

電子計算機의 한글 入出力에 關한 研究現況과 한글 반 풀어쓰기 提案(2)

安 秀 桔

<서울工大 助教授>

8. 한글 모아쓰기를 爲한 매카니즘

한글 字素 24個를 가지고 第6節에서 말한 바 30가지 構成方式로 組合해 가면 可能한 文字의 數가 11,000字를 超過한다(유시정, 참고문헌 7, p. 36)고는 하지만 印刷所에서 흔히 쓰는 活字數는 2,000個 程度이고 公판인쇄기의 境遇도 2,000字를 넘지 못한다. 그 中 가장 複雜한 文字가 文字發生器의 最小의 크기를 決定지어줄 것이니 「ㄷ」字의 境遇를 例로 심는다면 枝路數가 24 枝路數가 18로서 Alphabet 中 가장 複雜한 E文字의 6枝路와 5枝路에 比해 거의 4倍임을 알 수 있다.

動員된 Read Only Memory 素子が 클수록 收容可能한 圖案의 複雜性은 더 急히 增加하기 때문에 한글 모아쓰기를 爲한 最大 記憶素子の 크기는 10×14 程度이면 足하리라고 推산된다.

하나의 $64 \times 7 \times 5$ 素子 크기의 文字發生器는 대강 16個의 모아쓰기 文字를 收容하게 된다는 計算이 되니 公판 程度의 印刷나 display 를 爲해서는 100個 程度의 文字發生器가 必要한 것이다

4個 程度의 上記 規模의 Read Only Memory 로서 모아쓰는 窮理가 없는 것은 아니나(안수길 참고문헌 14) 原모아쓰기의 Read Only Memory 의 必要經費는 文字發生器 單價가 \$10程度라면 全部해서 \$1,000이 必要되는 것으로서 이것만으로서 는 비싼듯 하지 않으나 함께 所要되는 全體 材料만 해도 特別히 이러한 많은 ROM에서 올바른 address의 文字무늬를 찾아 내가는 Demultiplexer 만 해도 거의 不可能에 가까운 複雜性을 띠게 되는 것이다. 客談이지만, 人間頭腦의 記憶容量의 구석구석에 숨은 情報을 적은 Energy

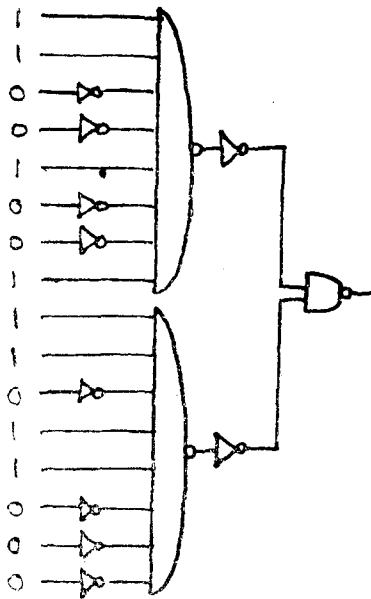
※ 技術士 (電氣部門)

로서 찾아내는 能力에 다시 感嘆하게 된다. 한 마디로 해서 한글 모아쓰기의 問題는 Alphabet 의 境遇에 比해서 數十倍 내지 數百倍 더 비싸게 먹힌다는 이야기이다. 이곳에서 數百倍로 잡는 것은 素子が 많아져 가는 것에 比해 같은 Screen 上에 의 모든 文字를 表示시키는 데에 必要한 回路의 複雜性이 自乘에 比例하는 것으로 친 까닭이다.

電算機 內部에서 各各의 文字를 나타내는 內部 code 에 關해서는 IBM 과 UNIVAC 에서는 8 bit 를 基本單位로 하고 있기 때문에 code 의 個數가 많아($2^8=256$) 한글字素에 對해서 別途로 code 가 주어져 있으나, CDC 에서는 內部 code 가 6 bit 로 되어 있어서 Alphabet 의 內部 code 를 빌려 쓰고 있다.

한글의 境遇는 入力은 풀어쓰기로 하고 있지만 soft ware 的인 方法으로 모아쓰기로 變換해서 모아쓴 한글文字에 對한 內部 code 를 주어야 한다. 그 具體的인 방안이 提案되었는데(문명호 참고문헌 13 p. 77) 收容할 수 있는 한글의 文字數는 4725이고 必要 bit 數는 16이다. 이 內部 code 를 判別하여 어느 文字인가를 찾아내는 (decoding) 過程에서 材料의 個數가 역시 상당히 많아서 한 code 에 대해서 8 Input Nand T Gate 가 2個, Inverter 가 平均 10個로서(第5圖) 1500字만 decode해도 8 Input Nand Gate가 3000個, Hex, Inverter 가 2500個, Quad 2 Input Nand Gate가 約 400個가 드는데, 이러한 TTL IC 한 個에 \$1.00로 計算해도 約 \$6,000의 材料費가 들게 되는데 이것은 아직 decoding 단계 다시 말해서 지금 電算機의 어느 한 address의 memory 에 어느 文字가 들어 있는가를 判定해

第五圖 coding



第五圖 16bit decoder 의 一例

1100100111011000라는 신호가 들어올때만
신호를 내어준다.

내는데 쓰이는 것만 해서 그와같이 든다는 것으로, 該當되는 文字發生器에서 文字圖形의 走査分解에서 나오는 binary 信號를 얻어 내어 CRT에 display 시키기까지는 아직 도 먼 얘기이다. 上記方法은 並列처리의 方法이고, 하나의 주어진 code에 대해서 1500個 文字에 該當되는 code를 交代交代로 incoming code와 比較시켜서 判別을 行하는 方法도 있겠으나 最小限 1500過程을 行하는 時間이 필요하기 때문에 System의 能率을 害치게 될 것이다.

한편 soft ware 的인 方法을 써서 內部 code와 incoming code와를 比較시키는 것도 있으나 이 역시 直列動作이기 때문에 machine time이 消費되고, 그 때문에 結果되는 能率의 低下는 大型電算機일수록 심하다. 그런데 한글문제를 다루기를 願하는 것은 大型機의 境遇가 많기도 하려니와 한글을 다룬다는 事實自驗가 大型機作用을 必要하게 하기 때문에 能率의 低下는 더욱 심각하게 된다.

가장 能率的인 soft ware 的인 解決方法은 display나 printing 만을 爲해 小型電算機를 따로 쓰는데 있다. 근래 小型機의 價格低下(\$

2,000~3,000)로 보아 이것이 最善이 될 可能性이 있다.

24個 字母를 收容한 ROM에서 얻어진 直列信號를 CRT에 보내되 그 文字中心(그곳을 中心으로 微細走査를 하는) 位置를 display 하고자 하는 文字에 따라, 다시 말해서 그것이 垂直母音을 가졌는지 또는 水平母音을 가졌는지 또는 받침이 있는지 없는지에 따라서 알맞게 移動시켜 주는 機能을 갖고 있는 “Variable combination circnit”를 使用하여 24個 字素만을 갖고 모아쓰기를 하는 案이 提示되었는데(이주근, 참고문헌 10, p. 43), 이 境遇에 있어서 이 “Variable combination circnit”은 display 하고자 하는 文字의 類型에 따라서 X 및 Y軸에 必要한 偏向을 주는 Analogue 電壓을 Digital 量으로부터 만들어 주는 D-A 變換器가 必要하고(모든 row output character display의 階段走査에서 이미 X軸 Y軸에 하나씩 必要한 것이지만) 이 X軸 및 Y軸方向의 2個의 微細偏向用 D-A 變換器의 clock pulse rate는 原走査用보다 훨씬 높아야 하면서 이 變換器들에 보내지는 位置code는 文字에 따라 달라지기 때문에, 前述한 바, decoding 過程에서도 이미 상당히 複雜한 機構와 高價의 費用이 들게 되는 것이다.

따라서 모아쓴 한글文字의 code 配當에는 初聲中聲終聲의 三部分을 code 길이를 역시 三分하여 各各 獨立的으로 配當하여 decoding은 全體 bit에 걸치지 않고 局部的으로 할 수 있게 해야 기구가 간단하게 될 것이다.

中心點을 옮기지 않는 範圍에서의 解決方案을 (안수길 loccit) 찾아 내었는데 不遠 發表할 豫定이다. 다시 line printer의 境遇를 考察하면 soft copy의 境遇와는 달라서 한번에 display 해 줄 必要는 없고, 인쇄를 二段階나 三段階로 해서 各段階 사이에 많은 시간이 經過되더라도 最終的으로 모아 찍혀 고운 結果란 보여 주면 되니까 문제가 더 쉬운 듯 하지만 Drum printer의 境遇나 Chain printer의 境遇나 font 素子の 數가 많아질 수록 인쇄速度가 늦어져서 문제가 된다. 특히 入出力때문에 시간을 많이 빼앗기고 있는 것이 實情이고 보면 그것이 다시 더 느려지는 것은 相當한 短點이라 하겠다. 그러나 하

나의 한글文字를 풀어 놓았을 때에는 하나의 字素를 찾아 가기는 比較的 빠르다 하더라도 二段階, 三段階로 적음으로 해서 特別히 멀리 떨어져 있는 字素사이를 왔다 갔다 할 때에는, 오히려 chain이 길어져도 한꺼번에 모아쓰는 경우가 比較的 덜 不利할 境遇도 있으리라는 것이 짐작되어서 複合字母를 많이 가진 IBM의 方式도 평가될 만 하다. 그러나 이미 말한 바와 같이 인쇄된 글자가 적고 받침이 두 가지 모자라는 點이 있으며, 그밖의 메이커—printer들은 글자가 크나 글자의 大小가 고르지 못하며 均衡이 안잡혀 곱지 못한 感이 있다. 이 境遇도 字素의 中心點을 옮겨 줄 수 있다면(母音字素가 一字의 떨어지지 말고 自己 左側에 있는 子音字素 바로 옆에 가서 찍히는 등) 좋지만 그러한 line printer는 전혀 開發되어 있지도 않고, 開發된다손치더라도 字母를 판정하며 中心點을 옮겨주는 기구가 複雜해서 單價가 올라갈 것이며, 또한 수요가 Alphet line printer에 比해서 너무 적기 때문에 高價가 되는 癖이 있을 것이다.

結局 한글 모아쓰기는 歐文을 使用하는 各國에서는 想像 못할 程度의 비싼 欲望이 되고마는 것이다.

2벌식 程度의 簡單한 한글 打字機字板으로 적으면서도 4벌식 程度의 高출력을 내어주고 그러면서도 速度가 빠른 打字機專用 超미니 컴퓨터를 쓰는 方式이 研究되었는데(朴憲緒 참고문헌 15) 이는 글자가 高운 4벌식은 여러벌의 鍵中 어느편을 찍어야 하는지 등 찍을 때 이미 判斷力을 動員해야 하는데 그 불편을 없애서 打字速度가 커지며 손쉬워지는 利點이 있으나 이는 한글打字機用으로서, 모아쓴 한글이 一字分 한꺼번에 찍히는 것일 수는 없고 電算機用 line printer를 위한 研究는 아니나 console printer teletype에는 쓸 수 있을 것이다.

9. 한글 풀어쓰기

一國의 言語에 관련되는 限 不可避한 事情에 直面하기 前에는 너무 變경을 시키지 말아야 하는 것이기 때문에, 그리고 電算機가 導入되고 活用된 時日이 짧기 때문에 跳躍의 發展을 한素地를 안은 채 한글의 障壁을 뚫지 못한 狀態

로 남아 있으나, 이 分野의 研究者들을 한결같이 풀어쓰기를 提唱하고 있고(성기수, 참고문헌 1, pp.62~68), (이주근, 참고문헌 5, p.10), (박안기, 참고문헌 4, p.122), 나머지 參考文獻에서는 이미 풀어쓰기를 前提로 하여 研究를 하고 있다.

한글을 풀어 썼을 때, 10個 中聲字素中 垂直母音 ㅏ, ㅑ, ㅓ, ㅕ, ㅗ를 풀어 써도(橫書의 境遇) 形態는 그대로 있고, 子音과 母音字素간 거리만이 모아쓰기의 境遇에 比해 더 떨어져 있는 것이기 때문에 比較的 읽기가 쉽지만, 水平母音 ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ는 풀어쓰기를 했을 때 더 읽기가 어려워지는 것을 볼 수 있다. 第六圖에서 垂直母音이 많은 (a)의 境遇와 水平母音이 많은 (b)의 境遇를 比較해 보면 前者가 比較的 읽기가 쉽다는 것을 볼 수 있다. 이것은 垂直母

第六圖

(a) 이레바나저기 사아터기타 이나디타

(b) 이-오우타-레 나-라바기타-지-조조기타

音은 初聲과의 距離가 떨어질 따름이고 方向에는 變動이 없으나 水平母音의 경우는 두 字素사이 距離도 멀어지지만 下部에 있던 母音이 옆으로 오는 데에서 더욱 읽기가 힘들게 되는 까닭이다. 垂直母音의 境遇는 比較的 보기가 容易한 편이므로 제대로 풀어쓰기를 하되 水平母音의 경우는 풀어쓰지 말고 初聲과 中聲까지만 모아쓰기로 한 것을 水平式 반풀어쓰기라고 부르기로 한다면 完全 풀어쓰기와 水平式 반풀어쓰기의 比較를 第七圖에서 해 볼 수가 있다. (a)가 完全풀어쓰기이고 (b)가 水平式풀어쓰기이다. 이때 반풀어쓰기가 完全풀어쓰기의 경우보다는 越等히 쉬운 것을 볼 수 있다.

한편 反對로 세로쓰기의 경우를 볼것 같으면 終聲이 역시 距離만 떨어지고 위치는 自然스러운 下位에 있기 때문에 越等히 有利하나 中聲의 位置는 도리어 垂直母音이 不利하고 水平母音이 有利하다. 따라서 이번에는 水平母音을 띄어쓰고 垂直母音을 붙여 쓰면 第八圖와 같이 된다. 但 水平型과 垂直型이 섞인 複母音의 境遇 初聲에 水平型을 붙여쓰면서도 垂直母音을 띄어쓰

sembler language 를 사용한 것과 마찬가지로 우리에게 가장 便하고 보기 좋다고는 하지만 모아쓰기를 固執하면 너무나 비싸게 먹이기 때문에 우선 너무 비싸지 않은 範圍에서 우리들의 감각도 尊重한다는 뜻으로 타협點을 찾아 반쯤 모아쓰기를 고려하여야 할 것이다. (完)

참 고 문 헌

1. 성기수 "EDPS 한글화에 관한 연구" 1971년 1월
2. UNIVAC "한글 Data 처리에 관한 설명서" KD-7201 1968년 12월
3. IBM KOREA Inc. "Programmer's Guide for using Hangeul" Form No. Hangeul 1-1
4. 박안기 "중앙전자계산소 1970년도 연구개발사업 보고서" 1970년 12월
5. 이주근, 김영용 "한글문자의 전자계산조직에 적용하기 위한 특징抽出에 관한 연구(I)" 대한전자공학회지 제 6권 제 4 호 1969년 12월
6. 이주근, 이광우 "한글문자의 인식에 관한 연구

- (II)" 제 7권 제 3 호 1970년 11월
7. 유시정 "한글입력의 패턴화에 대한 고찰" 성균관대학교 경제개발대학원 학위논문 1973년 1월
8. 강필구, 이형세 "한글字體의 특징抽出의 한 방식" 대한전자공학회지 제 6권 제 2 호 1969년 9월
9. 문교부 우리말에 쓰인 글자의 짚기 조사 1955년
10. 이주근, 이광하 "組合方式의 character generator 에 관한 연구" 1972년도 추계전자학술연구발표회 논문요약집 1972년 9월
11. 차성부 "한글문헌정보검색시스템구성연구" 성균관대학교 경제개발대학원 학위논문
12. 이주근 "한글문자의 인식에 관한 연구(IV)" 대한전자공학회지 제 9권 제 4 호 1972년 9월
13. 문명호 "한글의 기계처리에 관한 고찰" 성균관대학교 경제개발대학원 학위논문 1971년 12월
14. 안수길 "한글문자모아쓰기 display 의 한 방안" 미발표
15. 학원서 "초소형 「한글 컴퓨터」개발" 동아일보 과학란 1973년 3월 9일자 (完)

技 術 相 談 室 案 內

韓國技術士會는

農業, 水業, 林業, 電氣, 機械, 化工, 纖維, 金屬, 鑛業, 船舶, 航空機, 建設, 應用理學의 13個 部門 297名(1回~9回)의 技術士로 構成, 技術士法에 依據하여 設立된 政府의 認可團體입니다.

技術士란?

國家考試에 合格하여 認定을 받은 科學技術界의 專門인 知識과 應用能力을 가진 技術의 權威이며 農業技術에서부터 工場管理에 이르는 相談·指導等에 關與하고 있습니다.

本誌는 讀者 諸位와 좀더 가까운 벗이 되고자 하여 여러분의 「技術相談室」을 마련하였습니다.

讀者에서 平常 技術的인 點에 對해 簡單히 問議하실 것이 있으시면, 本 相談室을 利用하여 주시기 바랍니다.

到首원 相談文은 內容에 따라 專門分野의 技術士에게 依頼하여 誠意있는 答을 드리겠습니다.

◇ 相談要領 ◇

問 議 書 : 200字 原稿紙 3枚 程度

相談方法 : 問議書의 解答은 本人에게 郵送通知하고, 本誌에 掲載 可能한 것은 次刊號에 掲載함.

相 談 料 : 無料

보내실 곳 : 서울特別市 中區 明洞 2街 2-7 電話 (22) 8265 · 5866

韓國技術士會 事務局