

---

# Impact des Réseaux Sociaux sur le Processus de Recherche d'Information

**Chahrazed Bouhini**

*Laboratoire Hubert Curien, UMR CNRS 5516  
18 rue du professeur Benoît Lauras  
42000 Saint-Etienne  
Chahrazed.Bouhini@univ-st-etienne.fr*

---

*RÉSUMÉ. L'explosion du Web 2.0 (blogs, wikis, sites de partage, réseaux sociaux, etc.) ouvre des perspectives inédites de partage et de gestion de l'information, en permettant la construction collaborative de contenus et le développement de réseaux sociaux ouverts. Notre travail s'articule autour des problématiques d'accès à l'information dans ce contexte où l'utilisateur est à la fois producteur et consommateur de contenus dans une structure qui représente les relations sociales sous forme de graphes.*

*Nous présentons un état de l'art sur l'impact des informations sociales sur le processus de Recherche d'Information (RI), en positionnant les travaux par rapport à certains aspects d'intégration de l'information sociale au sein du modèle de RI : approches basées sur l'intégration des relations sociales ou sur le contenu informationnel apporté par l'utilisateur.*

*ABSTRACT. The explosion of Web 2.0 (blogs, wikis, sharing Web sites, social networks, etc.) opens new perspectives sharing and dealing with information, allowing the collaborative content construction and the open social networks' development. Our work deals with information access issues in a context where the user is both producer and consumer of content. We present a state of the art about the impact of the social information on the Information Retrieval (IR) process, positioning the works in relation to certain aspects of integration of social information within the IR model: approaches based on social relationships or informational content provided by the user.*

*MOTS-CLÉS : recherche d'information sociale, réseaux sociaux, modèle de RI.*

*KEYWORDS: social information retrieval, social networks, IR model.*

---

## 1. Introduction

La croissance d'Internet a permis de former différents types de réseaux sociaux (RS) à grande échelle, qui sont maintenant reconnus comme un moyen important pour la diffusion de l'information (Saito 2010).

L'explosion des RS a permis l'émergence d'une nouvelle branche de la Recherche d'Information (RI) : la RI sociale. Il s'agit d'adapter les modèles et les algorithmes de la RI classique afin d'exploiter les informations sociales propres à ce nouveau cadre, selon les mesures classiques de Rappel / Précision.

Dans le présent document nous nous focalisons sur la problématique de l'impact des informations sociales sur le processus de RI, avec un objectif d'amélioration des résultats de RI classique. Les informations sociales (relations sociales, annotations, clics, profils, etc.) peuvent être exploitées au sein même du modèle de RI (modèle de document et de requête, fonction de pondération / de correspondance), ou en aval de ce modèle (reclassement de la liste des résultats). Nous allons détailler, dans la section 2, quelques exemples d'informations sociales utilisées pour améliorer les résultats de systèmes de RI. Par la suite nous présentons un état de l'art dans la section 3 sur les travaux effectués dans le domaine de RI sociale.

## 2. RI classique / RI sociale

Cette section présente, par rapport au modèle de RI classique, les principales informations sociales apportées par les utilisateurs et pouvant être intégrées dans la RI sociale.

### 2.1. Modèle de RI classique

La RI traite différents aspects, notamment la représentation, le stockage, l'organisation et l'accès à l'information. Un système de RI se base sur une comparaison entre la représentation interne de la requête et la représentation interne des documents du corpus (Boubekeur 2008), comme montre la **figure 1**.

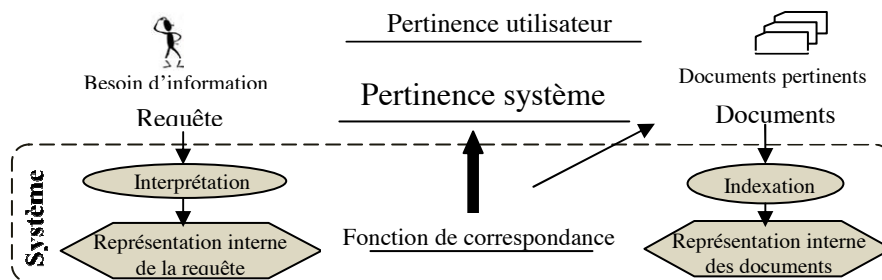


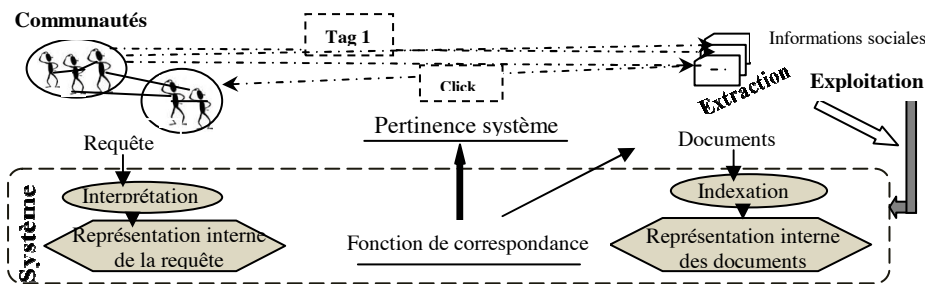
Figure 1. Modèle de RI classique

## 2.2. Informations sociales

Avec l'explosion des technologies du Web 2.0, les utilisateurs du Web produisent divers contenus, créent des annotations, manipulent les documents sur le Web et laissent des traces de leurs passages, etc. Par exemple, les *folksonomies*<sup>1</sup> représentent des informations d'une grande importance. La plupart des RS sont modélisés par des structures de graphe social dont les nœuds sont les utilisateurs du réseau et les arcs représentent les relations entre ces utilisateurs. Dans le cas des *folksonomies*, la structure sous-jacente peut être considérée comme un graphe tripartite : utilisateurs, annotations et nœuds de ressources. On observe dans les RS l'existence des informations suivantes :

- les tags, utilisés pour annoter des bookmarks, des pages Web, des images, etc. Ce type d'informations peut être considéré comme un avis de l'utilisateur, généralement positif, à propos des différentes ressources annotées. Il est aussi envisageable d'exploiter ces données pour en extraire des informations thématiques à propos des ressources annotées (ex : ressources liées à un domaine d'intérêt).
- les traces des utilisateurs (navigation sur le Web, visualisation des pages Web et documents, etc.). L'exploitation de ces traces permet également d'extraire des informations, éventuellement thématiques, sur les préférences des utilisateurs.
- les relations entre utilisateurs au sein du réseau social : amis, co-auteurs, etc.
- les profils d'utilisateurs.

D'une manière générale, ces informations sociales peuvent être exploitées pour incorporer le domaine d'intérêt de l'utilisateur dans la RI sociale (ex. à partir de ses avis sur un thème précis) mais aussi pour améliorer l'évaluation des ressources sur le Web par rapport à un thème donné suite aux retours positifs que donne l'utilisateur. Elles peuvent être prises en compte dans un modèle de RI sociale à différents niveaux, comme montre la **figure 2**.



**Figure 2.** Modèle de RI sociale : l'utilisateur producteur et consommateur d'informations.

<sup>1</sup> *Folksonomies*: un système collaboratif d'étiquetage de ressources (pages Web, vidéos, images, etc.) à l'aide de mots clés tags (annotations).

### **3. Etat de l'art**

Dans cette section nous présentons des travaux de l'état de l'art de la RI exploitant les informations sociales, dans un premier temps ceux exploitant le contenu informationnel puis ceux exploitant les relations sociales elles-mêmes.

#### **3.1. Approche basée sur le contenu informationnel apporté par l'utilisateur (tags et traces)**

Peu de travaux ont été faits sur l'intégration en RI des précieuses informations apportées par les utilisateurs, telles que les annotations sociales. Dans cette approche, les annotations sociales sont exploitées pour l'amélioration de la RI sociale par l'incorporation du domaine d'intérêt extrait du contenu social de l'utilisateur.

(Zhou 2008), qui propose un modèle de génération d'annotations en supposant qu'un tag et un sujet du document sont pareils.

(Bao 2007) et (Xu 2006) exploitent les annotations sociales pour le calcul de similarités entre les requêtes et les tags sociaux et montrent qu'une utilisation des annotations dans le processus de RI sociale permet d'optimiser la RI sur le Web, ainsi ils proposent deux algorithmes pour la RI : « *Social Page Rank* » et « *Social Sim Rank* » pour calculer des scores de popularité des pages Web et de similarité (entre les annotations sociales et les requêtes du Web). Ils montrent par la suite qu'en combinant les deux algorithmes, la précision moyenne peut être améliorée significativement.

#### **3.2. Approche combinant le contenu informationnel et les relations sociales**

Dans cette approche, le score calculé dans un système de RI est amélioré en combinant le score de documents et un score social calculé à partir de l'exploitation des relations sociales et l'incorporation, dans le modèle de RI, de différentes mesures sur ces relations : (Kirsch 2005), (Konstas 2008), (Mutschke 2001), (Kirchhoff 2008) et (Ben-Jabeur 2010) notamment dans les réseaux d'accès aux ressources bibliographiques où on considère différents types de relations entre les auteurs du réseau bibliographique : (Mutschke 2001), (Kirchhoff 2008) et (Ben-Jabeur 2010) exploitent différents liens de coauteurs.

(Konstas 2008) considère l'exploitation des annotations sociales avec les relations sociales entre les utilisateurs pour l'amélioration de systèmes de recommandation. Il évalue le modèle proposé sur un jeu de données collecté à partir du réseau social «*last.fm*» incluant des relations d'amitié et des annotations collaboratives. Il montre par la suite que l'incorporation des relations d'amitié et des annotations sociales peut améliorer la performance d'un système de recommandation.

(Ben-Jabeur 2010) introduit un modèle de RI sociale pour l'accès aux ressources bibliographiques dans lequel la pertinence d'un document est estimée par combinaison de la pertinence thématique et de la pertinence sociale, qui est à son

tour dérivée de l'importance sociale des auteurs associés. Le modèle proposé exploite, en plus des liens de co-auteur, deux autres types de relations telles que la citation et l'annotation sociale. Des poids sont attribués à ces relations qui tiennent compte de la position des acteurs dans le réseau et de leurs mutuelles collaborations. Le modèle est évalué sur une collection d'articles scientifiques dont les annotations sociales sont extraites depuis le réseau académique CiteULike.org. Ces relations au sein des RS sont généralement pondérées pour le calcul des distances sociales.

#### 4. Discussion et conclusion

Les différentes approches de RI classique ignorent l'influence des relations sociales et des interactions de l'utilisateur au sein de son contenu social, sur le processus global de RI. Les travaux effectués dans la RI sociale montrent l'intérêt d'exploiter les informations sociales pour la RI. Nous avons vu dans le présent document deux catégories d'approches suivies pour les différents travaux d'état de l'art en RI sociale. La première catégorie consiste à exploiter les informations sociales relatives au contenu (annotations, traces, etc.). La deuxième catégorie permet de combiner ce contenu social avec les relations entre les utilisateurs des RS.

Ces travaux ouvrent de nombreuses perspectives. En particulier, nous pensons que cette combinaison ne peut être satisfaisante que si elle intervient au sein même du modèle de RI, ce qui nécessite une adaptation des composants de ce modèle dans un cadre social.

Un problème important réside dans le fait qu'il n'existe pas de collection de test standardisée pour la RI sociale. En effet, dans les travaux présentés, plusieurs jeux de données sont utilisés pour évaluer les approches (Del.icio.us, Citeulike.org, last.fm, etc.) mais aucun d'entre eux n'utilise une collection de test standardisée. La construction d'une telle collection est un problème fondamental de la RI sociale. Il existe différentes collections de test basées sur les données du Web, par exemple *TREC Blog*. Cependant, ces collections ne tiennent pas compte des spécificités des RS, en particulier l'existence de relations explicites entre les utilisateurs, ou encore les communautés qui se forment entre eux.

Notons qu'une tâche "*Social Networks Search*" est proposée dans le cadre de la nouvelle piste *MicroBlog* de l'édition 2011 de TREC<sup>2</sup>, basée sur des données issues de Twitter.

## 12. Bibliographie

Bao S., Wu X., Fei B., Xue G., Su. Z., Yu Y., « Optimizing web search using social annotations », *World Wide Web Conference 2007*, p. 943-952.

<sup>2</sup> TREC 2011 « MicroBlog Track » : <http://trec.nist.gov/pubs/call2011.html>

Chahrazed Bouhini

- Ben-Jabeur L., Tamine-Lechani L., Boughanem L., « Un modèle de Recherche d'Information Sociale pour l'Accès aux Ressources Bibliographiques : Vers un réseau social », *Atelier Recherche et Recommandation d'Information dans les Réseaux Sociaux, REISO10*, 2010.
- Boubekeur F., Contribution à la définition de modèles de recherche d'information flexibles basés sur les CP-Nets, Thèse de doctorat. Université Toulouse III - Paul Sabatier, 2008.
- Kirchhoff L., Stanoevska-Slabeva K., Nicolai T., et Fleck M., « Using social network analysis to enhance information retrieval systems », *Applications of Social Network Analysis (ASNA)*, 2008.
- Kirsch S.M., Social Information Retrieval, Thèse de doctorat, Université de Rheinische Friedrich-Wilhelms, 2005.
- Konstas I., Stathopoulos V., Jose J.M., « On social networks and collaborative recommendation », *Conference on Research and Development in IR 2009*, p. 195-202.
- Mutschke P., « Enhancing information retrieval in federated bibliographic data sources using author network based stratagems », *Reserach and Advanced Technology for Digital Libraries, 5<sup>th</sup> European Conference, ECDL 2001*, p. 287-299.
- Saito K., Kimura M., Ohara K., Motoda H., « Selecting Information Diffusion Models over Social Networks for Behavioral Analysis », *European Conference, ECML PKDD 2010*, p. 180-195.
- Xu S., Bao S., Cao Y., Yu Y., « Using social annotations to improve language model for information retrieval », *Conference on Information and Knowledge Management, CIKM 2007*, p. 1003-1006.
- Zhou D., Bian J.,Zheng S., « Exploring Social Annotations for Information Retrieval », *World Wide Web Conference 2008*, p. 715-724.