
Vers un modèle de Recherche d'Information Sociale pour l'accès aux Ressources Bibliographiques

Lamjed Ben Jabeur et Lynda Tamine

IRIT - UMR 5505, Université Paul Sabatier, Toulouse

{jabeur, tamine}@irit.fr

1. Le modèle de recherche d'information sociale

La recherche d'information sociale estime la pertinence d'un document à partir de l'importance sociale des auteurs associés (Kirsch *et al.*, 2006) (Mutschke, 2001). Dans ce contexte et en se basant sur la représentation introduite par Baeza-Yates, nous proposons de définir le modèle de recherche d'information sociale par un quintuplet $[D, Q, G, F, R(q, d, G)]$ où la fonction de classement R tient compte de la topologie du réseau social G et intègre divers facteurs de la pertinence sociale.

2.1 Topologie du réseau social : Le réseau social regroupe les acteurs et les données sociaux qui interagissent dans le cadre de la production et la consommation de l'information. Le réseau social des ressources bibliographiques est représenté par un graphe $G(V, E)$ où les nœuds V correspondent aux auteurs, aux annoteurs, aux documents et aux tags. L'ensemble des arcs $E \in (V \times V)$ représente les relations sociales entre ces nœuds. Nous illustrons dans la figure 1 le réseau social des ressources bibliographiques avec les différentes relations sociales qui peuvent exister entre les nœuds.

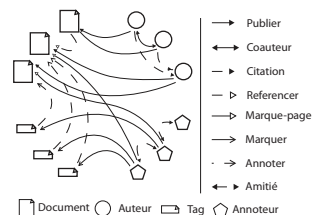


Figure 1. Le réseau social des ressources bibliographiques

2.2 Composante quantitative : Afin de mesurer la similarité, l'influence et le transfert de l'information entre les nœuds du réseau, nous pondérons les relations sociales comme suit :

- **Coauteur** : $Co(i, j) = \frac{2A(i, j)}{A(i) + A(j)}$

- **Citation** : $Ci(i, j) = \frac{C(i, j)}{C(j)}$

- **Publication** : $w(a_k, d) = [1 - \frac{1}{\|T^k\|} H_d^k(T^k)] - \frac{1}{A(k)} \theta$

Avec : $A(i, j)$ le nombre des documents co-écrits par les auteurs a_i et a_j ; $C(i, j)$ le nombre de fois que a_i cite a_j ; $H_d^k(T^k)$ l'entropie des *tags* assignés aux publications de a_k ; $1 - \theta$ le poids par défaut attribué aux auteurs ayant une seule publication.

Nous calculons pour chaque auteur un score d'importance sociale $C_G(a_i)$ à l'aide des mesures suivantes : *Betweeness*, *Closeness*, *PageRank* et les scores d' "Aurority" et de "Hub" calculés par l'algorithme *HITS*. Un score sera drivé pour chaque document par la somme pondérée des scores de ses auteurs comme suit $Imp_G(d) = \sum_{k=1}^m w(a_k, d)C_G(a_i)$. Enfin, le score d'importance social du document $Imp_G(d)$ est combiné avec une métrique de la pertinence thématique basé sur le contenu (Tamine *et al.*, 2009). Nous définissons ainsi la fonction de classement $R(q, d, G) = \alpha RSV(q, d) + (1 - \alpha) Imp_G(d)$ avec $\alpha \in [0, 1]$ est un paramètre de pondération.

2. Évaluation expérimentale

Dans le but d'évaluer l'efficacité de notre modèle, nous avons construit une collection d'articles scientifiques issus de la conférence ACM SIGIR et annotés par le réseau social académique CITEULIKE. Nous avons choisis 25 requêtes à partir des *tags* les plus populaires et nous supposons comme pertinents les documents annotés au moins une fois par le *tag* (requête).

Nous avons comparé les précisions obtenues par notre modèle avec celles obtenues par un système de recherche d'information classique et cela pour la meilleure mesure d'importance sociale (*Hub*) et pour les meilleures valeurs du paramètre α . Comme décrit dans la figure 2, notre modèle atteint une amélioration de 15% à 55% par rapport au baseline.

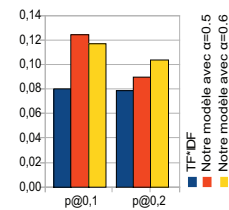


Figure 2. Évaluation l'efficacité de notre modèle

3. Conclusion

Nous avons proposé un modèle de recherche d'information sociale pour l'accès aux ressources bibliographiques qui a la spécificité d'intégrer et de pondérer les relations sociales de citation, de production et d'annotation. En perspective, nous envisageons de comparer notre modèle aux autres modèles de recherche d'information sociale.

4. Bibliographie

- Kirsch S. M., Gnasa M., Cremers A. B., « Beyond the Web : Retrieval in Social Information Spaces », *ECIR 2006*, Springer, 2006.
- Mutschke P., « Enhancing Information Retrieval in Federated Bibliographic Data Sources Using Author Network Based Stratagems », *ECDL 2001, Darmstadt, Germany*, 2001.
- Tamine L., Jabeur A., Bahsoun W., « An Exploratory Study on Using Social Information Networks for Flexible Literature Access », *FQAS*, chapter 8, p. 88-98, 2009.