

UTILISATION DE SYSTEMES A BASE DE FOLKSONOMIES DANS UN PROCESSUS DE VEILLE : PROPOSITIONS POUR L'ORGANISATION ET L'EXPLOITATION DES INFORMATIONS COLLECTEES

Fabrice PIROLI

fabrice.pirolli@iut-dijon.u-bourgogne.fr

Université de Bourgogne

[Laboratoire CIMEOS / LIMSIC - EA 4177](#)

IUT Département Information-Communication

Boulevard du Dr Petitjean - BP 17867

21078 Dijon cedex

Tél : 03 80 39 65 29

Mots clefs : veille stratégique, folksonomies, taxonomies, organisation de l'information, associations folksonomiques.

Keywords: strategic intelligence, folksonomy, taxonomy, information organisation, folksonomic related terms.

Palabras clave : escudriñar estratégica, folksonomía, taxonomía, organización de la información, folksonomica asociacón.

Résumé

L'utilisation de folksonomies dans le cadre d'un processus de veille est un phénomène qui peu à peu est intégré aux pratiques de la veille sur l'Internet. Les folksonomies lorsqu'elles sont considérées comme sources d'information présentent un grand intérêt dans la mesure où elles permettent d'identifier et de surveiller l'émergence de nouveaux concepts ainsi que leurs représentations. Leur dimension collaborative et sociale facilite l'identification de cibles et d'experts. La diversité des ressources décrites au moyen de folksonomies permet d'élargir le champ des recherches d'information à des domaines parfois mal couverts par les outils plus traditionnels que sont les moteurs de recherche et les portails spécialisés. Cependant pour être utilisées dans le cadre d'un processus de veille les informations traitées par de tels systèmes doivent être analysées, traitées, organisées et stockées de façon spécifique. Nos propositions concernent particulièrement les aspects de d'organisation et de mémorisation des informations en présentant une approche hybride mêlant folksonomies et taxonomies fondée notamment sur la notion « d'association folksonomique ».

1. Introduction

L'utilisation de folksonomies dans le cadre d'un processus de veille est un phénomène qui peu à peu s'intègre aux pratiques de la veille sur l'Internet. Cependant les usages observés, bien souvent guidés par une approche empirique, ne semblent exploiter que partiellement les apports potentiels d'une telle source d'information [13]. De même ils sont rarement le prolongement d'une réflexion générale sur les spécificités en terme de contenu informationnel des systèmes à base de folksonomies. Dans le cadre de nos travaux nous proposons dans un premier temps, après avoir défini la notion de folksonomie, une analyse de leurs apports et de leurs limites pour un travail de veille stratégique. Dans un deuxième temps, compte tenu de nos constatations, nous exposerons une approche théorique reposant sur l'utilisation d'un système hybride diachronique mêlant taxonomies et folksonomies en plaçant la variable temporelle, capitale dans le domaine de la veille, au centre de l'étude. Enfin nous terminerons en présentant l'architecture d'un système dédié à la veille exploitant cette approche théorique.

2. Les apports des folksonomies pour la veille stratégique scientifique et technologique

2.1 Définitions

Le terme de folksonomie, néologisme issu de la contraction des termes anglais « folk » et « taxonomy » initialement proposé par Thomas Vander Wal [17], désigne des pratiques d'indexation de contenu à l'aide de mots-clés, les *tags*, choisis librement par les usagers du système. Les trois composants d'une folksonomie sont l'étiquetage, l'objet étiqueté et l'identité de l'étiqueteur. Celle-ci s'établit dans un environnement partagé et ouvert : l'information ajoutée par un individu est mise à la disposition de tous les membres appartenant au même réseau social. Les systèmes à base de folksonomies s'appuient sur les principes du web 2.0 [11] et plus particulièrement sur les notions de partage, de développement collaboratif et d'intelligence collective.

Dans les systèmes utilisant les folksonomies l'ensemble des *tags* constitue un mode de représentation du corpus traité, indépendamment de la nature des ressources qui le composent (texte, images, sons,...). Ainsi l'utilisation des folksonomies concerne plusieurs grandes catégories d'applications actuelles de l'internet : partage de signets, identification de billets sur les blogs, partage d'images, partage de fichiers, etc. Au delà de la représentation du corpus un système à base de folksonomies propose une organisation et une représentation d'un ensemble de connaissances fondées sur la mise en commun et l'agglomération des représentations individuelles au moyen du réseau d'échange formé par la communauté des utilisateurs [3].

2.2 Émergence et surveillance de concepts

Du point de vue de la veille, la première caractéristique remarquable des folksonomies est liée à leur grande simplicité d'utilisation et à la totale liberté laissée à leurs utilisateurs lors de la phase de description des ressources informationnelles, parfois qualifiée « d'indexation sociale » [8]. En effet les systèmes à base de folksonomies n'imposent aucune contrainte formelle ou sémantique à leurs usagers, tant lors des phases de saisies que lors des phases de recherche

d'information. L'ensemble des *tags* constitue un système de représentations non contrôlées dont le contenu n'est pas organisé conformément à un modèle pré-établi et dont l'utilisation n'est pas régie par des règles préexistantes (contrairement aux différents types de langages d'indexation utilisés traditionnellement en documentation). Dans de nombreux cas les auteurs des documents représentés par les folksonomies sont directement responsables de l'indexation de leurs documents et du choix des termes les représentant. De ce fait les folksonomies facilitent l'expression de nouveaux concepts dans un environnement donné car elles ont une très faible force d'inertie face à la nouveauté. Elles constituent un terrain favorable à l'émergence et à la propagation de nouvelles acceptions, généralement liées à l'usage, de termes existants ou encore à l'apparition de néologismes.

Dans la mesure où les folksonomies correspondent à une technologie dite *bottom-up* (dont les concepts émergent de la pratique) l'observation de l'évolution du corpus de *tags* reflète l'évolution de la production documentaire et des termes utilisés pour la décrire dans un environnement donné. Cette évolution peut se traduire par l'engouement ou a contrario la désaffection des utilisateurs pour certains termes dont la fréquence d'utilisation varie fortement au cours du temps. Ainsi les modifications du corpus folksonomique peuvent refléter une modification de l'environnement (technique, juridique, économique).

2.3 Dimension sociale

Les folksonomies ne peuvent être étudiés indépendamment de leurs contextes d'utilisation. Elles proposent une organisation et une représentation d'un ensemble de connaissances fondées sur la mise en commun et l'agglomération des représentations individuelles au moyen des réseaux d'échange formés par la communauté des utilisateurs. Chaque nœud de ces réseaux peut potentiellement être considéré comme une source d'information ou une cible à surveiller lors d'une démarche de veille [15]. L'identification de ces personnes-cibles peut être réalisée sur la concordance entre leurs activités sur le système de partage d'informations (thèmes abordés, domaine d'activité,...) et les points stratégiques sur lesquels porte la veille informationnelle ou encore sur le profil propre à chaque utilisateur [8]. Cette notion « d'identification d'experts » constitue l'un des apports clairement identifiés des folksonomies et plus largement des réseaux sociaux et des technologies du web 2.0 en terme de veille [13][17]. Il nous semble indispensable de prendre en compte cette dimension en adoptant une définition multi-critères de profils d'utilisateurs fondée sur l'identité de la personne, son secteur d'activité mais également ses pratiques en tant qu'utilisateur du système.

2.4 Richesse des ressources, nouvelles modalités d'accès à l'information

Les folksonomies permettent d'accéder à des ressources traditionnellement mal indexées par les systèmes de recherche d'information en raison de leur nature, de leur format, de leur structure : photos, vidéos, présentations assistées par ordinateurs, etc. Face à ce type de ressources les moteurs de recherche ne fournissent pas de résultats réellement satisfaisants car les procédés d'indexation automatique ne peuvent s'appliquer. Le recours à une médiation humaine effectuée par des spécialistes (annuaires, portails,...) donne des résultats satisfaisants mais les performances restent en inadéquation avec le volume et la rapidité d'évolution des ressources à traiter. Le recours aux folksonomies constitue dans de nombreux cas le moyen le plus rapide de proposer un accès à de telles ressources.

Enfin les folksonomies permettent un accès original à l'information selon deux modalités. La première concerne l'accès à l'information selon un processus de recherche dont la mesure de l'efficacité a déjà fait l'objet de travaux de recherche [8][10]. La seconde modalité de recherche s'inscrit dans le cadre d'une approche navigationnelle nourrie principalement par le principe de la découverte. Un tel mode d'accès à l'information s'appuie sur des mécanismes de

réenchantement / désenchantement que peut connaître l'utilisateur du système lors de son parcours et sur l'identification de nouveaux liens sémantiques. Les modes de présentation des *tags* souvent associés aux folksonomies en particulier les nuages de mots-clés exploitent ce principe de navigation. Les travaux de Sinclair et Cardew-Hall [16] montrent que ces modes de présentations ne peuvent se substituer à une interface de recherche classique par mot-clé. Néanmoins ils sont particulièrement utiles lors de phases exploratoires leur avantage principal étant de synthétiser de façon simple et visuelle le contenu des ressources indexées. Enfin l'exploitation de nuages de mots-clés implique pour l'utilisateur une charge cognitive moins importante que l'expression de requêtes.

L'ensemble de ces particularités a rapidement donné aux folksonomies l'image d'outils facilitateurs de sérendipité, notion particulièrement importante dans le domaine de la veille. A ce titre nous considérons que les folksonomies permettent de définir des relations d'associations sémantiques spontanées et contextualisées en fonction de la co-occurrence des *tags* employés, de l'identité de leur auteur et de la ressource informationnelle impliquée.

2.5 Limites du modèle théorique

Les folksonomies peuvent donc être considérées comme une source d'informations dont les spécificités peuvent apporter une réelle plus value lors d'un travail de veille sous réserve que les traitements qui leur sont appliqués prennent en compte toutes leurs particularités. Pour cela il nous semble important d'énoncer quelques réserves portant sur le modèle théorique afin de les prendre en compte lors de nos propositions.

2.5.1 Absence de contrôle du vocabulaire

La première critique couramment formulée à l'encontre des folksonomies concerne l'absence de contrôle du vocabulaire. Cette absence de contrôle est l'un des éléments qui expliquent la simplicité de l'utilisation des folksonomies néanmoins elle expose ses utilisateurs à tous les problèmes sémantiques et syntaxiques rencontrés dans la langue naturelle tels que la polysémie, la synonymie, auxquels s'ajoute l'absence de normalisation des termes qui rend l'identification d'un concept délicate. Ces difficultés sont traditionnellement levées par l'utilisation de langages d'indexations contrôlés de type listes d'autorités, lexique, taxonomies, thésaurus dont le mode de fonctionnement est à l'opposé de celui des folksonomies. Dans une démarche de veille il faut donc bien prendre en compte le fait qu'un système à base de folksonomies nourrit de nombreuses ambiguïtés langagières, d'autant plus présentes s'il s'agit d'un système multilingue.

2.5.2 Influences intra-personnelles et inter-personnelles

Notre seconde remarque concerne la neutralité supposée de l'outil. Les systèmes à base de folksonomies doivent uniquement traduire la représentation d'une ressource informationnelle telle qu'elle a été construite par un utilisateur. L'outil technique et son interface ne doivent être qu'un support permettant l'expression de cette représentation. Or dans la pratique un utilisateur subit dans son rapport au système un ensemble d'influences endogènes ou exogènes qui conditionnent son comportement. Les travaux menés par Golder et Huberman [5] indiquent que le taux d'accroissement du nombre de *tags* employés diffère très fortement entre individus au sein d'une population d'utilisateurs d'un même système.

Plusieurs facteurs expliquent ces variations parmi lesquels nous identifions particulièrement :

- la perception de l'outil technique et ses représentations : l'utilisateur d'un système à base de folksonomies en fera un usage correspondant à sa représentation propre des fonctionnalités de l'outil. Les tâches de description de ressources qui nécessitent un travail d'analyse et de synthèse ne seront pas menées de façon identique selon que l'utilisateur les considère comme une aide au classement, un procédé d'annotation, un mode de diffusion, etc. Une vision trop idiosyncrasique sera nécessairement très réductrice. Golder et Huberman [5] identifient sept types de d'usages différents des folksonomies : décrire le contenu de la ressource, le type de support, l'auteur des tags, des opinions sur la ressource, une appropriation des contenus, ou enfin une activité associée à la ressources. Les folksonomies ne peuvent donc être uniquement considérées comme un procédé d'indexation mais plutôt comme une forme particulière de métadonnées.
- l'influence du groupe : l'utilisateur de folksonomies s'inscrit dans le cadre d'un travail collaboratif et a conscience du fait qu'il se construit peu à peu une image dans cette communauté. Son comportement sera alors conditionné par un ensemble de facteurs psychologiques : jeux d'influences (majoritaire ou minoritaire), attitude face au groupe, tendance à la normalisation, au conformisme, etc. [1]
- l'interface : l'outil technique peut également directement influencer sur la construction des représentations. Particulièrement s'il suggère des termes aux utilisateurs (*tags* utilisés les plus fréquemment par exemple) favorisant de la sorte une convergence des représentations qui peut être perçue comme un processus d'uniformisation.

Tous ces jeux d'influences ont un impact sur les pratiques constatées et sur les modes d'utilisation des outils. L'hétérogénéité des résultats produits, tant dans leur forme que dans leurs objectifs, constitue l'un des principaux freins quant à l'utilisation des folksonomies comme unique modèle de représentation de connaissances. Néanmoins dans une approche centrée sur un travail de veille, la prise en compte de ces éléments nous paraît indispensable dans la mesure où ils conditionnent aussi bien le niveau de vocabulaire utilisé que le type et le mode de représentation de concepts qu'ils traduisent.

3. Propositions pour un système d'aide à l'exploitation de folksonomies dans une démarche de veille

Selon Le Deuff [8] et Pédaque [12] les folksonomies « *s'inscrivent dans un processus de désintermédiation dans le sens où la recherche d'information et son organisation se trouvent personnalisées et passent outre la médiation traditionnelle* ». Afin d'exploiter ce processus de personnalisation pour un travail de veille nous proposons de ré-introduire un étape de médiation, au sens documentaire du terme, qui s'appuiera sur l'utilisation de taxonomies. Dans un premier temps nous présenterons les éléments sur lesquels sera fondée cette médiation puis dans un second temps nous caractériserons les étapes de mémorisation et d'organisation des informations collectées.

3.1 Ré-introduction d'une étape de médiation

Sur la base des réflexions théoriques présentées en paragraphe 2 il nous semble intéressant que, dans le cadre d'une démarche de veille exploitant les folksonomies, les éléments suivants soient explicitement pris en compte :

- l'étude temporelle de l'évolution des *tags* (fréquence d'utilisation, cycle de vie,...) ;
- le comportement et le profil des utilisateurs du système ;
- l'identification et l'expression des concepts, ainsi que les relations qu'ils entretiennent, au sein du vocabulaire construit ;
- les ambiguïtés langagières et les particularités liées à la libre formulation des *tags* (polysémie, multilinguisme, néologismes,...).

Pour cela nous proposons une approche, que nous qualifions d'hybride et de diachronique, mobilisant simultanément l'usage de folksonomies et de taxonomies tout en accordant une place centrale à la variable temporelle. Nous proposons ainsi de réintroduire une démarche de médiation documentaire afin de pallier certains inconvénients inhérents au principe même de folksonomie et ainsi de proposer un mode d'organisation des informations plus propice à leur exploitation dans une démarche de veille. Compte tenu des spécificités des pratiques en matière d'indexation sociales cette approche ne peut s'appliquer que dans des contextes de veille très précis sur des thématiques clairement pré-établies.

Les taxonomies sont définies comme « *un cadre d'organisation pour des ressources numériques de toutes natures destiné à en permettre une présentation ordonnée et y donnant accès par navigation hypertextuelle* » [9]. Le choix de l'utilisation de taxonomies composées de descripteurs permet d'envisager l'emploi de relations hiérarchiques et de relations d'associations, tout en laissant des possibilités d'adaptation de l'outil aux besoins spécifiques de la veille. Le recours aux taxonomies permet en partie d'écarter certaines sources d'ambiguïtés langagières par la sélection des termes retenus pour décrire et organiser les ressources. Cette sélection est accompagnée de principes de normalisation du vocabulaire employé.

3.2 Mémorisation et organisation des informations collectées

3.2.1 Les informations collectées

L'approche hybride présentée pour objectif de fournir une aide à l'exploitation des informations représentées par les folksonomies dans une démarche de veille. Elle s'appuie pour cela sur la collecte d'informations caractérisant des ensembles de ressources identifiées sur le *web* au moyen d'outils de partage de signets. Les informations traitées concernent à la fois les ressources en tant qu'unités documentaires, le vocabulaire constitué par l'ensemble des *tags* et les personnes partageant ces ressources (nœuds des réseaux sociaux).

Les ressources informationnelles jugées pertinentes ou potentiellement pertinentes par rapports aux besoins exprimés en terme de veille sont caractérisées par un ensemble de données bibliographiques (auteurs du document, titre, adresse...) auxquelles s'ajoutent le contenu informationnel des tags. Afin de tenir compte des différents types d'informations que peuvent représenter les *tags* nous proposons de les qualifier contextuellement. Pour cela nous nous appuyons sur la typologie, résumée sur la figure 1, présentée par Roulleaux et Dugage [14] pour caractériser différents types de métadonnées. Nous distinguons ainsi informations de sujet, informations d'usage, information d'actif et informations relationnelles. Enfin chaque ressource est qualifiée par l'identifiant de la personne à l'origine de son partage.

↑	<p>Métadonnées de sujet Quoi : Sujet, description, mots-clés,...</p>	<p>Métadonnées d'usage Usages possibles : Habilitations, droits,...</p>
---	---	--

Indépendamment des informations issues de la recherche et de la collecte de ressources sur le *web*, l'approche présentée implique l'utilisation de taxonomies. Le processus de veille s'appuiera à chaque étape de son déroulement sur des taxonomies organisées et contrôlées par un ou des experts responsables du travail de veille. Lors de la phase d'initiation de ce travail les responsables de la veille constituent et organisent un ensemble de taxonomies qui constitueront le point de départ du processus de collecte. Ces taxonomies permettent de préciser l'environnement sémantique et les champs de connaissances dans lesquels s'inscriront les démarches de recherche d'information. Ce cadre d'organisation sera progressivement enrichi par les folksonomies liées aux ressources collectées. La particularité de cette approche est donc de combiner dans un même système de représentations descripteurs et *tags*. Le lexique ainsi constitué est composé de termes normalisés et d'expressions non contrôlées.

3.2.2 La relation d'association folksonomique

Au cours des phases de recherche, de sélection puis de collecte toutes les informations présentées en 3.2.1 sont intégrées au système de veille. Les *tags* subissent un traitement spécifique dans la mesure où ils seront qualifiés en fonction de leur nature avant d'être mémorisés. Deux situations types sont envisagées :

- le *tag* correspond à un descripteur contenu dans l'une des taxonomies : le concept qu'il représente est déjà connu ;
- le *tag* ne correspond pas à un descripteur déjà présent dans le système : le *tag* traduit un concept nouveau qui n'est pas encore représenté par les taxonomies ou bien il représente une variante orthographique ou syntaxique d'un concept déjà identifié.

Dans cette deuxième situation, et afin d'inclure le nouveau terme dans un système relationnel, nous définissons la notion de relation d'**association folksonomique** présentée dans la figure 2. Une telle relation sera établie entre *tags* - traduisant de la sorte la co-occurrence de termes en rapport à une ressource identifiée - ou entre *tag* et descripteur. Dans les deux cas cette association folksonomique n'est pas nécessairement de nature sémantique mais traduit l'existence d'un lien entre différents concepts dont l'origine est fondée à la fois par le contenu et la nature de l'unité informationnelle et par les caractéristiques propres de l'auteur de la folksonomie. Elle exprime donc la part de subjectivité induite par la pratique d'une indexation sociale sur la base du triptyque fondateur des folksonomies : étiquetage, objet étiqueté et identité de l'étiqueteur.

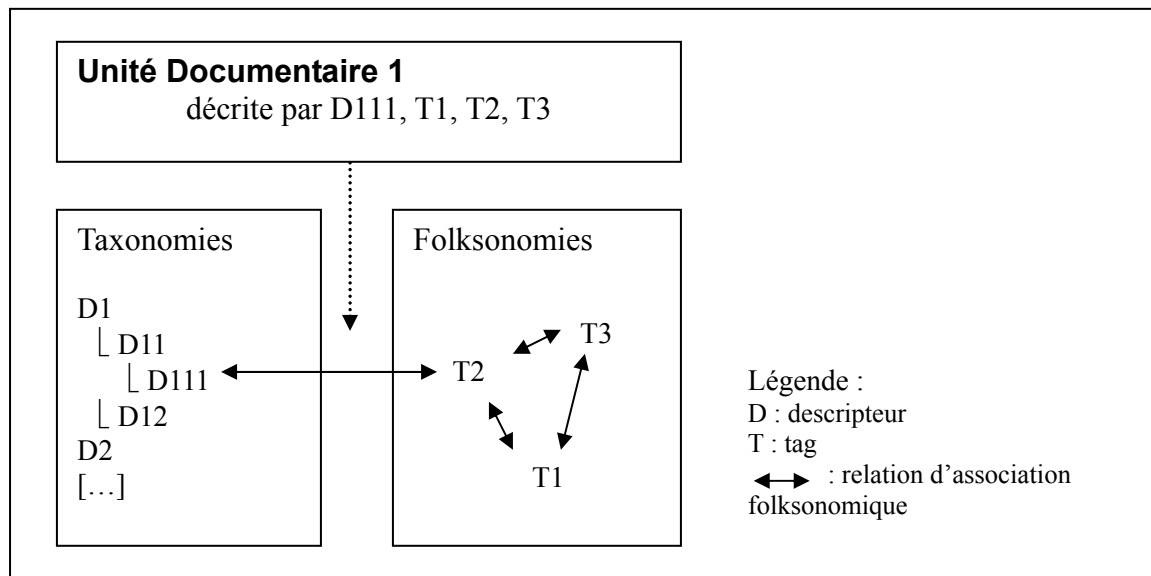


Figure 2 - La notion d'association folksonomique

Afin de permettre l'expression et l'intégration de nouveaux concepts dans le modèle taxonomique la possibilité est laissée aux responsables du travail de veille de modifier le statut d'un *tag* pour l'intégrer au lexique des taxonomies et ainsi lui conférer le statut de descripteur, sous réserve que celui-ci réponde aux critères de normalisation établis. Cette démarche doit également être périodique, le changement de statut d'un terme de *tag* à celui de descripteur doit être envisagé de façon récurrente sur la base d'éléments prédéfinis tels que la fréquence d'apparition du terme. Les étapes de médiation correspondent aux tâches d'évaluation et de qualifications des *tags* importé dans le système d'aide à la veille. Les relations entre les ressources documentaires, les tags et leurs auteurs sont gérées par le système. La figure 3 synthétise la séquence complète d'importation et de traitement des informations.

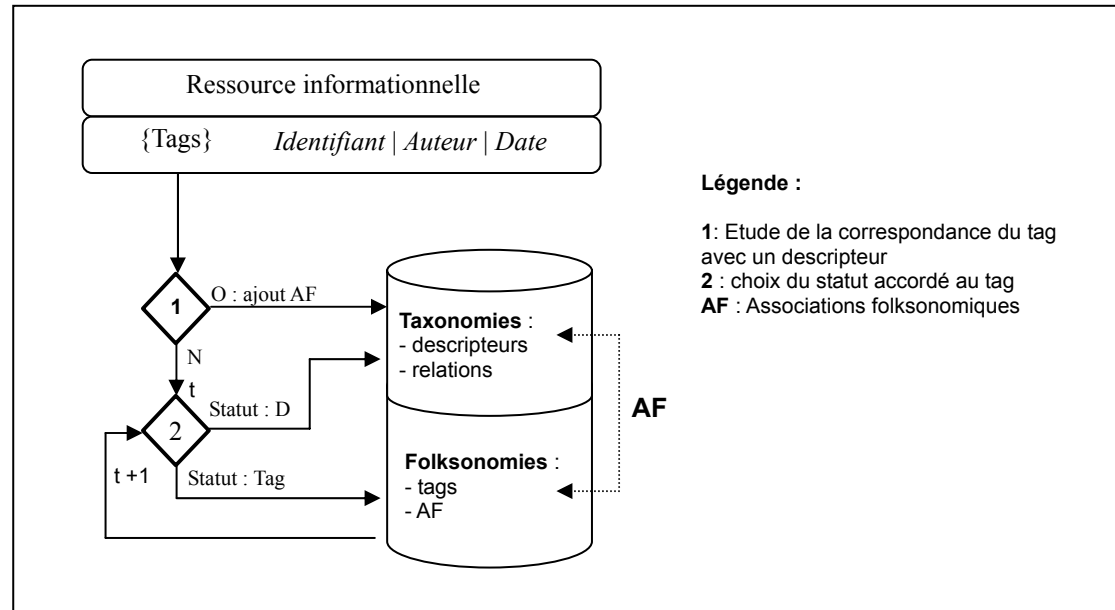


Figure 3 - Organisation et mémorisation des informations

3.2.3 Exploitation des informations

Les fondements techniques du système d'aide à la veille présentés ci-dessus permettent d'envisager diverses modalités de traitement pour l'exploitation des données recueillies afin de répondre aux objectifs présentés dans le paragraphe 3.1 : étude diachronique, prise en compte du comportement et du profil des utilisateurs, identification et représentation de concepts et enfin levée des ambiguïtés langagières.

3.2.3.1 Etude diachronique

L'étude diachronique de l'évolution du corpus des *tags* et des descripteurs est rendue possible par la prise en compte de la date d'indexation des ressources par les utilisateurs. Cette étude permet notamment la surveillance de :

- l'apparition de nouveaux termes qui peut traduire l'émergence de nouveaux concepts ou de nouvelles pistes de recherche ;
- la fréquence d'utilisation des *tags* ou des descripteurs traduisant les phénomènes d'engouement ou de désaffection vis à vis des concepts représentés ;

- l'apparition de nouvelles associations folksonomiques et leurs évolutions qui permettent de situer tags et descripteur dans un univers relationnel enrichi de relations autres que sémantiques ;
- l'ajout de *tags* pour décrire une ressource informationnelle donnée.

Golder et Huberman [5] identifient deux tendances dans les pratiques d'ajout de tags par les utilisateurs. L'expérience menée montre que dans 67% des cas le nombre de tags connaît une croissance forte et constante pendant une période de 10 jours. Puis dans un deuxième temps le nombre d'ajouts diminue progressivement jusqu'à devenir nul sur une période de 30 jours. Cette tendance traduit l'attrait des utilisateurs pour la ressource indexée ainsi que sa nouveauté. Du point de vue de la veille l'identification de ce type de phénomène facilite l'identification des évolutions du milieu observé. Néanmoins dans 17% des cas les ressources connaissent un pic de popularité en moyenne six mois après leur première indexation. L'intérêt du veilleur est dans ce cas précis d'identifier ces ressources et les concepts qui leurs sont liés le plus tôt possible.

3.2.3.2 Comportement et profil des utilisateurs

Le comportement et le profil des utilisateurs peuvent être pris en compte du fait que chaque *tag* est systématiquement lié à son auteur. Cette approche permet l'identification de cibles ou d'experts, non seulement sur la base des unités documentaires indexées mais également en fonction de leur utilisation du système de partage de ressources. Il devient possible d'identifier les personnes à l'origine de l'énonciation de nouveaux concepts, de mener des études quantitatives sur l'ensemble des *tags* partagés ainsi qu'en fonction du contenu informationnel de ces tags (informations de sujet, informations d'usage, information d'actif et informations relationnelles). Ce dernier élément nous semble particulièrement pertinent dans la mesure où il traduit les pratiques individuelles des utilisateurs permettant ainsi une surveillance fondée sur des catégories d'utilisateurs. Guy et Tonkin [6] ont montré que la distributions des tags connaît un effet de « longue traîne » traduisant le fait qu'une majorité d'utilisateurs mobilisent une faible quantité de tags différents alors qu'une partie réduite d'entre eux mobilise un grand nombre de tags pour indexer [2].

3.2.3.3 Identification et représentation de concepts

Le modèle de représentation que constitue l'ensemble des descripteurs et des *tags*, ainsi que leurs relations, exprime simultanément des relations sémantiques et des relations contextualisées entre les différents termes présents. Les ressources informationnelles sélectionnées opèrent un rôle de médiation entre les représentations subjectives exprimées par l'utilisateur et l'ensemble du vocabulaire mobilisé pour indexer le corpus sélectionné. Cette contextualisation permet d'envisager différentes modalités d'accès à l'information laissant au veilleur la possibilité d'exploiter ou non cet environnement relationnel élargi par rapport à un terme, une ressource ou un individu.

3.2.3.4 Ambiguïtés langagières

Les problèmes posés par les ambiguïtés langagières sont traités par l'utilisation simultanée de deux modes de représentation de façon conjointe : un vocabulaire libre constitué par les folksonomies et un vocabulaire contrôlé constitué par les taxonomies. Il est ainsi possible d'intégrer les différentes variantes d'un terme : les descripteurs correspondent à des formes normalisées, les tags associés traduisent les différentes variantes identifiées sur le corpus traité. Ces variantes peuvent résulter de phénomènes de polysémie ou d'homographie mais également de variations d'usages de termes. L'étude diachronique permet

éventuellement à terme en fonction des usages constatés de substituer un terme à un autre, opération se traduisant par la modification du statut d'un terme. Les problèmes posés par les phénomènes de synonymies au de quasi-synonymies ne sont envisagés qu'au moyens de liens entre descripteurs et *tags*. L'évolution du modèle des taxonomies vers une organisation thésaurale est envisagée dans la mesure où l'utilisation d'un thésaurus, constitué de descripteurs et de non-descripteurs, permet un traitement plus élaboré des problèmes liés à la clarification du sens des termes.

4. Conclusion et perspectives

Les informations traitées par les systèmes à base de folksonomies présentent de nombreuses caractéristiques qui répondent aux exigences d'un travail de veille : diversité des contenus accessibles, identification de sources, rapidité de diffusion des informations et des nouveautés, identification d'experts, constitution de réseaux. Cependant le recours aux folksonomies ne peut constituer qu'un élément d'une démarche plus générale. Un processus de veille sur le *web* ne peut s'appuyer uniquement sur l'utilisation de folksonomies sans se heurter à certaines limites inhérentes au principe d'indexation sociale.

La première limite est de nature technique et concerne le taux de couverture des ressources disponibles sur le *web*. Si les systèmes folksonomiques facilitent l'indexation et l'identification de ressources insuffisamment ou incomplètement indexées par les outils de recherche traditionnels – moteurs ou annuaires – ils ne couvrent néanmoins qu'une infime partie des ressources disponibles. Ils ne peuvent en aucun cas être considérés comme unique source d'information. La deuxième limite concerne la nature du travail d'indexation collaborative. Celui-ci résulte d'une démarche volontaire, les ressources et les *tags* mis à la disposition des internautes correspondent à ce que l'auteur accepte de partager avec son réseau de relations. Cette notion de sélection et d'arbitrage peut s'opposer à la notion d'exhaustivité qui demeure un élément essentiel dans une démarche de recherche d'information.

L'approche que nous présentons dans cette communication regroupe un ensemble d'éléments théoriques et techniques visant à faciliter l'exploitation des folksonomies pour la veille. Le modèle d'organisation et de traitement des informations collectées est actuellement en phase de test au sein de notre laboratoire dans le cadre de travaux portant sur le travail collaboratif, les réseaux sociaux et le *web 2.0*. Les retours d'expériences qui seront collectés à l'issue de la phase de test permettront de valider ou d'invalider certaines de nos propositions.

5. Bibliographie

[1] **BLANCHET A.**, *La psychologie des groupes*, 2ème édition, Armand Colin, Collection universitaire de poche.

[2] **CREPEL M.**, *Les folksonomies comme support émergent de navigation sociale et de structuration de l'information sur le web*, Réseaux 2008/6, n°152, P. 169-204

[3] **ERTZCHEILD O., GALLEZOT G.**, *Etude exploratoire des pratiques d'indexation sociale comme une renégociation des espaces documentaires. Vers un nouveau big bang documentaire ?*, document numérique et société, ADBS éditions, collection sciences et techniques de l'information 2006, 344p.

- [4] **FRANCIS E., QUESNEL O.**, *Indexation Collaborative et folksonomies*, Documentaliste-Sciences de l'information 2007/1, volume 44, p. 58-63.
- [5] **GOLDER S.A., HUBERMEN B.**, *The structure of collaborative tagging systems*, Journal of Information Science, vol 32.
- [6] **GUY M., TONKIN E.**, *Folksonomies tidying up tags ?*, D-Lib Magazine, Vol 12, n°1, 2006.
- [7] **LAINÉE F.**, *La veille technologique : de l'amateurisme au professionnalisme*, Eyrolles, 1991, 138p.
- [8] **LE DEUFF O.**, *Folksonomies et communautés de partage de signets, vers de nouvelles stratégies de recherche d'informations*, Collaborer, échanger, Inventer : expériences de réseaux., Hammamet, Tunisie, septembre 2008
- [9] **MENON B.**, *Les langages documentaires – un panorama , quelques remarques critiques et un essai de bilan*, Documentaliste-Sciences de l'information 2007/1, volume 44, p. 18-27.
- [10] **MORRISON J.**, *Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web*, Information Processing and Management 44 (2008) 1562–1579.
- [11] **O'REILLY T.**, *What is web 2.0 ? - Design Patterns and Business Models for the Next Generation of Software*.
- [12] **PEDAUQUE R.T.**, *Document et modernités*, 2006, <http://rtp-doc.enssib.fr/IMG/pdf/pedauque3-V4.pdf>
- [13] **POIRAUD-LAMBERT F.**, *De l'usage des Tags (1) - Etude de l'usage de 131 veilleurs*, <http://www.tecoman.info/article-25766414.html>
- [14] **ROULLEAUX DUGAGE M.**, *Organisation 2.0, le Knowledge Management nouvelle génération*, Eyrolles, éditions d'organisation, 2007, Paris, 258p.
- [15] **SAMIER H., SANDOVAL V.**, *La veille stratégique sur l'Internet*, Hermes Science, Lavoisier, 2002, 191p.
- [16] **SINCLAIR, CARDEW-HALL**, *The folksonomy tag cloud: when is it useful?*, Journal of Information Science.2008; n°34, p. 15-29
- [17] **VANDER WAL T.**, *Folksonomy Coinage and Definition*, <http://vanderwal.net/folksonomy.html>