Entrez Direct (EDirect) is a software package that provides access to NCBI’s interconnected databases (publication, sequence, structure, gene, variation, expression, etc.). EDirect enables searching, linking to related records, retrieval of information, and extraction of specific data elements. You can easily combine individual operations in multi-step workflows to create custom datasets and generate reports.

EDirect is designed to run in a Unix command line environment and requires the Perl language interpreter to run its commands.

Unix command line environment: Unix has been around a long time and is still a good operating system for issuing instructions to a computer by typing commands on a line. Although we now have more modern ways of interacting with a computer (graphical user interface with the ability to do mouse clicks or touch the screen), using a command line interface (in a terminal window) is still a powerful and flexible way of telling a computer what to do.

Perl programming language: Perl, developed in the late 1980s and often referred to as the Swiss Army chainsaw of scripting languages, was adopted early and often as a common language for bioinformatics as it is powerful, accessible, and adaptable. EDirect uses some Perl commands, so to install and use EDirect, you will make sure that the Perl programming language is installed on your system.

1. Install EDirect

To install EDirect, first open your Mac or Linux terminal window. Then, go to the EDirect help document (http://bit.ly/EDirectHelpDoc) and scroll down to “Getting Started > Installation”. Copy and paste the text below into the terminal window.

cd ~
/bin/bash
perl -MNet::FTP -e \
  '$ftp = new Net::FTP("ftp.ncbi.nlm.nih.gov", Passive => 1);
  $ftp->login; $ftp->binary;
  $ftp->get("/entrez/entrezdirect/edirect.tar.gz");'

gunzip -c edirect.tar.gz | tar xf -
rm edirect.tar.gz
builtin exit
export PATH=$PATH:$HOME/edirect >& /dev/null || setenv PATH "${PATH}:$HOME/edirect"
./edirect/setup.sh
In order to complete the configuration process, please execute the following:
echo "export PATH=\$PATH:\$HOME/edirect" >> $HOME/.bash_profile

Copy and paste the line shown above (beginning with echo) into your terminal window.

Once the installation is complete, you may the following message in your terminal window:

Entrez Direct has been successfully downloaded and installed.

② Test your EDirect installation

To confirm that EDirect is installed and working properly, run your first EDirect script!

Type the following into your terminal window and press Enter.

esearch -help

You should see this printed on the screen:

esearch 8.30
Query Specification
   -db Database name
   -query Query string
......

Then type the following into your terminal window and press Enter.

xtract -help

You should see this printed on the screen:

xtract 8.30
Overview
   Xtract uses command-line arguments to convert XML data into a tab-delimited table.
   -pattern places the data from individual records into separate rows.
   -element extracts values from specified fields into separate columns.
   -group, -block, and -subset limit element exploration to selected XML subregions.

Processing Flags
   -strict Remove HTML and MathML tags
   -mixed Allow PubMed mixed content (Please note, the 8.30 version shown on your screen
......

You have now installed EDirect successfully and can begin using it!

Questions or feedback?
info@ncbi.nlm.nih.gov

U.S. National Library of Medicine
National Center for Biotechnology Information