

## Stockage, compression, archivage

### 1. HTML et son évolution

Il semblerait qu'il y ait aujourd'hui deux évolutions parallèles et non forcément concurrentes du langage de description de pages nommé communément HTML, à savoir : d'une part, le passage de HTML 4.01 à HTML 5 et, d'autre part, le passage de XHTML 1.0/XHTML 1.1 à XHTML 2.

Donner des arguments justifiant qu'il faut privilégier la seconde transformation (le passage de XHTML 1.0/XHTML 1.1 à XHTML 2) au détriment de la première. Votre rédaction devra comporter au minimum 10 lignes avec au moins 5 mots de 4 syllables ou plus. Vos arguments devront refléter votre culture du Web avec ses langages, ses applications et ses contraintes actuelles. On ne se limitera pas uniquement à des exemples pour le B2C (*Business to Consumer*).

### 2. Compression "conceptuelle" de données binaires

On s'intéresse à des fichiers ne contenant que les caractères "0", "1" et le caractère "espace". On pourra imaginer que les lignes d'un tel fichier correspondent à un individu (au sens statistique du terme) et que les valeurs 0 et 1 d'une même ligne, séparées par un ou plusieurs espaces, correspondent aux réponses de cet individu à un ensemble de questions (par exemple "0" signifie *Oui* et "1" signifie *Non*). On trouvera sur la page suivante un exemple de tel fichier. La ligne 1 et la colonne 1 n'en font pas partie mais servent seulement ici à repérer les lignes et les colonnes.

Lig01	Col01	Col02	Col03	Col04	Col05	Col06
Lig02	1	0	0	1	0	0
Lig03	1	0	1	0	1	0
Lig04	1	0	1	0	1	0
Lig05	1	0	0	0	1	1
Lig06	1	1	1	0	0	0
Lig07	1	0	0	1	1	0
Lig08	1	0	0	1	1	0
Lig09	1	0	0	1	1	0
Lig10	1	0	0	1	1	0
Lig11	1	0	0	1	1	0
Lig12	1	0	0	1	1	0
Lig13	1	0	0	1	1	0
Lig14	1	0	0	1	1	0
Lig15	1	0	0	1	1	0

- Sachant qu’il y a 42 caractères dans la ligne 1 du ”L” de Lig01 au ”6” de Col06, combien y a-t-il de caractères en tout dans le fichier d’exemple ?
- On décide tout d’abord de représenter chaque ligne ou chaque colonne par le codage suivant, facilement lisible par un être humain, nommé **codage 1** : chaque suite de caractères ”0” ou ”1” égaux est codée par le caractère commun, suivi du caractère espace, de son nombre d’occurrences et d’encore un autre caractère espace. Donner le codage de la ligne 1 et de la colonne 6.
- On décide ensuite de représenter chaque ligne ou chaque colonne par le codage suivant, facilement lisible par un être humain, nommé **codage 2** : chaque suite de caractères ”0” ou ”1” égaux est codée par un même caractère (”Z” pour ”0” et ”U” pour ”Un”), auquel on concatène le nombre d’occurrences de ce caractère (et donc sans aucun caractère espace). Donner là encore le codage de la ligne 1 et de la colonne 6.
- Ces deux codages sont-ils vraiment des compressions du fichier ? Peut-on calculer à partir de combien de lignes et de colonnes ces codages compressent à plus de 80 % le fichier ?
- Comparez les avantages et les inconvénients de deux codages.
- Si on devait programmer ces codages à la fois en Perl et en Php, lequel serait le plus facile à programmer ?

### 3. Scriptez-le !

Traduire l'extrait de pseudo-algorithme suivant en *Rexx* et en *Perl* :

```
LIRE LA LIGNE lig SUR LE FICHER fic
# on calcule le nombre de mots
AFFECTER NBMOTS(lig) A nbm
# on passe chaque mot en revue à l'aide
# de son indice de mot (idm)
POUR idm DE1A nbm
  AFFECTER MOT(lig,idm) A mc
  # pour le codage 1, on additionne les...
  SI mc EST UN ENTIER
    ALORS AFFECTER mc + nbCar A nbCar
  FINSI
FINPOUR # idm
```

### 4. Un peu de DTD et de XSL

Dans le fichier <http://www.info.univ-angers.fr/~gh/tuteurs/titres1.xml> l'élément *racine* FILMS contient des éléments nommés FILM et des éléments nommés ARTISTE. Voici un exemple d'élément FILM et un exemple d'élément ARTISTE :

```
<FILM Annee="1997">
  <TITRE>Le cinquième élément</TITRE>
  <GENRE>Science-fiction</GENRE><PAYS>FR</PAYS><MES idref="111"/>
  <ROLES><ROLE><PRENOM>Bruce</PRENOM><NOM>Willis</NOM><INTITULE>Korben Dallas
    </INTITULE></ROLE><ROLE><PRENOM>Milla</PRENOM><NOM>Jovovich</NOM>
    <INTITULE>Leeloo</INTITULE></ROLE>...
  </ROLES>
  <RESUME>
    Au XXIIIème siècle, dans un univers étrange et coloré...
  </RESUME>
</FILM>

<ARTISTE id="3">
  <ACTNOM>Hitchcock</ACTNOM><ACTPNOM>Alfred</ACTPNOM>
  <ANNEENAISS>1899</ANNEENAISS>
</ARTISTE>
```

- Donner l'extrait de grammaire DTD permettant de définir l'élément **ARTISTE** avec son attribut et ses sous-éléments, attribut et sous-éléments qu'on considérera comme obligatoires.
- On voudrait remplir une base de données *MySql* avec les données de `titres1.xml` et on supposera qu'on a déjà une table **ARTISTE** et une table **FILM** avec les « bons » champs. Donner une transformation XSL permettant de fournir à partir de `titres1.xml` un fichier qui contient des instructions *MySql* réalisant l'insertion des données. A défaut de la syntaxe XSL exacte, vous pouvez donner les grandes lignes de la transformation, les "TEMPLATE" à repérer *etc.*

## 5. Questions au choix

Vous traiterez deux questions au choix parmi toutes celles présentées. Vos réponses devront mettre en évidence votre culture et votre recul en matière de modélisation et de traitement de l'information. Chaque réponse devra faire 5 lignes au minimum, sans limite de maximum.

1. *Ce qu'on nomme ORM (mapping objet relationnel) est-il aujourd'hui la seule bonne voie pour la programmation évoluée des données persistantes ?*
2. *Faut-il comparer LINQ to XML et XQUERY ?*
3. *La compression vidéo va-t-elle encore progresser de façon significative dans les 10 ans à venir ?*
4. *L'externalisation (des .js, des .css etc.) est-elle vraiment un agent fort dans la progression vers des pages plus "sémantiques" ?*
5. *Les logiciels libres sont-ils un choix raisonnable aujourd'hui (2009) ?*
6. *SVG est-il déjà un standard pour l'entreprise ? Si non, va-t-il le devenir dans les 10 prochaines années ?*