

1 Site de vente en ligne

1.1 Contexte

Un vendeur de livres, CD et DVD en ligne, pour améliorer ses ventes, souhaite compléter son système d'information d'un outil de recommandation : lorsqu'un client achète un album de Franz Ferdinand, un livre d'Oscar Wilde, un film avec Sylvester Stallone, il pourrait être pertinent de lui proposer automatiquement un disque des Strokes, un livre d'Edgar Allan Poe, un film avec Arnold Schwarzenegger. Pour cela, il faut, d'une part, préciser la description des articles mis en vente (au-delà d'une simple référence, d'un titre, d'un auteur ou réalisateur) et, d'autre part, mémoriser le comportement des utilisateurs (au-delà d'un ensemble de références, d'une date d'achat et d'une tarification).

1.2 Description fine des articles

Comment qualifier un article relativement aux goûts d'un utilisateur ? Après une première réflexion, le vendeur pense aux données suivantes :

- *époque* (moyen-âge, renaissance, contemporaine, ...),
- *origine* (sud-américain, asiatique, britannique, français, américain, ...),
- *genre* (SF, drame, romance, rock, rock indépendant, classique, baroque, ...),
- *thème* (pour les livres et les films, *eg.*, guerre, affaires criminelles, famille, ...),
- autres informations caractérisant un article, telles que auteurs et interprètes, date de sortie, ouvrage en promotion, œuvre primée.

1.2.1 Catégorisation des œuvres

Pour ces différentes informations, **les catégories peuvent dépendre du type d'ouvrage**, par exemple : "*rock*" est une catégorie des disques, "*SF*" est une catégorie des livres et des films, "*Asiatique*" est une catégorie des livres, des films et des disques. De plus, **les valeurs que prennent ces catégories sont hiérarchisées**. Par exemple : "*rock*" contient ou *généralise* "*pop-rock*", "*électro*" et "*rock indépendant*"; "*Britpop*" spécialise "*pop-rock*"; "*fantastique*" *généralise* "*science-fiction*" et "*horreur*"; "*horreur*" *spécialise* "*gore*"; "*Chine*", "*Japon*", "*Corée*" *spécialisent* "*asiatique*". Pour l'instant, le vendeur souhaite aller jusqu'à trois niveaux hiérarchiques pour "*genre*" (*eg.*, *SF* → *horreur* → *gore*), deux pour "*origine*" (continent et pays), un pour "*thème*" et un pour "*époque*".

Un même ouvrage peut avoir **plusieurs thèmes** (livre, film, opéra), **plusieurs genres** (livre, album, film); *eg.* : "*famille*" et "*affaires criminelles*", "*drame*" et "*fantastique*". Lorsque c'est le cas, le vendeur souhaite pouvoir donner un poids (de 1 à 4) à ces thèmes, comme à ces genres. Par exemple, le vendeur considère que le film *La vie est belle* de Roberto Benigni a pour thème la guerre, et pour genres, "*comédie*" (genre principal de poids 1) et "*histoire*" (genre secondaire de poids 2).

Un article peut être **une œuvre originale** (cas d'un livre) ou **l'interprétation d'une œuvre originale**; par exemple : *la 9ème symphonie de Beethoven* est une œuvre originale, *Sir Simon Rattle dirigeant l'orchestre philharmonique de Vienne en 2003* en est une interprétation; les deux albums *True Love Waits : Christopher O'Riley Plays Radiohead* et *Hold Me to This : Christopher O'Riley Plays Radiohead* du pianiste Christopher O'Riley sont des reprises de morceaux de divers albums originaux du groupe Radiohead; le film *Entre les murs* de Laurent Cantet est une mise à l'écran du roman du même titre de François Bégaudeau. De façon plus indirecte, certaines œuvres sont inspirées par d'autres : par exemple, le film *Galaxy Quest* de Dean Parisot est une parodie de la série et des films *Star Trek*. Enfin, pour les ouvrages inspirés d'un fait, divers ou historique, il faudra pouvoir associer ce fait à l'ouvrage.

1.2.2 Catégorisation des interprètes

Pour un film, **le type des personnages** qu'il met en scène peut également intervenir (*eg.*, une blonde plantureuse, un(e) comique de service, un séducteur patenté, un tueur psychopathe). Le vendeur souhaite

alors créer une base d'acteurs avec, pour chaque acteur, le type de rôles tenus dans les films qu'il a tournés (*eg.*, 4 fois l'idiote du village, 1 fois l'amoureux éperdu, 2 fois le détective décadent, ...).

De même, pour les œuvres musicales, au-delà du compositeur et de l'interprète principal (l'interprète principal pouvant être un groupe, un chanteur, un orchestre, un instrumentiste selon le type d'œuvre et d'interprétation), il faut prendre en compte **les instruments** (cordes, piano, guitare, ...), ainsi que **les interprètes** (membres du groupe, instrumentistes). Un interprète peut jouer de différents instruments (*eg.*, voix et guitare), et faire partie de différents ensembles (*eg.*, Damon Albarn est chanteur et pianiste dans le groupe Blur, compositeur dans Gorillaz; Thom Yorke est chanteur, compositeur, guitariste et pianiste pour Radiohead, mais seulement chanteur pour UNKLE). De plus, un même artiste peut être conjointement interprète et auteur/compositeur (c'est de nouveau le cas de Damon Albarn, mais aussi de Jacques Brel, Cali, ...). Enfin, un artiste peut être conjointement impliqué dans des livres, des albums et des films (par exemple, Björk est avant tout connue pour sa carrière de musicienne, mais aussi pour son rôle principal dans le film *Dancer in the Dark*).

1.2.3 Récapitulatif des informations à modéliser

- Une œuvre a un titre, une époque, une date de création, une origine, un genre, un thème (le cas échéant), un auteur, une langue de composition, une inspiration (fait ou autre œuvre originale, le cas échéant). Pour un film (*resp.*, une œuvre musicale originale, l'interprétation d'une œuvre musicale originale), c'est le scénariste (*resp.*, le compositeur ou le groupe, le chef d'orchestre) qui est considéré comme étant l'auteur de l'œuvre. Pour les films, on considère également le réalisateur.
- Une œuvre interprétée a un interprète principal et des interprètes secondaires. L'interprète principal est l'orchestre, l'instrumentiste, le groupe, l'acteur en tête d'affiche (selon les cas).
- Les articles vendus, identifiés par une référence unique, sont les parutions, à une date et en une langue donnée (avec mention du traducteur pour les livres), d'œuvres.
- Les œuvres peuvent être primées (Goncourt, Prix d'interprétation au festival de Cannes, Académie Charles-Cros, ...); les articles peuvent être en promotion (à un moment donné); en édition particulière (collection).
- Un artiste a un nom, un prénom, une date de naissance, une date de décès (le cas échéant), une liste d'instruments dont il joue (le cas échéant), une liste de rôles qu'il a joués (le cas échéant).
- Un genre peut se décliner en sous-genres et sous-sous-genres; une origine (continent) se décline en sous-origines (pays).
- Un instrument a une catégorie (cordes, cuivre); une prestation musicale peut être acoustique ou non, enregistrée en studio ou en public, en audio mono, stereo, 5.1, ...
- Si le vendeur ne souhaite pas mémoriser les formations philharmoniques, il veut néanmoins mémoriser les orchestres (orchestre philharmonique de Vienne, orchestre philharmonique de Radio France); par ailleurs, les groupes (rock, pop, ...) sont, eux, mémorisés.
- Un film peut être un long-métrage ou un court-métrage, couleur ou noir et blanc, filmé ou non en numérique, en audio mono, stereo, 5.1, ..., en video 2D ou 3D; certains films ont une bande originale; les films d'animation ont des personnages qui ont un nom et une origine (*eg.*: Mickey – Walt Disney, le petit poucet – Charles Perrault, Apollon – mythologie, ...).

1.3 Mémorisation du comportement des internautes

Il ne s'agit évidemment pas de mémoriser les informations personnelles, mais contextuelles : lors d'une navigation sur le site du vendeur, un utilisateur a consulté $\frac{1}{2}$ des livres de Nick Hornby et des films avec Hugh Grant; un autre utilisateur, au cours d'une commande, a acheté des articles pour le moins hétéroclytes, mais tous en promotion ou en version collection. Les informations collectées alors permettent de parfaire la description des ouvrages par de nouvelles associations (émergence de profils utilisateur), créées a posteriori : dans les exemples donnés ci-dessus, il s'agirait des profils "*amateur de culture britannique contemporaine*" et "*promotion*", "*collection*".

Pour l'instant, il ne s'agit pas encore de générer ces profils (cette partie sera sous-traitée à un laboratoire universitaire), mais seulement de créer le modèle permettant de mémoriser le comportement, soit, une navigation : point d'entrée (moteur de recherche, site publicitaire), enchaînement des consultations (une page pouvant être consultée par clic sur un lien, ou par interrogation sur le site), articles correspondants. Pour les achats : liste des articles achetés au cours d'une même commande. Toutes ces informations sont datées.

1.3.1 Récapitulatif des informations à modéliser

- Une navigation est caractérisée par une date et une heure, une durée, une origine, un point d’entrée (première page visitée du site), un point de sortie (dernière page visitée du site), un enchaînement de pages consultées sur le site du vendeur. L’utilisateur est identifié par son adresse IP.
- Pour chaque page visitée au cours d’une même navigation, il faut connaître la date, heure et durée de la consultation, la page précédemment visitée et la page visitée à sa suite au cours de la même navigation (la page succédant la page de sortie étant extérieure au site du vendeur, elle est considérée comme nulle).
- L’origine d’une navigation est le lien qui a mené à la première page visitée sur le site du vendeur. Il peut s’agir par exemple d’une recherche google (www.google.com/search?q=radiohead), ou de la saisie directe de l’URL dans le navigateur (origine alors nulle).
- Une page (ou lien qui y a mené) est toujours une URL. Une page dynamique résulte d’une recherche sur le site du vendeur ; celui-ci permet de chercher un article par type (film, livre, disque), auteur, titre, prix, date de sortie, genre (on ne considère que les genres principaux, *ie.*, de premier niveau dans la hiérarchie des genres). Il faut alors mémoriser les paramètres de la requête.
- Un achat est caractérisé par une date, une liste d’articles, un montant, un code postal (adresse communiquée par l’utilisateur). Un achat est lié à une navigation.
- Dans le cadre d’un achat, le vendeur souhaite connaître quels sont les articles en promotion, les ouvrages primés, les articles en format collection, les disques de type compilation (ou “*best-of*”).

1.4 Exemples de requêtes que le système devra traiter

Combien d’utilisateurs ayant acheté, au cours d’une commande, *High fidelity* de Nick Hornby ont-ils également acheté le film *Quatre mariages et un enterrement* de Mike Newell ? Quelle proportion d’utilisateurs a-t-elle effectué des commandes composées d’au moins 60% d’articles en promotion ? Quels sont les articles disponibles à la vente dont les auteurs sont dramaturges américains ? Quels sont les groupes de rock indépendant britanniques achetés au moins une fois lors de l’achat d’un album de Radiohead ?

2 Travail demandé

2.1 Modélisation-Conception de la base

1. Modélisez l’énoncé ci-dessus sous forme d’un modèle entité-association. Consignes :
 - Explicitiez toute hypothèse que vous devez faire sur ce que le vendeur souhaite réellement modéliser et qui induit un choix de conception.
 - Justifiez/commentez chacun de vos choix de conception *eg.*, choix d’une cardinalité 0 ou 1 pour une association, choix de représentation d’un concept sous forme d’entité *vs.* d’association, choix de faire porter une information donnée sur un concept plutôt qu’un autre, contraintes d’intégrité, ...
 - Concevez votre modèle en ayant à l’esprit que celui-ci est amené à évoluer (d’autant s’agissant d’une nouvelle application !) : le modèle doit être le plus souple possible.
2. Déduisez du modèle E/A le modèle relationnel, en indiquant bien, pour chaque relation :
 - ses clés ;
 - ses contraintes ;De nouveau, pour la conversion, justifiez vos choix (plus particulièrement pour la conversion des associations).

2.2 Interrogation de la base

Un schéma relationnel corrigé de la base vous sera remis le vendredi 3 décembre en TP. La deuxième partie du devoir, qui vous sera donnée lors du TD-TP du 10 décembre, consistera à exprimer des requêtes en algèbre relationnelle, ainsi qu’en SQL (script “.sql” exécutable composé du schéma de la base et des requêtes demandées). La troisième partie de ce devoir, consacrée à PHP-MySQL, vous sera donnée lors du TP du 7 janvier.

2.3 Organisation

- le travail est à réaliser en **binôme** (éventuellement monôme) ;

- la première partie est à rendre le **3 décembre avant 14h**, par envoi électronique, à votre chargé de TD-TP (letocart@lipn.fr, borne@lipn.fr) et/ou sous forme papier lors du TP ou dans la boîte aux lettres Licence ;
- la seconde partie sera à rendre pour le **3 janvier**, par envoi électronique, à votre chargé de TD-TP (letocart@lipn.fr, borne@lipn.fr).
- la troisième partie sera à rendre pour le **14 janvier**, par envoi électronique, à votre chargé de TD-TP (letocart@lipn.fr, borne@lipn.fr).