

```
/Applications/Anaconda-Navigator.app/Contents/MacOS/python /Users/kycheung/Desktop/Capstone-master/evaluation_word_doc.py
```

Character accuracy in Tesseract/Annual_report

```

Characters predicted 199134.0
Errors              206.0
Character accuracy   99.9

```

Throughput in Tesseract/Annual_report

```

Characters predicted 199134.00
Errors              206.00
OCR time            332.54
Page                174.00
OCR time per page   1.91

```

	Penalty	Throughput
0	0	598.83
1	1	598.21
2	2	597.59
3	3	596.97
4	4	596.35
5	5	595.73
6	6	595.11
7	7	594.49
8	8	593.87
9	9	593.25

Accuracy by Character Class in Tesseract/Annual_report

```

Characters predicted 199134
Errors 206

```

	Class_type	Count_predicted	Count_error	Accuracy
0	Lowercase	180565	97	99.95
1	Uppercase	7122	10	99.86
2	Number	7785	48	99.38
3	Symbol	3662	51	98.61

Accuracy by frequency in Tesseract/Annual_report

	Character	Count_predicted	Count_error	Accuracy_by_frequency
0	!	1	0	100.00
1	"	93	5	94.62
2	#	5	1	80.00
3	\$	477	3	99.37
4	%	127	3	97.64
5	&	73	0	100.00
6	'	172	8	95.35
7	(439	14	96.81
8)	440	2	99.55
9	*	4	0	100.00
10	+	24	0	100.00
11	,	1614	5	99.69
12	/	45	0	100.00
13	0	1855	2	99.89

14	1	1437	11	99.23
15	2	1495	14	99.06
16	3	440	6	98.64
17	4	401	2	99.50
18	5	514	2	99.61
19	6	459	3	99.35
20	7	385	4	98.96
21	8	388	2	99.48
22	9	411	2	99.51
23	:	37	0	100.00
24	;	95	0	100.00
25	@	6	0	100.00
26	A	505	3	99.41
27	B	254	0	100.00
28	C	413	0	100.00
29	D	265	0	100.00
30	E	325	0	100.00
31	F	342	0	100.00
32	G	352	0	100.00
33	H	538	0	100.00
34	I	621	6	99.03
35	J	22	0	100.00
36	K	414	0	100.00
37	L	143	0	100.00
38	M	289	0	100.00
39	N	109	1	99.08
40	O	164	0	100.00
41	P	382	0	100.00
42	Q	10	0	100.00
43	R	317	0	100.00
44	S	523	0	100.00
45	T	745	0	100.00
46	U	86	0	100.00
47	V	64	0	100.00
48	W	210	0	100.00
49	X	5	0	100.00
50	Y	21	0	100.00
51	Z	3	0	100.00
52	^	10	10	0.00
53	a	13327	9	99.93
54	b	2003	0	100.00
55	c	6646	5	99.92
56	d	6326	5	99.92
57	e	22792	16	99.93
58	f	3937	4	99.90
59	g	4135	3	99.93
60	h	6102	0	100.00
61	i	14648	4	99.97
62	j	229	0	100.00
63	k	633	0	100.00
64	l	8157	0	100.00
65	m	4673	1	99.98
66	n	15753	6	99.96
67	o	14645	8	99.95
68	p	4649	0	100.00
69	q	165	0	100.00
70	r	12407	9	99.93

71	s	10831	5	99.95
72	t	15607	8	99.95
73	u	5216	7	99.87
74	v	2795	3	99.89
75	w	2032	0	100.00
76	x	618	4	99.35
77	y	2190	0	100.00
78	z	49	0	100.00

Grouped error in Tesseract/Annual_report

Count_error

Character Predicted

"	{}	5
#	{}	1
\$	8	3
%	{}	3
'	{}	8
(2	1
	G	6
	Q	1
	d	1
	{}	5
)	{}	2
,	{}	5
0	1	1
	{}	1
1	0	1
	2	2
	[1
	{}	7
2	1	1
	9	1
	{}	12
3	4	1
	{}	5
4	{}	2
5	{}	2
6	{}	3
7	1	1
	{}	3
8	{}	2
9	{}	2
A	{}	3
I	1	1
	D	2
	[1
	{}	1
		1
N	{}	1
^	{}	10
a	{}	9
c	{}	5
d	d~	1
	{}	4
e	{}	16
f	n	1

	{}	3
g	{}	3
i	t	1
	{}	3
m	{}	1
n	m	1
	{}	5
o	e	1
	{}	7
r	m	1
	{}	8
s	S	2
	{}	3
t	{}	8
u	U	1
	{}	6
v	{}	3
x	{}	4

Note: {} indicates no prediction made by the OCR engine
Number of missed prediction: 171

Error table with prediction confidence for Tesseract/Annual_report

	Character	Predicted	OCR confidence
0	I	D	0.0
1	d	d~	0.0
2	(d	0.0
3	f	n	0.0
4	I		19.0
5	(Q	22.0
6	n	m	28.0
7	I	[42.0
8	(2	43.0
9	\$	8	44.0
10	\$	8	44.0
11	\$	8	44.0
12	o	e	53.0
13	3	4	53.0
14	2	9	61.0
15	7	1	65.0
16	(G	68.0
17	1	[68.0
18	(G	68.0
19	(G	68.0
20	(G	71.0
21	I	D	71.0
22	1	2	71.0
23	(G	72.0
24	(G	72.0
25	r	m	75.0
26	i	t	76.0
27	I	1	82.0
28	s	S	92.0
29	u	U	92.0
30	s	S	92.0
31	1	2	95.0

32	0	1	95.0
33	2	1	96.0
34	1	0	96.0

Average confidence: 58.23

Process finished with exit code 0