

TP n°3

JFlex

Dans ce TP nous utilisons JFlex (voir <http://jflex.de/manual.html>).

Exercice 1 Les commentaires :

On considère un langage de programmation (comme JAVA) avec des commentaires entre `/*` et `*/`. A partir du fichier en format JFlex `commentaire.flex` (à récupérer sur DIDEL) on obtient un programme qui est censé éliminer ces commentaires d'un fichier donné comme argument. Ce programme JFlex est en mode `standalone`. Cela signifie que JFlex génère dans la classe générée (ici `Lexer`) une fonction `main` qui attend comme argument un nom de fichier sur la ligne de commande et qui fait tourner l'analyseur sur ce fichier. Les valeurs retournées par l'analyseur sont ignorées et tout texte non couvert par les règles est imprimé sur la sortie standard. En exécutant `jflex commentaire.flex` on obtient le fichier `Lexer.java` qu'on peut compiler et ensuite exécuter.

- Tester le programme sur `test.java`.
- L'option `--dot` permet d'obtenir les automates produits par `jflex` en format `dot`. A partir d'un fichier en format `dot` on peut produire un fichier en format `pdf` en utilisant `dot -Tpdf fichier.dot > fichier.pdf`. Regarder les trois automates produits par JFlex à partir de `commentaire.flex`.
- Modifier le programme pour qu'il élimine aussi les commentaires du genre `//....`.
- Si on applique le programme sur des fichiers contenant des définitions de chaînes de caractères qui contiennent `/*` cela ne marche pas (voir `test2.java`). Modifier le programme pour régler ce genre de situation.
- Tester sur `test3.java` et modifier si nécessaire.
- Modifier le programme pour qu'il affiche la proportion des commentaires dans le fichier.

Exercice 2 Les chiffres dans un fichier.

- Écrire à l'aide de JFlex un programme JAVA qui calcule la somme de tous les chiffres (un chiffre est un caractère de 0 à 9) qui apparaissent dans un fichier.

Exercice 3 Wordcount :

Écrire à l'aide de JFlex un programme JAVA qui affiche le nombre de caractères, le nombre de mots et le nombre de lignes d'un fichier donné comme argument. Votre programme devrait donner le même résultat que la commande `wc` de unix. Un mot est une suite de caractères autre qu'un blanc, un retour à la ligne et une tabulation. Définissez dans JFlex une expression régulière `mot` qui correspond aux mots. Définissez également une expression régulière `finligne` qui correspond à une fin de ligne (ici soit `\n` (Unix), soit `\r`, soit `\r\n` (Windows)).

- Donner une solution à l'aide en utilisant le mode `standalone` comme dans l'exercice ???. Écrire uniquement un fichier `wcstandalone.flex` qui contiendra tout ce qu'il faut. Regardez les automates créés par JFlex.
- Enlever l'option `standalone` et regarder la différence dans le fichier généré.

- Écrire ensuite une classe permettant d'obtenir le même résultat sans utiliser l'option `standalone`. On rappelle que la méthode `yylex()` permet de lancer le scanner. Par défaut elle est censé renvoyer un token de type `Ytoken`. A la fin du fichier, elle renvoie `null`. Par conséquent, il faut définir une classe `Ytoken`, qui peut être vide, si on ne renvoie jamais rien.
- On veut utiliser maintenant les tokens pour résoudre le problème. On va utiliser le fichier `wc.flex`. Ce fichier **ne doit pas être modifié**. Définir (dans un nouveau fichier) la classe `Sym` qu'il faut (à déduire de `wc.flex`) ainsi que les classes `Token` et `StringToken` (qui hérite de `Token`). Ensuite écrire une classe `WC` avec `main` qui utilise les méthodes de `Lexer` pour le reste.

Exercice 4 Mettre au pluriel :

- Écrire un programme `JFlex` qui lit des mots d'un fichier et les met au pluriel. On ne tiendra pas compte des accents. On utilisera les règles (non exhaustives) suivantes :
 - Les mots finissant par `x`, `s` ou `z` sont invariables. Exemple : `gaz` → `gaz` ;
 - les terminaisons en `al` se transforment en `aux`, sauf pour les mots `bal`, `carnaval`, `festival` et `récitation` qui prennent un `s`. Exemple : `cheval` → `chevaux`, mais `bal` → `bals` ;
 - les mots terminant en `au` et `eu` prennent un `x`, sauf `bleu` et `pneu` ;
 - les mots en `ou` prennent un `s`, sauf `bijou`, `caillou`, `chou`, `joujou`, `genou`, `pou`, et `hibou`. Exemple : `chou` → `choux`, mais `cou` → `cous` ;
 - les terminaisons en `ail` prennent un `s` sauf pour les mots `bail`, `travail`, `soupirail`, `émail`, `vitrail` et `corail`. Exemple : `détail` → `détails`, mais `bail` → `baux` ;
 - enfin, le pluriel de `œil` est `yeux` et celui de `ciel` est `cieux` ;
 - dans tous les autres cas, on met un `s` à la fin du mot.

Tester sur le fichier `pluriel`.

- Ensuite, écrire un programme pour mettre des phrases très simples au pluriel. Les phrases sont toutes de la forme :

`<article> <nom> <verbe du premier groupe> <complément>`

`<pronom personnel> <verbe du premier groupe> <complément>`

Le complément ne sera pas mis au pluriel. Exemples :

- Le chat mange du poisson et des viandes.
Les chats mangent du poisson et des viandes.
- Un animal marche dans la forêt.
Des animaux marchent dans la forêt.
- Je souhaite du silence.
Nous souhaitons du silence.

Tester sur le fichier `plurielphrase`.

Exercice 5 Sous-titres :

Vous rappelez-vous les sous-titres du premier TP ? Maintenant, on veut les traiter plus en détails.

Écrire un fichier `JFlex` qui sépare dans un fichier `.srt` les différentes phrase de sous-titre, et qui détermine leur temps de début de l'affichage, temps de fin d'affichage, le positionnement du sous-titre sur l'écran (s'il est précisé) leur type (c'est-à-dire, s'il s'agit d'une description pour malentendant ou d'un sous-titre à proprement parler), le nom du personnage qui parle (s'il est précisé), les balises et finalement le texte.

- Utiliser le lexeur engendré pour calculer le temps total de parole dans le film, In Bruges

à partir des sous-titres qui vous sont fournis pour le premier TP. Attention aux descriptions de sons !

- *Les sous-titres en cursif indiquent que les paroles (ou les sons) transcrites (ou décrits) ont été dites (ou produits) par un personnage (ou un objet) qui n'était pas présent sur le plan. Calculez le temps total des voix off dans le film.*
- *Pour les voix off dans les sous-titres pour malentendants, on met le nom de personnage avant de citer ses paroles. Quels sont les noms de personnage qui ont parlé en voix off dans le film ?*