[Query algorithms for unstructured data in MIMIC II are performed using NLP processing techniques. [Perl]

--The following is an example (also truncated) using text processing techniques of certain programming language to deal with unstructured data, which is referred to as “Natural Language Processing” in this book. Most medical researches that need information from clinician notes, nursing notes, or radiology reports do not require the program to have real natural language processing ability, but rather the ability to identify certain keywords in certain parts of the free text data. For example, certain medication in the “Home Medication” section of the discharge summaries, or certain diagnosis in the “Past Medical History” section of the discharge summaries. Please note that, if information is being extracted from the “History of Present Illness” section of discharge summaries or radiology reports, the program does have to be more sophisticated, since those text are more “unstructured” than the previous two examples.

#look for certain medication in “Home Medication Section” of the discharge summaries

# load the list of drugs that we are looking for...

sub load\_drugs {

while ($line = <DFILE>) {

#print $line;

$line = &remcr($line);

if (length($line) > 1) {

# $line =~ tr/[a-z]/[A-Z]/; # Convert the line to upper case.

$line = lc $line;

#PPI:

if ( $line =~ /omeprazole|prilosec|esomeprazole|nexium|pantoprazole|protonix|lansoprazole|prevacid|rabeprazole|aciphex/i ) {

# H2B:

if ( $line =~ /ranitidine|zantac|famotidine|pepcid|cimetidine|tagamet|axid|nizatidine/i ) {

# DIURETICS:

if ( $line =~ /furosemide|lasix|hydrochlorothiazide|hctz|spironolactone|aldactone|torsemide|demadex|acetazolamide|diamox|nizatidine|triamterene|dyrenium|bumetanide|bumex|ethacrynic|edecrin|eplerenone|inspra|amiloride|midamor|metolazone|mykrox|zaroxolyn|chlorthalidone|hygroton|thalitone/i ) {

$drugs{$line} = 0b0010;

} else { #should never happen

print "drugs list does not match what's in perl code '$line' \n";

print "$line \n";

die;

}

}

}

close DFILE;

}

sub remcr {

my ($line) = @\_;

while ($line =~ /[\n\r]$/){chop($line);}

return ($line);

}

sub find\_drugs {

my $meds = ""; # Contains the section header which indicates that we're still in a medications section

my $sect = "";

my $type = "";

my $ty = 0;

my $medgroup = 0;

my $found = 0;

my $group = 0;

my $admit = 'unk';

my $disch = 'unk';

my $other = 'unk';

my @words = ();

my $sect\_head\_line\_index = 0;

my $line\_count = 0;

$admit = 'unk'; $disch = 'unk'; $other = 'unk'; $ty = 0;

my $prev\_line\_blank = 0;

my $found\_home\_meds = 0;

while ($line = <FILE>) {

if ($line =~ /(\d+)\_:-:\_(\d+)\_:-:\_/) {

$hadmid = $1; $case = $2;

#print "$hadmid\n";

$found\_home\_meds = 0;next;

}

chomp($line);

$line\_count++;