



PicMiner

智慧字元辨識系統

丘祐瑋 – David Chiu

EMAIL: david@largitdata.com

網站: www.largitdata.com

電話: +886929094381

文字辨識系統分類

■ 光學文字辨識系統

(Optical Character Recognition)

處理掃描後文件資料的圖檔，對圖檔上的印刷體進行辨識並分析文件內容的系統

■ 智慧型字元辨識系統

(Intelligent Character Recognition)

針對手寫辨識，處理不同字型及書寫風格，此外在辨識時系統也會繼續學習已增加自己的準確率

人工智慧讓辨識精度大大耀進



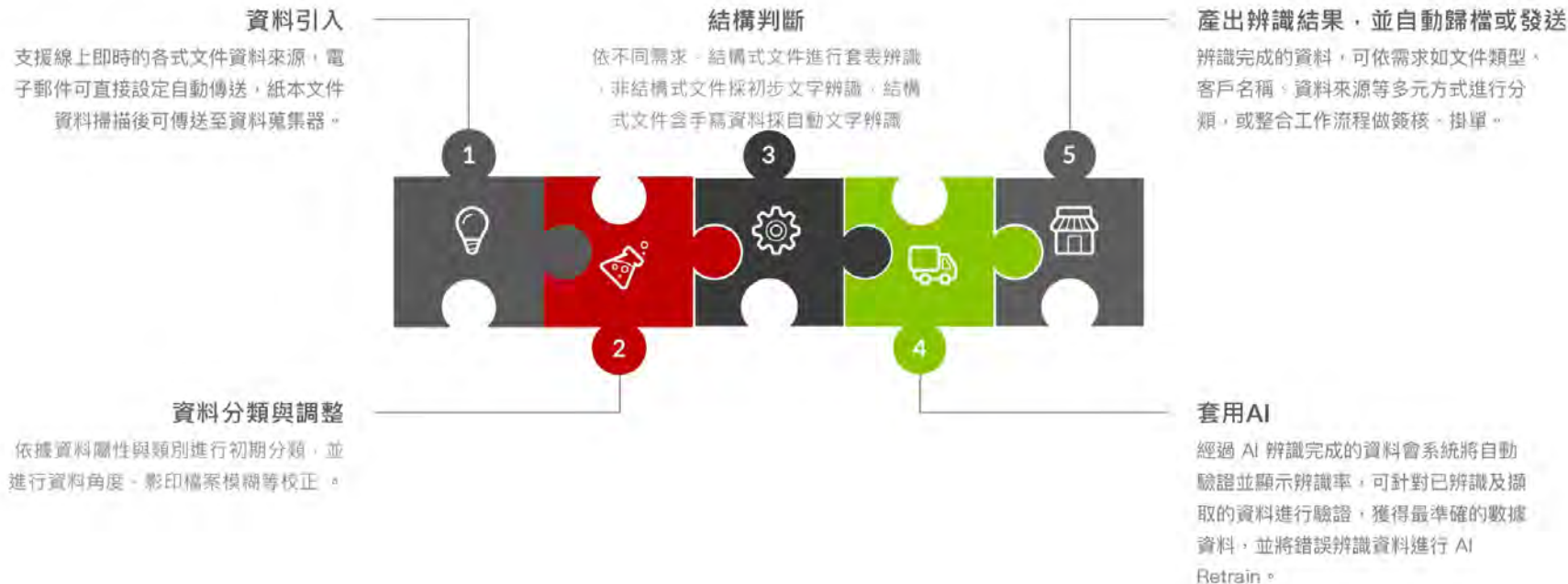
Real-Time Recognition of Handwritten Chinese Characters Spanning a Large Inventory of 30,000 Characters

Vol. 1, Issue 5 • September 2017

by Handwriting Recognition Team

Source: <https://machinelearning.apple.com/2017/09/12/handwriting.html>

ICR 解決方案-資料擷取辨識流程



手寫文字辨識範例

原圖



辨識結果

有些事就是這麼無奈 遲到一分鐘就算遲到 加班一分鐘不算加班 吃一口就會變胖 少吃一口 不會變瘦! HONY a 2017'02'02 八耐舜子

文字	座標
有些事就是這麼無奈	[(x: 415, y: 149), (x: 1087, y: 149), (x: 1087, y: 258), (x: 415, y: 258)]
遲到一分鐘就算遲到	[(x: 619, y: 309), (x: 1129, y: 309), (x: 1129, y: 385), (x: 619, y: 385)]
加班	[(x: 632, y: 396), (x: 759, y: 398), (x: 758, y: 475), (x: 631, y: 473)]
-	[(x: 767, y: 397), (x: 789, y: 397), (x: 788, y: 474), (x: 766, y: 474)]
分鐘不算加班	[(x: 798, y: 398), (x: 1169, y: 403), (x: 1168, y: 480), (x: 797, y: 475)]
吃一口就會變胖	[(x: 683, y: 527), (x: 1128, y: 527), (x: 1128, y: 608), (x: 683, y: 608)]

多元字樣辨識範例

demo images		
	available	Available
	shakeshack	SHARESHACK
	london	London
	greenstead	Greenstead
	toast	TOAST

	merry	MERRY
	underground	underground
	ronaldo	RONALDO
	bally	BALLY
	university	UNIVERSITY

財報識別

行	目	109年9月30日			108年12月31日			108年9月30日		
		金額	單位	匯率	金額	單位	匯率	金額	單位	匯率
1100	現金及現金等價物(附註六)	\$ 454,195,113	25		\$ 455,399,336	20		\$ 452,436,300	21	
1110	透過損益按公允價值衡量之金融資產(附註七)	\$ 252,681	-		\$ 376,839	-		\$ 322,089	-	
1120	透過其他綜合損益按公允價值衡量之金融資產(附註六)	\$ 128,608,752	6		\$ 127,396,577	6		\$ 132,133,452	6	
1136	附屬機構或事業之金融資產(附註八)	\$ 2,095,162	-		\$ 290,800	-		\$ -	-	
1139	應收之金融資產(附註九)	\$ -	-		\$ -	-		\$ 27,000	-	
1170	應收票據及帳款(附註六)	\$ 143,760,935	4		\$ 140,006,509	4		\$ 144,340,859	7	
1180	應收關係人款項(附註六)	\$ -	-		\$ -	-		\$ -	-	
1210	其他流動資產(附註六)	\$ -	-		\$ -	-		\$ -	-	
1310	其他非流動資產(附註六)	\$ -	-		\$ -	-		\$ -	-	
1476	其他流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
1479	其他非流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
15XX	流動資產合計	\$ 726,564,880	25		\$ 723,802,224	20		\$ 728,930,691	21	
1600	非流動資產	\$ -	-		\$ -	-		\$ -	-	
1606	透過損益按公允價值衡量之金融資產(附註七)	\$ -	-		\$ -	-		\$ -	-	
1609	透過其他綜合損益按公允價值衡量之金融資產(附註六)	\$ -	-		\$ -	-		\$ -	-	
1610	附屬機構或事業之金融資產(附註八)	\$ -	-		\$ -	-		\$ -	-	
1619	應收之金融資產(附註九)	\$ -	-		\$ -	-		\$ -	-	
1670	應收票據及帳款(附註六)	\$ -	-		\$ -	-		\$ -	-	
1680	應收關係人款項(附註六)	\$ -	-		\$ -	-		\$ -	-	
1689	其他非流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
1690	其他非流動資產(附註六)	\$ -	-		\$ -	-		\$ -	-	
1699	非流動資產合計	\$ -	-		\$ -	-		\$ -	-	
17XX	資產總計	\$ 726,564,880	25		\$ 723,802,224	20		\$ 728,930,691	21	
1800	負債	\$ -	-		\$ -	-		\$ -	-	
1806	短期借款(附註六)	\$ -	-		\$ -	-		\$ -	-	
1809	應付票據及帳款(附註六)	\$ -	-		\$ -	-		\$ -	-	
1810	應付關係人款項(附註六)	\$ -	-		\$ -	-		\$ -	-	
1819	其他流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
1870	其他非流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
1890	其他非流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
1899	負債合計	\$ -	-		\$ -	-		\$ -	-	
1900	權益	\$ -	-		\$ -	-		\$ -	-	
1906	資本公積金(附註六)	\$ -	-		\$ -	-		\$ -	-	
1909	盈餘公積金(附註六)	\$ -	-		\$ -	-		\$ -	-	
1910	未分配盈餘(附註六)	\$ -	-		\$ -	-		\$ -	-	
1919	其他權益(附註六)	\$ -	-		\$ -	-		\$ -	-	
1970	其他非流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
1990	其他非流動負債(附註六)	\$ -	-		\$ -	-		\$ -	-	
1999	權益合計	\$ -	-		\$ -	-		\$ -	-	
20XX	負債及權益合計	\$ 726,564,880	25		\$ 723,802,224	20		\$ 728,930,691	21	

透過OCR 技術識別文字後，
再將資料直接匯出成Excel 檔案，
大幅減少財報key 單、審閱的時間

A	B	C	D	E	F	G	H	I
1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	27
30	31	32	33	34	35	36	37	38
43	44	45	46	47	48	49	50	51
54	55	56	57	58	59	60	61	62
67	68	69	70	71	72	73	74	75
82	83	84	85	86	87	88	89	90
97	98	99	100	101	102	103	104	105
112	113	114	115	116	117	118	119	120
127	128	129	130	131	132	133	134	135
142	143	144	145	146	147	148	149	150
157	158	159	160	161	162	163	164	165
172	173	174	175	176	177	178	179	180
187	188	189	190	191	192	193	194	195
202	203	204	205	206	207	208	209	210
217	218	219	220	221	222	223	224	225
232	233	234	235	236	237	238	239	240
247	248	249	250	251	252	253	254	255
262	263	264	265	266	267	268	269	270
277	278	279	280	281	282	283	284	285
292	293	294	295	296	297	298	299	300
307	308	309	310	311	312	313	314	315
322	323	324	325	326	327	328	329	330
337	338	339	340	341	342	343	344	345
352	353	354	355	356	357	358	359	360
367	368	369	370	371	372	373	374	375
382	383	384	385	386	387	388	389	390
397	398	399	400	401	402	403	404	405
412	413	414	415	416	417	418	419	420
427	428	429	430	431	432	433	434	435
442	443	444	445	446	447	448	449	450
457	458	459	460	461	462	463	464	465
472	473	474	475	476	477	478	479	480
487	488	489	490	491	492	493	494	495
499	500	501	502	503	504	505	506	507
512	513	514	515	516	517	518	519	520
527	528	529	530	531	532	533	534	535
542	543	544	545	546	547	548	549	550
557	558	559	560	561	562	563	564	565
572	573	574	575	576	577	578	579	580
587	588	589	590	591	592	593	594	595
599	600	601	602	603	604	605	606	607
612	613	614	615	616	617	618	619	620
627	628	629	630	631	632	633	634	635
642	643	644	645	646	647	648	649	650
657	658	659	660	661	662	663	664	665
672	673	674	675	676	677	678	679	680
687	688	689	690	691	692	693	694	695
699	700	701	702	703	704	705	706	707
712	713	714	715	716	717	718	719	720
727	728	729	730	731	732	733	734	735
742	743	744	745	746	747	748	749	750
757	758	759	760	761	762	763	764	765
772	773	774	775	776	777	778	779	780
787	788	789	790	791	792	793	794	795
799	800	801	802	803	804	805	806	807
812	813	814	815	816	817	818	819	820
827	828	829	830	831	832	833	834	835
842	843	844	845	846	847	848	849	850
857	858	859	860	861	862	863	864	865
872	873	874	875	876	877	878	879	880
887	888	889	890	891	892	893	894	895
899	900	901	902	903	904	905	906	907
912	913	914	915	916	917	918	919	920
927	928	929	930	931	932	933	934	935
942	943	944	945	946	947	948	949	950
957	958	959	960	961	962	963	964	965
972	973	974	975	976	977	978	979	980
987	988	989	990	991	992	993	994	995
999	1000	1001	1002	1003	1004	1005	1006	1007

11



電視監控



後，發現
器中又有
軸心斷裂！
17原定
出查朝
，恐再
延…
56支單
檢測30支
未能檢測

德華最愛臺灣線！？

宜蘭西德山
84.5

說·播·批評

更多相關新聞請訂

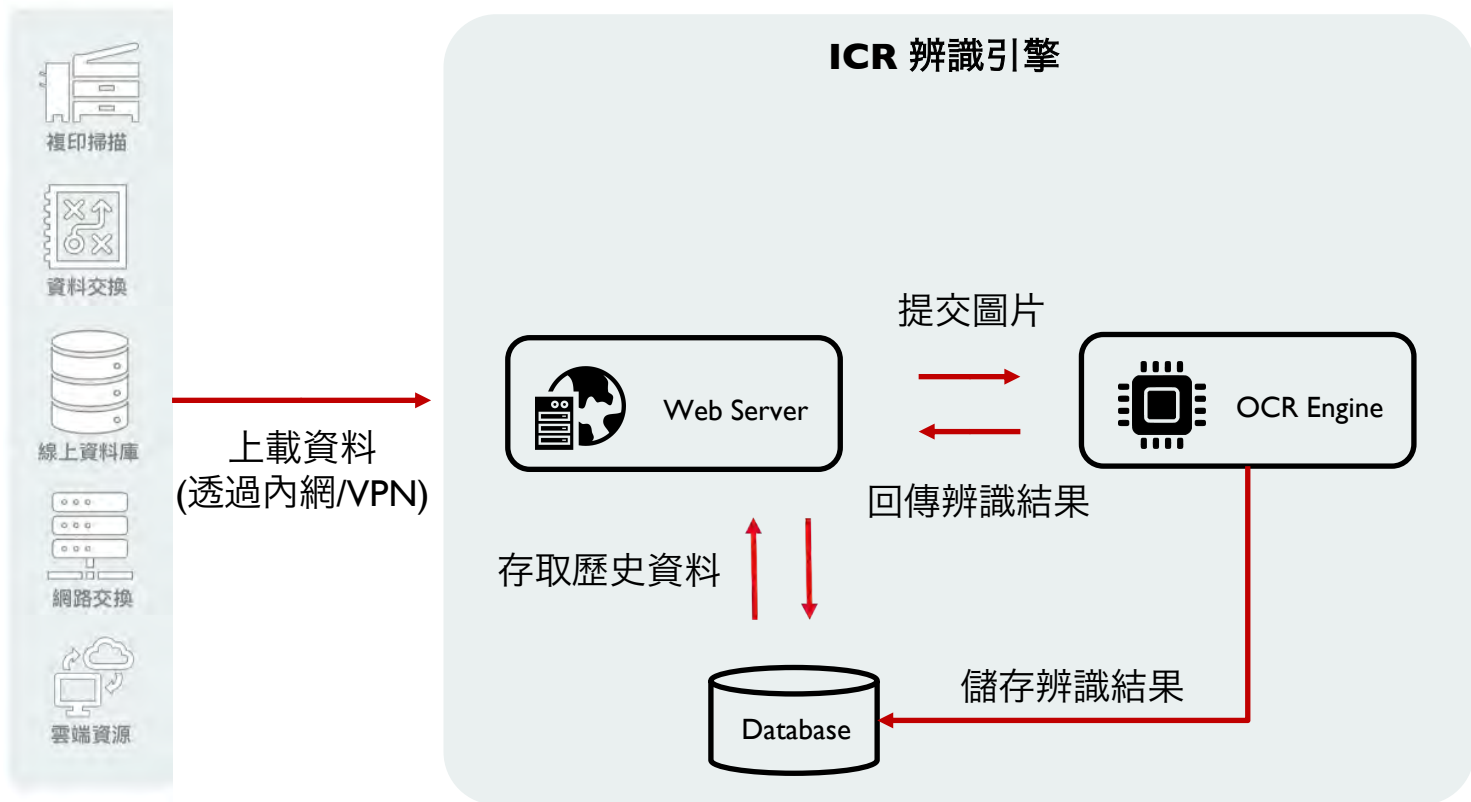
張雅琴挑戰新聞

年代新聞

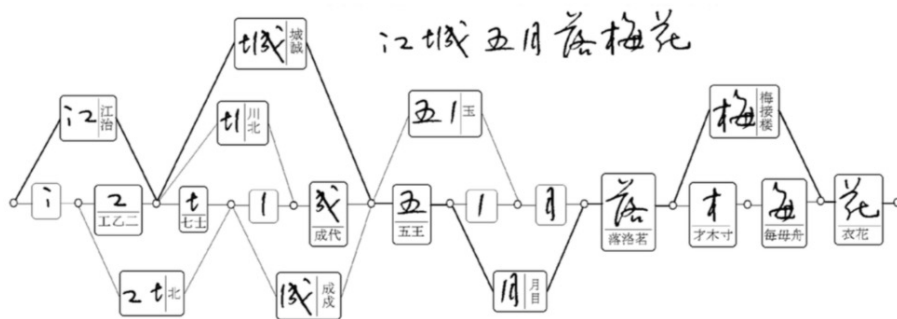
年代

利用OCR 針對視頻中的文字進行分析
即時監控最新電視新聞

ICR 應用端資料處理架構



AI 解決方案帶來的效益



辨識目標
30000字
模型大小
15MB

辨識效果：Top-1
85%

辨識效果：Top-4
95%

辨識效果：Top-10
97%