

Données du Web : XML Relationnel

TD en binômes à rendre le 25 octobre.

o) Warm-up, Fonctionnalités Oracle pour XML : Stockage CLOB et Binary XML, Interrogation

1. Récupérer la DTD ainsi que le document XML que vous avez proposés pour la gestion des Tweets.
2. Connectez vous à ORACLE avec le commande `rlwrap sqlplus login/mot_de_passe@venus/master`
3. Créer des tables pour enregistrer vos documents XML en utilisant le différentes options de stockage.

```
CREATE TABLE nom_de_la_table_CLOB (nom_colonne_texte varchar(20), nom_colonne_xml XMLTYPE)
XMLTYPE nom_colonne_xml STORE AS CLOB;
CREATE TABLE nom_de_la_table_binaryxml (nom_colonne_texte varchar(20), nom_colonne_xml XMLTYPE)
XMLTYPE nom_colonne_xml STORE AS BINARY XML;
```
4. Enregistrez votre document XML pour les Tweets dans la base avec l'instruction

```
INSERT INTO nom_de_la_table (nom_colonne_texte,nom_colonne_xml)
VALUES ('test', sys.xmltype.createxml('<mettreXMLici/>')) );
```
5. Exécuter 3 requêtes XPath et 3 requêtes XQuery parmi celles du TD précédent, avec les commandes :

```
SELECT EXTRACT(nom_colonne_xml,'/mettre/ici/la/requête/XPath') FROM nom_de_la_table;
SELECT XMLQUERY('/mettre/ici/la/requête/XQuery' PASSING nom_colonne_xml RETURNING CONTENT)
FROM nom_de_la_table;
```

1) Stockage Monet

1. Implémenter en SQL le stockage Monet vu en cours.
2. Peupler les tables avec des valeurs correspondants à votre document XML pour la presse.
3. Exprimez en SQL trois requêtes XPath de votre choix.

2) Stockage schema-aware

1. À partir de la DTD proposée pour les batiments, définir un schéma de stockage relationnel.
2. Remplissez les tables avec des valeurs correspondants à votre document XML.
3. Traduisez en SQL trois requêtes à choix.

3) Interval-encoding avec SAX

1. Donnez l'encodage begin/end de l'XML pour les bâtiments, puis enregistrez-le dans la table NODE.
2. À l'aide de la classe `SaxParser` (.java sur Moodle), programmez l'encodage par intervalles begin/end. Cette classe permet de parser un document XML en streaming (noeud après noeud). Modifiez les méthodes `startElement` et `endElement` qui sont déclenchées à l'ouverture et fermeture d'un noeud de type `element`, ainsi que la méthode `characters` qui est déclenchée lorsque on rencontre un noeud de texte. L'output du programme sera une liste de mises à jour pour la table NODE.

```
INSERT INTO NODE (begin,end,parent,tag,nodtyp) VALUES(1,16,null,'racine','element')
```
3. Testez le programme sur le document XML concernant les Tweets, et reformulez sur la table NODE les requêtes que vous avez proposé dans l'exercice 1.
4. Testez le programme avec le fichier XML <http://www.ins.cwi.nl/projects/xmark/Assets/standard.gz> et reportez le temps d'exécution.

(extra) Modifier le programme afin d'implémenter l'encodage pre/post du document.

(extra-extra) Modifier le programme afin d'implémenter l'encodage Dewey du document.