# LE SYSTÈME DE RECOMMANDATION : UN NOUVEAU MÉDIATEUR DES CONTENUS ?

Gérald KEMBELLEC, CNAM-INTD, Paragraphe
Ghislaine CHARTRON, CNAM-INTD, Dicen

11 juin 2012, Paris



LE SYSTÈME DE RECOMMANDATION:

UN PHÉNOMÈNE OMNIPRÉSENT SUR LA TOILE







### **Amazon**



a

www.amazon.fr/Le-Deuxième-Sexe-Simone-Beauvoir/dp/2070323528/ref=sr\_1\_2?s=books&ie=UTF8&qid=1338236346&sr= 1/275





amazon recommandation



Page 1 sur 17



#### Les clients ayant acheté cet article ont également acheté



Le Deuxième Sexe, tome 1 : Les faits et les... de Simone de Beauvoir

\*\*\*\*\* (5) EUR 9,12



La mondialisation de Olivier Dollfus EUR 9,97



Les origines de la pensée grecque de Jean-Pierre Vernant EUR 9.50



La Crise de la culture de Hannah Arendt AAAAA (5) EUR 7,69



Rapports de force : Histoire, rhétorique, preuve de Carlo Ginzburg EUR 18,34



Le Deuxième Sexe, tome 1 : Les Faits et les... de Simone de Beauvoir (5) EUR 28,98



#### Descriptions du produit

Présentation de l'éditeur

Dans ce second volume, Simone de Beauvoir entreprend " d'étudier avec soin le destin traditionnel de la femme ", c'est à dire de " situer " la femme. " Comment la femme fait-elle l'apprentissage de sa condition, comment l'éprouve-t-elle, dans quel univers se trouve-t-elle enfermée, quelles évasions lui sont permises, voilà ce que je chercherai à décrire. " D'abord sa formation : dans l'enfance, dans l'adolescence, dans l'initiation sexuelle, tout semble disposé, agencé, pour creuser davantage le fossé naturel qui la sépare de l'homme, pour transformer des différences en inégalité, et cette inégalité en infériorité. Ensuite sa situation : Simone de Beauvoir décrit la femme dans le mariage, avec ses prémisses, ses traditions, ses conséquences ; dans la maternité ; dans la prostitution, dans la société ; dans le vieillissement et la vieillesse. Enfin elle envisage les problèmes qui se posent aux femmes qui " héritant d'un lourd passé, s'efforcent de forger un avenir nouveau ". --Ce texte fait référence à / lédition Broché.

#### Détails sur le produit

Broché: 663 pages

Editeur: Gallimard (1 mai 1973)

Collection: Folio
Langue: Français
ISBN-10: 2070323528
ISBN-13: 978-2070323524

Moyenne des commentaires client : ★★★☆☆ № (4 commentaires client)

Classement des meilleures ventes d'Amazon: 2.907 en Livres (Voir les 100 premiers en Livres)

nº1 dans Livres > Sciences humaines > Philosophie > Grandes doctrines et courants philosophiques > Existentialisme

nº32 dans <u>Livres</u> > <u>Littérature</u> > <u>Littérature française</u> > **XXe siècle** 

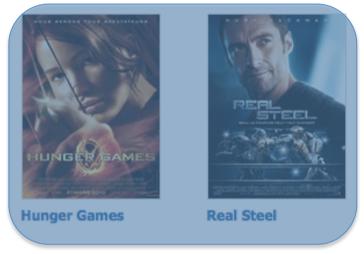
nº45 dans Livres > Dictionnaires, langues et encyclopédies > Langues étrangères

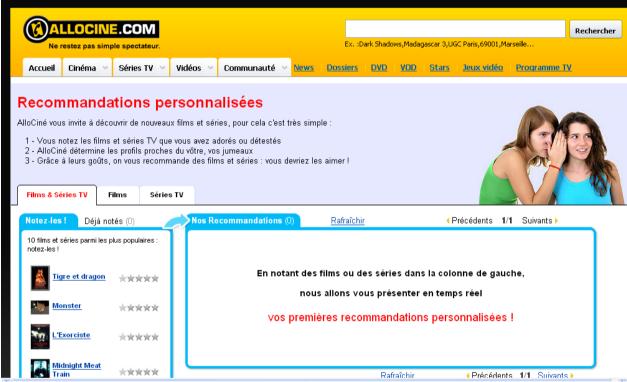
Table des matières complète

Souhaitez-vous compléter ou améliorer les informations sur ce produit ? Ou faire modifier les images?

## Allociné: cinéma

Si vous aimez ce film, vous pourriez aimer ...





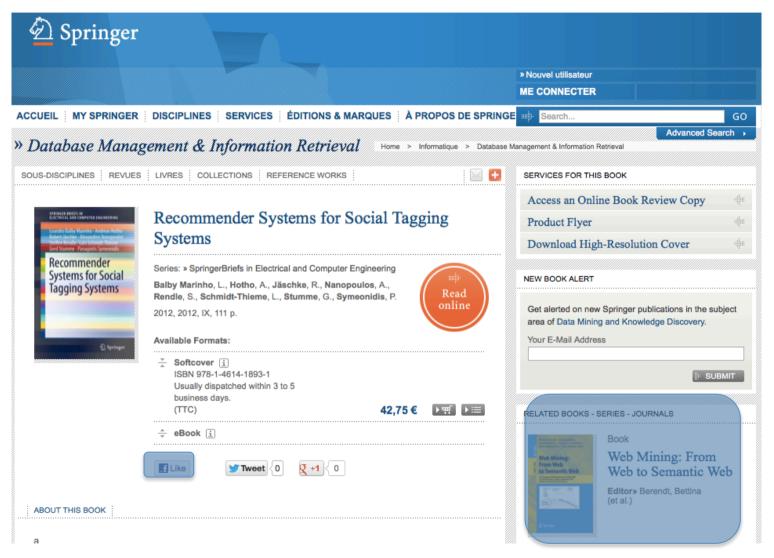
# Babelio: lectures partagées



### Elsevier: édition



# Springer: édition



# Mendeley: Bibliographie collaborative



# Nos objectifs: un travail exploratoire en SIC

- Cerner les enjeux
- Comprendre: ouvrir la « boite noire »
- Comparer les différents systèmes
- Décrypter les services
- Ouvrir des pistes de recherche

# Recherche d'information (vs) Recommandation

COMPARAISON	Système de recherche	Système de recommandation
Demande de l'utilisateur	Il sait ce qu'il veut, demande précise	Demande de découverte, de conseils
Expression du besoin	Explicite (mots, phrase, équation)	Non, définition parfois d'un profil
Technique dominante pour les réponses, propositions fournies	Basée sur le contenu et ses métadonnées (filtrage thématique)	Basée sur les pratiques collaboratives (filtrage collaboratif, recherche de similarités)
Données exploitées dominantes	Contenu et métadonnées	Principales: Logs de parcours, achats, notes données par l'utilisateur  Optionnelles: métadonnées, tags,
Données personnelles exploitées (logs, achats, profils)	non	oui

# E-commerce (vs) « Information retrieval »

Les enjeux du *e-commerce* :

Analyse prédictive pour les outils de Customer Relationship Management (CRM) et d'e-mailing.

Economie de l'attention (retenir le client)

# E-commerce (vs) « Information retrieval »

### Les enjeux de la recherche d'information:

- Aide au choix (critères de contenus, similarités de goûts).
- Réduction de la « surcharge informationnelle »
- Offre de services à valeur ajoutée.

# Types de recommandation

- Les personnes qui ont acheté/vu/aimé cet item ont aussi acheté/vu/aimé ces items...
- Liste d'items selon votre profil.
- Liste de personnes ayant des goûts similaires
- Liste de personnes qui peuvent être intéressées par cet item...

# Système de recommandation: Données explicites, données implicites

- Explicites (données fournies):
  - score donné par l'utilisateur
  - classement d'items par l'utilisateur (selon l'ordre de préférence)
  - choix entre deux items
  - fourniture d'une liste d'items choisis

# Système de recommandation: Données explicites, données implicites

- Implicites: (les traces)
  - items visualisés, écoutés par un utilisateur
  - le temps de visualisation d'un item
  - Liste des achats réalisés

• • •

# Typologie dominante des SR

- CBF (Content Based Filtering):
   utilise les métadonnées sur les contenus
   couplés aux profils des utilisateurs, « filtrage
   thématique »
- CF (Collaborative Filtering): utilise les logs d'usages, « filtrage collaboratif »
- Hybride: mixte, les 2 types de filtrage

# Filtrage collaboratif (CF)



- Requiert la notation des items par les usagers
- Les notes peuvent aider à prédire l'appréciation d'un usager  $\alpha$  sur un item en se basant sur celle d'un autre utilisateur  $\beta$  ayant voté régulièrement de manière similaire.
- Pour déterminer quel sera l'utilisateur β le plus similaire à α, la corrélation de Pearson est souvent utilisée (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994). « bouche à oreille » (Shardanand & Maes, 1995) ou « People-to-people correlation » (Schafer, Konstan, & Riedi, 1999).

$$r = \frac{\sum (\alpha - \overline{\alpha})(\beta - \overline{\beta})}{\sqrt{\sum (\alpha - \overline{\alpha})^2 \sum (\beta - \overline{\beta})^2}}$$

**Équation 1. Corrélation de Pearson** 

Chaque utilisateur pourra bénéficier des appréciations d'au moins un autre usager au profil similaire au sien

(corrélation r tendant vers 1).

Votes	Utilisateur A	Utilisateur B	Utilisateur C	Utilisateur D	Utilisateur E
Objet 1	9	10	7	10	9
Objet 2	7	6	2	1	1
Objet 3	5	1	5	5	4
Objet 4	3	5	3	2	2
Objet 5	1	3	5	7	6

Tableau 1. Exemple d'échantillon de votes

Corrélation	Utilisateur A	Utilisateur B	Utilisateur C	Utilisateur D	Utilisateur E
Utilisateur A	X	0,699	0,243	0,215	0,246
Utilisateur B	0,699	Χ	0,265	0,341	0,413
Utilisateur C	0,243	0,265	X	0,977	0,669
Utilisateur D	0,215	0,341	0,977	X	0,996
Utilisateur E	0,246	0,413	0,669	0,996	Χ

Tableau 2. Proximité des usagers basée sur la corrélation de Pearson

# Filtrage collaboratif (CF)



# **Avantages**

- agnostique au contenu (indépendant)
- sérendipité : découverte
- qualité de la recommandation croit avec l'utilisation du système.

# Filtrage collaboratif (CF)

#### **Inconvénients**

- Le principe de popularité :
  - Plus un objet sera noté positivement, plus il sera recommandé et donc sera de nouveau évalué.
- Manipulation des votes (shilling)
- Démarrage à froid : sans note, pas de recommandation possible.
- Éparpillement :
  - avec un nombre trop faible d'évaluations pour un corpus vaste, les données sont trop éparses pour établir des corrélations

# Filtrage par le contenu (CBF)

- Basé sur l'analyse des contenus et leur description
- Fondé sur la relation entre le profil de l'usager et les métadonnées (mots-clés associés aux items)
- L'indexation sur la fréquence des termes entre un item et l'ensemble du corpus (Salton & Buckley, 1988; Jones, 1972)
- Documents, requêtes et utilisateurs sont représentés par des vecteurs. L'algorithme compare la distance entre les vecteurs et retient les écarts les plus faibles.

# Filtrage par le contenu

### **Avantages:**

- Opérationnel dès le 1<sup>er</sup> utilisateur
- La qualité des recommandations système évolue avec la taille du corpus.
- Très efficace et simple les données textuelles: valorisation des catalogues de bibliothèques ou de librairies.

# Filtrage par le contenu

#### Inconvénients:

- nécessite la disponibilité de métadonnées
- moins de «sérendipité», faute de corrélation avec les usagers
- tendance à l'hyperspécialisation dans les recommandations

# Hybride...

L'hybridation de systèmes de recommandation CBF + CF (R Burke, 2002):

- Pondération :

somme des méthodes disponibles.

- Bascule :

soit une méthode orientée données, soit un filtrage social selon le contexte de recherche

- Combinaison:

possibilité d'enrichir les données intégrées a priori dans le système avec les votes des utilisateurs

### Quelle acceptation par l'usager?

# Données privées :

Les systèmes de recommandation requièrent des données personnelles, explicites.

- fournies par l'usager
- à son insu (exploitation des traces laissées comme la navigation les logs, les achats, les scores ...), par des cookies souvent.

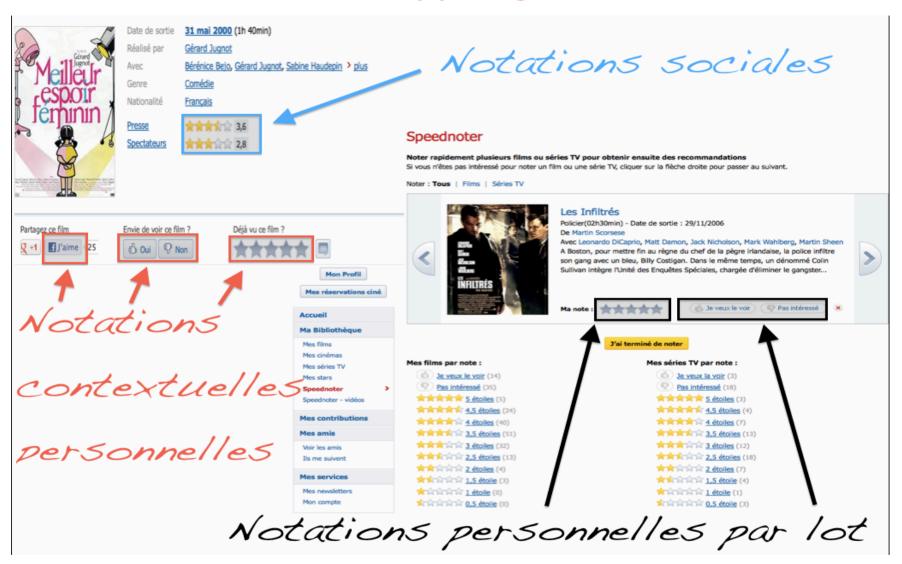
### Quelle acceptation par l'usager?

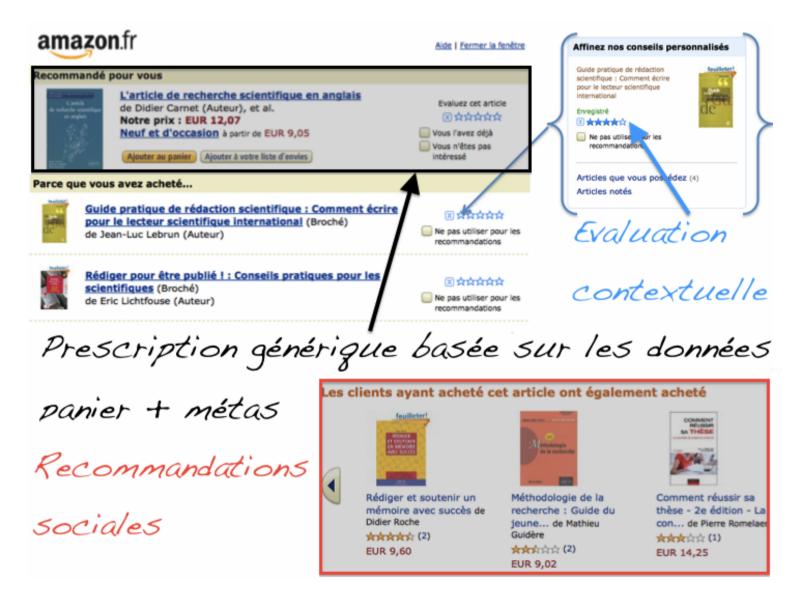
### La transparence, la confiance :

Les calculs de similarité basés sur les scores où des logiques d'influence, voire de manipulation peuvent être présentes (shilling).

Faire confiance, c'est accepter de s'en remettre à autrui que l'on juge compétent et bien intentionné.

La <u>qualité du service rendu</u> peut remporter l'adhésion de l'usager pour fournir des données privées (Li & Unger, 2012).







**Prescriptions de Goodreads** 



	Amazon	Goodreads	Allociné
Objectif majeur	Vente	Catalogue social	Services
Type de recommandation	<ul> <li>- Les clients ayant acheté cet article ont également acheté</li> <li>- articles que les clients achètent après avoir regardé cet article?</li> </ul>	"If you liked this, you might like this" par les amis	« Films similaires »
Type de système de recommandation (CBF, CF, hybride)	Hybride	Hybride	CF
Type de données exploitées	-Métadonnées de contenu -Visualisation et achat des internautes -Commentaires -Scores	-Métadonnées d'Amazon jusqu'en janvier 2012 -Aujourd'hui : score usagers (8 millions) + métadonnées livres d'une autre source	Scores
Participation de l'usager (scores, préférences à donner)	Pas obligatoire	Oui, 20 ouvrages à sélectionner au départ pour initier votre profil	Pas obligatoire (compte FaceBook`)

# Investigations (SIC)...

- Décryptages des plateformes éditoriales et culturelles, les hybridations
- Evaluation des systèmes de recommandation (qualité, efficience, découverte...)
- Effets induits en termes de pratiques culturelles et informationnelles
- Acceptabilité par les usagers: comment se met en place la confiance ?
- Innovations de services pour la recherche d'information

# Quelques références...

- BURKE, R. (1999). Knowledge-based recommender systems. Library, 1-23.
- BURKE, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, *12*(4), 331-370. Retrieved from http://www.springerlink.com/index/N881136032U8K111.pdf
- MEYER Franck (2012), Recommender systems in industrial contexts, doctorat,
   Université de Grenoble, 25 janvier 2012
- SALTON, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513-523. Elsevier.
- WAKETING Simon & all, Readers who borrowed this also borrowed...: recommender systems in UK Library, *Library Hi Tech*, vol.30, n°1, 2012, pp.134-150
- LI, T., & UNGER, T. (2012). Willing to pay for quality personalization? Trade-off between quality and privacy. *European Journal of Information Systems*, (February), 1-22. Nature Publishing Group. doi:10.1057/ejis.2012.13
- MARTINEZ-LOPEZ Francisco, J. Inma Rodríguez-Ardura, Juan Carlos Gázquez-Abad, Manuel J. Sánchez-Franco, Claudia C. Cabal, (2010) "Psychological elements explaining the consumer's adoption and use of a website recommendation system: A theoretical framework proposal", *Internet Research*, Vol. 20 Iss: 3, pp.316 – 341
- FING, Nouvelles approches de la confiance numérique, Février 2011,
   <a href="http://doc.openfing.org/CONFIANCE/ConfianceNumerique\_SyntheseFinale\_Fevrier2011.pdf">http://doc.openfing.org/CONFIANCE/ConfianceNumerique\_SyntheseFinale\_Fevrier2011.pdf</a>
   G.Chartron, G. Kembellec

# Travaux en synergie

- CHARTRON G., MOREAU F., rapport « Culture-medias & numérique », in ANR, Atelier de Recherche Prospective, "Quelles innovations, quelles ruptures dans la société et l'économie numériques ?", 20p + annexes (30p.), Octobre 2011. Article associé, CIDE 2011 http://archivesic.ccsd.cnrs.fr/sic\_00634783
- Projet ISCC-DICEN: Open access, services, interdisciplinarité et expertise (Oasie, 2012): http://culture.cnam.fr/agenda/mars/open-access-services-interdisciplinarite-et-expertise-oasie--479308.kjsp (Emma BESTER, Hans DILLAERTS, doctorants DICEN)
- KEMBELLEC, G. (2012). Technologie et pratiques bibliographiques associées à l'écriture scientifique en milieu universitaire. Arxiv preprint arXiv:1201.4574 (pp. 1-25). Paris, France. Retrieved from <a href="http://hal.archives-ouvertes.fr/hal-00661550/">http://hal.archives-ouvertes.fr/hal-00661550/</a> Doctorant laboratoire Paragraphe, équipe Citu (dir I. Saleh)
   Sujet: Visualisation et cartographie de vastes bases de documents scientifiques en vue de la réalisation semi-automatisée de bibliographies. Thèse en cours sur la navigation par les ontologies

# Merci pour votre attention

