```
#!python
# -*- coding:utf8 -*-

import re,sys

num = 0
def spannumber(m):
    global num
    span = m.group(1)
    if span == 'span':
        num += 1
        return '<span' + str(num)
    else:
        temp = '</span' + str(num)
        num -= 1
        return temp

file_name = sys.argv[1]
input_file = open(file_name, "rU")
all = input_file.read()
input_file.close()

all = unicode(all, "sjis")

re_kaigyo = re.compile(r'¥n{2,}')
all = re_kaigyo.sub('{¥kaigyou}', all)
all = all.replace("¥n", " ")
all = all.replace("{¥kaigyou}", "¥n¥n")

re_generator = re.compile(r'<(meta name|link).+?>')
all = re_generator.sub('', all)

all = all.replace('<style>','<meta http-equiv="content-style-type" content="text/css">')

all = all.replace('</style>','<link rel="stylesheet" type="text/css" href="style.css">')

re_two_spaces = re.compile(r' {2,}')
all = re_two_spaces.sub(" ", all)

re_comment = re.compile(r'<!--.+?-->')
all = re_comment.sub('', all)

re_nbsp = re.compile(r'( )+')
all = re_nbsp.sub(' ', all)

re_xmlns = re.compile(r' xmlns.+?>')
all = re_xmlns.sub('>', all)

re_blocktag = re.compile(r'</(p|li|div|table|tr|th|body|head|h.+?|title)>')
all = re_blocktag.sub("¥g<0>¥n¥n", all)

re_blocktag2 = re.compile(r'(<(body.+?|link.+?|br|head|div.+?|html.+?|!DOCTYPE.+?|>))')
all = re_blocktag2.sub("¥g<0>¥n¥n", all)

re_blocktag3 = re.compile(r'(<(meta.+?)>)')
all = re_blocktag3.sub("¥g<0>¥n", all)

re_if = re.compile(r'<!¥[(if.+?|endif)¥]>')
all = re_if.sub('', all)

re_div_style = re.compile(r'<div.+style=¥'layout-grid.+>')
all = re_div_style.sub('', all)

re_class = re.compile(r' (class|lang)=[^ >]+')
all = re_class.sub('', all)

re_html = re.compile(r'<html>')
all = re_html.sub('<html lang="ja">', all)

re_text = re.compile(r'(text¥-|mso¥-|margin¥-|font¥-family:|background:|tab¥-).+?¥'')
all = re_text.sub("'", all)

re_color_black = re.compile(r'(style=¥')?color:black.+?¥'')
all = re_color_black.sub("",all)

re_span_langdir = re.compile(r'<span (lang|dir).+?>')
all = re_span_langdir.sub("<span>",all)

re_kigou1 = re.compile(r'¥&(sup2|Oslash);')
all = re_kigou1.sub("",all)

re_msorm = re.compile(r' ¥!msorm')
```

```
all = re_msorm.sub("",all)

re_op = re.compile(r'</?o:p>')
all = re_op.sub("",all)

re_null_quotation = re.compile(r'[-a-z0-9]+=¥'¥'')
all = re_null_quotation.sub('', all)

re_span_style = re.compile(r"<span style='font-[^>]+?'></span>")
all = re_span_style.sub('', all)

all = all.replace(' >', '>')

re_spannum = re.compile(r'<(/?span)')
all = re_spannum.sub(spannumber, all)

re_null_span = re.compile(r'<span(-?¥d+)>(.+?)</span¥1>')
m = re_null_span.search(all)
while m:
    all = re_null_span.sub('¥g<2>',all)
    m = re_null_span.search(all)
re_unspannum = re.compile(r'(</?span)-?¥d+')
all = re_unspannum.sub('¥g<1>', all)

re_div = re.compile(r'</?div.*?>')
all = re_div.sub('', all)

re_karatag = re.compile(r'<([^ >]+)[^>]*> *</¥1>')
m = re_karatag.search(all)
while m:
    all = re_karatag.sub('',all)
    m = re_karatag.search(all)

all = re_kaigyo.sub('¥n¥n', all)

all = '<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">¥n' + all

all = all.encode('sjis')
print all
```