```perl
#!c:\Dwimperl\perl\bin

# Author: Jordan Schoenfeld;
# "Contracts between Firms and Shareholders";
# Journal of Accounting Research;
# Description: This file produces a pipe-delimited data set that identifies whether a 13D contains a shareholder contract and the other attributes of the contracts. Results are printed to FILE2 as defined below.;

# First create list of all files in subdirectories;
use File::Find;
use HTML::Strip;
use List::MoreUtils 'true';

# This is the directory where I want to look for all the 13D files.;
my $localdir = 'D:\jordan\data\edgar\13\new_13d';

# This is where I store the list of file names in the $localdir.;
my $filelist = 'D:\jordan\data\edgar\13\new_13d_results\filelist.txt';

# This is where I put the results.;
my $resultfile = 'D:\jordan\data\edgar\13\new_13d_results\results.txt';

# Now parse through all the individual files using the locations in the text file I made above;
# FILE2 is where I write the results to. This will wipe out whatever is in FILE2.;
open FILE2, ">", $resultfile or die$!;
# Print column headings;
print FILE2 "filename|filetype|doccount|target_cik|target_name|target_city|target_state|target_zip|target_cusip|target_cusip2|filer_cik|filer_name|filer_city|filer_state|filer_zip|lines|words|chars|filedate|conformdate|dateofchange|investor_agreement|investment_contract|stockholder_contract|cooperation|rights|exchange|covenant|investor_contract|any_agreement|contract|support_agreement|investment_agreement|stockholder_agreement|merger|acquisition|tender|tender_offer|shelf|block_size|compensation|exec_compensation|salary|source_sc|source_bk|source_af|source_wc|source_pf|source_oo|reporting_type_bd|reporting_type_bk|reporting_type_ic|reporting_type_iv|reporting_type_ia|reporting_type_ep|reporting_type_hc|reporting_type_sa|reporting_type_cp|reporting_type_co|reporting_type_pn|reporting_type_in|reporting_type_oo|fcovenant|acovenant|ncovenant|info_confidential|info_confidentiality|info_informationrights|info_accessandinfo|info_accesstoinfo|info_accesstodocs|info_visitinspect|info_communication|info_communicate|info_communicationacct|info_nda|info_fullaccess|info_books|investment_purpose|shareholder_contract|roa|roe|net_income|earnings|revenue|sales|cash_flow|interest_coverage|interest|net_worth|debt|leverage|current_ratio|debt_to_equity|assets|debt_to_cash_flow|ebit|ebitda|debt_service|book_value|shareholder_agreement|dividend|payout|distribut\n";

# Set skip equal to 1 if you want the script to analyze only a subset of 13D filings ($skip = 0 is the default).;
$skip = 0;
if ($skip == 0) {
    # Create list of all files within the directory and subdirectories of $localdir;
    open FILE3, ">", $filelist or die$!;
    # This looks into $localdir and returns the path to all files in that folder and its subfolders;
    find(
    sub {print FILE3 $File::Find::name, "\n" if /\.txt$/}, $localdir);
    close (FILE3);
}

# Now we analyze all the files in FILE4.;
open FILE4, "<", $filelist or die$!;
# Determine number of files that need to be parsed in the filelist file.;
$totallines = 0;
while (<FILE4>) {
    $totallines++;
}
```

```
close (FILE4);

# Open the file containing the list of files to be parsed.;
open FILE4, "<", $filelist or die$!;
$filecount = 0; # This is used to keep track of how many files have been parsed;

# Now we start processing each 13D at a time (the code at the end of this file creates one observation row for each 13D).;
while (<FILE4>) {
        open (RES, $_) or die "could not open $_\n";
        $filecount++;
        #### Reset ALL the variables for each 13D as they are processed here.;
        my ($lines, $words, $chars) = (0,0,0);
        my $filename = $_;
        $filename =~ s/\s*$//; # Take file name and trim whitespace from the right;
        $filedate = '.';
        $conformdate = '.';
        $dateofchange = '.';
        $doccount = '.';
        $cik_num = '.';
        $cik_num_filer = '.';
        $first_cik = 0;
        $first_name = 0;
        $first_zip = 0;
        $count = 0;
        $count2 = 0;
        $fls = 0;
        $item_info = '.';
        $item_info2 = '.';
        $item_info3 = '.';
        $item_info4 = '.';
        $item_info5 = '.';
        $item_info6 = '.';
        $item_info7 = '.';
        $item_info8 = '.';
        $item_info9 = '.';
        $item_info10 = '.';
        $item_info11 = '.';
        $item_info12 = '.';
        $item_info12_search = '.';
        $item_info12_count = 0;
        $agreement = 0;
        $cooperation = 0;
        $rights = 0;
        $exchange = 0;
        $covenant = 0;
        $any_agreement = 0;
        $contract = 0;
        $merger = 0;
        $acquisition = 0;
        $tender = 0;
        $shelf = 0;
        $support_agreement = 0;
        $investment_agreement = 0;
        $clean2 = '.';
        $percent_extract = '.';
        $percent_count = 0;
        $percent_search = 0;
        $tender_offer = 0;
        $compensation = 0;
        $salary = 0;
        $exec_compensation = 0;
        $stockholder_agreement = 0 ;
        $fund_source_count = 0;
        $fund_source_search = 0;
        $source_of_funds = '.';
```

```
111         $source_sc = 0;
112         $source_bk = 0;
113         $source_af = 0;
114         $source_wc = 0;
115         $source_pf = 0;
116         $source_oo = 0;
117         $reporting_type_count = 0;
118         $reporting_type_search = 0;
119         $reporting_type = '.';
120         $reporting_type_bd = 0;
121         $reporting_type_bk = 0;
122         $reporting_type_ic = 0;
123         $reporting_type_iv = 0;
124         $reporting_type_ia = 0;
125         $reporting_type_ep = 0;
126         $reporting_type_hc = 0;
127         $reporting_type_sa = 0;
128         $reporting_type_cp = 0;
129         $reporting_type_co = 0;
130         $reporting_type_pn = 0;
131         $reporting_type_in = 0;
132         $reporting_type_oo = 0;
133         $fcovenant = 0;
134         $acovenant = 0;
135         $ncovenant = 0;
136         $info_confidential = 0;
137         $info_confidentiality = 0;
138         $info_informationrights = 0;
139         $info_accessandinfo = 0;
140         $info_accesstoinfo = 0;
141         $info_accesstodocs = 0;
142         $info_visitinspect = 0;
143         $info_communication = 0;
144         $info_communicate = 0;
145         $info_communicationacct = 0;
146         $info_nda = 0;
147         $info_fullaccess = 0;
148         $info_books = 0;
149         $investment_purpose = 0;
150         $shareholder_contract = 0;
151         $stockholder_contract = 0;
152         $investor_contract = 0;
153         $investment_contract = 0;
154         $shareholder_agreement = 0;
155         # Accounting terms (see pg. 1101 Dichev and Skinner, JAR);
156         $roa = 0;
157         $roe = 0;
158         $net_income = 0;
159         $earnings = 0;
160         $revenue = 0;
161         $sales = 0;
162         $cash_flow = 0;
163         $interest_coverage = 0;
164         $interest = 0;
165         $net_worth = 0;
166         $debt = 0;
167         $leverage = 0;
168         $current_ratio = 0;
169         $debt_to_equity = 0;
170         $assets = 0;
171         $debt_to_cash_flow = 0;
172         $ebit = 0;
173         $ebitda = 0;
174         $debt_service = 0;
175         $book_value = 0;
176         $dividend = 0;
```

```
177         $payout = 0;
178         $distribut = 0;
179         # Set counts that stop from parsing for CIK, etc. when these items have already
            been found in the code below;
180         $done1 = 0;
181         $done2 = 0;
182         $done3 = 0;
183         $done4 = 0;
184         $done5 = 0;
185         $done6 = 0;
186         $done7 = 0;
187         $done8 = 0;
188         $done9 = 0;
189         $done10 = 0;
190         $done11 = 0;
191         $done12 = 0;
192         $done13 = 0;
193         $done14 = 0;
194         $done15 = 0;
195         $done16 = 0;
196         $done17 = 0;
197         $done18 = 0;
198         $done19 = 0;
199         $done20 = 0;
200         $done21 = 0;
201         $done22 = 0;
202         $done23 = 0;
203         $done24 = 0;
204         $done25 = 0;
205         $done26 = 0;
206         $done27 = 0;
207         $done28 = 0;
208         $done29 = 0;
209         $done30 = 0;
210         $done31 = 0;
211         $done32 = 0;
212         $done33 = 0;
213         $done34 = 0;
214         $done35 = 0;
215         $done36 = 0;
216         $done37 = 0;
217         $done38 = 0;
218         $done39 = 0;
219         $done40 = 0;
220         $done41 = 0;
221         $done42 = 0;
222         $done43 = 0;
223         $done44 = 0;
224         $done45 = 0;
225         $done46 = 0;
226         $done47 = 0;
227         $done48 = 0;
228         $done49 = 0;
229         $done50 = 0;
230         $done51 = 0;
231         $done52 = 0;
232         $done53 = 0;
233         $done54 = 0;
234         $done55 = 0;
235         $done56 = 0;
236         $done57 = 0;
237         $done58 = 0;
238         $done59 = 0;
239         $done60 = 0;
240         $done61 = 0;
241         $done62 = 0;
```

```
        $done63 = 0;
        $done64 = 0;
        $done65 = 0;
        $done66 = 0;
        $done67 = 0;
        $done68 = 0;
        $done69 = 0;
        $done70 = 0;
        $done71 = 0;
        $done72 = 0;
        $done73 = 0;
        $done74 = 0;
        $done75 = 0;
        $done76 = 0;
        $done77 = 0;
        $done78 = 0;
        $done79 = 0;
        $done80 = 0;
        $done81 = 0;
        $done82 = 0;
        $done83 = 0;
        $done84 = 0;
        $done85 = 0;
        # Begin reading in each line from opened file;
        while (<RES>) {
            # First I need to make sure that I avoid PDFs and JPGs embedded within the files, which show up as long strings of garbled text;
            # Split $_ by space and if that is greater than 55 assume it is garbled.;
            @garblecheck = split / /, $_;
            if (length($garblecheck[0]) > 50) {
                next;
            }
            # Strip html as each line is read in;
            my $hs = HTML::Strip->new();
            my $clean = $hs->parse($_);
            $hs->eof;
            # End html strip;
            chomp $clean;
            # Trim white spaces horizontally and vertically;
            $clean =~ s/[\h\v]+/ /g;
            # Remove non-UTF8 chars;
            $clean =~ s/[^[:ascii:]]{2}(.*?[ ]{3})/$1 /g;
            # Create line, character, and words count;
            $lines++;
            $chars += length($clean);
            $words += scalar(split(/\s+/, $clean));
            # This tells the program to differentiate between target and filer CIK;
            $first_cik = 0;
            $first_name = 0;
            $first_zip = 0;
            # @wordsplit = split(/ /, $clean);
            # Split sentence into an array by period, space, etc. so I can use word dictionaries later;
            @wordsplit = split /[., ,:,;,(,),!,\,]+/, $clean;
            $slength = length $clean;
                # Extract document identifier information.
                # NOTE: these data will be cross-checked later (in SAS) with data from WRDS SEC Analytics.;
                # NOTE: the code segments below simply repeat themselves with each segment extracting a new piece of data from the filing.;
                if ($count == 0)
                {
                    if ((index($clean, 'FILED AS OF DATE:') != -1) && $count == 0 && $done1 == 0) {
                        $filedate = $clean;
                        $filedate =~ s/[^0-9]//g;
```

```
                    $done1 = 1;
                    if ($filedate eq "")
                    {
                      $filedate = '.';
                    }
                  }
                  if ((index($clean, 'CONFORMED PERIOD OF REPORT:') != -1) && $count == 0 && $done2 == 0) {
                    $conformdate = $clean;
                    $conformdate =~ s/[^0-9]//g;
                    $done2 = 1;
                    if ($conformdate eq "")
                    {
                      $conformdate = '.';
                    }
                  }
                  if ((index($clean, 'PUBLIC DOCUMENT COUNT:') != -1) && $count == 0 && $done3 == 0) {
                    $doccount = $clean;
                    $doccount =~ s/[^0-9]//g;
                    $done3 == 1;
                    if ($doccount eq "")
                    {
                      $doccount = '.';
                    }
                  }
                  if ((index($clean, 'CENTRAL INDEX KEY:') != -1) && $count == 0 && $done4 == 0) {
                    $cik_num = $clean;
                    $cik_num =~ s/[^0-9]//g;
                    $done4 = 1;
                    $first_cik = 1;
                    if ($cik_num eq "")
                    {
                      $cik_num = '.';
                    }
                  }
                  if ((index($clean, 'CENTRAL INDEX KEY:') != -1) && $count == 0 && $done13 == 0 && $first_cik == 0) {
                    $cik_num_filer = $clean;
                    $cik_num_filer =~ s/[^0-9]//g;
                    $done13 = 1;
                    if ($cik_num_filer eq "")
                    {
                      $cik_num_filer = '.';
                    }
                  }
                  if ((index($clean, 'DATE AS OF CHANGE:') != -1) && $count == 0 && $done9 == 0) {
                    $dateofchange = $clean;
                    $dateofchange =~ s/[^0-9]//g;
                    $done9 = 1;
                    if ($dateofchange eq "")
                    {
                      $dateofchange = '.';
                    }
                  }
                  if ((index(lc($clean), 'investor agreement') != -1) && $count == 0 && $done10 == 0) {
                    $done10 = 1;
                    $agreement = 1;
                  }
                  if ((index(lc($clean), 'investor contract') != -1) && $count == 0 && $done84 == 0) {
                    $done84 = 1;
                    $investor_contract = 1;
```

```
                    }
                    if ((index(lc($clean), 'stockholder contract') != -1) && $count == 0 && $done85 == 0) {
                        $done85 = 1;
                        $stockholder_contract = 1;
                    }
                    if ((index(lc($clean), 'cooperation agreement') != -1) && $count == 0 && $done11 == 0) {
                        $done11 = 1;
                        $cooperation = 1;
                    }
                    if ((index(lc($clean), 'investor rights') != -1) && $count == 0 && $done12 == 0) {
                        $done12 = 1;
                        $rights = 1;
                    }
                    if ((index(lc($clean), 'exchange agreement') != -1) && $count == 0 && $done14 == 0) {
                        $done14 = 1;
                        $exchange = 1;
                    }
                    if ((index(lc($clean), 'investment agreement') != -1) && $count == 0 && $done40 == 0) {
                        $done40 = 1;
                        $investment_agreement = 1;
                    }
                    if ((index(lc($clean), 'investment contract') != -1) && $count == 0 && $done83 == 0) {
                        $done83 = 1;
                        $investment_contract = 1;
                    }
                    if ((index(lc($clean), 'covenant') != -1) && $count == 0 && $done25 == 0) {
                        $done25 = 1;
                        $covenant = 1;
                    }
                    if ((index(lc($clean), 'agreement') != -1) && $count == 0 && $done26 == 0) {
                        $done26 = 1;
                        $any_agreement = 1;
                    }
                    if ((index(lc($clean), 'contract') != -1) && $count == 0 && $done27 == 0) {
                        $done27 = 1;
                        $contract = 1;
                    }
                    if ((index(lc($clean), 'merger') != -1) && $count == 0 && $done28 == 0) {
                        $done28 = 1;
                        $merger = 1;
                    }
                    if ((index(lc($clean), 'acquisition') != -1) && $count == 0 && $done29 == 0) {
                        $done29 = 1;
                        $acquisition = 1;
                    }
                    if ((index(lc($clean), 'tender') != -1) && $count == 0 && $done30 == 0) {
                        $done30 = 1;
                        $tender = 1;
                    }
                    if ((index(lc($clean), 'tender offer') != -1) && $count == 0 && $done34 == 0) {
                        $done34 = 1;
                        $tender_offer = 1;
                    }
```

```
415                     if ((index(lc($clean), 'shelf') != -1) && $count == 0 && $done31 ==
                        0) {
416                         $done31 = 1;
417                         $shelf = 1;
418                     }
419                     if ((index(lc($clean), 'support agreement') != -1) && $count == 0 &&
                         $done33 == 0) {
420                         $done33 = 1;
421                         $support_agreement = 1;
422                     }
423                     if ((index(lc($clean), 'compensation') != -1) && $count == 0 &&
                        $done37 == 0) {
424                         $done37 = 1;
425                         $compensation = 1;
426                     }
427                     if ((index(lc($clean), 'executive compensation') != -1) && $count ==
                         0 && $done38 == 0) {
428                         $done38 = 1;
429                         $exec_compensation = 1;
430                     }
431                     if ((index(lc($clean), 'salary') != -1) && $count == 0 && $done39 ==
                         0) {
432                         $done39 = 1;
433                         $salary = 1;
434                     }
435                     if ((index(lc($clean), 'stockholder agreement') != -1) && $count ==
                        0 && $done40 == 0) {
436                         $done40 = 1;
437                         $stockholder_agreement = 1;
438                     }
439                     if ((index(lc($clean), 'financial covenant') != -1) && $count == 0
                        && $done42 == 0) {
440                         $done42 = 1;
441                         $fcovenant = 1;
442                     }
443                     if ((index(lc($clean), 'affirmative covenant') != -1) && $count == 0
                         && $done43 == 0) {
444                         $done43 = 1;
445                         $acovenant = 1;
446                     }
447                     if ((index(lc($clean), 'negative covenant') != -1) && $count == 0 &&
                         $done44 == 0) {
448                         $done44 = 1;
449                         $ncovenant = 1;
450                     }
451                     if ((index(lc($clean), 'confidential') != -1) && $count == 0 &&
                        $done45 == 0) {
452                         $done45 = 1;
453                         $info_confidential = 1;
454                     }
455                     if ((index(lc($clean), 'confidentiality') != -1) && $count == 0 &&
                        $done46 == 0) {
456                         $done46 = 1;
457                         $info_confidentiality = 1;
458                     }
459                     if ((index(lc($clean), 'information rights') != -1) && $count == 0
                        && $done47 == 0) {
460                         $done47 = 1;
461                         $info_informationrights = 1;
462                     }
463                     if ((index(lc($clean), 'access and information') != -1) && $count ==
                         0 && $done48 == 0) {
464                         $done48 = 1;
465                         $info_accessandinfo = 1;
466                     }
467                     if ((index(lc($clean), 'access to information') != -1) && $count ==
```

```
0 && $done49 == 0) {
    $done49 = 1;
    $info_accesstoinfo = 1;
}
if ((index(lc($clean), 'access to documents') != -1) && $count == 0 && $done50 == 0) {
    $done50 = 1;
    $info_accesstodocs = 1;
}
if ((index(lc($clean), 'visit and inspect') != -1) && $count == 0 && $done51 == 0) {
    $done51 = 1;
    $info_visitinspect = 1;
}
if ((index(lc($clean), 'communication') != -1) && $count == 0 && $done52 == 0) {
    $done52 = 1;
    $info_communication = 1;
}
if ((index(lc($clean), 'communicate') != -1) && $count == 0 && $done53 == 0) {
    $done53 = 1;
    $info_communicate = 1;
}
if ((index(lc($clean), 'communication with accountant') != -1) && $count == 0 && $done54 == 0) {
    $done54 = 1;
    $info_communicationacct = 1;
}
if ((index(lc($clean), 'non-disclosure agreement') != -1) && $count == 0 && $done55 == 0) {
    $done55 = 1;
    $info_nda = 1;
}
if ((index(lc($clean), 'full access') != -1) && $count == 0 && $done56 == 0) {
    $done56 = 1;
    $info_fullaccess = 1;
}
if ((index(lc($clean), 'books') != -1) && $count == 0 && $done57 == 0) {
    $done57 = 1;
    $info_books = 1;
}
if ((index(lc($clean), 'investment purpose') != -1) && $count == 0 && $done58 == 0) {
    $done58 = 1;
    $investment_purpose = 1;
}
if ((index(lc($clean), 'shareholder contract') != -1) && $count == 0 && $done59 == 0) {
    $done59 = 1;
    $shareholder_contract = 1;
}
if ((index(lc($clean), 'return on asset') != -1) && $count == 0 && $done60 == 0) {
    $done60 = 1;
    $roa = 1;
}
if ((index(lc($clean), 'return on equity') != -1) && $count == 0 && $done61 == 0) {
    $done61 = 1;
    $roe = 1;
}
if ((index(lc($clean), 'net income') != -1) && $count == 0 && $done62 == 0) {
```

```
                        $done62 = 1;
                        $net_income = 1;
                    }
                    if ((index(lc($clean), 'earning') != -1) && $count == 0 && $done63 == 0) {
                        $done63 = 1;
                        $earnings = 1;
                    }
                    if ((index(lc($clean), 'revenue') != -1) && $count == 0 && $done64 == 0) {
                        $done64 = 1;
                        $revenue = 1;
                    }
                    if ((index(lc($clean), 'sales') != -1) && $count == 0 && $done65 == 0) {
                        $done65 = 1;
                        $sales = 1;
                    }
                    if ((index(lc($clean), 'cash flow') != -1) && $count == 0 && $done66 == 0) {
                        $done66 = 1;
                        $cash_flow = 1;
                    }
                    if ((index(lc($clean), 'interest coverage') != -1) && $count == 0 && $done67 == 0) {
                        $done67 = 1;
                        $interest_coverage = 1;
                    }

                    if ((index(lc($clean), 'interest') != -1) && $count == 0 && $done68 == 0) {
                        $done68 = 1;
                        $interest = 1;
                    }

                    if ((index(lc($clean), 'net worth') != -1) && $count == 0 && $done69 == 0) {
                        $done69 = 1;
                        $net_worth = 1;
                    }

                    if ((index(lc($clean), 'debt') != -1) && $count == 0 && $done70 == 0) {
                        $done70 = 1;
                        $debt = 1;
                    }

                    if ((index(lc($clean), 'leverage') != -1) && $count == 0 && $done71 == 0) {
                        $done71 = 1;
                        $leverage = 1;
                    }

                    if ((index(lc($clean), 'current ratio') != -1) && $count == 0 && $done72 == 0) {
                        $done72 = 1;
                        $current_ratio = 1;
                    }

                    if ((index(lc($clean), 'assets') != -1) && $count == 0 && $done73 == 0) {
                        $done73 = 1;
                        $assets = 1;
                    }

                    if ((index(lc($clean), 'debt to cash') != -1) && $count == 0 &&
```

```
$done74 == 0) {
    $done74 = 1;
    $debt_to_cash_flow = 1;
}

if ((index(lc($clean), 'ebit') != -1) && $count == 0 && $done75 == 0) {
    $done75 = 1;
    $ebit = 1;
}

if ((index(lc($clean), 'ebitda') != -1) && $count == 0 && $done76 == 0) {
    $done76 = 1;
    $ebitda = 1;
}

if ((index(lc($clean), 'debt service') != -1) && $count == 0 && $done77 == 0) {
    $done77 = 1;
    $debt_service = 1;
}

if ((index(lc($clean), 'book value') != -1) && $count == 0 && $done78 == 0) {
    $done78 = 1;
    $book_value = 1;
}
if ((index(lc($clean), 'shareholder agreement') != -1 or index(lc($clean), "shareholders' agreement") != -1 or index(lc($clean), 'shareholders agreement') != -1) && $count == 0 && $done79 == 0) {
    $done79 = 1;
    $shareholder_agreement = 1;
}

if ((index(lc($clean), 'dividend') != -1) && $count == 0 && $done80 == 0) {
    $done80 = 1;
    $dividend = 1;
}

if ((index(lc($clean), 'payout') != -1) && $count == 0 && $done81 == 0) {
    $done81 = 1;
    $payout = 1;
}

if ((index(lc($clean), 'distribut') != -1) && $count == 0 && $done82 == 0) {
    $done82 = 1;
    $distribut = 1;
}
# This gets the item information attribute from the 8k ;
# NOTE: As above, much of this information is later cross-checked with data from WRDS SEC Analytics.;
if ((index($clean, 'CONFORMED SUBMISSION TYPE:') != -1) && $count == 0 && $done5 == 0) {
    $item_info = lc(substr($clean, 27, $slength)); # This gets the item info and makes it lower case for ease of coding later;
    # Trim extra whitespace from the item info;
    $item_info =~ s/^\s+//; # From left;
    $item_info =~ s/\s*$//; # From right;
    $item_line = $lines;
    $item_info =~ s/[|]+//g;
    $done5 = 1;
    if ($item_info eq "")
```

```
                        {
                          $done5 = 3; # I don't want the program to look any further if there isn't a second item;
                          $item_info = '.';
                        }
                    }
                    if ((index($clean, 'COMPANY CONFORMED NAME:') != -1) && $count == 0 && $done15 == 0) {
                        $item_info2 = lc(substr($clean, 25, $slength)); # This gets the item info and makes it lower case for ease of coding later;
                        # Trim extra whitespace from the item info;
                        $item_info2 =~ s/^\s+//; # From left;
                        $item_info2 =~ s/\s*$//; # From right;
                        $item_line = $lines;
                        $item_info2 =~ s/[|]+//g;
                        $done15 = 1;
                        $first_name = 1;
                        if ($item_info2 eq "")
                        {
                          $done15 = 3; # I don't want the program to look any further if there isn't a second item;
                          $item_info2 = '.';
                        }
                    }
                    if ((index($clean, 'COMPANY CONFORMED NAME:') != -1) && $count == 0 && $done16 == 0 && $first_name == 0) {
                        $item_info3 = lc(substr($clean, 25, $slength)); # This gets the item info and makes it lower case for ease of coding later;
                        # Trim extra whitespace from the item info;
                        $item_info3 =~ s/^\s+//; # From left;
                        $item_info3 =~ s/\s*$//; # From right;
                        $item_line = $lines;
                        $item_info3 =~ s/[|]+//g;
                        $done16 = 1;
                        if ($item_info3 eq "")
                        {
                          $done16 = 3; # I don't want the program to look any further if there isn't a second item;
                          $item_info3 = '.';
                        }
                    }
                     if ((index($clean, 'CITY:') != -1) && $count == 0 && $done4 == 1 && $done17 == 0) {
                        $item_info4 = lc(substr($clean, 7, $slength)); # This gets the item info and makes it lower case for ease of coding later;
                        # Trim extra whitespace from the item info;
                        $item_info4 =~ s/^\s+//; # From left;
                        $item_info4 =~ s/\s*$//; # From right;
                        $item_line = $lines;
                        $item_info4 =~ s/[|]+//g;
                        $done17 = 1;
                        if ($item_info4 eq "")
                        {
                          $done17 = 3; # I don't want the program to look any further if there isn't a second item;
                          $item_info4 = '.';
                        }
                    }
                     if ((index($clean, 'CITY:') != -1) && $count == 0 && $done16 == 1 && $done18 == 0) {
                        $item_info5 = lc(substr($clean, 7, $slength)); # This gets the item info and makes it lower case for ease of coding later;
                        # Trim extra whitespace from the item info;
                        $item_info5 =~ s/^\s+//; # From left;
                        $item_info5 =~ s/\s*$//; # From right;
                        $item_line = $lines;
```

```
                    $item_info5 =~ s/[|]+//g;
                    $done18 = 1;
                    if ($item_info5 eq "")
                    {
                      $done18 = 3; # I don't want the program to look any further if there isn't a second item;
                      $item_info5 = '.';
                    }
                  }
                  if ((index($clean, 'STATE:') != -1) && $count == 0 && $done19 == 0) {
                    $item_info6 = lc(substr($clean, 8, $slength)); # This gets the item info and makes it lower case for ease of coding later;
                    # Trim extra whitespace from the item info;
                    $item_info6 =~ s/^\s+//; # From left;
                    $item_info6 =~ s/\s*$//; # From right;
                    $item_line = $lines;
                    $item_info6 =~ s/[|]+//g;
                    $done19 = 1;
                    if ($item_info6 eq "")
                    {
                      $done19 = 3; # I don't want the program to look any further if there isn't a second item;
                      $item_info6 = '.';
                    }
                  }
                  if ((index($clean, 'STATE:') != -1) && $count == 0 && $done16 == 1 && $done20 == 0) {
                    $item_info7 = lc(substr($clean, 8, $slength)); # This gets the item info and makes it lower case for ease of coding later;
                    # Trim extra whitespace from the item info;
                    $item_info7 =~ s/^\s+//; # From left;
                    $item_info7 =~ s/\s*$//; # From right;
                    $item_line = $lines;
                    $item_info7 =~ s/[|]+//g;
                    $done20 = 1;
                    if ($item_info7 eq "")
                    {
                      $done20 = 3; # I don't want the program to look any further if there isn't a second item;
                      $item_info7 = '.';
                    }
                  }
                  if ((index($clean, 'ZIP:') != -1) && $count == 0 && $done21 == 0) {
                    $item_info8 = lc(substr($clean, index($clean, 'ZIP:')+4)); # This gets the item info and makes it lower case for ease of coding later;
                    # Trim extra whitespace from the item info;
                    $item_info8 =~ s/^\s+//; # From left;
                    $item_info8 =~ s/\s*$//; # From right;
                    $item_line = $lines;
                    $item_info8 =~ s/[|]+//g;
                    $done21 = 1;
                    $first_zip = 1;
                    if ($item_info8 eq "")
                    {
                      $done21 = 3; # I don't want the program to look any further if there isn't a second item;
                      $item_info8 = '.';
                    }
                  }
                  if ((index($clean, 'ZIP:') != -1) && $count == 0 && $done16 == 1 && $done22 == 0 && $first_zip == 0) {
                    $item_info9 = lc(substr($clean, index($clean, 'ZIP:')+4)); # This gets the item info and makes it lower case for ease of coding later;
```

```
735                         # Trim extra whitespace from the item info;
736                         $item_info9 =~ s/^\s+//; # From left;
737                         $item_info9 =~ s/\s*$//; # From right;
738                         $item_line = $lines;
739                         $item_info9 =~ s/[||]+//g;
740                         $done22 = 1;
741                         if ($item_info9 eq "")
742                         {
743                           $done22 = 3; # I don't want the program to look any further
                              if there isn't a second item;
744                           $item_info9 = '.';
745                         }
746                       }
747                       if ((index($clean, 'CUSIP:') != -1) && $count == 0 && $done23 == 0)
                           {
748                         $item_info10 = lc(substr($clean, 7, 17)); # This gets the item
                              info and makes it lower case for ease of coding later;
749                         # Trim extra whitespace from the item info;
750                         $item_info10 =~ s/^\s+//; # From left;
751                         $item_info10 =~ s/\s*$//; # From right;
752                         $item_line = $lines;
753                         $item_info10 =~ s/[||]+//g;
754                         $done23 = 1;
755                         if ($item_info10 eq "")
756                         {
757                           $done23 = 3; # I don't want the program to look any further
                              if there isn't a second item;
758                           $item_info10 = '.';
759                         }
760                       }
761                       if ((index($clean, 'CUSIP No.') != -1) && $count == 0 && $done24 ==
                           0) {
762                         # Trim extra whitespace from the item info;
763                         $clean2 = $clean;
764                         $clean2 =~ s/^\s+//; # From left;
765                         $clean2 =~ s/\s*$//; # From right;
766                         $item_info11 = lc(substr($clean2, 8, 18)); # This gets the item
                              info and makes it lower case for ease of coding later;
767                         # Remove any symbols from the line;
768                         $item_info11 =~ s/[\$#@~!&*()\[\];.,:?^ `\\\/]+//g;
769                         # Trim extra whitespace from the item info;
770                         $item_info11 =~ s/^\s+//; # From left;
771                         $item_info11 =~ s/\s*$//; # From right;
772                         $item_line = $lines;
773                         $item_info11 =~ s/[||]+//g;
774                         $done24 = 1;
775                         if ($item_info11 eq "")
776                         {
777                           $done24 = 3; # I don't want the program to look any further
                              if there isn't a second item;
778                           $item_info11 = '.';
779                         }
780                       }
781                       # Now look for the percentage ownership.;
782                       # NOTE: Ownership data is later cross-checked with Audit Analytics.;
783                       if ((((index(lc($clean), 'percent of class') != -1) || (index(lc(
                           $clean), 'item 13') != -1) || (index(lc($clean), 'percent  of
                           class') != -1) || (index(lc($clean), 'class represented') != -1))
                           && $count == 0 && $done32 == 0 && $percent_count < 25) || (
                           $percent_search == 1 && $percent_count < 25)) {
784                         # This will cause the loop to search to the next line if it
                              doesn't find the percentage held initially;
785                         $percent_count++; # This keeps track of how many lines we've
                              searched for the percentage for.  We want to max out at 3 lines;
786                         $percent_search = 1;
787                         $clean2 = $clean;
```

```
788                         # Trim extra whitespace from the item info;
789                         $clean2 =~ s/^\s+//; # From left;
790                         $clean2 =~ s/\s*$//; # From right;
791                         $percent = "%";
792                         $colon = ":";
793                         # First, start looking at string after the number 11 appears,
                            which comes up on the line for ownership size;
794                         $percent_location = index($clean2, $percent);
795                         $colon_location = index($clean2, $colon);
796                         if ($percent_location != -1)
797                         {
798                             $done32 = 1;
799                             $percent_search = 0;
800                             $percent_locationm = $percent_location - 6;
801                             $percent_length = 7;
802                             if ($colon_location != - 1) { $percent_locationm =
                                $colon_location + 1; }
803                             if ($percent_locationm < 0) { $percent_locationm = 0; }
804                             $percent_length = $percent_location - $percent_locationm;
805                             $percent_extract = lc(substr($clean2, $percent_locationm,
                                $percent_length));
806                             $percent_extract =~ s/[^0-9.]//g;
807                             $percent_extract =~ s/^\s+//; # From left;
808                             $percent_extract =~ s/\s*$//; # From right;
809                             $percent_extract =~ s/[|]+//g;
810                         }
811                     }
812                     #Now fook for fund source;
813                      if ((((index(lc($clean), 'source of') != -1) || (index(lc($clean),
                      'source  of') != -1) || (index(lc($clean), 'of funds') != -1) || (
                      index(lc($clean), 'of  funds') != -1)) && $count == 0 &&
                      $fund_source_count < 5) || ($fund_source_search == 1 &&
                      $fund_source_count < 11)) {
814                         # This will cause the loop to search to the next line if it
                            doesn't find the percentage held initially;
815                         $fund_source_count++; # This keeps track of how many lines
                            we've searched for the percentage for.  We want to max out at 3
                            lines;
816                         $fund_source_search = 1;
817                         # If this is its last try then make sure it stops looking;
818                         if ($fund_source_count == 11) { $fund_source_count = 0;
                            $fund_source_search = 0; }
819                         # Now check for the various types of funds;
820                         if ((index($clean, 'SC') != -1)) { $source_of_funds = 'SC';
                            $source_sc = 1; $fund_source_search = 0; $fund_source_count= 0; }
821                         if ((index($clean, 'BK') != -1)) { $source_of_funds = 'BK';
                            $source_bk = 1; $fund_source_search = 0; $fund_source_count= 0;
                            }
822                         if ((index($clean, 'AF') != -1)) { $source_of_funds = 'AF';
                            $source_af = 1; $fund_source_search = 0; $fund_source_count= 0;
                            }
823                         if ((index($clean, 'WC') != -1)) { $source_of_funds = 'WC';
                            $source_wc = 1; $fund_source_search = 0; $fund_source_count= 0;
                            }
824                         if ((index($clean, 'PF') != -1)) { $source_of_funds = 'PF';
                            $source_pf = 1; $fund_source_search = 0; $fund_source_count= 0;
                            }
825                         if ((index($clean, 'OO') != -1) || (index($clean, '00') != -1))
                            { $source_of_funds = 'OO'; $source_oo = 1; $done35 = 1;
                            $fund_source_search = 0; $fund_source_count= 0; }

826                     }
827                     #Now fook for type of reporting person;
828                      if ((((index(lc($clean), 'type of reporting') != -1) || (index(lc(
                      $clean), 'of reporting') != -1) || (index(lc($clean), 'reporting
                      person') != -1) || (index(lc($clean), 'reporting persons:') != -1)
```

```
                        || (index(lc($clean), 'reporting persons') != -1) || (index(lc($clean), 'reporting person:') != -1) || (index(lc($clean), 'reporting  person') != -1) || (index(lc($clean), 'type  of reporting') != -1)) && $count == 0 && $reporting_type_count < 11)
                        || ($reporting_type_search == 1 && $reporting_type_count < 11)) {
829                        # This will cause the loop to search to the next line if it doesn't find the percentage held initially;
830                        $reporting_type_count++; # This keeps track of how many lines we've searched for the percentage for.  We want to max out at 3 lines;
831                        $reporting_type_search = 1;
832                        # If this is its last try then make sure it stops looking;
833                        if ($reporting_type_count == 10) { $reporting_type_count = 0; $reporting_type_search = 0; }
834                        # Now check for the various types of funds;
835                        if ($clean =~ /\bBD\b/) { $reporting_type = 'BD'; $reporting_type_bd = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
836                        if ($clean =~ /\bBK\b/) { $reporting_type = 'BK'; $reporting_type_bk = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
837                        if ($clean =~ /\bIC\b/) { $reporting_type = 'IC'; $reporting_type_ic = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
838                        if ($clean =~ /\bIV\b/) { $reporting_type = 'IV'; $reporting_type_iv = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
839                        if ($clean =~ /\bIA\b/) { $reporting_type = 'IA'; $reporting_type_ia = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
840                        if ($clean =~ /\bEP\b/) { $reporting_type = 'EP'; $reporting_type_ep = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
841                        if ($clean =~ /\bHC\b/) { $reporting_type = 'HC'; $reporting_type_hc = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
842                        if ($clean =~ /\bSA\b/) { $reporting_type = 'SA'; $reporting_type_sa = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
843                        if ($clean =~ /\bCP\b/) { $reporting_type = 'CP'; $reporting_type_cp = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
844                        if ($clean =~ /\bCO\b/) { $reporting_type = 'CO'; $reporting_type_co = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
845                        if ($clean =~ /\bPN\b/) { $reporting_type = 'PN'; $reporting_type_pn = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
846                        if ($clean =~ /\bIN\b/) { $reporting_type = 'IN'; $reporting_type_in = 1; $reporting_type_search = 0; $reporting_type_count = 0; }


847                        if ($clean =~ /\bOO\b/ || $clean =~ /\b00\b/) { $reporting_type = 'OO'; $reporting_type_oo = 1; $reporting_type_search = 0; $reporting_type_count = 0; }
848                    }
849                }
850
851            ## End textual analysis;
852        }
853        # Perl counts a new line as a space, so I adjust for this;
854        $charsadjusted = $chars - $lines + 1;
855        print FILE2 "$filename|$item_info|$doccount|$cik_num|$item_info2|$item_info4|$item_info6|$item_info8|$item_info10|$item_info11|$cik_num_filer|$item_info3|
```

```
            $item_info5|$item_info7|$item_info9|$lines|$words|$charsadjusted|$filedate|
            $conformdate|$dateofchange|$agreement|$investment_contract|$stockholder_contract
            |$cooperation|$rights|$exchange|$covenant|$investor_contract|$any_agreement|
            $contract|$support_agreement|$investment_agreement|$stockholder_agreement|
            $merger|$acquisition|$tender|$tender_offer|$shelf|$percent_extract|$compensation
            |$exec_compensation|$salary|$source_sc|$source_bk|$source_af|$source_wc|
            $source_pf|$source_oo|$reporting_type_bd|$reporting_type_bk|$reporting_type_ic|
            $reporting_type_iv|$reporting_type_ia|$reporting_type_ep|$reporting_type_hc|
            $reporting_type_sa|$reporting_type_cp|$reporting_type_co|$reporting_type_pn|
            $reporting_type_in|$reporting_type_oo|$fcovenant|$acovenant|$ncovenant|
            $info_confidential|$info_confidentiality|$info_informationrights|
            $info_accessandinfo|$info_accesstoinfo|$info_accesstodocs|$info_visitinspect|
            $info_communication|$info_communicate|$info_communicationacct|$info_nda|
            $info_fullaccess|$info_books|$investment_purpose|$shareholder_contract|$roa|$roe
            |$net_income|$earnings|$revenue|$sales|$cash_flow|$interest_coverage|$interest|
            $net_worth|$debt|$leverage|$current_ratio|$debt_to_equity|$assets|
            $debt_to_cash_flow|$ebit|$ebitda|$debt_service|$book_value|
            $shareholder_agreement|$dividend|$payout|$distribut\n";
            print "$filename done. ($filecount of $totallines)\n";
            close (RES);
    }
    close (FILE2);
    close (FILE4);
    print "\nEnd\n\n";
```