

Parsing para informes de radiologia

Convirtiendo a TXT

Lo primero es obtener los informes en un pdf legible. Hay dos tipos de informes disponibles en PDF.

1. PDFs escaneados 📄
2. PDFs impresos desde *MSWord* o similar

Para los primero tenemos que usar OCR. Primero hay que convertirlos en una imagen mas o menos grande y sobre esta ejecutar el reconocimiento de caracteres,

```
$ convert -density 300 -units PixelsPerInch informe.pdf tmp.tif
$ tesseract tmp.tif informe
```

En el segundo caso, esto no funciona. Pero el texto en si se puede facilmente,

```
$ pdftotext informe.pdf informe.txt
```

En el caso de la v0, cuando proceden de *MSWord* el archivo .docx existe en el directorio. Entonces es facil hacer,

```
for x in facehbi_informes_mri/v0/*.pdf; do n=$(echo ${x} | awk -F"/" '{print $3}'); if [[ -e ${x%.pdf}.docx ]]; then pdftotext ${x} txts/v0/${n%.pdf}.txt; else convert -density 300 -units PixelsPerInch ${x} tmp.tif; tesseract tmp.tif txts/v0/${n%.pdf}; fi; done
```

Variables a buscar

[parse.pl](#)

```
#!/usr/bin/perl

use strict;
use warnings;
use File::Find::Rule;
use File::Basename qw(basename);
use Data::Dump qw(dump);

my %rois = ('medial' => 'ATM',
           'cortical parietal' => 'ACP');
my $parse_dir = shift;
my @txts = find(file => 'name' => "*.txt", in => $parse_dir);
my %info;
foreach my $report (sort @txts){
    my $name = basename $report;
    $name =~ s/\.txt$//;
}
```

```

$info{$name}{'Fazekas'} = "NA";
open IDF, "<$report";
my $this_line;
while(<IDF>) {
    foreach my $roi (sort keys %rois){
        if (/ $roi/){
            if (/bilateral/){
                if(/ \d\s*bilateral/){
                    my ($grado) = /(\d)\s*bilateral/;
                    $info{$name}{$rois{$roi}.'_I'} = $grado;
                    $info{$name}{$rois{$roi}.'_D'} = $grado;
                }
                if(/bilateral\s*grad\w\s*\d/){
                    my ($grado) = /bilateral\s*grad\w\s*\d*-(\d)/;
                    $info{$name}{$rois{$roi}.'_I'} = $grado;
                    $info{$name}{$rois{$roi}.'_D'} = $grado;
                }
            }
            if (/izqui/){
                if(/ \d\s*izqui/){
                    my ($grado) = /(\d)\s*izqui/;
                    $info{$name}{$rois{$roi}.'_I'} = $grado;
                }
            }
            if (/dere/){
                if(/ \d\s*dere/){
                    my ($grado) = /(\d)\s*dere/;
                    $info{$name}{$rois{$roi}.'_D'} = $grado;
                }
            }
        }
    }
    if (/azekas\s*\d/){
        (my $fazekas) = /azekas\s*\d*-(\d)/;
        $info{$name}{'Fazekas'} = $fazekas;
    }
    if (/cortical global/){
        if(/ \d/){
            my ($grado) = / \d*-(\d)/;
            $info{$name}{'ACG'} = $grado;
        }
    }
}
close IDF;
}

print "Subject,";
foreach my $roi (sort keys %rois){
    my $ch = $rois{$roi}.'_I','.$rois{$roi}.'_D',';
    print "$ch";
}

```

```
print "ACG,Fazekas\n";
foreach my $subject (sort keys %info){
    my $fazekas =
exists($info{$subject}{'Fazekas'})?$info{$subject}{'Fazekas'}:"NA";
    my $acg =
exists($info{$subject}{'ACG'})?$info{$subject}{'ACG'}:"NA";
    print "$subject,";
    foreach my $roi (sort keys %rois){
        my $ad =
exists($info{$subject}{$rois{$roi}.'_D'})?$info{$subject}{$rois{$roi}.'_D'}:"NA";
        my $ai =
exists($info{$subject}{$rois{$roi}.'_I'})?$info{$subject}{$rois{$roi}.'_I'}:"NA";
        print "$ad,$ai,";
    }
    print "$acg,$fazekas\n";
}
```

From:
<https://mail.fundacioace.com/wiki/> - **Detritus Wiki**

Permanent link:
https://mail.fundacioace.com/wiki/doku.php?id=parse_informe

Last update: **2021/04/22 16:05**

