extraites de DEL.ICIO.US

	Base 1	Base 2
Type	Dense	Eparsé
# Triplets	67 290	99 989
# Utilisateurs	6 822	43 419
# Tags	671	18 066
# Ressources	13 102	53 397

6.1. Compacité de la base générique BGR_T

Afin d'évaluer la compacité de BGR_T par rapport à l'ensemble total des RAT, nous avons préalablement exécuté l'algorithme TRICONS séparément sur les deux bases, *i.e.*, dense et éparsé. Le tableau 7 résume les différents résultats obtenus. Nous avons ensuite procédé à l'extraction de BGR_T . Pour déterminer l'ensemble des RAT valides, nous avons considéré les valeurs de seuils suivantes : $minsupp_c = 0.10$ et $minconf_c = 0.10$.

Tableau 7. Nombre de tri-concepts extraits à partir des bases de DEL.ICIO.US

Base 1 (Type)	# Triplets	# TC	Base 2 (Type)	# Triplets	# TC
Base 1 (Dense)	67 290	3 877	Base 2 (Éparsé)	99 989	1 432

En examinant la compacité de BGR_T sur les deux bases, nous avons relevé que pour toutes les valeurs de TC , la taille de BGR_T est largement inférieure à l'ensemble total des règles extraites. La figure 2 met en évidence que lorsque le nombre de règles triadiques de RT grimpe jusqu'à 2 742 644, en considérant la base dense, et jusqu'à 4 602 pour la base éparsé, le nombre de règles triadiques de notre base générique demeure, respectivement, en dessous de 11 328 et de 2 903, *i.e.*, ce qui représente, respectivement, 0.42 % et 63.08 % de l'ensemble total de règles triadiques. En effet, la compacité de notre base générique est principalement due à ces deux facteurs :

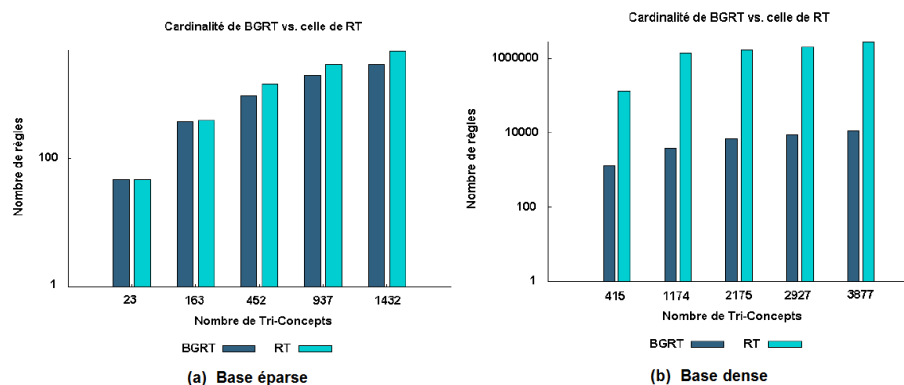


Figure 2. Taille de la $GBTR$ vs. Taille de RT

– **La taille de l'ensemble des tags T des tri-concepts qui peut être assez élevée (de 2 jusqu'à 10 tags)** : d'un côté, la taille de $BGRT$ ne s'en trouve pas affectée (le nombre des règles génériques étant égal à la taille de T). D'un autre côté, la cardinalité de RT augmente sensiblement de quelques centaines de règles. Ceci est principalement dû aux multiples combinaisons possibles de tags pour les $RATs$ de RT .

– **La forme particulière des $RATs$ de $BGRT$** : constituées d'un seul tag en pré-misse et d'un nombre maximal de tags en conclusion, est un avantage pour notre base générique puisque ce type de règles permet non seulement de véhiculer le maximum d'information mais également d'éviter la redondance d'un très grand nombre de règles.

Par ailleurs, nous observons que la base générique $BGRT$ réduit en moyenne le nombre des règles triadiques extraites à partir des tri-concepts de 36. 92 % sur la base éparse et de 99. 58 % sur la base dense. Pour la base dense, le grand nombre de règles extraites revient à la taille des tri-concepts, *i.e.*, contenant jusqu'à 10 tags qui induit un nombre élevé de règle. En revanche, dans la base éparse, l'écart observé entre la taille de RT et celle de $GBRT$ est largement moins évident. Ceci est dû au faible nombre de tags contenus dans les différents tri-concepts, *i.e.*, une moyenne 2 tags par tri-concept.

6.2. Évaluation de l'autocomplétion de requêtes par $BGRT$

Afin d'évaluer le processus d'autocomplétion de requêtes par $BGRT$, nous avons mené pour chaque base une série d'expérimentations visant à mettre en évidence le taux de couverture moyen des requêtes par la base générique. Nos expérimentations ont porté sur l'étude des deux éléments suivants :

– Le nombre de requêtes soumises par les différents utilisateurs et ce, pour chaque base : ces requêtes sont fournies en entrée à l'algorithme `COMP_USER_QUERY`.

– Le nombre de requêtes couvertes : le nombre de requêtes qui ont été réellement complétées par l’algorithme COMP_USER_QUERY, *i.e.*, les requêtes ayant un tag appartenant à l’une des règles de $\mathcal{BGR}\mathcal{T}$.

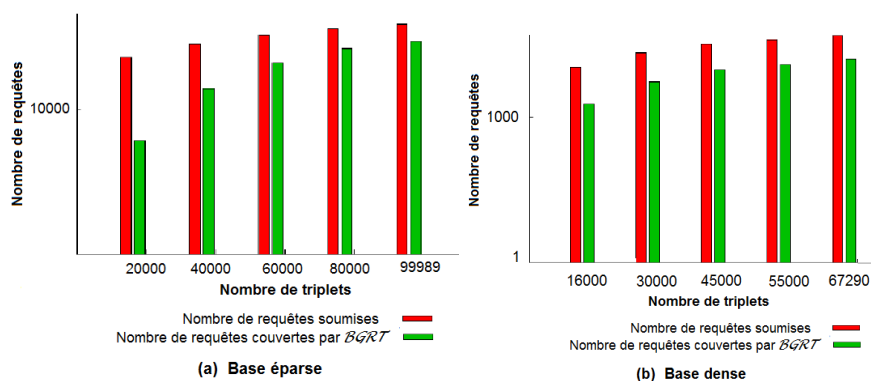


Figure 3. Requetes soumises vs. Requetes couvertes par $\mathcal{BGR}\mathcal{T}$

En examinant le nombre de requêtes couvertes par $\mathcal{BGR}\mathcal{T}$ sur les deux bases, nous avons relevé que pour toutes les valeurs de triplets l’écart entre le nombre de requêtes couvertes et le nombre total de requêtes soumises est relativement réduit. La figure 3 met en exergue que lorsque le nombre de requêtes soumises grimpe jusqu’à 15 223, en considérant la base dense, et jusqu’à 14 700 pour la base épars, le nombre de requêtes couvertes par notre base générique demeure élevé respectivement, de 13 224 et de 12 989, *i.e.*, ce qui représente 99. 86 % et 99. 88 % de l’ensemble total des requêtes.

$$\text{Taux de couverture moyen} = 1 - \frac{\text{Nombre de requêtes couvertes par } \mathcal{BGR}\mathcal{T}}{\text{Nombre de requêtes soumises}},$$

Par ailleurs, suite à une estimation menée sur les deux bases, du taux de couverture moyen des requêtes soumises par $\mathcal{BGR}\mathcal{T}$ au moyen de la formule énoncée ci-dessus, nous avons constaté que le taux de couverture moyen des requêtes sur la base dense est égale à 83. 48 %, soit 22. 43 % plus élevé que celui de la base épars. En effet, puisque le nombre de tags partagés par les différents utilisateurs est considérablement plus important sur la base dense, la probabilité qu’une requête soit couverte par $\mathcal{BGR}\mathcal{T}$ est forcément plus grande.

D’un autre côté, un même tri-concept peut couvrir jusqu’à $T \times U$ requêtes ce qui explique le très grand nombre de requêtes couvertes et ce, sur les deux bases considérées. Considérons, par exemple, la règle générique $nasa \Rightarrow_{(u_1, u_6, u_7), (r_1, r_4)} satellite, mars, images, science$ ($Conf_c = 0. 88$), et la requête $q = nasa$. La requête complétée $C_Q = \{(nasa, satellite, mars, images, science), (u_1, u_6, u_7)\}$ permet de couvrir à la fois les quatre requêtes suivantes : $q_1 = satellite$, $q_2 = mars$, $q_3 = images$ et $q_4 = science$.

7. Conclusion et perspectives

Afin de permettre aux usagers l'accès le plus pertinent à l'information à travers un processus d'autocomplétion de la requête soumise, nous avons proposé dans ce papier d'exploiter la structure des folksonomies, *i.e.*, les triplets (ressources, utilisateurs, tags), pour introduire une nouvelle définition d'une base générique de règles d'association triadiques, appelée $BGR\mathcal{T}$. Nous avons ainsi montré que $BGR\mathcal{T}$ remplit les conditions de non perte d'information et de non redondance des règles triadiques. L'étude expérimentale menée sur deux bases de différents types, *i.e.*, dense et éparses, extraites à partir de la folksonomie DEL.ICIO.US a mis en exergue que $BGR\mathcal{T}$ apporte d'importants gains en termes de compacité. Les perspectives de prolongement du présent travail sont résumées dans les points suivants :

- la définition d'une nouvelle base générique adéquate aux "représentations concises" : en effet, un courant émergent de l'algorithmique d'extraction des règles d'association préconise l'extraction efficace de ce type de représentations, *e.g.*, les motifs essentiels ou les motifs non dérivables (Hamrouni *et al.*, 2009). À cet effet il serait intéressant de préconiser la définition d'un système axiomatique générique, qui pourrait être instancié pour chaque base générique extraite à partir d'une représentation concise ;

- l'exploitation des relations non taxonomiques entre tags : en effet, la prise en considération des relations non taxonomiques entre tags peut s'avérer d'un grand intérêt pour l'autocomplétion de requêtes et ce notamment par l'exploitation des ontologies en ligne et des ressources lexicales externes, *i.e.*, WORDNET et WIKIPÉDIA (Trabelsi *et al.*, 2010). Par exemple, la relation non taxonomique suivante "java provides library", extraite à partir de WIKIPÉDIA entre les tags "java" et "library" pourrait fort bien être exploitée pour améliorer le taux de couverture moyen des requêtes soumise. En effet, dans le cas de la requête $q = \text{java}$, COMP_USER_QUERY aura non seulement à explorer les règles de $BGR\mathcal{T}$ contenant en prémisse le tag "java" mais également celles qui contiennent le tag "library".

Remerciements

Ce travail a été partiellement financé par le projet CMCU 11G1417.

Bibliographie

- Agrawal R., Imielinski T., Swami A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the ACM-SIGMOD international conference on management of data*, p. 207–216. Washington D. C., USA.
- Amer-Yahia S., Galland A., Stoyanovich J., C.Yu. (2008). From del.icio.us to x.qui.site: recommendations in social tagging sites. In *Proceedings of the ACM SIGMOD international conference on management of data*, p. 1323-1326. Vancouver, ACM Press.
- Bach T. L. (2006). *Construction d'un web sémantique multi-points de vue*. Dissertation de doctorat d'état en sciences, École des Mines de Nice, Sophia Antipolis, France.

- Bastide Y., Pasquier N., Taouil R., Stumme G., Lakhil L. (2000). Mining minimal non-redundant association rules using frequent closed itemsets. In *Proceedings of the 1st international conference on computational logic*, vol. 1861, p. 972–986. London, UK, Springer-Verlag.
- Beeri C., Dowd M., Fagin R., Statman R. (1984). On the structure of armstrong relations for functional dependencies. *Journal of the ACM*, vol. 31, p. 30–46.
- BenYahia S., Gasmi G., Nguifo E. M. (2009a). A new generic basis of factual and implicative association rules. *Intelligent Data Analysis - An International Journal (IDA)*, vol. 13, n° 4, p. 633-656.
- BenYahia S., Gasmi G., Nguifo E. M. (2009b). A new generic basis of "factual" and "implicative" association rules. *Intelligent Data Analysis*, vol. 13, n° 4, p. 633-656.
- BenYahia S., Nguifo E. M. (2004). Approches d'extraction de règles d'association basées sur la correspondance de Galois. In J.-F. Boulicaut, B. Cremilleux (Eds.), *Revue d'ingénierie des systèmes d'information*, vol. 9, p. 23-55. Hermès-Lavoisier.
- Bischoff K., Firan C. S., Nejd W., Paiu R. (2008). Can all tags be used for search? In *Proceedings of the 17th ACM conference on information and knowledge management*, p. 193-202. Napa Valley, California, ACM Press.
- Cuppens F., Demolombe R. (1989). How to recognize interesting topics to provide cooperative answering. *Information Systems Journal*, vol. 14, n° 2, p. 163-173.
- D.Yin, Z.Xue, L.Hong, Davison B. (2010). A probabilistic model for personalized tag prediction. In *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*, p. 959-968. Washington, USA, ACM Press.
- Feno D. R. (2007). *Mesures de qualité des règles d'association : normalisation et caractérisation des bases*. Doctorat de l'université de la réunion, L'Université de La Réunion.
- Frantz V., Shapiro J. (1991). Algorithms for automatic construction of query formulations in boolean form. *Journal of the American Society for Information Science*, vol. 1, n° 42, p. 16-26.
- Garg N., Weber I. (2008, October). Personalized tag suggestion for flickr. In *Proceedings of the international conference on world wide web*, p. 1063-1064. Beijing, ACM Press.
- Hamrouni T., BenYahia S., Nguifo E. M. (2009). Sweeping the disjunctive search space towards mining new exact concise representations of frequent itemsets. *International Journal of Data and Knowledge Engineering*, vol. 68, n° 10, p. 1091-1111.
- Hotho A. (2010). Data mining on folksonomies. In *Intelligent information access*, vol. 301, p. 57-82. Springer.
- Hotho A., Jäschke R., Schmitz C., Stumme G. (2006). Information retrieval in folksonomies : Search and ranking. In *Proceedings of the 3rd european semantic web conference*, vol. 4011, p. 411-426. Springer, Heidelberg.
- Jäschke R., Hotho A., Schmitz C., Ganter B., G.Stumme. (2008). Discovering shared conceptualizations in folksonomies. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, p. 38-53.
- Karen H. L., Sutter T., Marinho L. B., Schmidt-Thieme L. (2008). Tag-aware recommender systems by fusion of collaborative filtering algorithms. In *Proceedings of the ACM symposium on applied computing*, p. 1995–1999. New York, USA, ACM.

- Krause B., Hotho A., Stumme G. (2008). A comparison of social bookmarking with traditional search. In *Proceedings of the 30th european conference on IR research, advances in information retrieval*, vol. 4956, p. 101-113. Springer.
- Kryszkiewicz M. (2002). Concise representations of association rules. In *Proceedings of the ESF exploratory workshop on pattern detection and discovery in data mining*, vol. 2447, p. 92–109. London, UK, Springer-Verlag.
- Motro A. (1987). Extending the relational Databases Model to support Goal queries. In *Expert database systems, lerschberg, l., (eds.)*, p. 129–150. The Benjamin/Cummings Publishing Company.
- Nguyen T., Cerf L., Plantevit M., Boulicaut J.-F. (2010). Discovering Inter-Dimensional Rules in Dynamic Graphs. In *Workshop on dynamic networks and knowledge discovery, co-located with ECML PKDD 2010*. Barcelona, Spain, CEUR Workshop Proceedings.
- Nguyen T., Cerf L., Plantevit M., Boulicaut J.-F. (2011). Multidimensional association rules in boolean tensors. In *Proceedings of 11th SIAM international conference on data mining*. Arizona, USA.
- Pasquier N. (2000). *Datamining : Algorithmes d'extraction et de réduction des règles d'association dans les bases de données*. Thèse de doctorat, Université Clermont Ferrand 2, France.
- Schmitz C., Hotho A., Jäschke R., Stumme G. (2006). Mining association rules in folksonomies. In *Proceedings of data science and classification*. Springer.
- Trabelsi C., Jelassi N., BenYahia S. (2012). Scalable mining of frequent tri-concepts from folksonomies. In *Proceedings of the 16th pacific-asia conference on knowledge discovery and data mining*, p. 231-242. Kuala Lumpur, Malaysia, Springer-Verlag.
- Trabelsi C., Jrad A. B., BenYahia S. (2010). Bridging folksonomies and domain ontologies: Getting out non-taxonomic relations. In *Proceedings of the 10th IEEE international conference on data mining workshops*, p. 369-379. Sydney, Australia, IEEE Computer Society.