

INFO-H-509 : Technologies XML

Projet : XSLT

Introduction

DBLP¹ est une base de données de publications en sciences informatiques contenant près de 1,6 million de références. Son contenu est disponible de manière publique au format XML². Un extrait de celui-ci sera utilisé pour ce projet (voir la page web du cours).

Le document `dblp.xml` décrit une collection d'enregistrements DBLP :

```
<?xml version="1.0" encoding="ISO-8859-1">
<!DOCTYPE dblp SYSTEM "dblp.dtd">
<dblp>
  enregistrement 1
  enregistrement 2
  ...
  enregistrement n
</dblp>
```

Chaque enregistrement appartenant à la collection décrit la référence bibliographique d'une publication. Ces publications peuvent être de 8 types : `article`, `inproceedings`, `proceedings`, `book`, `incollection`, `phdthesis`, `mastersthesis`, et `www`, respectant le format BibTeX. Suivant le type de publication, un enregistrement peut contenir les champs suivants : `author`, `editor`, `title`, `booktitle`, `pages`, `year`, `address`, `journal`, `volume`, `number`, `month`, `url`, `ee`, `cdrom`, `cite`, `publisher`, `note`, `crossref`, `isbn`, `series`, `school`, et `chapter`. Notez que tous les champs ne sont pas permis dans chaque type de publication ; référez vous au DBLP DTD donné dans le fichier `dblp.dtd` ainsi qu'au document <http://dblp.uni-trier.de/xml/docu/dblpxml.pdf> pour plus d'informations sur les types de publications et leurs champs associés.

Projet

L'objectif de ce projet est d'écrire un document au format XSLT 2.0 qui génère, à partir du document `dblp-excerpt.xml`, un ensemble de fichiers HTML produisant une partie du site DBLP³.

Concrètement, pour chaque personne *P* se trouvant dans les champs `author` ou `editor`, d'un enregistrement du fichier `dblp-excerpt.xml`, le document XSLT devra générer un fichier dont le nom respecte le format suivant :

`a-tree/first-letter-of-lastname/lastname.firstname.html`

Par exemple, pour l'auteur "David Maier", il faudra créer le fichier `a-tree/m/Maier.David.html`, alors que pour "Michael Ley", il faudra créer le fichier `a-tree/l/Ley.Michael.html`. Les caractères d'espace devront être remplacés par des underscores ; tout autre caractère non-alphanumérique devra être remplacé par le caractère '='. Ceci permet d'éviter des URLs invalides.

Le contenu du fichier HTML de la personne *P* devra contenir les parties suivantes. Référez vous au fichier `Maier.David.html` pour un exemple illustratif.

1. <http://www.informatik.uni-trier.de/~ley/db/>
2. <http://dblp.uni-trier.de/xml/>
3. Utilisez la commande `xsl:result-document` afin de générer plusieurs fichiers.

1. Le nom de la personne dans un tag `h1`.
2. Une table de toutes les publications de cette personne, regroupées par année (triées de manière descendante par rapport à l'année et pour chaque groupe par rapport au titre). Outre la ligne indiquant l'année, chaque ligne de cette table devra être sous le format suivant :

publication number, link to online version, publication reference

Référez vous au document `Maier.David.html` pour plus de détails (quelques explications y sont incluses sous la forme de commentaires du document HTML).

3. Le titre "Co-author index" dans un tag `h2`.
4. Une table listant toutes les autres personnes avec qui *P* a écrit une publication. Une ligne de cette table contenant la paire (*co-author-name*, *list-of-references-to-joint-publications*) sera consacrée à chacun de ces co-auteurs. Chaque référence dans cette liste devra renvoyer vers la publication correspondante dans la table des publications. La table devra être triée selon le nom de famille du co-auteur. (Référez vous au fichier HTML nommé `Maier.David.html` pour plus de détails.)

Modalités

Concevez un seul fichier au format XSLT 2.0 nommé `generate-author-pages.xslt` permettant de générer les documents HTML. L'ensemble des fichiers (`dblp-excerpt.xml`, `dblp.dtd`, et `dblp.xml.pdf`) nécessaire à la réalisation du projet sont disponibles sur la page web du cours.

Tout comme lors du premier projet, celui-ci contribuera à 2 points sur 20 de la note finale (il y aura encore un projet contribuant à 2 points sur 20), l'examen écrit vaudra par conséquent 14 points sur 20.

Ce projet *doit* être réalisé par groupe de deux personnes. Nous vous demandons d'envoyer les noms des membres de votre groupe à M. Julien Roland (`juroland@ulb.ac.be`) au plus tard le 15 Avril 2011. Si vous n'arrivez pas à constituer un groupe, envoyez un email à M. Julien Roland.

Nous vous demandons de rendre un rapport contenant l'ensemble de vos hypothèses, ainsi que le document XSLT. Le rapport doit être rendu à M. Julien Roland sous format papier (au 2N3.211C ou dans son casier). De plus celui-ci sera envoyé avec le document XSLT par courrier électronique à l'adresse `juroland@ulb.ac.be` **au plus tard le 4 Mai 2011 à midi**.