



HAL
open science

Approche temporelle pour la génération personnalisée de profils folksonomiques

Tahar-Rafik Boudiba, Rachid Ahmad-Ouamer

► **To cite this version:**

Tahar-Rafik Boudiba, Rachid Ahmad-Ouamer. Approche temporelle pour la génération personnalisée de profils folksonomiques. *INFormatique des Organisations et Systemes d'Information et de Decision (INFORSID 2017)*, May 2017, Toulouse, France. pp.263-273. <hal-02641182>

HAL Id: hal-02641182

<https://hal.science/hal-02641182v1>

Submitted on 28 May 2020

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

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



HAL Authorization




Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is a publisher's version published in: <http://oatao.univ-toulouse.fr/22184>

To cite this version:

Boudiba, Tahar-Rafik  and Ahmad-Ouamer, Rachid *Approche temporelle pour la génération personnalisée de profils folksonomiques*. (2017) In: *INformatique des Organisations et Systemes d'Information et de Decision (INFORSID 2017)*, 30 May 2017 - 2 June 2017 (Toulouse, France).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Approche temporelle pour la génération personnalisée de profils folksonomiques

Tahar-Rafik Boudiba^{(1) (2)}, Rachid Ahmed-Ouamer⁽¹⁾

1. LARI laboratory Department of computer science
Mouloud Mammeri University 15000, Tizi-Ouzou, Algeria.
Rachid.ahmedouamer@yahoo.fr
2. IRIT IRIS, 118 Route de Narbonne
31062, Toulouse CEDEX 9 France.
Tahar-Rafik.Boudiba@irit.fr

RÉSUMÉ. L'annotation collaborative offre aux utilisateurs la possibilité de décrire des ressources avec des mots-clés (tags). Un tag décrit l'intérêt que l'utilisateur porte à la ressource. L'ensemble des triplets {utilisateur, ressource, tags} constitue une folksonomie. Dans cet article, nous proposons de construire des profils utilisateurs à partir des tags. Ces profils (folksonomiques) sont des clusters contenant des ressources correspondant aux divers centres d'intérêts de l'utilisateur. Comme les intérêts utilisateurs évoluent avec le temps, nous estimons qu'en pondérant les tags en fonction non seulement de leur qualité supposée, mais aussi de leur "fraîcheur", on améliore le classement des ressources. Pour évaluer la "pertinence" de ces profils, nous les avons utilisés dans un contexte de recherche d'information personnalisée. En comparant notre approche avec une approche de base (sans facteur temporel), nous avons mené une série d'expérimentations sur des données extraites à partir de MovieLens. Les résultats confirment l'efficacité de notre approche.

ABSTRACT. Collaborative annotation offers users the possibility to describe resources with keywords (tags). A tag describes the user's interest to the resource. The set of triplets {user, resource, tags} constitutes a folksonomy. In this paper, we propose to construct user profiles from tags. These (folksonomy based) user profiles are clusters containing classified resources corresponding to the various interests of the user. As user interests evolve over time, we estimate that by weighting the tags according to not only their supposed quality but also their "freshness", the classification of resources is improved. To evaluate the "relevance" of these profiles, we used them in a personalized information retrieval context. By comparing our approach with a basic one (without time factor), we conducted a series of experimentations extracted from MovieLens Dataset. The results confirm the effectiveness of our approach.

MOTS-CLÉS : Folksonomies, recherche d'information, temporalité, profil utilisateur, clustering.

KEYWORDS: Folksonomies, information retrieval, temporality, user profile, clustering.

1. Introduction

Avec le développement fulgurant que connaît Internet, à travers les systèmes d'annotations collaboratives, le Web 2.0 est devenu de plus en plus populaire, ce qui a fait passer les utilisateurs de consommateurs passifs en producteurs actifs de contenus Web (Zanardi, Capra, 2008).

Les usagers de ces systèmes génèrent une énorme quantité de données personnelles. Ils utilisent des mots-clés ou tags leur permettant de catégoriser, d'indexer, ou de classer diverses ressources qui concordent avec les intérêts de ces usagers. L'ensemble des triplets { utilisateur, ressource, tags } constitue une folksonomie.

Les tags décrivent les intérêts que portent les utilisateurs aux ressources. Ils fournissent des informations utiles à la construction de profils utilisateurs. En général, ces profils sont représentés sous forme de vecteurs de termes (tags) auxquels on associe un poids w à chaque terme t . Certains travaux (Xu *et al.*, 2008 ; Vallet *et al.*, 2010), utilisent *tf-idf* ainsi que des variantes de ce modèle comme schéma de pondération.

En partant du constat que l'ensemble des tags utilisateurs décrivent les préférences de celui-ci à l'égard d'une ressource, la dimension temporelle de l'annotation fournit quant à elle, une indication sur l'évolution de l'intérêt de l'utilisateur à travers le temps (Zheng, Li, 2011). Certains travaux intègrent cette dimension lors de la construction de profils utilisateurs (Kacem *et al.*, 2014 ; Cheng *et al.*, 2008 ; Abel *et al.*, 2011). De manière générale, la temporalité associée à l'action d'annotation est quantifiée à l'aide d'une stratégie de pondération visant à attribuer un poids à chaque ressource en se basant sur les informations que fournissent les tags (popularité ou timestamps) ou une combinaison des deux (Zheng, Li, 2011).

Cependant les travaux exploitant ce type d'information ne nous renseignent pas à quel moment un contenu devient obsolète. De plus, si certains utilisateurs actifs annotent des ressources, et cela, de manière plus fréquente que d'autres utilisateurs moins actifs, alors les poids du tag de l'utilisateur le plus actif seront plus élevés que ceux du moins actif. En ce qui concerne les ressources, les utilisateurs actifs influencent la fréquence avec laquelle elles ont été annotées, les rendant plus populaires du fait de la tendance disproportionnée des utilisateurs actifs à annoter ces ressources. Cela affecte de ce fait le classement global de celles-ci, qui subissent une distorsion systématique causée par des utilisateurs plus actifs que d'autres.

Dans cet article nous proposons de construire des profils qui prennent en considération l'aspect temporel des annotations, nous regroupons en catégorie les ressources en intégrant une formule de pondération normalisée modérée par une fonction temporelle. Ces profils (folksonomiques) sont des clusters contenant des ressources classées correspondant aux divers intérêts de l'utilisateur ainsi que de leurs évolutions dans le temps. Nous évaluons par la suite notre approche en la comparant avec une approche

de base sur une collection de données extraites à partir de MovieLens¹ qui confirment l'efficacité de notre approche.

L'article est structuré de la manière suivante: nous présentons dans la section 2 un aperçu de certains travaux connexes, après quoi nous détaillons notre approche dans la section 3. La section 4 est consacrée à l'expérimentation. Enfin, la section 5 conclut l'article et énonce certaines perspectives.

2. État de l'art

Dans cette section, nous présentons certaines approches de personnalisation exploitant les folksonomies pour la construction de profils, nous nous intéresserons aussi aux travaux intégrant la temporalité pour la construction de profils utilisateurs ou visant à prendre en compte l'évolution de ces derniers. Nous faisons une synthèse de quelques-unes de ces approches en les présentant en deux sous-sections, l'une traitant des approches de personnalisation basée sur les folksonomies, tandis que l'autre portera sur les approches temporelles se basant sur les tags.

2.1. Approche de personnalisation basé sur les folksonomies

Les approches décrites ci-dessous utilisent le modèle vectoriel pour la génération de profils. Dans (Xu *et al.*, 2008), les auteurs utilisent la mesure du cosinus afin de calculer la similarité entre profils utilisateurs et profils de documents. L'approche est basée sur le modèle vectoriel. Comme schéma de pondération (Xu *et al.*, 2008) utilisent *tf-idf* qui est basé sur le modèle *BM25*. Noll et Meinel (Noll, Meinel, 2007) supposent que les tags plus fréquemment utilisés sont plus intéressants et plus pertinents pour un utilisateur que ceux utilisés rarement, l'idée est de tirer parti des annotations fournies par la communauté pour identifier les perceptions communément admises des documents. (Vallet *et al.*, 2010) présentent une approche de personnalisation avec comme schéma de pondération *tf-idf*. Celle-ci est similaire à celle de (Xu *et al.*, 2008), cependant Vallet et al. éliminent le facteur de normalisation de la longueur du document. Dans le modèle vectoriel le but d'une telle manœuvre est de pénaliser le score des documents qui contiennent un grand nombre d'informations, et qui de ce fait sont susceptibles de correspondre à la requête par hasard. En ce qui concerne les systèmes d'annotations sociales, le fait qu'un grand nombre de tags soit connexes est lié à la popularité des ressources. Selon (Cai, Li, 2010) les schémas *tf*, *tf-idf* ou *BM25* utilisés pour pondérer des tags, se révèlent être insuffisant pour indiquer à quel point un utilisateur présente de l'intérêt pour les tags décrivant son profil. De plus, des utilisateurs actifs influencent les fréquences avec lesquelles les ressources ont été annotées, les rendant plus populaire du fait de la tendance disproportionnée des utilisateurs actifs à annoter ces ressources. Cela affecte de ce fait le ranking (classement) des ressources à la recherche.

1. <http://movielens.org/>

2.2. Approches temporelles pour la construction de profils utilisateurs

La motivation principale de ce type d'approche réside dans le fait que les intérêts des utilisateurs évoluent avec le temps. En effet ils déclinent au fur et à mesure que le temps passe (Zheng, Li, 2011). Les travaux décrivant ces approches (Abel *et al.*, 2011 ; Zheng, Li, 2011 ; Kacem *et al.*, 2014) exploitent des fonctions temporelles qui simulent l'atténuation des intérêts utilisateurs au cours du temps.

Dans (Zheng, Li, 2011), les auteurs développent une fonction exponentielle adaptative qui simule l'atténuation graduelle des intérêts utilisateurs. Zheng et al. combinent deux mesures de poids; l'une concerne la ressource par rapport au tag qui le décrit, l'autre fait état de la valeur temporelle attribuée à la ressource.

Dans (Kacem *et al.*, 2014), le profil utilisateur est représenté à l'aide d'un vecteur de termes. L'importance de chaque terme est ajusté en fonction de sa date d'apparition. Cette mesure de "fraîcheur" de terme est calculée grâce à une fonction à noyaux gaussiens, qui permet le calcul d'une distance temporelle entre la date courante et la date où apparaît le terme (l'interaction sociale). Abel et al. (Abel *et al.*, 2011) Analysent les activités individuelles sur Twitter². Ils comparent différentes stratégies pour la création de profils utilisateurs, étudient comment ces profils changent au fil du temps, et comment la dynamique temporelle influence l'exactitude du processus de personnalisation. Les auteurs considèrent deux types de profil utilisateur: les utilisateurs actifs qui interagissent pendant une longue période de temps, et ceux qui sont moins actifs et qui interagissent de manière plus irrégulière.

3. Approche temporelle pour la génération de profils folksonomiques

3.1. Notations

Une folksonomie F peut être définie comme étant un 4-uplet $F = (U, T, R, A)$, où U est l'ensemble des utilisateurs annotant les ressources de l'ensemble R avec $U = \{u_1, u_2, \dots, u_m\}$ où chaque u_i est un utilisateur; T est l'ensemble des tags que comprend le vocabulaire exprimé par la folksonomie $T = \{t_1, t_2, \dots, t_l\}$; R est l'ensemble des ressources annotées par les utilisateurs $R = \{r_1, r_2, \dots, r_n\}$; $A = \{u_m, t_l, r_n\} \in U \times T \times R$ est l'ensemble des annotations de chaque tag t_l à une ressource r_n par un utilisateur u_m . Le profil utilisateur u_m peut être défini comme un vecteur $\vec{u}_m = (u_{m,1}, \dots, u_{m,l})$, où $u_{m,l} = |\{(u_m, t_l, r) \in A / r \in R\}|$ est le nombre de fois que l'utilisateur a annoté une ressource avec le tag t_l .

3.2. Description générale de l'approche

On considère un système non personnalisé S retournant un ensemble de ressources R , et qui fournit une liste de ressources classées $S(q) \subseteq R$. On suppose que $\forall q, S(q)$

2. <https://twitter.com/>

satisfait une requête q de l'utilisateur. Le classement suit l'ordre implémenté par le système $[S = r_1 \geq r_2 \geq \dots r_k]$ où $r_k \in R$.

Notre approche consiste à adopter une formule de pondération normalisée modérée par une fonction temporelle afin d'attribuer un poids à chaque tag. Ce poids sera représentatif de la qualité supposée du tag à décrire la ressource, mais aussi de la "fraîcheur", de celui-ci. Dans le but de proposer un contenu personnalisé à l'utilisateur, nous catégorisons les ressources annotées par celui-ci via un clustering ; les ressources seront regroupées en catégories d'intérêts $C = (C_1, C_2 \dots C_h)$. Enfin, nous définissons (section 4.2) une signature propre à chaque cluster comme étant la proportion de tags qui apparaît le plus dans le cluster.

Nous considérons les signatures des clusters comme des requêtes représentées par des vecteurs de tags $\vec{q} = (t_1, t_2, \dots t_n)$. Nous les utiliserons pour récupérer les ressources en comparant les tags des signatures à l'ensemble des tags décrivant les ressources contenues dans le cluster. Afin de classer ces ressources par similarités, nous calculons pour chaque ressource appartenant à C représentée comme un vecteur de tags $\vec{e} = (t_1, t_2, \dots t_m)$ la similarité entre les deux vecteurs \vec{e} et \vec{q} , cette approche personnalisée fournit une liste de ressources classées $S'(q, e) \subseteq R$, en accord avec les préférences de l'utilisateur u . Formellement $[S' = r_1 \geq r_2 \geq \dots r_k]$, en définissant la relation d'ordre \geq centrée sur l'utilisateur u , par la condition:

$$r_i \geq r_j \Leftrightarrow sim(u, r_i, \vec{q}) \geq sim(u, r_j, \vec{q})$$

Où $sim(u, r, \vec{q})$ est la fonction de similarité entre u et r en prenant en considération le classement de r dans $S(q)$.

3.3. Schéma de Pondération

Nous proposons de représenter initialement le profil utilisateur sous forme vectorielle. Nous attribuons à chaque tag une valeur $v_{i,k}$ qui correspond à une fréquence de tag normalisée. \vec{U}_i est un vecteur de tags: valeurs, i.e

$$\vec{U}_i = (t_{i,1} : v_{i,1}, t_{i,2} : v_{i,2} \dots t_{i,n} : v_{i,n})$$

Où $t_{i,k}$ est un tag annoté par u_i , n est le nombre total de tags annotés par u_i . $v_{i,k}$ est le degré de préférence de l'utilisateur u_i au tag $t_{i,k}$, qui peut être intuitivement obtenu comme suit:

$$v_{i,k} = \frac{N_{i,k}}{N_i} \tag{1}$$

Où $N_{i,x}$ est le nombre de fois qu'un utilisateur i annoté avec le tag x , et N_i est le nombre de ressources annotées par l'utilisateur i .

En tenant compte du fait que les préférences des utilisateurs évoluent avec le temps, nous proposons de revisiter la notion de fréquence de tag normalisée en biaisant cette dernière par une fonction temporelle. Nous utilisons une fonction à noyaux

Gaussien (Kacem *et al.*, 2014 ; Badache, Boughanem, 2015) afin d'estimer la distance entre la date courante et la date à laquelle la ressource a été annotée:

$$K(S^c, S_j) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{(S^c - S_j)^2}{2\sigma^2}\right] \quad (2)$$

Où σ est le coefficient d'interpolation, S^c est la date courante, S_j est la date qui correspond à l'action d'annotation.

Par conséquent, nous revisitons la notion de profil utilisateur, en attribuant les pondérations suivantes au vecteur de tags:

$$\vec{U} = (t_1^{S_j} : W_1^{S_j}, t_2^{S_j} : W_2^{S_j}, \dots, t_m^{S_j} : W_m^{S_j})$$

Où le poids $W(t)^{S_j}$ d'un tag t dans un profil utilisateur est la somme de ses fréquences normalisées et biaisées par la fonction temporelle, et est définie comme suit:

$$W(t_k)^{S_c} = \sum v_{i,k}(t_k)^{S_j} K(S^c, S_j) \quad (3)$$

3.4. Clustering

Afin de regrouper les ressources en catégories d'intérêts, nous construisons d'abord un graphe biparti représentant l'assignation des tags utilisateurs à l'ensemble des ressources utilisateurs. Nous attribuons à chaque tag du graphe biparti la pondération normalisée et biaisée par la fonction temporelle $W(t_k)^{S_c}$. Par la suite, en faisant une projection 1-mode sur l'ensemble des ressources, nous obtenons un graphe de ressources où chaque nœud correspond à une ressource. Dès lors, nous pondérons les arêtes entre les ressources par la somme des pondérations $W(t_k)^{S_c}$ des tags en commun reliant les ressources entre elles.

3.5. Description de la collection de test

MoviesLens est un système de recommandation de films où les utilisateurs peuvent attribuer des tags et des notes aux films en fonction de leurs préférences. La collection de test que nous avons extrait, contient des informations sur l'identifiant de l'utilisateur, le tag, ainsi que le film annoté. Elle contient aussi la date à laquelle l'utilisateur a annoté le film (exprimé en format Timestamp). De plus elle fournit l'appréciation (rating) de chaque utilisateur à propos d'un film. Cette "note" attribuée par l'utilisateur évalue le film de 1 à 5 étoiles selon l'appréciation de celui-ci.

La collection contient en outre 2000063 notes(rating) et 465564 tags et 27278 films et a été généré le 31 mars 2015. La collection de données MovieLens a été publiée par le groupe de recherche GroupLens à l'Université de Minisota. Afin d'évaluer notre approche, nous donnons un bref aperçu sur l'ensemble de données. Le **Tableau 1** y présente les principaux détails.

Tableau 1 – Statistiques sur la collection

Nombre d'utilisateurs	138493
Nombre de films	27278
TAS(Tag assignement)	465564
Rating	2000063
Periode	January 09,1995 - March 31, 2015

3.6. Expérimentation

à partir de cette collection de données, nous extrayons 100 utilisateurs ayant au moins 100 tags dans leurs activités d'annotations. Puis nous évaluons la performance de l'approche personnalisée en fonction de la pertinence graduée des ressources pour chaque utilisateur et la comparons à celle d'une approche qui ne tient pas compte du paramètre temporel.

En ce qui concerne le regroupement des ressources en communautés, nous utilisons un algorithme de clustering dit glouton "*greedy algorithm*". Cet algorithme de clustering est basé sur une optimisation de la modularité, en effet la mesure quantitative qui garantit la qualité des clusters est la modularité (Newman, 2004 ; 2006). La modularité d'un clustering est le nombre d'arêtes intraclusters, comparé au nombre moyen d'arêtes intraclusters que l'on aurait après brassage aléatoire du graphe.

Par la suite, nous comparons notre méthode de génération de profils utilisateurs, à deux autres approches. La première (Yeung *et al.*, 2008) **Baseline 1**, construit des profils utilisateurs en regroupant les ressources sans attribuer préalablement de poids aux tags. La deuxième, **Baseline 2**, regroupe les ressources en sommant les degrés de préférence de l'utilisateur pour chaque tag en commun entre deux ressources i.e: la somme des $v_{i,k}$ en commun entre deux ressources.

En outre, **Baseline 2** adopte un schéma de pondération normalisée des termes, le but d'une telle manœuvre est de montrer l'impact séparé de la pondération des tags et l'atténuation au cours du temps. Yeung et.al utilisent aussi un algorithme de clustering pour regrouper les ressources, mais ne classent pas les ressources selon les intérêts de l'utilisateur. De plus (Yeung *et al.*, 2008) ne prennent pas en compte l'aspect temporel des annotations.

Ci-après un aperçu du nombre de clusters par utilisateur que génère notre approche (figure 1).

Une fois les clusters générés, nous définissons une signature propre à chaque cluster comme étant la proportion de tags qui apparaît le plus dans celui-ci. La signature fournit un aperçu du regroupement en cluster des ressources, mais aussi un moyen d'identifier chaque cluster grâce à un ensemble de tags qui lui est propre. Nous fixons cette proportion à $\rho=40\%$. les similarités seront calculées entre chaque ressource et classées. Ce classement des ressources correspondra aux préférences de l'utilisateur à travers le temps. à l'issue des expérimentations, nous fixons aussi la valeur du coeffi-

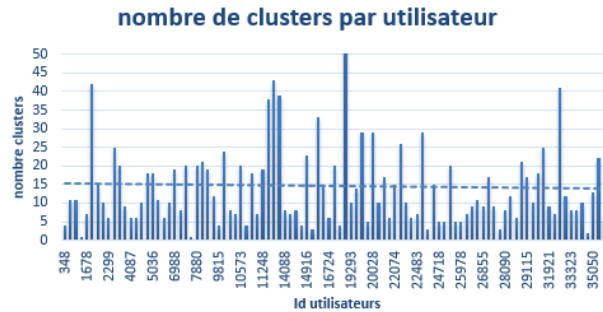


FIGURE 1 – Un aperçu du nombre de clusters générés par utilisateur

cient d'interpolation σ à 4 (l'augmentation de la valeur du paramètre σ réduit sensiblement l'amplitude de la fonction gaussienne 2 section 3.3).

3.7. Évaluation

Les *ratings* (notes) fournis par l'utilisateur, nous apportent un moyen utile de collecter les ressources que l'utilisateur a appréciées. Ils seront utilisés comme jugement de pertinence. Dans le but d'évaluer la qualité du classement des ressources, nous définissons la fonction "*rating*" $f_{u_i}(m_i)$ qui retourne la valeur de la note du film m_i attribuée par l'utilisateur u_i .

La valeur maximale de la fonction est : $Max(F_{u_i}(m_i)) = 5$, cela indique une très bonne appréciation de l'utilisateur à l'égard du film, la valeur minimale de la fonction est quant à elle égale à 1.

Pour une requête donnée, notre système doit fournir une liste classée de résultats, les ressources les mieux classées doivent correspondre aux préférences de l'utilisateur. Dans notre cas pour chaque signature de cluster obtenue grâce à notre approche, une liste de films classés est générée.

Le principe du DCG (Discount Cumulative gain) stipule que les films très pertinents apparaissant au bas d'un résultat de recherche doivent être pénalisés, car la valeur de pertinence graduée est réduite de manière logarithmique et proportionnelle à la position du résultat. Plus la position d'un film pertinent est basse, moins il sera utile pour l'utilisateur, car il est moins susceptible d'être examiné.

Le score de pertinence de chaque film est utilisé comme mesure de la valeur acquise pour sa position classée dans le résultat et le gain est additionné à la position classée de 1 à n. La liste des films classés devient une liste de valeurs acquises en remplaçant les ID du film par leur score de pertinence.

Les scores de pertinence sont définis sur une échelle de 5 étoiles: de 0 à 5, 0 pour des films non notés, 1 à 3 pour des films non pertinents et 4 à 5 pour des films pertinents. Nous présentons dans ce qui suit, le résultat de notre évaluation grâce au calcul du nDCG (Normalized Discounted Cumulative Gain (Jarvelin, Kekalainen, 2002)) (**Tableau 2**). Le DCG idéal étant les films qui sont classés d'après les notes de l'utilisateur.

3.8. Résultats et analyse

On observe (**Figure 2**) une nette amélioration de la position de classement en faveur de notre approche à $r@ = 5$ à $r@ = 10$ **Tableau 2**. La valeur nDCG de cette dernière augmente pour atteindre 0.69 à $r@ = 90$ et $r@ = 100$, tandis que la **baseline 1** atteint la valeur maximale de 0,62 lorsque $r@ = 90$ et $r@ = 100$. La valeur nDCG de la **baseline 2** indique quant à elle une amélioration qui surpasse celle de la **Baseline 1** et atteint une valeur maximale de 0.64 lorsque $r@ = 90$ et $r@ = 100$.

De notre point de vue, cela met en évidence d'une part, l'impact de l'attribution de poids qualitatifs représentant les préférences de l'utilisateur, mais aussi l'importance de prendre en compte l'évolution des intérêts de celui-ci. Par conséquent, la combinaison d'un schéma de pondération normalisée de termes avec une mesure permettant l'atténuation progressive de l'importance de ces termes au cours du temps, améliore le classement personnalisé des ressources retournées à l'utilisateur lors du processus de recherche.

Tableau 2 – Comparaison NDCG

Rank	NDCG_baseline1	NDCG_baseline2	NDCG_Our
r@5	0.426	0.45	0.56
r@10	0.446	0.5	0.57
r@20	0.465	0.48	0.586
r@30	0.49	0.53	0.587
r@40	0.51	0.55	0.61
r@50	0.53	0.591	0.624
r@60	0.55	0.59	0.63
r@70	0.57	0.611	0.6423
r@80	0.59	0.612	0.67
r@90	0.61	0.63	0.685
r@100	0.62	0.64	0.698

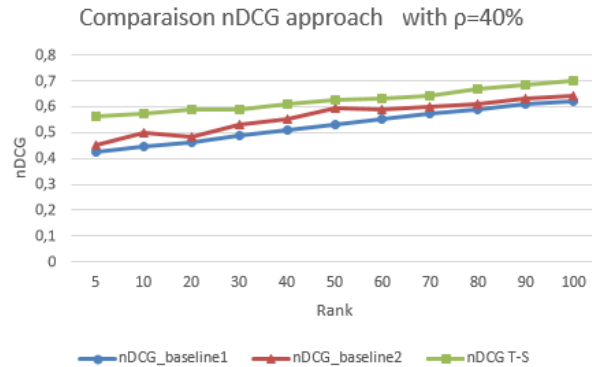


FIGURE 2 – Comparaison nDCG

4. Conclusion

Nous avons proposé une approche visant à générer des profils folksonomiques. Ces profils sont des clusters correspondant aux intérêts de l'utilisateur en tenant compte aussi de l'évolution de ses préférences. Nous avons mis en place une formule de pondération qui prend en compte les aspects qualitatifs et temporels d'une folksonomie. Le modèle de pondération utilisé a été défini grâce à une méthode de clustering. Cette méthode nous a permis d'extraire des ressources classées selon l'ordre de préférence de l'utilisateur et son évolution dans le temps.

En perspective, nous pensons concevoir une étude plus approfondie du paramètre ρ , dans cette optique, des expérimentations sont en cours de validation afin de déterminer avec exactitude la proportion de tags apparaissant le plus dans les clusters générés, et de ce fait, fournir une meilleure description de ces clusters.

Remerciements

Les expériences présentées dans cet article ont été réalisées en utilisant la plateforme OSIRIM qui est administrée par l'IRIT et soutenue par CNRS, la région Midi-Pyrénées, le gouvernement français, et le FEDER (voir <http://osirim.irit.fr/site/fr>).

Bibliographie

- Abel F., Gao Q., Houben G.-J., Tao K. (2011). Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In *Proceedings of the 3rd international web science conference*, p. 2.
- Badache I., Boughanem M. (2015). Document Priors Based on Time-Sensitive Social Signals. In *European Conference on Information Retrieval*, p. 617–622.

- Cai Y., Li Q. (2010). Personalized search by tag-based user profile and resource profile in collaborative tagging systems. In *Proceedings of the 19th acm international conference on information and knowledge management*, p. 969–978.
- Cheng Y., Qiu G., Bu J., Liu K., Han Y., Wang. (2008). Model bloggers' interests based on forgetting mechanism. In *Proceedings of the 17th international conference on world wide web*, p. 1129–1130.
- Jarvelin K., Kekalainen J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, vol. 20, n° 4, p. 422–446.
- Kacem A., Boughanem M., Faiz R. (2014). Time-sensitive user profile for optimizing search personalization. In *International conference on user modeling, adaptation, and personalization*, p. 111–121.
- Newman M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical review E*, vol. 69, n° 6, p. 066133.
- Newman M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, vol. 103, n° 23, p. 8577–8582.
- Noll M. G., Meinel C. (2007). Web search personalization via social bookmarking and tagging. In *The semantic web*, p. 367–380. Springer.
- Vallet D., Cantador I., Jose J. M. (2010). Personalizing web search with folksonomy-based user and document profiles. In *European conference on information retrieval*, p. 420–431.
- Xu S., Bao S., Fei B., Su Z., Yu Y. (2008). Exploring folksonomy for personalized search. In *Proceedings of the 31st annual international acm sigir conference on research and development in information retrieval*, p. 155–162.
- Yeung A., Man C., Gibbins N., Shadbolt N. (2008). A study of user profile generation from folksonomies.
- Zanardi V., Capra L. (2008). Social ranking: uncovering relevant content using tag-based recommender systems. In *Proceedings of the 2008 acm conference on recommender systems*, p. 51–58.
- Zheng N., Li Q. (2011). A recommender system based on tag and time information for social tagging systems. *Expert Systems with Applications*, vol. 38, n° 4, p. 4575–4587.