

# NFP108: devoir sur les automates

F. Barthélemy

Année 2011/2012

## 1 Présentation du projet

On veut réaliser un outil d'aide au tri d'annonces immobilières sur internet au moyen de machines finies. On veut trier les annonces qui concernent une maison avec au moins 3 chambres et une cheminée. On voudra de plus sélectionner le prix, la surface et le code postal de chaque annonce retenue. Vous trouverez sur le site web du cours une page d'exemple tirée d'un site réel. On l'utilisera comme référence pour la présentation des annonces.

On vous propose une démarche en deux temps : d'abord trier les informations pertinentes et éliminer les autres au moyen d'un transducteur. Pour le prix, on considère que c'est toute séquence de chiffres suivie du signe euro, pour la surface, toute séquence de chiffres suivie de  $m^2$ . Le code postal sera repéré au moyen de la balise HTML div de classe `rech_ville`. Le nombre de chambres sera repéré par un chiffre suivi du mot chambre au singulier ou au pluriel. Pour simplifier, on supposera le nombre de chambres inférieur à 10. Cheminée sera recherchée textuellement. Une autre information importante à utiliser sera la séparation entre annonces.

Chaque information sera traduite par le transducteur en mettant d'abord son type, puis sa valeur entre parenthèses. Par exemple,  $85 m^2$  sera traduit en `surface(85)`. On pourra utiliser soit un seul symbole pour "surface", soit une chaîne. Toutes les informations non pertinentes seront oubliées, c'est à dire réécrites en la chaîne vide  $\epsilon$ . Le transducteur s'appliquera sur le code source HTML de la page web, vu comme une séquence de caractères.

Dans un second temps, on voudra sélectionner dans les sorties du transducteur les annonces qui correspondent à notre recherche, c'est à dire avec cheminée et entre 3 et 9 chambres. Pour cela, on utilisera un automate fini dont on fera l'intersection avec la sortie du transducteur. Cet automate devra prévoir que l'ordre des différentes informations n'est pas fixé.

## 2 Travail à faire

Vous spécifierez le transducteur et l'automate au moyen d'expressions régulières combinées par des opérations régulières ou ensemblistes. Vous utiliserez Lextools et Fsm (les outils vus en TP) pour compiler ces deux machines. Vous pouvez télécharger ces outils sur le site web du cours. Pour ceux qui n'ont pas la possibilité de les exécuter (ils existent uniquement pour Unix), un service web sera prochainement fourni sur le site du cours.

Vous devrez rendre une archive .zip avec les fichiers sources lextools et un fichier texte décrivant les opérations à effectuer pour produire le transducteur et l'automate demandés. Cette archive est à rendre via un formulaire qui sera accessible sur le site du cours. La date limite est le dimanche 15 janvier 2012.