Safir Medjeber
Marc Gedik
Master 1

Rapport du projet XML

Usage

Pour tester le projet il faut d'abord compiler le parseur en suivant les étapes ci-dessous :

- se rendre dans le dossier « ./src »,
- lancer le makefile avec la commande suivante : -> make
- supprimer les fichiers inutiles générés à la compilation : -> make clear

Pour convertir l'ensemble des fichiers Gedcom en documents XML, il faut procéder ainsi :

- se rendre dans le dossier racine « . »
- lancer le script gedToXml à l'aide de la commande suivante : -> ./gedToXml

Les documents XML correspondants à tous les fichiers Gedcom présents dans le dossiers « ./ged » sont alors automatiquement générées dans le dossier « ./xml »
Mais pour convertir les fichiers un par un, suivez ces étapes :

- se rendre dans le dossier racine « . »
- -> ./src/convert file.ged

Ensuite un moyen rapide de tester si les fichiers XML valident la DTD ainsi que le schéma est également à votre disposition en procédant ainsi :

- se rendre dans le dossier racine « . »
- lancer le script xmlValid à l'aide de la commande suivante : -> ./xmlValid

Mais pour tester les fichiers un par un, suivez ces étapes :

- se rendre dans le dossier racine « . »
- -> xmllint --noout --dtdvalid gedcom.dtd xml/file.xml (teste si le fichier est validé par la DTD)
- -> xmllint --noout --schema gedcom.xsd xml/file.xml (teste si le fichier est validé par le schéma)

Enfin pour convertir l'ensemble des fichiers Gedcom en documents XML, il faut procéder ainsi :

- se rendre dans le dossier racine « . »
- lancer le script xmlToHtml à l'aide de la commande suivante : -> ./xmToHtml

Les fichiers HTML correspondants à tous les fichiers XML présents dans le dossier « ./xml » sont alors automatiquement générées dans le dossier « ./html »
Mais pour tester les fichiers un par un, suivez ces étapes :

- se rendre dans le dossier racine « . »
- -> ./xsltproc xml2html.xsl xml/file.xml > file.html

Traduction des fichiers Gedcom en document XML

Pour convertir les fichiers Gedcom en document XML, nous avons choisi le langage *OCAML*. Nous avons de ce fait utiliser l'analyseur lexical *ocamllex* pour traduire un fichier Gedcom en une liste de lexème. Cette liste contient l'ensemble des niveaux, identificateurs, tags et valeurs présents dans le fichier Gedcom fourni.

Nous traitons une première fois ces données en les organisant selon le niveau. On obtient ainsi une liste de chaque ligne au sens de Gedcom.

Nous effectuons alors un second passage qui va réorganiser ces données sous la forme d'un arbre qui a pour racine une balise <root>. Lors de la construction de cet arbre, nous procédons à une insertion des nœuds, triées selon l'ordre lexicographique. Au finale, le parcours préfix de l'arbre permet la génération d'un fichier XML dont les balises sont ordonnées. (N.B : les fils directs de la balises <root> n'ont pas été triés car l'ordre lexicographique ne s'y prêtais pas. En effet la balise <header> se doit d'être en première position avant les balises <indi> ou <fam> par exemple.)

Choix effectués dans le projet

Nous avons donc fait le choix de trier les balises durant la conversion des fichiers. Ainsi quel que soit le fichier Gedcom fourni, une fois converti, la structure du fichier XML sera toujours ordonné selon l'ordre lexicographique.

Ce choix est motivé avant tout afin de simplifier l'élaboration de la DTD et la rendre plus précise. En effet, il constitue le meilleur compromis entre une DTD simple mais très permissive (par exemple : plusieurs femmes ou maris dans une même famille) et une DTD plus précise mais beaucoup plus complexe (en faisant toutes les permutations)

En outre nous supportons tous les tags indiqués dans le sujet ainsi que ceux présents dans les fichiers Gedcom fournis. Mais nous avons également ajouté des balises <lname> et <fname> qui permettent au sein d'une balise <name> de distinguer le nom de famille du prénom et rend ainsi plus précis le fichier XML généré par la suite.

Enfin de limiter le nombre de possibilités concernant le remplissage de la balise <date>, nous avons décidé de *parser* son contenu à l'aide d'une expression régulières.

Elaboration de la DTD

Le fait d'ordonner les balises pendant la génération du fichier XML permet durant l'élaboration de la DTD de traiter toutes les balises du fichier en les séquençant. Ainsi la gestion du nombre de répétition d'une balise donnée est beaucoup plus fine. En effet dés lors que les balises ne sont plus séquencés, les outils fournit par la DTD ne permettent plus la gestion précise du nombre d'occurrence d'une balise.

Elaboration du schéma

Concernant l'élaboration du schéma, on a veillé à être le plus précis possible en utilisant les restrictions ainsi que les expressions régulières.

Traduction des documents XML en documents HTML

Nous avons structuré les documents html sous forme de deux tableaux et un index alphabétique :

- Le premier tableau contient l'ensemble des individus triés selon l'ordre alphabétique. Pour un individu données, nous avons les liens vers ses familles (fams / famc) et des informations complémentaires accessible par le bouton more.
- Le second tableau contient l'ensemble des familles avec un lien vers chacun de ses individus.
- L'index alphabétique est construit à partir de la première lettre des prénoms de l'ensemble des individus.