Programme d'extraction des balises DESCRIPTION dans les fils RSS

Pour commencer, le texte du programme que vous pourrez réutiliser *en le modifiant* dans votre programme personnel pour traiter l'arborescence complète de tous les fils RSS du Monde pour l'année 2008 :

```
1  #!/usr/bin/perl
2  use XML::RSS;
3  my $file="$ARGV[0]";
4  my $rss=new XML::RSS;
5  #-----------------------------------------------
6  my $TEXTEDUFIL="";
7  $TEXTEDUFIL.="<FILE=\"$file\">\n";
8  #-----------------------------------------------
9  $rss->parsefile($file);
10 my $nombredechampdescription=0;
11 foreach my $item (@{$rss->{'items'}}) {
12     $nombredechampdescription++;
13     $TEXTEDUFIL.="<article=\"$nombredechampdescription\">\n";
14     my $description=$item->{'description'};
15     $TEXTEDUFIL.="§ ".$description."\n";
16 }
17 #-----------------------------------------------
18 print $TEXTEDUFIL;
19
```

Ce programme est lancé de la manière suivante :

```
perl extract-txt-avec-xml-rss.pl 0,2-3208,1-0,0.xml
```

le programme prend en entrée un fil RSS passé en argument au programme dans la ligne de commandes.

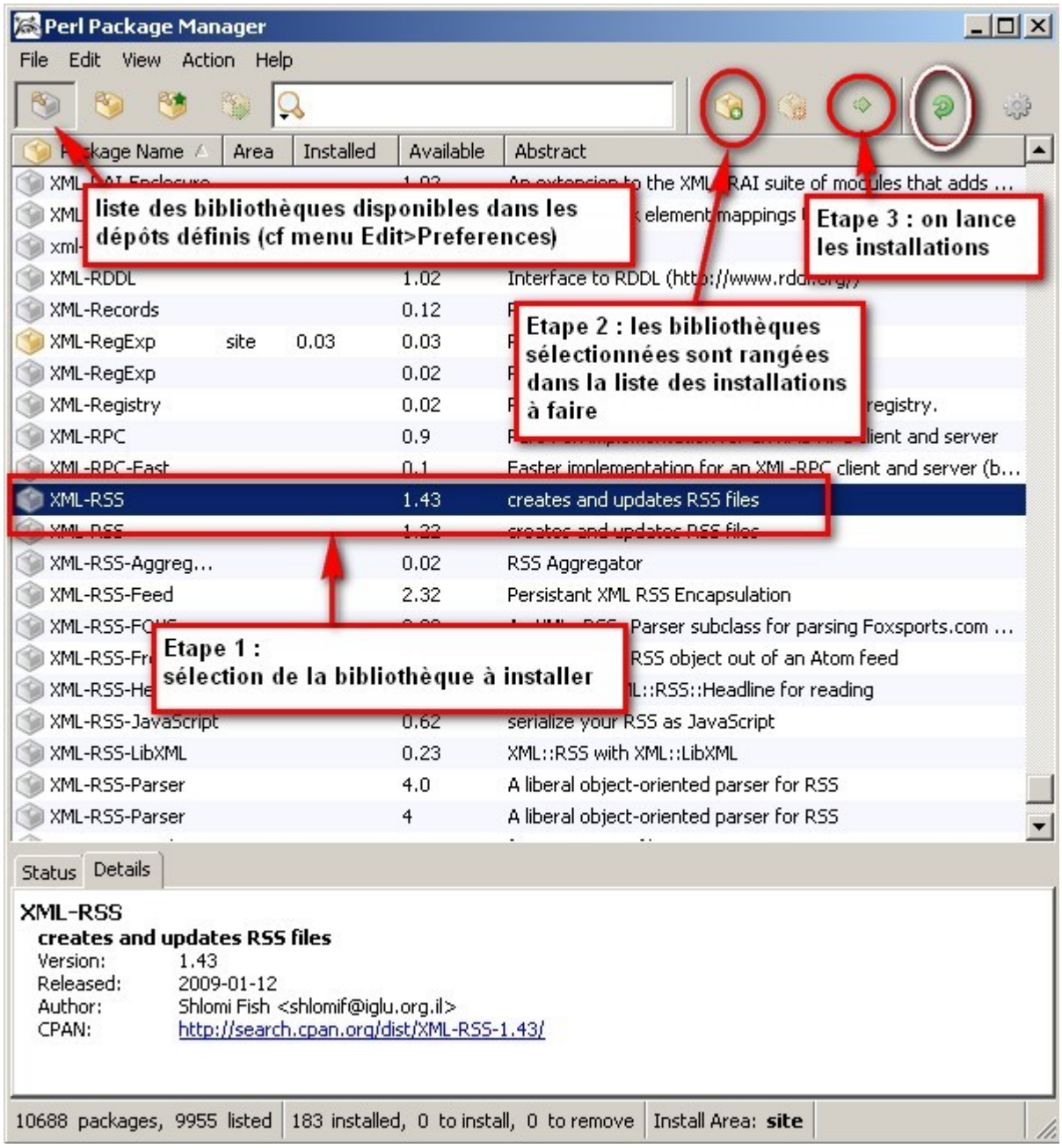
Contrairement à ce que nous avons fait dans les séances précédentes (lecture du fil ligne à ligne et repérage par des expressions régulières du contenu de la balise description de chaque item du fil RSS), ce programme utilise une bibliothèque nommée XML ::RSS (cf ligne 2) dont l'objectif est justement de manipuler « facilement » ce type de fichier pour accéder aux « différentes zones » définies par le format RSS utilisé par ce fichier.
(cf http://www.perl.com/pub/a/2000/01/rss.html)

Cette bibliothèque n'est pas installée par défaut. Pour tester ce programme, il faut donc au préalable l'installer.

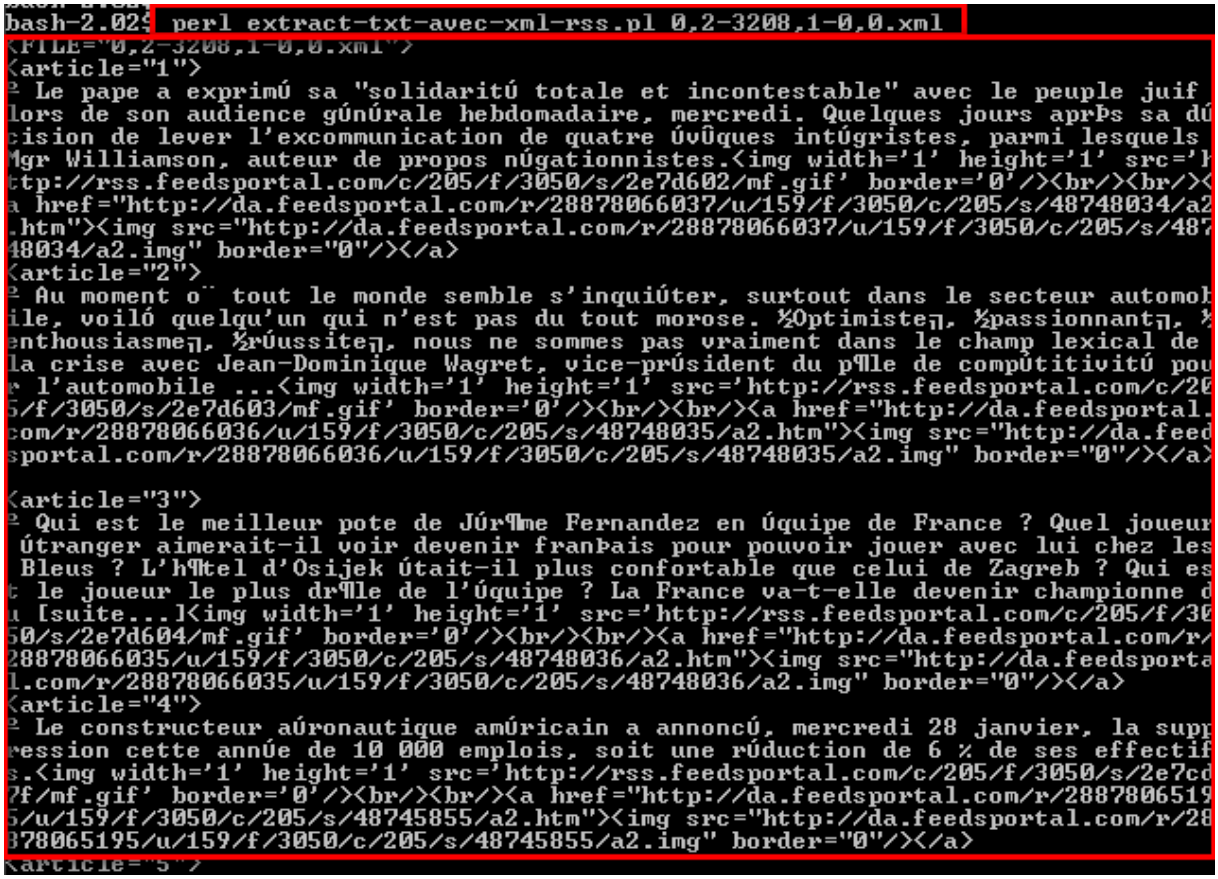
La figure qui suit montre comment réaliser cette installation pour la version de Perl Activestate installée sur les machines du labo à l'ILPGA : (Perl Activetstate 5.8 pour windows). Le principe est le même pour toutes les versions Perl ActiveState (macosx…).

Pour installer une (ou plusieurs bibliothèques), on va « se servir » dans les dépôts (de bibliothèques) disponibles en ligne (sur le site d'ActiveState ou ailleurs).

Pour faire « ce marché », on utilise le programme « *Perl Package Manager* » (qui fait partie de la distribution Perl ActiveState) et qui est en fait le gestionnaire des bibliothèques pour Perl. PPM donne à voir les bibliothèques installées sur notre machine de travail et celles disponibles dans les différents dépôts disponibles (cf menu Edit>Preferences). Dans la figure qui suit, PPM est lancé (menu démarrer>ActivePerl), on sélectionne la bibliothèque à installer, puis on l'installe…



On peut ensuite l'utiliser…



Comment ça marche…

le fonctionnement est décrit sur les sites suivants :

http://www.perl.com/pub/a/2000/01/rss.html
http://search.cpan.org/~shlomif/XML-RSS-1.43/lib/XML/RSS.pm#DESCRIPTION
http://perl-rss.sourceforge.net/

On en reparle au cours de la prochaine séance…