LANGBOX-ARA 3.3

The Bilingual English/Arabic TTY Environment

of UNIX

Technical Product Description

April 1998

LangBox International
Immeuble Space, Bat. B - 208 / 212 Route de Grenoble - 06200 Nice - France
Tel: +33 (0) 4 93 71 14 10 - Fax: +33 (0) 4 93 71 15 6 - http://www.langbox.com

1. Introduction

1.1 Background

Arabic, as a calligraphic language, presents major processing problems.

An Arabic character might take one, two, or sometimes four different shapes, yet it is represented by one code. The shape of the character is determined depending on its position in the word. This is but one problem called "**Character Shaping**".

Another problem is the direction of writing. Arabic text is written from right to left. This conflicts with Latin, which is written in the opposite direction. When mixing text languages, characters are **added** in one language and **pushed** in the other. This is called "**Bi-Directionality**".

Some users speak only Arabic. They will not accept a cursor positioned at the leftmost position of the screen. They want to have an option allowing them to start at the rightmost position of the line, i.e., in brief, a mirror image of the screen. The implication is that, in this mode, Latin characters are pushed from right to left. This is called "**Screen Mirroring**"

Yet one more complication is vocalization (or **diacritics**). These characters, like their counterparts in Latin, the vowels, are a linguistic necessity, even if they are not used in most of basic publications. Yet, in Arabic, they appear on top or below their respective consonants.
These linguistic complications -- and more -- make Arabic a difficult language to handle.

Another aspect of the problem is that standard UNIX Operating system support of Internationalization (local) supplies only support for European Languages generally based on the ISO 8859-1 character codeset. This National Language Support (NLS or XLOCALE) allows keyboard mapping, handling of collating sequences and character types, date and time format. It could also include, in the near future, a specific extension for the Japanese language, but cannot handle Arabic characters, and, in general "right-to-left" writing direction languages.

LangBox International is specialized in the design and the development of bilingual and multilingual Operating Systems. It has implemented bilingual capabilities on a large number of machines operating under UNIX, XENIX, AIX, SUN/OS, ISC, SINIX, DG/UX, EP/IX, CLIX,.SOLARIS, IRIX, OSF/1, etc…

In response to a clear market demand in the Arab countries, LangBox International has developed a Bilingual System supporting both the Arabic and English languages. **LANGBOX for Arabic (named LANGBOX-ARA)** has thus been conceived with the following principles in mind:

- Providing a global solution to Arabic and English simultaneously.
- Transparency of applications.
- Total transparency to storage and display of data in national languages.
- Total UNIX application transparency for usage in Arabic.
- Ease of internationalization of applications.
- Conformity with national and international standards.

LANGBOX-ARA is a bilingual Operating System and a bilingual development environment.
LANGBOX-ARA is based on the ISO 8859-6 codeset. The languages supported are:

- English
- Arabic

1.2 The LANGBOX-ARA Working Environment

The **LANGBOX-ARA** system is built around **TTY** driver of the UNIX System V. When loaded onto the UNIX system, **LANGBOX-ARA** provides the user with a full bilingual environment in Operating System interface, systems development and character based applications runtime operations.

The **LANGBOX-ARA** system supports any application running on TTY terminal screen, from the dump terminal (such as VT100 RS-232 connected terminal) to any PTY connection through the Network (Xterm emulation or PC based product running telnet sessions).
To conform to the UNIX operating system, **LANGBOX-ARA** is also designed to run in a multi-user and multi-tasking environment and to co-reside with the standard UNIX facilities. Having **LANGBOX-ARA** added to a UNIX system does not prohibit its users from operating in a pure UNIX environment.

While operating under the **LANGBOX-ARA** bilingual environment, users can select and set their default language, English or Arabic. Users can login to the system in the language of their choice and communicate with the host using the standard UNIX commands and utilities. The commands could be entered either in Latin or in Arabic and are executed by the **LANGBOX-ARA** shell command interpreter. The system responses are displayed in the language chosen by the user.

Although set with a default language prior to login, a **LANGBOX-ARA** user can start multiple work sessions, (shell child processes), each with a different base language. He will be able to alternate languages within the same work session, at the command line, directory and file level, or when running a standard UNIX application (ex: text processing, spreadsheets, data base management, etc.).

In addition to the bilingual UNIX user interface, **LANGBOX-ARA** provides a comprehensive bilingual software development environment for programmers. The programmer will be able, with or without the knowledge of the Arabic language, to develop with ease bilingual applications or to adapt, with minimal effort, current English software packages to run in a bilingual mode under the **LANGBOX-ARA** environment.

1.3 Internationalization of Application

LANGBOX-ARA provides an enhanced environment for the internationalization of applications as compared with classical techniques.
Usually, application programmers incorporate international character strings' manipulation, "context analysis" (automatic shape determination), and display processing within the application. Under **LANGBOX-ARA**, 8 bit clean character based application will run with no modification.
The following table lists the benefits obtained by comparing a full LANGBOX-ARA bilingual work environment with a standard software internationalization approach:

LANGBOX-ARA BILINGUAL ENVIRONMENT	SOFTWARE INTERNATIONALIZATION
Operating system user interface.	Not available.
Language & I/O processing handled by **LANGBOX-ARA** with no application overhead.	Language & I/O processing implemented in every application program, bigger overhead.
English only software developer able to provide bilingual products.	Sophisticated developers required with knowledge of the Arabic language particularities.
Bilingual software environment is uniform.	Potential inconsistency in software internationalization.
Bilingual UNIX Mail easily implemented under LANGBOX-ARA.	Major effort required.
Communication and networking easily implemented under LANGBOX-ARA.	Major effort required.

2. Product Description

LANGBOX-ARA supports exclusively the TTY interface, used under shell sessions under Dump terminals, telnet or rlogin connection to a UNIX operating system. It is composed of two packages:

- The Runtime System
- The Development System

The Runtime System must be installed prior to the installation of the Development System package. These packages include the standard UNIX V modules to which a set of **LANGBOX-ARA-**specific facilities is added.

2.1 The Runtime System

2.1.1 The Shells

LANGBOX-ARA includes two national language shells. They are differentiated by the first characters prefixing their label. Thus

- **ash** and **aksh** are the Arabic Bourne shell and Korn shell.
- **esh** and **eksh** are the English Bourne shell and Korn shell.
- **Araexec** allows to launch any Operating shell.

Each is a special shell version created to provide.

- An interpretation of the bilingual commands.
- A full bilingual environment management.
- Interface facilities with the kernel.
- Messages in the national language.

When the **LANGBOX-ARA** shell is invoked

- The environment variable **LANG** is assigned to the national language value (ex: “ar”) and connects all commands to the messages in the respective language.
- The keyboard of the logged-in terminal is reconfigured to the appropriate national setting. The terminal is bilingual. With a single keystroke, the keyboard is back to the original configuration. A set of keyboard stickers is supplied to help differentiate the keys in the dual keyboard.
- The shell prompt indicates the base language of the invoked shell.
- The cursor is positioned at the rightmost position on the screen. Arabic characters are added, English characters are pushed on.

Each shell can be called from another language shell and achieves the same effect. Exiting from a shell is done via the traditional "control-d." The national language variables are created with a new shell and restored when the spawned sub-shell is exited.
The main advantages of the **LANGBOX-ARA** shells are:

- The user decides and selects his/her base language.
- Each shell provides a bilingual work session in a bilingual environment.
- Command messages (as well as those of the shell) are in the language of the session.

2.1.2 Commands and utilities

There are two groups of commands and utilities supplied with LANGBOX-ARA.

2.1.2.1 The UNIX-like Group:

This is a set of executable modules that have the same calling sequence as their UNIX counter parts, except that they work in 8-bits and are bilingual. These commands are differentiated by their prefix **"a"**. Example: aa2ps, avi, auemacs, alp...
They also share a set of characteristics :

- They have the same name irrespective of the session language (**avi** works in English under **esh**, in Arabic under **ash**).
- The messages of these commands are in the selected national shell language.

Many other UNIX commands can run under **LANGBOX-ARA** without modification.

2.1.2.2 The LANGBOX-ARA specific group:

This is a set of commands and utilities supplied with **LANGBOX-ARA** and aimed at servicing the bilingual community of users. They include:
The **LANGBOX-ARA** specific commands are the following :

- Character management

acharset	display the character set on the terminal
aload	download Arabic character set on terminal
aloadp	download Arabic character set on printer
asmo449	convert a file to ASMO 449 codeset

- Display processing and shape management

afps	Arabic floating point symbol definition
amask	video line language attribute
amode	manage the space character
arabic	select the Arabic screen display mode
asetup	general LANGBOX-ARA function setup
astatus	display the status of the Arabic environment
context	enable context analysis
dataproc	force numeral to 7-bits for computation
english	select the english screen display mode
hindi	control the display of Arabic numerals
months	select months names
neutral	set neutral characters in the display processing
nocontext	disable context analysis
nohindi	control the display of Arabic numerals
notashkil	disable the diacritics management
tm	display the LANGBOX-ARA driver level release
tashkil	enable the diacritics management
wordproc	record numerals in their language

- Terminals management

atic	Arabic terminal information compiler
toggle	define the toggle key for the bilingual keyboard
aresize	adjust the Arabic screen after a resizing action

2.2 The Development System

2.2.1 The LANGBOX-ARA commands:

In addition to the standard development power of UNIX, **LANGBOX-ARA** provides a set of tools to develop bilingual applications. They include string management functions reconfigured to service the 8-bit character sets, messages extraction and handing tools, sorts and conversion functions, etc.

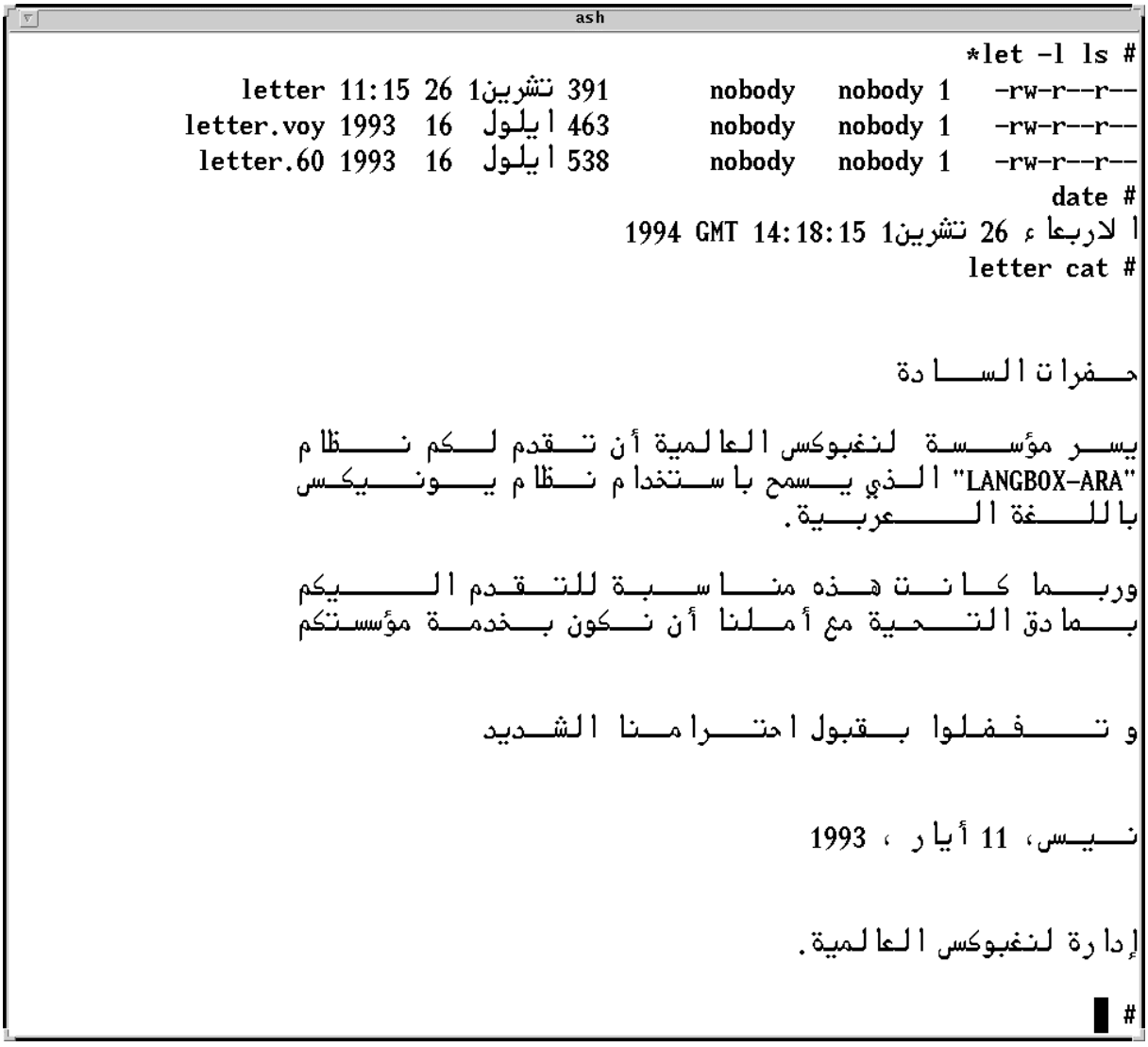
The message extraction capabilities provided permit the separation of strings out of "C" program into files, and a formatter for these messages to simplify the subsequent translations.
These tools are extremely powerful when "internationalization" of applications are envisaged.

2.2.2 The LANGBOX-ARA function libraries

These libraries include:

- Character types.
- Character transformation.
- Date and calendar routines.
- Regular expressions.
- Sort.
- Internationalization functions.

Bilingual applications can benefit by recompiling under **LANGBOX-ARA** as these libraries have the same calling sequence as their UNIX counterparts, thus generating object code capable of handling bilingual messages.



Sample screen output of a Sun cmdtool window, running ash shell

3. Standards

LANGBOX-ARA is adapted to the standards as set forth by AT&T's SVID and ISO.

The standards adopted under **LANGBOX-ARA** are related to the following:

- Character sets.
- Standard display conventions.
- Arabic level support standards.

3.1 The Character Sets

The Arabic characters are 8-bits wide and conform to the following standard:

- ISO 8859-6 (ASMO 708)
- ASMO 449+

3.2 Display Conventions

Characters will be displayed according to their language specific conventions. Latin characters will always appear separately, while Arabic characters will be contexted and displayed in their composite form.
The display technique adopted depends on the base and current languages chosen by the user.

When the environment is Latin, the initial cursor position is at the left of the line and characters are added to the right as they are entered. Arabic characters are inserted, and pushed to the right as they are entered.

When the environment is Arabic, the reverse phenomenon occurs: the initial cursor position is at the right most position of the line, Arabic characters are added and English are inserted.
Vocalization is fully supported on terminals capable of displaying 256 downloadable characters.

3.3 Neutral Characters

These characters are not context sensitive and do not, therefore, affect the shape of Arabic characters. Yet they have, in certain instances, opposite meaning in Arabic, due to the direction in which this language is written. Typical examples are: ( ) { } < > [ ] etc…
A special command handles the meaning the user wants to assign to these special characters.

3.4 Numerals

The user is provided with a command allowing the display of numerals in Arabic shape (Latin shape) or Hindi shape (used in the Middle East). All known complications associated with this subject have been solved and incorporated in the package.

3.5 Automatic Shape Determination

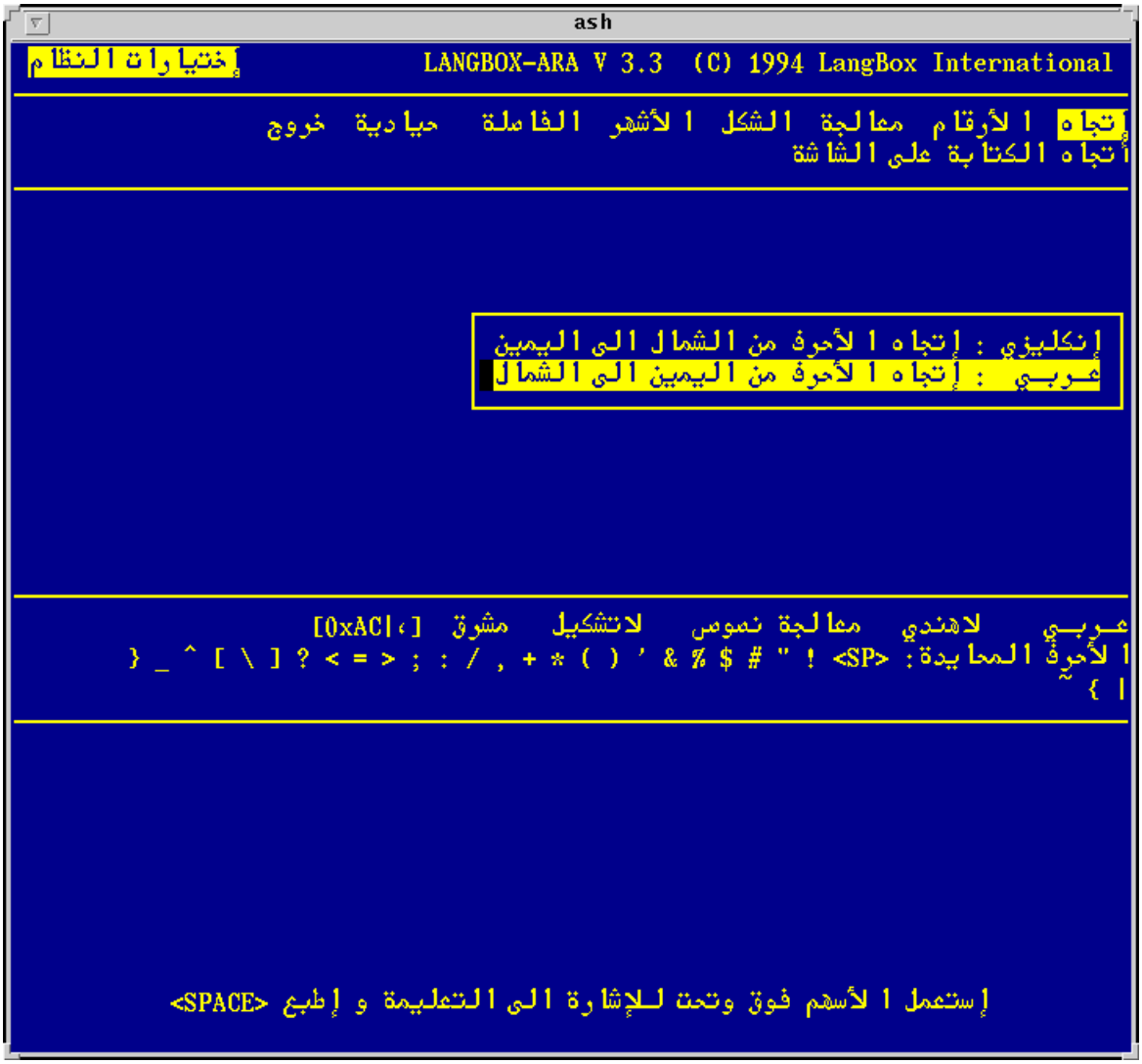
The standard rule is to display characters in the way calligraphy requires it. It was, however, found that this feature should be optional, since system users are frequently in debug work sessions and prefer to have their characters displayed in their original, base form. A pair of commands has been included to inhibit or restore Automatic Shape Determination.

3.6 LANGBOX-ARA Level Support Standards

LANGBOX-ARA capabilities will not change when installed on different hardware system configurations. The following standards have been adopted, and will be supported even if the hardware does not have the required features:

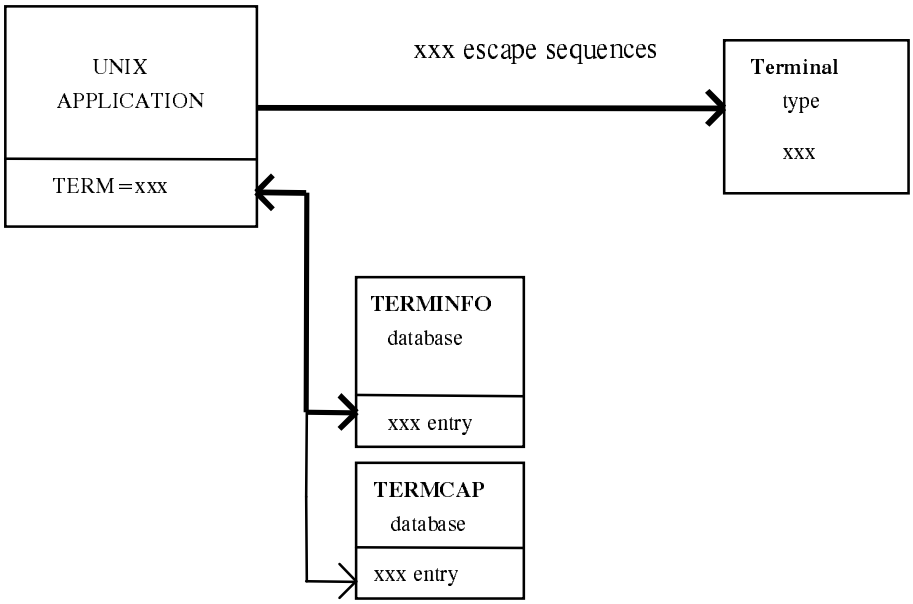
- Full **VT220** Terminal emulation support : The TERM environment variable is swtiched autoamtically and the application addresses a full VT220 pseudo terminal environment.
- Keyboards, display and print devices will support the full ISO 8859-6 character set.
- All I/O device drivers support diacritics (vocalization characters and some Arabic special characters). Display or print devices unable to output these characters are recognized by LANGBOX-ARA and receive only the character set they can handle.
- Keyboards capable of returning diacritic codes but attached to display or print devices that do not have these features are supported, even if these characters are not echoed. The diacritic codes are accepted into the input stream for further processing.
- Context Analysis is supported on all output devices.

It is further stressed here that **LANGBOX-ARA** generates and supports a Bilingual environment. **LANGBOX-ARA** allows users to set a SESSION LANGUAGE and to intersperse English and Arabic any time in the session, even within one command line.

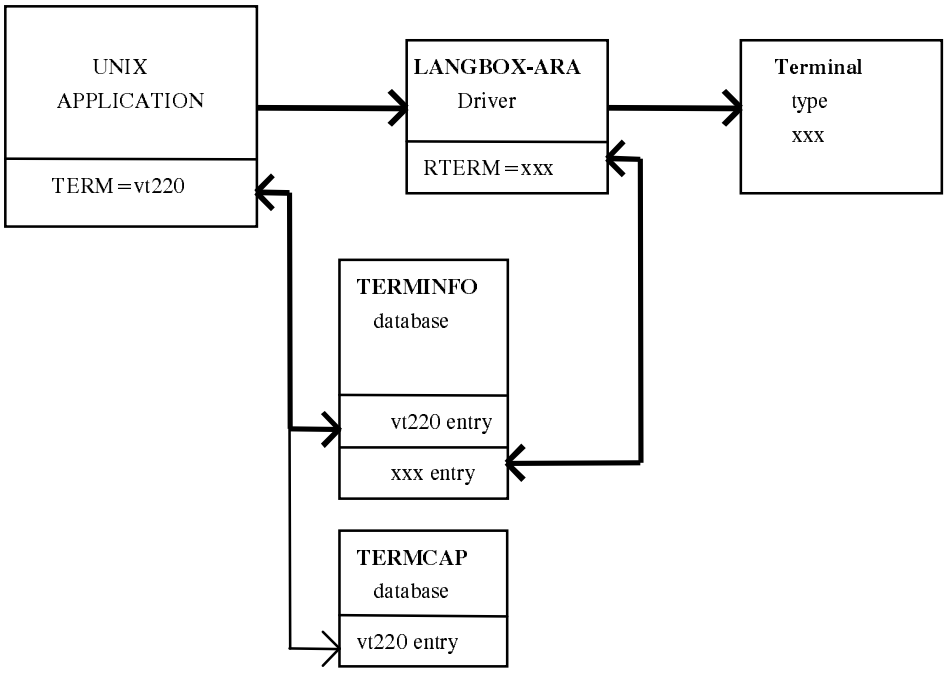


Sample screen copy of a Curses(3X) based application ruuning under LANGBOX-ARA

Standard Application Process



LANGBOX-ARA Application process



This diagram shows how the LANGBOX-ARA kernel driver take controls of all TTY line I/O within the UNIX kernel and is able to perform exception process from the Application flow of character to the physical real terminal screen.

In the same way, keyboard codes sent by the terminals are mapped into a dual logical keyboard management system and are sent back to the application.

4. Availability

LANGBOX-ARA 3.3 is available on the following operating systems:

- SCO UNIX System V release 3.2 (AT and MCA versions).
- INTERACTIVE 386/IX release 2.0.
- IBM AIX PS/2 model 70 and 80 releases 1.1 and 1.2.
- IBM AIX RISC System/6000 release 3.1.
- BULL BOS.
- INTERGRAPH CLIX.
- CDC's EP/IX.
- Data General's DG/UX.
- Siemens' SINIX.
- Solaris 2.3, 2.4, 2.5.1 and 2.6 on SPARC series.
- Solaris X86 2.6 on PC platform
- Silicon Graphics IRIX 5.3 and 6.2.

LANGBOX-ARA 3.3 is available on the following target machines:

- PC/AT Architecture on the Intel platforms.
- IBM PS/2 model 60, 70 and 80.
- IBM RISC System/6000 Family.
- BULL DPX/2.
- INTERGRAPH INTERPRO series.
- ALTOS 386 series 1000.
- SUN's SPARC series.
- Control Data's Mips series.
- Data General's AViiON.
- Siemens' MX series.
- Silicon Graphics INDY, INDIGO, CHALENGE and ONYX family.

LANGBOX-ARA 3.3 provides full support for the following terminals or compatible:

- VGA console terminal (PC only)
- DEC VT320
- Wyse wy60
- Opus 3n1
- HP hp43

Terminal emulation programs (such as xterm, dtterm, cmdtool, aixterm, winterm…) under the X Window system or other windowing environments, such as CDE, Open Windows, DECwindows, Environment V, etc., are fully supported under LANGBOX-ARA version 3.3 in VT220 emulation.

LANGBOX-ARA 3.3 provides full support for the following printer or compatible printer:

- IBM 4201us (9 pin)
- IBM 4201 (9 pin)
- Epson lq1000 (24 pin)
- Fujitsu dl3400 (24 pin)
- HP deskjet 500+
- HP laserjet (PCL)
- DEC LA75+
- Postscript

Note: This document includes some terms such as UNIX, X Window, Postscript, VT100, VT220, WYSE, Laserjet, Deskjet… that are trademarks or Copyrights from their respective authors.