
TD-6 : Relax-NG avancé

Benoît Valiron

1. Contenu mixte

(Reprise du TD 5)

Le format XML pour une biographie est le suivant : la racine est `bio` ; ses fils sont dans l'ordre un élément `nom` et du texte. Le texte peut contenir des morceaux en italique et en gras, encadrés par des balises `i` et `b`.

Écrivez une feuille de validation Relax-NG pour ce format (version XML et version compacte).

Modifiez vos schémas pour donner comme espace de noms `bio:court` aux éléments `bio` et `nom` et l'espace de noms `http://www.w3.org/1999/xhtml` aux éléments `b` et `i`.

Donnez un titre et une description à vos schémas en utilisant le format Dublin core (<http://purl.org/dc/1.1/elements/>). Documentez aussi les éléments `bio` et `nom` avec la balise `description` de Dublin core.

Quelle serait une DTD correspondante ? Quels problèmes rencontre-t-on ?

2. Sitemaps

(Reprise du TD 4)

Le moteur de recherche google propose un service aux webmasters pour l'indexation : les sitemaps. Il s'agit d'un format XML pour donner des informations succinctes au robot d'indexation sur les pages qu'il peut rencontrer. Le format est le suivant (pris sur wikipedia) :

- Élément `<urlset>`. Obligatoire. Racine du document.
- Élément `<url>`. Obligatoire. Élément parent pour chaque entrée. Les éléments restant sont tous fils de cet élément.
- Élément `<loc>`. Obligatoire. Contient l'adresse internet d'une page, incluant le protocole (`http://` ou `https://`). Doit faire au maximum 2048 caractères de long.
- Élément `<lastmod>`. Facultatif. La date de dernière modification du fichier, en format ISO ou plus simplement YYYY-MM-DD.
- Élément `<changefreq>`. Facultatif. Fréquence à laquelle la page est modifiée en général : `always`, `hourly`, `daily`, `weekly`, `monthly`, `yearly`, `never`.
- Élément `<priority>`. Facultatif. L'importance relative de cette page par rapport aux autres. Valeur entre 0.0 et 1.0, valeur par défaut de 0.5.

Est-il possible d'écrire une DTD qui prenne en compte toutes les contraintes ?

Écrivez une feuille de validation Relax-NG qui réponde aussi précisément que possible à la spécification.

3. Document composite

(Reprise du TD 2)

On reprend ici le flux RSS de thèses en ligne. Un format DTD (très) simplifié est le suivant :

```
<!ELEMENT rss (channel)>
```

```

<!ATTLIST rss xmlns:dc CDATA #FIXED
                "http://purl.org/dc/elements/1.1/"
                version CDATA #REQUIRED>
<!ELEMENT channel (title,link,description,item*)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT link (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT item (link|pubDate|title|dc:creator|dc:description|
                dc:title|description)*>
<!ELEMENT dc:creator (#PCDATA)>
<!ELEMENT dc:description (#PCDATA)>
<!ELEMENT dc:title (#PCDATA)>
<!ELEMENT pubDate (#PCDATA)>

```

Écrivez un schéma Relax-NG en syntaxe simplifiée et en syntaxe XML qui corresponde à cette DTD. (Attention aux espaces de noms!)

Modifiez les schémas pour forcer l'élément pubDate à avoir comme type date, l'élément link le type anyURI et l'attribut version le type décimal et comme valeur soit 1.0, soit 2.0.

4. Encore des espaces de noms.

(Reprise du TD 2)

Donnez un schéma Relax-NG pour le document suivant. Utilisez les type XML-Schema pour être aussi précis que possible.

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:pcv="http://prismstandard.org/namespaces/pcv/1.0/"
         xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description
    rdf:about="http://wanderlust.com/2000/08/Corfu.jpg">
    <dc:identifiant
      rdf:resource="http://wanderlust.com/content/2357845" />
    <dc:creator>
      <pcv:Descriptor rdf:about="http://wanderlust.com/emp3845">
        <pcv:label>John Peterson </pcv:label>
      </pcv:Descriptor>
    </dc:creator>
    <dc:coverage>
      <pcv:Descriptor
        rdf:about="http://prismstandard.org/vocabs/ISO-3166/GR">
        <pcv:label xml:lang="en">Greece</pcv:label>
        <pcv:label xml:lang="fr">Grece</pcv:label>
      </pcv:Descriptor>
    </dc:coverage>
  </rdf:Description>
</rdf:RDF>

```