

# Extract Important Information

```
function infolist = ExtractImportantInfo(filename, previousInfo)
```

## Function to create or add onto a spreadsheet of PO information

### Includes school name, PO number, and total PO amount

```
%Bring PDF contents into matlab for use
filecontents = extractFileText(filename);

%if this is the first bit of information we're collecting
%add placeholder to start the list
if size(previousInfo) == 0
    previousInfo = ["000" "000000000" "0000"];
else
    previousInfo = previousInfo;
end

%find marker in each PO to count number of POs in PDF
%findmarker: list of locations of marker in each PDF
findmarker = strfind(filecontents, 'Where applicable, these instructions to'...
    'contractors supersede those in PART 1. ');
%if the marker isnt there, there's only one PO
if isempty(findmarker)
    PDFendPrelim = strlength(filecontents);
else

    PDFendPrelim = findmarker;
end
%ID number of PDFs
numPDFs = length(PDFendPrelim);

%Go through each PO one by one
for pag = 1:numPDFs
    %if it's the first page
    if pag == 1
        %if the marker isnt there
        if isempty(findmarker)
            %the length of the PDF is the length of the string
            PDFend = PDFendPrelim;
        %if the marker is there
        else
            %the PDF ends 77 characters after the marker
            PDFend = PDFendPrelim(pag)+77;
        end
        PDFbeg = 1;

        %if it's not the first page
    else
```

```

    %PDF begins 80 characters after the last one ends
    PDFbeg = PDFendPrelim(pag-1)+80;
    %PDF ends 77 characters after the marker
    PDFend = PDFendPrelim(pag)+77;

end
%content of PO found between beginning and end location
pagerange = extractBetween(filecontents, PDFbeg, PDFend);

```

## Extract PO number

```

lookforpo = 'WO2000';
if isempty(strfind(pagerange, lookforpo))
    lookforpo = 'WO1900';
end
if isempty(strfind(pagerange, lookforpo))
    lookforpo = 'WO1901';
end
%Pull out PO number
POnum = strfind(pagerange, lookforpo);
PObegin = [];
PObegin = POnum(1);
POend = [];
POend = PObegin+10;
%PONumber = [];
PONumber = extractBetween(pagerange, PObegin, POend);

%If we already have the PONumber recorded return the PO Number
if sum(contains(previousInfo(:,2), PONumber)) > 0
    PONumber

    if pag == numPDFs
        infolist = ["000" "000000" "000"];
    end
    continue

else

```

## Extract total PO amount

```

%Extract total PO amount

dollabillz = strfind(pagerange, '$');
billindic = strfind(pagerange, 'Total Amount');
begintotal = [];
begintotal = dollabillz;
endtotal = [];

```

```

endtotal = billindic(1);
changedend = endtotal - 3;
%string that contains PO amount
PrelimPOamount = extractBetween(pagerange, changedend-9, changedend);
%pull out only string containing amount
totalAmount = extractAfter(PrelimPOamount, '$');

%Extract school name
finddistrict = strfind(pagerange, 'District');
districtbegin = finddistrict(1);
schoolnamebegin = districtbegin + 110;
schoolnameend = schoolnamebegin+3;
%findschoolname = strfind(filecontents, 'HS01');
SchoolName = extractBetween(pagerange, schoolnamebegin, schoolnameend);

```

```
end
```

## Output important information to a list

```

%output important information to a list
infolist(pag,:) = [SchoolName, PONumber, totalAmount];

```

```
end
```

```
end
```