



한글 코드 체계

국민대학교 소프트웨어학부

강 승 식

Topics

- 영문 코드 -- ASCII, EBCDIC
- 한글 코드
 - N 바이트 코드, 3 바이트 코드
 - 상용 조합형 (삼보 조합형)
 - KS 완성형 -- 한글, 한자, 2바이트 부호 등
 - CP949 -- MS 윈도
- 각 한글 코드 체계의 문제점은 무엇인가?

영문 코드

- EBCDIC : IBM

- 정보 교환을 위한 Extended BCD 코드

- ASCII 코드

- 영문자 52개
 - 대문자 0x41~0x5A, 소문자 0x61~0x7A
 - 숫자 10개
 - 0x30~0x39
 - 특수문자 34개
 - 제어문자 32개
 - 0x00~0x1F

한글 코드

- N 바이트, 3 바이트, 상용 조합형
- KS 완성형
 - KS C 5601-1987(ISO 2022-KR) → **KS X 1001**
 - DBCS: <0xA1~0xFE, 0xA1~0xFE>
 - 2,350 음절
 - KS 완성형 한자 4,888자
 - KS C 5657-1991 : 확장 한글 코드 8,822 음절
 - KS C 5601-1987에 누락된 한글-한자 코드 보완
 - **KS X 1002**:1991에서 누락된 기호-한글-한자 등이 확장됨
- **EUC-KR = KS 완성형 + ASCII**
- 유니코드 2.0 → KS X 1005-1:1995

N 바이트, 3 바이트 코드

- N 바이트 코드

- 자음 14+5+11자 : 영문 대문자, [, \,], ^ 로 mapping
- 모음 21자 : 영문 소문자로 mapping
- 한 음절에 대해 2~5 바이트까지 가변적
 - 예: ‘안녕’ → ^NWbDDjW^O

- 3 바이트 코드

- 복자음(예: ㄱㅅ, ㄴㅈ, ㄹㅎ), 복모음(예: ㅗㅜ, ㅜㅓ, ㅓㅓ)에 코드 부여
- 모든 음절을 3 바이트로 표기

- 한글 앞뒤에 SI, SO 문자로 영문과 구분

- SI(shift-in) : ctrl-O
- SO(shift-out) : ctrl-N

N 바이트 한글 코드표

| 자음 | | 모음 | |
|----|---|----|---|
| ㄱ | A | ㅏ | b |
| ㄲ | B | ㅑ | c |
| ㄴ | D | ㅓ | d |
| ㄷ | G | ㅕ | e |
| ㄸ | H | ㅗ | f |
| ㄹ | I | ㅛ | g |
| ㅁ | Q | ㅜ | j |
| ㅂ | R | ㅠ | k |
| ㅃ | S | ㅡ | l |
| ㅅ | U | ㅝ | r |
| ㅆ | V | ㅟ | s |
| ㅇ | W | ㅢ | w |
| ㅈ | X | ㅥ | z |
| ㅉ | Y | ㅦ | |
| ㅊ | Z | | |
| ㅋ | [| | |
| ㆁ | ₩ | | |
| ㅌ |] | | |
| ㅎ | ^ | | |

예) N 바이트 한글 코드

- ‘국어와 영어’

<SO>AsAWfWlb WjWWf<SI>

<SO>AsAWfWlb<SI> <SO>WjWWf<SI>

<SO>AsAWf<SI><SO>Wlb<SI> <SO>WjWWf<SI>

- ‘와’

Wlb 또는 Wm

- ‘앞’

WbDX 또는 WbE

상용 조합형 한글 코드

- 초성/중성/종성
 - 각각 5 bits 코드 부여
- 모든 음절을 2 바이트(1 + 15 bits)로 조합
 - MSB 1 bit : 한-영 구분
 - 5 bits : 초성
 - 5 bits : 중성
 - 5 bits : 종성
- 국제 표준을 따르지 않음

상용 조합형과 3바이트 한글코드표

| 십진수 | 이진수 | 초성 | 중성 | 종성 |
|-----|-------|-----|-----|-----|
| 0 | 00000 | | | |
| 1 | 00001 | 채움 | | 채움 |
| 2 | 00010 | ㄱ A | 채움 | ㄱ A |
| 3 | 00011 | ㄴ B | ㅏ b | ㄴ B |
| 4 | 00100 | ㄷ D | ㅑ c | ㄴ C |
| 5 | 00101 | ㄹ G | ㅓ d | ㄷ D |
| 6 | 00110 | ㄷ H | ㅕ e | ㅓ E |
| 7 | 00111 | ㄹ I | ㅗ f | ㅕ F |
| 8 | 01000 | ㅁ Q | | ㄹ G |
| 9 | 01001 | ㅂ R | | ㄹ I |
| 10 | 01010 | ㅅ S | ㅛ g | ㅂ J |
| 11 | 01011 | ㅇ U | ㅜ j | ㅅ K |
| 12 | 01100 | ㅈ V | ㅠ k | ㅈ L |
| 13 | 01101 | ㅊ W | ㅡ l | ㅈ M |
| 14 | 01110 | ㅊ X | ㅓ m | ㅈ N |
| 15 | 01111 | ㅊ Y | ㅓ n | ㅈ O |

상용 조합형과 3바이트 한글코드표

| 십진수 | 이진수 | 초성 | 중성 | 종성 |
|-----|-------|-----|-----|-----|
| 16 | 10000 | ㄷ Z | | ㄹ P |
| 17 | 10001 | ㅋ [| | ㅁ Q |
| 18 | 10010 | ㅌ \ | ㅅ o | |
| 19 | 10011 | ㅍ] | ㅈ b | ㅂ R |
| 20 | 10100 | ㅎ ^ | ㅊ s | ㅅ T |
| 21 | 10101 | | ㅊ t | ㅅ U |
| 22 | 10110 | | ㅊ u | ㅅ V |
| 23 | 10111 | | ㅊ v | ㅇ W |
| 24 | 11000 | | | ㅈ X |
| 25 | 11001 | | | ㄷ Z |
| 26 | 11010 | | ㅍ w | ㅋ [|
| 27 | 11011 | | ㅍ z | ㅌ \ |
| 28 | 11100 | | ㅍ { | ㅍ] |
| 29 | 11101 | | ㅍ | ㅎ ^ |
| 30 | 11110 | | | |
| 31 | 11111 | | | |

예) 상용 조합형 코드

- ‘가’
 - 1 00010 00011 00001 → 0x8861
- 중성 ‘ㅏ’
 - 1 00001 00011 00001 → 0x8461
- 초성 ‘ㄱ’
 - 1 00010 00010 00001 → 0x8841
- 종성 ‘ㄱ’
 - 1 00001 00010 00010 → 0x8642

KS 완성형 한글 코드

- 11,172자 중 2,350자에 대한 코드 부여
- ‘가’(0xB0A1) ~ ‘힉’(0xC8FE)
 - 상위 바이트 변이 : 0xB0 ~ 0xC8 (25가지)
 - 하위 바이트 변이 : 0xA1 ~ 0xFE (94가지)
- 자모 코드
 - 자음 30자 : 0xA4A1 ~ 0xA4BE
 - 모음 21자 : 0xA4BF ~ 0xA4D3
 - 옛 한글 자모 : 0xA4D5 ~ 0xA4FE
- 2 바이트 특수 문자
 - 상위 바이트 변이 : 0xA1 ~ 0xAC
 - 하위 바이트 변이 : 0xA1 ~ 0xFE (94가지)
- 정의되지 않은 문자 예 : 똥/뽕/웁/굶/슌/웁/굶/툼

KS 완성형 한자 코드

- 한자 4,888자
 - 상위 바이트 변이 : 0xCA ~ 0xFD (52가지)
 - 하위 바이트 변이 : 0xA1 ~ 0xFE (94가지)
- 음가가 2 이상인 한자에 대해 별도 코드 부여
 - 更 : ‘갱’(0xCBD6), ‘경’(0xCCDA)
 - 樂 : ‘락’(0xD5A5), ‘악’(0xE4C5), ‘요’(0xE8F9)
 - 한자 → 한글 --- 1:1 자동 변환 가능
 - 한글 → 한자 --- 자동 변환은 안됨

CP949: Microsoft

- $11,172 - 2,350 = 8,822$ 자에 대한 코드 정의
 - $[81 \sim A0][41 \sim 5A, 61 \sim 7A, 81 \sim FE]$: $32 \times (26 + 26 + 126)$
 - $[A1 \sim C5][41 \sim 5A, 61 \sim 7A, 81 \sim A0]$: $37 \times (26 + 26 + 32)$
 - C6 [41-52] : 18자

