

紙の統計表をテキストデータ化する ネ申Excelのルーツ・芸術・情報抽出

2021年12月3日@2021年度「NDLデジタルライブラリーカフェ」
有本 寛（一橋大学）
arimotoy@ier.hit-u.ac.jp

自己紹介

- 名前: 有本 寛 (一橋大学経済研究所)
- 仕事: 経済学者 (開発経済学, 経済史, 農業経済学)
 - 現地フィールド調査でデータ収集の例
 - ベトナムで粗製肥料, マダガスカルで米流通
 - 歴史統計データを使った分析の例
 - 『帝国統計年鑑』の府県別データ
 - 官報の行旅死亡人公告データ (身長データの収集)
- 好きなエディタ: `vscode` > `emacs`
- 普段使っている言語: `Stata`, `Python`, `R`

はじめに

- 社会科学分野でも, 統計データ分析の重要性が増している
 - 日本には, 140年以上に渡る長期の政府統計がある
 - ただし, ごく近年まで紙媒体, よくて電子画像で提供.
 - 電子画像の統計表を, Excelファイルにテキストデータ化したい
 - 手入力? 膨大な時間と費用がかかる → 21世紀なので機械で
1. どこまで機械的に統計表を情報抽出できるか紹介
 2. その過程で出会った, 元祖「ネ申Excel」を観賞する

- はじめに
- **OCRの活用と課題**
- ocrstatsの概要
- 元祖ネ申Excelとの邂逅
- 今後の課題と展望

OCRの活用と課題

- OCR (optical character recognition: 光学的文字認識)
- 統計表をOCRする際のポイント
 1. 文字(列)を正確に認識する
 2. 文字(列)を正確な位置 (セル) に配置する

表レイアウト解析

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4

B3

	A	B	C	D	E	F	G	H
1			キロワット					
2	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
3	94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
4	715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
5	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
	104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
	241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
	638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
	118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
	75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
	191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
	5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
	2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
	8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
	6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
	9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
	20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
	22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
	26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
	15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
	19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
	27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
	15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
	24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
	503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
	108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
	39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
	29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
	24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
	21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
	42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
	36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
	25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4

統計表をOCRにかけた例

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275	1	1
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511	4,737	6,409	15	1,762
715,021	2,251,597	62,078	19,374,221	22.7	105.8	1,083.4	9,324.3	7,828	8,881	21,543	28,521	151	5,551
115,250	265,434	7,406	2,617,500	6.5	64.2	162.4	1,601.8	1,876	1,975	5,083	6,743	5	534
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9	1,234	1,339	2,701	3,592	16	4,419
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1	3,489	3,637	9,067	12,057	24	3,279
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4	6,457	6,860	21,916	29,894	11	380
118,978	273,262	6,201	2,471,397	5.8	48.2	137.2	1,240.8	726	787	3,006	3,907	166	163
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062	6	29
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703	7	4,802
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20	—	—
2,844,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,876	79,964	107,273	102	20,925

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275	1	1
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511	4,737	6,409	15	1,762
715,021	2,251,597	62,078	19,374,221	22.7	105.8	1,083.4	9,324.3	7,828	8,881	21,543	28,521	151	5,551
115,250	265,434	7,406	2,617,500	6.5	64.2	162.4	1,601.8	1,876	1,975	5,083	6,743	5	534
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9	1,234	1,339	2,701	3,592	16	4,419
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1	3,489	3,637	9,067	12,057	24	3,279
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4	6,457	6,860	21,916	29,894	11	380
118,978	273,262	6,201	2,471,397	5.8	48.2	137.2	1,240.8	726	787	3,006	3,907	166	163
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062	6	29
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703	7	4,802
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20	—	—
2,844,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,876	79,964	107,273	102	20,925

統計表をOCRにかけた例

JUST PDF4 (!!)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1			IFN										777.	
2	43.641	128.961	3.4671	1.438,5311	7.81	87.2	21.2	236.0			524	3.214	4.1	
3	43.641	256.954	6.708		4.5	44.9	80.2					4.737	6	
4		2,251.59	2,520.59				593.5					1,433	7,823E11	21,243.3
5	3	7	11			22.7	198.8	1,083.4					15	151 5
6		965.434	62.07811		6.51		9.374.31						3.069	0
7	1	244.913	5.723		7.71	80.5	368.4	3.270.11	1.234	3.4891		2.701	3	16 24
8	638.282	1.510.96	42.231	14.141.1251	208.4		1,026.79	5,484.4	6.4571			21.911	29.11	26
9	118.978	273.262	16.201	2.471.393	5.31	48.2	137.2	1,240.8	726			3.006	3	166
10	75.9901	179.064	4.021	1.695.0531	5.5	32.1	15.7	1,435.7	588.			2.302	3.6	
11	191.5251	483.199	10.108	4,999.445	6.3	64.9	180.9	1,871.8	1.188			6.394	5, 7	
12	5.670	11.0991	204	82.5521	2.1	15.3	70.7	526.1	2.			15	20	
13	2,344.735:	6,127.421	160,865	56,726.6761	11.6	107.2	248.0	2,287	25,27.376			79.96	1	402
14	43.641	128.961	3.467	1.438.3311	7.8	87.2	21.2	23			524	573	170	4,275 935
15	8.097	40.001	800		3.9	38.7	90.1	6.0			140			
16	6.993	21.507	344	222.590.	2.6	26.9	23.9	247.5			49	51		214
17	9.433	24.582	574	248.489.	2.7	27.0	32.6	329.6			60	69		68
18	20.854	47.811	1.203.	488.888	5.0	51.4	79.7	814.6			323	330		1,419
19	22.541	67.881	1.8901	600.428	7.9	65.7	125.5	1,110.3			4350	351		1,344
20	26.509	71.049	2.088.	670.834	5.6	53.2	84.0	792.9			506	510		2,029
21	15.731	40.987	981	544.706	3.2	42.2	106.4	1,414.21			1641	1691		566
22	19.946	50.374	822	521.084	3.1	32.3	122.3	1,265.3			210.	221		1,107
23	27.653	69.412	1.241	683.354	7.1	68.4	170.4	1,678.0			3091	319		1,819
24	15.474	37.9911	6431		2.9		143.1	11			1551			761
25	24.509	65.2901	1.177	373.219	4.8	28.6	200.2	40			207	172	2091	414
26	503.227	1,683.65	47.976	14,007.656	59.9	44.4	10,322.2	110,872.71	5.938			359	19,296	
27	108.481	303,888	9.265	2,637,395.	26.5	230.3	1,952.1	16,942.2	1.385			3,426		4,558
28	39,754	101,454	2.705		5.4	60.9	123.0	1.3			592	301		2,098
29	29.878	58.997	1.586	1,144.249	7.3	65.9	221.5	87.						2,801 982
30	24.2721	61.422	2.015	509.960	7.5	60.5	226.9	2,124.3:			306	739		1,762
31	21.346	43.561	1.103	388.194	6.8	60.9	159.9	1,423.11			587	909		1,698
32	42.699	122.236	3.394	1,471.821	8.4	101.6	143.2	1,723.8:	434	1.233		1,643		1,135
33	36,345	71,381	1.706	626.2671	6.7	58.8	106.3	932.7:	552		854			
34	25.508	51.296	623	464.781	7.6	69.2	198.5	98.				612		814
35	14.951	36,3191	961	349,412'	6.2	59.7	125.3	4.		248 147		882		509
36														
37	89,331	186.109	4.146	1,787.704	12.7	122.3	369.4	3.5		1,157 1,838	327	3,630		4,825 5,585 1,138
38	78,286	210,985	5,949	2,866,0861	17.3	234.4	711.5	9,664.8		992		4,001		6,053
39	142.693	314,804	8,469	2,808,627	18.8	137.1	582.6	4,966.3		1,169		4,153		5,528
40			26,309	7,388,938	40.3	339.6	7,682.9	63.					11,740	
41	368,219	877,496	874	492,785.	8.6	84.9	240.1	80			3,612	275 409	720	15,625 958 1,730
42		630		584,6891	7.6	77.2	183.1	1.9				1,302		
43	10,919	24,4031	557	237,3291	5.4	52.1	108.9	1.0					237	316
44	19,455	39,467	960	311,597	5.3	41.5	90.6	58.		98 69 35		423	961	
45	53,258	118.261	3,056	1,056,17511	7.3	36.5	113.8	2,028.1		488		1,987		2,643
46														
47	16.581	43.2381	862		4.1	39.1	110.9	64.				147		
48	11,550	27.064	542	265,976	4.7	36.7	99.8	9		36101		353		196 469
49		51.996												
50	23,992	99,0481	1,668	599,660	6.9	80.6	454.5	1		129 204 154	712		948 1,308 337	
51	20,918	198,8851	985	567,093:	5.6	63.0	123.3			120	1391 458			
52	83,826		4,754	2,039,922:	11.0	112.8				414				2,814
53														
54	13,973	44,1561	700	450,231	6.5	66.3	5	2.8		108		117 39		1,417
55	8,001	29,7381	585	376,157	2.3	29.4	63.9	81		89		89 101		250
56	7,624	20,738	602	219,758	3.5	37.5	42.6	8.1		84				771
57	21,5481	53,066	1,170	500,3361	3.9	36.6	88.1	830.7		220		245		1,625
58	5,6701	11,099.	204		2.1	15.3	70.7	526.1		2				20
59	2,344.73516.	148.4201	160,86556.	726.676.	11.6	107.2	248.012	287.	25,27.3761			79.964	1	402

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
				α α' y bj								r 4 α v bj			
2	43641	128961		3.467	1438531	7.8	87.2	21.2	236	524	573	3,214	4275	1	bj 1
3	94357	256554		6708	2520591	4.5	44.9	60.2	593.5	1433	1511	4737	6409	15	1762
4	715021	2251597		62078	19374221	22.7	195.8	1083.4	9324.3	7828	8881	21543	28521	151	5556
5	115250	265434		7409	2617500	6.5	64.2	162.4	1601.8	1876	1975	5069	6743	5	534
6	104552	244913		5723	2562869	7.7	80.5	137.3	1436.9	1234	1339	2701	3592	16	4419
7	241469	543375		12715	4823396	10.7	94.8	368.4	3270.1	3489	3637	9067	12057	24	3279
8	638282	1510962		42231	14141125	22.4	208.4	1026.7	9548.4	6457	6860	21916	29894	11	380
9	118978	273262		6201	2471393	5.3	48.2	137.2	1240.8	726	787	3006	3997	166	163
10	75990	179064		4021	1,695,053	5.5	52.1	151.7	1435.7	588	623	2302	3062	6	29
11	191525	483199		10108	4999445	6.3	64.9	180.9	1871.8	1074	1188	6394	8703	7	4802
12	5670	11099		204	82552	2.1	15.5	70.7	526.1	2	2	15	20		
	2344735	6127420		160865	56726676	11.6	107.8	248.0	2287.9	25231	27376	79964	102723	402	20925
13															
14	43641	128961		3467	1438531	7.8	87.2	21.2	236	524	573	3214	4275	1	1
15	8027	23716		600	289362	3.2	38.7	39.1	476.7	140	170	703	935	1	1
16	6993	21507		344	222590	2.6	26.9	23.9	247.5	49	51	161	214	3	180
17	9133	24582		574	218,489	2.7	27	32.6	329.6	60	69	352	468	—	—
18	20854	47811		1203	488888	5	51.4	79.7	814.6	323	330	1067	1419	—	—
19	22541	67889		1899	600,428	7.4	65.7	125.5	1110.3	655	351	1010	1344	4	680
20	26509	71049		2088	670834	5.6	53.2	84	792.9	506	510	1444	2029	7	901
21	15731	40987		951	544706	3.2	42.2	106.4	1414.2	164	169	425	566	—	—
22	19946	50374		822	521084	5.1	52.3	122.3	1265.5	210	221	831	1107	—	—
23	27653	69412		1242	683354	7.1	69.4	170.4	1678	309	319	1324	1819	—	—
24	15474	37991		645	373219	2.9	28.6	143.1	1406.0	155	172	670	761	—	—
25	24509	65290		1177	606807	4.8	44.4	200.2	1860.5	207	209	359	414	—	—
26	503227	1683655		47976	14007656	59.9	498.6	13322.2	110873	5398	6175	14508	19296	123	866
27	108481	303888		9265	2637395	26.5	230.3	1952.1	16942.2	1385	1616	3426	4558	28	4690
28	39754	101454		2705	1144249	5.4	60.9	123	1387.7	592	629	2098	2801	2	224
29	29878	58997		1586	509960	7.5	65	221.5	1914.2	301	306	739	982	—	—
30	24272	61422		2015	575097	7.9	73.8	226.9	2124.3	396	428	1323	1762	3	310
31	21346	43561		1103	388194	6.8	60.9	159.9	1425.1	587	612	909	1198	—	—
32	42699	122236		3394	1471821	8.4	101.6	143.2	1723.9	434	522	1235	1643	16	4419
33	36345	71381		1706	626267	6.7	58.8	106.3	932.7	552	567	854	1135	—	—
34	25508	51296		623	464781	7.6	69.2	198.5	1798.4	248	250	612	814	—	—
35	14951	36319		961	349412	6.2	59.7	125.3	1205.5	147	156	382	509	—	—
36	89331	186109		4146	1787704	12.7	122.3	369.4	3548.3	1157	1207	3630	4825	2	100
37	96899	239550		5558	1,996,2451	12.2	101.7	765.9	6382.3	1858	1938	4199	5585	19	3172
38	40288	81397		2050	690035	8.3	64	220.9	1872.3	327	336	856	1138	3	7
39	78286	210985		5949	2866086	17.3	234.4	711.5	9664.8	992	1040	4001	6053	—	—
40	142693	314804		8469	2808627	15.8	137.1	582.6	4956.3	1169	1244	4153	5528	5	1
41	368219	877496		26309	7388938	40.3	339.6	7582.9	63851.9	3612	3882	11740	15625	6	379
42	23022	50178		874	492785	8.6	84.9	249.1	2446.6	275	275	720	958	—	—
				630		7.4	77.7	198.1		400	410				

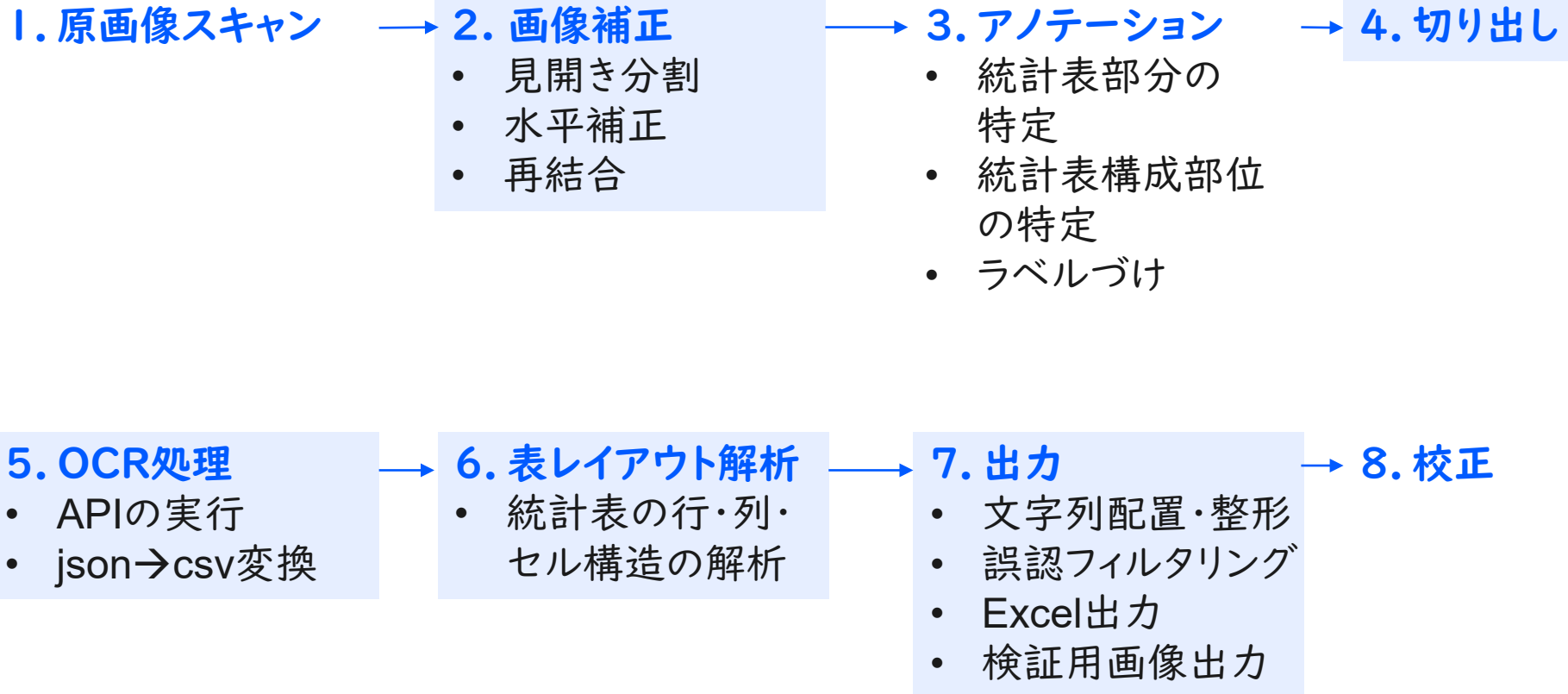
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
			「キロワット」								「キロワット」			「キロワット」	
2	43,641	126,961	3,467	1,438,531	7.0	87.2	21.2	236.0	524	573	3,214	4,275	1	1	
3	94,357	256,554	6,788	2,526,591	4.5	44.9	68.2	503.5	1,433	1,511	4,737	6,489	15	1,762	
4	715,821	2,251,597		6,287,819,374,221	22.7	195.8	1,683.3	9,324.3	7,828	8,881	21,543	28,521	151	5,556	
5	115,258	265,434	7,429	2,617,508	6.5	64.2	162.4	1,681.8	1,876	1,975	5,869	8,743	5	534	
6	104,352	244,913	5,783	2,622,869	7.7	80.5	137.3	1,436.9	1,284	1,339	2,701	3,592	18	4,419	
7	241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1	3,489	3,637	9,067	12,957	24	3,279	
8		6,382,821,510,962		42,231,14,141,125	22.4	208.4	1,026.7	9,564.8	6,457	6,868	21,916	29,894	11	388	
9	118,978	273,262	6,281	2,471,393	5.3	48.2	137.2	1,246.8	726	767	3,086	3,997	166	163	
10	179,064	4,021	1,695,958	5.5	52.1	151.7		1,435.7	588	623	2,382	3,062	6	29	
11	191,525	483,199	10,188	4,999,445	6.3	64.9	188.9	1,871.8	1,074	1,188	6,394	8,783	7	4,892	
12	5,879	11,099	204	82,552	2.1	15.5	70.7	526.2	2	2	15	20			
13	2,314,735	6,127,420	160,855	56,726,576	11.6	107.2		24,882,287.9	25,231	27,376	79,961	107,273	482	20,925	
14	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275	1	1	
15	8,827	23,716	606	260,362	3.2	38.7	39.1	476.7	140	178	783	935		1	
16	6,993	21,587	344	222,598	2.6	26.9	23.9	247.5	40	51	161	214	3	188	
17	9,433	24,582	574	248,489	2.7	27.0	32.6	329.6	86	69	352	468			
18	28,854	47,811	1,293	488,888	5.8	51.4	79.7	814.6	323	338	1,067	1,419			
19	22,541	67,889	1,899	609,428	7.4	65.7	125.5	1,110.3	355	351	1,818	1,341	4	688	
20	26,509	71,949	2,088	678,834	5.6	53.2	84.0	792.9	556	548	1,444	2,029	7	591	
21	15,731	46,987	951	544,796	3.2	40.2	186.4	1,414.2	164	160	425	566			
22	10,046	56,374	822	521,884	5.1	52.3	122.3	1,265.5	210	221	831	1,187			
23	27,653	69,412	1,242	693,354	7.1	69.4	170.4	1,678.0	399	319	1,324	1,819			
24	15,474	37,991	645	573,219	2.9	28.6		143.1	1,408.8	155	172	878			
25	24,509	65,290	1,177	606,807	4.8	44.4	280.2	1,869.5	287	290	359	414			
26	503,227	1,683,655	47,976	14,887,656	59.9	498.6		133,222,110,872.7	5,398	6,175	14,588	19,296	123	866	
27	168,481	383,888	9,265	2,637,395	24.5	238.3	1,952.1	16,942.2	1,385	1,616	3,426	4,558	28	4,698	
28	39,754	181,454	2,795	1,444,249	5.4	68.9	123.8	1,387.7	592	620	2,698	2,881	2	224	
29	29,878	58,997	1,586	590,968	7.5	65.0	221.5	1,914.2	381	386	799	982			
30	24,272	61,422	1,015	575,897	7.9	73.8	226.9	2,124.3	396	428	1,323	1,762	3	318	
31	21,316	43,561	2,183	388,194	6.8	68.9	159.9	1,425.1	587	612	909	1,288			
32	42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9	434	522	1,235	1,643	16	4,419	
33	36,345	71,381	1,766	626,267	6.7	57.8	186.3	932.7	552	567	854	1,135			
34	25,588	51,296	623	464,781	7.6	69.2	198.5	1,798.4	248	258	612	814			
35	14,951	36,319	961	399,412	6.2	59.7	125.3	1,285.5	147	156	382	509			
36	89,331	286,189	4,146	1,187,704	12.7	122.8	389.4	8,548.3	1,157	1,297	3,838	4,825	2	108	
37	96,899	239,550	5,558	1,896,245	12.2	101.7		765.96,382.3	1,858	1,938	4,199	5,585	19	3,172	
38	48,288	81,397	2,050	699,835	8.3	64.0	220.9	1,872.3	327	336	856	1,138	3	7	
39	78,265	216,985	5,949	2,866,866	17.3	234.4	711.5	96,668.0	992	1,048	4,881	6,653			
40	142,693	314,884	8,459	2,888,627	15.8	137.1	582.6	4,956.3	1,169	1,244	4,153	5,528	5	1	
41	368,219	877,496	26,399	7,788,938	48.3	399.6		7,582,963,851.9	3,612	3,882	11,748	15,625	8	379	
42	23,822	50,178	874	492,785	8.6	84.9		249.12,446.6	275	275	729	958			
43	28,862	57,499	630	584,889	7.6	77.2	185.1	1,882.3	459	418	1,302	1,739			
44	18,919	24,483	557	237,329	5.4	52.1	188.9	1,058.7	98	197	237	316			
45	10,455	39,467	968	311,597	5.3	41.5	98.6	715.0	69	71	423	561			
46	18,765	47,803	756	458,064	3.9	36.5		113.81,871.3	35	40	212	281			
47	53,258	118,261	3,056	1,856,175	7.3	659.0	227.1	2,028.1	488	525	1,987	2,643	166	163	
48	16,381	43,238	852	413,308	4.1	39.1	110.9	1,064.9	36	42	147	196			
49	11,556	27,064	542	265,978	3.7	35.7	99.8	980.4	101	108	353	469	5	25	
50	23,952	51,580	1,668	599,668	6.9	80.6	454.5	5,283.3	129	142	712	948			
51	27,364	69,948	1,112	1,16,397	6.3	47.2	282.4	1,513.6	264	216	984	1,388			
52	13,118	31,362	699	313,828	4.5	45.4	69.8	688.4	154	165	253	337	1	4	
53	26,918	50,457	985	567,893	5.6	63.0		125.3	1,488.1	120	139	1,219	1,621		
54	83,825	298,885	4,784	2,839,922	11.8	112.8		62,588,418.7	414	458	2,484	2,814			
55	13,973	44,156	790	459,231	6.5	68.3		275.8	2,812.5	188	117	1,034	1,477		
56	35,636	86,159	1,312	845,948	8.8	78.1	366.4	3,597.5	39	38	109	145			
57	8,801	29,738	585	376,157	2.3	29.4	63.9	868.1	89	88	188	250			
58	7,624	26,738	682	210,758	3.5	37.5	42.6	459.8	84	181	582	771			
59	21,548	53,066	1,178	508,336	3.9	38.6	88.1	830.7	220	245	1,222	1,625	7	4,892	
60	5,879	11,099	204	82,552	2.1	15.5	70.7	526.2	2	2	15	20			

- はじめに
- OCRの活用と課題
- **ocrstatsの概要**
- 元祖ネ申Excelとの邂逅
- 今後の課題と展望

ocrstatsの概要

- 何？
 - 統計表画像のテキストデータ化を体系的に支援するPythonツール
 - 大量の統計表を体系的にテキストデータ化することを想定
- 特徴・要件
 - (1) 一括処理, (2) 定型工程の自動化, (3) 外部OCRの活用
- 実用的な精度で(とりあえず)動いて使えるものを作る
 - 使える外部ツールは躊躇なく, 節操なく, なんでも使う
 - アノテーションツール, OCRエンジン

作業の流れ



※青枠はocrstatsで自動処理

原画像（見開き分割→水平補正→再結合後）

13. 電気事業及其事業

165. 原動力、電圧、發電力別事業数及發電力 (全國) 大正二年 續

		電氣供給及電氣線道事業				自家用電氣工作物		官廳施設 電氣工作物	合 計
		電氣供給	電氣線道	電氣線道 延長(キロメートル)	計	第一種	第二種		
		發 電 力 (單位: キロワット)							
原動力	水力	213,547	180	98,525	212,552	3,850	31,704	400	278,296
	火力	53,124	3,300	59,425	115,849	11,842	62,655	14,571	204,917
	風力	6,365	—	434	7,369	1,875	10,684	270	20,228
	計	273,036	3,480	88,384	365,800	17,567	105,133	15,241	509,541
	電氣力	50,737	3,870	55,557	90,164	1,496	17,147	3,937	112,744
電 壓	供電設備	464	2,680	6,335	9,479	17,367	22,796	6,960	56,092
	配電設備	52,769	800	41,549	95,118	—	53,650	4,881	153,640
	特別高壓	220,703	—	40,500	261,203	—	28,697	3,400	293,200
	計	273,036	3,480	88,384	365,800	17,367	105,133	15,241	509,541

二種以上ノ原動力ヲ用ヒタルモノハ其ノ主ナル方ニ計上セリ
電氣線道及電氣供給事業ノ事業數中鐵道ノ因襲ノモノヲ、供給ノ事業數中ノモノニテハ
第二種自家用電氣工作物ノ事業數中電氣事業法ヲ採用スルモノ同第三種又電氣力ノ一部ヲ使用シタル容量不明ノ事業數五ヶ所計
ノニ加算セリ

166. 事業別發電所數、發電機箇數及其容量 (全國)

	原 動 力			電 壓			發 電 力			計
	水力	火力	風力	低 壓	高 壓	特 別 高 壓	五 十 キ ロ ワ ッ ト 以 上	五 十 キ ロ ワ ッ ト 以 上	五 十 キ ロ ワ ッ ト 以 上	
	永力	火力	風力	低 壓	高 壓	特 別 高 壓	五 十 キ ロ ワ ッ ト 以 上	五 十 キ ロ ワ ッ ト 以 上	五 十 キ ロ ワ ッ ト 以 上	
發 電 所 數 (自明治四十年至大正二年)										
明治四十年末	74	61	2	137	19	99	19	137	19	137
同 四十二年	94	74	2	170	16	128	26	170	20	201
同 四十四年	109	88	4	201	16	149	36	201	38	239
同 四十六年	131	107	17	255	23	189	51	255	50	305
同 四十八年	174	113	47	334	18	239	77	334	83	417
同 五十年	215	118	91	424	29	294	102	424	121	545
同 五十二年	247	130	117	494	24	365	105	494	171	665
發 電 機 箇 數 (大 正 二 年 末)										
電 氣 機 箇 數	216	83	100	408	13	315	60	408	165	573
電 氣 機 箇 數	1	7	1	9	4	48	25	77	6	113
電 氣 機 箇 數	30	40	7	77	4	48	25	77	6	113
計	247	130	117	494	24	365	105	494	171	665
發 電 機 容 量 (自明治四十年至大正二年)										
明治四十年末	—	—	—	—	63	247	15	325	56	80
同 四十二年	—	—	—	—	56	290	26	372	43	90
同 四十四年	—	—	—	—	55	325	36	416	44	95
同 四十六年	—	—	—	—	72	394	48	514	47	107
同 四十八年	—	—	—	—	73	499	53	625	48	128
同 五十年	—	—	—	—	79	578	61	718	57	156
同 五十二年	—	—	—	—	82	655	114	851	92	166
發 電 機 容 量 (自明治四十年至大正二年)										
明治四十年末	—	—	—	—	5,668	54,881	12,050	74,499	1,569	59,878
同 四十二年	—	—	—	—	7,183	68,270	29,020	104,473	1,137	93,311
同 四十四年	—	—	—	—	7,323	85,246	32,323	124,792	1,374	93,866
同 四十六年	—	—	—	—	10,359	117,139	48,673	176,071	1,451	107,074
同 四十八年	—	—	—	—	12,686	173,059	49,281	234,128	1,508	124,128
同 五十年	—	—	—	—	11,947	207,406	147,393	365,746	1,640	140,431
同 五十二年	—	—	—	—	13,174	230,826	243,768	487,768	2,631	170,811

本表ノ發電所數ノ使用認可モノノミヲ指シタル一發電所内ニ二種以上ノ電壓アルモノハ其高圧方ヲ計上ス

13. 電気事業及其事業 317

167. 電 氣 供 給 (地方別) 大正二年

地 方	電 氣 供 給				電 力				其他電力 供給	其他電力 消費
	供給戸數	電燈 電氣力	電燈 電氣力	電燈 電氣力	供給戸數	電燈 電氣力	電燈 電氣力	電燈 電氣力		
	人口百々付	面積一 方里々付	電燈 電氣力	電燈 電氣力	人口百々付	面積一 方里々付	電燈 電氣力	電燈 電氣力		
統計範圍	43,641	128,061	3,487	1,438,531	7.8	87.2	21.2	236.0	524	573
北海道	34,357	256,534	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511
東北道	715,021	2,251,597	62,079	19,374,221	22.7	195.8	1,083.4	4,824.3	7,828	8,881
関東道	115,263	385,434	7,400	2,617,560	6.5	64.2	102.4	1,601.8	1,576	1,275
中部道	104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,498.9	1,234	1,359
近畿道	241,469	543,375	12,719	4,823,336	10.7	94.8	308.4	3,270.1	3,480	3,637
中国道	638,222	1,510,962	42,331	14,111,125	22.4	208.4	1,026.7	5,548.4	6,457	6,903
四国	118,978	273,392	6,301	2,471,393	5.3	48.2	137.2	1,340.8	726	757
九州	75,960	179,064	4,921	1,695,053	5.5	52.1	151.7	1,435.7	588	623
計	191,525	483,199	10,108	4,909,445	6.3	64.9	180.9	1,871.8	1,074	1,188
計	6,670	11,000	244	82,558	2.1	15.3	70.7	526.1	2	15
總 計	2,844,735	6,127,430	160,855	56,726,076	11.6	107.2	218.0	2,247.0	25,201	27,376
北海道	43,641	128,061	3,487	1,438,531	7.8	87.2	21.2	236.0	524	573
東北道	8,027	28,710	260	289,362	3.2	38.7	39.1	476.7	140	170
関東道	6,308	21,307	344	293,587	2.6	29.9	32.9	347.3	49	51
中部道	9,433	24,582	574	248,490	2.7	27.0	32.6	329.6	60	69
近畿道	30,854	47,811	1,203	888,888	5.0	51.4	79.7	814.6	323	330
中国道	22,541	67,889	600	600,426	5.4	65.7	125.5	1,110.3	435	351
四国	26,409	71,049	2,088	670,834	6.6	53.2	84.0	792.9	506	540
九州	15,731	40,957	951	544,706	3.2	42.2	106.4	1,414.9	164	169
計	19,940	50,374	822	521,084	5.1	52.8	122.3	1,295.5	210	221
計	37,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0	303	319
計	15,474	37,991	373	373,219	23.6	134.1	1,405.6	135	172	172
計	24,508	65,244	1,177	608,807	4.8	41.4	270.2	1,800.5	207	209
計	503,227	1,683,555	47,376	10,077,656	59.9	498.6	1,022.0	5,512.7	5,339	6,175
計	108,481	303,888	9,360	6,637,395	26.5	220.3	1,052.1	5,943.2	1,365	1,616
計	39,754	101,454	2,708	1,144,249	5.4	60.9	134.0	1,387.7	502	629
計	29,878	58,997	1,580	509,860	7.5	63.0	221.5	1,914.2	301	306
計	24,272	61,432	2,015	575,007	7.9	73.8	226.9	1,214.3	396	428
計	21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1	587	612
計	42,699	129,236	3,304	1,471,821	8.4	101.6	143.2	1,733.9	434	522
計	36,345	71,381	1,706	626,267	6.7	58.8	108.3	1,934.7	552	567
計	25,508	51,246	625	464,781	7.6	69.2	109.5	1,708.4	248	250
計	14,931	36,319	961	348,412	6.2	59.7	125.2	1,205.5	147	156
計	89,331	186,100	4,146	1,787,704	12.7	122.2	369.4	3,548.3	1,157	1,207
計	96,809	229,550	5,558	1,996,245	12.2	101.7	765.9	6,382.3	1,858	1,938
計	10,288	81,897	2,060	690,035	8.3	64.0	220.9	1,872.3	327	339
計	78,386	210,965	5,949	2,886,096	17.3	234.4	711.5	5,664.8	992	1,040
計	142,629	314,304	9,469	2,908,497	15.8	137.1	582.4	5,506.3	1,199	1,244
計	368,319	877,496	26,339	7,388,038	40.3	329.6	1,582.9	8,511.9	3,512	3,882
計	23,022	56,178	874	492,785	8.6	84.9	249.1	1,446.6	275	275
計	26,062	57,490	630	584,680	7.6	77.2	185.1	1,892.3	409	419
計	10,919	24,403	557	237,329	5.4	52.1	108.9	1,058.7	98	107
計	19,455	46,407	800	311,587	5.5	41.5	80.6	715.0	89	71
計	18,765	47,809	760	450,984	3.9	36.5	113.8	1,071.3	35	42
計	53,258	116,361	3,056	1,066,173	7.3	65.0	227.1	2,028.1	488	523
計	16,881	49,238	862	415,308	4.1	39.1	119.9	1,004.9	36	42
計	11,556	27,064	546	265,976	3.7	36.7	89.8	980.4	101	100
計	39,392	51,594	1,648	549,860	6.9	60.6	454.5	3,283.1	129	142
計	27,364	69,048	1,112	516,897	6.3	47.2	202.4	1,513.6	204	216
計	13,118	31,302	690	313,020	4.5	45.4	69.0	688.4	154	165
計	90,318	50,457	985	567,093	5.6	63.0	135.3	1,408.1	136	139
計	83,825	198,895	4,714	4,030,932	11.6	112.8	625.8	5,413.7	434	458
計	13,973	44,156	700	450,321	8.5	65.3	273.8	2,912.5	116	117
計	35,366	86,159	1,312	845,498	8.0	73.1	366.4	3,597.5	39	39
計	8,001	29,738	585	376,157	23.3	23.4	63.9	808.1	89	89
計	77,363	207,736	5,555	2,735,848	35.5	37.5	41.6	434.3	84	70
計	21,548	53,066	1,170	500,336	3.9	36.6	88.1	830.7	228	245
計	5,670	11,000	204	82,552	2.1	15.5	70.7	526.1	2	15
總 計	2,844,735	6,127,430	160,855	56,726,076	11.6	107.2	218.0	2,247.0	25,201	27,376

本表ハ一事業ノシテ一以上ノ府縣ニ亙リ供給スルモノハ其設備ヲ區別シテ各府縣別ニ掲テ
電力ハ電氣事業業者ノ供給ノ分ヲ合サス

表体のアノテーションと切り出し（クロッピング）

107. 電氣供給（地方別）大正二年

[illegible]

注釈 青木表ハ一事業ニシテ二以上ノ府縣ニ亘リ供給スルモノハ其設備ヲ區別シテ各所屬府縣ニ揚テ電力ハ電氣事業者ニ供給ノ分ヲ含メス

地 方	電 燈						電 力						
	供給戶數	電燈箇數	換電氣力	算燭光數	人口百=付		面積一方里=付		供給戶數	電 動 機 裝置數	換 算 換馬力	其他電力裝置	
					電燈 箇數	燭光數	電燈 箇數	燭光數				裝置數	換 算 換電氣力

[illegible]

者本表ハ一事業ニシテ二以上ノ府縣ニ亘リ供給スルモノハ其設備ヲ區別シテ各所屬府縣ニ掲ク
電力ニハ電氣事業者ニ供給ノ分ヲ含マス

OCR处理(文字認識) | json → csv

Visual Studio Code interface showing the conversion of OCR results from JSON to CSV.

Left Panel (JSON): Displays the JSON structure for a text element. A red box highlights the bounding box coordinates and the text value.

```
15_p315_corrected_表体1.json > [ ] images > { } 0 > [ ] fields > { } 60
{
  "lineBreak": false
},
{
  "valueType": "ALL",
  "boundingPoly": {
    "vertices": [
      {
        "x": 444.0,
        "y": 235.0
      },
      {
        "x": 559.0,
        "y": 235.0
      },
      {
        "x": 559.0,
        "y": 277.0
      },
      {
        "x": 444.0,
        "y": 277.0
      }
    ]
  },
  "inferText": "5,723",
  "inferConfidence": 0.9998,
  "type": "NORMAL",
  "lineBreak": false
},
{
  "valueType": "ALL",
  "boundingPoly": {
    "vertices": [

```

Right Panel (CSV): Displays the corresponding CSV data. A red box highlights the row corresponding to the JSON element above.

text	confidence	bbox1.x	bbox1.y	bbox2.x	bbox2.y	bbox3.x	bbox3.y	bbox4.x	bbox4.y
6,743	1	1878	193	1993	193	1993	235	1878	235
8,881	1	1545	151	1655	147	1657	190	1546	194
9,324.3	1	1201	154	1348	154	1348	193	1201	193
5	0.98	2097	193	2131	193	2131	235	2097	235
12.7	1	823	1698	909	1698	909	1737	823	1737
122.3	1	940	1705	1047	1705	1047	1742	940	1742
147	1	1434	1653	1507	1653	1507	1695	1434	1695
156	1	1588	1656	1664	1656	1664	1695	1588	1695
382	1	1765	1658	1841	1658	1841	1698	1765	1698
509	1	1917	1656	1995	1656	1995	1698	1917	1698
534	1	2233	196	2309	196	2309	238	2233	238
1,234	1	1395	235	1507	235	1507	277	1395	277
1,339	1	1542	237	1658	233	1659	277	1544	280
1,436.9	1	1207	238	1348	238	1348	277	1207	277
104,552	1	42	232	198	232	198	274	42	274
12,057	1	1862	282	1987	282	1987	321	1862	321
2,701	1	1726	238	1839	238	1839	279	1726	279
244,913	1	225	235	379	235	379	274	225	274
3,279	1	2196	282	2309	282	2309	324	2196	324
3,637	1	1546	279	1656	279	1656	321	1546	321
42,231 14,141,12	0.95	426	324	765	324	765	363	426	363
5,723	1	444	235	559	235	559	277	444	277
7.7	1	844	238	906	238	906	277	844	277
10.7	1	823	282	906	282	906	319	823	319
16	1	2076	235	2128	235	2128	277	2076	277
80.5	1	958	238	1047	238	1047	277	958	277

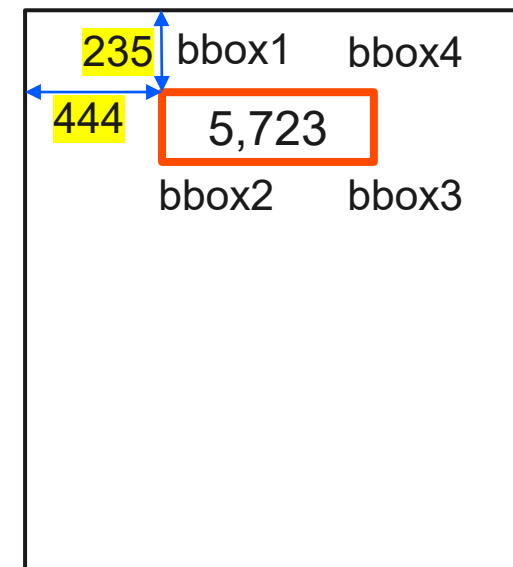
Ln 1635, Col 18 Spaces: 4 UTF-8 CRLF JSON

OCR处理

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275	1	1
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511	4,737	6,409	15	1,762
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3	7,828	8,881	21,543	28,521	151	5,556
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	1,876	1,975	5,069	6,743	5	534
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9	1,234	1,339	2,701	3,592	16	4,419
241,469	543,375	12,715	4,823,396	10.7	94.8	288.4	3,270.1	3,489	3,637	9,067	12,057	24	3,279
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,346.4	6,457	6,860	21,916	29,894	11	380
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8	725	787	2,006	3,937	166	163
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062	6	29
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703	7	4,802
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20	—	—
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925

3,467	1,438,531	7.8	87.2
6,708	2,520,591	4.5	44.9
62,078	19,374,221	22.7	195.8
7,409	2,617,500	6.5	64.2
5,723	2,562,869	7.7	80.5
12,715	4,823,396	10.7	94.8
42,231	14,141,125	22.4	208.4
6,201	2,471,393	5.3	48.2
4,021	1,695,053	5.5	52.1
10,108	4,999,445	6.3	64.9
204	82,552	2.1	15.5
160,865	56,726,676	11.6	107.2

Text: 5,723
 Confidence: 1
 bbox1.x: 444
 bbox1.y: 235
 bbox2.x: 559
 bbox2.y: 235
 bbox3.x: 559
 bbox3.y: 277
 bbox4.x: 444
 bbox4.y: 277



表レイアウト解析(≒自動で罫線を引くこと)

43,641	128,961	「キロワット」	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554		6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3	
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9	
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1	
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4	
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8	
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8	
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7	
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5	
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6	
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6	
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3	
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9	
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2	
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5	
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0	
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0	
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5	
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7	
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2	
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7	
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2	
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3	
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1	
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9	
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7	
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4	

43,641	128,961	「キロワット」	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554		6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3	
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9	
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1	
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4	
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8	
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8	
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7	
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5	
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6	
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6	
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3	
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9	
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2	
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5	
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0	
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0	
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5	
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7	
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2	
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7	
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2	
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3	
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1	
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9	
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7	
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4	

表レイアウト解析1 | OCR処理をする

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4

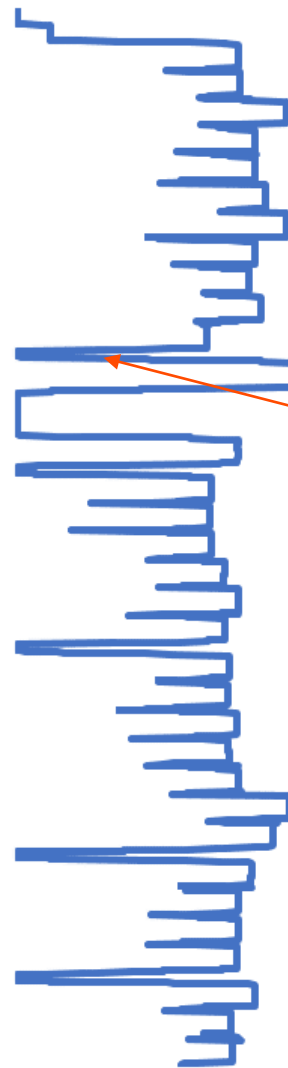
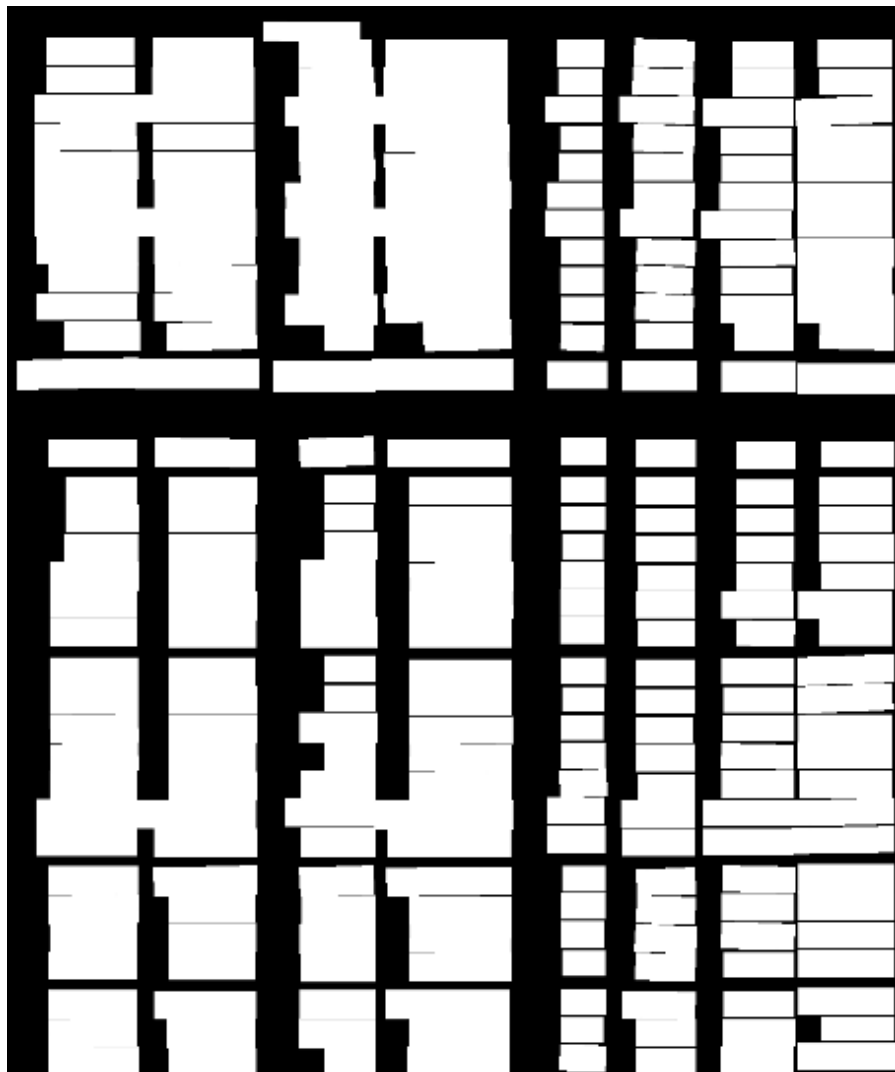
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4

表レイアウト解析2|モノクロ画像を作成

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	13,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4



表レイアウト解析3 | 水平・垂直投影プロフィール

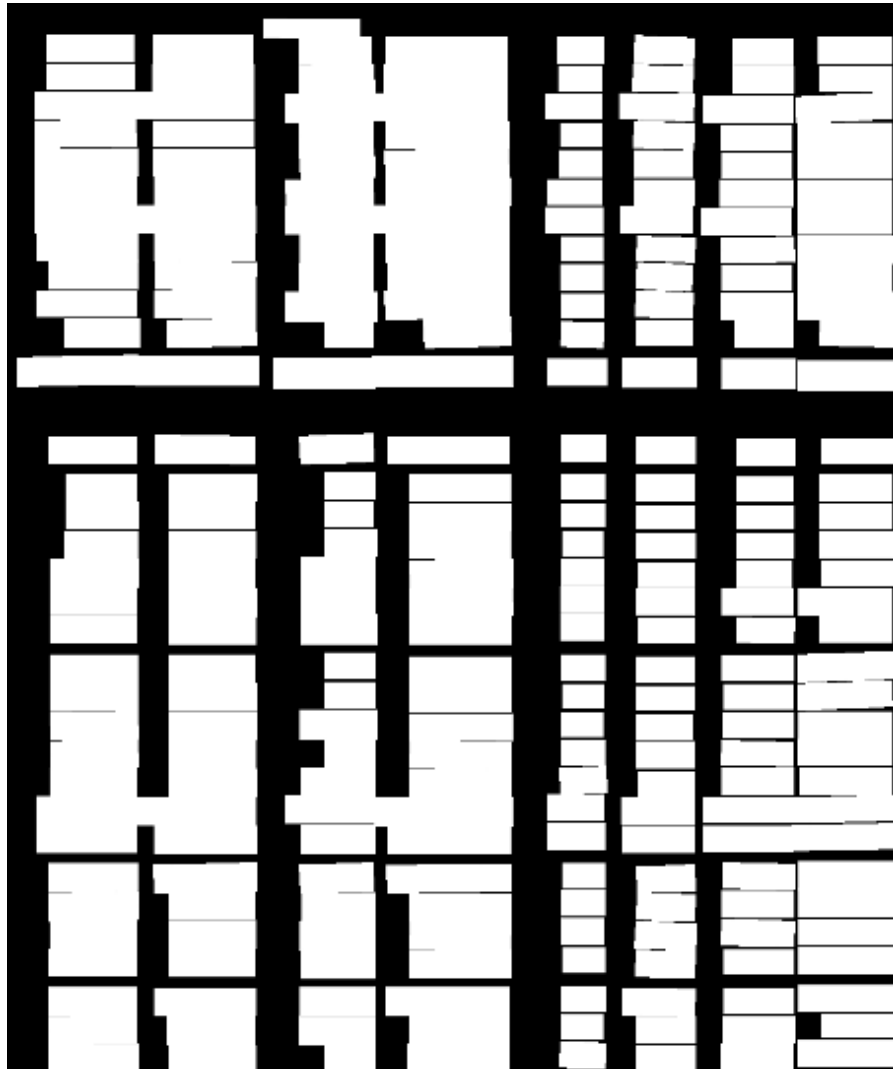


水平投影プロフィール
(projection profile)

行方向に、Y座標ごとに、
白いピクセル数を数えて、
ヒストグラムにする

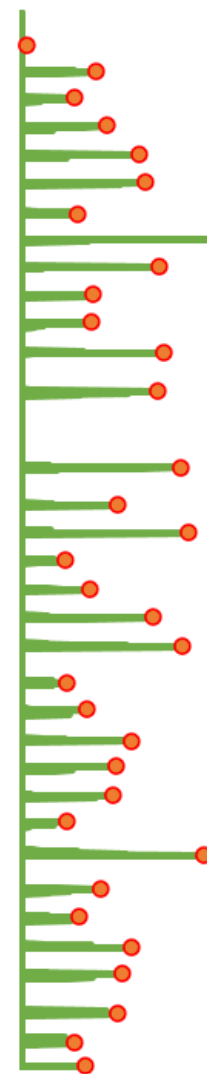
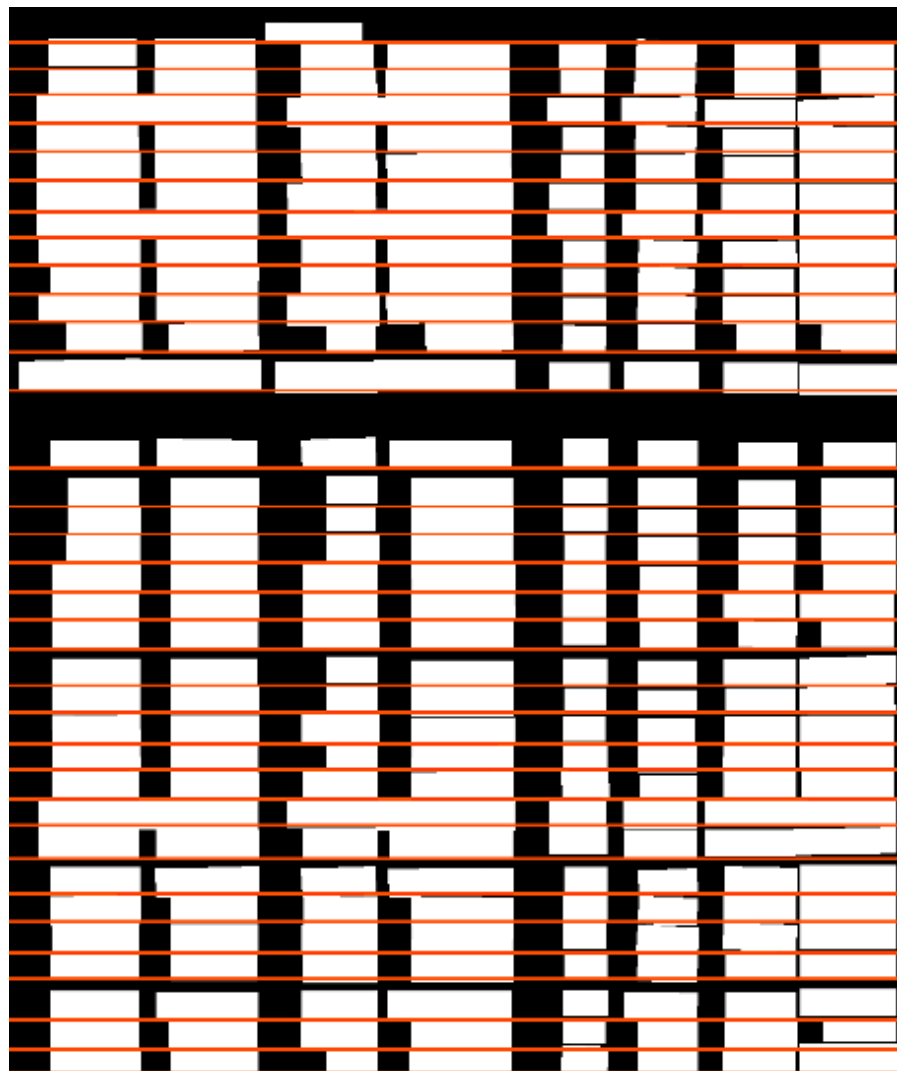
改行で急減

表レイアウト解析4|ヒストグラムの差分をとる



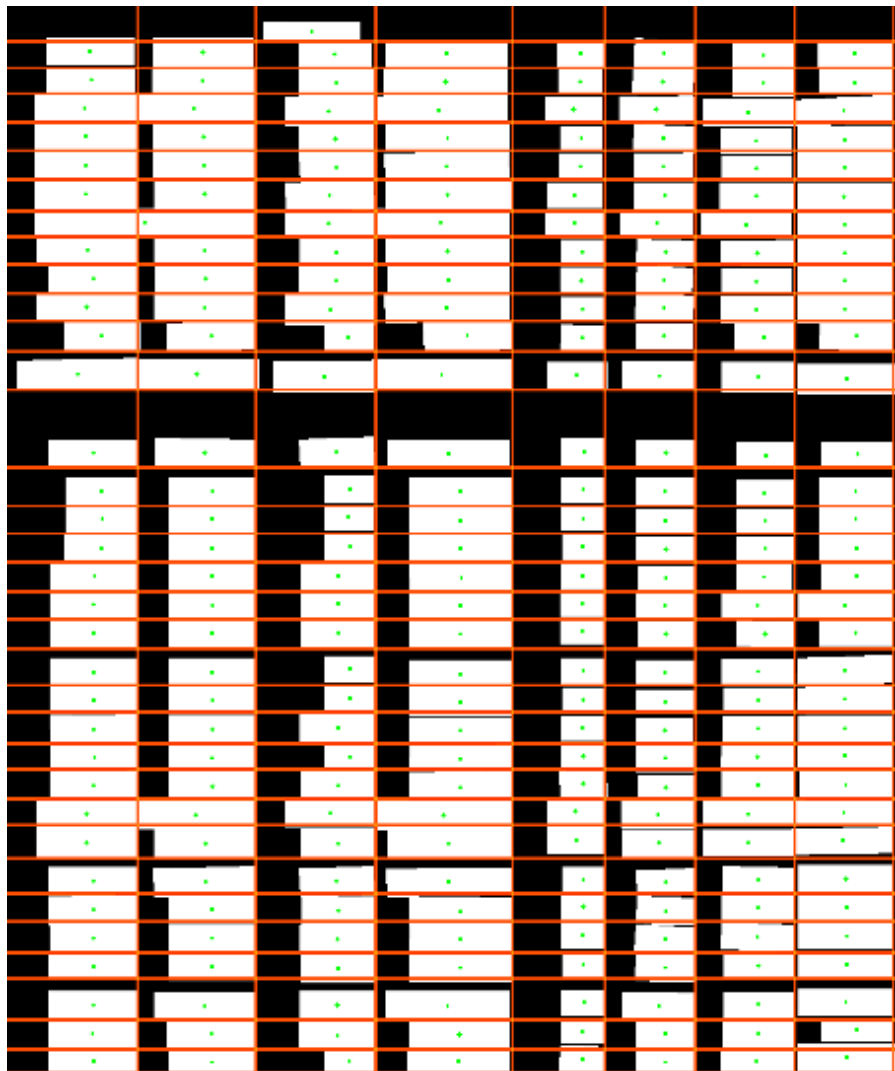
改行位置＝
減少幅が大きい
ところ

表レイアウト解析5 | ピーク検出・改行位置の特定



↑ ピーク間の幅：
Bboxの高さ以上

表レイアウト解析6 | 文字列の配置



各boundingboxの
重心が含まれるセルに配置

統計表 (Excel形式) と検証用画像の出力

キロワット						
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9
5,670	11,099	204	82,552	2.1	15.5	70.7
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
8,027	23,716	600	289,362	3.2	38.7	39.1
6,993	21,507	344	222,590	2.6	26.9	23.9
9,433	24,582	574	248,489	2.7	27.0	32.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7
22,541	67,889	1,899	600,428	7.4	65.7	125.5
26,509	71,049	2,088	670,834	5.6	53.2	84.0
15,731	40,987	951	544,706	3.2	42.2	106.4
19,946	50,374	822	521,084	5.1	52.3	122.3
27,653	69,412	1,242	683,354	7.1	69.4	170.4
15,474	37,991	645	373,219	2.9	28.6	143.1
24,509	65,290	1,177	606,807	4.8	44.4	200.2
503,227	1,683,655	47,976	14,007,656	59.9	498.6	1,322.2
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0
29,878	58,997	1,586	509,960	7.5	65.0	221.5
24,272	61,422	2,015	575,097	7.9	73.8	226.9
21,346	43,561	1,103	388,194	6.8	60.9	159.9

	A	B	C	D	E	F	G
1			「キロワット」				
2	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
3	94,357	256,554	6,708	2,520,591	4.5	44.9	60.2
4	715,021	2,251,597		6,207,819,374,221	22.7	195.8	1,083.4
5	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4
6	104,552	244,913	5,723	2,562,869	7.7	80.5	137.3
7	241,469	543,375	12,715	4,823,396	10.7	94.8	368.4
8		6,382,821,510,962		42,231 14,141,125	22.4	208.4	1,026.7
9	118,978	273,262	6,201	2,471,393	5.3	48.2	137.2
10	75,990	179,064	4,021	1,695,053	5.5	52.1	151.7
11	191,525	483,199	10,108	4,999,445	6.3	64.9	180.9
12	5,670	11,099	204	82,552	2.1	15.5	70.7
13	2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	
14	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
15	8,027	23,716	600	289,362	3.2	38.7	39.1
16	6,993	21,507	344	222,590	2.6	26.9	23.9
17	9,433	24,582	574	248,489	2.7	27.0	32.6
18	20,854	47,811	1,203	488,888	5.0	51.4	79.7
19	22,541	67,889	1,899	600,428	7.4	65.7	125.5
20	26,509	71,049	2,088	670,834	5.6	53.2	84.0
21	15,731	40,987	951	544,706	3.2	42.2	106.4
22	19,946	50,374	822	521,084	5.1	52.3	122.3
23	27,653	69,412	1,242	683,354	7.1	69.4	170.4
24	15,474	37,991	645	373,219	2.9	28.6	143.1
25	24,509	65,290	1,177	606,807	4.8	44.4	200.2
26	503,227	1,683,655	47,976	14,007,656	59.9	498.6	
27	108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1
28	39,754	101,454	2,705	1,144,249	5.4	60.9	123.0
29	29,878	58,997	1,586	509,960	7.5	65.0	221.5
30	24,272	61,422	2,015	575,097	7.9	73.8	226.9
31	21,346	43,561	1,103	388,194	6.8	60.9	159.9

校正

地 方	電 燈					
	供給戸数	電燈箇数	換 算 電 氣 力	燭 光 数	人口百ニ付	
					電燈 箇数	燭光数
統計區畫			「キロワット」			
北海道	43,641	128,961	3,467	1,438,531	7.8	87.2
東北區	94,357	256,554	6,708	2,520,591	4.5	44.9
關東區	715,021	2,251,597	62,078	19,374,221	22.7	195.8
北陸區	115,250	265,434	7,409	2,617,500	6.5	64.2
東山區	104,552	244,913	5,723	2,562,869	7.7	80.5
東海區	241,469	543,375	12,715	4,823,396	10.7	94.8
近畿區	638,282	1,510,962	42,231	14,141,125	22.4	208.4
中國區	118,978	273,262	6,201	2,471,393	5.3	48.2
四國區	75,990	179,064	4,021	1,695,053	5.5	52.1
九州區	191,525	483,199	10,108	4,999,445	6.3	64.9
沖繩縣	5,670	11,099	204	82,552	2.1	15.5
總 計	2,344,735	6,127,420	160,865	56,726,676	11.6	107.2
道府縣						
北海道	43,641	128,961	3,467	1,438,531	7.8	87.2
東北區						
青森縣	8,027	23,716	600	289,362	3.2	38.7
岩手縣	6,993	21,507	344	222,590	2.6	26.9
秋田縣	9,433	24,582	574	248,489	2.7	27.0
山形縣	20,854	47,811	1,203	488,888	5.0	51.4
宮城縣	22,541	67,889	1,899	600,428	7.4	65.7
福島縣	26,509	71,049	2,088	670,834	5.6	53.2
關東區						
茨城縣	15,731	40,987	951	544,706	3.2	42.2
栃木縣	19,946	50,374	822	521,084	5.1	52.3
群馬縣	27,653	69,412	1,242	683,354	7.1	69.4
埼玉縣	15,474	37,991	645	373,219	2.9	28.6
千葉縣	24,509	65,290	1,177	606,807	4.8	44.4
東京都	503,227	1,683,655	47,976	14,007,656	59.9	498.6

	A	B	C	D	E	F	G	H
1	統計區畫	府県	電灯	電灯	電灯	電灯	電灯	電灯
2			供給戸数	電灯箇数	換算電氣力	燭光数	人口百ニ付	人口百ニ付
3					キロワット		電灯箇数	燭光数
4								
5	北海道		43,641	128,961	3,467	1,438,531	7.8	87.2
6	東北區		94,357	256,554	6,708	2,520,591	4.5	44.9
7	關東區		715,021	2,251,597	62,078	19,374,221	22.7	195.8
8	北陸區		115,250	265,434	7,409	2,617,500	6.5	64.2
9	東山区		104,552	244,913	5,723	2,562,869	7.7	80.5
10	東海区		241,469	543,375	12,715	4,823,396	10.7	94.8
11	近畿區		638,282	1,510,962	42,231	14,141,125	22.4	208.4
12	中国區		118,978	273,262	6,201	2,471,393	5.3	48.2
13	四國區		75,990	179,064	4,021	1,695,053	5.5	52.1
14	九州區		191,525	483,199	10,108	4,999,445	6.3	64.9
15	沖繩區		5,670	11,099	204	82,552	2.1	15.5
16	總計		2,344,735	6,127,420	160,865	56,726,676	11.6	107.2
17								
18	北海道區	北海道	43,641	128,961	3,467	1,438,531	7.8	87.2
19	東北區	青森	8,027	23,716	600	289,362	3.2	38.7
20	東北區	岩手	6,993	21,507	344	222,590	2.6	26.9
21	東北區	秋田	9,433	24,582	574	248,489	2.7	27.0
22	東北區	山形	20,854	47,811	1,203	488,888	5.0	51.4
23	東北區	宮城	22,541	67,889	1,899	600,428	7.4	65.7
24	東北區	福島	26,509	71,049	2,088	670,834	5.6	53.2
25								
26	關東區	茨城	15,731	40,987	951	544,706	3.2	42.2
27	關東區	栃木	19,946	50,374	822	521,084	5.1	52.3
28	關東區	群馬	27,653	69,412	1,242	683,354	7.1	69.4
29	關東區	埼玉	15,474	37,991	645	373,219	2.9	28.6
30	關東區	千葉	24,509	65,290	1,177	606,807	4.8	44.4
31	關東區	東京	503,227	1,683,655	47,976	14,007,656	59.9	498.6

- はじめに
- OCRの活用と課題
- ocrstatsの概要
- **元祖ネ申Excelとの邂逅**
- 今後の課題と展望

主なエラーのパターン

—	2	2
1	4	5
1	5	6
2	11	13
4	13	17
3	9	12
3	6	9
—	7	7

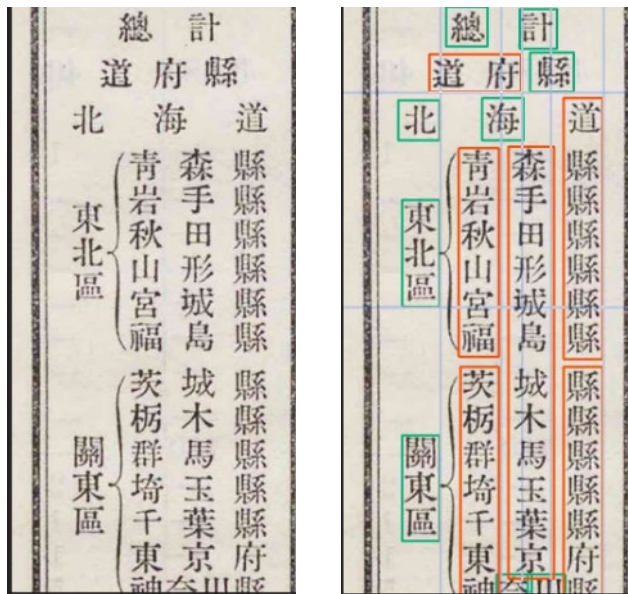
1	292	154,784	73,645	228,429
—	2,497	110,427	39,081	149,508
—	14	4,617	1,661	6,278
—	353	15,261	3,372	18,633
5	769	683,390	205,023	888,413
—	5,354	204,180	67,620	271,800
244	21,474	14,584,125	7,038,490	21,622,615
1	118,042	3,559,084	1,208,908	4,767,992
245	139,516	18,143,209	8,247,398	26,390,607
239	127,418	17,203,790	8,207,970	25,411,760
287	117,174	19,009,784	7,777,504	26,787,288
212	100,450	14,105,585	7,223,092	21,328,677
174	100,918	12,334,778	6,399,725	18,734,503
147	95,227	10,616,784	8,252,228	18,869,012

22	—
26	3
4	—
16	—
20	—
14	—
12	—
28	—
19	—
10	—
23	—
23	—
16	2
10	—
243	5

文字列範囲の認識不良

認識漏れ

表側は「縦読み」してしまう



そのままだと「縦読み」してしまう

	A	B	C	D	E
1		統計	區	畫	
2	北東關北東東近中四九沖	海北東陸山海畿國國州繩總道府	計		道區づ區區區區區區區縣縣
3	北東北區	青岩秋山宮鯛	海		道・縣縣縣縣縣縣
4	關東區	茨栃群埼千東神	奈	森手田形城島城木馬玉葉京川	縣縣縣縣縣府縣
5	北陸區	新富石w			縣縣縣縣縣縣縣
6	東山區	長岐滋			
7	東海區	山部愛三		潟山川井 野阜賀 梨岡知重都庫阪民	縣縣縣縣
8	近畿區	京兵大奈和	歌	山	
9	中國區	鳥島岡廣山		取根山島口	府市三ス縣縣オウツウツカウツー 縣縣
10	四國區	德香愛高			縣m間縣
11	九州區	大福佐長熊宮鹿		島川媛知分岡賀崎本崎	縣縣入縣縣縣縣
12	沖		繩	兒島	縣

表頭の入れ子構造は難しい

地 方	電 燈								電 力					
	供給戸数	電燈筒数	換 電 氣 力	燭 光 數	人口百ニ付		面積一方里ニ付		供給 戸数	電 動 機			其他電力装置	
					電燈 筒数	燭光數	電燈 筒数	燭光數		装置數	換 電 氣 力	換 馬 力	装置數	換 算 電 氣 力

地 方	電 燈								電 力					
	供給戸数	電燈筒数	換 電 氣 力	燭 光 數	人口百ニ付		面積一方里ニ付		供給 戸数	電 動 機			其他電力装置	
					電燈 筒数	燭光數	電燈 筒数	燭光數		装置數	換 電 氣 力	換 馬 力	装置數	換 算 電 氣 力

	A	B	C	D	E	F	G	H	I	J	K
1	電	燈					電		力		
2	地 方換		入口百ニ付		面積一方里ニ付	供給	電動	機			其他電力装置
3	供給戸数電燈筒数電気 取引	燭光數電燈筒数	燭光數	電燈筒数	燭光數	戸數	装置数換電氣力	算	か 芳	装置数	換算電氣力

ネ申Excelとは

- 「紙へ印刷することを前提に、セルの結合や罫線機能などをフルに使い、見栄えを優先して作ったExcelファイル」¹⁾
- 機械処理が難しく、データとしての再利用性が低い

「ネ申 Excel」問題

奥村晴彦@三重大学



<http://www.louvre.fr/en/oeuvre-notices/archaic-writing-tablets>

2)

		E	F	G	H	I	J
4	区 分	給 与 所 得 者 数					給 与
5		3月末	6月末	9月末	12月末	年間月平均	総 額
6		千人	千人	千人	千人	千人	百万円
7	平成18年度	54,752	55,013	54,065	53,396	54,506	200,034,648
8	19	54,957	55,311	54,417	53,768	54,817	201,272,200
9	20	54,672	55,269	55,000	54,739	55,124	201,317,695
10	21	56,232	54,909	54,035	53,884	54,967	192,474,249
11	22	55,817	54,887	54,367	54,153	54,792	194,372,236
12	23	55,169	54,688	54,459	54,273	54,647	195,799,665
13		人	人	人	人	人	百万円
14	10人未満	10,010,414	10,022,094	9,900,419	9,627,198	9,890,052	30,235,033
15	10人以上	8,333,724	7,886,247	7,863,073	7,735,850	7,954,723	25,820,091
16	30人 "	8,725,936	8,668,139	8,630,664	8,653,029	8,669,395	28,749,537
17	100人 "	11,820,473	11,718,332	11,658,236	11,727,615	11,731,189	41,499,618
18	500人 "	4,039,942	4,061,021	4,072,913	4,081,849	4,063,946	15,959,270
19	1,000人 "	7,178,060	7,244,851	7,248,209	7,290,875	7,240,466	30,937,597
20	5,000人 "	5,060,628	5,086,941	5,085,593	5,156,124	5,097,320	22,598,520
21	計	36,825,039	36,779,284	36,695,615	36,909,492	36,802,316	139,744,542
22	合 計	55,169,177	54,687,825	54,459,107	54,272,540	54,647,091	195,799,665

- 1) <https://atmarkit.itmedia.co.jp/ait/articles/1612/26/news032.html>
- 2) <https://oku.edu.mie-u.ac.jp/~okumura/SSS2013slide.pdf>

総務省のガイドライン

- 総務省（2020）「統計表における機械判読可能なデータ作成に関する表記方法について」*

データ・項目の取扱い

- 1セル1データとなっているか（1-2）
- 数値データは数値属性とし、文字列を含まないこと（1-3）
- セルの結合をしていないか（1-4）
- スペースや改行等で体裁を整えていないか（1-5）
- 項目名等を省略していないか（1-6）
- オブジェクトを使用していないか（1-8）
- データの単位を記載しているか（1-9）

表の構成の取扱い

- データが分断されていないか（2-1）
- 1シートに複数の表が掲載されていないか（2-2）

「統計表における機械判読可能なデータ作成に関する表記方法」令和2年12月18日統計企画会議申合せ

ガイドラインのNG項目が、
約100年前の政府統計書で
ことごとく確認できる



「ネ申Excel」のルーツは
100年遡れる！

* https://www.soumu.go.jp/menu_news/s-news/01toukatsu01_02000186.html

数値・文字列・単位の混在, オブジェクトの利用

1セルに複数のデータが入力されている

	企業等数	売上金額 (費用総額)	
総計	1188389	391445	(355943)
A	50384	69565	(63883)
B	154138	50468	(45332)
C	86522	44098	(33210)
D	85983	22258	(20237)
E	580003	37652	(35016)
F	27456	15288	(14633)
G	119085	115499	(110735)
H	84818	36617	(32897)

気					温	
平 均	累年平均 (観測年数)	最 高 平 均	最 低 平 均	最高最低 ノ 差	最 高 極 (年 月 日)	最 低 極 (年 月 日)
度	度	度	度	度	度 年 月 日	度 年 月 日
24.4	24.3(10)	28.1	27.7	6.4	34.4(39.V.17)	10.1(34.II.12)
23.6	23.4(10)	28.2	20.1	8.1	37.9(36.VI.16)	8.2(34.II.13)
23.4	23.0(10)	28.6	19.5	9.1	36.2(34.VII.22)	3.0(34.II.13)
23.1	22.5(10)	25.8	21.3	4.5	33.1(41.X.6)	7.3(34.II.12)
22.6	22.0(10)	28.5	18.4	10.1	37.2(35.VIII.1)	-1.0(34.II.13)
22.1	21.6(10)	26.7	18.5	8.1	37.0(42.VIII.16)	-0.2(34.II.13)
21.9	21.5(10)	25.0	19.3	5.7	35.9(42.VI.9)	3.0(34.II.13)
23.4	23.2(10)	26.5	20.0	6.5	34.3(43.VII.12)	7.1(34.II.13)
22.4	22.0(20)	26.0	19.1	6.9	35.1(43.VII.21)	5.2(34.II.12)
21.2	20.8(10)	24.8	18.0	6.7	34.6(39.VIII.7)	3.1(34.II.12)
17.4	16.8(25)	21.4	14.0	7.4	36.2(26.VII.26)	-6.1(19.II.2)

「円」、「▲ (マイナス表記)」、「, (カンマ)」が
文字列として入力されている

	単価	前回差分	生産台数
サンプル1	10,000円	130	12,000
サンプル2	9,100円	▲200	29,000
サンプル3	8,020円	▲350	37,000
サンプル4	7,500円	500	43,000
SUM関数	0	630	0
+(加算演算)	#VALUE!	#VALUE!	121000

↑黄色の行は関数で合計を表示した例

年 次	数 量	價 額	数 量	價 額	数 量
魚 鱈 春 鰯 鰯					
明治三十二年		48,538,428円		6,532,917円	9,688,513
同 三十三年		48,672,087円		7,284	
同 三十四年	石	54,272,807円		7,019	
同 三十五年	石	62,313,363円		7,206	
同 三十六年		33,560,367円		5,715	
同 三十七年		39,748,566円		6,842	
同 三十八年		31,296,516円		5,692,690円	8,643,197円
同 三十九年		28,288,888円		4,881,811円	8,271,552円
同 四十年		28,288,888円		4,881,811円	8,271,552円

オブジェクトを使用している

名称	内容	数量
商品 A	サンプル 1	100
	サンプル 2	200
	サンプル 3	300

セル内改行

改行で体裁を整えている

薬剤名	出荷 本数	単価
鎮静剤 A-1	429	756
鎮静剤 A-2	321	648
鎮静剤 A-3	384	438
鎮静剤 A-4	408	775

横書き

縦書き

結 果							人口萬ニ付 新受件數	判所番號 年次及裁
國席判決	拋棄認諾ニ 基ク判決	其他ノ終 局判決	取 下	和 解	訴狀差戻	其 他 結 果	合 計	

入れ子構造

セルが結合されている

都道府県	エリア	市区町村	導入台数	増減数
	特別区部	千代田区	58406	11291
	特別区部	中央区	141183	18421
	特別区部	港区	243283	38152
S				
東京都	特別区部	葛飾区	442913	327
	特別区部	江戸川区	681298	2331
	市町村	八王子市	577513	-2540
	市町村	立川市	176295	-3373
	市町村	武蔵野市	144730	5996
	市町村	三鷹市	186936	853
	市町村	青梅市	137381	-1958
	市町村	府中市	260274	4768
S				
	市町村	青ヶ島村	-	
	市町村	小笠原村	-	

48. 海外在留本邦人 (内地人)

關東州及南滿洲鐵道附屬地						支那					
總數		本業者		本業外労働者及家事従事人		總數		本業者		本業外労働者及家事従事人	
男	女	男	女	男	女	男	女	男	女	男	女
1,027	821	665	81	362	740	443	307	344	106	99	201

紡織工業

製絲業
絹絲紡績業
絹絲紡績業
麻絲紡績業
毛絲紡績業
製綿業
絹織物業
絹織物業及絹織交織物業
麻織物業及其交織物業
毛織物業及其交織物業
其他ノ織物業
染色及精練整理業
莫大小業
其他ノ紡織工業

計

機械及器具工業

各種五軸及各種機械製造業
電氣機械器具製造業
農業土木採掘機械運搬用機器製造業
製造加工用機械器具製造業
度量衡計測計測用機器製造業
車輛製造業
造船業
其他ノ機械器具製造業

計

金屬工業

金屬精鍊業
金屬材料品製造業
鑄物製造業
鑄物以外ノ金屬製品製造業
鍍金業
其他ノ金屬工業

計

工

陶磁器製造業
硝子製造業
セメント及石灰製造業
煉瓦製造業
其他ノ窯業

化學工業

製紙業
人造絹絲製造業
燐ノ製造業
護謨製造業
油脂類製造業
セルロイド製造業
醫藥及賣藥製造業
工業藥製造業
染料製造業
塗料及顔料製造業

石鹼化材料品製造業
人造肥料製造業
其他ノ化學工業

計

均等割り付け

スペースで体裁を整えている

地域名	出荷本数	在庫本数
津□□市	429	756
四日市市	321	648
伊勢市	384	438
松阪市	408	775

道府県		
北海道	青森県	岩手県
東北區	秋田県	山形県
	宮城県	福島県
関東區	茨城県	栃木県
	群馬県	埼玉県
	千葉県	東京都
北陸區	新潟県	富山県
	石川県	福井県

地方	
本州	東京府
	神奈川府
	埼玉縣
	千葉縣
	茨城縣
	栃木縣
	群馬縣
	山梨縣
	長野縣
	岐阜縣
	愛知縣
	三重縣
	滋賀縣
	福井縣
	石川縣
	富山縣
	新潟縣
	山形縣
	秋田縣
	岩手縣
	青森縣
	北海道

職業	業
總數	
1	農耕ニ従事スル者
2	畜産ニ従事スル者
3	蠶業ニ従事スル者
4	林業ニ従事スル者
5	漁業ニ従事スル者
6	採炭ニ従事スル者
7	採鑛ニ従事スル者
8	石油鑛業ニ従事スル者
9	土石採取ニ従事スル者
10	窯業、土石加工ニ従事スル者
11	金屬工業、機械器具製造、造船、運搬用具製造ニ従事スル者
12	精巧工業ニ従事スル者
13	化學製品ノ製造ニ従事スル者
14	紡織工業ニ従事スル者
15	被服、身製品製造ニ従事スル者

表側の省略

同じ名称を空白で省略している

薬剤名	出荷本数	在庫本数
鎮静剤 A-1	429	756
2	321	648
3	384	438
4	408	775

		發	送
年度			
大正	13	19,603,080	
	14	20,954,554	
昭和	1	22,584,405	
	2	22,236,780	
	3	24,564,036	
	4	25,806,686	
	5	26,380,272	
	6	24,608,728	
	7	24,703,934	
	8	30,147,919	

		普通	特殊	無料	合
		總 數			
明治四十五年度大正元年度		15,607	7,564	1,222	2
大正	二 年度	16,613	7,553	1,204	2
	三 年度	16,638	8,292	1,273	2
	四 年度	17,107	7,645	1,377	2
	五 年度	19,081	10,060	1,438	3
	六 年度	21,637	10,091	1,516	3
同	七 年度	25,082	13,313	1,851	4
同	八 年度	24,399	14,751	2,004	4
同	九 年度	24,043	17,343	1,937	4
同	十 年度	26,860	20,013	2,065	4
同	十一年度	27,878	21,357	2,121	5
同	十二年度	26,324	20,154	1,980	4

地 方		
本 地	東 京	明治三十七年度
		同 三十六年度
	神 奈 川	〃
		〃
	埼 玉	〃
		〃
	千 葉	〃
		〃
	茨 城	〃
		〃

長 崎	長 崎	本部及支部 管内區裁判所
	佐 賀	本 部 管内區裁判所
	福 岡	本部及支部 管内區裁判所
	大 分	〃
		〃
	熊 本	〃
		〃
	鹿 兒 島	〃
		〃
	宮 崎	〃
崎	那 覇	本 部 管内區裁判所
	計	本部及支部 管内區裁判所
		本部及支部

レイアウト解析の難しさ

第一二表 民有有 製造反別加償 昭和十三年一月一日

品名	単位	数量	金額	備考
...

見開きで
ページを
またぐ

1ページに複数の表

107. 肥料營養者及販賣價額 (全國)

年次	年 末 營 業 者				製 造 販 賣 價 額 (円)					
	製造者	輸入者	移入者	賣買者	動物糞	植物糞	鐵物糞	調 合	其 他	合 計
明治四十二年	18,516	448	414	35,018	6,026,327	6,494,810	6,364,274	6,618,632	13,972	29,617,111
同 四十三年	20,256	543	403	37,600	7,545,271	7,516,779	7,322,940	12,194,174	27,570	34,676,734
同 四十四年	21,794	696	409	39,719	7,530,454	9,797,916	8,967,203	14,308,988	22,297	40,677,353
同 四十五年	22,857	632	387	41,946	9,733,419	10,655,313	14,108,122	10,862,529	27,196	51,444,536
大 正 二 年	23,633	692	379	44,618	11,011,340	11,767,115	15,803,202	19,125,912	51,412	57,759,981
同 三 年	23,700	678	381	45,681	10,562,157	11,977,941	16,314,035	17,481,472	54,169	56,389,771
同 四 年	24,205	697	398	45,868	8,816,941	13,656,778	14,671,265	19,330,102	56,632	49,481,658
同 五 年	24,353	687	413	45,470	4,483,988	18,565,280	19,841,508	12,565,735	85,703	61,111,974
同 六 年	25,007	745	413	46,355	13,230,814	37,581,737	27,341,229	17,961,069	156,921	98,761,770
同 七 年	25,775	834	423	48,765	19,965,197	48,208,400	38,218,706	26,526,613	277,113	130,066,713
同 八 年	26,187	944	438	51,774	27,397,511	67,115,330	60,371,516	49,032,444	234,993	204,149,794
同 九 年	25,948	1,052	476	53,492	25,359,428	52,852,250	58,793,075	34,564,051	1,332,131	172,900,935
同 十 年	25,689	1,135	490	50,427	16,657,467	31,234,339	38,021,853	22,420,739	99,979	107,524,877
同 十 一年	25,338	?	?	?	16,720,342	35,830,758	47,570,297	24,837,437	62,939	125,071,673
同 十 二 年	25,347	?	?	?	20,510,641	37,927,032	56,735,398	29,760,459	136,565	138,569,003

大正元年以後ニ肥料取締法施行規則ノ改正ニ依リ製造價額ヲ改メ

108. 内國製造煙草生産高 (種類別)

年次	口付紙巻煙草		兩切紙巻煙草		葉巻煙草		割 煙 草		価額合計
	生産高	價 額	生産高	價 額	生産高	價 額	生産高	價 額	
大 正 二 年	5,378,893	21,381,763	2,017,162	7,134,162	825	38,223	6,556,577	14,335,631	83,039,200
同 三 年	5,942,536	22,224,260	2,745,088	7,144,457	850	43,229	6,907,207	14,601,323	83,412,685
同 四 年	5,963,098	22,745,204	1,847,396	6,530,193	897	38,261	7,047,792	16,066,175	80,398,513
同 五 年	6,332,909	29,132,700	2,242,425	8,791,294	1,180	51,540	6,425,456	14,736,799	84,712,933
同 六 年	8,305,313	34,648,248	2,639,893	11,507,239	2,287	115,643	6,498,473	15,896,503	102,157,633
同 七 年	11,256,218	52,408,476	3,110,211	14,627,940	2,267	163,231	6,945,195	62,478,592	129,678,209
同 八 年	14,460,647	78,099,983	2,932,240	16,743,565	3,182	234,245	6,968,792	77,044,489	172,123,292
同 九 年	18,775,071	108,546,363	3,988,734	21,434,699	3,206	193,235	7,129,693	82,746,849	212,834,337
同 十 年	20,087,975	123,095,700	3,229,019	20,892,371	3,075	150,101	7,408,431	83,841,280	227,923,357
同 十 一 年	21,905,943	128,809,445	4,022,569	22,305,127	3,225	139,828	7,007,541	73,370,562	224,714,962
同 十 二 年	21,699,683	134,953,604	5,241,159	29,527,567	1,214	220,651	6,214,925	75,772,992	249,774,774

108. 内國製造煙草生産高 (種類別) 大正十二年度

口付紙巻煙草				兩切紙巻煙草				葉巻煙草				割 煙 草			
製品名	生産高	價 額		製品名	生産高	價 額		製品名	生産高	價 額		製品名	生産高	價 額	
	千本	円			千本	円			千本	円			千本	円	
御 料 煙 草	170	4,110	ナ イ ル マ	1,376	45,592	御 料 ハ バ ナ	4	一 水 府	2,356	129,804					
ス ー パ ー	28,840	381,303	ア ル マ ス	3,899	110,682	イ ム コ フ ア ス	2	2,000	薩 摩 刺	583	53,657				
ベ ル ト	131,166	1,481,392	セ ャ	2,056	47,493	ス ャ フ フ ェ ー	251	22,417	蘭 毒 草	14,199	627,189				
不 費	13,985,396	104,197,704	ス タ	126,601	1,173,287	レ ガ リ ャ	1	0,354	白 梅	227,738	5,854,072				
や ま ひ	77,421	589,638	エ ー シ ャ	580,863	4,938,699	ブ リ ン セ ャ	2	2,650	さ つ 子	1,314,164	23,452,475				
大 朝 日	5,268,593	29,550,708	ニ ャ ン ト	3,319,616	18,808,807	ニ ャ ン ト	38	3,220	あ や め	1,881,396	20,980,974				
カ メ ャ	177,382	683,947	ハ ー	543,555	2,347,271	ロ ン ド レ ス	506	15,667	な ぎ	1,834,065	17,087,103				
カ ャ	45,881	193,891	は ま れ	610,860	1,771,891	コ ン ト	506	55,340	な ぎ	1,834,065	17,087,103				
か を り		347	南 切 版		396	セ ニ ャ	510	49,803	富 貴 煙	13,209	76,506				
			シ ャ	2,900	65,362	シ ャ									
合 計	21,699,683	134,953,604	合 計	5,241,159	29,527,567	合 計	1,214	220,651	合 計	6,214,925	75,772,992				

本表ニ於テ價額トハ實地代金額ナリ

表側だけ
左ページ

粗密の妙

				梅	桃	梨	柿	
				(石)			生 柿	干 柿
				全				國
大	正	二	年	471,725	10,796,235	21,441,768	47,040,623	4,472,500
同	三	年	436,009	10,463,266	20,847,354	38,911,224	3,707,462	
同	四	年	413,471	12,116,250	22,336,633	47,913,990	3,413,002	
同	五	年	341,744	12,964,142	24,635,620	40,039,291	2,215,552	
同	六	年	332,132	13,399,913	29,429,812	45,285,869	2,523,857	1
同	七	年	383,327	12,407,970	27,823,932	30,225,623	1,590,979	
同	八	年	411,927	12,809,160	27,726,480	43,617,962	2,251,680	
同	九	年	401,353	13,168,224	27,659,238	45,523,740	2,115,623	
同	十	年	405,823	11,065,483	30,128,501	49,615,279	2,857,540	
同	十一	年	383,794	11,321,854	30,230,244	47,609,824	2,865,061	1
同	十二	年	377,937	11,235,927	27,554,423	44,275,382	2,332,804	
				地 方 別				(大 正 十

89. 鑛產物價額											
地 方	金	銀	銅	鉛	錫	亞鉛	鐵	硫 化 鐵	格魯 鐵	諸 他 鐵	亞硫酸
1 北 海 道	915,060	205,989	452,410	—	—	—	2,097,237	—	12,324	40,058	—
東北區	2 3 青 島 縣 6,845 14,777 90,639 4 5 萊 陽 縣 10,300 6 7 招 遠 縣 218 20,738 76,720 8 9 龍 口 縣 412,843 533,178 8,450,832 27 10 11 煙 台 市 332 21,426 293,017 12 13 福 山 縣 4,902 6,034 36,932 14 15 牟 平 縣 239,291	—	—	—	—	—	5,302,643	17,432	—	4,347 306 690	—
關東區	8 茨 城 縣 2,131,135 842,947 6,009,178 9 栃 木 縣 178,321 711,435 9,415,391 10 群 馬 縣 11 千 葉 縣 12 東 京 縣 13 神 奈 川 縣	—	—	—	—	—	531,052	—	2,048 1,450 351	290,214	—
北陸區	15 新 潟 縣 611,315 218,439 14,705 16 富 山 縣 17 石 川 縣 1,446,557 18 福 井 縣	—	—	—	—	—	100,666	—	32,272 385	—	—
東山區	19 山 梨 縣 14,800 20 長 野 縣 21 群 馬 縣 46,054 260,903 6,983 574,566	—	—	—	—	—	131,981	—	17,649 13,517	31,490	—
東海區	22 靜 岡 縣 23 愛 知 縣 24 岡 崎 縣	—	—	—	—	—	42,730	—	2,420 126	—	—
近畿區	25 滋 賀 縣 26 京 都 府 27 和 歌 山 縣 28 大 阪 府 29 兵 庫 縣 30 神 戶 市	—	—	—	—	533,763	—	4,092	173 34,965	—	—
中國區	31 鳥 取 縣 32 島 根 縣 33 岡 原 縣 34 廣 島 縣 35 山口縣	—	—	—	—	—	71,263	—	181,500	—	—
四國區	36 德 島 縣 37 香 川 縣 38 愛 媛 縣 39 高 知 縣	—	—	—	—	—	1,203,893 2,596 697	122,851 7,564	—	—	—
九州區	40 福 岡 縣 41 北 九 州 市 42 長 崎 縣 43 大 分 縣 44 宮 崎 縣 45 鹿 兒 島 縣 46 那 霸 市	—	—	—	—	—	724,095	10,698	7,033	—	—
47 沖 繩 縣	—	—	—	—	—	—	—	—	—	—	—
總 計	10,209,068	4,862,401	43,628,979	691,713	558,637	5,052,910	7,851,105	2,991,829	153,437	157,780	1,596,265

あえて、ネ申Excelを擁護

- そこに美を見出す
 - 高度な活版印刷技術, ヒューマンな誤植, 罫線を多用した様式美
- データの利用可能性？
 - そんなことすんの, 一部の~~マニアック~~な学者とかデータ屋だけ
 - 大多数は, 集計された, 見やすい「表」がほしい
 - ネ申Excelは「見やすい表」として, 100年以上の伝統と合理性, 美を兼備

認識範囲不良は短冊切り出しで改善

3.2	42.2	106.4	1,414.2	3.2	42.2	106.4	1,414.2	3.2	42.2	106.4	1,414.2	3.2	42.2	106.4	1,414.2
5.1	52.3	122.3	1,265.5	5.1	52.3	122.3	1,265.5	5.1	52.3	122.3	1,265.5	5.1	52.3	122.3	1,265.5
7.1	69.4	170.4	1,678.0	7.1	69.4	170.4	1,678.0	7.1	69.4	170.4	1,678.0	7.1	69.4	170.4	1,678.0
2.9	28.6	143.1	1,406.0	2.9	28.6	143.1	1,406.0	2.9	28.6	143.1	1,406.0	2.9	28.6	143.1	1,406.0
4.8	44.4	200.2	1,860.5	4.8	44.4	200.2	1,860.5	4.8	44.4	200.2	1,860.5	4.8	44.4	200.2	1,860.5
59.9	498.6	13,322.2	110,872.7	59.9	498.6	13,322.2	110,872.7	59.9	498.6	13,322.2	110,872.7	59.9	498.6	13,322.2	110,872.7
26.5	230.3	1,952.1	16,942.2	26.5	230.3	1,952.1	16,942.2	26.5	230.3	1,952.1	16,942.2	26.5	230.3	1,952.1	16,942.2
5.4	60.9	123.0	1,387.7	5.4	60.9	123.0	1,387.7	5.4	60.9	123.0	1,387.7	5.4	60.9	123.0	1,387.7
7.5	65.0	221.5	1,914.2	7.5	65.0	221.5	1,914.2	7.5	65.0	221.5	1,914.2	7.5	65.0	221.5	1,914.2
7.9	73.8	226.9	2,124.3	7.9	73.8	226.9	2,124.3	7.9	73.8	226.9	2,124.3	7.9	73.8	226.9	2,124.3
6.8	60.9	159.9	1,425.1	6.8	60.9	159.9	1,425.1	6.8	60.9	159.9	1,425.1	6.8	60.9	159.9	1,425.1
8.4	101.6	143.2	1,723.9	8.4	101.6	143.2	1,723.9	8.4	101.6	143.2	1,723.9	8.4	101.6	143.2	1,723.9
6.7	58.8	106.3	932.7	6.7	58.8	106.3	932.7	6.7	58.8	106.3	932.7	6.7	58.8	106.3	932.7
7.6	69.2	198.5	1,798.4	7.6	69.2	198.5	1,798.4	7.6	69.2	198.5	1,798.4	7.6	69.2	198.5	1,798.4
6.2	59.7	125.3	1,205.5	6.2	59.7	125.3	1,205.5	6.2	59.7	125.3	1,205.5	6.2	59.7	125.3	1,205.5
12.7	122.3	369.4	3,548.3	12.7	122.3	369.4	3,548.3	12.7	122.3	369.4	3,548.3	12.7	122.3	369.4	3,548.3
12.2	101.7	765.9	6,382.3	12.2	101.7	765.9	6,382.3	12.2	101.7	765.9	6,382.3	12.2	101.7	765.9	6,382.3
8.3	64.0	220.9	1,872.3	8.3	64.0	220.9	1,872.3	8.3	64.0	220.9	1,872.3	8.3	64.0	220.9	1,872.3
17.3	234.4	711.5	9,664.8	17.3	234.4	711.5	9,664.8	17.3	234.4	711.5	9,664.8	17.3	234.4	711.5	9,664.8
15.8	137.1	582.6	4,956.3	15.8	137.1	582.6	4,956.3	15.8	137.1	582.6	4,956.3	15.8	137.1	582.6	4,956.3
40.3	339.6	7,582.9	63,851.9	40.3	339.6	7,582.9	63,851.9	40.3	339.6	7,582.9	63,851.9	40.3	339.6	7,582.9	63,851.9
8.6	84.9	249.1	2,446.6	8.6	84.9	249.1	2,446.6	8.6	84.9	249.1	2,446.6	8.6	84.9	249.1	2,446.6
7.6	77.2	185.1	1,882.3	7.6	77.2	185.1	1,882.3	7.6	77.2	185.1	1,882.3	7.6	77.2	185.1	1,882.3
5.4	52.1	108.9	1,058.7	5.4	52.1	108.9	1,058.7	5.4	52.1	108.9	1,058.7	5.4	52.1	108.9	1,058.7
5.3	41.5	90.6	715.0	5.3	41.5	90.6	715.0	5.3	41.5	90.6	715.0	5.3	41.5	90.6	715.0
3.9	36.5	113.8	1,071.3	3.9	36.5	113.8	1,071.3	3.9	36.5	113.8	1,071.3	3.9	36.5	113.8	1,071.3
7.3	65.0	227.1	2,028.1	7.3	65.0	227.1	2,028.1	7.3	65.0	227.1	2,028.1	7.3	65.0	227.1	2,028.1
4.1	39.1	110.9	1,064.9	4.1	39.1	110.9	1,064.9	4.1	39.1	110.9	1,064.9	4.1	39.1	110.9	1,064.9
3.7	36.7	99.8	980.4	3.7	36.7	99.8	980.4	3.7	36.7	99.8	980.4	3.7	36.7	99.8	980.4
6.9	80.6	454.5	5,283.3	6.9	80.6	454.5	5,283.3	6.9	80.6	454.5	5,283.3	6.9	80.6	454.5	5,283.3
6.3	47.2	202.4	1,513.6	6.3	47.2	202.4	1,513.6	6.3	47.2	202.4	1,513.6	6.3	47.2	202.4	1,513.6
4.5	45.4	69.0	688.4	4.5	45.4	69.0	688.4	4.5	45.4	69.0	688.4	4.5	45.4	69.0	688.4
5.6	63.0	125.3	1,408.1	5.6	63.0	125.3	1,408.1	5.6	63.0	125.3	1,408.1	5.6	63.0	125.3	1,408.1
11.0	112.8	625.8	6,418.7	11.0	112.8	625.8	6,418.7	11.0	112.8	625.8	6,418.7	11.0	112.8	625.8	6,418.7
6.5	66.3	275.8	2,812.5	6.5	66.3	275.8	2,812.5	6.5	66.3	275.8	2,812.5	6.5	66.3	275.8	2,812.5
8.0	78.1	366.4	3,597.5	8.0	78.1	366.4	3,597.5	8.0	78.1	366.4	3,597.5	8.0	78.1	366.4	3,597.5
2.3	29.4	63.9	808.1	2.3	29.4	63.9	808.1	2.3	29.4	63.9	808.1	2.3	29.4	63.9	808.1
3.5	37.5	42.6	450.9	3.5	37.5	42.6	450.9	3.5	37.5	42.6	450.9	3.5	37.5	42.6	450.9
3.9	36.6	88.1	830.7	3.9	36.6	88.1	830.7	3.9	36.6	88.1	830.7	3.9	36.6	88.1	830.7
2.1	15.5	70.7	526.1	2.1	15.5	70.7	526.1	2.1	15.5	70.7	526.1	2.1	15.5	70.7	526.1
11.6	107.2	248.0	2,287.9	11.6	107.2	248.0	2,287.9	11.6	107.2	248.0	2,287.9	11.6	107.2	248.0	2,287.9

表側

總 計		
道 府 縣		
北 海 道	青 森 縣	森 手 縣
東北區	青 岩 縣	山 形 縣
	山 宮 縣	城 島 縣
	福 宮 縣	城 木 縣
關東區	茨 城 縣	群 馬 縣
	埼 玉 縣	千 葉 縣
	東 京 府	神 奈 川 縣

總 計		
道 府 縣		
北 海 道	青 森 縣	森 手 縣
東北區	青 岩 縣	山 形 縣
	山 宮 縣	城 島 縣
	福 宮 縣	城 木 縣
關東區	茨 城 縣	群 馬 縣
	埼 玉 縣	千 葉 縣
	東 京 府	神 奈 川 縣

總 計		
道 府 縣		
北 海 道	青 森 縣	森 手 縣
東北區	青 岩 縣	山 形 縣
	山 宮 縣	城 島 縣
	福 宮 縣	城 木 縣
關東區	茨 城 縣	群 馬 縣
	埼 玉 縣	千 葉 縣
	東 京 府	神 奈 川 縣

水平罫線で画像を短冊切り出し
(表体画像の表レイアウト解析で
特定した改行座標情報を使う)

13	總 計
14	道府縣北海道
15	青森縣
16	岩手縣
17	秋 田 縣
18	東北画山形縣
19	宮城縣
20	福島縣
21	茨城縣
22	栃木縣
23	關群馬縣
24	東埼玉縣
25	區千葉縣
26	東 京 府
27	神奈川縣
28	新潟縣
29	北富山縣
30	陸石川縣
31	固福井縣
32	東(長野縣
33	山岐阜縣
34	區滋賀縣
35	梨縣
36	東海江海塩岡縣
37	縣
38	三重縣
39	京都府
40	近兵庫縣
41	畿大阪府
42	區奈 長 縣
43	和歌山縣

- はじめに
- OCRの活用と課題
- ocrstatsの概要
- 元祖ネ申Excelとの邂逅
- 今後の課題と展望

現状・課題・展望

● 現状

- ocrstatsで、そこそこの精度の自動テキストデータ化を実現
- 現在、『日本帝国統計年鑑』のテキストデータ化進行中

● 課題

1. レイアウト解析の自動化

- 画像レイアウト解析（表の検出） + 表構成部位解析（表頭, 表側, 表体）

2. 表頭・表側の処理

3. OCRの改善: 活字, 旧字, 「統計表」データセットの機械学習

セマンティクスの利用

374. 問 投訴海軍人及年機

年次	總數	對內投資總額		對國外投資總額		對內及對國外投資		總士官		
		總數	單位數	總數	單位數	總數	單位數			
民國 25	13,021	7,472,075	230	1,529,314	5,942	9,718,395	23,974	17,477,450	2,174	2,027,404
26	15,019	29,472,418	222	1,441,165	12,797	23,031,253	13,024	35,503,671	2,434	2,187,207
27	17,047	31,731,775	216	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
28	17,054	32,131,877	214	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
29	17,047	33,460,000	216	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
30	17,054	32,131,877	214	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
31	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
32	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
33	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
34	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
35	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
36	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
37	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
38	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
39	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
40	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
41	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
42	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
43	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
44	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
45	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
46	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
47	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
48	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
49	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
50	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
51	17,047	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313
52	17,054	—	217	1,361,175	14,676	30,370,600	14,842	39,102,375	2,716	2,494,313

375. 國有鐵道職員

年次	總數	男					女						
		藝文	學術	社會服務	技術	其他	藝文	學術	社會服務	技術	其他		
78	138,707	15,424	—	17,678	6,135	2,903	4,940	64,636	93,729	7,181	—	2,147	4,783
79	176,682	17,047	—	40,732	12,537	5,349	7,141	72,567	102,774	7,580	—	2,147	6,025
80	210,267	20,249	—	51,606	19,797	8,983	12,014	88,939	125,842	7,840	—	2,147	6,077
81	237,126	21,736	—	57,121	21,414	10,000	13,707	100,343	141,176	8,140	—	2,147	6,539
82	262,281	24,111	—	61,104	23,114	11,000	16,000	108,277	156,828	8,498	—	2,147	6,848
83	286,311	26,011	—	64,602	24,602	12,000	17,602	116,425	169,426	8,840	—	2,147	7,181
84	311,191	28,111	—	68,111	26,111	13,000	19,111	124,426	185,426	9,181	—	2,147	7,539
85	336,191	30,111	—	71,111	28,111	14,000	20,111	132,426	199,426	9,481	—	2,147	7,890
86	361,191	32,111	—	74,111	30,111	15,000	21,111	140,426	215,426	9,781	—	2,147	8,241
87	386,191	34,111	—	77,111	32,111	16,000	22,111	148,426	232,426	10,081	—	2,147	8,592
88	411,191	36,111	—	80,111	34,111	17,000	23,111	156,426	248,426	10,381	—	2,147	8,943
89	436,191	38,111	—	83,111	36,111	18,000	24,111	164,426	264,426	10,681	—	2,147	9,294
90	461,191	40,111	—	86,111	38,111	19,000	25,111	172,426	282,426	10,981	—	2,147	9,645

376. 湯 德 麟 編

[illegible]

大正十一年末總數中ニハ一、二等船ニ分關シ觀テ船員 341 ヲ含ム。昭和三年末總數中ニハ觀望取巻船ノ船員ヲ含ム。

377. 紫斑寶墨及附屬

支店名	支店番号	郵便官署			郵便月報						従事人員(人)
		郵便局 支店	郵便局 支店	郵便局 支店	郵便局 支店	郵便局 支店	郵便局 支店	郵便局 支店	郵便局 支店		
大正	1,284	5,075	14,303	17	25,562	995	1,676	5,693	53,916	1,000	
昭和	1,285	5,076	14,304	18	25,563	996	1,677	5,694	53,917	1,001	
昭和	1,286	5,077	14,305	19	25,564	997	1,678	5,695	53,918	1,002	
昭和	1,287	5,078	14,306	20	25,565	998	1,679	5,696	53,919	1,003	
昭和	1,288	5,079	14,307	21	25,566	999	1,680	5,697	53,920	1,004	
昭和	1,289	5,080	14,308	22	25,567	1,000	1,681	5,698	53,921	1,005	
昭和	1,290	5,081	14,309	23	25,568	1,001	1,682	5,699	53,922	1,006	
昭和	1,291	5,082	14,310	24	25,569	1,002	1,683	5,700	53,923	1,007	
昭和	1,292	5,083	14,311	25	25,570	1,003	1,684	5,701	53,924	1,008	
昭和	1,293	5,084	14,312	26	25,571	1,004	1,685	5,702	53,925	1,009	
昭和	1,294	5,085	14,313	27	25,572	1,005	1,686	5,703	53,926	1,010	
昭和	1,295	5,086	14,314	28	25,573	1,006	1,687	5,704	53,927	1,011	
昭和	1,296	5,087	14,315	29	25,574	1,007	1,688	5,705	53,928	1,012	
昭和	1,297	5,088	14,316	30	25,575	1,008	1,689	5,706	53,929	1,013	
昭和	1,298	5,089	14,317	31	25,576	1,009	1,690	5,707	53,930	1,014	
昭和	1,299	5,090	14,318	32	25,577	1,010	1,691	5,708	53,931	1,015	
昭和	1,300	5,091	14,319	33	25,578	1,011	1,692	5,709	53,932	1,016	
昭和	1,301	5,092	14,320	34	25,579	1,012	1,693	5,710	53,933	1,017	
昭和	1,302	5,093	14,321	35	25,580	1,013	1,694	5,711	53,934	1,018	
昭和	1,303	5,094	14,322	36	25,581	1,014	1,695	5,712	53,935	1,019	
昭和	1,304	5,095	14,323	37	25,582	1,015	1,696	5,713	53,936	1,020	
昭和	1,305	5,096	14,324	38	25,583	1,016	1,697	5,714	53,937	1,021	
昭和	1,306	5,097	14,325	39	25,584	1,017	1,698	5,715	53,938	1,022	
昭和	1,307	5,098	14,326	40	25,585	1,018	1,699	5,716	53,939	1,023	
昭和	1,308	5,099	14,327	41	25,586	1,019	1,700	5,717	53,940	1,024	
昭和	1,309	5,100	14,328	42	25,587	1,020	1,701	5,718	53,941	1,025	
昭和	1,310	5,101	14,329	43	25,588	1,021	1,702	5,719	53,942	1,026	
昭和	1,311	5,102	14,330	44	25,589	1,022	1,703	5,720	53,943	1,027	
昭和	1,312	5,103	14,331	45	25,590	1,023	1,704	5,721	53,944	1,028	
昭和	1,313	5,104	14,332	46	25,591	1,024	1,705	5,722	53,945	1,029	
昭和	1,314	5,105	14,333	47	25,592	1,025	1,706	5,723	53,946	1,030	
昭和	1,315	5,106	14,334	48	25,593	1,026	1,707	5,724	53,947	1,031	
昭和	1,316	5,107	14,335	49	25,594	1,027	1,708	5,725	53,948	1,032	
昭和	1,317	5,108	14,336	50	25,595	1,028	1,709	5,726	53,949	1,033	
昭和	1,318	5,109	14,337	51	25,596	1,029	1,710	5,727	53,950	1,034	
昭和	1,319	5,110	14,338	52	25,597	1,030	1,711	5,728	53,951	1,035	
昭和	1,320	5,111	14,339	53	25,598	1,031	1,712	5,729	53,952	1,036	
昭和	1,321	5,112	14,340	54	25,599	1,032	1,713	5,730	53,953	1,037	
昭和	1,322	5,113	14,341	55	25,600	1,033	1,714	5,731	53,954	1,038	
昭和	1,323	5,114	14,342	56	25,601	1,034	1,715	5,732	53,955	1,039	
昭和	1,324	5,115	14,343	57	25,602	1,035	1,716	5,733	53,956	1,040	
昭和	1,325	5,116	14,344	58	25,603	1,036	1,717	5,734	53,957	1,041	
昭和	1,326	5,117	14,345	59	25,604	1,037	1,718	5,735	53,958	1,042	
昭和	1,327	5,118	14,346	60	25,605	1,038	1,719	5,736	53,959	1,043	
昭和	1,328	5,119	14,347	61	25,606	1,039	1,720	5,737	53,960	1,044	
昭和	1,329	5,120	14,348	62	25,607	1,040	1,721	5,738	53,961	1,045	
昭和	1,330	5,121	14,349	63	25,608	1,041	1,722	5,739	53,962	1,046	
昭和	1,331	5,122	14,350	64	25,609	1,042	1,723	5,740	53,963	1,047	
昭和	1,332	5,123	14,351	65	25,610	1,043	1,724	5,741	53,964	1,048	
昭和	1,333	5,124	14,352	66	25,611	1,044	1,725	5,742	53,965	1,049	
昭和	1,334	5,125	14,353	67	25,612	1,045	1,726	5,743	53,966	1,050	
昭和	1,335	5,126	14,354	68	25,613	1,046	1,727	5,744	53,967	1,051	
昭和	1,336	5,127	14,355	69	25,614	1,047	1,728	5,745	53,968	1,052	
昭和	1,337	5,128	14,356	70	25,615	1,048	1,729	5,746	53,969	1,053	
昭和	1,338	5,129	14,357	71	25,616	1,049	1,730	5,747	53,970	1,054	
昭和	1,339	5,130	14,358	72	25,617	1,050	1,731	5,748	53,971	1,055	
昭和	1,340	5,131	14,359	73	25,618	1,051	1,732	5,749	53,972	1,056	
昭和	1,341	5,132	14,360	74	25,619	1,052	1,733	5,750	53,973	1,057	
昭和	1,342	5,133	14,361	75	25,620	1,053	1,734	5,751	53,974	1,058	
昭和	1,343	5,134	14,362	76	25,621	1,054	1,735	5,752	53,975	1,059	
昭和	1,344	5,135	14,363	77	25,622	1,055	1,736	5,753	53,976	1,060	
昭和	1,345	5,136	14,364	78	25,623	1,056	1,737	5,754	53,977	1,061	
昭和	1,346	5,137	14,365	79	25,624	1,057	1,738	5,755	53,978	1,062	
昭和	1,347	5,138	14,366	80	25,625	1,058	1,739	5,756	53,979	1,063	
昭和	1,348	5,139	14,367	81	25,626	1,059	1,740	5,757	53,980	1,064	
昭和	1,349	5,140	14,368	82	25,627	1,060	1,741	5,758	53,981	1,065	
昭和	1,350	5,141	14,369	83	25,628	1,061	1,742	5,759	53,982	1,066	
昭和	1,351	5,142	14,370	84	25,629	1,062	1,743	5,760	53,983	1,067	
昭和	1,352	5,143	14,371	85	25,630	1,063	1,744	5,761	53,984	1,068	
昭和	1,353	5,144	14,372	86	25,631	1,064	1,745	5,762	53,985	1,069	
昭和	1,354	5,145	14,373	87	25,632	1,065	1,746	5,763	53,986	1,070	
昭和	1,355	5,146	14,374	88	25,633	1,066	1,747	5,764	53,987	1,071	
昭和	1,356	5,147	14,375	89	25,634	1,067	1,748	5,765	53,988	1,072	
昭和	1,357	5,148	14,376	90	25,635	1,068	1,749	5,766	53,989	1,073	
昭和	1,358	5,149	14,377	91	25,636	1,069	1,750	5,767	53,990	1,074	
昭和	1,359	5,150	14,378	92	25,637	1,070	1,751	5,768	53,991	1,075	
昭和	1,360	5,151	14,379	93	25,638	1,071	1,752	5,769	53,992	1,076	
昭和	1,361	5,152	14,380	94	25,639	1,072	1,753	5,770	53,993	1,077	
昭和	1,362	5,153	14,381	95	25,640	1,073	1,754	5,771	53,994	1,078	
昭和	1,363	5,154	14,382	96	25,641	1,074	1,755	5,772	53,995	1,079	
昭和	1,364	5,155	14,383	97	25,642	1,075	1,756	5,773	53,996	1,080	
昭和	1,365	5,156	14,384	98	25,643	1,076	1,757	5,774	53,997	1,081	
昭和	1,366	5,157	14,385	99	25,644	1,077	1,758	5,775	53,998	1,082	
昭和	1,367	5,158	14,386	100	25,645	1,078	1,759	5,776	53,999	1,083	
昭和	1,368	5,159	14,387	101	25,646	1,079	1,760	5,777	54,000	1,084	
昭和	1,369	5,160	14,388	102	25,647	1,080	1,761	5,778	54,001	1,085	
昭和	1,370	5,161	14,389	103	25,648	1,081	1,762	5,779	54,002	1,086	
昭和	1,371	5,162	14,390	104	25,649	1,082	1,763	5,780	54,003	1,087	
昭和	1,372	5,163	14,391	105	25,650	1,083	1,764	5,781	54,004	1,088	
昭和	1,373	5,164	14,392	106	25,651	1,084	1,765	5,782	54,005	1,089	
昭和	1,374	5,165	14,393	107	25,652	1,085	1,766	5,783	54,006	1,090	
昭和	1,375	5,166	14,394	108	25,653	1,086	1,767	5,784	54,007	1,091	
昭和	1,376	5,167	14,395	109	25,654	1,087	1,768	5,785	54,008	1,092	
昭和	1,377	5,168	14,396	110	25,655	1,088	1,769	5,786	54,009	1,093	
昭和	1,378	5,169	14,397	111	25,656	1,089	1,770	5,787	54,010	1,094	
昭和	1,379	5,170	14,398	112	25,657	1,090	1,771	5,788	54,011	1,095	
昭和	1,380	5,171	14,399	113	25,658	1,091	1,772	5,789	54,012	1,096	
昭和	1,381	5,172	14,400	114	25,659	1,092	1,773	5,790	54,013	1,097	
昭和	1,382	5,173	14,401	115	25,660	1,093	1,774	5,791	54,014	1,098	
昭和	1,383	5,174	14,402	116	25,661	1,094	1,775	5,792	54,015	1,099	
昭和	1,384	5,175	14,403	117	25,662	1,095	1,776	5,793	54,016	1,100	
昭和	1,385	5,176	14,404	118	25,663	1,096	1,777	5,794	54,017	1,101	
昭和	1,386	5,177	14,405	119	25,664	1,097	1,778	5,795	54,018	1,102	
昭和	1,387	5,178	14,406	120	25,665	1,098	1,779	5,796	54,019	1,103	
昭和	1,388	5,179	14,407	121	25,666	1,099	1,780	5,797	54,020	1,104	
昭和	1,389	5,180	14,408	122	25,667	1,100	1,781	5,798	54,021	1,105	
昭和	1,390	5,181	14,409	123	25,668	1,101	1,782	5,799	54,022	1,106	
昭和	1,391	5,182	14,410	124	25,669	1,102	1,783	5,800	54,023	1,107	
昭和	1,392	5,183	14,411	125	25,670	1,103	1,784	5,801	54,024	1,108	
昭和	1,393	5,184	14,412	126	25,671	1,104	1,785	5,802	54,025	1,109	
昭和	1,394	5,185	14,413	127	25,672	1,105	1,786	5,803	54,026	1,110	

前年暫定額の暫定係数が過去中の過去年度1及び過去18を含む。
 過去年度中の過去年度は、07を含む。内港の割合は2年未満。

374. 現役陸海軍人及年俸

[illegible]

375. 國有鐵道職員

[illegible]

376. 通信工程

國家行政學院										
年	次	成績	一等獎	三等獎	總分	英文	國文	算術	常識	
學 生 名 單										
		男 生		女 生		共 計		最高	最低	
		姓名		姓名		姓名		分數	分數	
第一學期	1	105.55	23.22	11.55	66.78	5.13	48.74	16.33	55.53	15.55
	2	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	3	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	4	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	5	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	6	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	7	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	8	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	9	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77
	10	137.22	29.77	16.22	81.23	5.77	51.22	18.23	57.22	16.77

[illegible]

377. 警察官署及職員

[illegible][illegible]

神太無失中ニハ無主移長 33 ヲ合ム。内地外ハ昭和三年末ナリ。

現状・課題・展望

● 今後の展望

- テキストデータ化する統計表の拡大
 - 『日本帝国統計年鑑』, 各省庁統計表, 府県統計書など
- (政府) 統計レポジトリの作成
 - 統計の目録, メタデータ, 電子画像, テキストデータ, 作業情報の集約化
- クラウドソーシング型プロジェクトの展開?
 - 校正をクラウドソーシング (例: 「みんなで翻刻 <https://honkoku.org/> 」)