Proposition de stage L3

Laboratoire d'accueil
Laboratoire de Recherche en Informatique
bâtiment 490, Université Paris Sud 91405 ORSAY Cedex

Responsable du stage
Jean-Christophe Filliâtre (Jean-Christophe.Filliâtre@lri.fr)
Juin–Juillet 2005

Vérification formelle de l'algorithme d'Ukkonen

L'*arbre des suffixes* est une structure de données qui permet de représenter de manière compacte tous les suffixes d'un mot. Les applications de l'arbre des suffixes sont innombrables (recherche de motifs exacts, de répétitions, du plus long facteur commun à plusieurs mots, etc.). Une propriété remarquable de l'arbre des suffixes est qu'il peut être construit en temps et en espace *linéaire* (selon de la longueur du mot). L'algorithme d'Ukkonen est l'une des méthodes connues pour effectuer cette construction optimale. On pourra consulter l'ouvrage de Dan Gusfield [1] pour une introduction aux arbres de suffixes et à leurs applications, et pour une présentation détaillée de l'algorithme d'Ukkonen.

L'objectif de ce stage est de réaliser une vérification formelle d'une implantation en C de l'algorithme d'Ukkonen. Le code C pourra être un code libre existant ou être développé par le stagiaire. La vérification formelle du code C sera réalisée à l'aide de l'outil Caduceus [2] et de l'assistant de preuve Coq [3]. Un premier objectif consistera à spécifier formellement le code (et donc à développer la théorie logique minimale sur les mots et leurs suffixes nécessaire à cette spécification). Si le temps le permet, le second objectif sera de mener à bien la preuve proprement dite.

Références

[1] Dan Gusfield. *Algorithms on Strings, Trees, and Sequences.* Cambridge University Press, 1997.

[2] *The Caduceus tool for the verification of C programs.* `http://why.lri.fr/caduceus/`

[3] *The Coq Proof Assistant.* `http://coq.inria.fr/`