

電子計算機의 한글 入出力에 關한 研究現況과

한글 반 풀어 쓰기提案 (I)

安 秀 桔*

1. 序 論

韓國에도 1960年 末頃부터 적지 않은 臺數의 電子計算機가 導入되어 活用되고 있다.

이들 電子計算機는 初期의 잘못된 一般의 期待와 같이 거의 超人的인 神秘的인 일을 行하는데 使用되고 있는 것은 아니나, 計算 速度가 빨라서 計算量이 많아도 時間이 적게 걸리는 點과 大量의 計算을 해도 錯誤가 없다는 點에서 代置가 不可能한 獨自的인 役割을 擔當하고 있어서 지금에 와서는 이미 電子計算機의 도움이 없이는 事實上 行할 수 없는 業務가 增加해 가고 있다.

그럼에도 不拘하고 우리들 日常生活은 直接的인 影響을 거의 받지 않고 있다고 생각하기 쉬운 程度로 우리와는 關聯이 적어 보인다. 또 나아가서는 電子計算機의 導入趨勢도 最近의 傾向으로 보아서는 鈍化한 느낌이다.

2. 電子計算機와 人間의 Interface

現在로서 電子計算機 (以上 電算機라 略할)의 強點은 速度와 正確성에 있다. 人間은 이에 比하면, 너무 느리기도 하지만 일의 量이 어느 限界를 넘으면 또는 고단하면 error가 續出하게 된다 따라서 二個의 計算過程 中間에서 人間이 介入하였을 때, 예를 들어 桌上計算機를 가지고 어느 두數를 곱한 것에 第三의 數를 더한 경우와 같이 첫 段階의 計算 結果를 눈으로 읽고 손으로 종이에 記錄한 다음 다시손으로 key를 눌러 計算機에 入力시켜서 다음 段階의 計算을 行할 경우 計算機는 上記 二大強點을 잃어 버리게 되는 것

이다.

科學한 發達이 오늘날 렌즈를 取扱하는 사람의 立場에서 사람의 눈을 볼 때 그것이 너무나도 엉성한 裝置로 보이는 것과 마찬가지로 電算機는 nano 秒單位(10^{-9} 秒)의 計算 速度가 入出力機械 때문에 速度가 늦어지고 더욱 人間의 判別能力의 速度 制限때문에 다시 能力의 低下를 일으키면서 error가 混入하게 되는 것이다. 따라서 電算機의 能率向上을 爲해서는 人間이 中間에 介入할 必要가 없게 만들어야 한다.

그럼에도 電算機의 最大의 活用은 電算機關係 사람들이 介入할 必要없이 되도록이면 많은 一般 人들이 容易하게 電算機와의 接觸을 할 수 있게 하는데 있고, 이는 또한 機械語의 境遇와 같이 人間에게는 어렵고 機械에게는 쉬운 作動方法이 아니고, 오히려 人間에게 쉬운 言語, 다시 말해서 韓國人의 境遇에는 한글을 통한 機械와의 interface가 可能하게 되어야 한다.

電算機가 여러가지 必要한 計算을 行하고, 아울러 告知書를 印刷하고 住所도 印刷하여 郵送하고, 다시 領收證을 發付하며 納付여부를 再確認하고 督促狀도 같은 模樣으로 郵送하는 段階까지 또는 文書를 찍어 주고 新聞을 印刷해 주고 부랴운管에서 메모를 display 해 주는 段階까지 다 해 줄때 電算機는 제대로 많이 活用되고 있다고 말할 수 있고, 또 한편 제대로의 投資効率が 올랐다고 말할 수 있게 되는 것이다.

3. 電算機로 부터의 文字出入

現在 電算機로 부터의 文字出力은 前부터 쓰여 온 bar printer 以外에 line printer 등으로 文字가

* 正會員, 서울大學校 工科大學

6단위 BCD 코드

Bits						열			
b ₆	b ₅	b ₄	b ₃	b ₂	b ₁	0	1	2	3
0	0	0	0	0	0	0	+	—	△
0	0	0	1	0	0	1	A	J	/
0	0	1	0	0	0	2	B	K	S
0	0	1	1	0	0	3	C	L	T
0	1	0	0	0	0	4	D	M	U
0	1	0	1	0	0	5	E	N	V
0	1	1	0	0	0	6	F	O	W
0	1	1	1	0	0	7	G	P	X
1	0	0	0	0	0	8	H	Q	Y
1	0	0	1	0	0	9	I	R	Z
1	0	1	0	0	0	10	:	<	!
1	0	1	1	0	0	11	=	.	\$
1	1	0	0	0	0	12)	*
1	1	0	1	0	0	13	&	^	#
1	1	1	0	0	0	14	%	"	/
1	1	1	1	0	0	15	{	:	>

第1表 6單位 BCD 符號

印刷되어 나오는 것으로 代表되는 hard copy 와 Braun 管 등에 비치게 하는 것으로 代表되는 따라서 印刷物의 形態로 남지 않는 soft copy 의 二種으로 代表되는데, 이 두가지가 다 Alphabet 에 對해서는 거의 解決되어 있다. 前者의 境遇는 drum printer (UNIVAC)와 train printer (CDC, IBM)가 있는데, 우리가 比較的 볼 機會가 많은 電算機 出力結果로서 活字의 크기가 항상 一定하고 大文字밖에 없다.

電算機로부터 printer 에 對해서 各文字의 一連 番號에 該當되는 binary(二進法) 符號에 依해서 적어야 할 文字가 指定되는데, 初期에는 6單位인 BCD 符號를 쓰다가 近來에는 8單位의 BCD 또는 EBCDIC 符號를 使用하고 있다. (第1表 및 第2表 참조)

Braun管 (以下 CRT 라 略함) 등에 依하여 눈 에 보이게 하는 文字 display 의 境遇는 畫面上을 光點이 左에서 右로 한字分 한字分 階段的으로 뛰어 가게 하기 爲해서 水平偏向과 垂直偏向板에

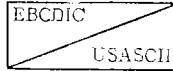
第一圖와 같이 階段的 波形을 가진 電壓을 印加한다. 電壓이 一定하게 머무른 時間동안 CRT 光點도 그 screen 上에 멈추어 서 있는데 그 時間사이에 微細하게 다시 그 點을 中心으로 하여 左右上下로 走査시키고 明暗을 바꿔서 數字나 文字가 나타나게 하는데 어느 點은 어둡게 하고, 어느 點을 밝게 하는가 하는 것은 文字發生器에서 나오는 binary 信號에 依해서 한다. 이 文字發生器는 實際文字의 그림을 5列 7行으로 分解하여 各各의 面素가 어두운가 밝은 가에 따라 "0" 또는 "1"의 binary 信號를 내주는 read only memory 이다.

4. 文字發生器(Character Generator)

最近 MOS 의 發達에 따라 IC 의 package density가 커져서 24 pin 程度의 하나의 IC package에 ASCII font(活字)를 다 收容시킨 LSI 製品들이 있다.

前述한바와 같이 spot가 段階的으로 走査되는

EBCDIC TO USASCH (Continued)



				0	1	2	3	4	5	6	7
				1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0
4	5	6	7	COL							
↓	↓	↓	↓	↓							
ROW				0	1	2	3	4	5	6	7
0	0	0	0								0
0	0	0	1	a	j			A	J		1
0	0	1	0	b	k	s		B	K	S	2
0	0	1	1	c	l	t		C	L	T	3
0	1	0	0	d	m	u		D	M	U	4
0	1	0	1	e	n	v		E	N	V	5
0	1	1	0	f	o	w		F	O	W	6
0	1	1	1	g	p	x		G	P	X	7
0	0	0	0	h	q	y		H	Q	Y	8
1	0	0	1	i	r	z		I	R	Z	9
1	0	1	0								
1	0	1	1								
1	1	0	0								
1	1	0	1								
1	1	1	0								
1	1	1	1								

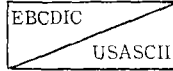
第2表 2-1 EBCDIC/USASCH 符號

境遇는 그 點을 中心으로 하여 다시 TV의 境遇와 같이 左右로 急速度로 그리고 上下로 천천히 走査하기 때문에 5列 7行으로 記憶되어 있는 (明暗에 該當되는) "1" 또는 "0"의 信號가 行方向 順序로 나오게 되어 있는 row output character generator의 實際 商品 例로서는 TMS 2403,

TMS 2404 (Texas Instruments), 2513NX/CM 2140 (Signetics), CRC 3504, (大文字), CRC 3505 (小文字) (Collins), MCM 1120 (Motorola), FDR 116Z1 (Philips) 등이 있다.

近來에 와서는 文字와 圖形을 같은 CRT screen에 display 할 수 있게 하기 위해서 spot를 左右

EBCDIC TO USASCII



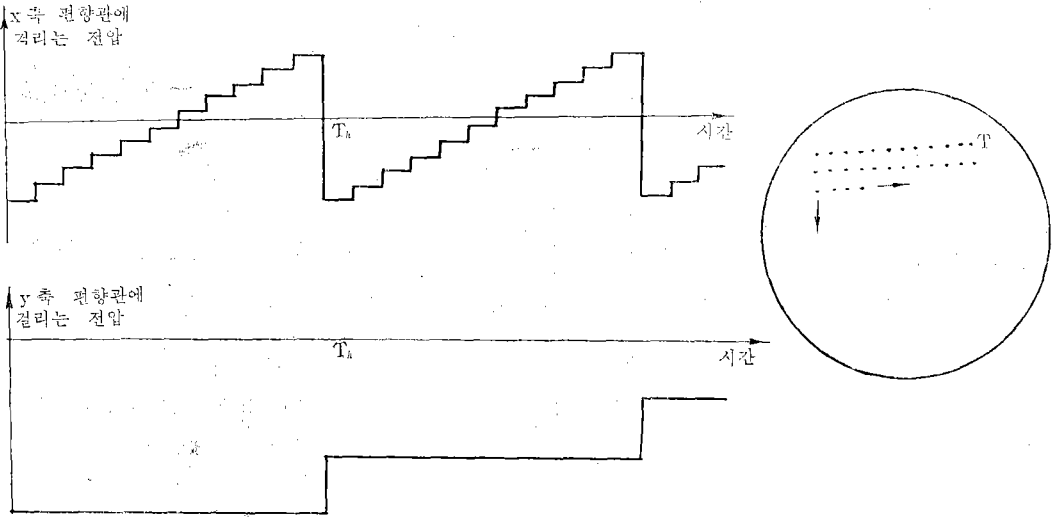
0 → 1 → 2 → 3 →		0 0 0	0 0 1	0 0 1 0	0 0 1 1	0 1 0 0	0 1 0 1	0 1 1 0	0 1 1 1			
4 ↓	5 ↓	6 ↓	7 ↓	ROW →								
				0	1	2	3	4	5	6	7	
0	0	0	0	0	NUL	DLE	DC		SP	&		
					NUL	DLE	NUL		SP	&		
0	0	0	1	1	SOH	DC1	SOS			/		
					SOH	DC1	NUL			/		
0	0	1	0	2	STX	DC2	FS	SYN				
					STX	DC2	NUL	SYN				
0	0	1	1	3	ETX	DC3						
					ETX	DC3						
0	1	0	0	4	PF	RES	BYP	PN				
					NUL	NUL	NUL	NUL				
0	1	0	1	5	HT	NL	LF	RS				
					HT	NUL	LF	NUL				
0	1	1	0	6	LC	BS	EOB	UC				
					NUL	BS	ETE	NUL				
0	1	1	1	7	DEL	IL	PRE	EOT				
					DEL	NUL	ESC	EOT				
1	0	0	0	8		CAN						
						CAN						
1	0	0	1	9		EM						
						EM						
1	0	1	0	10	SMM	CC	SM		!	:		
					NUL	NUL	NUL		NUL	!	:	
1	0	1	1	11	VT				\$		#	
					VT				\$		#	
1	1	0	0	12	FF	IFS		DC4	<	%	Ⓐ	
					FF	FS		DC4	<	%	Ⓐ	
1	1	0	1	13	CR	IGS	ENO	NAK	()	-	/
					CR	GS	ENO	NAK	()	-	/
1	1	1	0	14	SO	IRS	ACK		+	;	>	
					SO	RS	ACK		+	;	>	
1	1	1	1	15	SI	IUS	BEL	SUB		∟	?	?
					SI	US	BEL	SUB		∟	NUL	?

--Continued

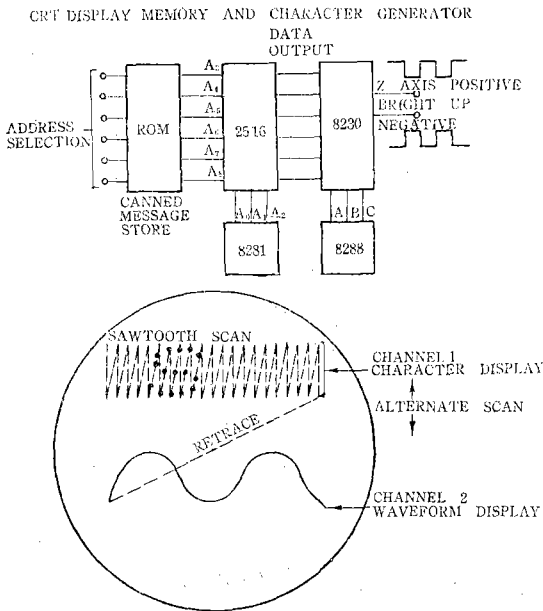
第2表 2-2 EBCDIC/USASCII 符號(continued)

로 連續的으로 走査시키면서 (勿論 上에서 下로 는 더욱 느리게 階段的으로 走査하고 있다.) 다 시 上下로 빠르게 saw tooth wave에 依해서 微細走査하면서 (第2圖) 上下方向으로 文字의 圖形을 그려 가는, 따라서 5列 7行 面素를 垂直方 向 順序에 따라서 binary 符號를 送出해 주는

series character generator가 있는데 그 製品 例로서는 TMS 4103 (T.I.), 2516 NX/CM2510 (6×8 signetics), MCM 1130 column select, USACII character generator (Motorola)등이 있다. 이러한 character generator에 收容되어 있는 文數字를 第3圖에 提示한다.



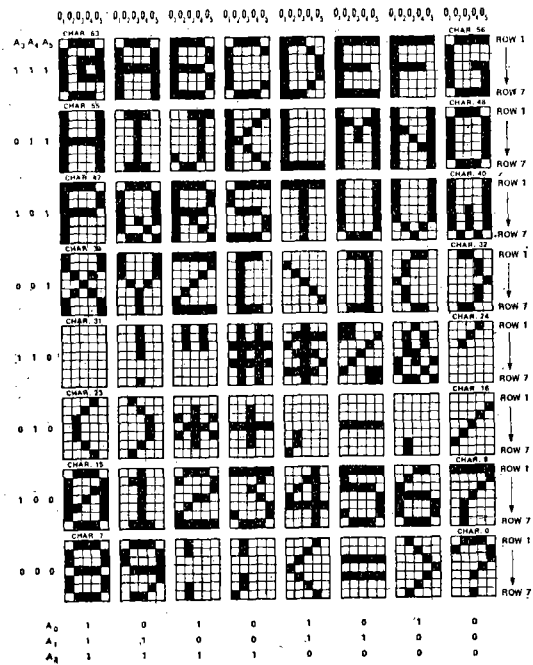
第1圖 문자 display 를 위한 CRT beam 의 주사 (Row output) screen 상의 각 점이 문자를 위한 미세주사의 중심이 된다.



第2圖 CRT Display character generator (column wise)

그밖에 上記 row output character generator 와 마찬가지로 spot 가 左右로는 段階的으로 走査 되면서 微細走査는 米字形으로 되는 star burst

ASCII CHARACTER FONT



第3圖 ASCII character font

character generator 에는 FDR 106Z1 (Philips) 가 있다.

第3圖에서 보는 바와 같이 display 된 文數字는 곱지 않기 때문에 TV 방송의 境遇는 5×7로 하지 않고, 10×16으로 하여 resolution 을 높여 사용하고 있다.

遇와는 달리 入力되는데로 出力시키면 읽기가 힘드므로 다시 別途 software 的인 方法을 써서 모아쓰게 되는데, 모아 쓰는데에 있어서 各 메이커의 方法이 相當히 다르다.

UNIVAC 은 上記 33字 以外에 垂直母音 延長用(終聲子音이 없을때 한글 文字가 너무 짧게 보이기 때문에 延長) 活字 2個와, 역시 終聲子音이 없을때 水平母音을 내리고 키우기 위한 ㄱ, ㄴ, ㄷ, ㄹ, ㅡ 等 9字를 더 採擇하여 42字素를 使用하고 있고²⁾ IBM 에서는 上記 32個字素(짧은 母音 ㅈ를 除外) 以外에 終聲子音이 없을때의 모든 單母音 10字素 以外에 ㅈ, ㅊ, ㅌ, ㅋ (但 終聲子音이 없을때) 4個, 역시 ㄱ, ㅅ, ㄷ, ㄹ, ㅈ, ㅊ, ㄱ, ㄴ 두벌(終聲子音이 있을 때와 없을때) 14個, 其他 終聲單子音 13個(ㅋ을 除外) 및 終聲 重子音(ㅍ ㄷ) 2個, 終聲複子音(ㄱ, ㅈ, ㅊ, ㅌ, ㄴ, ㄷ, ㅌ, ㄷ, ㅌ, ㅌ) 10個, 都合 85字素를 쓰고 있다³⁾. 이 境遇는 終聲으로 可能한 27個 子音 字素에서⁴⁾ ㄱ 및 ㄴ이 빠져 있으나, 이 두字素는 使用빈도가 0.001% 以下이기 때문에 큰 支障은 없을 것이다.

CDC 는 이미 電話料金 告知書를 통해서 우리가 불 機會를 많이 가진 것으로서 單母音 및 單子音 24字素와 初聲雙字音 5字素 및 垂直複母音 4字素로 形成된 33字를 使用한다.

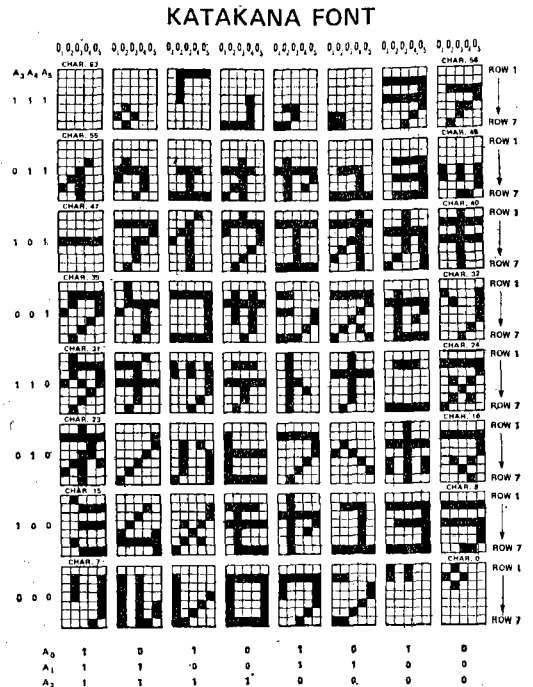
이 모든 境遇에 있어서 初聲子音이 中聲母音의 位置에 關係없이 (水平의 境遇나 垂直의 境遇) 같은 位置이기 때문에 우리가 一般印刷에서 보는 바와 같은 高운 形態가 되어 있지는 않다. 字素를 많이 使用한 IBM 은 越等히 高우나, 하나의 한글文字 全劃이 하나의 活字 크기로 印刷되기 때문에 print 된 結果文字가 他社에 比해서 醜劣하다.

7. 한글 모아쓰기의 問題

한글은 表音文字이기 때문에 母音과 子音이 獨立된 字素(ㄱ, ㄴ, ㄷ, ㅌ等)에 依해서 나타나고, 이들을 合成해서 모든 發音을 表現할 수 있게 되어 있으나, 中國의 影響을 받아 文字를 形成할 때에는 音節單位로 이들 字素를 모아서 漢文式으로 한 字가 한 音節이 되게 되어 있다. 漢文의 絶對的인 影響下에 있으면서도 한 音節을 다시 分解

해서 母音과 子音을 찾아 내고, 그러한 字素가 24個(너무 分解해버린 感이 있으나) 밖에 안된다는 것을 發見한 것은 대단한 일이었으나, 그것을 다시 音節別로 모아 쓰고, 또한 水平母音과 垂直母音을 두어 位置가 上下 또는 左右로 되게 한 것은 오늘날 機械化段階에서 어려움을 준 것이다.

OCR 研究의 一環으로서 文字로 절점과 枝路로 나누어 그 사이의 關聯을 matrix로 나타내는 方法이 提案되었는데⁸⁾, 이 境遇 枝路數와 절점(시작점, 끝점, 굴절점, 교점) 數가 그 字形의 複雜性을 나타내게 될 것이므로 한글의 境遇와 Alphabet 을 비교하면 다음과 같다.(第4表), 결국 比較的 單純한 形態를 가진 基本 24字素의 境遇로 限定해도, 한글 한 字素가 절점數로 보아서 Alphabet 의 1.23倍, 枝路數로 보아서는 1.22가 된다는 말이 되며, 이러한 한글字素가 2 內至 7個 모여서 한글文字가 되기 때문에 一般的으로 한글文字 한 字는 Alphabet 한 字의 3.6倍 以上 複雜하다는 말이 된다. 풀어쓰기를 할 境遇에도 이미 Alphabet 에 比해서 20%以上 複雜하다는 것이니 5×7의 文字發生器로서 한글字素(ㄱ, ㄴ等 單字



第4圖 カタカナ ROM

母의 경우)를 display 하면 Alphabet (第3圖)나 日本 가나가나 文字 (第4圖)의 境遇보다 그만큼 보기 좋하게 된다는 뜻이고, 5×7로 分割하는 것이 最小의 必要限界이기 때문에 擇한 것이라면, 20% 以上 더 複雜한 한글字素의 境遇는 몹시 보기 좋하여 訓練된 사람만이 便하게 읽게 될 可能性이 있다.

	단모음	단자음	중 자 음	복 모 음
CDC	10	14	ㄱ, ㄷ, ㅂ, ㅅ, ㅈ, ㅊ	ㄱ, ㅋ, ㆁ, ㆅ
UNIVAC	10	14	"	"
IBM	10	14	"	없 음

第3表 card puch 기계에 있는 자모

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	평균
절 점	5	3	2	2	6	5	5	6	2	2	5	3	5	4	1	3	3	4	2	4	2	3	5	5	4	7	3.77
지로수	5	3	1	2	5	4	4	5	1	1	4	2	4	3	1	3	3	3	1	3	1	2	4	4	3	6	3.00

	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㆅ	ㅌ	ㅍ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ	평균
절 점	3	3	4	6	4	6	4	1	5	7	5	6	8	5	4	6	4	6	4	6	4	6	2	2	4.63
지로수	2	2	3	5	4	6	3	1	4	5	4	4	8	3	3	5	3	5	3	5	3	5	1	1	3.66

第4表 한글과 Alphabet의 복잡성 비교

3×5로 한글 24字素를 分割하여 水平으로 또는 垂直으로 走査하는 方法이 提案되었으나⁵⁾, 著者의 말과 같이 ㅍ, ㅎ, ㅊ, ㅌ等 4字가 不自然스럽다⁶⁾.

前述한바 character generator는 64×7×5의 記憶素子를 가졌으니 한글 24字素를 收容시키기에 充分하나 別途 註文이 必要하다.

이 모든 것이 풀어쓰기 段階에서 이미 일어난 難點들인 것이다. (다음號에 계속)

참 고 문 헌

1. 성기수 "EDPS 한글화에 관한 연구" 1971년 1월
2. UNIVAC "한글 Data 처리에 대한 설명서" KD-7201

3. IBM KOREA Inc. "Programmer's Guide for using Hangeul" Form No. Hangeul 1-1
4. 박안기 "중앙전자계산소 1970년도 연구개발사업보고서" 1970년 12월
5. 이주근, 김영웅 "한글문자의 전자계산조직에 적용하기 위한 특징추출에 관한 연구(I)" 대한전자공학회지 제6권 제4호 1969년 12월
6. 이주근, 이광우 "한글문자의 인식에 관한 연구(II)" 대한전자공학회지 제7권 제3호 1970년 11월
7. 유시정 "한글입력의 패턴화에 대한 고찰" 성균관 대학교 경제개발대학원 학위논문 1973년 1월
8. 강린구, 이행세 "한글자체의 특징추출의 한 방식" 대한전자공학회지 제6권 제2호 1969년 9월