

INFO-H-303 - Bases de données

Projet

Professeur : Esteban Zimanyi

Assistant : Michaël Waumans

<http://cs.ulb.ac.be/public/teaching/infoh303>

Année académique 2012-2013

Etude de cas

On vous demande de créer une application de consultation de références en utilisant les données fournies par *The DBLP Computer Science Bibliography*¹.

L'application stocke la description d'un ensemble d'articles, de livres, de thèses de master ou de doctorat. Pour chaque source, le système retiendra l'auteur, l'éditeur, le titre, l'année, le journal, l'url, l'isbn, le chapitre et le "publisher".

Le système acceptera deux types d'utilisateurs : les utilisateurs anonymes et les administrateurs. Les utilisateurs anonymes auront uniquement le droit de consultation. Les administrateurs auront en plus le droit d'insertion, de suppression et de modification. Les administrateurs auront un identifiant (leur adresse email) ainsi qu'un mot de passe haché (à l'aide par exemple de SHA-256) et pourront s'authentifier sur le système.

Tous les utilisateurs devront pouvoir consulter facilement toutes les informations présentes dans la base de données. Notamment, ils devront pouvoir :

- Consulter toutes les informations d'un article, d'un livre ou d'une thèse.
- Rechercher des articles, des livres ou des thèses sur base de leur nom, de leur année de publication, des auteurs, l'isbn,...
- Consulter toutes les informations d'un auteur ainsi que toutes les publications auxquelles il a contribué.

Format des données

Les fichiers que vous devez importer dans votre base de données relationnelle se trouvent à l'adresse <http://dblp.uni-trier.de/xml/>. Le fichier à employer est nommé `dblp.xml`. A vous de choisir quels outils vous utilisez pour extraire les données et les importer dans votre base de données.

Pour des raisons de performance, veillez à n'importer que les données concernant les publications comprises qu'entre 2008 et 2011 inclus.

On vous demande de créer un script d'insertion de ces données dans votre base de données.

1. <http://dblp.uni-trier.de/xml/>

Déroulement du projet

Première partie

Pour cette partie, on vous demande de modéliser le problème à l'aide du formalisme entité-association et de préciser les contraintes d'intégrité nécessaires. Ces contraintes doivent être exprimées en français et utiliser les mêmes noms d'entités, d'associations ou d'attributs que dans votre modèle conceptuel. Vous pouvez également exprimer et justifier des hypothèses sur votre modèle. Ces hypothèses peuvent résulter par exemple d'ambiguïtés dans l'énoncé du problème.

Déduisez ensuite de ce modèle conceptuel le modèle relationnel correspondant ainsi que ses contraintes. Vos choix de modélisation doivent être justifiés.

Ces modèles doivent être suffisamment riches pour pouvoir servir de support à l'application décrite ci-dessus ainsi que pour pouvoir répondre aux requêtes de la deuxième partie.

Deuxième partie

Création

On vous demande tout d'abord de déduire de votre modèle relationnel un script SQL DDL de création de la base de données et de ses différentes tables ainsi que de créer cette base de données.

Initialisation

On vous demande d'écrire un script permettant d'importer dans votre base de données les fichiers nécessaires. Ces données devront être présentes dans votre base de données lors de la défense.

Application

On vous demande ensuite de développer une interface graphique (web ou client lourd) pour votre base de données, permettant au minimum de réaliser les opérations ci-dessous :

- Consulter les informations des publications
- Consulter les auteurs et leurs publications
- Rechercher des publications sur base de leur nom
- Rechercher des publications sur base de critères avancés comme l'année de publication, l'isbn, l'éditeur, ...
- Afficher les résultats des requêtes demandées (voir plus loin)
- Connecter un administrateur. Les administrateurs pourront :
 - Enregistrer un administrateur (à l'aide d'un autre compte administrateur)
 - Ajouter, supprimer et modifier les données (publications, auteurs,...)

Votre application devra veiller à ce que la base de données reste cohérente en utilisant autant que possible le moteur de base de données (contraintes de clé étrangère, etc...)

Vous pouvez bien sûr ajouter des fonctionnalités à votre application comme par exemple :

- Voter sur une publication
- Trier les publications selon des critères (genre, année, éditeur,...)
- Afficher l'ensemble des co-auteurs d'un auteur
- Prendre en compte plus de données de DBLP
- etc...

Ces apports personnels seront valorisés à hauteur de 4 points sur 20.

Requêtes

Nous vous demandons d'écrire en algèbre relationnelle et calcul relationnel tuple les requêtes R1,R2,R3 et R4 ainsi que toutes les requêtes en SQL. Si une requête vous semble imprécise, indiquez dans votre rapport les hypothèses que vous avez faites pour lever ces imprécisions. Attention, pour les requêtes en algèbre relationnelle et en calcul relationnel, vous ne pouvez pas utiliser de fonction d'agrégation.

- R1 : Tous les auteurs qui ont chacun publié au moins une fois chaque année entre 2008 et 2010 inclus
- R2 : Les auteurs qui ont écrit au moins deux articles pendant la même année
- R3 : Les auteurs Y qui sont à une distance 2 d'un auteur X. Un auteur Y est à une distance 1 d'un auteur X si ces deux auteurs ont co-écrit un article.
- R4 : Les articles qui n'ont aucun docteur parmi leurs auteurs, où un docteur est défini comme un auteur pour lequel la base de données contient une thèse de doctorat dont il est le seul auteur
- R5 : Les auteurs ayant publié dans le plus de journaux différents
- R6 : Les journaux avec leur nombre total d'articles, le nombre d'articles moyen par an et le nombre d'auteurs moyen par article depuis leur année de création et ce pour tous les journaux dont le nombre de volumes par an est supérieur au nombre de volumes moyen par an des journaux.

Rapports

Pour la première partie, on vous demande les documents suivants :

- Un diagramme entité-association modélisant le projet ainsi que ses contraintes
- Une traduction relationnelle de ce diagramme et ses contraintes
- Vos hypothèses et la justification de vos choix de modélisation

Le formalisme utilisé doit être un de ceux vus au cours ou aux TPs.

Pour la deuxième partie, vous rendrez un rapport contenant :

- Les documents de la première partie tenant compte des remarques faites
- Une explication de votre méthode d'extraction des données
- Les requêtes demandées
- Les explications et justifications de vos choix et hypothèses

Le code source de votre script d'insertion et de votre application devra être déposé sur le *RhodeCode* pour le 16 mai à 23 heures 59.

Informations pratiques

- Le projet se fera obligatoirement par groupe de deux.
- La première partie, en version papier, devra être rendue pour le 15 avril.
- Le rapport de la seconde partie devra être rendu en version papier lors de votre défense. Votre code source devra être déposé sur le *RhodeCode* pour le 16 mai à 23 heures 59. La défense du projet aura lieu le lendemain durant les plages horaires attribuées au cours selon un ordre à déterminer. On vous demandera de présenter votre application pendant 10 minutes et de répondre à quelques questions.
- Le projet comptera pour 25 % de la note finale du cours.
- Les apports personnels au projet seront valorisés à hauteur de 4 points sur 20. En d'autres mots, un projet parfait sans apport personnel aura une valeur de 16 points sur 20.
- Sauf mention explicite, vous pouvez utiliser les langages et outils de votre choix (MySQL, PostgreSQL, SQLite, Python, PHP, Java,...) Toutefois, les requêtes SQL doivent être réalisées explicitement dans votre programme.

- Vous pouvez développer sur votre propre machine et présenter vos projets sur un ordinateur portable que vous apporterez lors de la défense.