```perl
#!/usr/bin/perl

use POSIX;

$stamp = `date`;
&get_old;
&make_dates;
$linkroot = "http://www.ca6.uscourts.gov";
open(META, ">meta-cir6.txt");
open(WGET, "wget http://www.ca6.uscourts.gov/cgi-
bin/opinions.pl?FROMDATE=$last_month%20TODATE=$to_day -q -O -|");
$flag = 0;
READIT: while(<WGET>){
&find_stuff;
}
close WGET;

# ------------ Subroutines ----------------

sub find_stuff{
    if(/PUBLISHED OPINIONS/){
        $status = "Status: Published";
    }
    if(/NOT RECOMMENDED/){
        $status = "Status: Not Published";
    }
    if(/<a href=\"?\/opinions/ && $flag == 1){
        print "CHOVAH $file\n";
    }
    if(/<font size.*?>(.*?)<\/font>/ && $flag == 4){
            $summary = "$1, $status";
            $caption =~ s/\'/\\'/sg;
            $caption =~ s/\"/&quot;/sg;
            $caption =~ s/\r//sg;
            $flag = 0;
            print META "$fname\"$fname\_cir6\"$docket\"$decided\"Sixth
Circuit\"$caption\"On Appeal From: $summary\n";
            &mod_html;
            undef $fname;
            undef $docket;
            undef $link;
            undef $decided;
            undef $caption;
            undef $summary;
            ++$count5;
    }
    if(/<td>\&nbsp\;(.*)/ && $flag == 3){
            $caption = $1;
            $flag = 4;
            ++$count4;
    }
    if(/<td align=center>(\d{2,4})\/(\d{1,2})\/(\d{1,2})/ && $flag ==
2){
            $year = $1;
```

```perl
			$month = $2;
			$day = $3;
			$decided = "$year-$month-$day";
			&date_test;
			$flag = 3;
			++$count3;
	}
	if(/<a href=\"?\/paceruser\.html\"?>\W+(\d.*)<\/a>/ && $flag == 1){
			$docket = $1;
			$flag = 2;
			++$count2;
	}
	if(/pacer\.ca6\.uscourts\.gov\/cgi-bin/ && $flag == 1){
			 $flag = 9;
	}
	if(/(\d\d-\d{3,6})\W*<\/a>/ && $flag == 9){
		$docket = $1;
		&dup_test;
		$flag = 2;
		++$count2;
	}
	if(/<a href=\"?(\/opinions\.pdf\/.*?\.pdf)\"?>(.*)<\/a>/){
		$link = "$linkroot$1";
		$fname = $link;
		$fname =~ s/.*?\/(\d.*-\d\d)\.pdf/$1/;
		$docket = $fname;
		&dup_test;
#		 if($go == 1){
		&get_dock;
		  print "got doc\n";
#			$go = 0;
#		  }
		$flag = 1;
		++$count;
	}


} # end sub find_stuff

sub get_dock{
system("wget $link");
system("pdftotext -layout -htmlmeta $fname.pdf $fname\_cir6.html");
} # end sub get_dock.

sub date_test{
unless($year =~ /\d{4}/){
$year = "20$year";
}
}

 sub dup_test{
 @test = grep /$docket/, @raw_data;
 if($test[0] ne ""){
 print "we already got this one: $docket compare $test[0]\n";
```

```perl
$flag = 0;
undef @test;
unlink("$fname.pdf");
unlink("$fname\_cir6.html");
next READIT;
}else{
  ++$count7;
  $go = 1;
  print "$docket is a newbie. Get it. $count7\n";
 }

} # end sub dup_test.

sub make_dates{
@date1 = (localtime)[3..5]; # grabs day/month/year values
$date1[1] = ($date1[1] +1);
$date1[2] = ($date1[2] + 1900);
$date1[1] = sprintf "%2.2d%s", $date1[1];
$date1[0] = sprintf "%2.2d%s", $date1[0];
$to_day = "$date1[1]\/$date1[0]\/$date1[2]";
$lmonth = localtime(time - 2629743);
@date2 = split / /, $lmonth;
print "$date2[1]: @date2\n";
%months = ( "Jan", 1,"Feb", 2,"Mar", 3,"Apr",
4,"May", 5,"Jun", 6,"Jul", 7,"Aug", 8,"Sep",
9,"Oct", 10,"Nov", 11,"Dec", 12 );
$mon = $months{$date2[1]};
$mon = sprintf "%2.2d%s", $mon;
if($date2[5] =~ /\d\d(\d\d)/){
$date2[5] = $1;
}else{
$date2[5] = $date2[4];
$date2[5] =~ s/^\d\d//;
}
$date2[3] = sprintf "%2.2d%s", $date2[3];
$last_month = "$mon/$date2[3]/$date2[5]";
print "today: $to_day  last month=$last_month\n";
} # end sub make_dates

sub get_old(){
open(DOCKS, "wget
http://camlaw.rutgers.edu/~jjoerg/ponies/puppies.cgi?court=Sixth -q -O -
|");
@raw_data = <DOCKS>;
close DOCKS;
chomp @raw_data;
} # end sub get_old

sub mod_html{
# first, make a unix-date stamp so swish-e can do dates.
$umon = ($month - 1);
$uyear = ($year - 1900);
$unixtime = mktime (0, 0, 0, $day, $umon, $uyear, 0, 0);
# now, deal with the page breaks and insert metadata.
```

```perl
open(NEW, "$fname\_cir6.html") || die "cannot open $fname: $!\n";
read(NEW, $buff, 80000);
close NEW;
$buff =~ s/<title>.*<\/title>/<title>$fname.html<\/title>/;
$buff =~ s/<meta name=.*?>//sg;
$buff =~ s/\n{2,}(
\d{1,3}\W*\n)/\n\n_____\n$1/sg;
$buff =~ s/\f/_____\n\n/sg;
$buff =~ s/<head>/<head>
<META NAME=\"cite\" CONTENT=\"$docket\">
<META NAME=\"docket\" CONTENT=\"$docket\">
<META NAME=\"decided\" CONTENT=\"$year-$month-$day\">
<META NAME=\"udate\" CONTENT=\"$unixtime\">
<META NAME=\"court\" CONTENT=\"Sixth Circuit">
<META NAME=\"caption\" CONTENT=\"$caption\">
<META NAME=\"uri\"
CONTENT=\"\/resource.org\/fed_reporter\/F3\/cir6\/$fname\_cir6.html\">
<META NAME=\"summary\" CONTENT=\"$summary\">
<META NAME=\"revisionDate\" CONTENT=\"$stamp\">
<META NAME=\"revisionDesc\" CONTENT=\"Downloaded from the Circuit Court
website, converted from original PDF format using wptotext with -htmlmeta
and -layout. Citations converted to hypertext links.\">\n/i;
open(MOD, ">$fname\_cir6.html");
print MOD $buff;
close MOD;
# undef $status;
undef $summary;
undef $year;
undef $month;
undef $day;
undef $unixdate;
undef $umon;
undef $uyear;
undef $caption;
undef $decided;
undef $docket;
undef $fname;
undef $link;
} # end sub mod_html.
```