

Documents structurés – TALA540a

Projet fin de semestre

Informations générales

Date de rendu : **01 juin 2021** jusqu'à 23 h 59.

Dépôt : sur votre Github, dans un nouveau projet appelé « projet-final ».

Présentation

L'objectif de ce projet est de réaliser une application web pour présenter des documents qui ont été traduits et alignés.

Deux grandes étapes seront mises en œuvre.

La première étape consistera à transformer des documents HTML vers le format XML TMX¹. Vous aurez accès à une DTD permettant de valider la structure de votre document XML.

Une fois cette étape terminée, les documents seront stockés dans une base de données XML (eXist-db). À partir de celle-ci, vous allez devoir réaliser un site internet à l'aide de xquery et xslt. Sur ce site seront présentées trois interfaces : une page pour présenter le projet, une autre pour afficher vos documents alignés et finalement une pour la recherche d'informations dans les documents.

Explications

Architecture

Un dossier nommé « projet-final » qui contient deux sous-dossiers :

- « partie-1 » : pour la transformation HTML vers TMX
- « partie-2 » : pour votre application (Exide > Application > Download App)

Étape 1 – 10 points

À partir d'un corpus en HTML, vous allez devoir transformer les documents en XML TMX. Chaque document est un extrait de la bible, soit en français, soit en anglais. La correspondance se fait grâce aux noms des documents. Ainsi, « bible_eng-1.html » va de pair avec « bible_fre-1.html ».

La DTD fournie est là pour vous aider à construire le document attendu en sortie (voir le cours 02-DTD).

Les transformations sont à réaliser grâce à une feuille (ou plusieurs) de style XSL. Il est nécessaire d'utiliser XSLT en version 3.0. Vous pouvez vous servir d'Oxygen XML Editor et de toutes les ressources en ligne.

De plus, afin de préparer le terrain pour notre application, vous baliserez les mots suivants dans un élément « hi ». Chaque pair de mots se verra attribué un identifiant similaire ce qui facilitera la création de notre interface de recherche.

¹ https://en.wikipedia.org/wiki/Translation_Memory_eXchange et <https://www.gala-global.org/tmx-14b>

- god / dieu
- spirit / esprit
- day / jour
- night / nuit
- lord / éternel
- blood / sang

Voir : <https://www.saxonica.com/documentation10/index.html#!xsl-elements/analyze-string> et <https://www.saxonica.com/documentation10/index.html#!functions/fn/document>

Étape 2 – 10 points

Les documents XML générés vont être manuellement ajoutés dans eXist-db, dans un dossier « data ». Exist-db permet de créer des sites à l'aide de xquery. L'objectif de cette étape est de réaliser un site pour présenter vos documents. Il y aura trois interfaces à réaliser.

Index.html

La première interface sera la page d'accueil de votre site. Votre contrainte pour cette page est de la construire à partir d'un document XML qui sera transformé par XSLT pour un affichage HTML. Vous présenterez sur cette page :

- le projet et les étapes de mise en œuvre
- la navigation sur le site et ce qu'on peut y trouver
- les difficultés rencontrées : expliquer en détail et les solutions trouvées

Ce document XML sera stocké dans « data » également.

Voir : <http://exist-db.org/exist/apps/fundocs/view.html?uri=http://exist-db.org/xquery/transform&location=java:org.exist.xquery.functions.transform.TransformModule&de tails=true> et la fonction « transform:transform ».

search.html

La seconde interface sera une page où l'on pourra effectuer de la recherche plein texte sur vos documents. Par exemple, pour un mot donné, retourner le nom du document dans lequel il apparaît. Vous êtes libre de proposer d'autres choses.

Néanmoins, il faudra absolument réaliser au moins une interface où les résultats sont montrés comme sur Linguee (<https://www.linguee.fr/francais-anglais/search?source=auto&query=c%C5%93ur>). C'est à ce moment qu'on pourra se servir des éléments « hi ».

Cette partie sera réalisée en xquery.

Display.html

Finalement, la troisième interface sera pour vous l'occasion de proposer un rendu graphique de vos documents XML. Le but est de proposer un affichage pour refléter l'alignement de nos segments. On peut se baser sur le site original² pour se donner des idées. Pour cela, il faudra employer un peu de CSS. Vous pouvez utiliser à la fois xslt et/ou xquery pour le reste.

² <https://www.biblica.com/bible/niv/genesis/1/bds/>