

# ProjetUMINM141 d'analyse syntaxique : analyseur XML

Michel Meynard

10 octobre 2006

## 1 Introduction

XML (eXtensible Markup Language) est un dialecte de SGML défini par le World Wide Web Consortium (W3C). SGML (Standard Generalized Markup Language) est le standard international de l'ISO définissant la structure de différents types de documents électroniques. HTML (HyperText Markup Language) est le langage du Web.

SGML est très complexe et admet de nombreuses possibilités et options dans la définition de la syntaxe des documents structurés (manuscrits irlandais, documentation technique, partitions musicales, ...). Au contraire, HTML est un langage simple définissant une classe de documents publiables sur le Web. XML est une version simplifiée de SGML permettant de :

- définir facilement un type de document par une DTD (Document Type Definition) ou un Schéma XML,
- d'écrire rapidement des programmes :
  - vérifiant la syntaxe d'une instance de document par rapport à son type;
  - traduisant un document dans différents formats (HTML, PDF, RTF, ...).

## 2 Le projet

Le projet consiste à écrire un validateur XML, c'est à dire un outil vérifiant la syntaxe d'un document XML simple (sans DTD). Ce projet peut être implémenté :

- de différentes façons (analyse descendante ou ascendante);
- en utilisant ou pas différents outils (flex, bison);
- dans un langage ou un autre (C ou C++);

### 2.1 Spécifications fonctionnelles du projet A RESPECTER

Le validateur XML devra s'appeler `valxml` et devra pouvoir être utilisé à la ligne de commande de la façon suivante : `valxml monfic.xml`. Si le fichier est correct, le message suivant devra être affiché : `fichier XML valide`. Sinon, il devra afficher les champs suivants séparés par des “ : ” :

- le nom du fichier erroné (`monfic.xml`);
- le numéro de la ligne du fichier où l'erreur syntaxique a été détectée;
- le mot clé “ error ”;
- le type d'erreur;

Cela pourrait donner par exemple, l'affichage suivant : `monfic.xml :17 : error : erreur de syntaxe dans l'élément <lieu>`.

### 2.2 Extension du projet

Ceux qui seront intéressés pourront étendre le projet à la validation de fichiers XML avec DTD.

### 2.3 Documentation à rendre

Le projet sera effectué en binôme et donnera lieu à l'écriture d'un mini-rapport (20 pages) contenant notamment :

- vos choix d'implémentation;
- éventuellement un diagramme de classes (UML);
- l'API des fonctions ou des classes produites;
- une conclusion sur les extensions futures;

Une démonstration de 5 minutes aura lieu en salle machine avec les fichiers XML de mon choix! Tout validateur ne répondant pas aux spécifications ci-dessus obtiendra une note inférieure à la moyenne.