
Auto-complétion de requêtes par une base générique de règles d'association triadiques

Ch. Trabelsi, N. Jelassi et S. Ben Yahia

Faculté des Sciences de Tunis, Tunis, Tunisie.

RÉSUMÉ. L'exploitation des réseaux sociaux et systèmes collaboratifs dans le cadre d'un processus de recherche d'information est un phénomène qui peu à peu est intégré aux 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 légères représentées comme des hypergraphes tripartites d'utilisateurs, de tags et de ressources, appelées folksonomies. Dans ce papier, nous exploitons ces triplets (Resources, Utilisateurs, Tags) 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 non redondantes pour l'auto-complétion de requêtes permet de mettre en exergue la pertinence des folksonomies et leur intérêt réel dans le cas de la recherche d'information. Les premiers résultats obtenus sur une folksonomie réelle, collectée à partir du réseau social DELICIOUS, s'avèrent prometteurs et ouvrent de nombreuses perspectives.

ABSTRACT. The exploitation of social networks and collaborative systems as part of a process of information retrieval is a phenomenon that is gradually integrated into the practices of Internet research. These systems, emerging from the era of Web 2.0, allow participants to collaborate using keywords, tags, creating lightweight structures represented as tripartite hypergraphs of users, tags and resources, aka folksonomies. 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 non-redundant rules for auto-completion feature allows applications to highlight the relevance of folksonomies and their real interest in the case of information retrieval. The first results obtained on a real folksonomy, collected from the social network DELICIOUS, are promising and offer many opportunities.

MOTS-CLÉS : folksonomies, auto-complétion de requêtes, règles d'association triadiques

KEYWORDS: folksonomies, query auto-completion, triadic association rules

1. Contexte et motivations

Le social tagging s'est récemment imposé dans le paysage du web social et collaboratif (Web 2.0) comme un support à l'organisation de ressources partagées en permettant aux utilisateurs de catégoriser leurs ressources en leur associant des mots-clés, appelés *tags*. La structure, ainsi créée, résultante de la collecte de tags, appelée *folksonomie*, peut être considérée comme un hypergraphe tripartite d'utilisateurs, de tags et de ressources. Cette structure spécifique aux *folksonomies* a rendu leur exploitation d'un grand intérêt pour la recherche d'information dans la mesure où elles permettent d'identifier et de surveiller l'émergence de nouveaux concepts. 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'informations (*SRI*) typique retourne la même liste de résultats pour une même requête soumise par des utilisateurs ayant pourtant des besoins d'informations différents. Par exemple, pour 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 dans la mesure où des algorithmes du domaine de l'extraction de connaissances des bases de données sont exploités.

Par ailleurs, ces *SRI*s s'avèrent inadaptés au contexte du web collaboratif dans la mesure où elles ne prennent pas en compte l'aspect social et comportemental des usagers des *folksonomies*. Le domaine de la recherche d'information dans les *folksonomies* est ainsi apparu avec la promesse de permettre aux usagers l'accès le plus pertinent à l'information.

Plusieurs approches ont alors été proposées. Certaines tentent d'étendre les approches de recherches 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 *et al.*, 2008), (Karen *et al.*, 2008) et (Amer-Yahia *et al.*, 2008)). En effet, 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 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.

Cependant, d'autres travaux proposent d'adapter les techniques de suggestion ou de complétion de requêtes sur le web, généralement basées sur les techniques de fouille des fichiers logs, à la complétion des requêtes dans les *folksonomies* et ce, en exploitant 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) et (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, les auteurs dans (Schmitz *et al.*, 2007), 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. Cependant, mise 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, dans la mesure où le changement de dimension causera inéluctablement une perte d'information. En effet, 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, ayant une utilité moindre pour l'utilisateur, l'amènent à relancer son processus de recherche afin d'accéder aux ressources correspondantes.

De manière générale, nous nous proposons de répondre à deux principaux défis pour permettre à l'utilisateur l'accès le plus pertinent à l'information, à savoir : (i) l'extraction d'un sous-ensemble réduit de règles d'association triadiques, appelée base générique contenant les règles qui véhiculent le maximum d'information utiles ; (ii) l'utilisation de ces règles pour la complétion de requêtes dans les *folksonomies*.

Dans ce papier, nous proposons une approche d'auto-complétion 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'auto-complétion de requêtes. En effet, l'objectif de l'auto-complétion est d'ajouter des informations complémentaires et utiles à la requête d'un utilisateur ayant interrogé une *folksonomie*.

Le reste de l'article est organisé comme suit. La section 2 introduit les notions de l'extraction des règles d'association triadiques et leur connexion avec l'Analyse Formelle des Concepts. Nous consacrons la section 3 à la définition d'une nouvelle base informative des règles d'association triadiques, appelée *BGRT*. 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 4 qui décrit le processus d'auto-complétion de requêtes par la *BGRT* argumenté par un exemple illustratif. Les résultats des expé-

rimentations menées sur un jeu de données réel sont présentés dans la section 5. La conclusion et les travaux futurs font l'objet de la section 6.

2. Extraction des règles d'association triadiques (RATs)

Avant de définir les RATs et de décrire leur processus d'extraction, nous allons tout d'abord commencer par donner les définitions y 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 $\mathcal{Y} \subseteq \mathcal{U} \times \mathcal{T} \times \mathcal{R}$ représente une relation triadique dont chaque $y \in \mathcal{Y}$ peut être représenté par un triplet :

$$y = \{(u, t, r) \mid u \in \mathcal{U}, r \in \mathcal{T}, r \in \mathcal{R}\}.$$

ce qui signifie que l'utilisateur u a annoté la ressource r par le tag t .

Exemple 1 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} .

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	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

Tableau 1. Une folksonomie.

Nous présentons dans ce qui suit la définition d'un concept triadique.

Définition 2 (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 $\mathcal{TC}_{\mathcal{F}} = \{\mathcal{TC}_i \mid \mathcal{TC}_i = (U, T, R) \in \mathcal{Y} \text{ est un tri-concept}, i = 1 \dots n\}$.

Ainsi, pour l'extraction des tri-concepts à partir d'une *folksonomie*, nous avons utilisé l'algorithme TRIAS proposé par (Jäschke *et al.*, 2006). En effet, TRIAS prend en entrée la *folksonomie* $\mathcal{F} = (\mathcal{U}, \mathcal{T}, \mathcal{R}, \mathcal{Y})$ ainsi que les valeurs de supports minimaux à savoir, $u - minsupp$, $t - minsupp$ et $r - minsupp$ et retourne l'ensemble des tri-concepts fréquents. L'exécution de TRIAS sur la *folksonomie* illustrée par le Tableau 1, avec $u - minsupp = 2$, $t - minsupp = 2$ et $r - minsupp = 2$, permet d'obtenir les tri-concepts \mathcal{TC}_1 , \mathcal{TC}_2 et \mathcal{TC}_3 avec $\mathcal{TC}_1 = \{(u_5, u_7), (t_2, t_3, t_4), (r_1, r_2, r_3)\}$, $\mathcal{TC}_2 = \{(u_2, u_3, u_4, u_7), (t_1, t_4), (r_2, r_3)\}$ et $\mathcal{TC}_3 = \{(u_2, u_3, u_7), (t_1, t_2, t_3, t_4), (r_2, r_3)\}$. Ainsi, le tri-concept \mathcal{TC}_1 , par exemple, signifie que la communauté d'utilisateurs identifiés par (u_5, u_7) ont tous annoté les ressources (r_1, r_2, r_3) par les tags (t_2, t_3, t_4) .

Par ailleurs, nous admettons que deux tri-concepts sont dits *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 .

Nous définissons, dans ce qui suit, ces deux notions de similarité :

La 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) :

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

On définit ainsi, la similarité normalisée par : $Sim_1(\mathcal{V}_{U_1}, \mathcal{V}_{U_2}) = \frac{|\mathcal{V}_{U_1} \cap \mathcal{V}_{U_2}|}{Sim(\mathcal{V}_{U_1}, \mathcal{V}_{U_2})}$.

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

La 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. Ainsi, pour calculer la similarité entre r_i et r_j , nous avons utilisé la mesure de similarité normalisée *PLSC*, énoncée dans (Bach, 2006), permettant de calculer la longueur de la plus longue sous-chaîne commune entre r_i et r_j . On peut alors estimer la similarité entre deux ressources comme suit :

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

Nous admettons qu'un bon degré de similarité est atteint lorsque la plus longue sous chaîne commune entre r_i et r_j représente au moins 75% de la longueur de r_i .

2.1. Génération des règles d'association triadiques

La formalisation du problème d'extraction des règles d'association a été introduite par Agrawal *et al.* dans (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 $minsupp^1$ fixé par l'utilisateur, I est dit *fréquent*.

Une *règle d'association* est définie comme une implication entre deux itemsets. Elle est de la forme suivante : $R : X \Rightarrow Y$ tel 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 jeu les corrélations entre les tags d'un tri-concept. En effet, la structure la plus intéressante à étudier dans les *folksonomies* concerne l'ensemble des tags étant donné leur richesse sémantique et le fait qu'ils constituent le mode de représentation des *folksonomies* et ce, indépendamment de la nature des ressources qui la composent (Schmitz *et al.*, 2006).

Définition 3 (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*

1. Le nombre de transactions minimal auxquelles doit appartenir un *itemset* I pour être qualifié de fréquent.

$T_1 \rightarrow_{U,R} (T_2 \setminus T_1)$ 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 RATs.

Nous introduisons, par conséquent, deux nouvelles mesures de qualité des RATs, à savoir le *support triadique conditionné* et la *confiance triadique conditionnée* que nous définissons comme suit :

Définition 4 (SUPPORT TRIADIQUE CONDITIONNÉ) Soit T un ensemble de tags, $TC = (U, T, R)$ un tri-concept et \mathcal{TC}_C l'ensemble des tri-concepts similaires à 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 1 La mesure $supp_c$ est une mesure qui vérifie la propriété d'anti-monotonie.

Preuve 1 Soient T_1 et T_2 deux ensembles de tags tel que $T_1 \subseteq T_2 \subseteq T$. Montrons que $supp_c(T_1) \geq supp_c(T_2)$. Nous avons : $supp_c(T_1) = \frac{\prod_{t_i \in T_1} Supp(t_i) \mid t_i \in \mathcal{TC}_C.T}{|\mathcal{TC}_C.T|}$ et $supp_c(T_2) = \frac{\prod_{t_j \in T_2} Supp(t_j) \mid t_j \in \mathcal{TC}_C.T}{|\mathcal{TC}_C.T|}$. Ainsi, $\frac{supp_c(T_1)}{supp_c(T_2)} = \frac{\prod_{t_i \in T_1} Supp(t_i) \mid t_i \in \mathcal{TC}_C.T}{\prod_{t_j \in T_2} Supp(t_j) \mid t_j \in \mathcal{TC}_C.T} = \frac{1}{\prod_{t_k \in (T_2 \setminus T_1)} Supp(t_k) \mid t_k \in \mathcal{TC}_C.T} = \frac{1}{supp_c(T_2 \setminus T_1)} \geq 1$ puisque $supp_c(T_2 \setminus T_1) \in]0, 1]$. Par conséquent, $supp_c(T_1) \geq supp_c(T_2)$.

Définition 5 (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 6 (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 : $conf_c(R) = \frac{supp_c(T)}{supp_c(T_1)} = supp_c(T \setminus T_1)$

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

Dans cette section, nous introduisons notre base générique pour les RATs. Cette approche repose sur l'extraction d'un sous-ensemble compact, de toutes les RATs, appelé *base générique*, tout en satisfaisant 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 RATs 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 RATs.

Définition 7 (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 (de) $R_2 : T''_2 \Rightarrow_{U,R} T'_2$ 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 nous intéressons qu'aux requêtes d'un seul tag lors du processus, nous définissons la Base Générique de RATs comme suit :

Définition 8 (BASE GÉNÉRIQUE DE RÈGLES TRIADIQUES) *Soit \mathcal{TC} l'ensemble des concepts triadiques. Une base générique de règles triadiques BGR_T ² est l'ensemble des règles triadiques $t \rightarrow_{U,R} (T \setminus t)$ où $(U, T, R) \in \mathcal{TC}$.*

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

Ainsi, les règles génériques de BGR_T représentent des implications entre des prémisses minimales et des conclusions maximales (en terme de nombre de tags). En effet, dans (Pasquier, 2000), l'auteur 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 BGR_T , à savoir la *dérivabilité* et l'*informativité*.

2. BGR_T est l'acronyme de Base Générique de Règles Triadiques.

1) **Dérivabilité** : Afin de dériver l'ensemble de toutes les règles valides, nous utilisons dans la suite le système axiomatique d'Armstrong (Beeri *et al.*, 1984) et nous prouvons qu'il est valide et complet. Par ailleurs, étant donné la particularité des RATs considérées, *i.e.*, constituées d'un seul tag en prémisses, nous nous suffisons aux seuls axiomes d'Augmentation et de Décomposition.

Augmentation : Si $conf_c(R_1 : T_1 \Rightarrow_{U,R} T \setminus T_1) \geq minconf_c$ alors, $conf_c(R_2 : (T_1 \cup T_2) \Rightarrow_{U,R} (T \setminus T_2)) \geq minconf_c ; T_2 \subset T$.

Décomposition : Si $conf_c(R_1 : T_1 \Rightarrow_{U,R} T_2 \cup T_3) \geq minconf_c$ alors, $conf_c(R_2 : (T_1) \Rightarrow_{U,R} (T_2)) \geq minconf_c$ et $conf_c(R_3 : (T_1) \Rightarrow_{U,R} (T_3)) \geq minconf_c ; T_2, T_3 \subset T$.

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

Proposition 1 Soient \mathcal{BGRT} 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 2 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 $supp_c(R_1) = supp_c(R_2)$. Par ailleurs, nous avons $conf_c(R_1) = \frac{supp_c(T)}{supp_c(T_1)}$. D'autre part, nous avons $conf_c(R_2) = \frac{supp_c(T)}{supp_c(T_1 \cup T_2)}$. Puisque $T_1 \subseteq T_1 \cup T_2$ alors $supp_c(T_1 \cup T_2) \leq supp_c(T_1)$, nous déduisons que $\frac{supp_c(T)}{supp_c(T_1 \cup T_2)} \geq \frac{supp_c(T)}{supp_c(T_1)}$. Par conséquent, $conf_c(R_2) \geq conf_c(R_1)$ et $supp_c(R_2) = supp_c(R_1)$.

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 2 Le système axiomatique proposé est complet : toutes les règles pouvant être extraites de l'ensemble \mathcal{TC} sont dérivables de \mathcal{BGRT} moyennant le système axiomatique proposé.

Preuve 3 Soit \mathcal{BGRT} 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{BGRT} 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{BGRT} étant donné qu'elle contient un seul tag en prémisses 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 $R_2 \in \mathcal{BGRT}$ contenant le tag t de T_1 en prémisses et l'ensemble $(T \setminus T_1)$ en conclusion.

- Sinon, il existe une règle $R_2 \in \mathcal{BGRT}$ contenant en prémisses 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} 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 la \mathcal{BGR} pour prouver l'exactitude de ces mesures.

Proposition 3 La \mathcal{BGR} est informative i.e., les mesures de support et de confiance conditionnés des règles redondantes peuvent être déterminées avec exactitude.

Preuve 4 Soit la règle $R_1 : t \Rightarrow_{U,R} (T \setminus t) \in \mathcal{BGR}$ et $R_2 : T_1 \Rightarrow_{U,R} (T_2 \setminus T_1)$ une règle redondante appartenant à \mathcal{RT} . Peut-on déterminer les support et confiance 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, d'ailleurs, 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 la \mathcal{BGR} , qui contiennent tous un seul tag. Par conséquent, le support conditionné et la confiance conditionnée de toutes les règles triadiques peuvent être déterminés avec exactitude.

Exemple 2 Soit le tri-concept $TC_1 = (\{u_5, u_7\}, \{t_2, t_3, t_4\}, \{r_1, r_2, r_3\})$ extrait à partir du Tableau 1. L'ensemble total des RATs 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 $conf_c(R) = \frac{conf_c(R_1)}{supp_c(t_2)} \geq conf_c(R_1)$

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

4. Auto-Complétion de requêtes par la *BGRT*

L'objectif de la complétion de requêtes, domaine initialement abordé par (Motro, 1987), (Cuppens *et al.*, 1989) et (Frantz *et al.*, 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 9 (AUTO-COMPLÉTION DE REQUÊTES) *Une auto-complétion de requête (ou complétion de requête) \mathcal{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 ainsi complétée.*

4.1. Problématique de l'Auto-Complétion de requêtes

Étant donné 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'auto-complétion de requêtes revient par conséquent à :

1. 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 prémisses le tag de la requête Q .
2. Trouver l'ensemble maximal d'utilisateurs relatifs à cet ensemble maximal de tags. Cela revient à étudier les groupes d'utilisateurs contenant u_s et relatifs aux règles triadiques ayant en prémisses le tag de la requête Q .

Le processus d'auto-complétion de requêtes à partir de la *BGRT* est décrit par l'algorithme `COMP_USER_QUERY` donné par l'algorithme 1. `COMP_USER_QUERY` prend en entrée la *BGRT* ainsi que la requête Q soumise par l'utilisateur u_s et retourne la requête complétée \mathcal{C}_Q . `COMP_USER_QUERY` procède par un parcours de l'ensemble des RATs de la *BGRT* 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 triadique R_c .
- 2) le tag de la requête Q est contenu dans la prémisse de la règle triadique R_c .

Ainsi, après détermination de 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 \mathcal{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 \mathcal{C}_Q .

Données : - La base générique des règles triadiques \mathcal{BGR}_T , Q et u_s .

Résultats : - La requête complétée \mathcal{C}_Q .

```

1 début
2    $BestConfidence := 0; \mathcal{C}_Q := Q;$ 
3   pour chaque Règle Triadique  $RT$  de  $\mathcal{BGR}_T$  faire
4     si  $u_s \in RT.U$  alors
5       si  $Q \in RT.T.Premisse$  alors
6         si  $Confiance(RT) \geq BestConfidence$  alors
7            $RT_c := RT;$ 
8            $BestConfidence := Confiance(RT);$ 
9   si  $RT_c \neq null$  alors
10     $\mathcal{C}_Q.M_c = \mathcal{C}_Q \cup RT_c.U;$ 
11     $\mathcal{C}_Q.T_c = \mathcal{C}_Q \cup RT_c.T.Conclusion;$ 
12  retourner  $\mathcal{C}_Q;$ 
13 fin

```

Algorithme 1. COMP_USER_QUERY

Le processus d'auto-complétion de requêtes se base ainsi sur la \mathcal{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 les mêmes centres intérêts dans les ressources recherchées.

4.2. Exemple illustratif du processus d'Auto-Complétion de requêtes par la \mathcal{BGR}_T

Afin d'illustrer le processus d'auto-complétion de requêtes par la \mathcal{BGR}_T , nous prenons comme exemple un ensemble de règles extraites, décrites dans le Tableau 3, 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 définie par le tag *nasa*, i.e., $Q := \{nasa\}$. Afin de compléter Q , COMP_USER_QUERY explore la \mathcal{BGR}_T pour identifier les RATs dont le groupe d'utilisateurs contient *Jacob* et dont le tag en prémisses 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, ainsi générées à partir du Tableau 3 sont :

$$nasa \Rightarrow_{(u_1, u_5), (r_1, r_2, r_3)} photos, space, images, free, stock, science, Conf_c = 0, 37,$$

$$nasa \Rightarrow_{(u_1, u_6, u_7), (r_1, r_2)} satellite, mars, images, science, Conf_c = 0, 71,$$

$$nasa \Rightarrow_{(u_1, u_3, u_4, u_7), (r_2, r_3)} publicdomain, images, science, Conf_c = 0, 55.$$

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 : $C_Q = \{(nasa, satellite, mars, images, science), (u_1, u_6, u_7)\}$.

5. Etude expérimentale

Afin d'évaluer la compacité de la BGR_T par rapport à l'ensemble total des RATs, nous avons mené une série d'expérimentations sur un jeu de données collecté à partir du site collaboratif DEL.ICIO.US ³

Nous avons ainsi considéré cinq *snapshots* différents appartenant à la même période (Janvier 2007) et exécuté, par la suite l'algorithme TRIAS séparément. Le tableau 2 résume les différents résultats obtenus.

Une fois les tri-concepts extraits, nous avons procédé à l'extraction de BGR_T et ce, pour chaque *snapshot*. Pour déterminer l'ensemble des RATs valides, nous avons utilisé les valeurs de seuils suivantes : $minsupp_c = 0, 10$ et $minconf_c = 0, 10$.

$ \mathcal{F} $	100.000	200.000	300.000	400.000	500.000
$ \mathcal{TC} $	30	132	256	353	450

Tableau 2. Le nombre de tri-concepts extraits à partir des snapshots de DEL.ICIO.US

Nous avons ainsi relevé que pour toutes les valeurs de TC , la taille de la BGR_T est très largement inférieure à l'ensemble total des règles extraites. Tandis que le nombre de règles triadiques de \mathcal{RT} grimpe jusqu'à 2, 357, 836 au terme de cette période, le nombre de règles triadiques de notre base générique demeure en dessous de 1, 387 ce qui représente seulement 0, 05% de l'ensemble total de règles triadiques. Cette première constatation est illustrée par la Figure 1.

La compacité de notre base générique est dû principalement à deux facteurs :

- La taille de l'ensemble de tags T des tri-concepts peut être grand (jusqu'à 10 tags) : si la taille de BGR_T ne s'en trouve pas affectée, le nombre de ses règles étant égal à la taille de T , la cardinalité de \mathcal{RT} augmente sensiblement de quelques centaines de règles. Ceci revient aux multiples combinaisons possibles de tags pour les RATs de \mathcal{RT} .

- La forme particulière des RATs de la BGR_T , constituées d'un seul tag en pré-misse et un d'un nombre maximal de tags en conclusion est un avantage certain étant donné que ce type de règles véhicule le maximum d'informations tout en évitant la redondance d'un très grand nombre de règles (Pasquier, 2000).

3. Le jeu de données est disponible sur le site <http://data.dai-labor.de/corpus/delicious/>

Par ailleurs, nous observons que la base générique BGR_T réduit entre 99, 11 % et 99, 95 % le nombre des règles triadiques extraits à partir des tri-concepts.

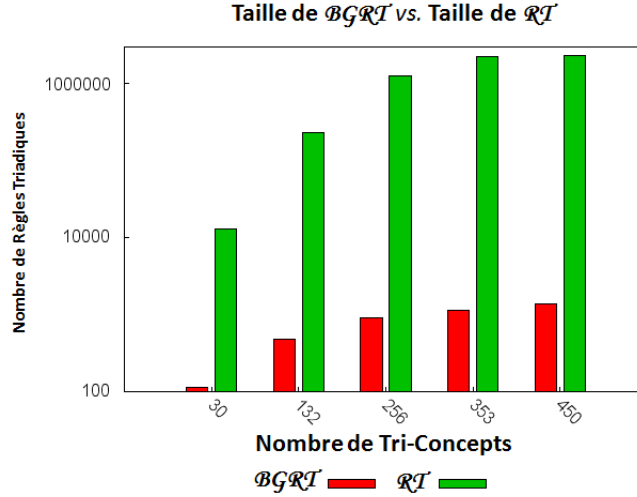


Figure 1. Taille de la base générique BGR_T vs. l'ensemble total des règles triadiques RT .

U	BGR_T	R	$supp_c$	$conf_c$
Angie, Lynavo	sitemap->service,answering,online,business,phone,internet		0, 819	0, 452
	answering->service,sitemap,online,business,phone,internet	R_1	0, 819	0, 726
	online->sitemap,answering,service,business,phone,internet	R_2	0, 819	0, 887
	business->service, sitemap,answering,online,phone,internet	R_3	0, 819	0, 717
	phone->service,sitemap,answering,online,business,internet		0, 819	0, 213
Jacob, Brson	space->photos,nasa,images,free,stock,science		0, 571	0, 653
	photos->space,nasa,images,free,stock,science	R_4	0, 571	0, 592
	nasa->photos,space,images,free,stock,science	R_5	0, 571	0, 371
	images->photos,nasa,space,free,stock,science	R_6	0, 571	0, 653
	science->photos,nasa,images,free,stock,space		0, 571	0, 592

Tableau 3. La base générique de règles triadiques extraite à partir d'un jeu de données de DEL.ICIO.US

6. Conclusion et Perspectives

Dans ce papier, et afin de permettre aux usagers l'accès le plus pertinent à l'information à travers un processus d'auto-complétion de la requête soumise, nous avons exploité 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_T . Nous avons ainsi montré que la BGR_T remplit les conditions de non perte d'information et de non redondance des règles triadiques. Par

<i>Ressource</i>	URL du Site Web
R_1	http://answeringserviceonline.info/answer-service/
R_2	http://respondinghelp.info/telephone-answering-service/
R_3	http://answeringassistance.info/answering-service-uk/
R_4	http://visibleearth.nasa.gov/
R_5	http://www.nasa.gov/multimedia/highlights/index.html
R_6	http://www.unclesamsphotos.com/

Tableau 4. Les URL correspondant aux ressources du Tableau 3

ailleurs, étant donné le nombre considérable des triplets dans les *folksonomies* et le nombre de règles d'association pouvant être générées, nous avons utilisé les concepts triadiques. L'étude expérimentale menée sur la *folksonomie* DEL.ICIO.US a mis en exergue que la *BGR*T apporte d'importants gains en terme de compacité. Les perspectives de prolongement du présent travail sont comme suit : (i) L'intégration de la base *BGR*T dans un environnement de visualisation (Yahia *et al.*, 2004). Ceci permet d'améliorer l'interaction Homme/Machine où l'utilisateur aura des informations additionnelles relatives à une règle particulière (*e.g.*, règles redondantes à la règle générique sélectionnée); (ii) La considération des requêtes composées, *i.e.*, contenant plus d'un tag. Ceci revient à adapter le processus d'auto-complétion notamment en y incluant un module d'indexation de requêtes; (iii) L'intégration d'un nouveau module de stématisation et de désambiguïsation sémantique des tags et ce notamment, par l'exploitation des ontologies en ligne et des ressources lexicales externes, *i.e.*, WORDNET et WIKIPÉDIA.

7. Bibliographie

- Agrawal R., Imielinski T., Swami A., « Mining Association Rules between sets of items in large Databases », *ACM-SIGMOD International Conference on Management of Data*, Washington D. C., USA, p. 207-216, 1993.
- Amer-Yahia S., Galland A., Stoyanovich J., C. Yu, « From del.icio.us to x.qui.site : recommendations in social tagging sites », *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008*, ACM Press, Vancouver, p. 1323-1326, 2008.
- Bach T. L., Construction d'un web sémantique multi-points de vue, Dissertation de doctorat d'état en sciences, École des Mines de Nice, Sophia Antipolis, France, 2006.
- Beeri C., Dowd M., Fagin R., Statman R., « On the structure of Armstrong relations for functional dependencies », *Journal of the ACM*, vol. 31, p. 30-46, 1984.
- Bischoff K., Firan C. S., Nejd W., Paiu R., « Can all tags be used for search ? », *Proceedings of the 17th ACM conference on Information and knowledge management, CIKM 2008*, ACM Press, Napa Valley, California, p. 193-202, 2008.
- Cuppens F., Demolombe R., « How to recognize interesting topics to provide cooperative answering », *Information Systems Journal*, vol. 14, n° 2, p. 163-173, 1989.

Ch. Trabelsi, N. Jelassi et S. Ben Yahia

- D.Yin, Z.Xue, L.Hong, Davison B., « A probabilistic model for personalized tag prediction », *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, SIGKDD 2010*, ACM Press, Washington, USA, p. 959-968, 2010.
- Feno D. R., Mesures de qualité des règles d'association : normalisation et caractérisation des bases, Doctorat de l'université de la réunion, L'Université de La Réunion, Décembre, 2007.
- Frantz V., Shapiro J., « 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, 1991.
- Garg N., Weber I., « Personalized tag suggestion for Flickr », *Proceedings of the International Conference on World Wide Web, WWW 2008*, ACM Press, Beijing, p. 1063-1064, October, 2008.
- Hotho A., *Studies in Computational Intelligence*, vol. 301/2010, Springer, chapter Data Mining on Folksonomies, p. 57-82, 2010.
- Hotho A., Jäschke R., Schmitz C., Stumme G., « Information Retrieval in Folksonomies : Search and Ranking », *Sure, Y., Domingue, J. (eds.) The Semantic Web : Research and Applications. LNCS, vol. 4011, pp. 411-426*, Springer, Heidelberg, 2006.
- Jäschke R., Hotho A., Schmitz C., Ganter B., G.Stumme, « Discovering shared conceptualizations in folksonomies », *Web Semantics : Science, Services and Agents on the World Wide Web*, vol. 6, p. 38-53, 2008.
- Jäschke R., Hotho A., Schmitz C., Ganter B., Stumme G., « TRIAS - An Algorithm for Mining Iceberg Tri-Lattices », *Proceedings of the 6th IEEE International Conference on Data Mining, ICDM 2006*, IEEE Computer Society, Hong Kong, p. 907-911, 2006.
- Karen H., Sutter T., Marinho L. B., Schmidt-Thieme L., « Tag-aware recommender systems by fusion of collaborative filtering algorithms », *In ACM symposium on Applied computing. SAC 2008. ACM, pp. 1995-1999*, New York, NY, USA, 2008.
- Krause B., Hotho A., Stumme G., « A Comparison of Social Bookmarking with Traditional Search », *Proceedings of the 30th European Conference on IR Research, Advances in Information Retrieval, ECIR 2008*, vol. 4956, Springer, p. 101-113, 2008.
- Kryszkiewicz M. *Concise representations of association rules*, Springer-Heidelberg, September, 2002.
- Motro A., « Extending the relational Databases Model to support Goal queries », *Expert Database Systems, Lerschberg, L., (eds.)*, The Benjamin/Cummings Publishing Company, p. 129-150, 1987.
- Pasquier N., DataMining : algorithmes d'extraction et de réduction des règles d'association dans les bases de données, Doctorat d'université, Université de Clermont-Ferrand II, 2000.
- Schmitz C., Hotho A., Jäschke R., Stumme G., « Mining Association Rules in Folksonomies », *Proceedings of Data Science and Classification (IFCS 2006)*, Springer-Heidelberg, 2006.
- Schmitz C., Hotho A., Jäschke R., Stumme G., « Mining Association Rules in Folksonomies », *Proceedings of Data Science and Classification, IFCS 2006*, Springer, Heidelberg, Germany, p. 261-270, 2007.
- Yahia S. B., Nguifo E. M., « Emulating a cooperative behavior in a generic association rule visualization tool », *In Proceedings of 16th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2004*, Boca Raton, Florida, p. 148-155, November, 2004.