# 유닉스상의 한글 처리

- 출력을 중심으로 -

최준호, *junker@jazz.snu.ac.kr*

1998년 2월 8일

# 순서

1. 유닉스상의 한글처리
   (a) 한글의 표현
   (b) 한글의 입력
2. 한글 출력
   (a) 콘솔에서의 출력
   (b) X윈도우에서의 출력
   (c) 프린터로 출력
   (d) 글꼴
3. 기타
4. 끝으로

# 유닉스상의 한글처리

## 국제화(I18N)

- InternationalizatioN
- 소프트웨어를 국가와 언어, 문화의 차이에 대응하도록
- ISO, ANSI-C, POSIX.1, POSIX.2, X/Open
- locale, Netscape

## 지역화(L10N)

- LocalizatioN
- 소프트웨어를 특정 국가, 언어, 문화에 맞도록 수정
- 해당 지역(언어) 사용자의 노력
- hanterm, hanemacs

# 유닉스상의 한글처리(2)

## locale

실행 시간에 어플리케이션이 결정하는 언어 환경. 언어, 지역, 코드셋의 지정을 포함.

- ANSI C: setlocale()함수에 의해 지정
- 환경변수 LANG과 LC_*
- 한국어는 보통 LANG=ko_KR.EUC

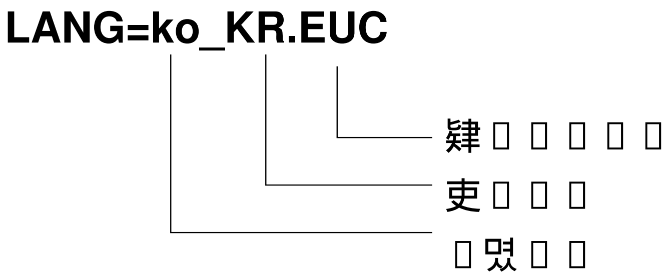


그림 1: LANG환경변수

# 한글 처리가 필요한 곳

- 한글의 표현(코드)
- 한글의 입력
- 한글의 출력
- 기존 프로그램과의 조화
- 인터넷(전자우편, 웹, 뉴스)

기타: 다국어 메시지 처리(e.g. GNU-NLS), 문서 처리 도구 등

공개 유닉스의 경우, 대부분의 문제를 사용자 스스로가 해결해야 함

# 한글의 표현

## 코드셋

- 한글을 컴퓨터 안에서 표현하는 방법
- 유닉스에서는 보통 완성형 코드(KSC5601, EUC-KR)을 사용
- ISO-2022-KR: 전자우편 교환에 사용. KSC5601의 또 다른 표현 방법
- CCS(Coded Character Set)과 CES(Character Encoding Scheme)의 차이

## 유니코드

- http://www.unicode.org
- ISO/IEC 10646-1의 부분집합. 16비트.
- 한글 조합형 11172자를 모두 포함
- Windows NT, Java등에서 지원

# 한글의 입력

- 텍스트 기반 터미널 - 출력을 위한 하드웨어적 배려
- PC의 콘솔 - han
- X윈도우
  - X11R5이후 국제화된 방법 제공
  - HanX - L10N hack!
    - 한글(euc-kr)코드가 나오는 경우 한글 글꼴으로 출력
    - 국제화 방식과 맞지 않음
    - Dynamic HanX?
  - 입력기(X Input Method) - KIMS, hanIM, 벼루
- hanterm, hanemacs - 조합형 지원, L10N

# 한글의 출력 - Intro

- 정해진 방법이나 표준의 부재
- 지원 프로그램에 따라서 방법이 다름
- 대부분 포스트스크립트로 출력

## PostScript

- Adobe Corp. 의 페이지 기술 언어
- stack-based, postfix
- Ghostscript: PostScript인터프리터

# 한글의 출력

## 터미널과 콘솔

- hscreen
- HAN - KON(Kanji ON console)의 한글판
- 한글 콘솔 드라이버

## X윈도우

- 화면 출력에 대해서는 잘 정의되어 있음
- Fontset: 국제화된 출력 처리(한글 포함) 지원
- 비트맵(bdf)과 윤곽선(Type1, Speedo) 글꼴 지원

# X윈도우의 한글 출력

**X11R6 - XOM(X Output Method)의 지원**

- X11R5의 Fontset의 확장
- XIM과 대칭적인 관계
- 어플리케이션 뿐 아니라 윈도우 매니저도 국제화/지역화되어야

**hanterm**

- 조합형,완성형 폰트 사용: 완성형, 조합형, UTF-8지원
- hanemacs에도 같은 출력방법 사용
- 가장 일반적인 한글 관련 프로그램

# 출력물

프린터로 출력을 보내는 경우, 일반적으로 사용하는 시스템이 없음

- 프로그램 각자의 해결방법

**X11**

- XPRINT(X11R6부터): 최근의 방법
- 화면 글꼴과 출력 글꼴의 불일치: 비트맵 글꼴 사용

# 포스트스크립트와 한글

- Type 1 글꼴의 한계 - 256자
- Type 0 Composite(8/8) - 2바이트 인코딩 가능

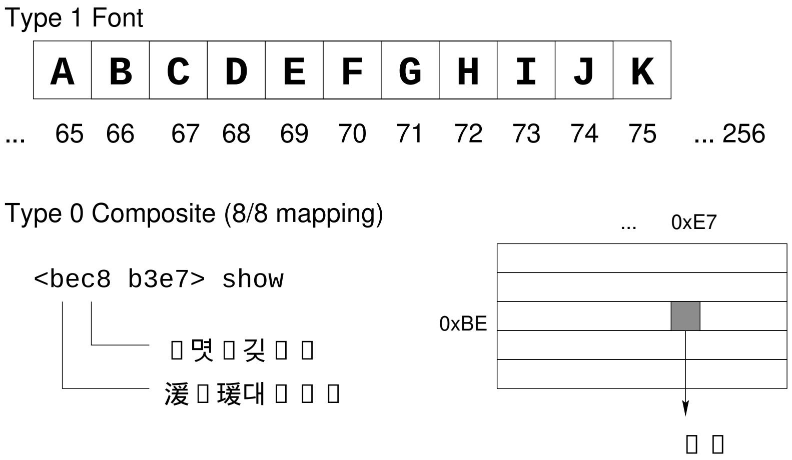


그림 2: Type 1과 Type 0글꼴

## 포스트스크립트 한글 글꼴

- 상용 버전의 경우 Type 0 Composite(8/8 mapping)
- 유닉스의 경우 프로그램에 따라 다를 수 있음
- hpscat: 조합형/Type 0
- nh2ps: HLaTeX/Type 0
- h2ps: ncode Type 1

Type 0가 가장 바람직한 형태이나 공개된 글꼴이 부족

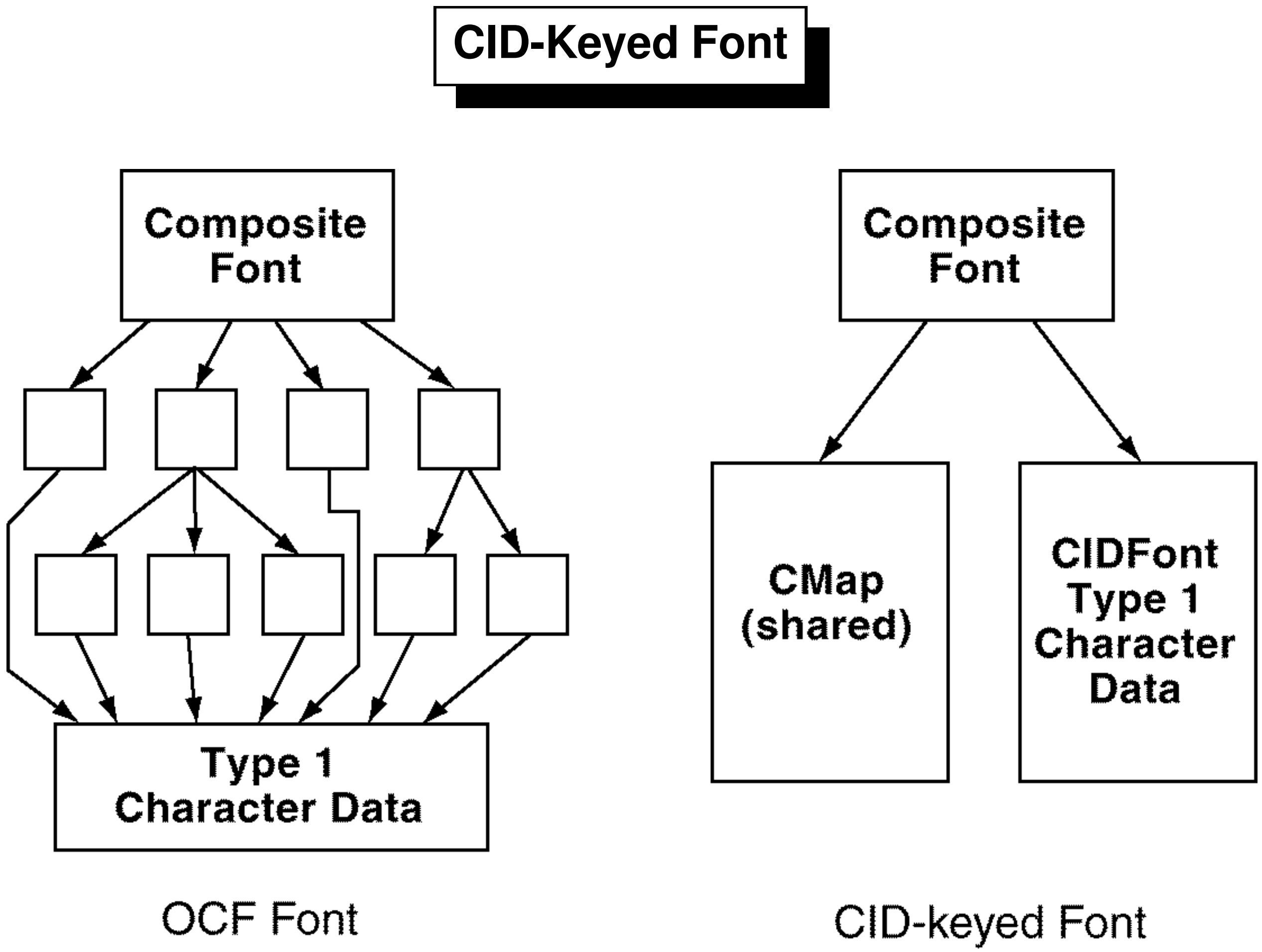


그림 3: OCF(Original Composite Font)와 CID-Keyed Font

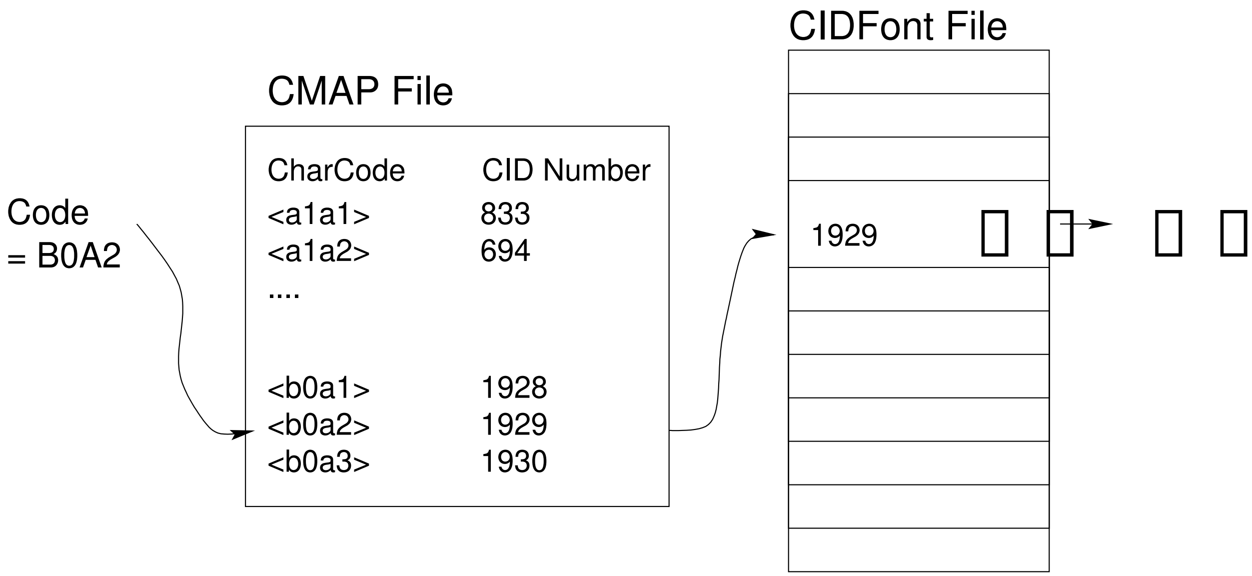


그림 4: CID-Keyed Font의 구조

- CMap: EUC-KR, KSC5601-1992(조합), ISO-2022-KR, 유니코드, UTF-8, ...
- 글꼴 제작자는 CIDFont만을 작성, CMap은 Adobe가 제공
- Munhwa글꼴이 공개되어 있음 - Ghostscript에서 사용 가능
- EUC-KR CMap을 사용하는 경우에는 Type 0 OCF같이 사용할 수 있음
- 글꼴 지정: *글꼴이름-인코딩* 예) Batang-Medium-KSC-EUC-H

# TeX의 한글처리

모두 기본적으로 METAFONT를 사용하여 한글 출력 - 256자 제한

## hTeXp/hLaTeXp

- KAIST수학과
- 최초의 한글 TeX에서 발전
- 글꼴의 저작권 문제 - pk글꼴만 배포
- 한글 오류 메시지 출력

## HLaTeX

- 은광희씨가 개발
- LaTeX의 한글화
- 완전한 Free 한글 LaTeX패키지(글꼴, 시스템)

# 텍스트 출력 프로그램

**hpscat**

- 조합형 Type 3글꼴과 Type 0 OCF사용 가능
- 조합형 글꼴을 사용하는 경우 글꼴을 미리 읽어들여야

**h2ps**

- ncode Type 1글꼴 사용
- 크기가 작고 기존 PS프린터에서 출력 가능
- 가변폭 글꼴

**nh2ps**

- Type 0 OCF와 HLaTeX PS글꼴 사용 가능
- 글꼴은 반드시 외부에서 읽어들여야 함 - ghostscript

## ncode의 조합 글꼴

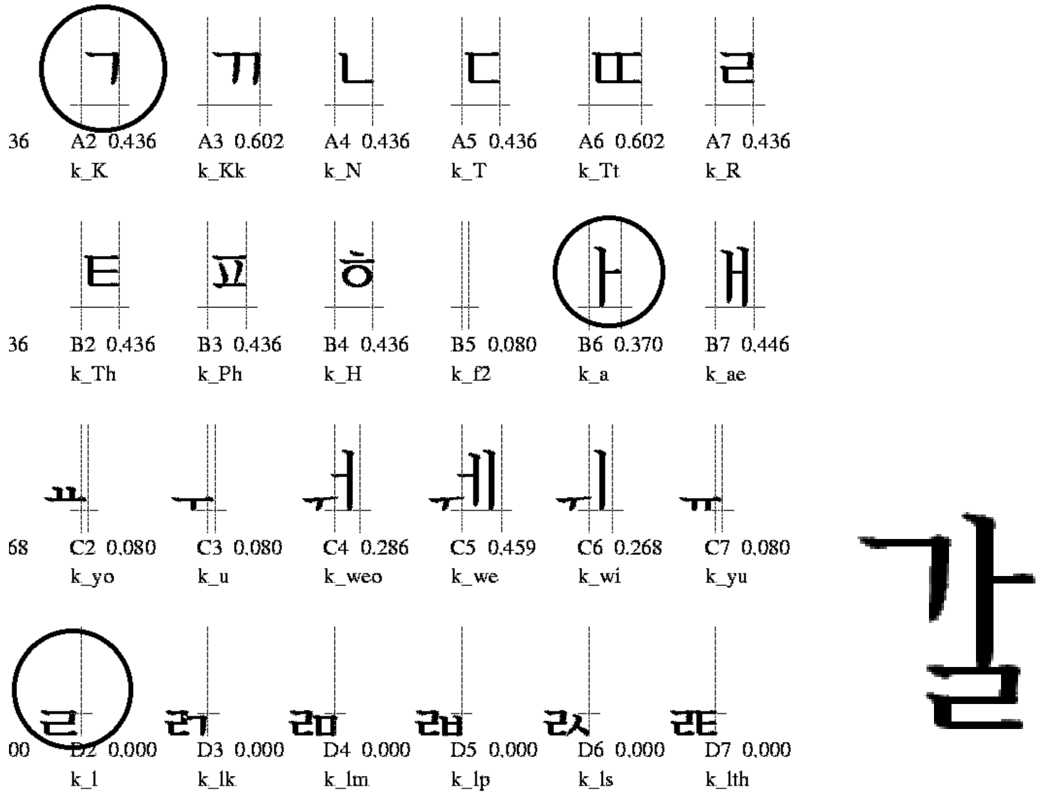


그림 5: n3f글꼴과 ncode표(일부)에 의한 조합과정

# HLaTeX의 완성형 Type 1 글꼴

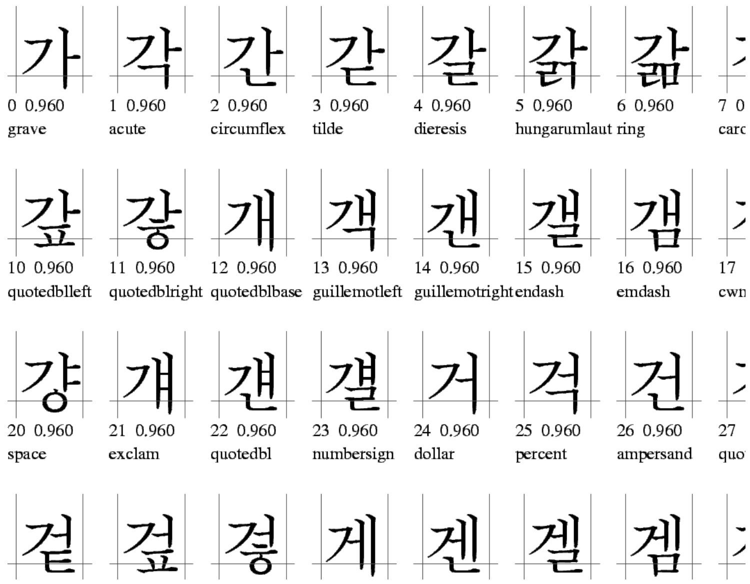


그림 6: HangulMyoungjoA-Regular글꼴

- 256개씩 10개 글꼴이 한 벌(한글 2350자)
- 완성형(한글, 전각기호, 한자) 글꼴 모두 지원
- 특별한 조합 방법 필요(KSC5601순서가 아님)

# Netscape의 WWW문서 출력

한글 인코딩 상태에서 한글을 출력하면 가능

- “Print”와 “Save as... Postscript”의 출력방식이 다름
- 3.0 이후로 변하지 않았음

**“Print”**

- 한글과 영문 글꼴을 구분(Type 0 OCF/Type 1 iso encoding)
- 한글의 글꼴 속성을 잃어버림(기울임이나 굵은글씨 등)
- 한글글꼴 내장의 PS프린터에서 직접 출력가능

**“Save as... PostScript”**

- 한글과 영문 글꼴을 구분하지 않음(모두 영문으로 생각)
- 필터링이 불가피하지만 글꼴의 속성 보존 가능

# Netscape출력 보조 프로그램

사용 가능한 한글 글꼴을 이용하여 출력하도록

**nhpf**

- Print와 Save as두 경우 모두 사용 가능
- n3f글꼴 사용 - 한글만 출력 가능

**nhppf**

- Print와 Save as두 경우 모두 사용 가능
- HLaTeX의 한글 글꼴 사용
- KSC5601의 모든 글자 사용 가능

웹브라우저에서

<b> □ □湲</b> □ □由 □ □듬 □ □

□ □湲 □ □ □ □ □由 □ □듬 □ □

Save as의 경우

```
0 708.7 moveto 12  f1
(\307\321\261\333\300\272) show
45.4 708.7 moveto 12   f0
( \276\306\270\247\264\344\264\331) show
```

f1 → 援듬 □ □湲 □ □ □

f0 → 蹂댁 □ □

Print..의 경우

```
0 749.7 moveto 12  of
<c7d1b1dbc0ba> show
36 749.7 moveto 12 f0
( ) show
12  of
<bec6b8a7b4e4b4d9> show
```

of → □ □湲 □湲 □瑗 □

# Ghostscript와 한글

- Type 0 OCF글꼴의 부족
- CID-Keyed Font? - gs5부터 지원
- gs-hfftype
  - FreeType Project의 공개 트루타입 래스터라이저 라이브러리 사용
  - gs-kfftype(일본어)에서 출발
  - 트루타입 글꼴을 사용하여 Type 0 OCF(euc-kr인코딩)을 에뮬레이션
  - 초기 단계 - 다른 인코딩 지원 필요(조합, 유니코드 등)
- 한글 PDF(Portable Document Format)은 아직 정립이 안되어 있음

# 글꼴

- 공개된 한글 글꼴의 부족 - 출력의 어려움
- X11 - 대우 mincho, gothic, 한양글꼴, 웹바탕 등
- 윤곽선 - ked글꼴(hpscat), HLaTeX, 일부 트루타입 글꼴
- 크기와 모양이 다양하지 못함 - 대부분 명조 정도에 그치고 있음
- 한자나 KSC5601전각기호, 조합형 글꼴의 부족
- 다양한 모양 부족(italic, bold, oblique...) - 에뮬레이션

**해결책**

- 공개 글꼴의 증가(자작 또는 기부)
- 다른 글꼴로부터 변환하여 사용 - 변환 도구 필요

# 공개 글꼴의 제작

## X11글꼴

- bdf의 경우 편집기로 작성 가능(xfed, xmbdfed 등)
- SERI 한글 글꼴 - HLaTeX PS글꼴을 변환하여 사용

## 윤곽선 글꼴

- Type 1은 한글에 적합하지 않음
- Type 0를 만들려면 특별한 도구나 PS에 대한 지식 필요
- HLaTeX PS글꼴 - 문화부 글꼴의 변환, 자작 등

## 글꼴 변환 도구

- FreeType Project - 트루타입 글꼴의 다양한 사용 가능
- ttf2bdf-k - ttf글꼴을 bdf글꼴로 변환 가능
- 윈95와 똑같게? - 힌팅

## 출력물

- Type 0 OCF, CID-Keyed Font, HLaTeX, n3f, ked, 트루타입 글꼴 등
- 각자 사용방법이 틀림 - 출력 어플리케이션이 처리해야 함
- gs-hfftype은 편리하게 트루타입 글꼴을 사용할 수 있도록 해줌
- CID-Keyed Font의 사용이 바람직함

## X11 글꼴 래스터라이저 라이브러리 - libfont

- 래스터라이저를 libfont(X11의 글꼴 처리기 라이브러리)에 추가 가능
- 트루타입 글꼴 - Solaris 2.6, FreeType Project등
- Accelerated-X의 경우 동적으로 글꼴 처리기 로드 가능
- libfont는 X서버와 xfs(글꼴 서버)에 동일하게 사용
- 글꼴의 제공 및 재배치 가능

## UNIX에서 다른 OS의 글꼴 사용

- 글꼴에 대한 라이센스 문서를 잘 읽어야 함
- 재배포 불가 여부, 다른 OS에서 사용 여부, 출력 가능 여부 등
- 트루타입 - Bitstream Cyberbit가 적당
- 웹바탕, 변환된 글꼴 등은 제한된 라이센스 내에 있음

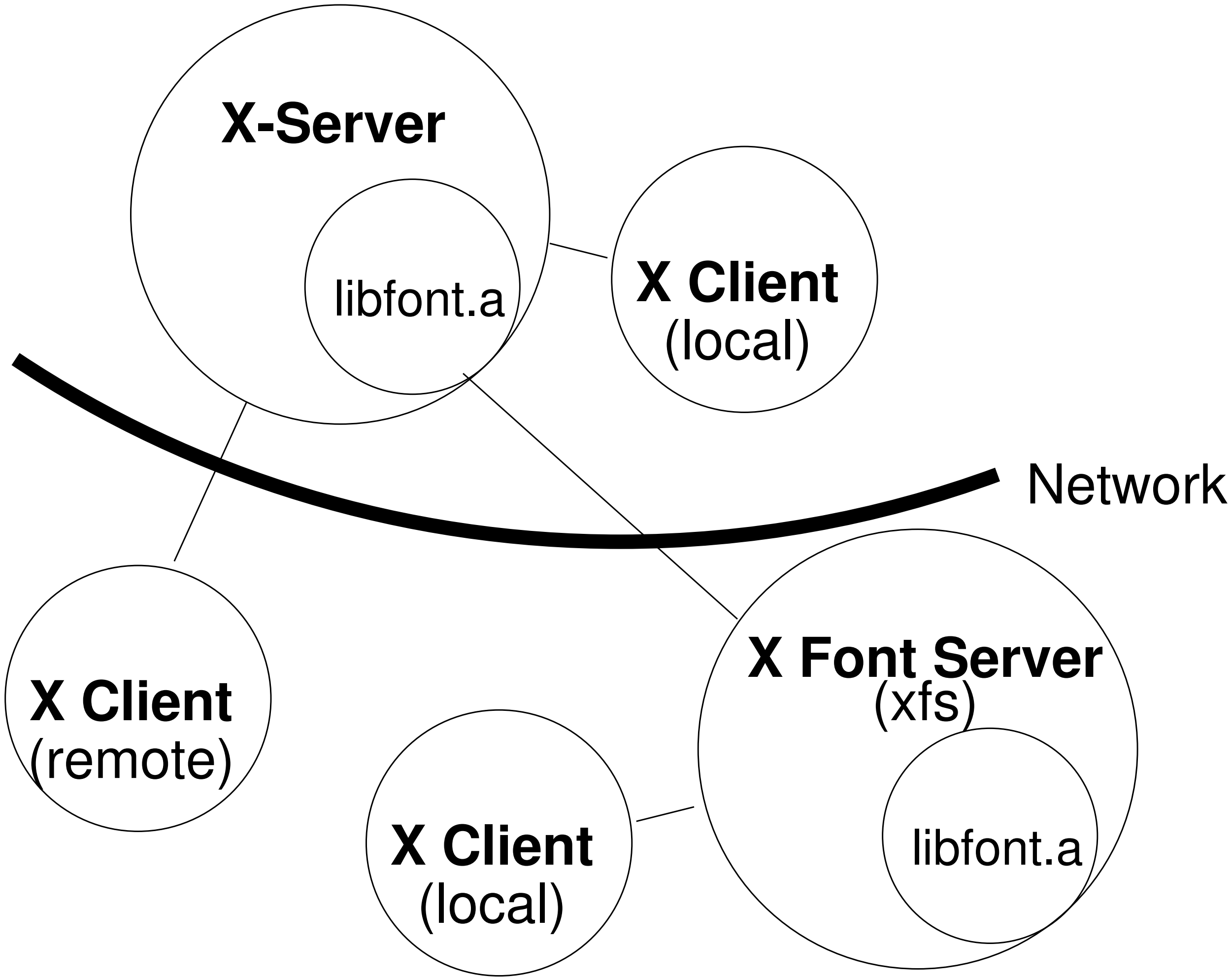


그림 7: libfont.a와 글꼴 서버

# 기타

- 인터넷 - 전자우편, 웹, 뉴스 등에서의 한글 인코딩
- 편집기 - helvis, nvi-m17n, emacs 20, pico...
- 문서의 구조화 - SGML-Tools(구 Linuxdoc-SGML), KLDP의 문서 포맷(?)
- 쉘 - bash, tcsh의 한글 입력/메시지 문제
- GNU-NLS - 프로그램 메시지의 국제화(한국어 팀 있음)
- 상용/소스 비공개(?) 소프트웨어 - HWP/X, KIMS, hanIM, ...
- 잘 알려지지 않은 수많은 한글 도구들...

# 끝으로

- Viva Free Unix(Linux, FreeBSD, OpenBSD, NetBSD)!
- 자발적 참여 없이는 이루어지지 않음 - 능동적으로 행동하자
- 신토불이 vs 국제화 제품
- R.T.F.M!
- 진실과 대담(?)