

# Les systèmes de recommandations

*Bibliographie Scientifique*

Projet Transversal

Orange - Polytech'Nantes

Fabien Richard

Valentin Proust



# Introduction

Définition wikipédia :

*“Les systèmes de recommandation sont une forme spécifique de filtrage de l'information visant à présenter les éléments d'information (films, musique, livres, news, images, pages Web, etc) qui sont susceptibles d'intéresser l'utilisateur.”*

Pour comprendre et approfondir cette définition, nous présenterons dans un premier temps les différentes fonctions des moteurs de recommandation. Puis nous verrons en quoi les moteurs de recommandations peuvent apporter des informations plus ou moins ciblées au travers des techniques utilisées. Enfin, nous tenterons d'expliquer leur principe de fonctionnement, sans entrer dans les détails techniques. Mais tout d'abord, demandons nous pour quelles raisons autant de moyens sont mis en oeuvre pour recommander de l'information.

## **Pourquoi les systèmes de recommandation ?**

De plus en plus de choix en contenu (films, livres, musiques, articles, etc) et en équipement (vêtements, électronique, etc) est offert aux consommateurs sur les sites marchands et d'information. Il est dans l'intérêt du consommateur et du vendeur de pouvoir proposer du contenu personnalisé qui correspond à ses goûts, ses humeurs, les tendances etc. L'intérêt pour le consommateur est de disposer de l'information qui l'intéresse le plus et d'accéder rapidement à ce qu'il désirait trouver en se rendant sur le système d'information ou même de trouver ce qu'il n'avait pas encore considéré. L'intérêt pour le vendeur est de cibler le comportement de ses clients pour pouvoir lui vendre plus de produits ou encore augmenter le niveau de satisfaction de l'utilisateur pour le fidéliser ou augmenter le bouche à oreille.

# **Sommaire**

**I) Outils d'aide à l'utilisation des systèmes d'information basés sur un système de recommandation**

**II) Techniques qui implémentent ces outils**

**III) Principe du fonctionnement des moteurs de recommandation**

# I) Outils d'aide à l'utilisation des systèmes d'information basés sur un système de recommandation

Différents outils d'aide à l'utilisation des systèmes d'information sont basés sur des moteurs de recommandation. Nous nous intéressons ici à leur principe et à la valeur ajoutée pour le système d'information, mais aussi à la façon dont la recommandation est visualisée.

## 1. Aide à la navigation

Le principe d'aide à la navigation est de proposer à l'utilisateur de visualiser d'autres informations (souvent en rapport à l'information qu'il est actuellement en train de consulter). Cet outil est notamment utilisé dans le e-commerce. L'utilisateur navigue ainsi au gré des recommandations. L'information se présente pour la plupart du temps sous forme d'images avec une brève description, l'objectif étant d'attirer l'oeil du lecteur.

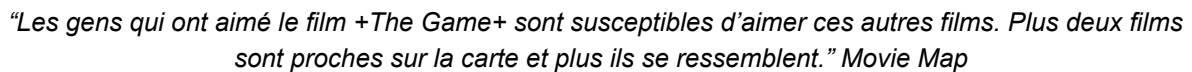
En 2009, l'outil d'aide à la navigation a généré 30% des bénéfices d'Amazon. Cette part importante du chiffre d'affaire montre bien qu'un grand nombre d'utilisateurs *naviguent* sur les systèmes d'information sans réellement savoir ce qu'ils cherchent, à la manière du "lèche-vitrine" du monde réel. Ceci révèle l'importance de la mise en valeur des produits, comme pour la vitrine d'un magasin, à la différence qu'un système d'information peut posséder un très grand nombre de renseignements sur l'utilisateur et donc orienter sa "mise en vitrine" en fonction de son profile, ce qu'un magasin réel ne peut pas faire.

La visualisation peut aussi se faire à l'aide d'une simple liste comme sur les sites de presse.



Sur *LeParisien.fr*, le lecteur peut naviguer d'articles en articles

L'outil d'aide à la décision permet de faire des prédictions sur les valorisations ou préférences qu'un utilisateur attribuerait à un objet ou à un élément social qu'il n'a pas encore considéré. La valeur ajoutée pour le système d'information est d'aider l'utilisateur dans sa prise de décision, l'incitant à la consommation. Les moyens de visualisation peuvent être des listes (ordonnées ou non), des cartes, des imageries, sans oublier les commentaires laissés par les autres utilisateurs qui peuvent aider dans la prise de décision.

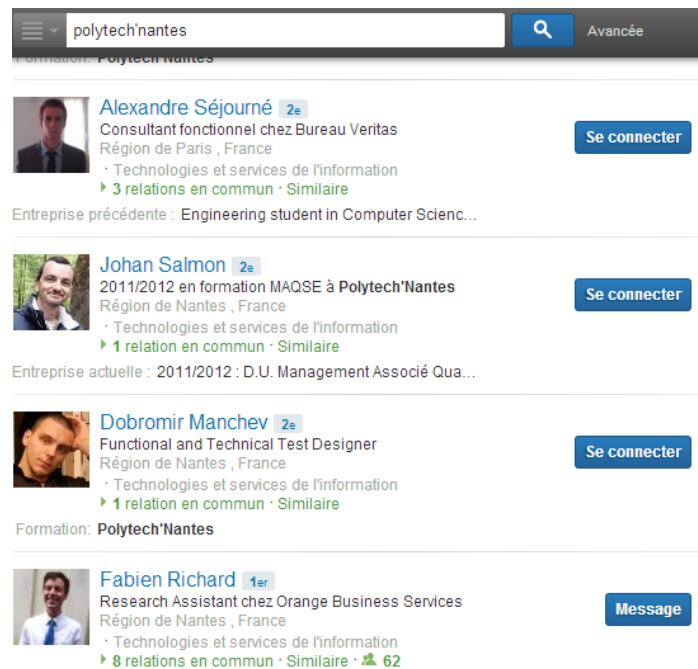


5

### 3. Aide à la comparaison

Cet outil permet d'ordonner ou réordonner les résultats d'une recherche effectuée par un utilisateur pour lui proposer les résultats les plus pertinents. La visualisation se fait souvent à l'aide d'une liste.

Par exemple, si l'on recherche des contacts professionnels sur LinkedIn, le moteur de recherche nous proposera d'abord des contacts avec qui on a des relations en commun, qui parlent la même langue, etc..



*Fonctionnalité de recherche sur le réseau social professionnel LinkedIn*

### 4. Aide à la découverte

Cet outil permet à l'utilisateur de consulter des informations auxquelles il n'aurait pas pensé, de lui recommander un objet ou une personne qu'il n'avait pas considéré auparavant. C'est le principe de beaucoup de systèmes de recommandation.

On peut ainsi trouver sur le service de clips musicaux en ligne de Vevo une rubrique "staff picks" (*sélection du personnel*) qui recommande aux utilisateurs des clips qui a priori ne sont pas personnalisés. On peut en effet questionner la question de neutralité de ces suggestions puisque rien nous permet de vérifier que la sélection des vidéos de cette rubrique n'est en fait pas biaisée par les préférences de l'utilisateur.

## Staff Picks



**Papaoutai**  
Stromae  
76,880,050 views



**J'me tire (Official Video)**  
Maître Gims  
38,976,549 views



**Ailleurs**  
Black M  
4,651,523 views



**I Love You (Official Video)**  
Woodkid  
5,489,665 views

*Catégorie "recommandation du personnel" sur le site de clips musicaux Vevo*

On retrouve un système similaire d'aide à la découverte sur le réseau social Facebook pour la suggestion d'amitiés :



*Fonctionnalité de suggestion d'amitiés sur le réseau social Facebook*

## **II) Techniques qui implémentent ces outils**

Nous avons vu qu'il existe différents outils basés sur des systèmes de recommandations. Nous allons maintenant aborder quelles techniques sont mises en oeuvre pour effectuer ces recommandations.

### **1. Recommandation “éditoriale”**

C'est la méthode de recommandation la plus simple. Elle consiste pour l'administrateur du système d'information à mettre des informations en avant. Par exemple des promotions, Top 10 des ventes, Top 10 apprécié. Cette recommandation n'est pas personnalisée pour l'utilisateur. L'avantage de cette technologie est qu'elle est simple à mettre en oeuvre, en revanche elle ne permet pas de cibler directement l'utilisateur.

### **2. Recommandation “sociale”**

Cette technologie consiste à mettre à disposition de l'utilisateur les avis des autres utilisateurs sur une information. Par exemple coupler un catalogue de produits à un forum. Si l'utilisateur est connecté à un réseau social, il peut consulter l'avis de ses contacts. L'avantage de cette technologie est qu'elle est simple à mettre en oeuvre et qu'elle est personnalisée à l'utilisateur. En revanche, l'inconvénient est que la recommandation n'est pas “automatique”.

### **3. Recommandation contextuelle**

Cette technologie consiste à proposer des informations en rapport avec l'information visualisée. C'est cette technologie qui est utilisée dans les outils d'aide à la navigation. La recommandation se fait en Item-to-Item (mots clés communs, catégorie commune, etc..) ou Item-to-Users ( “les personnes qui ont acheté ces produits ont également aimé ceux ci”).

### **4. Recommandation automatique personnalisée**

C'est la méthode la plus efficace. C'est aussi la méthode la plus difficile à mettre en oeuvre. Elle permet de sélectionner l'information pour l'utilisateur en se basant sur son profil. L'avantage est que la recommandation est ciblée et donc plus pertinente. L'inconvénient est que l'utilisateur doit être connecté au système d'information, que son profil doit contenir des informations.



### III) Principe du fonctionnement des moteurs de recommandation

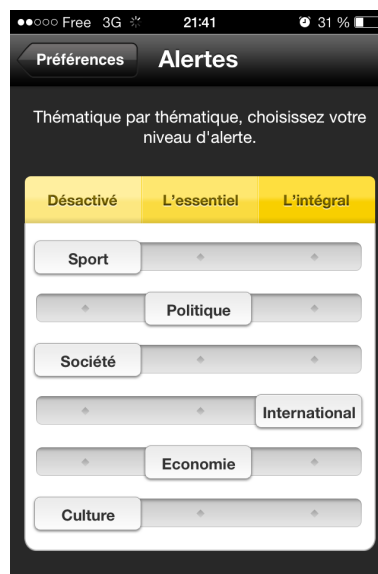
Dans cette partie, nous commencerons par traiter des deux façons de recueillir de l'information sur les produits et sur les utilisateurs pour pouvoir générer des recommandations pertinentes. Puis nous proposons une explication accompagnée d'exemples, d'avantages et d'inconvénients pour les deux types de moteur de filtrage : collaboratif et thématique.

#### 1. Le recueil de l'information

##### 1.1 Recueil actif

La première méthode de recueil est le recueil actif. Elle consiste à recueillir l'information sur les utilisateurs et les objets de manière explicite. Ce type de recueil est très rare car l'utilisateur ne souhaite en générale pas consacrer de temps à donner ses goûts au système d'information mais préfère plutôt fouiller dans le site pour rechercher ce qu'il désire. Pourtant, ce type de recueil est le plus précis et le moins biaisé puisqu'on demande explicitement à l'utilisateur ses goûts sans faire aucune déduction erronée.

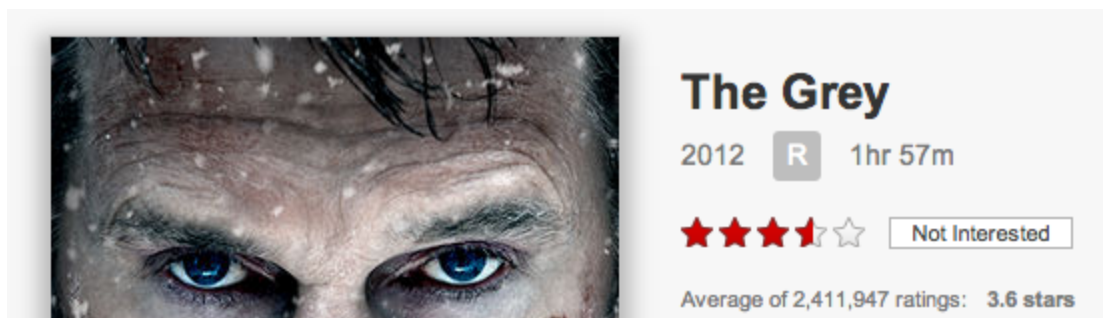
Par exemple, pour recueillir de l'information sur les utilisateurs on peut leur demander leurs goûts. L'application du journal Le Monde est l'un des rares systèmes d'information à proposer à l'utilisateur d'indiquer ses préférences en matière de contenu :



*Choix des préférences pour les alertes sur l'application iPhone du journal Le Monde*

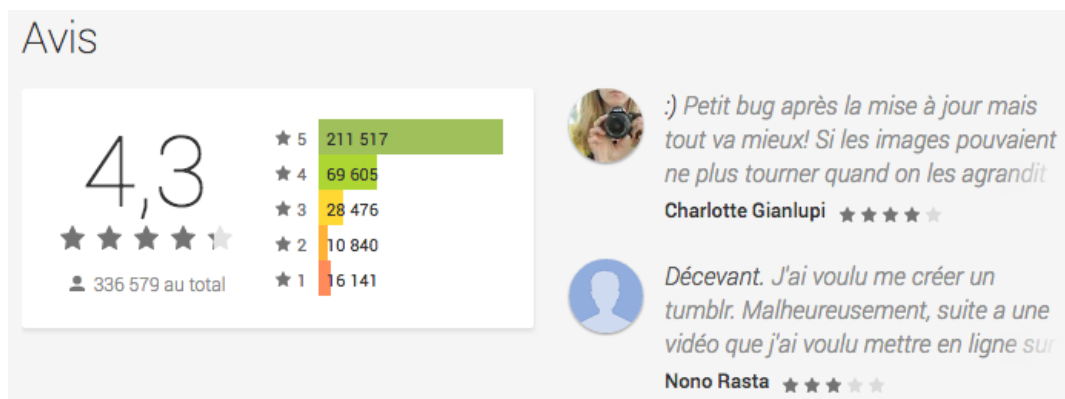
De la même façon, le recueil actif pour les information sur un objet peut se faire de plusieurs manières. Certains site comme Amazon, Netflix ou encore le Google Play pour les applications

Android ont choisi un système de notation allant de 0 étoile pour signifier que l'utilisateur n'aime pas du tout le contenu ou l'objet à 5 étoiles qui veut dire que l'utilisateur l'a adoré.



*Système de notation du site américain de visionnage de films en ligne Netflix*

Ce système de notation simple permet à l'utilisateur de rapidement connaître l'avis des autres utilisateurs sur ce film. Mais il permet surtout au système d'information d'affiner le profil de l'utilisateur en connaissant mieux ses goûts pour pouvoir lui faire des recommandations toujours plus pertinentes. Certains systèmes d'information accompagnent la notation d'un commentaire ou avis qui permet de justifier la note attribuée à l'objet. C'est le cas des trois systèmes d'information cités ci-dessus.



*Illustration de la possibilité d'ajouter un avis à la note attribuée à une application sur le Google Play*

L'avantage de ce type de recueil d'information est qu'il est très précis et très proche de la réalité puisque l'information est explicitement demandée à l'utilisateur. En revanche, il peut être difficile de recueillir un grand volume d'information de cette manière car il faut réussir à encourager les utilisateurs à vouloir mettre une note et/ou un avis sur un produit qu'il a consommé. Des techniques ont été mises en place pour faciliter la notation comme le système de notation en un clic de Netflix qui n'implique pas le rechargement de la page web. Amazon a fait le choix d'envoyer un email après la réception de la commande qui invite à laisser une note au produit :

So Richard Fabien, how did this item meet your expectations?



Disney Assorted Advent Calendar  
3Pack 1.76oz each  
Frankford

Start by rating it



*Email envoyé par le site marchand Amazon invitant à noter le produit consommé*

Un inconvénient au recueil d'information actif est qu'il peut être biaisé par l'échelle subjective de notation de chaque utilisateur. En effet, certains utilisateurs auront tendance à donner facilement 5 étoiles sur 5 à un produit consommé alors que certains n'attribueront cette note qu'en cas de satisfaction totale. On utilise alors des outils mathématiques comme le recentrage des valeurs pour rapporter ces notations à une échelle "standard" de notation. Un autre paramètre important à prendre en compte dans cette technique de recommandation est le temps. En effet, si un utilisateur a aimé un produit ou un contenu à un instant  $t$ , cela ne veut pas dire qu'il l'aimera toujours à un instant  $t+1$ . Ce changement de goût peut être du par exemple au fait que l'utilisateur a grandi et donc que ces centres d'intérêts ont évolué. Il peut également être du à une simple évolution de sa manière de penser, en fonction de l'environnement dans lequel il vit ou des tendances. Il faut donc garder à l'esprit que les modèles utilisateurs qui représentent leurs goûts doivent être mis à jour régulièrement.

### 1.2 Recueil passif

En opposition au recueil actif, le recueil passif consiste à recueillir de l'information sans la demander explicitement à l'utilisateur. On peut par exemple se baser sur son historique de consultation ou d'achat, sur son comportement de navigation. Le site internet de visionnage de films en ligne Netflix analyse par exemple en temps réel le déplacement de la souris de l'utilisateur sur la page web. Ainsi, si un utilisateur survole plusieurs fois avec sa souris un titre de film ou son illustration, on pourra en déduire qu'il a hésité à cliquer dessus et donc qu'il a une préférence pour ce film. Amazon, quant à lui, conserve un historique des commandes passées par un utilisateur pour pouvoir établir son profile.

Une fois l'information recueillie, il faut utiliser un moteur de filtrage pour pouvoir recommander du contenu ou des objets de manière pertinente et personnalisée.

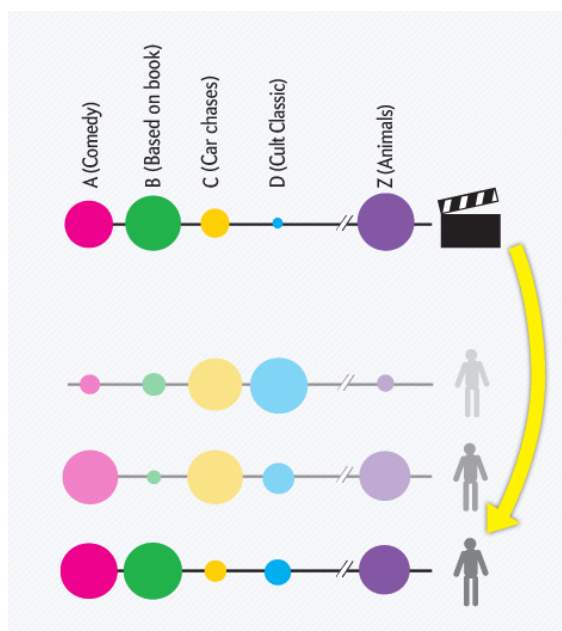
## 2. Les moteurs de filtrage

### 2.1 Moteur de filtrage collaboratif

Le filtrage de contenu consiste à dresser un profil des produits en leur affectant différents attributs (genre, acteurs, popularité, etc) et des utilisateurs (âge, sexe, origine démographique, etc) à partir de leurs usages. A partir de ces profils, les algorithmes tentent de faire correspondre utilisateurs et produits. Les données viennent pour la plupart de logs d'usage. Pour améliorer son moteur de recommandation collaboratif, Netflix avait mis en place un prix, le *Netflix Prize*, qui invitait les concurrents à tenter de trouver des algorithmes capables de recommander les films mieux que le fait le moteur de recommandation actuel de Netflix. Pour cela, les concurrents travaillaient tous sur le même jeu de données. Le prix de 1M de dollars récompensait seulement l'équipe avec le meilleur algorithme et sa valeur dépendait de la qualité de l'algorithme. Ainsi, Netflix pouvait profiter du travail de recherche de milliers de personnes mais ne récompenser que la meilleure équipe.

## 2.2 Moteur de filtrage thématique

Le filtrage thématique ne fait pas intervenir les habitudes d'usage de l'utilisateur du système d'information. Il fonctionne à partir de thèmes ou caractéristiques des informations. Ce moteur de recommandation est plus simple et son fonctionnement est transparent. Il sert en général de complément. L'inconvénient est que le système d'information nécessite un catalogue riche et bien renseigné et de mettre à jour régulièrement le modèle utilisateur pour prendre en compte ses changements de goûts au cours du temps.



*Illustration du principe de filtrage thématique. Source : scientificamerican.com*

# Conclusion

Les systèmes d'informations prennent une part de plus en plus conséquente dans les systèmes d'information, notamment dans le e-commerce. Il existe différents outils de recommandation qui se basent sur des informations aussi différentes que des "mots-clé" sur des produits, des logs d'achat ou des notes d'utilisateurs.

## Références bibliographiques

FRANCESCO Ricci, LIOR Rokach and BRACHA Shapira, Introduction to Recommender Systems Handbook

KOREN Yehuda, BELL Robert, VOLINSKY Chris, Matrix factorization techniques for recommender systems, IEE Computer society 2009

MATHIEU N. Les Algorithmes de recommandations - Podcast [en ligne]. In : Podcast Science, France. Site disponible sur : <http://www.podcastscience.fm/dossiers/2012/04/25/les-algorithmes-de-recommandation/> (Page consultée le 10/11/2013)

MEYER Franck. Recommender systems in industrial contexts. Thèse de doctorat d'université. Grenoble : Université de Grenoble, 2012.