

# Form extractor from word to Excel

Last Updated Thursday, 18 March 2010

I had a bunch of filled out word documents with word forms in them and needed them in Excel. initially i tried CSV but it didn't play nice with encodings. So I decided to write directly to XLS.

"""

Copyright 2009 Konrads Smelkovs  
 UTF8Recorder and UnicodeWriter come from python docs  
 """

```
import sys,os, csv
import win32com.client
import pywintypes

class ExcelWriter(object):
    def __init__(self,excelfile):
        self.excelapp=win32com.client.DispatchEx('Excel.Application')
        self.excelapp.Visible=0
        self.excelapp.Application.AskToUpdateLinks=0
        self.workbook=self.excelapp.Workbooks.Add()
        os.unlink(excelfile) #TODO: remove for release
        self.workbook.SaveAs(excelfile)
        # Only worksheet 1 is used.
        self.worksheet=self.workbook.Worksheets.Item(1)
        self.currentrow=1

    def _getrow(self,row):
        """Convert integer row index to Alphabetical:
        1 -> A
        2 -> B
        ...
        """
        if row>sys.stderr,"Range: %s" % range
            cell=self.worksheet.Range(range)
            cell.Value=data[col-1]
            self.currentrow+=1

def main():
    if len(sys.argv)> sys.stderr, "Processing %s" % docfile
        worddoc=wordapp.Documents.Open(os.path.join(directory,docfile))
        for i in range(1,worddoc.FormFields.Count+1):
            try:
                form=worddoc.FormFields.Item(i)
                name=form.Name
                value=form.Result
                thisdocresults.append((name,value))
            try:
                print >>sys.stderr, "%s: %s" % (name,value)
            except UnicodeEncodeError,e:
                print >>sys.stderr, "Error decoding charset,%s" % e
            except pywintypes.com_error,e:
                print >>sys.stderr, "Exception: %s" % str(e)
            results.append(thisdocresults)
        worddoc.Close()
    wordapp.Quit()
    writer=ExcelWriter(os.path.abspath(sys.argv[2]))
    print >>sys.stderr, "Writing to Excel"
    for docres in results:
        data=[]
        for (n,v) in docres:
            data.append(v)
        writer.writerow(data)

if __name__=="__main__":
```

main()