```
#!usr/bin/perl

#updated may 2012 to deal with ncbi's eutils system
#makes two tables: estunigene, which connects genbank accession numbers
to unigene cluster IDs as well as preserving UGIDs, and UNIGENE, which
connects cluster IDs to genedB IDs.
    #would it be more efficient to make this a single table?
#outstanding issues: none

use CGI;
use LWP::Simple;

open (ACCESSION, "nrsortedacclist.txt");
@gbAcc = (<ACCESSION>);
close ACCESSION;

foreach $gbAcc (@gbAcc) {
    chomp $gbAcc;
    $URL =
"http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?&db=unigene&te
rm=$gbAcc";
    $file = get($URL);
    @file = split("\n", $file);
    $ugid = "";

#writes estunigene and clusterid.only (temp table for use later on in
this program)
    foreach $file (@file) {
      if ($file =~ /\<Id\>(\d*?)\<\/Id\>/) {$ugid = $1; next}
    }


    $URL2 =
"http://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=unigene&id
=$ugid";
    $file2 = get($URL2);
    @file2 = split("\n", $file2);
    $cluster = "";

    foreach $file2 (@file2) {
      if ($file2 =~ /CLUSTERID" Type\="String"\>(.*)\<\/Item\>/)
{$cluster = $1; next}
    }

    foreach $file2 (@file2) {
    if ($file2 =~ /GENEID" Type\="String"\>(\d*)\<\/Item\>/) {$gene = $1;
next}
    }

#checks to make sure the cluster is human, skips to the next accession
number on the list
    unless ($cluster =~ /Hs\.(\d*)/) {open ERROROUT, ">>bad.clusters";
print ERROROUT $cluster, "\n"; close ERROROUT; next}
```

```
#checks to make sure there was an entry for that accession number
    unless ($cluster) {print "NO DATA $gbAcc\n"; open NO_ENTRY,
">>no.clusters"; print NO_ENTRY $gbAcc, "\n"; close NO_ENTRY; next}


    open (ESTUNIGENE, ">>tableestunigene");
    print ESTUNIGENE $gbAcc, "\t";
    print ESTUNIGENE $ugid, "\t";
    print ESTUNIGENE $cluster, "\n";
    print $gbAcc, "\t", $ugid, "\t", $cluster, "\t";
    close ESTUNIGENE;

    open (CLUSTERLIST, ">>clusterid.only");
    if ($cluster) {print CLUSTERLIST $cluster, "\n"}
    close CLUSTERLIST;

    open (UNIGENE, ">>tableunigene");
    print UNIGENE $cluster, "\t";
    print UNIGENE $gene, "\n";
    print $gene, "\n";
    close UNIGENE;

    open (GENEID, ">>geneid.only");
    print GENEID $gene, "\n";
    close GENE;
}
```