Proposition de stage de DEA : manipulation sûre de données XML

Encadrants :

Roberto Di Cosmo Université de Paris 7 et INRIA Rocquencourt
`http://www.dicosmo.org`, E-mail : `roberto@dicosmo.org`
Tel : 01 44 27 86 55.

Michel Mauny INRIA Rocquencourt
`http://pauillac.inria.fr/~mauny`, E-mail : `Michel.Mauny@inria.fr`
Tel : 01 39 63 53 30.

Jerôme Vouillon Université de Paris 7
`http://www.pps.jussieu.fr/~vouillon`, E-mail : `jerome@vouillon.org`
Tel : 01 44 27 61 99.

Laboratoires d'accueil : PPS (Université de Paris 7) et/ou INRIA Rocquencourt.

Cadre de la recherche : Ces dernières années ont vu un essor spectaculaire des technologies liées au Web. Nous nous intéressons ici en particulier au formalisme XML, qui est désormais largement préconisé pour l'échange de contenu. Pour manipuler des données représentées en XML et dont la structure est définie par une DTD, on trouve une grande variété de propositions, mais nous considérons que le bagage théorique de notre communauté peut apporter un plus significatif dans ce domaine trop souvent colonisé par des produits commerciaux développés sans bases théoriques. Un bel exemple de réussite dans cette direction est celui du langage XDuce [1] (`http://xduce.sourceforge.net/`), qui vérifie par le typage [2] la bonne formation des données XML traitées dans un programme (en éliminant ainsi les contrôles en début et fin de traitement traditionnellement présents dans les systèmes largement utilisés aujourd'hui), et qui a déjà eu une influence sur les normes du Web (notamment RELAX NG et XML Query).
Cependant, XDuce est un langage expérimental, loin de langages de programmation complets comme Java, Python ou C++, et il serait une erreur que de vouloir se lancer dans une course poursuite en espérant rivaliser avec ces langages dans tous les domaines.
Mais il est par contre tout à fait intéressant d'étudier la possibilité de marier XDuce avec un langage déjà largément utilisé, comme par exemple Ocaml ou Java. Il y a aussi un domaine dans lequel XDuce peut déjà presque rivaliser avec l'existant : les pages Web dynamiques (actuellement souvent écrites en PHP).

Dans ce cadre, nous proposons les deux sujets de stage suivants

Sujet 1 : étude de la correction de OcamlDuce Nous avons proposé une façon très efficace de mélanger XDuce avec Ocaml, dans un système appelé

OcamlDuce.
Cela implique des modifications du système de typage de XDuce et de son interpréteur.
Le but de ce stage est de valider cette proposition, en donnant une preuve de correction de l'interpréteur modifié, et du système de typage modifié de XDuce.

Sujet 2 : intégration de OcamlDuce et Apache pour la gestion de pages dynamiques dont la correction syntaxique est assurée par le typage...
En utilisant OcamlDuce, il est déjà possible aujourd'hui de créer des pages Web dynamiques dont la correction vis à vis d'une spécification (ex : DTD) est garantie statiquement à la compilation, ce qui rend OcamlDuce extrêmement intéressant en pratique.
Mais il y a quelques obstacles à franchir pour que ce système puisse devenir très attrayant pour un grand nombre d'utilisateurs :

Librairies pour rivaliser avec des systèmes établis, comme PHP, il est nécessaire d'identifier un certain nombre de librairies de base qui sont utiles dans la construction de sites dynamiques.

Apache une exécution efficace de code OcamlDuce dans un serveur Web comme Apache nécessite le développement d'un module équivalent au `mod_perl` existant pour Perl. Il s'agit d'un travail à la fois d'implémentation et de recherche.

Références

[1] H. Hosoya and B. C. Pierce. XDuce : A typed XML processing language. In *Proceedings of Third International Workshop on the Web and Databases (WebDB2000)*, May 2000.

[2] H. Hosoya, J. Vouillon, and B. C. Pierce. Regular expression types for XML. In *Proceedings of the International Conference on Functional Programming (ICFP)*, pages 11–22, Sept. 2000.