Les Cahiers du CEDIMES
2025, Volume 20, numéro 2, https://doi.org/10.69611/cahiers20-2-11
ISSN 2110-6045

La conception et la réalisation d'un prototype de système de recommandations à base de filtrage collaboratif au sein de la bibliothèque nationale du royaume du Maroc

Amine SENNOUNI

Conservatoire National des Arts et des Métiers et l'Ecole des Sciences de l'Information, Paris, France et Rabat, Maroc, sennouni.amine@gmail.com

Résumé : Les systèmes de recommandations sont parmi les volets prometteurs de l'apprentissage automatique, qui ont révolutionné la recherche de l'information. Ce concept reste nouveau dans les services d'information documentaire, et rare sont les travaux qui traitent ce point-là.

Notre travail consiste à concevoir et mettre en œuvre un système de recommandation basé sur les données implicites des usagers de la bibliothèque, en s'appuyant sur l'approche par filtrage collaboratif.

Mots-clés : Système de recommandation ; Apprentissage automatique ; Données implicites ; Filtrage collaboratif ; Spark ; ALS ; service d'information documentaire.

Abstract: The search for information is an activity whose objective is to enable the user to obtain relevant documents, so that it meets his information needs. Indeed, the corpus composed of documents, subject of an analysis during the indexing phase which produces representations of the documents interpretable by a computer system. While the user carries his need for information and formulates it in a natural language in the form of a request, the latter is in turn interpreted during the indexing of the corpus in order to obtain a representation of the request close to that of the documents.

In our work, we aspire to design a model recommendation system to optimize the selection and access of users to documentary resources, where information is an inevitable instrument for the academic path and the production scientist.

Keywords : Personalization of information ; Recommender system ; Supervised classification ; Modeling ; Library.

Classification JEL : O3

1. Introduction

Aujourd'hui, la quantité d'informations échangées sur le web augmente de façon colossale au point qu'il n'est plus facile pour les internautes d'y trouver des contenus pertinents, qui répondent à leurs attentes. Face à cette multitude d'informations, il est alors évident que les internautes éprouvent une difficulté à prendre les décisions sur les contenus à consulter sur le web, cette surcharge pourra leur être coûteuse en matière de temps, d'énergie et même d'argent. Et puisque les usagers ne sont

pas forcément dans les mêmes contextes et avec les mêmes préférences, il serait judicieux de personnaliser leur expérience de navigation, en offrant à chacun d'eux les contenus dont il aura besoin. Les systèmes de recommandation font partie de la famille des systèmes hypermédia adaptatifs et viennent apporter de l'aide aux internautes en leur proposant des contenus pertinents pour leurs besoins d'informations.

Le concept du système de recommandation commence à émerger dans les services documentaires, qui, en se contentant seulement des moteurs de recherches classiques, n'arrivent pas à mieux servir leurs utilisateurs. Ce problème est dû, d'une part, au fait qu'ils ignorent leurs besoins et, d'autre part, parce que ce besoin est évolutif à travers le temps. Il est devenu alors nécessaire de mettre en place des outils permettant aux utilisateurs d'accéder aux documents qui les intéressent de manière personnalisée, le plus rapidement possible, et sans avoir à les interroger sur leurs besoins informationnels ou de naviguer d'une page à une autre les résultats retournés par le catalogue en ligne.

2. SID et web collaboratif : quel niveau de personnalisation visé ?

Les pratiques informationnelles liées au développement du web 2.0 ou le web collaboratif ont révolutionné les pratiques informationnelles des individus, qui sont passés du stade d'un simple consommateur de l'information à un consommateur actif placé au cœur d'une démarche interactive, personnalisée et ultra-ciblée. Les notions d'intelligence collective, de partage des savoirs et d'interopérabilité des applications ont dessiné de nouveaux contours à l'information en ligne ou en intranet, dans la mesure où les modes de collecte, de traitement et de diffusions des ressources informationnelles et documentaires associés au web classique sont ainsi fondamentalement remis en cause.

En effet, les outils du web 2.0, notamment les blogs, les sites communautaires, les outils de partage de signets et les agrégateurs de fils RSS donnent la priorité à l'usager plutôt qu'à la ressource diffusée. Ainsi, les usagers se procurent de plus en plus un savoir-faire en matière de collecte, de traitement et, éventuellement, de dissémination de l'information. Par conséquent, le rôle classique des professionnels de l'information et de la documentation a été questionné, du moment où leurs publics deviennent de plus en plus exigeants quant à leur attentes et aspirations[1].

De manière empirique, les usagers sont en passe d'abandonner le mode linéaire dans le modèle de recherche d'information et de documentation, qui consiste en l'expression d'une requête pour laquelle des résultats sont retournés, pour s'orienter plutôt vers les modèles navigationnels dont la pierre angulaire est les notions du lien et de l'individualisation des nœuds présents sur les différents réseaux sociaux, et où la navigation ne se limite guère de document à document, mais s'étend pour concerner aussi les relations entre personnes, groupes ou communautés. En outre, ces nœuds aident au tri, la régulation et l'optimisation des flux d'information, comme en témoigne l'exemple des flux RSS.

De surcroît, la description et l'analyse des ressources informationnelles n'est plus l'apanage du professionnel de l'information exclusivement, qui développe les chemins d'accès aux ressources par le biais des langages documentaires universels servant à l'indexation, la classification et l'interprétation des contenus. L'usager peut aussi et sans avoir besoin de maîtriser ces langages, qui freinaient son exploitation optimale des catalogues des services d'information documentaire, décrire des ressources à travers le contenu d'un billet sur un blog, ou caractériser une photo ou une vidéo sur des plateformes dédiées, voire même indexer librement des contenus par l'entremise des tags (Folksonomie). Dans cette optique, chaque individu mobilise ses propres schémas de représentations mentales et ses propres habitudes au lieu de se référer à un cadre théorique et à des outils préétablis et poussés techniquement. Ainsi, les internautes favorisent aussi la popularisation de certaines

[1]BEER, David. Making Friends with Jarvis Cocker : Music Culture in the Context of Web 2.0. Cultural Sociology.2008, vol. 2, n° 2, p. 222-241.

représentations visuelles de l'information dont le « nuage de tags » qui mettent au-devant les occurrences les plus fréquentes des tags attribués par les individus sur un site ou un catalogue donné.

Sur un autre registre, l'usager devient aussi une partie intégrante de l'évaluation des produits et services informationnels qui lui sont proposés. En fait, il peut apprécier et émettre son avis qualitatif ou quantitatif sur les différents contenus, ce qui lui confère le statut d'un évaluateur voire d'un prescripteur. Pratiquement parlant et de point de vue collectif, une ressource popularisée sur les réseaux sociaux contribue à sa visibilité et élargit sa diffusion à une grande échelle. Ce processus rentre dans le cadre du phénomène du « buzz » qui investit le web 2.0, sur le plan individuel, les feedbacks, les appréciations et les activités d'un internaute peuvent être suivis par l'ensemble de la communauté[1].

Néanmoins, Les systèmes de recherche d'information documentaire classiques portent un lot important de limites, dont la principale demeure l'engagement effectif de l'usager qui lance la requête, et par conséquent, il est à l'origine du processus de recherche établi. D'ailleurs, l'absence de l'action de l'usager est synonyme qu'aucun document ne sera restitué. L'interprétation de la requête par le système est un autre défi relevé, puisque la requête est souvent imprécise, voire ambiguë, ce qui rend son interprétation et mise en relation avec la représentation des documents difficile, et de là ne pas pouvoir sélectionner les documents pertinents.

La recherche sur le catalogue en ligne d'un système intégré de la gestion des bibliothèques ne plaide pas en faveur d'une précision en termes des résultats de notices de documents retournées aux utilisateurs, du moment où, dans les deux modes de recherche (simple ou avancée), l'usager se trouve face à un flot important de notices, réparties en lot sur six à dix pages de résultats à l'image d'un moteur de recherche, dans le cas où la requête est fructueuse, et ceci dans le cadre d'un système de navigation linéaire à base de pagination, qui s'oppose à celui du scroll infini reposant sur la présentation de l'ensemble du contenu sur une page unique, qui convient juste de la parcourir du haut jusqu'au bas[2].

En fait, ce scénario expose les usagers à une surcharge cognitive, qui rend la tâche de décider dans quel document visualiser la notice et l'emprunter ou le réserver d'une extrême délicatesse. Par conséquent, un temps énorme risque d'être investi par l'usager et sans avoir la garantie de repérer le bon document.

3. Travaux liés

Pour pallier les manquements avérés et évoqués précédemment, Il existe les systèmes hypermédia adaptatifs qui tendent à personnaliser les contenus en fonction du contexte de l'usager, dont la famille des systèmes de recommandations présentes dans plusieurs sphères (e-commerce, la musique, les vidéos, la presse, les jeux etc.). Quant aux services d'information documentaire, nous soulignons qu'il existe un certain nombre de systèmes liés notamment aux bibliothèques numériques, dont nous citons 3 exemples.

3.1. BookPsychic.

BookPsychic est un système de recommandation simple pour les usagers de la bibliothèque. A l'image de Netflix ou Amazon, il permet d'évaluer les livres et les DVD et enregistre les préférences des utilisateurs, pour déboucher sur des listes de recommandation.

En effet, il s'agit de choisir une bibliothèque qui est inscrite dans book psychic, et noter les livres par genre, en vue d'obtenir des recommandations.

[1]ANAND, S. et B, MOBASHER. Intelligent techniques for web personalization. In: IntelligentTechniques for Web Personalization. Springer. 2005, p. 1-36.

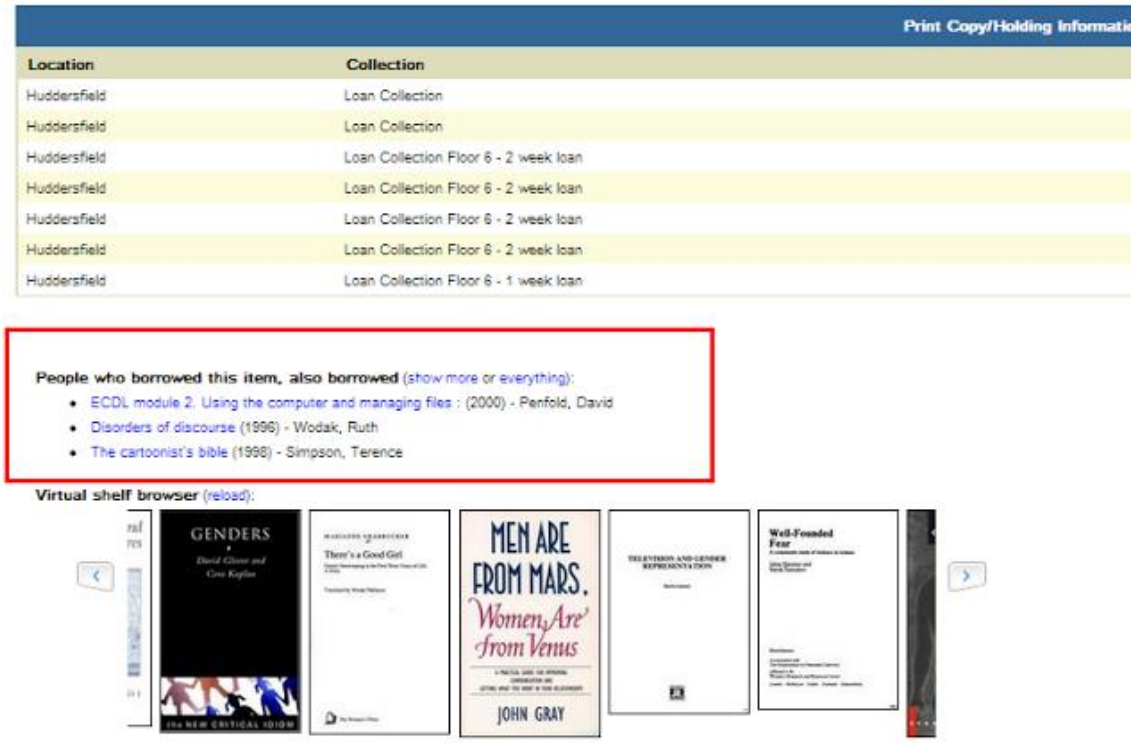
[2]CHANTEPIE, Philippe et Alain, LE DIBERDER. Révolution numérique et industries culturelles. Paris : La Découverte, 2010, 111 p.

Figure n° 1 : Interface de BookPsychic système de recommandation pour une bibliothèque numérique[1]

3.2. Huddersfield Book Recommender system

Il s'agit d'un système qui permet aux utilisateurs de contribuer explicitement dans la génération des recommandations, en ajoutant manuellement des titres similaires à des livres, ou de permettre à d'autres utilisateurs de suivre ce qu'ils prêtent et évaluent.

Figure n° 2 : Interface de Huddersfield Book Recommender system (système de recommandation pour une bibliothèque numérique)[2]

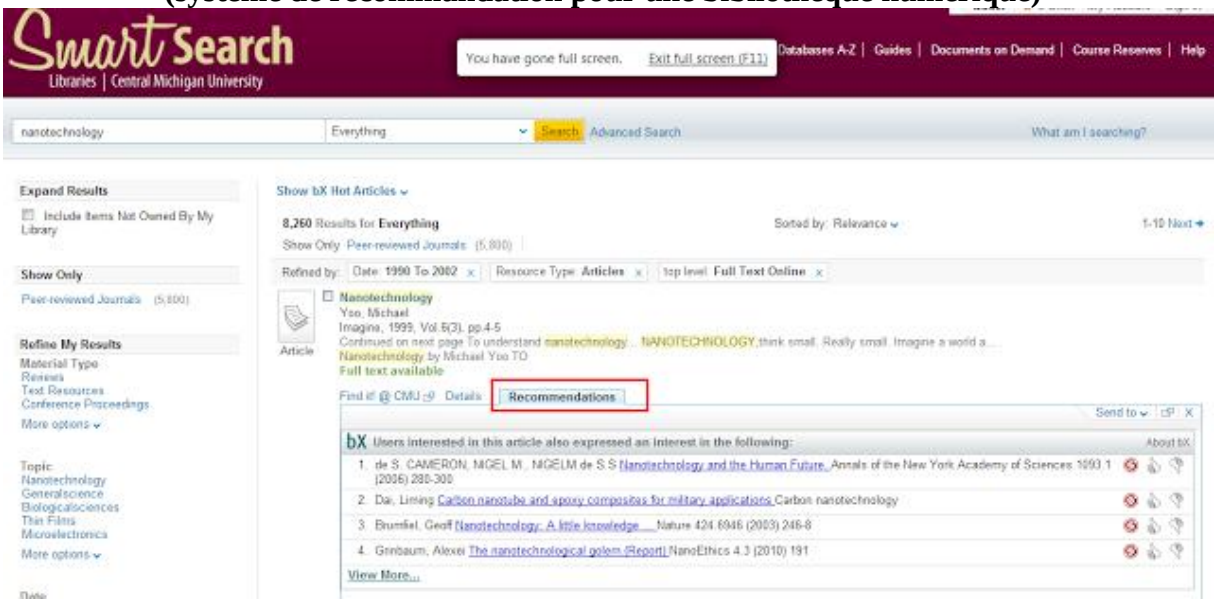


[1] PARK, Jung-ran et al. Catalogingprofessionals in the digital environment: A content analysis of job descriptions. Journal of the American Society for Information Science and Technology. 2009. Vol. 60, n° 4, p. 844–857.
[2] Ibid., p.5.

3.3. Ex LibrisBxRecommender

Les services d'information des bibliothèques universitaires à l'instar des bibliothèques publiques ont fait appel aux systèmes de recommandations, en utilisant le ScholarRank, qui établit un classement des ressources disponibles, et les résultats sont ordonnés par la pertinence dépendant aussi bien de la spécialité de l'étudiant, que du cycle dans lequel il suit ses études.

Figure n° 3 : Interface de Ex LibrisBxRecommender (système de recommandation pour une bibliothèque numérique)[1]



Ainsi, après avoir mis sous la loupe certains systèmes de recommandation incorporés dans des bibliothèques numérique, nous allons procéder à leur comparaison avec la solution que proposons :

Tableau n° 1 : Comparatif des fonctionnalités des solutions de recommandations en services documentaires avec notre solution

Critères / Exemples de solutions de SR	Bibliothèque physique	Bibliothèque numérique	Données explicites	Données implicites	Recommandation sur l'OPAC	Usager authentifié	Usager anonyme	Filtrage collaboratif à base de modèle	Filtrage collaboratif à base de mémoire
BookPsychic	✗	✓	✗	✓	✗	✗	✓	✗	✓
Huddersfield Book Recommender system	✗	✓	✗	✓	✗	✗	✓	✗	✓
Ex LibrisBxRecommender	✗	✓	✗	✓	✗	✗	✗	✗	✓
Notre solution proposée	✓	✗	✓	✓	✓	✓	✓	✓	✗

[1]BOUQUILLON, Philippe. Les industries de la culture et de la communication. Les stratégies du capitalisme. Presses Universitaires de Grenoble, 2008.

Ainsi et après avoir dressé le comparatif ci-dessus, nous déduisons que notre solution se distingue par plusieurs critères, qui ne sont pas partagés par les autres exemples :

- Ce modèle est développé à la base des logs d'usage des usagers, représentant des données implicites, plutôt que leurs retours d'appréciation explicites constituant le nerf de l'écrasante majorité des systèmes de recommandation déployés dans la sphère des industries culturelles et créatives en général, et les services d'information documentaire plus précisément ;
- Notre modèle prend en compte deux cas de figure, à savoir un usager authentifié disposant déjà d'un compte et celui anonyme qui vient d'interroger le système ;
- Notre modèle est implémenté sur l'OPAC du service documentaire et non pas sur une plateforme dédiée ;
- Notre solution repose sur l'approche de filtrage collaboratif à base de modèle, faisant appel aux données du système pour construire un modèle de prédiction sans avoir besoin à consulter les données d'entrée, contrairement à l'approche à base de mémoire, sollicitant toujours les données sur les items et les usagers stockées dans la mémoire du système, en vue d'établir la prédiction ;
- Ce modèle est conçu à partir d'une collection physique de ressources documentaires et non pas de la recommandation appliquée aux ressources numériques disponibles en texte intégral.

4. Conception du modèle de recommandation : Cas de la BNRM

Afin d'orienter notre modèle, trois objectifs ont été fixés :

- Exploiter les données logs implicites des usagers et les convertir en données explicites.
- Faire un profiling des usagers.
- Créer un modèle de recommandation et l'utiliser pour pouvoir générer des recommandations.

4.1. Interfaçage des résultats de recommandations.

Ces objectifs constituent des modules qui composent notre application[1].

Il nous a été demandé par le service informatique de développer une interface pour pouvoir visualiser ces recommandations et tester le système. Une fois testé, le service pourra s'occuper de son intégration dans le SGIB de la BNRM.

Notre solution se base sur les fichiers logs générés par le SIGB, qui contiennent l'activité et les traces des usagers, pour leur proposer des contenus qui peuvent leur être pertinents. Etant basé sur le filtrage collaboratif comme approche de recommandation, notre système va permettre :

- D'échanger des idées entre usagers en faisant profiter les uns des préférences des autres.
- De favoriser le transfert des connaissances
- De faciliter la sélection de l'information.

4.2. Scénario de recommandation

Le scénario de recommandation que l'application doit garantir fait distinguer deux cas de traitement : Un utilisateur abonné peut s'authentifier via son login et mot de passe attribués par l'administrateur. Ainsi, le système charge son profil qu'il a créé, et, soit le système lui présente des

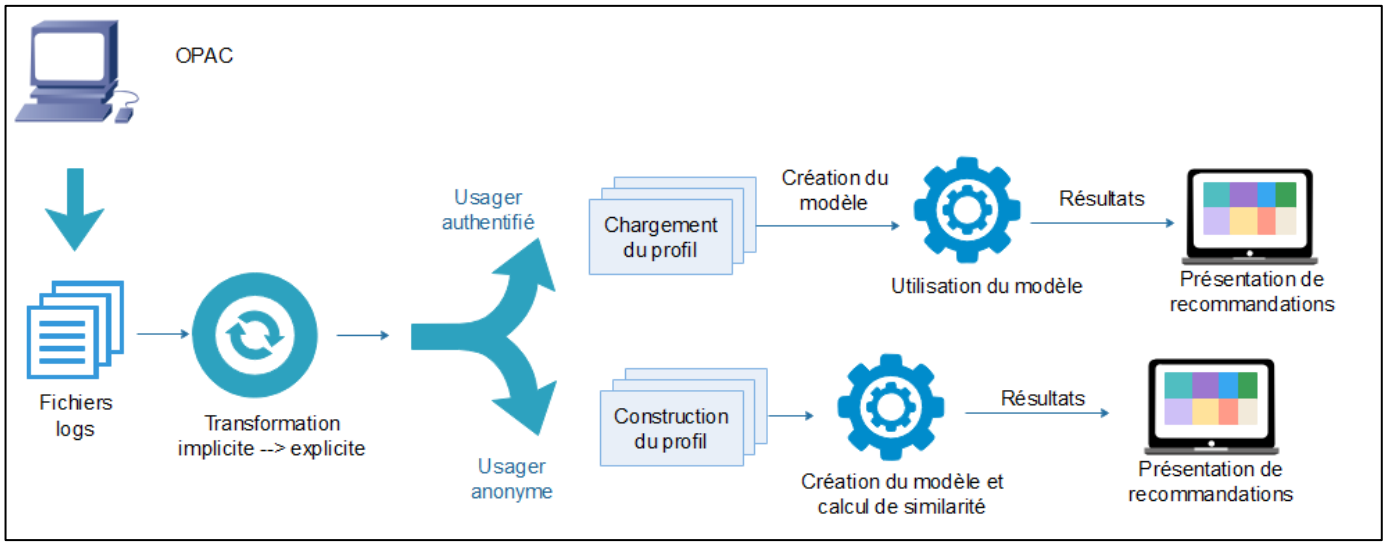
[1]LECHANI, Tamine Lynda et Mohand, BOUGHANEM. Accès personnalisé à l'information : Apporches et techniques. Rapport interne. RIT, 2005. Disponible sur :< ftp://ftp.irit.fr/IRIT/SIG/rapport_Perso_0904_VF.pdf> [consulté le 6 Octobre 2016].

ecommandations en se basant sur ses besoins antérieurs, soit, à travers l'analyse de similarité à base de profil, le système lui fait des recommandations qui lui sont personnalisés

4.3. Architecture conceptuelle de la solution

Notre système exploite les données contenues dans les logs, à travers un processus de plusieurs étapes, pour pouvoir générer des recommandations, on pourra schématiser le processus comme suit :

Figure 4 : Schématisation du processus de recommandation à travers l'application[1]



Comme mentionné dans la figure ci-dessus, le processus de recommandation se fait en plusieurs étapes :

L'extraction de fichiers logs

Dans notre cas, et pour faire des simulations de recommandations, on utilisera un jeu de données que le service a généré depuis les fichiers logs par PMB, qui contient des informations de navigation d'usagers, structurées comme suit :

- Identifiant de l'utilisateur : identifiant donné par le système à un usager connecté, qui permet de l'identifier.
- Dernier document consulté : c'est le document qui met fin à la recherche, c'est le besoin que l'usager cherche à satisfaire.
- Nombre de rebond de la requête : nombre de tentative de recherche avant de trouver le document voulu, ce nombre nous renseigne sur l'intérêt que cet usager actif a vis-à-vis du document trouvé.
- Affichage du document dans les résultats : nombre de fois que le document se génère parmi les résultats, c'est entre autres le nombre de consultation de documents non pertinents à la place du document trouvé.

Le temps passé sur l'OPAC : c'est le temps investi dans la recherche.

La Transformation des données implicites en données explicites

Comme déjà mentionné dans l'état de l'art, le système de recommandations peut se baser sur deux types d'informations :

- Des informations explicites qui représentent les notations que les utilisateurs ont attribuées aux livres, sur une échelle allant de 1 à 10 points, celles-ci sont directement exploitables par le système de recommandation.

[1]Figure générée dans le cadre de notre étude

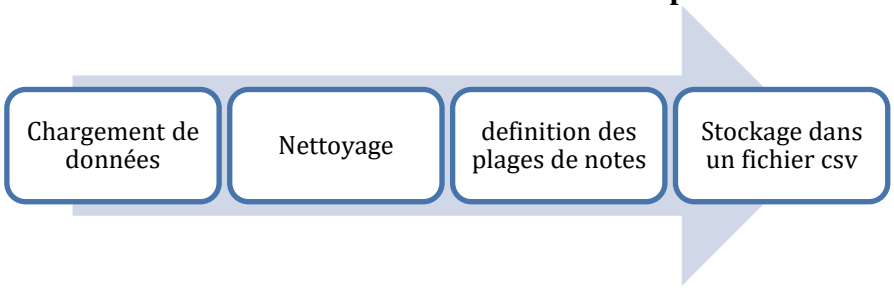
- Des informations implicites qui caractérisent le comportement de navigation de l'usager, que le SIGB collecte durant sa navigation, il s'agit du temps de connexion sur l'OPAC, le nombre de rebond de la requête, et le nombre d'affichage du document dans les résultats.

Les informations implicites collectées restent hétérogènes et à l'état brut. Le temps, par exemple, ne peut pas être rapproché au nombre de rebond de la requête et au nombre d'affichage du document dans les résultats .De plus la comparaison entre plusieurs profils d'utilisateur reste inaccessible, un utilisateur par exemple qui s'est connecté 20 minutes, en lançant 13 requêtes et générant un document 7 fois, ne peut pas être comparé à un autre utilisateur ayant navigué sur l'OPAC pendant 11 minutes, générant le même document que le premier 8 fois en lançant un nombre total de 13 requêtes.

La méthode répandue dans le traitement des données implicites est de les transformer en notations semblables à celles collectées explicitement[1]. Or, il ne s'agit souvent que d'un seul paramètre, le temps de connexion par exemple.

La particularité de notre projet réside donc dans la combinaison de trois paramètres implicites pour estimer une note générale, qui exprime l'intérêt de l'usager vis-à-vis du document. Leschéma ci-dessous présente les étapes du processus de transformation de données implicites en notations :

Figure n° 5 : Processus de transformation de données implicites en notations



Pour convertir les données implicites en explicites, on commence par nettoyer les données, définir ensuite les plages de notes à appliquer, puis stocker le fichier dans un format qu'on pourra exploiter par la suite.

Chargement de données

Les données employées sont souvent sous format texte (fichier log), les informations contenues dans ce fichier log ont été exportés sous forme de base de connaissance, ayant la forme ci-dessous :

Tableau n° 3 : Forme du fichier à exploiter.

Identifiant de l'utilisateur	Dernier document consulté	Nombre de rebond de la requête	Affichage du document dans les résultats	Le temps passé sur l'OPAC
1	Doc1	5	3	11
2	Doc 2	6	11	13

Le nettoyage de données

C'est l'opération de détection et correction des erreurs présentes sur le fichier, en plusieurs étapes :

Traitement des valeurs manquantes : à travers l'analyse du fichier, on n'a trouvé que 3 enregistrements contenant des valeurs manquantes. Puisqu'il s'agit d'une faible portion qui représente moins de 5 % de la taille du fichier, nous avons enlevé ces enregistrements et nous n'avons conservé que les enregistrements complets.

[1]OARD, Douglas W., KIM, Jinmook, et al. Implicit feedback for recommender systems. In : Proceedings of the AAAI workshop on recommender systems. 1998. p. 81-83.

Nous avons limité chaque session de connexion à 30 minutes, ce qui est raisonnable, du moment où un utilisateur ne pourra pas se connecter au-delà de cette durée, et on a éliminé toutes les valeurs qui contiennent des durées dépassant 30 minutes, le fichier résultant contient 500 enregistrements, après avoir été de l'ordre de 600 lignes (enregistrements).

Définition des plages de notes

Pour définir les plages de notes, nous devons diviser les données en intervalles de même amplitude. Le nombre d'intervalles est conditionné par l'échelle de notation à adopter. Ainsi, deux échelles de notations se présentent : l'une sur 5 points et une autre sur 10 points. Cette dernière a été adoptée pour garantir plus de précisions en ayant des intervalles de 10 au lieu de 5, chaque intervalle sera remplacé par une note, ce qui rend plus fiable l'estimation de la notation qu'on se propose d'établir.

Après la définition des plages de notes, on remplace chaque valeur contenue dans la plage par la notation correspondante, comme illustré dans le schéma ci-dessous :

Tableau n° 4 : Définition des notations correspondantes à chaque plage de données

	Nombre de rebond de la requête		Affichage du document dans les résultats		Le temps passé sur l'OPAC	
	Borne inférieure	Borne supérieure	Borne inférieure	Borne supérieure	Borne inférieure	Borne supérieure
Note 1	[0;	0,8[	[0;	2,3[	[5;	7,4[
	[0,8;	1,6[	[2,3;	4,6[	[7,4;	9,8[
	[1,6;	2,4[	[4,6;	6,9[	[9,8;	12,2[
	[2,4;	3,2[	[6,9;	9,2[	[12,2;	14,6[
	[3,2;	4[	[9,2;	11,5[	[14,6;	17[
	[4;	4,8[	[11,5;	13,8[	[17;	19,4[
	[4,8;	5,6[	[13,8;	16,1[	[19,4;	21,8[
Note 8	[5,6;	6,4[	[16,1;	18,4[	[21,8;	24,2[
	[6,4;	7,2[	[18,4;	20,7[	[24,2;	26,6[
Note10	[7,2;	8]	[20,7;	23]	[26,6;	29]

Cette opération garantit l'homogénéisation des données, puisqu'on aura des notations sur une échelle unifiée. On obtient vers la fin un fichier structuré, contenant des notations sur une échelle de 10 points.

Stockage sous format finale

L'étape suivante est de calculer une note générale qui prend en compte les notations produites. Après des discussions avec l'administrateur de la base de données, on s'est mis d'accord sur le fait que l'affichage du document dans les résultats aura une pondération élevée, puisqu'il exprime au mieux l'intérêt de l'utilisateur vis-à-vis du document.

La formule adoptée pour calculer la note générale est :

Note générale = Note (Nombre de rebond de la requête)*30% +Note (Affichage du document dans les résultats)*40% + Note (Le temps passé sur l'OPAC)*30%.

Le fichier résultant sera ainsi sous cette forme :

Tableau n° 5 : Format final du fichier à exploiter.

Id de l'utilisateur	Document consulté	Note globale estimée
1	Doc 1	Note 1
2	Doc 2	Note 2

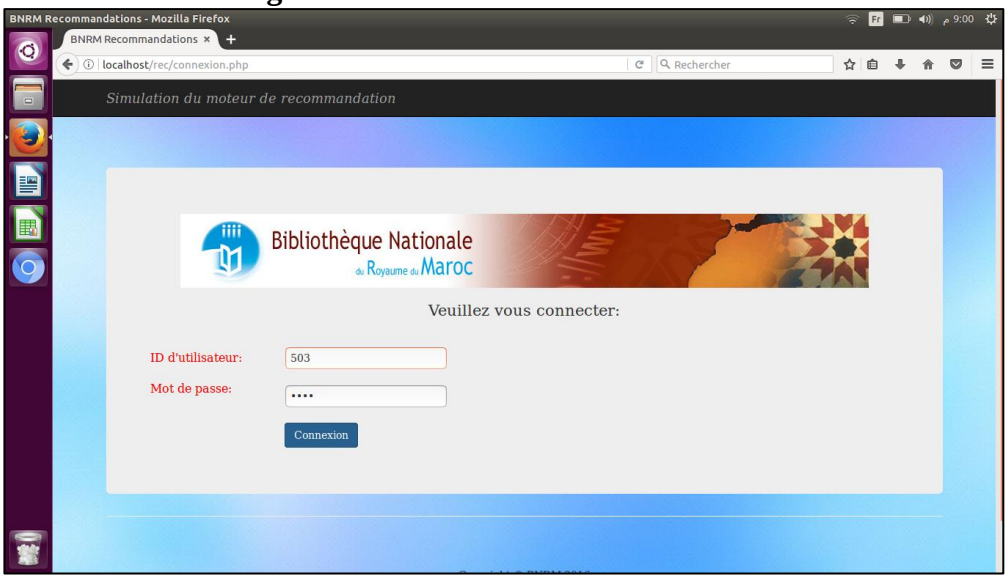
Ce fichier va être transformé par la suite en fichier CSV que l'application exploitera.

5. La validation du modèle

5.1. Authentification de l'usager

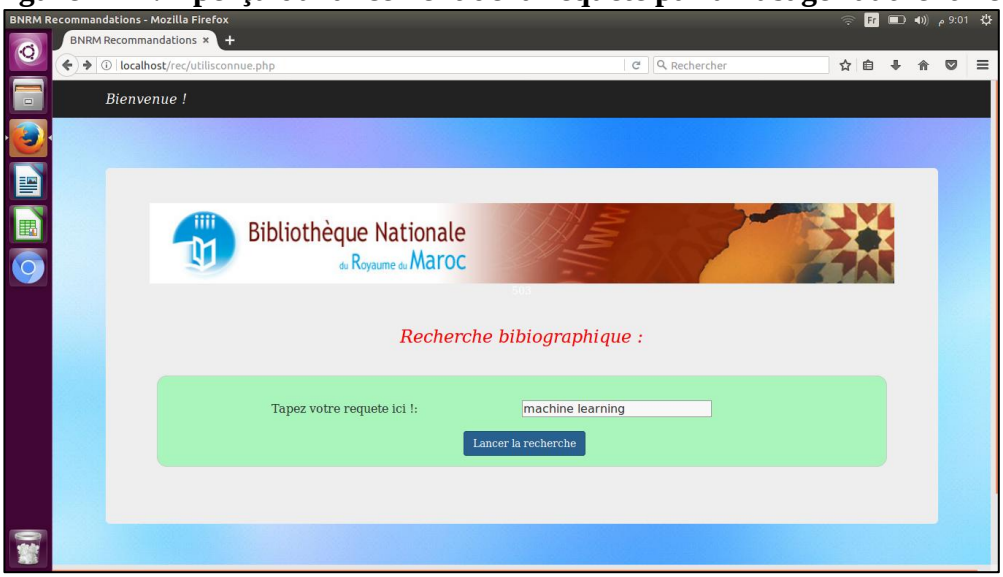
L'usager s'authentifie à travers l'application PHP, à l'aide de son identifiant et son mot de passe, l'interface d'accueil de l'application se présente comme suit :

Figure n° 6 : L'interface de connexion[1]



L'utilisateur s'authentifie via son identifiant et son mot de passe pour bénéficier des recommandations liées à son profil. Une fois authentifié, l'identifiant de cet utilisateur sera communiqué à Spark où se poursuivront les traitements.

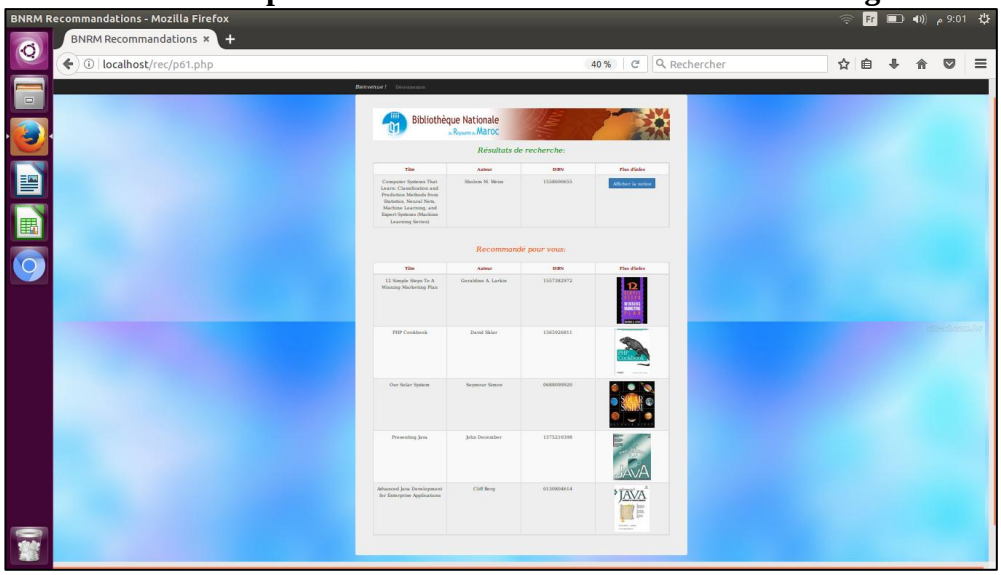
Figure n° 7 : Aperçu du lancement de la requête par un usager authentifié[2]



Cette interface permet à l'utilisateur d'effectuer des recherches sur la base de données, une fois une recherche lancée, il reçoit des recommandations qui lui sont appropriés. La figure ci-dessous montre un aperçu sur les résultats de recherche accompagnés de recommandations en bas.

[1] Figure générée dans le cadre de notre étude
[2] Ibid.

Figure n° 8 : Zoom de la présentation des recommandations à un usager authentifié[1]



Les tops 5 recommandations sont affichés directement après les résultats de recherche. Ces recommandations contiennent le titre du livre, l'auteur, L'ISBN, et un aperçu sur sa couverture.

5.2. L'évaluation par le calcul des fonctions d'erreur (RMSE, MSE, K MAPK)

Les fonctions d'erreurs permettent de mesurer la distance moyenne entre les prévisions estimées et les observations correspondantes. Ainsi, nous pouvons déduire qu'une prédiction est parfaite si la valeur de la fonction est proche de 0. Les fonctions les plus populaires dans l'évaluation des systèmes de recommandations, et qu'on peut exécuter sur Spark, sont :

- La fonction d'erreur moyenne quadratique (MSE : « MeanSquaredError »), c'est une fonction qui calcule le carré de la distance entre la prévision et l'observation, elle est connue par sa grande sensibilité à la précision.
- La fonction RMSE : « RootMeanSquaredError » : c'est la racine de l'erreur moyenne quadratique (MSE), elle est largement utilisée, à la place du MSE, pour évaluer les systèmes de recommandation.
- La précision moyenne en K MAPK « Meanaverageprecisionat K », mesure les scores de pertinence moyenne d'un ensemble de documents top-K présentés en réponse à une requête, cette mesure est la plus adaptée aux données implicites.

Après plusieurs tests, en jouant sur les paramètres du modèle, on s'est arrêté sur des valeurs optimales qui minimisent le RMSE. Les valeurs des paramètres de ce modèle optimal ainsi que les résultats retournés par les fonctions métriques d'évaluation sont récapitulées dans le tableau ci-dessous :

Tableau n° 6 : Valeurs des fonctions d'évaluations

Paramètres du modèle	RMSE	La précision moyenne en K
Itérations=110. Rang=10. Lambda=0.0.1. Alpha=1.	2.712	K=1 →0.820 K=3 →0.277 K=5 →0.166 K=7 →0.118 K=9 →0.092

1 Figure générée dans le cadre de notre étude

La fonction RMSE n'est pas une bonne mesure pour prédire la préférence des utilisateurs, notamment lorsqu'il s'agit des données implicites. Cette mesure doit être combinée et, dans certains cas, remplacée par d'autres mesures issues de la récupération d'informations telles que la précision[1].On tiendra plutôt compte dans notre cas des valeurs de la précision moyenne en k, et d'après les résultats de cette mesure qui tendent vers 0, la marge d'erreur de notre système est minime, on peut dire qu'il donne des recommandations pertinentes[2].

5.3. Calcul du temps d'exécution du programme

Un autre critère qu'on peut utiliser pour évaluer la performance de notre système est de connaître le temps d'exécutions du script de recommandation. On a procédé par plusieurs tests de calcul du temps d'exécution, et on a pris les mesures ci-dessous :

Tableau n° 7 : valeurs des tests de durées d'exécution du script de recommandation

Test	Durée d'exécution
1	6.64
2	6.40
3	6.23
4	6.12
5	6.37
6	5.83
7	5.48
8	5.59
9	5.71
10	5.20

La moyenne des valeurs générées par le test est égale à 5.95 secondes, donc notre système offre les recommandations dans plus au moins 6 secondes. Nous déduisons ainsi que notre système est donc performant, et pertinent

6. Conclusion

Notre projet a abouti au prototypage et l'implémentation d'un système de recommandations adapté au sein d'une unité documentaire en l'occurrence, la Bibliothèque Nationale du Royaume du Maroc (BNRM). De fait, le système a été implémenté sur Apache Spark en langage Scala. Notre système procède à une analyse des données de navigation des usagers à travers plusieurs algorithmes, notamment l'algorithme ALS, pour ensuite présenter les résultats de l'analyse sous forme de recommandations, de plus le système distingue entre un usager abonné et un usager anonyme en terme de recommandations, car si le premier bénéficie des suggestions de livres selon son besoin connu par le système, le second usager (anonyme) bénéficie également des recommandations, mais sur la base des notices qu'il consulte. Après plusieurs tests, il s'est avéré que notre système est pertinent.

Sur un autre registre, des améliorations de notre travail peuvent être envisagées. Il s'agit d'automatiser le processus de transformation de données implicites en données explicites, et d'intégrer une composante qui permettra de tirer parti des données accumulées par le système de recommandation. Ces données de masse pourront être utiles pour produire des statistiques sur la consultation des collections et identifier encore davantage les domaines susceptibles d'intéresser les

[1] PARRA, Denis, KARATZOGLOU, Alexandros, AMATRIAIN, Xavier, et al. Implicit feedback recommendation via implicit-to-explicit ordinal logistic regression mapping. Proceedings of the CARS-2011, 2011.

[2] NAAK, Amine. Papyres : Un système de gestion et de recommandation d'articles de recherche. Mémoire d'obtention de maitrise en sciences informatiques. Université de Montréal, 2009

usagers de la BNRM. Ainsi, la structure trouvera de la matière pour orienter et rationaliser sa politique d'acquisition.

Bibliographie

1. ANAND, S. et B, MOBASHER. Intelligent techniques for web personalization. In: IntelligentTechniques for Web Personalization. Springer. 2005, p. 1-36.
2. AURAY, Nicolas, La consommation en régime d'abondance. La confrontation aux offres culturelles dites illimitées. Revue française de socio-économie. 2011.vol. 2, n° 8, 2011, p. 85-102.
3. BEER, David. Making Friends with Jarvis Cocker: Music Culture in the Context of Web 2.0.Cultural Sociology.2008, vol. 2, n° 2, p. 222-241.
4. BOUQUILLON, Philippe. Les industries de la culture et de la communication. Les stratégies du capitalisme. Presses Universitaires de Grenoble, 2008.
5. BOUQUILLON, Philipp. Industries, économie créatives et technologies d'information et de communication. Tic & Société. 2010, vol.4, n° 2, p. 7-40.
6. CHANTEPIE, Philippe et Alain, LE DIBERDER. Révolution numérique et industries culturelles. Paris : La Découverte, 2010, 111 p.
7. LECHANI, Tamine Lynda et Mohand, BOUGHANEM. Accès personnalisé à l'information : Apporches et techniques. Rapport interne. RIT, 2005. Disponible sur :< ftp://ftp.irit.fr/IRIT/SIG/rapport_Perso_0904_VF.pdf> [consulté le 6 Octobre 2016].
8. KOP, R. The Challenges to Connectivist Learning on Open Online Networks : Learning Experiences during a Massive Open Online Course.European Distance and E-learning Network annual Conference.2010, p.4- 32.
9. NAAK, Amine. Papyres : Un système de gestion et de recommandation d'articles de recherche. Mémoire d'obtention de maitrise en sciences informatiques. Université de Montréal, 2009